

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

О.Я. Калюжний

Математичні алгоритми приймання та оброблення сигналів

**Частина II. Основи статистичної теорії
приймання сигналів.**

Навчальний посібник

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського як навчальний
посібник для студентів, які навчаються за спеціальністю 172 «Телекомунікації та
радіотехніка»*

Київ
КПІ ім. Ігоря Сікорського
2021

Рецензенти: Дружинін Володимир Анатольович, професор, доктор технічних наук,
професор Радіофізичного факультету КНУ ім. Т. Шевченко

Семенов Василь Юрійович, доктор фізико-математичних наук,
завідувач відділу математичних алгоритмів НВП «Дельта»

Відповідальний

редактор: Дубровка Федір Федорович, доктор технічних наук, професор
кафедри радіоінженерії

*Гриф надано Методичною радою КПП ім. Ігоря Сікорського (протокол № 4 від 07.04.2022 р.) за
поданням Вченої ради радіотехнічного факультету (протокол № 1/2022 від 31.01. 2022 р.)*

Електронне мережне навчальне видання

Калюжний Олександр Якович

Математичні алгоритми приймання та оброблення сигналів

Частина II. Основи статистичної теорії приймання сигналів.

Навчальний посібник для студентів спеціальності 172 «Телекомунікації та радіотехніка» /
О. Я. Калюжний; КПП ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 10,65
Мбайт). – Київ : КПП ім. Ігоря Сікорського, 2021. – 302 с.

Посібник присвячено розгляду математичних алгоритмів приймання та оброблення сигналів, тобто, послідовності дій над прийнятим сигналом від входу приймача до одержання корисної інформації, яку цей сигнал несе. Посібник складається з трьох частин перша з яких була опублікована у 2017 р. та яка містить загально теоретичний матеріал, що є необхідним для сприйняття наступних, більш спеціалізованих розділів. Друга частина присвячена розгляду статистичної теорії приймання та оброблення сигналів. У цій частині дана загальна постановка завдань статистичного приймання та оброблення сигналів для низки прикладних задач, розглянуті теоретичні засади вирішення таких завдань та приклади їх вирішення. Зокрема, розгляд охоплює завдання виявлення сигналів, вимірювань їх параметрів та демодуляції сигналів із цифровою модуляцією. При цьому наголос зроблено на нерозривній єдності оброблення сигналів не тільки у часі, але й як функції просторових координат по виходу приймальних елементів антенних решіток.

Посібник розраховано на студентів старших курсів, аспірантів та молодих фахівців, які спеціалізуються в області розробки і моделювання алгоритмів приймання та оброблення сигналів фізичного рівня для систем радіозв'язку та телекомунікацій.

Перелік скорочень

АР	авторегресія
АЧХ	амплітудно-частотна характеристика
АКФ	автокореляційна функція
ВКФ	взаємно кореляційна функція (сигналів)
ВФМ	відносна (різницева) фазова модуляція
ДПФ	дискретне перетворення Фур'є
ІВ	імпульсний відгук (лінійної стаціонарної системи)
ІПФ	інтегральне перетворення Фур'є
КАМ	квадратурна амплітудна модуляція (сигналу)
КФР	кумулятивна функція розподілу (ймовірностей)
ЛЧМ	лінійна частотна модуляція (сигналу)
НВП	науково-виробниче підприємство
МСІ	між символна інтерференція
ОП	оперативна пам'ять
ПЛІС	програмована логічна інтегральна схема
ПШ	псевдошумовий (сигнал)
СЩП	спектральна щільність потужності (сигналу)
ТТВ	типовий тракт виявлення (сигналів)
ФМ	фазова модуляція
ФЧХ	фазочастотна характеристика
ЧМ	частотна модуляція
ЦОТ	цифрова обчислювальна техніка
ЦФ	цифрова фільтрація, цифровий фільтр
ЧМБФ	цифрова модуляція з безперервною фазою
ШПФ	швидке перетворення Фур'є

Перелік іншомовних скорочень

ASK	<i>Amplitude Shift Keying</i> - маніпуляція амплітудним зсувом
APSK	<i>Amplitude-Phase Shift Keying</i> - амплітудно-фазова маніпуляція
BPSK	<i>Binary Phase Shift Keying</i> - бінарна фазова маніпуляція
CE	<i>Complex Envelope</i> - комплексна обвідна (сигналу)
CIC	<i>Cascaded Integrator-Comb</i> - каскадний інтегрально-гребінчастий (фільтр)
CDMA	<i>Code Division Multiple Access</i> - кодове розділення при багатосторонньому доступі
DFT	<i>Discrete Fourier Transform</i> - дискретне перетворення Фур'є
IDFT	<i>Inverse Discrete Fourier Transform</i> - зворотне дискретне перетворення Фур'є
IIR	<i>Infinite Impulse Response</i> - нескінченний імпульсний відгук (системи)
ISI	<i>Inter Symbol Interference</i> - між символна інтерференція
FFT	<i>Fast Fourier Transform</i> - швидке перетворення Фур'є
FIR	<i>Finite Impulse Response</i> - скінченний імпульсний відгук (системи)
FM	<i>Frequency Modulation</i> - частотна модуляція
FSK	<i>Frequency Shift Keying</i> - маніпуляція частотним зсувом
LTI	<i>Linear Time Invariant</i> - лінійна стаціонарна (система)
MSK	<i>Minimum Shift Keying</i> - маніпуляція з мінімальним частотним зсувом
O-BPSK	<i>Offset Binary Phase Shift Keying</i> - офсетна бінарна фазова маніпуляція
OFDM	<i>Orthogonal Frequency Division Multiplexing</i> - мультиплексування з ортогональним частотним поділом
PM	<i>Phase Modulation</i> - фазова модуляція
PSK	<i>Phase Shift Keying</i> - фазова маніпуляція
QAM	<i>Quadrature Amplitude Modulation</i> - квадратурна амплітудна модуляція
QPSK	<i>Quadrature Phase Shift Keying</i> - квадратурна фазова маніпуляція

Зміст частини II

	Стор.
Передмова	8
Розділ 9. Завдання статистичної теорії приймання сигналів	14
9.1. Формалізація завдань приймання сигналів.	14
9.2. Априорні характеристики шумів та завад	22
9.3. Опис полів сигналів	27
9.4. «Багатопроменева» модель поля сигналу	31
<i>Контрольні запитання та завдання до розділу 9.</i>	34
Розділ 10. Частина I. Теорія перевірки гіпотез	36
10.1. Постановка задачі перевірки гіпотез	36
10.2. Оптимальні правила статистичного рішення.....	39
10.3. Перевірка двох гіпотез	44
<i>Контрольні запитання та завдання до розділу 10, частина I.</i>	51
Розділ 10. Частина II. Теорія статистичного оцінювання.....	52
10.4. Оптимальне статистичне оцінювання.....	52
10.5. Похибки статистичного оцінювання.....	60
<i>Контрольні запитання та завдання до розділу 10, частина II</i>	69
Розділ 11. Загальна структура функціоналу відношення правдоподібності поля сигналу.....	70
11.1. Статистична постановка завдань приймання полів сигналів.....	70
11.2. Приймання полів наперед відомих (детермінованих) сигналів	73
11.3. Приймання полів квазидетермінованих сигналів.....	79
11.4. Приймання полів шумових сигналів.	89
<i>Контрольні запитання та завдання до розділу 11</i>	92
Розділ 12. Кореляційні характеристики випадкових полів.....	94
12.1. Загальні відомості	94
12.2. Поле випадкового локального джерела	98
12.3. Розклади кореляційної функції випадкових полів	102
<i>Контрольні запитання та завдання до розділу 12</i>	110

Розділ 13. Оптимальні та квазіоптимальні алгоритми лінійного оброблення сигналів у часі та просторі	111
13.1. Приймання квазідетермінованих сигналів на тлі білого шуму.....	111
13.2. Приймання квазідетермінованих сигналів на тлі корельованих завад	117
13.3. Приймання шумових сигналів.	125
13.4. Приймання сигналів на тлі локальних завад	129
<i>Контрольні запитання та завдання до розділу 13</i>	136
Розділ 14. Підсилення просторової обробки сигналів	138
14.1. Загальні відомості	138
14.2. Підсилення просторової обробки на тлі ізотропних завад.	148
14.3. Підсилення просторової обробки на тлі локальних завад.	152
<i>Контрольні запитання та завдання до розділу 14</i>	157
Розділ 15. Ефективність виявлення сигналів	158
15.1. Загальні відомості.	158
15.2. Виявлення точно відомих сигналів....	160
15.3. Виявлення сигналів із невідомими параметрами	167
15.4. Виявлення шумових сигналів	178
15.5. Приклади розрахунків ефективності виявлення сигналів	184
<i>Контрольні запитання та завдання до розділу 15</i>	190
Розділ 16. Вимірювання параметрів сигналів	192
16.1. Загальні положення	192
16.2. Методи вимірювань	197
16.3. Сигнальна функція обробки та функція невизначеності	204
16.4. Аналіз похибок вимірювань	211
16.5. Оптимальні алгоритми вимірювання параметрів детермінованих сигналів	216
16.6. Висновки	226
<i>Контрольні запитання та завдання до розділу 16</i>	227

Розділ 17. Базові принципи цифрової демодуляції сигналів	229
17.1. Демодуляція сигналів як завдання розрізнення статистичних гіпотез.....	229
17.2. Проекційна (координатна) технологія цифрової демодуляції сигналів ...	234
17.3. Когерентне приймання повідомлень «у цілому», послідовна та посимвольна демодуляція	242
17.4. Демодуляція за алгоритмом Вітербі	248
17.5. Некогерентна демодуляція	252
<i>Контрольні запитання та завдання до розділу 17</i>	265
Розділ 18. Аналіз ефективності демодуляції сигналів в обчислювальному середовищі Matlab	267
18.1. Теоретичний аналіз ефективності за допомогою інструменту BERTool ..	267
18.2. Моделювання систем передачі інформації у командному режимі	274
18.3. Моделювання систем передачі інформації у середовищі Simulink	286
<i>Контрольні запитання та завдання до розділу 18</i>	299
Література до частини II	301

Передмова

Однією з найбільш характерних рис сьогодення є надзвичайно стрімке та масштабне впровадження новітніх телекомунікаційних технологій. Важко навіть назвати ті сфери життя, яких би це впровадження не торкалося: економіка, промисловість, наука, культура, побут, тощо. У багатьох джерелах [] цей процес вже навіть називають не інакше, як науково-технічною революцією XXI сторіччя. Звісно, досягнутий масштаб змін був б неможливий, якби не спирався на вже накопичений науково-технічний потенціал, що створювався на протязі десятиріч зусиллями багатьох фахівців у різноманітних галузях науки та технологій. Також безперечним є внесок закладів освіти, які зуміли підготувати відповідні за кількістю та кваліфікаційним рівнем науково-інженерні кадри.

Цілком очевидно, що вітчизняна наука та освіта, не дивлячись на усі ті проблеми, які ми переживаємо на протязі останніх десятиріч, не мають залишатися осторонь цього глобального процесу. Звичайно, ми поки що не можемо конкурувати зі світовими лідерами, але той величезний потенціал, який був свого часу створено зусиллями провідних вітчизняних фахівців, дає нам можливість не залишатися сторонніми спостерігачами процесів наукових та технологічних перетворень. Треба лише творчо переосмислювати досягнення минулих поколінь та наповняти їх сучасним змістом. Саме спробою такого переосмислення і є друга частина Посібника *«Математичні алгоритми приймання та оброблення сигналів»*, яка присвячена розгляду базових положень статистичної теорії приймання сигналів (СТПС). Нагадаємо, що приймання сигналів – це сукупність методів та процедур DSP, які забезпечують виділення корисної інформації із сигналів на вході деякого приймального пристрою. Зміст корисної інформації, її призначення та фізичні умови роботи приймальних систем дуже різняться поміж собою. Тому й фізичні постановки завдань прийому теж дуже різні. Але їх загальною рисою є те, що майже усі сигнали, з якими доводиться мати справу в практичних застосуваннях, є випадковими. Ця обставина докладно розглядалася в першій частині Посібника [1], де, зокрема, відзначалося, що випадковість сигналів зумовлена, по-перше,

випадковістю інформації, яку вони несуть, по-друге, впливом на сигнал фізичного середовища передачі, по-третє, технічними особливостями приймально-передавальних пристроїв. Там же було проілюстровано опис випадкових сигналів на основі апарату теорії ймовірностей та теорії випадкових процесів.

Отже, сигнали на вході приймального пристрою безвідносно до їхньої фізичної природи можна розглядати як деякий набір статистичних даних, а завдання прийому – як оброблення таких даних на основі відповідних математичних процедур. Це і є головною ідеєю СТПС, розгляду основних положень якої й присвячено цю частину Посібника. Дана теорія має універсальний характер, а сфера її застосування охоплює практично усі галузі техніки, в яких можна говорити про існування сигналів – носіїв тієї чи іншої корисної інформації. До цих галузей, зокрема, відносяться завдання зв'язку у самому широкому сенсі, радіолокаційного, гідроакустичного, оптичного спостереження, технічної або медичної діагностики та багато іншого. Важливою особливістю даної частини Посібника є те, що розгляд завдань оброблення сигналів проводиться у нерозривній єдності властивостей сигналів як функцій не тільки часу, але й просторових координат. Тобто, якщо у традиційній теорії сигналом звичайно вважали вихід деякої антени, то, відповідно до сучасних поглядів, сигналом є деяке поле (електромагнітне, акустичне, оптичне, тощо), що спостерігається у заданій області простору. Саме такий підхід був характерною ознакою відомої у 70 – 80 роки «Київської наукової школи статистичної радіотехніки», яку складали фахівці декількох провідних кафедр НТУУ «КПІ», серед яких чільне місце займали й кафедри радіотехнічного факультету. Отже, даний Посібник витримано в дусі традицій вказаної наукової школи, до послідовників якої автор відносить й себе.

Тепер розглянемо побудову даної частини Посібника, яка загалом містить 10 розділів. Розділи мають наскрізну нумерацію по всім частинам Посібника, тобто, якщо останній розділ першої частини мав номер 8, то перший розділ другої частини починається з номера 9 і т.д. Отже, для посилань на попередні частини Посібника достатньо вказати лише номери відповідних розділів. Зокрема, першій частині

відповідають розділи с 1 до 8. Тому нагадування про номер частини Посібника надалі будуть опускатися.

Перші розділи другої частини (тобто, розділи 9 та 10) містять матеріал загального характеру, зокрема, постановку та класифікацію завдань СТПС, методи опису апіорних даних, які стосуються характеристик полів шумів, сигналів та завад. Також розглянуто деякі базові поняття математичної статистики, що необхідні для розуміння наступних розділів.

Розділ 11 є теоретичною основою теорії СТПС, на якій базуються майже усі наступні розділи. Тут розглядаються загальні співвідношення для функціоналів правдоподібності полів сигналів, що спостерігаються у заданій області простору. З методичних міркувань цей розгляд «прив'язано» до завдань виявлення сигналів, які у загальній проблематиці СТПС посідають одне з чільних місць. Але основні теоретичні співвідношення, що тут наведені стосовно завдань виявлення, далі застосовуються й при розгляді усіх інших завдань, яким присвячено наступні, більш спеціалізовані розділи Посібника. Вказані співвідношення конкретизовано для типових класів завдань СТПС, а саме, приймання наперед відомих (детермінованих) сигналів, приймання квазидетермінованих (тобто, з неповністю відомими параметрами) сигналів та приймання випадкових шумових сигналів.

Одним з важливих результатів розділу 11 є висновок про те, що загальна структура функціоналів правдоподібності прийнятих сигналів багато у чому визначається їх просторово-часовими кореляційними характеристиками. Тому розгляду цих характеристик присвячено наступний розділ 12. Тут наведено як загальна постановка завдань обчислення вказаних характеристик, так й їх деталізація для деяких типових умов приймання сигналів. Зокрема, розглянуто поняття статистично однорідних та неоднорідних полів, кутові спектри поля, тощо. Також розглядаються розклади просторово-часових кореляційних характеристик випадкових полів по сукупності плоских хвиль як для умов однорідного безмежного простору, так й для хвилеводів. Наведено приклади розрахунків кореляційних

характеристик, що мають теоретичне та практичне застосування та застосовуються у наступних розділах.

У розділі 13 на основі результатів розділів 11 та 12 проведено деталізацію загальних співвідношень для функціоналів правдоподібності полів на просторовій апертурі приймальної системи. Розглянуто умови так званого квазістаціонарного наближення для компактних просторових апертур, які дають можливість розділення загального алгоритму приймання сигналів на просторову та часову частини. Ця можливість проілюстрована прикладами вирішення деяких типових завдань за тематикою Посібника.

Розділ 14 ілюструє ефективність просторової частини обробки на основі показника її підсилення. Поняття підсилення ілюструється прикладами розрахунків для типових завдань приймання сигналів на тлі ізотропних та анізотропних у просторі завад.

Розділ 15 присвячено методам аналізу ефективності виявлення сигналів для типових моделей каналів передачі сигналу, а саме, каналу із випадковою фазою сигналу, релеєвського та райсівського каналів. Вказані методи ілюструються чисельними розрахунками основних показників виявлення сигналів, зокрема, робочих характеристик приймача, характеристик виявлення, порогових відношень сигнал/завада.

У розділі 16 розглядаються завдання вимірювання параметрів сигналів, які становлять важливу частину СТПС. Теоретичною основою таких вимірювань є теорія статистичного оцінювання, що розглядалися у розділі 10. У даному розділі розглядається наповнення методів статистичного оцінювання фізичним змістом та проводиться конкретизація алгоритмів вимірювання найважливіших параметрів. Детально розглянуті основні методи вимірювань та їх характеристики: сигнальні функції обробки, функції невизначеності, статистичні похибки вимірювань. Усі ці характеристики ілюструються прикладами розрахунків для типових практичних завдань.

Останні розділи 17 та 18 присвячено розгляду завдань демодуляції сигналів зв'язку та аналізу їх ефективності. У розділі 17 демодуляція сигналів розглянута як завдання розрізнення статистичних гіпотез, до вирішення якого можуть бути застосовані загально теоретичні методи, що були наведені у попередніх розділах. Наступний матеріал розділу 17 базується на застосуванні так званої проекційної технології демодуляції, яка використовує представлення прийнятих сигналів коефіцієнтами розкладу по координатним функціям сигнального простору. Сукупність таких коефіцієнтів утворює множину так званих достатніх статистик, що суттєво спрощує як теоретичні, так й практичні аспекти побудови підсистем демодуляції. Потім розглядається низка когерентних та некогерентних алгоритмів демодуляції. Також проілюстрована можливість вирішення завдань демодуляції за допомогою відомого алгоритму Вітербі, який свого часу був розроблений як потужний засіб декодування згорткових завадостійких кодів.

Розділ 18 присвячено розкладу завдань аналізу ефективності підсистем цифрової демодуляції та питань синхронізації цифрових приймачів. Наведені загально теоретичні співвідношення, які необхідні для проведення такого аналізу. Потім застосування вказаних співвідношень проілюстровано прикладами завдань аналізу ефективності демодуляції засобами середовища Matlab, зокрема, прикладного пакету VerTool. Розглянуто основні методи синхронізації цифрових приймальних пристроїв, зокрема, синхронізації носійної частоти та символної синхронізації. Розглянута синхронізація як за допомогою спеціальних службових (пілот) сигналів, так й безпосередньо на основі інформаційних символів. Наведені приклади вирішення завдань синхронізації цифрових приймачів для сучасних телекомунікаційних протоколів фізичного рівня.

Тепер про цільову аудиторію наших читачів: це магістранти, аспіранти, молоді науковці. Сподіваюсь, що деякі розділи будуть корисними й для більш досвідчених фахівців. Також, як й перша частина, друга частина Посібника «не прив'язана» до певної дисципліни чи навчального плану. Більше того, матеріал посібника відноситься до цілої групи дисциплін, об'єднаних загальним напрямком –

приймання та оброблення сигналів. Тому його використання в учбовому процесі може бути найбільш ефективним як допоміжного засобу при вивченні окремих розділів певних дисциплін або як матеріал для «швидкого» повторення деяких тем. З огляду на вказані обставини Посібник розраховано переважно на самостійну роботу студентів. Кожний розділ супроводжується відповідними контрольними завданнями, які слід ретельно виконувати у письмовій формі. По завершенню вивчення одного чи кількох розділів рекомендується періодично проводити семінарські заняття під керівництвом викладача або об'єднавшись зі своїми колегами по роботі.

Отже, бажаю усім своїм читачам успіхів у вивченні статистичної теорії приймання сигналів! Свої зауваження та пропозиції Ви можете надсилати мені через сторінку «*Контакти*» мого Web-сайту за посиланням:

<http://oleksa-site.blogspot.com/p/contacts.html>

З повагою, проф. Калюжний О.Я.

Розділ 9. Завдання статистичної теорії приймання сигналів.

У цьому розділі ми розглянемо статистичну постановку завдань приймання сигналів та вихідні дані, які є необхідними для їх вирішення.

9.1. Формалізація завдань приймання сигналів.

На фізичному рівні завдання приймання сигналів можуть бути представлені загальною схемою рис.9.1, до якої входять наступні складові:

- джерело сигналу, який несе корисну інформацію;
- джерела завад, які перешкоджають роботі приймальної системи;
- внутрішні шуми приймальної системи;
- приймальна система, яка складається з антени, підсистеми пристроїв аналогового Front-end та системи цифрового оброблення сигналів (DSP).

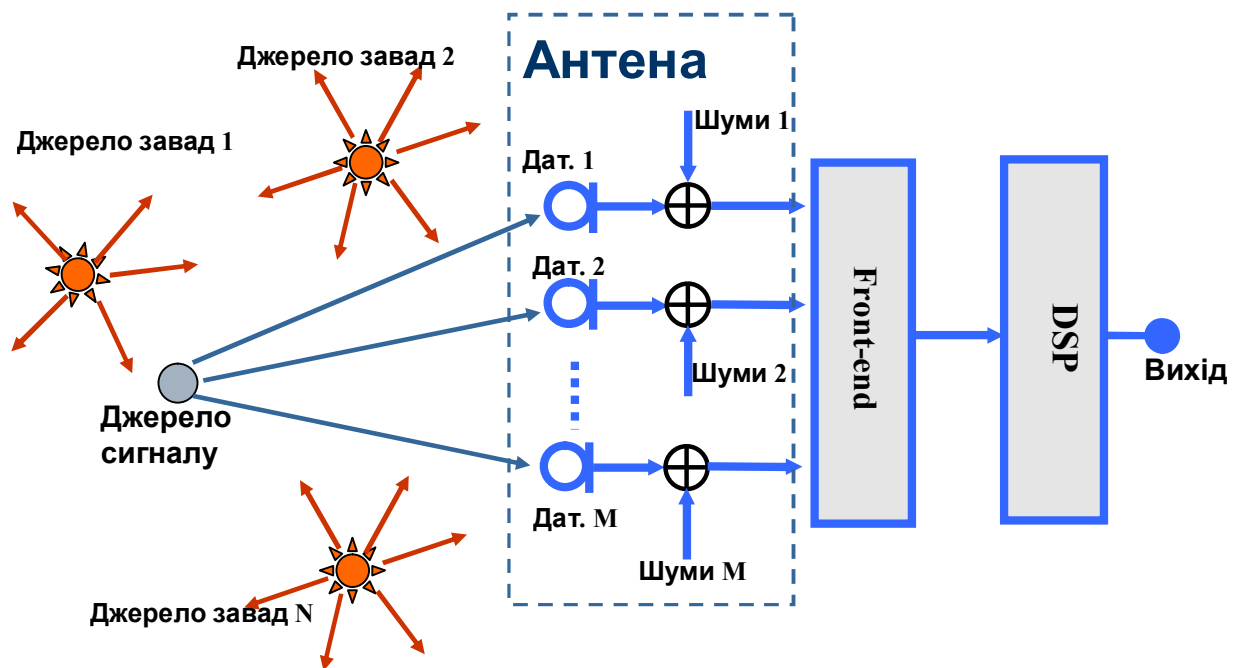


Рис.9.1. Фізична постановка завдань приймання сигналів.

Вважаємо, що антену формує сукупність¹ деяких датчиків, які перетворюють фізичні параметри поля сигналу в електричні сигнали. Такими фізичними параметрами, у залежності від типу приймальної системи, можуть бути напруженість

¹ Такі дискретні антени прийнято називати антенними решітками (АР).

електромагнітного поля, звуковий тиск акустичного поля, рівень оптичної освітленості, тощо.

Підсистема аналогового Front-end звичайно складається з різного роду пристроїв для узгодження та перетворення аналогових сигналів по виходу датчиків антени у вигляд, придатний для наступного цифрового оброблення. До її складу звичайно входять фільтри, підсилювачі, перетворювачі частоти, комутатори й т.п. Завершальною операцією Front-end є перетворення аналогових сигналів у масиви дискретних цифрових даних, які надалі й підлягають обробленню в системі DSP.

Важливими елементами цієї схеми є завади та внутрішні шуми. Завадами звичайно називають сигнали деяких сторонніх джерел різноманітного фізичного походження. Це можуть бути сигнали від схожих систем, що працюють поблизу, сигнали індустриальних та/або транспортних об'єктів, сигнали, що виникають унаслідок природних явищ (наприклад, грози), тощо. Усі ці сторонні сигнали поєднує та обставина, що вони за своєю суттю є фізично спорідненими з корисними сигналами, тобто, генеруються та поширюються в тому ж самому середовищі та за тими ж самими фізичними законами.

Внутрішні шуми – це завади особливого роду, які породжуються самою приймальною системою. Зокрема, до них відносяться власні шуми датчиків антени, пристроїв підсилення та перетворення сигналів, тощо. Також до їхнього складу часто включають і шуми, які породжуються взаємодією системи із зовнішнім середовищем (деякі приклади наведемо пізніше). Тому, хоча на схемі рис.9.1 додавання внутрішніх шумів умовно показано по виходу датчиків антени, це не слід сприймати буквально. У той же час, поділ факторів, які заважають роботі приймальних систем, на завади та внутрішні шуми має для побудови статистичної теорії приймання сигналів принципове значення, про що буде сказано далі.

Дана схема відображає основні складові завдань приймання сигналів, яких достатньо для розуміння статистичної теорії, що розглядається. Але це не означає, що фізичний зміст таких завдань не є важливим. Навпаки, успіх у вирішенні завдань приймання сигналів напряду залежить від того, наскільки повно враховані всі фізичні

«обставини» практичної роботи. Деякі ілюстрації щодо цього будуть зустрічатися у наступних розділах. Поки ж розглянемо приклади деяких систем, до яких ми будемо надалі звертатися.

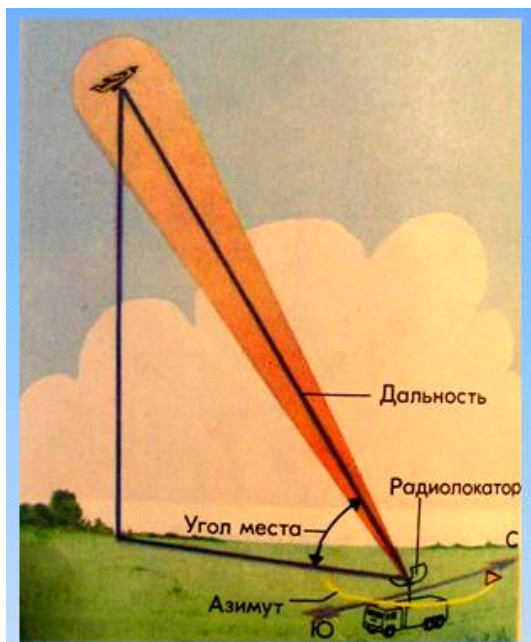


Рис.9.2. Система активної радіолокації (радар), що створює при взаємодії з об'єктом локації відбитий електромагнітний сигнал, який повертається до приймальної антени. Корисними параметрами прийнятого сигналу є його затримка в часі відносно моменту випромінювання зондувального сигналу, напрямок приходу, потужність, частота прийнятого сигналу та інше. Завданнями приймальної системи є, по-перше, встановлення самого факту існування відбитого сигналу, по-друге, вимірювання його корисних параметрів. Завадами для таких систем можуть бути сигнали інших радарів, штучно створені завади протидії, індустриальні та транспортні завади, атмосферні завади, тощо.

2. Системи пасивної радіолокації.

Відмінністю цього випадку від попереднього є те, що корисний сигнал створюється не за рахунок відбиття зовнішнього зондувального сигналу, а за рахунок власного електромагнітного випромінювання об'єктів локації, яке може створювати їх обладнання (двигуни, радіоелектронне обладнання, тощо).

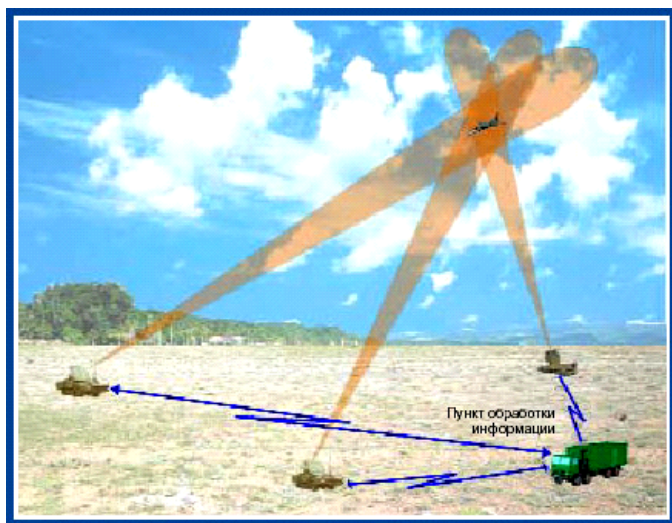


Рис.9.3. Система пасивної радіолокації

Перелік корисних параметрів сигналу в цілому схожий на попередній, але відсутній такий важливий параметр, як затримка сигналу. Відсутність цього параметра може бути компенсована за рахунок спільного оброблення сигналів з виходів кількох приймальних антен, рознесених у просторі, як показано на рис.9.3.

3. Системи радіонавігації (рис.9.4). Принцип дії таких систем зводиться до вимірювання тим чи іншим способом запізнення сигналів радіомаяків, розташувannya

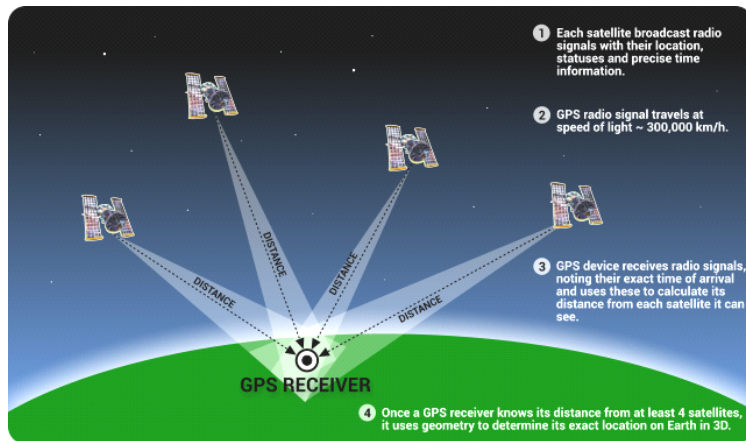


Рис.9.4. Система супутникової радіонавігації GPS

яких наперед відоме. Потім за результатами вимірювань запізнень обчислюється географічне положення точки приймання сигналів. До появи супутникових технологій радіомаяки розташовувались виключно на земній поверхні, у наш час найчастіше використовуються супутникові системи, як, наприклад, GPS, Glonass,

Galileo, Бэйдоу та інші.

4. Гідроакустичні системи (сонари) (рис.9.5). Такі системи за принципом дії подібні до радіолокаційних, але працюють у водному середовищі та використовують для локації акустичні хвилі.

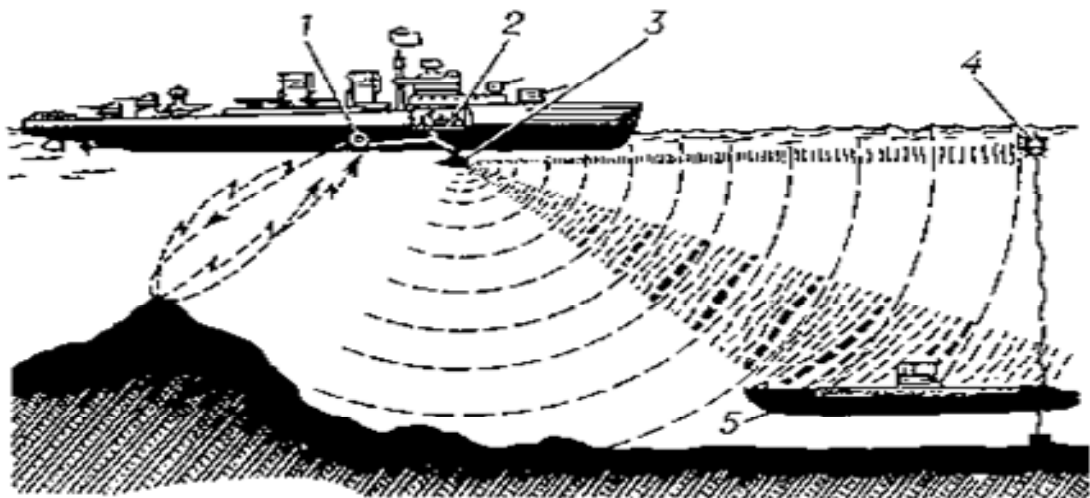


Рис.9.5. Гідроакустичні системи надводного корабля

Але слід зауважити, що фізичні умови роботи гідроакустичних систем часто набагато складніші, чим для відповідних радіолокаційних систем. Зокрема, у водних середовищах поширеним явищем є рефракція звуку, яка приводить до викривлення променевих траєкторій сигналу. У той же час, для переважної більшості радіолокаційних систем рефракція не відіграє суттєвої ролі². Також велике значення для гідроакустичних систем мають ефекти взаємодії звукових хвиль із границями середовища (дно та поверхня), а водне середовище незрівнянно більше насичене джерелами завад. Значно більшу роль у гідроакустиці, порівняно з радіолокацією, відіграє ефект Допплера, який зумовлений рухом об'єктів локації та/або самої приймальної системи. Внутрішні шуми також складніші: окрім власних шумів датчиків антени та пристроїв Front-end до їхнього складу звичайно включають й шуми гідродинамічного обтікання поверхні гідроакустичної антени, які виникають внаслідок руху носія, зокрема, судна. Це, до речі, є одним з прикладів шумів, які породжені взаємодією приймальної системи із зовнішнім середовищем.

5. Системи телекомунікацій (рис.9.6). Під цією лаконічною назвою насправді об'єднують силу-силенну всіляких систем передачі та прийому інформації, як,

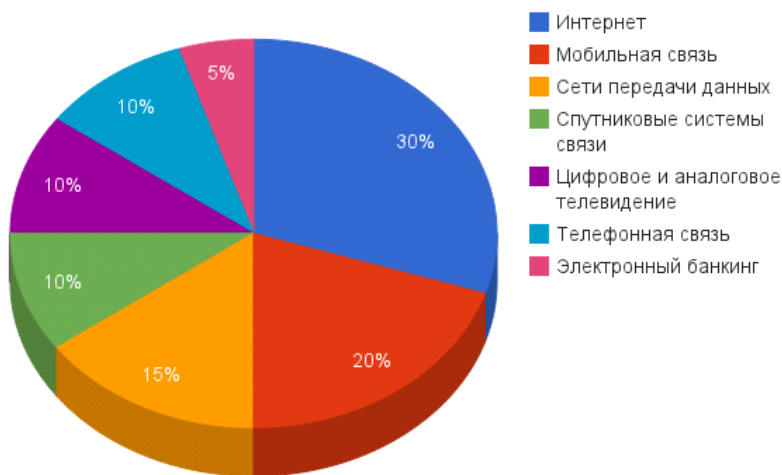


Рис.9.6. Системи телекомунікацій

зокрема, мережа Інтернет, системи мобільного зв'язку, супутникові системи передачі даних, телебачення, телефонний зв'язок, тощо. У наш час системи телекомунікацій стали окремою та великою галуззю економіки, що вже давно набула системоформуючих функцій. Дійсно, неможливо собі уявити сучасне

² Ми не розглядаємо спеціальні радари, як, наприклад, системи наддалекого виявлення для яких рефракційні ефекти вже є суттєвими.

життя без мобільного зв'язку, Інтернету та багатьох інших систем передачі інформації. За своїм масштабом та складністю системи телекомунікацій набагато перевершують усе те, що вже було згадано вище.



Рис.9.7. Апарат УЗД

6. Системи медичної або технічної діагностики, які забезпечують визначення стану об'єктів діагностики (виявлення захворювань, пошкоджень технічного стану, тощо). Цей клас систем виявляється дуже схожим на системи локації, про які йшла мова вище. У цьому, наприклад, мав змогу впевнитися кожен, кому доводилося проходити ультразвукове дослідження (рис.9.7). Але є й багато інших та значно складніших систем, як, скажімо, комп'ютерна чи магнітно-резонансна томографія, тощо.

То чи може бути щось спільне поміж усіма цими системами, які суттєво різняться як за призначенням, так за своїм технічним виконанням? Виявляється, що так, причому, такою спільною основою є методи й алгоритми приймання сигналів. Дійсно, у будь-якій з цих систем незалежно від її призначення необхідно перш за все вирішити завдання надійного приймання сигналів на тлі шумів та завад. Від того, наскільки успішним буде вирішення цього завдання, настільки й успішною буде робота будь-якої з цих систем у цілому. У свою чергу, теоретичною основою вирішення завдань приймання сигналів незалежно від їхнього фізичного змісту є СТПС, вивченню основ якої й присвячено дану частину нашого Посібника.

Тепер перейдемо до математичного формулювання розглянутої вище постановки завдань приймання сигналів. Відправною точкою для цього є так зване **рівняння спостережень**, що пов'язує сигнали по виходу датчиків антени з параметрами фізичних полів, які існують у середовищі передачі сигналу. Обмежимося при цьому розгляді практично важливим випадком скалярних спостережень. Зокрема, стосовно електромагнітних полів (ЕМП) це означає, що на вхід системи приймання надходить лише один поляризаційний вихід датчиків ЕМП (елементів АР), для акустичних полів – лише вихід датчиків акустичного тиску, тощо. Що ж стосується оброблення

векторних спостережень, то за відповідною інформацією слід звертатися до наукової літератури окремих галузей знань, наприклад, монографії [27].

Отже, для схеми рис.9.1 з урахуванням усього, що було сказано вище, скалярне рівняння спостережень може бути записано у вигляді:

$$u(t, \mathbf{r}) = s(t, \mathbf{r} / \boldsymbol{\theta}) + z(t, \mathbf{r}) + w(t, \mathbf{r}), \quad t \in (t_0, t_0 + T), \quad \mathbf{r} \in \mathbf{D}_A, \quad (9.1)$$

де $u(t, \mathbf{r})$ - поле сигналу по виходу деякого датчика³ антени, що розміщений у точці простору з координатами⁴ $\mathbf{r}$. Вказаний сигнал задано на часовому інтервалі $t \in (t_0, t_0 + T)$ тривалістю T , а датчики антени охоплюють деяку область середовища $\mathbf{r} \in \mathbf{D}_A$. Як бачимо, спостереження (9.1) складаються з трьох компонентів:

- поля корисного сигналу $s(t, \mathbf{r} / \boldsymbol{\theta})$, яке, у свою чергу, залежить від вектора деяких інформаційних параметрів $\boldsymbol{\theta}$;
- поля зовнішніх завад $z(t, \mathbf{r})$;
- поля внутрішніх шумів $w(t, \mathbf{r})$.

Рівняння спостережень (9.1) може бути застосовано для більшості завдань статистичної теорії приймання сигналів. Різниця полягає, головним чином, у складі та властивостях вектора інформаційних параметрів $\boldsymbol{\theta}$. Розглянемо декілька важливих випадків, про які піде мова далі.

1. **Завдання виявлення сигналів** полягає у встановленні лише факту наявності або відсутності корисного сигналу на заданому часовому інтервалі $t \in (t_0, t_0 + T)$ та в заданій області простору $\mathbf{r} \in \mathbf{D}_A$. Це типове завдання для всього роду локаційних систем. Аналогічне завдання може виникати також й у системах телекомунікацій на етапі встановлення з'єднання. Рівняння спостережень (9.1) у всіх цих випадках може бути представлено в наступному вигляді:

$$u(t, \mathbf{r}) = \theta \cdot s(t, \mathbf{r}) + z(t, \mathbf{r}) + w(t, \mathbf{r}), \quad t \in (t_0, t_0 + T), \quad \mathbf{r} \in \mathbf{D}_A, \quad (9.2)$$

³ Тут ми не торкаємося властивостей датчиків та їх структури, тобто, це можуть бути доволі складні пристрої. Але до подробиць будемо вдаватися лише там, де це доречно, та лише в міру необхідності.

⁴ Системи координат можуть вибиратися довільним чином, але найчастіше початок координат відповідає деякому умовному центру приймальної антени.

де θ - формальний параметр ситуації, який дорівнює нулю ($\theta = 0$) при відсутності сигналу та одиниці ($\theta = 1$) при його наявності.

2. **Завдання розрізнення сигналів**, яке треба вирішувати, наприклад, у системах телекомунікацій, коли з'єднання вже встановлено. Тоді факт наявності корисного сигналу достовірно відомий, але невідомо, який саме інформаційний символ на даному часовому інтервалі передається. У найпростішому випадку незалежного приймання символів рівняння спостережень (9.1) на поточному символьному інтервалі $t \in (t_0, t_0 + T)$ може бути представлено в наступному вигляді:

$$u(t, \mathbf{r}) = s_{\theta}(t, \mathbf{r} / \boldsymbol{\theta}) + z(t, \mathbf{r}) + w(t, \mathbf{r}), \quad t \in (t_0, t_0 + T), \quad \mathbf{r} \in \mathbf{D}_A, \quad (9.3)$$

де індекс поточного символу θ приймає значення з набору $\{1, 2, \dots, M\}$, M - загальний об'єм алфавіту символів. Зауважимо, що завдання розрізнення сигналів також виникає в деяких локаційних застосуваннях, коли необхідно визначити тип об'єкту, що спостерігається.

3. **Завдання оцінювання параметрів сигналів**, яке вирішується в багатьох приймальних системах та полягає у вимірюванні деяких інформаційних параметрів сигналів. Зокрема, у локаційних системах після того, як факт наявності сигналу вже встановлено, треба, як правило, визначити його затримку в часі, зсув по частоті відносно зондувального сигналу, напрямок приходу, тощо. У телекомунікаційних системах вимірювання параметрів сигналу найчастіше використовується з метою забезпечення різних видів синхронізації (частотної, символьної, пакетної, тощо), без якої будь-яка телекомунікаційна система працювати не може. Принципова відмінність завдання оцінювання від завдання розрізнення сигналів, полягає в тому, що в цьому випадку інформаційні параметри можуть приймати нескінченний (континуальний) набір значень на деякій множині $\boldsymbol{\theta} \in \boldsymbol{\Omega}$.

4. **Завдання фільтрації.** У цьому випадку інформаційні параметри є функціями деяких фізичних змінних, а завдання полягає у відновленні вигляду цих функцій. До даного класу відносяться, наприклад, завдання виділення сигналів радіо або телевізійного мовлення. Методи фільтрації також використовуються в локаційних системах для забезпечення стеження за об'єктами локації та в багатьох інших системах.

Для формування повної математичної постановки завдання приймання сигналів рівняння спостережень (9.1) має бути доповнено *апріорною інформацією*, тобто, описом характеристик його окремих складових, які є наперед відомими. Тож розглянемо деякі типові підходи до формування такого опису. Попередньо зауважимо, що кожна зі складових правої частини виразу (9.1) є випадковою функцією часу та просторових координат. Такі функції у фізиці прийнято називати випадковими полями. Отже, мова піде про статистичні характеристики випадкових полів. Окремим випадком є одно каналне приймання, коли випадкові поля переходять у випадкові процеси, про що вже йшла мова в розділі 4.

9.2. *Апріорні характеристики шумів та завад.*

Почнемо наш апріорний опис із характеристик полів шумів та завад. Перше питання, яке при цьому виникає: чи можливий такий опис з єдиних позицій, без прив'язки до конкретних технічних деталей? Виявляється, що так, можливий і ця можливість спирається на фундаментальні фізичні та ймовірнісні закони, про які піде мова нижче. При цьому внутрішні шуми та зовнішні завади будемо розглядати окремо, оскільки дані компоненти спостережень на фізичному рівні принципово різні.

Внутрішні шуми, як уже було сказано вище, генеруються, головним чином, датчиками антени та пристроями Front-end. Основною фізичною причиною існування таких шумів є тепловий рух електронів. Це явище було докладно вивчене фізичною наукою ще в 20-ті роки минулого сторіччя. Спершу зауважимо, що характеристики теплового шуму суттєво залежать від умов навколишнього середовища, зокрема, температури. Отже, якщо ці умови стабільні в часі, то й характеристики шуму є практично незмінними. Вказана обставина дозволяє прийняти гіпотезу щодо

квазістаціонарності внутрішніх шумів на інтервалі спостереження. Також зазначимо, що в діапазоні частот до kT/h (Гц) тепловий шум⁵ має практично рівномірну спектральну щільність потужності (СЩП). Неважко бачити, що для кімнатної температури $T = 300\text{ K}$ цей діапазон складає приблизно $6 \cdot 10^{12}$ Гц, що, як правило, перекриває робочі діапазони більшості сучасних радіоелектронних систем. Отже, щодо внутрішніх шумів можна з високим ступенем достовірності прийняти модель стаціонарного білого гавсівського шуму, яка вже розглядалася в розділі 4. Характерною властивістю такого шуму є некорельованість його миттєвих значень у часі, що математично формулюється у вигляді наступного виразу для його автокореляційної функції (АКФ):

$$K_w(\tau) = \langle w(t, \mathbf{r}) \cdot w(t + \tau, \mathbf{r}) \rangle = \frac{G_0}{2} \delta(\tau).$$

Тут G_0 - СЩП шуму, $\delta(\tau)$ - символічна дельта-функція, яка визначається співвідношенням (4.26). Отже, якщо обмежитися лише одним каналом приймання, то все, що треба знати про внутрішній шум – це значення константи G_0 . У свою чергу, це значення залежить від технічних характеристик аналогової частини приймального тракту та визначається окремо для кожного конкретного випадку вже з урахуванням технічних умов відповідного завдання.

Але у випадку багатоканальних приймальних систем додатково треба уточнити властивості шумів як функції просторових координат. У свою чергу, ці властивості визначаються переважно технічними характеристиками каналів антени та пристроїв Front-end. При належному інженерному проектуванні внутрішні шуми по виходах каналів можна вважати майже ідентичними за своїми характеристиками, а кореляція шумів поміж каналами практично відсутня. Отже, на підставі наведених вище міркувань приходимо до наступної статистичної моделі внутрішніх шумів $w(t, \mathbf{r})$:

⁵ Тут $k = 1,38064852 \cdot 10^{-23}$ - постійна Больцмана, $h = 6,626070040 \cdot 10^{-34}$ - постійна Планка, T - абсолютна температура провідника.

- шум $w(t, \mathbf{r})$ є стаціонарним у часі та однорідним у просторі гавсівським випадковим полем;
- АКФ шуму може бути записана у вигляді

$$K(t_i, t_k; \mathbf{r}_m, \mathbf{r}_l) = \langle w(t_i, \mathbf{r}_m) w(t_k, \mathbf{r}_l) \rangle = \frac{G_0}{2} \cdot \delta_{l,m} \cdot \delta(t_k - t_i), \quad (9.4)$$

де i, k - індекси відліків спостережень у часі, m, l - номери каналів приймального тракту, $\delta_{k,n}$ - символ Кронекера. Вказані припущення визначають будь-які статистичні характеристики поля внутрішніх шумів без прив'язки, як бачимо, до технічної та фізичної суті завдання. Єдиний невідомий параметр, що потребує визначення в конкретних випадках – це СЩП шумів G_0 . Але, як побачимо далі, багато загальних висновків можна зробити й без конкретизації цього параметра, тобто, лише на основі структури виразу (9.4).

Тепер обговоримо деякі підходи до формування статистичної моделі зовнішніх завад $z(t, \mathbf{r})$. На відміну від внутрішніх шумів тут без фізичних міркувань вже не обійтись, оскільки, як було сказано вище, зовнішні завади формуються в тому самому середовищі, що й корисний сигнал. Отже, діапазон припущень щодо їхніх характеристик майже настільки ж широкий, наскільки різноманітні фізичні умови та технічні особливості конкретних задач приймання сигналів. Тим не менш, і в цьому випадку є декілька достатньо загальних моделей, які мають дуже широке практичне застосування.

Головний з таких загальних підходів – це припущення про гавсівський (нормальний) розподіл миттєвих значень поля завад $z(t, \mathbf{r})$. Дане припущення ґрунтується на одному з найбільш фундаментальних результатів теорії ймовірностей, а саме, центральній граничній теоремі. Якщо не вдаватися до математичних нюансів, то суть теореми доволі проста. Вона полягає в тому, що розподіл випадкового сигналу, який формується як суперпозиція (сума) деяких незалежних елементарних сигналів, збігається зі зростанням кількості складових до нормального закону. Цей результат має фундаментальне теоретичне та практичне значення. Дійсно, фізичні умови формування різноманітних сигналів часто є такими, що до них може бути застосована

гіпотеза про нормальний розподіл імовірностей. Почнемо, перш за все, з тієї обставини, що розповсюдження сигналів у будь-якому фізичному середовищі підлягає певним рівнянням математичної фізики, зокрема, хвильовому рівнянню або, якщо розглядати спектральну область, рівнянню Гельмгольца. Рішення цих рівнянь щодо поля в області прийому може бути представлене інтегральним оператором Гріна відповідної крайової задачі від функції джерела. Таким чином, якщо джерело завади має певну протяжність у просторі, то поле, яке ним створюється на приймальній апертурі, є результатом суперпозиції полів від кожного елементу зони, що шумить. Уже ця обставина створює передумови для нормалізації розподілу завад таких джерел. Крім того, відзначимо, що функція Гріна крайової задачі також може бути представлена суперпозицією часткових рішень або нормальних хвиль (мод). Отже, вже в самій структурі хвильових полів присутні, як правило, суттєві фізичні чинники, які сприяють наближенню їх розподілу до нормального закону. Відзначимо й те, що пристрої Front-end приймальної системи також мають власну інерційність, тобто, вони чинять інтегруючий вплив на сигнал, що додатково нормалізує розподіл завад на їхніх виходах. З урахуванням викладених міркувань стає зрозумілою домінуюча роль гавсівської гіпотези щодо розподілу миттєвих значень зовнішніх завад.

Звичайно, нормалізація розподілу завад відбувається не завжди. Для того, щоб ефекти накладання елементарних сигналів дійсно мали місце, необхідні певні умови й щодо їх протяжності у часі, тобто, це повинні бути досить тривалі або часто повторювані сигнали. Якщо ж джерело завад генерує епізодичні короткотермінові сигнали, то ймовірність їх перекриття в часі дуже мала, що суттєво зменшує можливості нормалізації. Тому серед завад, що не відносяться до гавсівських, найбільш розповсюдженими є імпульсні завади різноманітного походження, якими також часто бувають насичені різноманітні фізичні середовища. Але питання оброблення сигналу на тлі негавсівських завад, зокрема, імпульсних, є достатньо складними та потребують окремого розгляду. Тому в даній частині Посібника ми обмежимося лише нормальними моделями завад. При цьому без обмеження загальності математичне сподівання завад можна вважати нульовим.

Ще одним важливим та практично поширеним припущенням є застосування гіпотези про квазістаціонарність завад у межах інтервалу спостереження $t \in (t_0, t_0 + T)$ та їх квазіоднорідність на просторовій апертурі $\mathbf{r} \in \mathbf{D}_A$. Нагадаємо, що у "вузькому" сенсі стаціонарність шуму означає⁶ інваріантність його статистичних характеристик щодо положення початкової точки відліку часу, однорідність – аналогічна інваріантність щодо розташування точки початку координат у просторі. Але в рамках нормального розподілу достатньо так званого "широкого" визначення, а саме, виконання наступного співвідношення щодо АКФ поля

$$K_z(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2) = \langle z(t_1, \mathbf{r}_1) \cdot z(t_2, \mathbf{r}_2) \rangle = K_z(t_2 - t_1, \mathbf{r}_2 - \mathbf{r}_1), \quad (9.5)$$

тобто, АКФ поля завад має залежати не від абсолютних значень аргументів, а лише від їхньої різниці. Відразу необхідно зауважити, що в реальному житті стаціонарні та однорідні поля, строго кажучи, існувати не можуть хоча б із тієї причини, що кожне фізичне поле має певні межі як у просторі, так й у часі. Теоретичне ж визначення стаціонарності та однорідності існування таких границь, взагалі кажучи, не передбачає. Крім того, характеристики шумів, вочевидь, залежать від стану навколишнього середовища, яке теж змінюється в часі та просторі. Тому практично можна говорити лише про інтервали квазістаціонарності $T_{\text{стац.}}$ та квазіоднорідності $L_{\text{однор.}}$ шумового поля. Інтервали $T_{\text{стац.}}$ та $L_{\text{однор.}}$ визначають відповідно як проміжок часу та розмір області простору, у межах яких приблизно зберігаються властивості стаціонарного та однорідного у "широкому" сенсі поля. Отже, якщо максимальні розміри просторово-часової апертури $\max|\mathbf{T}|$ та $\max|\mathbf{D}_A|$ не перевищують відповідно інтервалів $T_{\text{стац.}}$ та $L_{\text{однор.}}$, то приблизно можна вважати, що поле завад є квазістаціонарним та квазіоднорідним. Виконання цих умов у поєднанні з гавсівською гіпотезою значно полегшує справу, оскільки в цьому випадку розподіл завад та всі їх статистичні характеристики визначається лише кореляційною функцією (9.5).

Але слід зазначити, що навіть у цьому випадку залишаються суттєві складнощі, пов'язані з обчисленням конкретного вигляду АКФ (9.5). Це досить серйозне математичне та фізичне завдання, якому присвячено окремі розділи сучасної

⁶ Дивись розділ 4.

статистичної радіофізики. Деякі приклади таких обчислень ми розглянемо у розділі 12 нашого Посібника.

9.3. Опис полів сигналів.

Як зазначалося вище, всі компоненти рівняння спостережень (9.1) є випадковими полями. Але це зовсім не значить, що для опису їх характеристик треба використовувати однакові підходи. Почнемо з тієї очевидної обставини, що поле сигналу⁷, на відміну від поля шумів та завад, є носієм корисної інформації. Отже, завданням приймальної системи є виділення цієї інформації із прийнятого сигналу, тоді як стосовно завад воно навпаки, полягає в їхньому придушенні. Тому будь-який опис сигнальної компоненти $s(t, \mathbf{r}/\boldsymbol{\theta})$ має цю обставину брати до уваги. Зокрема, суто статистичний підхід, який було застосовано вище для опису шумів та завад, у даному разі не завжди є доцільним. Дійсно, такий статистичний опис не передає дрібних деталей, а відображає лише певні усереднені характеристики. Але у випадку корисного сигналу важливі, як правило, будь-які деталі та подробиці.

Тому для опису апріорних даних про сигнал звичайно застосовують наступний підхід: спочатку формують модель сигнального поля так, начебто вид сигналу, його параметри та умови передачі в середовищі точно відомі. Потім у модель можуть бути внесені корективи, пов'язані з відсутністю або неточним знанням певних факторів. До таких факторів, зокрема, можуть відноситись як деякі параметри сигналу джерела, так і взагалі його вигляд, характеристики середовища передачі, розташування джерела та приймальної системи, тощо. Отже, почнемо наш опис з огляду деяких детермінованих моделей, які будуть необхідні для подальшого розгляду. Потім, по мірі необхідності, будемо вносити до моделі ті чи інші випадкові фактори.

У рамках детермінованого підходу повний опис поля сигналу зводиться до вирішення деяких крайових задач математичної фізики. Зокрема, стосовно полів радіосигналів – це система рівнянь Максвелла, для звукових полів – акустичне наближення рівнянь аеро- або гідромеханіки. У математичному плані такі задачі є

⁷ Нагадаємо, що відповідно до загальної схеми рис.9.1 сигналом ми називаємо ті фізичні процеси, які виникають по виходу датчиків антени під дією випромінювання джерела корисної інформації.

достатньо складними та трудомісткими, а про існування «точних» рішень у більшості ситуацій годі й говорити. Але теоретичний аналіз моделей дозволяє виявити головні особливості поля сигналу, розуміння яких може бути корисним для побудови деяких наближених розв'язків, більш прийнятних для практичного застосування.

Ми не ставимо собі за мету систематичний розгляд усіх етапів побудови сигнальних моделей – це є предметом відповідних теоретичних розділів радіотехніки, оптики, акустики, тощо. Але коли мова заходить про якісь хвильові процеси, то безвідносно до фізичної природи в них виявляється багато спільного. Зокрема, існують певні елементарні хвильові моделі, структура яких не залежить від фізичної природи середовища передачі сигналу. Ці моделі відповідають рішенням хвильових задач для деяких ідеалізованих середовищ, наприклад, безмежного та однорідного простору. Потім із цих елементарних моделей можуть збиратися більш складні хвильові «конструкції» для умов, наближених до реального середовища.

Розглянемо деякі приклади. Однією з найбільш уживаних моделей є сферична хвиля. Так називають поле сигналу, яке утворює в однорідному та безмежному просторі точкове джерело⁸. Характерною особливістю хвиль цього типу є сферична форма поверхні хвильового фронту. Нагадаємо, що фронтом хвилі прийнято називати поверхню рівних затримок, тобто, на цій поверхні значення сигналу мають однакову затримку щодо джерела.

Звернемося до рис.9.8, який ілюструє даний приклад. Тут початок координат O_A вибрано в межах деякої приймальної апертури $\mathbf{r} \in \mathbf{D}_A$, джерело сигналу знаходиться в точці $\mathbf{r}_S$, точка прийому M має координати $\mathbf{r}$. Для вказаної геометрії задачі поле сигналу зі сферичним фронтом хвилі може бути записано [14] у наступному вигляді:

$$s(t, \mathbf{r}) = \frac{A_S}{|\mathbf{r} - \mathbf{r}_S|} \cdot s_0 \left(t - \frac{|\mathbf{r} - \mathbf{r}_S|}{c_0} \right). \quad (9.6)$$

У цьому співвідношенні c_0 - швидкість поширення хвилі в однорідному середовищі, $s_0(t)$ - функція збудження джерела, A_S - амплітуда сигналу, яка залежить

⁸ Фізичним наближенням точкового джерела звичайно вважають монополь малого хвильового розміру.

від фізичних та технічних умов задачі, $|\mathbf{r} - \mathbf{r}_S|$ - відстань від джерела до точки приймання M_R . Отже, як видно з виразу (9.6), сигнал на приймальній апертурі повторює за формою сигнал джерела та відрізняється від нього амплітудою та затримкою.

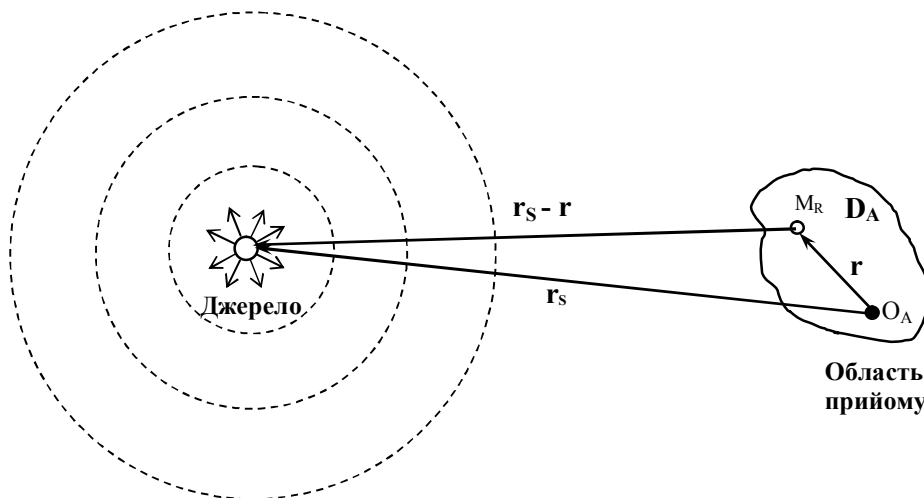


Рис.9.8. Сигнал зі сферичним фронтом хвилі

При цьому амплітуда сигналу спадає зворотно пропорційно відстані між джерелом та точкою прийому, а ця ж відстань, поділена на швидкість поширення хвилі, дає затримку сигналу, тобто,

$$\tau(\mathbf{r}) = \frac{|\mathbf{r} - \mathbf{r}_S|}{c_0}. \quad (9.7)$$

Неважко бачити, що рівняння для фронту сферичних хвиль $\tau(\mathbf{r}) = const$ відповідає сфері з центром у точці випромінювання $\mathbf{r}_S$, що й пояснює походження їх назви. Тепер запишемо спектр поля сигналу (9.6). Виконавши від виразу (9.6) перетворення Фур'є, маємо:

$$S(\omega, \mathbf{r}) = A_S \cdot S_0(\omega) \cdot G(\mathbf{r} / \mathbf{r}_S). \quad (9.8)$$

Тут $S_0(\omega)$ - спектр Фур'є функції збудження джерела $s_0(t)$, а

$$G(\mathbf{r} / \mathbf{r}_S) = \frac{1}{|\mathbf{r} - \mathbf{r}_S|} \exp\{-j \cdot \kappa_0 \cdot |\mathbf{r} - \mathbf{r}_S|\} \quad (9.9)$$

- так звана функція Гріна хвильової задачі, $\kappa_0 = \omega / c_0 = 2\pi / \lambda_0$ - хвильове число, $\lambda_0 = c / f_0$ - довжина хвилі сигналу. Зазначимо, що функція (9.9) у побудові багатьох хвильових моделей відіграє ключову роль. Часто саме вираз (9.9) у теоретичних працях і називають сферичною хвилею, розуміючи під фронтом хвилі поверхню рівних фаз.

На великих відстанях від джерела та в межах відносно невеликих приймальних апертур сферичний фронт хвилі (9.9) можна замінити площиною. Таке представлення поля точкового джерела називають [14] наближенням дальньої зони (наближенням Фраунгофера). Якщо відстань $r_S = |\mathbf{r}_S|$ від джерела до центру області прийому O_A значно перевищує відстань $r = |\mathbf{r}|$ від точки O_A до точки приймання M_R (тобто, $r_S \gg r$), то для відстані поміж джерелом та точкою M_R можна скористатися наступним ступеневим розкладом:

$$|\mathbf{r} - \mathbf{r}_S| = r_S - (\mathbf{n}_S, \mathbf{r}) + \frac{r^2 - (\mathbf{n}_S, \mathbf{r})^2}{2 \cdot r_S} - \dots,$$

де $\mathbf{n}_S = \mathbf{r}_S / r_S$ - одиничний вектор напрямку від точки O_A на джерело, $(\mathbf{n}_S, \mathbf{r})$ - скалярний добуток цього одиничного вектору на координатний вектор точки спостереження. Згідно наближення дальньої зони при обчисленні затримки сигналу (9.7) у даному розкладі залишають два перших члени, а при обчисленні амплітудного ослаблення – тільки перший член. Тоді хвиля (9.9) у вказаному наближенні приймає наступний вигляд:

$$G(\mathbf{r} / \mathbf{r}_S) \cong \frac{1}{r_S} \exp\{-j \cdot \kappa_0 \cdot (r_S - (\mathbf{n}_S, \mathbf{r}))\}. \quad (9.10)$$

Неважко бачити, що поверхня рівних фаз хвилі (9.10) є площиною, нормаль до якої задається вектором $\mathbf{n}_S$. Тому вираз (9.10) називають однорідною⁹ плоскою хвилею [14].

⁹ Визначення плоских хвиль як однорідних означає рівномірність їх амплітуди по фронту. В іншому випадку хвилі називають неоднорідними [14]. Неоднорідні плоскі хвилі виникають, наприклад, біля поверхонь розділу середовищ або поблизу тіл, що викликають дифракцію.

Плоску хвилю можна також записати у часовій області. Для цього підставимо (9.10) до виразу (9.8) та виконаємо зворотне перетворення Фур'є. У результаті маємо наступну модель поля сигналу в дальній зоні точкового джерела:

$$s(t, \mathbf{r}) \cong \frac{A_S}{r_S} \cdot s_0(t - \tau_S + (\mathbf{n}_S, \mathbf{r})/c_0) \quad (9.11)$$

де $\tau_S = r_S / c_0$ - час розповсюдження сигналу від джерела до фазового центру приймальної антени. Останній член в аргументі сигналу (9.11) відображає розкид затримок сигналу вздовж приймальної апертури відносно фазового центру антени. Залежність цього розкиду від координат точки прийому визначається, як бачимо, напрямком приходу хвилі сигналу та геометрією антени. Деякі приклади щодо цього ми розглянемо в наступних розділах Посібника.

Розглянуті найпростіші випадки сферичних та плоских хвиль становлять не лише теоретичний інтерес: ці хвилі часто є елементами більш складних хвильових «конструкцій», про які піде мова далі.

9.4. «Багатопроменева» модель поля сигналу.

Як уже зазначалося вище, практичні підходи до побудови моделей полів сигналів базуються, як правило, на застосуванні наближених методів рішення відповідних хвильових задач. У свою чергу, вибір відповідного наближення визначається фізичними умовами практичної задачі, серед яких геометрія середовища, параметри його неоднорідності, властивості границь, робочий діапазон частот та інше. Зокрема, для багатьох практичних завдань приймання радіосигналів у найбільш популярних на сьогодні діапазонах мобільного радіозв'язку, безпроводних мереж, тощо (~ від 800 до 2600 МГц) широке застосування має так зване **короткохвильове наближення** [14]. В основі цього наближення лежить асимптотичне рішення хвильових задач для випадку, коли хвильове число необмежено зростає, тобто, $k_0 \rightarrow \infty$, що еквівалентно $\lambda_0 \rightarrow 0$. Оскільки реальна довжина хвилі є скінченною, то звичайним критерієм для застосування цього наближення є співвідношення між довжиною хвилі та масштабом неоднорідності середовища R_M , а саме, повинна

виконуватися нерівність $\lambda_0 < R_M$. При цьому наближення тим краще, чим ця нерівність сильніша.

У теорії хвильових полів [14] показано, що при виконанні вказаних умов та для областей, що віддалені від джерел сигналів та границь середовища, поле сигналу може бути представлене суперпозицією плоских хвиль, подібних до (9.11), тобто:

$$s(t, \mathbf{r}) = \sum_{p=1}^L A_p \cdot s_0(t - \tau_p + (\mathbf{n}_p, \mathbf{r})/c). \quad (9.12)$$

Тут L - кількість енергетично значущих плоских хвиль, що потрапляють від джерела сигналу на приймальну антену, A_p , τ_p , $\mathbf{n}_p$ - відповідно їхні амплітуди, затримки та напрямки приходу.

У представлення поля сигналу (9.12) є й інша назва, а саме, *наближення геометричної оптики*. Походження цієї назви потребує додаткових пояснень. Спершу зауважимо, що напрямок приходу кожної з плоских хвиль суперпозиції (9.12) визначається відповідною нормаллю $\mathbf{n}_p$ до фронту хвилі. Цей напрямок в оптиці прийнято називати *променем*. В однорідному безмежному середовищі промені є прямими лініями, що поєднують точки випромінювання та прийому сигналу. Але при наявності в середовищі будь-яких неоднорідностей променева траєкторія може змінюватися. Для таких змін можуть бути дві основні причини:

- наявність у середовищі на шляху променів будь-яких «непрозорих» об'єктів;
- зміна фізичних параметрів середовища вздовж траси розповсюдження сигналу.

У першому випадку напрямок розповсюдження плоскої хвилі до перешкоди залишається прямолінійним, але після падіння на відбиваючий об'єкт напрямок змінюється за законами геометричної оптики (законами Снеліуса). Другий випадок значно складніший, оскільки зміна параметрів середовища призводить до явища *рефракції*, під дією якої траєкторії променів відхиляються від початкового напрямку або взагалі стають нелінійними.

Модель (9.12) охоплює достатньо широкий клас завдань приймання сигналів та є, мабуть, однією з найбільш загальних. Для розрахунку параметрів моделі звичайно використовують формули Снеліуса [14], характеристики відбиваючих поверхонь та рівняння переносу, що визначає енергетичні параметри моделі. Розроблено цілу низку алгоритмів та комп'ютерних програм для побудови променевих трас сигналу, орієнтованих на певні прикладні галузі, зокрема, радіотехніку та акустику [7]. В якості ілюстрації на рис.9.9 показано результати роботи однієї з таких програм при аналізі радіовидимості в районах щільної міської забудови.

На рисунку показані траєкторії (траси) деяких радіо променів, а загальні рівні сигналу позначено за допомогою стандартної шкали кольорів. Неважко бачити, що хоча передавальна антена розташована на даху високого будинку, пряма радіовидимість від антени до більшості ділянок забудови відсутня. Але незважаючи на це, рівні сигналу в «затінкових» зонах достатньо високі за рахунок багатократного відбиття променів. Зауважимо, що саме в таких умовах і працює сучасний мобільний радіозв'язок!

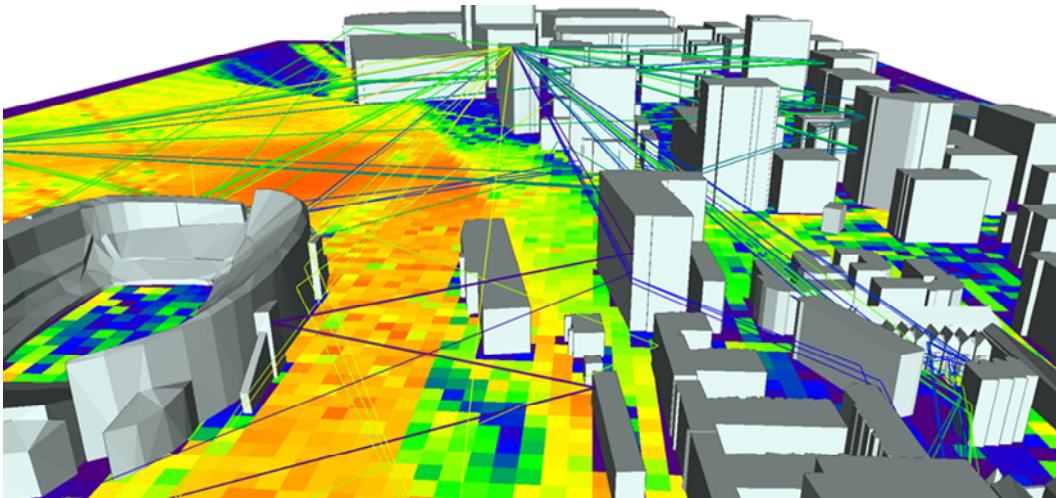


Рис.9.9. Трасування радіо променів в умовах міської забудови.

У наведеному прикладі основну проблему становлять елементи міської забудови, які перекривають пряму радіовидимість та створюють численні пере відбиття сигналу. Але поміж елементами забудови поширення радіосигналу все ж таки залишається прямолінійним. Значно складніше виглядає картина поширення акустичного сигналу в морському середовищі, де присутня значна рефракція звуку.

Джерелом такої рефракції є неоднорідність морського середовища, зумовлена зміною температури води, гідростатичного тиску та її солоності. Під дією вказаних факторів швидкість поширення звуку змінюється з глибиною, що й створює рефракційні ефекти. Ці ефекти ілюструє рис.9.10, на якому показана картина акустичних променів для типових умов глибоководних районів Світового океану.

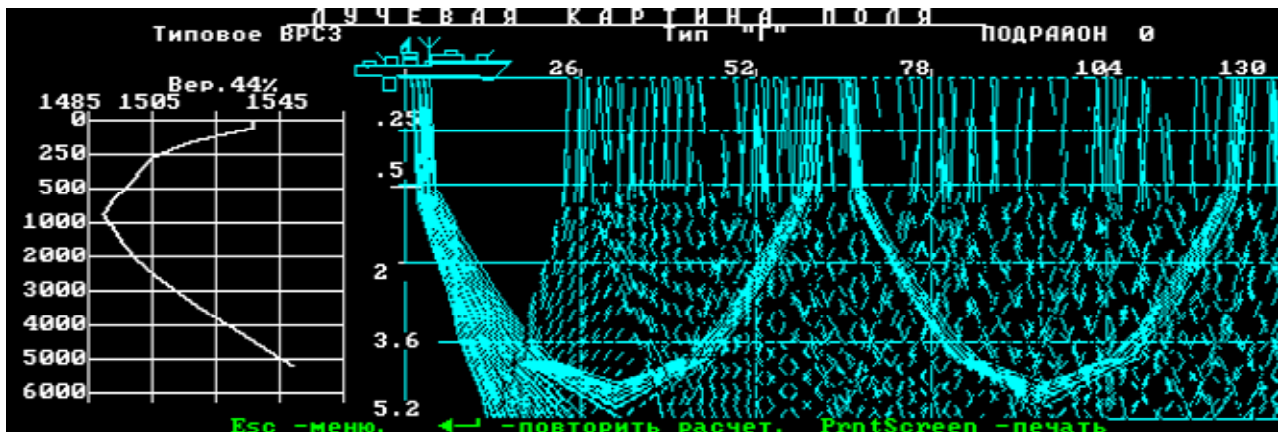


Рис.9.10. Променева картина поширення гідроакустичного сигналу для типових районів Світового океану.

На лівій частині цього рисунку представлена типова залежність швидкості поширення звуку (м/сек) від глибини (м), справа – відповідні їй променеві траєкторії вздовж траси довжиною до 130 км. З рисунку видно, що рефракція створює значну нерівномірність розподілу інтенсивності акустичного поля в зоні спостереження як за глибиною, так і по дистанції.

Контрольні запитання та завдання до розділу 9.

Дати в конспекті письмові відповіді на наступні запитання:

1. Пояснити загальну схему приймання сигналів рис.9.1 та призначення її окремих складових.
2. Що таке завади та внутрішні шуми? Яка поміж ними принципова різниця?
3. Що є сигналом у системах радіо та гідролокації? Яку корисну інформацію він несе? Які завади перешкоджають їх роботі?
4. Що таке системи телекомунікацій? Наведіть приклади таких систем та їх призначення.
5. Що таке рівняння спостережень? Які складові до нього входять? Наведіть приклади скалярних та векторних спостережень.
6. Що таке завдання виявлення та розрізнення сигналів? Яка поміж ними різниця? Де такі завдання виникають?

7. Що таке завдання оцінювання параметрів сигналів? Яка принципова різниця поміж завданнями оцінювання та розрізнення сигналів? Де завдання оцінювання виникають?
8. Що таке завдання фільтрації сигналів? Яка принципова різниця поміж завданнями фільтрації та оцінювання параметрів сигналів? Де завдання фільтрації виникають?
9. Що таке апіорна інформація? Що до неї входить та для чого вона потрібна?
10. У чому полягає типова апіорна інформація про внутрішні шуми?
11. У чому полягає типова апіорна інформація про завади? На яких загальних фізичних та математичних законах вона ґрунтується?
12. Як формується апіорний опис сигналу? У чому складність формування такого опису?
13. Що таке сферична хвиля? В яких умовах вона виникає?
14. Що таке наближення Фраунгофера? В чому його зміст?
15. Що таке плоска хвиля? При яких умовах вона виникає? Якими параметрами вона визначається?
16. У чому полягає багатопроменева модель поля сигналу? При яких умовах вона може бути застосована? Що таке трасування променів?

Розділ 10. Елементи математичної статистики.

У попередньому розділі ми розглянули фізичні аспекти завдань приймання сигналів та зробили деякі кроки до їхньої формалізації. Тепер подивимося на ці завдання так би мовити з іншого «боку», а саме, з точки зору застосування апарату теоретичної статистики. Математична (теоретична) статистика, яка на початкових етапах свого розвитку займалася обґрунтуванням статистичних методів обліку народонаселення та господарчої діяльності в наш час стала однією з основ інформаційних технологій, у тому числі й методів приймання та оброблення сигналів. З багатьох напрямків цієї теорії ми зупинимось лише на двох для нас найголовніших, а саме, в першій частині даного розділу розглянемо основні положення теорії перевірки гіпотез, а в наступній – статистичної теорії оцінювання.

Частина I. Теорія перевірки гіпотез.

10.1. Постановка задачі перевірки гіпотез

Теорія перевірки гіпотез складає, мабуть, найбільш відомий розділ математичної статистики. Зокрема, до прийнятих у цій теорії постановок зводяться задачі виявлення сигналів на тлі завад, задачі прийому сигналів зв'язку, задачі класифікації, тощо. Усі такі задачі розглядаються як деякий статистичний експеримент, який повинен підтвердити або спростувати певні статистичні гіпотези.

Нагадаємо, що статистичною гіпотезою H в математичній статистиці називають деяке припущення про ймовірнісні характеристики даних, що спостерігаються. Як зазначалося в розділі 3, найбільш повною ймовірнісною характеристикою деякого вектору¹ спостережень $\mathbf{u}$ є його густина розподілу ймовірностей. Оскільки цей розподіл залежить від висунутого припущення, то його звичайно позначають $p(\mathbf{u}/H)$, тобто, як умовний розподіл експериментальних даних у випадку, коли припущення H дійсно виконується. Розподіл $p(\mathbf{u}/H)$ залежить фактично від двох аргументів: вектору спостережень $\mathbf{u}$ та висунутої гіпотези H . По першому з цих аргументів функція $p(\mathbf{u}/H)$ задовольняє загальним вимогам для будь-якої густини ймовірностей.

¹ Тут і далі вектора спостережень, якщо не сказано інше, вважаються скінченновимірними.

Залежність же від висунутого припущення H прийнято називати **функцією правдоподібності гіпотез**. Ця функція посідає в математичній статистиці чільне місце і до неї ми будемо неодноразово звертатися.

Задача перевірки гіпотез виникає тоді, коли відносно одних і тих же спостережень u висувається декілька конкуруючих гіпотез $H_0, H_1, H_2, \dots, H_{M-1}$. Спираючись на експериментальні дані, одержані в результаті статистичних випробувань, треба визначитись, яка з цих гіпотез має місце насправді. Особливим випадком таких випробувань є перевірка двох гіпотез H_0 та H_1 . Особливість полягає в тому, що, по-перше, така ситуація на практиці зустрічається найчастіше, по-друге, для вибору між двома гіпотезами можуть бути застосовані деякі спеціальні методи, які для загального випадку перевірки кількох гіпотез не завжди придатні.

Методи вирішення всіх цих задач залежать від об'єму та характеру апіорної інформації про ймовірнісні характеристики спостережень, які формуються у вигляді статистичних гіпотез. Прийнято розрізняти три види статистичних гіпотез: прості, складні та непараметричні. **Проста гіпотеза** повністю визначає вид функції правдоподібності $p(u/H)$, тобто, як її функціональний вид, так і всю сукупність параметрів, що до неї входять. **Складна гіпотеза**, на відміну від простої, залишає невизначеними деякі параметри розподілу $p(u/H)$ при відомому загалом його функціональному вигляді. Нарешті **непараметрична гіпотеза** стосується тільки найбільш загальних властивостей розподілу $p(u/H)$, не конкретизуючи його вигляд.

Проілюструємо різницю поміж цими видами гіпотез на прикладі задачі перевірки гіпотез про математичне очікування спостережень. Задля простоти припустимо, що вибірка одновимірна, тобто, спостереження u - скалярна величина. Нехай у випадку гіпотези H_0 математичне очікування вибірових даних $\langle u/H_0 \rangle = 0$, а у випадку гіпотези H_1 - $\langle u/H_1 \rangle = a \neq 0$. Якщо додатково відомо, що для обох випадків вибірка має нормальний розподіл ймовірностей, то гіпотези H_0 та H_1 можуть бути записані у вигляді:

$$H_0: p(u/H_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{u^2}{2\sigma^2}\right\}; \quad (10.1a)$$

$$H_1: p(u/H_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(u-a)^2}{2\sigma^2}\right\}, \quad (10.1б)$$

де σ^2 - дисперсія вибірки. Подібна задача виникає, наприклад, у цифрових логічних схемах при розпізнаванні кодів "0" та "1" оскільки одиничний код звичайно передається високим рівнем напруги, а нульовий код – напругою, що близька до нуля (можливі й інші варіанти).

Гіпотези (10.1a) та (10.1б) є простими, якщо параметри σ^2 та a наперед відомі. Але у випадку, коли конкретне значення математичного очікування a невідомо, маємо задачу перевірки простої гіпотези H_0 проти складної гіпотези (альтернативи) H_1 . Якщо ж невідома й дисперсія вибірки σ^2 , але розподіл вибірки залишається нормальним, то маємо задачу перевірки двох складних гіпотез. Нарешті, коли, наперед відомо лише те, що математичне очікування вибірки при гіпотезі H_1 позитивне, а у випадку гіпотези H_0 нульове, то маємо задачу перевірки непараметричної гіпотези $a > 0$.

Для розв'язання кожної з цих задач треба побудувати деяку процедуру прийняття рішення на користь однієї з можливих гіпотез, яка базується на вибірці даних $\mathbf{u}$ та апріорній інформації про ймовірнісні характеристики статистичного ансамблю. Таку процедуру в математичній статистиці прийнято називати **статистичним критерієм** або **статистичним правилом (тестом)**. Розрізняють два класи статистичних критеріїв (або правил): нерандомізовані та рандомізовані. Нерандомізовані критерії мають ту визначальну особливість, що існує детермінована залежність поміж вхідними даними та результатом прийняття рішення. Таким чином, якщо одна й та ж вибірка даних $\mathbf{u}$ (тобто, незмінна) піддається статистичним випробовування за допомогою одного й того ж нерандомізованого статистичного критерію багато разів, то завжди буде прийматися одне й те ж рішення. Для рандомізованих же процедур залежність поміж вхідними даними та прийнятим рішенням є ймовірнісною, тобто, існує

ненульова ймовірність прийняття рішення на користь не одної, а кількох конкуруючих гіпотез при багатократному використанні незмінного набору вибірових даних. Слід зазначити, що рандомізація означає, по суті, внесення додаткового шуму в пристрій, що здійснює прийняття рішення. Тому ця техніка не може принципово покращити якість обробки інформації, а є суто формальною статистичною процедурою, яка дозволяє в окремих випадках лише підігнати достовірність прийняття рішення під заданий рівень. Виходячи з цієї обставини, надалі обмежимося розглядом лише нерандомізованих процедур.

Оскільки для нерандомізованих статистичних критеріїв існує детермінований зв'язок поміж вибіркою даних та прийнятим рішенням, то побудову статистичного правила можна трактувати як розбиття простору вибірових даних $\mathcal{U}$ на M підпросторів U_k , ($k = 0, 1, 2, \dots, M-1$), що не перетинаються. Таким чином, для всіх векторів даних $\mathbf{u} \in U_1$, які входять, до першого підпростору, рішення завжди приймається на користь гіпотези H_1 , для даних, які входять до другого підпростору ($\mathbf{u} \in U_2$) - рішення на користь гіпотези H_2 й т.д. Отже, побудова нерандомізованих статистичних критеріїв зводиться до вибору способу розбиття простору $\mathcal{U}$.

10.2. Оптимальні правила статистичного рішення

Тепер обговоримо питання, який саме статистичний критерій (спосіб розбиття простору $\mathcal{U}$) в умовах конкретної задачі є найкращим, тобто, оптимальним. Перш за все треба визначитися, в якому сенсі розуміти оптимальність правила вирішення. Класична теорія перевірки гіпотез у її сучасному вигляді спирається на концепцію так званого баєсового ризику. Сутність цього поняття полягає в тому, що вводиться деяка матриця втрат від помилок в прийнятті рішень (матриця штрафів), а оптимізація проводиться за критерієм мінімізації математичного очікування загальних втрат. Нехай C_{ik} - втрати від прийняття рішення на користь гіпотези H_i коли насправді має місце гіпотеза H_k , а $p(H_i / H_k)$ - ймовірність такого рішення. Тоді середній баєсовий ризик може бути записаний у вигляді:

$$R = \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} C_{ik} \cdot p(H_i / H_k) \cdot P(H_k), \quad (10.2)$$

де $P(H_k)$ - апіорні ймовірності гіпотез. У свою чергу ймовірність рішення на користь i -ої гіпотези коли насправді справедлива k -а гіпотеза можна обчислити шляхом інтегрування (підсумовування) по області $\mathbf{u} \in U_i$ функції правдоподібності $p(\mathbf{u} / H_k)$ для ситуації, що відповідає гіпотезі H_k , тобто:

$$p(H_i / H_k) = \int_{\mathbf{u} \in U_i} p(\mathbf{u} / H_k) d\mathbf{u}.$$

З урахуванням цього виразу для баєсового ризику остаточно маємо:

$$R = \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} C_{ik} \cdot P(H_k) \cdot \int_{\mathbf{u} \in U_i} p(\mathbf{u} / H_k) d\mathbf{u}.$$

Мінімізація даного критерію приводить [3] до наступного оптимального статистичного правила: рішення на користь деякої гіпотези H_i ($i = 0, 1, \dots, M-1$) приймається в тому випадку, коли задовольняється система $(M-1)$ нерівностей:

$$\sum_{k=0}^{M-1} C_{ik} \cdot P(H_k) \cdot p(\mathbf{u} / H_k) \leq \sum_{k=0}^{M-1} C_{jk} \cdot P(H_k) \cdot p(\mathbf{u} / H_k), \quad (10.3)$$

$$j = 0, 1, \dots, M-1; \quad j \neq i.$$

Правило (10.3) можна надати більш наочний вигляд, якщо скористатися теоремою множення ймовірностей:

$$p(\mathbf{u} / H_k) \cdot P(H_k) = P(H_k / \mathbf{u}) \cdot p(\mathbf{u}),$$

де $P(H_k / \mathbf{u})$ - апостеріорна ймовірність гіпотези H_k при відомих спостереженнях $\mathbf{u}$, $p(\mathbf{u})$ - безумовний розподіл спостережень по всій сукупності можливих гіпотез. З урахуванням останнього співвідношення а, також тієї обставини, що безумовний розподіл спостережень $p(\mathbf{u})$ від гіпотези не залежить, статистичне правило (10.3) може бути приведене до вигляду:

$$\sum_{k=0}^{M-1} C_{ik} \cdot P(H_k / \mathbf{u}) \leq \sum_{k=0}^{M-1} C_{jk} \cdot P(H_k / \mathbf{u}), \quad (10.4)$$

$$j = 0, 1, \dots, M-1; \quad j \neq i.$$

Отже, для прийняття оптимального рішення треба виконати наступні дії. Спочатку на підставі вектору спостережень $\mathbf{u}$ обчислити M значень апостеріорних ймовірностей $P(H_k / \mathbf{u})$ для кожної з гіпотез $H_0, H_1, H_2, \dots, H_{M-1}$. Далі ці значення

ймовірностей необхідно M раз підсумувати з наборами вагових коефіцієнтів, що відповідають M рядкам

$$(C_{i0}, C_{i1}, \dots, C_{i(M-1)}), \quad i = 0, 1, \dots, M-1$$

матриці втрат. Рішення приймається на користь тієї з гіпотез, для якої результат підсумовування виявиться найменшим.

Відзначимо, що тут ми зустрілись ще з одним базовим у статистичній теорії поняттям, а саме, **тестовою статистикою**. Тестовою статистикою називають деякий функціонал від спостережень $\mathcal{F}(\mathbf{u})$, який і визначає спосіб розбиття простору спостережень $\mathcal{U}$ на підпростори $\mathbf{U}_i$, $i = 0, 1, \dots, M-1$. Наприклад, для правила (10.4) тестовою статистикою є зважена сума апостеріорних імовірностей для всіх рядків матриці втрат, тобто:

$$\begin{aligned} \mathcal{F}(\mathbf{u}) &= \{\mathcal{F}_0(\mathbf{u}), \mathcal{F}_1(\mathbf{u}), \dots, \mathcal{F}_{M-1}(\mathbf{u})\}, \quad \text{де} \\ \mathcal{F}_i(\mathbf{u}) &= \sum_{k=0}^{M-1} C_{ik} \cdot P(H_k / \mathbf{u}), \quad i = 0, 1, \dots, M-1. \end{aligned}$$

Порівняння різних компонент цієї статистики поміж собою при кожній реалізації вектору спостережень і визначає конфігурацію простору рішень. Отже, побудову статистичного критерію можна також трактувати, як вибір відповідної тестової статистики. Зауважимо, що оптимальне правило (10.4) залежить від матриці втрат C_{ik} , вибір якої, взагалі кажучи, не є предметом статистичної теорії. Однак, існує досить значний діапазон умов, для яких вибір цієї матриці практично не має значення. Розглянемо спочатку окремий випадок, коли втрати від правильних рішень відсутні, а втрати від усіх неправильних рішень однакові, тобто,

$$C_{ik} \equiv \begin{cases} 0, & k = i \\ 1, & k \neq i \end{cases} \quad (10.5)$$

Таку штрафну матрицю іноді називають простою [3]. Неважко бачити, що баєсовий ризик (10.2) у цьому окремому випадку приймає вигляд

$$R = \sum_{i=0}^{M-1} \sum_{\substack{k=0 \\ k \neq i}}^{M-1} P(H_i / H_k) \cdot P(H_k) = P_{\text{error}}, \quad (10.6)$$

що, за визначенням, є повна ймовірність помилкових рішень. Таким чином, для простої матриці втрат (10.5) мінімізація баєсового ризику еквівалентна мінімізації середньої ймовірності помилок (10.6). Зауважимо, що критерій (10.6) є основним для задач прийому сигналів зв'язку, де його іноді називають [20] критерієм ідеального спостерігача.

Тепер розглянемо безпосередньо оптимальне правило рішення для цього окремого випадку. Після підстановки (10.5) у правило (10.4) одержимо:

$$\sum_{\substack{k=0 \\ k \neq i}}^{M-1} P(H_k / \mathbf{u}) \leq \sum_{\substack{k=0 \\ k \neq j}}^{M-1} P(H_k / \mathbf{u}), \quad j = 0, 1, \dots, M-1; \quad j \neq i.$$

Перепишемо це правило, виключивши обмеження на індекс підсумовування:

$$\sum_{k=0}^{M-1} P(H_k / \mathbf{u}) - P(H_i / \mathbf{u}) \leq \sum_{k=0}^{M-1} P(H_k / \mathbf{u}) - P(H_j / \mathbf{u}),$$

$$j = 0, 1, \dots, M-1; \quad j \neq i.$$

Після приведення подібних членів та зміни знаку дане правило приймає вигляд:

$$P(H_i / \mathbf{u}) \geq P(H_j / \mathbf{u}), \quad j = 0, 1, \dots, M-1; \quad j \neq i. \quad (10.7)$$

Таким чином, при виборі матриці втрат (10.5) ми приходимо до правила статистичного рішення за критерієм максимуму апостеріорної ймовірності гіпотез, яке в статистичній літературі прийнято позначати аббревіатурою MAP (від англійської назви методу Maximum A posteriori Probability).

Як видно з виразу (10.7), тестова статистика цього методу формується набором значень апостеріорних ймовірностей гіпотез, які обчислюються для вже відомої вибірки спостережень. Очевидно, що правило (10.7) є інваріантним до будь-яких монотонних перетворень обох його частин, які не змінюють відношень порядку "більше-менше". Найчастіше в якості такого перетворення використовують логарифмічну функцію. Отже, правило (10.7) може бути записано в еквівалентному вигляді:

$$\log\{P(H_i / \mathbf{u})\} \geq \log\{P(H_j / \mathbf{u})\}, \quad j = 0, 1, \dots, M-1; \quad j \neq i,$$

яке часто буває зручнішим для практичної реалізації. Критерій MAP відіграє в теорії статистичних рішень визначну роль, тому що, як виявляється, він придатний не

тільки для окремого випадку штрафної матриці (10.5), а для значно більш широкого діапазону умов [3]. Ця обставина пояснює роль та місце критерію MAP у загальній теорії статистичних рішень.

Окремим та, мабуть, найбільш важливим випадком статистичних процедур є метод максимальної правдоподібності ML (від слів Maximum Likelihood). Його важливість пояснюється зручністю практичного використання при забезпеченні високої ефективності, яка близька до оптимальної. Цей метод безпосередньо впливає з методу MAP, якщо припустити, що апіорні ймовірності гіпотез однакові, тобто, $P(H_k) \equiv 1/M$. Дійсно, у цьому випадку з формули Баєса

$$P(H_k / \mathbf{u}) = \frac{p(\mathbf{u} / H_k) \cdot P(H_k)}{p(\mathbf{u})} = \frac{p(\mathbf{u} / H_k)}{M \cdot p(\mathbf{u})}$$

витікає, що максимум апостеріорного розподілу має місце для тієї з гіпотез, для якої максимальною є й функція правдоподібності $p(\mathbf{u} / H_k)$. Отже, замість пошуку максимуму апостеріорної ймовірності можна відшукувати максимум функції правдоподібності, тобто, оптимальне правило прийняття рішення (10.7) приймає вигляд:

$$p(\mathbf{u} / H_i) \geq p(\mathbf{u} / H_j), \quad j = 0, 1, \dots, M-1; \quad j \neq i. \quad (10.8)$$

На перший погляд може бути незрозуміло, у чому тут полягає спрощення. Насправді ж річ у тім, що метод ML застосовується не лише при рівномірному розподілі гіпотез, а й у загальному випадку, коли цей розподіл може бути довільним та апіорі невідомим. Це, безумовно, є відходом від строгої оптимальності. Але виявляється [3], що для випадку **надійних** статистичних випробувань апіорний розподіл гіпотез практично не має значення. Критерії надійності ми уточнимо пізніше. Поки ж зауважимо, що надійними звичайно називають такі випробування, в яких імовірність помилкових рішень дуже мала. Таким чином, для надійних випробувань методи MAP та ML практично тотожні. У той же час метод ML не потребує знань про апіорний розподіл гіпотез. Саме цим пояснюється його широке практичне використання. Але в деяких випадках до використання методу ML слід підходити з обережністю, оскільки він усе ж таки не є строго оптимальним.

10.3. Перевірка двох гіпотез

У важливому окремому випадку перевірки двох гіпотез оптимальне правило статистичного рішення (10.3) приймає наступний вигляд:

$$C_{00} \cdot P(H_0) \cdot p(\mathbf{u} / H_0) + C_{01} \cdot P(H_1) \cdot p(\mathbf{u} / H_1) \leq \\ \leq C_{10} \cdot P(H_0) \cdot p(\mathbf{u} / H_0) + C_{11} \cdot P(H_1) \cdot p(\mathbf{u} / H_1).$$

Після приведення подібних членів і деякого перегрупування остаточно одержимо:

$$\lambda(\mathbf{u}) \underset{H_0}{\overset{H_1}{\geq}} \Pi, \quad (10.9)$$

$$\text{де} \quad \lambda(\mathbf{u}) = \frac{p(\mathbf{u} / H_1)}{p(\mathbf{u} / H_0)} \quad (10.10)$$

- так зване відношення правдоподібності гіпотез, яке часто позначають аббревіатурою LR (від англійської назви Likelihood Ratio), а

$$\Pi = \frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)}. \quad (10.11)$$

- поріг прийняття рішення. Виконання нерівності (10.9) вправо відповідає прийняттю рішення на користь гіпотези H_1 , а виконання протилежної нерівності – прийняттю рішення на користь гіпотези H_0 . Таким чином, структура оптимального правила статистичного рішення у випадку перевірки двох гіпотез зводиться до відомого **критерію відношення правдоподібності гіпотез**, а саме, до обчислення функціоналу (10.10) (або якогось монотонного перетворення від нього) та порівнянню його значення з порогом. Будь-які зміни параметрів C_{ik} та $P(H_i)$ впливають лише на величину порога (10.11), не зачіпаючи загальну схему прийняття рішення. Ця обставина сприяла появі кількох інших підходів до вибору порогу прийняття рішення, найбільш відомим з яких є **критерій Неймана-Пірсона** (критерій Н.-П.). У цьому критерії поріг Π встановлюється не за значеннями матриці втрат, яку практично досить складно вибрати, а на підставі заданої імовірності помилки у прийнятті однієї з гіпотез, наприклад H_0 . Ця гіпотеза називається **основною** або нульовою, а конкуруюча гіпотеза H_1 - **альтернативною**. Відхилення основної гіпотези H_0 коли вона насправді

вірна, називають **помилкою I-го роду**, а помилкове відхилення гіпотези H_1 - **помилкою II-го роду**. Критерій Неймана-Пірсона вимагає, щоб ймовірність P_I помилки I-го роду не перевищувала заданого рівня α_0 , а ймовірність P_{II} помилки II-го роду була мінімально можливою. Величину α_0 в математичній статистиці прийнято називати **рівнем значущості критерію**, а ймовірність правильного прийняття альтернативної гіпотези H_1 (тобто, $1 - P_{II}$) - **потужністю критерію**.

Критерій Н.-П. широко застосовується у тих прикладних задачах, де помилкове відхилення однієї з гіпотез тягне за собою значно більш серйозні наслідки, чим помилкове відхилення конкуруючої гіпотези. Як найбільш характерний приклад можна згадати системи попередження про напад: нульова гіпотеза H_0 - "нападу нема", альтернативна гіпотеза H_1 - "напад є". Оскільки помилкове приведення в дію засобів відбиття нападу (наприклад, ядерної або іншої потужної зброї) може призвести до катастрофічних наслідків, то, зрозуміло, що контроль за ймовірністю такої ситуації повинен бути як найсуворішим. Саме за таким принципом працюють системи виявлення цілей у радіо- та гідролокації. У цих системах ймовірність помилки I-го роду, тобто, прийняття рішення про наявність цілі коли її насправді нема, називають [20, 23] **ймовірністю хибних тривог** P_{XT} , а ймовірність помилки II-го роду тобто, прийняття рішення про відсутність цілі коли насправді вона є - **ймовірністю пропуску цілі** $P_{ПЦ}$. Відповідно, величину $P_{ПВ} = 1 - P_{ПЦ}$ називають **ймовірністю правильного виявлення** цілі. У локаційних системах параметр P_{XT} не повинен перевищувати заданого рівня α_0 , який, в залежності від складності та відповідальності системи, звичайно встановлюється у діапазоні 10^{-3} - 10^{-7} . Відповідність поміж термінами математичної статистики та тими, що вживаються в локаційних застосуваннях, показано в табл. 10.1.

Але не слід думати, що сфера застосування критерію Н.-П. обмежується лише системами військового призначення. Існує досить багато цивільних задач, для яких такий підхід також виправданий. Скажімо, у медицині найважливішою задачею є своєчасна постановка правильного діагнозу. Помилки в цій сфері (пропуск небезпечного захворювання або помилкове призначення лікування) також можуть

призводити до важких наслідків. Менш трагічні, але іноді теж дуже серйозні наслідки може мати помилкове прийняття хибної наукової теорії на підставі якихось не досить достовірних експериментальних даних. В усіх цих задачах доцільно вибирати в якості нульової гіпотези те з можливих припущень, помилкове прийняття якого більш небажано, чим його відхилення.

Таблиця 10.1

<i>Математична статистика</i>	<i>Системи локації</i>
Помилка I-го роду	Хибна тривога
Рівень значущості критерію	Рівень хибних тривог системи
Помилка II-го роду	Пропуск цілі
Потужність критерію	Імовірність правильного виявлення цілі

Приклад 10.1. На завершення цього параграфу проілюструємо розглянуті методи статистичних рішень на прикладі задачі перевірки гіпотез (10.1a) та (10.1б) про математичне очікування вибірки. Після підстановки функцій правдоподібності гіпотез (10.1a), (10.1б) у співвідношення (10.10) та з урахуванням співвідношень (10.9), (10.11) приходимо до наступного правила прийняття рішення:

$$\exp\left\{\frac{u^2 - (u - a)^2}{2\sigma^2}\right\} \begin{matrix} H_1 \\ \geq \\ H_0 \end{matrix} \frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)}$$

Дане правило, очевидно, інваріантне до будь-яких монотонних перетворень, що не змінюють відношень порядку. Тому взявши від обох частин цієї нерівності натуральний логарифм, одержимо еквівалентне правило:

$$u^2 - (u - a)^2 \begin{matrix} H_1 \\ \geq \\ H_0 \end{matrix} 2\sigma^2 \cdot \ln\left(\frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)}\right)$$

Після розкриття скобок та перегруповування членів остаточно знаходимо:

$$u \begin{matrix} H_1 \\ \geq \\ H_0 \end{matrix} \Pi, \quad (10.12)$$

$$\text{де} \quad \Pi = \frac{a}{2} + \frac{\sigma^2}{a} \cdot \ln \left(\frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)} \right) \quad (10.13)$$

- поріг прийняття рішення. Отже, оптимальне правило перевірки гіпотез (10.1) зводиться до порівняння одержаної вибірки спостережень u з порогом (10.13). Якщо поріг перевищено, то приймається гіпотеза про те, що математичне очікування дорівнює a , якщо вибірка u нижче порога – рішення про те, що математичне очікування дорівнює нулю.

Запишемо ймовірності помилок I-го та II-го роду правила (10.12). За визначенням помилка I-го роду – це відхилення гіпотези H_0 коли вона насправді вірна, тобто, коли вибірка має густину розподілу імовірності (10.1a). У свою чергу, гіпотеза H_0 відхиляється, коли вибірка u перевищує поріг Π . Отже, ймовірність помилки I-го роду дорівнює інтегралу від розподілу (10.1a) по області (Π, ∞) , тобто,

$$P_I = \frac{1}{\sqrt{2\pi}\sigma} \int_{\Pi}^{\infty} \exp\left\{-\frac{u^2}{2\sigma^2}\right\} du = 1 - \Phi\{\Pi/\sigma\}, \quad (10.14)$$

$$\text{де} \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{y^2}{2}\right) dy \quad (10.15)$$

- інтегральний закон нормального розподілу ймовірностей, який також відомий під назвою інтеграл Лапласа. Ця функція досить детально табульована [9] хоча при практичних обчисленнях замість $\Phi(z)$ частіше використовують іншу табульовану функцію

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-y^2) dy, \quad (10.16)$$

яка носить назву інтеграла ймовірності помилок або функції Крампа [3]. Функції (10.15) та (10.16) пов'язані простими співвідношеннями:

$$\Phi(z) = \frac{1}{2} \left(1 + \text{erf}(z/\sqrt{2}) \right), \quad \text{erf}(z) = 2\Phi(z\sqrt{2}) - 1. \quad (10.17)$$

Аналогічним чином знаходимо ймовірність помилки II-го роду. Для цього треба інтегрувати функцію правдоподібності (10.16) спостережень при гіпотезі H_1 по області, що відповідає рівням нижче порога, тобто

$$P_{II} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\Pi} \exp\left\{-\frac{(u-a)^2}{2\sigma^2}\right\} du = \Phi\{(\Pi-a)/\sigma\} = 1 - \Phi\{(a-\Pi)/\sigma\}. \quad (10.18)$$

Введемо до розгляду нормований параметр

$$q = \frac{a}{\sigma}. \quad (10.19)$$

Використовуючи це позначення та беручи до уваги співвідношення (10.13), (10.14), (10.18), для ймовірностей помилок I-го та II-го роду можна записати:

$$P_I = 1 - \Phi\left\{\frac{q}{2} + \frac{1}{q} \ln\left(\frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)}\right)\right\} \quad (10.20a)$$

$$P_{II} = 1 - \Phi\left\{\frac{q}{2} - \frac{1}{q} \ln\left(\frac{(C_{10} - C_{00}) \cdot P(H_0)}{(C_{01} - C_{11}) \cdot P(H_1)}\right)\right\} \quad (10.20б)$$

З розгляду цих виразів можна зробити наступні висновки. По-перше, при необмеженому зростанні параметру $q \rightarrow \infty$ ймовірності помилок (10.20a) та (10.20б) прямують до нуля. Таким чином, параметр (10.19) характеризує здатність до розрізнення конкуруючих гіпотез H_0 та H_1 : чим цей параметр більший, тим ефективність статистичного критерію вища. Тобто, величина параметру q може бути критерієм надійності статистичних випробувань. По-друге, для $q \gg 1$ ймовірності помилок (10.20a) та (10.20б) приймають приблизно одне й теж саме значення

$$P_I \approx P_{II} \approx 1 - \Phi(q/2) \quad (10.21)$$

що не залежить від вибору штрафної матриці та апіорних ймовірностей гіпотез $P(H_0)$, $P(H_1)$. З іншого боку, те саме значення ймовірності помилок відповідає й критерію ML. Дійсно, як було показано вище, критерій MAP переходить у критерій ML для випадку простої матриці втрат (10.5) та при однакових апіорних імовірностях гіпотез $P(H_0) = P(H_1) = 1/2$. Підстановка цих значень параметрів у співвідношення

(10.20) призводить до формули (10.21), що є підтвердженням асимптотичної оптимальності критерію ML для випадку надійних статистичних випробувань.

На завершення розглянемо цей же приклад з позицій критерію Н.-П. Перш за все необхідно знайти такий поріг прийняття рішення Π , який би забезпечував виконання умови $P_I \leq \alpha_0$, де α_0 - заданий рівень значущості критерію. З урахуванням співвідношення (10.14) ця задача зводиться до вирішення рівняння $1 - \Phi\{\Pi/\sigma\} = \alpha_0$, звідки витікає, що

$$\Pi = \sigma \cdot \Phi^{-1}\{1 - \alpha_0\}, \quad (10.22)$$

де $\Phi^{-1}\{\cdot\}$ - функція, що є зворотною до інтеграла Лапласа (10.15). Деякі значення функції $\Phi^{-1}\{1 - y\}$ наведено в табл.10.2.

Таблиця 10.2

y	10^{-1}	$5 \cdot 10^{-2}$	10^{-2}	$5 \cdot 10^{-3}$	10^{-3}	10^{-4}	10^{-5}
$\Phi^{-1}\{1 - y\}$	1.282	1.645	2.326	2.576	3.090	3.719	4.265

Після підстановки (10.22) у співвідношення (10.18) знаходимо вираз для ймовірності помилки II-го роду

$$P_{II} = 1 - \Phi\{q - \Phi^{-1}\{1 - \alpha_0\}\} \quad (10.23)$$

Як видно з цього співвідношення, й у випадку критерію Н.-П. імовірність помилки II-го роду є монотонно спадаючою функцією параметру (10.19). Цей параметр у теоретичній статистиці не має якоїсь загально вживаної назви, але він добре відомий у прикладних галузях, де його прийнято називати відношенням сигнал/завада по виходу приймальної системи. Попередній розрахунок цього параметру може бути виконано, як правило, із фізичних міркувань, що дозволяє краще планувати статистичний експеримент. Із цією ж метою часто використовують ще один важливий параметр - так зване порогове відношення сигнал/завада $q_{\text{порог.}}$. За визначенням цей параметр є вирішенням рівняння $P_{II}(q_{\text{порог.}}, \alpha_0) = \beta_0$, тобто, $q_{\text{порог.}}$ - це таке відношення сигнал/завада по виходу, яке при заданому рівні ймовірності помилок

I-го роду α_0 забезпечує значення ймовірності помилок II-го роду не вище β_0 . Для прикладу, що розглядається, з формули (10.23) знаходимо:

$$q_{\text{порог.}} = q_{\text{порог.}}(\alpha_0, \beta_0) = \Phi^{-1}(1 - \alpha_0) + \Phi^{-1}(1 - \beta_0). \quad (10.24)$$

Цей параметр відіграє важливу роль у задачах виявлення сигналів оскільки допомагає оцінити умови, що необхідні для забезпечення, наприклад, заданої дальності дії радіо чи гідролокатора. Крім того, порогове відношення сигнал/завада є зручним критерієм для порівняння ефективності систем виявлення сигналів: кращою буде та з них, котра забезпечує менші значення $q_{\text{порог.}}$.

Контрольні запитання та завдання до розділу 10, Частина I.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Що таке статистична гіпотеза? Що таке функція правдоподібності гіпотез? Які бувають гіпотези в залежності від об'єму апіорної інформації? Наведіть приклади.

2. Що таке статистичний критерій та статистичне правило (тест)? Що таке рандомізовані та нерандомізовані критерії? Як можна трактувати побудову нерандомізованого статистичного тесту?

3. Що таке баєсовий ризик? Як називають правила, що мінімізують цей ризик? Яку апіорну інформацію необхідно мати для того, щоб обчислити такий ризик?

4. Що таке апостеріорна та апіорна ймовірності гіпотези? Як у загальному випадку виглядає оптимальне баєсове правило перевірки гіпотез? Яка тестова статистика йому відповідає?

5. З якою характеристикою співпадає баєсовий ризик у випадку простої матриці втрат? Як у цьому випадку виглядає оптимальне баєсове правило статистичного рішення? Що таке критерії MAP та ML? За яких умов вони співпадають?

6. Як виглядає оптимальне баєсове правило для випадку перевірки двох гіпотез? Що таке відношення правдоподібності гіпотез?

7. Що таке критерій Неймана-Пірсона? Чи є цей критерій баєсовим? Що таке ймовірності помилок I-го та II-го роду? Як цей критерій формулюється стосовно локаційних систем?

8. Деякі дані таблиці 10.2 є помилковими. Зробити самостійно їх розрахунок та виявити хибні дані.

Вказівка: скористатися функціями *erf* та *erfinv* середовища Matlab та взяти до уваги співвідношення (10.17).

9. Побудувати характеристику виявлення цілі, тобто, залежність ймовірності правильного виявлення цілі $P_{ПВ}$ від параметру $20 \cdot \lg(q)$ для рівнів хибних тривог $\alpha_0 = 10^{-1}, 10^{-3}, 10^{-5}$. За графіками знайти відповідні порогові відношення сигнал/завада (дБ) для ймовірності $P_{ПВ} = 0.9$. Порівняти одержані значення з результатами розрахунків за формулою (10.24).

Розділ 10. Частина II. Теорія статистичного оцінювання.

10.4. Оптимальне статистичне оцінювання.

У першій частині розділу ми розглядали методи оптимального вибору поміж кількома конкуруючими гіпотезами. Для прикладу була розглянута задача перевірки гіпотез (10.1а), (10.1б), в яких математичне очікування спостережень могло приймати лише два значення: 0 та a . Маючи результат спостережень u , треба було визначитися, який саме варіант математичного очікування реалізується в поточному експерименті. Як було показано на цьому прикладі, задачі такого типу вирішуються фактично шляхом перебору можливих варіантів відповіді: рішення приймається на користь тої відповіді, міра вірогідності якої (апостеріорна ймовірність, функція правдоподібності, тощо) при даних спостереженнях є найбільшою.

У теорії ж статистичного оцінювання схожа задача виглядає наступним чином: математичне очікування a спостережень u наперед невідомо, а за результатами статистичних випробувань це значення треба оцінити (тобто, виміряти). Принципова відмінність цієї постановки від випадку перевірки гіпотез полягає в тому, що тепер параметр a може приймати безперервну множину значень на деякому інтервалі. Тому простий перебір можливих варіантів відповіді для вирішення цієї задачі не завжди є можливим, доцільним або прийнятним.

Отже, у загальному випадку постановка задачі статистичного оцінювання виглядає наступним чином: вектор спостережень $\mathbf{u}$ залежить від деякого вектора інформаційних параметрів $\mathbf{a}$, які при кожному статистичному випробуванні приймають наперед невідомі (тобто, випадкові) значення з деякої множини $\mathbf{a} \in A$. За результатами спостережень треба оцінити (визначити) конкретне значення вектору¹ $\mathbf{a}$.

Для вирішення даної задачі треба побудувати деяке правило оцінювання, яке буде ставити у відповідність кожному вектору спостережень $\mathbf{u}$ значення оцінки невідомих параметрів $\hat{\mathbf{a}}$. Як й у випадку перевірки гіпотез, вважаємо, що це правило є

¹ Нагадаємо, що тут, як і раніше, вектора позначаються малими, а матриці – великими товстими літерами.

нерандомізованим, тобто, існує деяка детермінована залежність $\hat{\alpha} = \hat{\alpha}(\mathbf{u})$ поміж векторами $\mathbf{u}$ та $\hat{\alpha}$. При цьому, як й у випадку перевірки гіпотез, мірою оптимальності правила оцінювання є середній баєсовий ризик. Але на відміну від завдання перевірки гіпотез мірою втрат буде вже не платіжна матриця C_{ik} , а деяка функція втрат $C(\alpha, \hat{\alpha})$. Математичне очікування цієї функції по всьому статистичному ансамблю спостережень та параметрів дає баєсовий ризик для задач статистичного оцінювання параметрів, тобто,

$$R = E\{C(\alpha, \hat{\alpha})\} = \int_{\mathcal{U}} \int_A C\{\alpha, \hat{\alpha}(\mathbf{u})\} \cdot p(\mathbf{u}, \alpha) d\alpha d\mathbf{u}, \quad (10.25)$$

де $E\{\cdot\}$ - оператор безумовного статистичного осереднення² по ансамблям спостережень та параметрів, $p(\mathbf{u}, \alpha)$ - спільна густина розподілу вектору спостережень $\mathbf{u}$ та вектору параметрів α . Будь-які правила оцінювання, що мінімізують ризик (10.25) для тих чи інших функцій втрат $C(\alpha, \hat{\alpha})$, називають баєсовими.

Беручи до уваги формулу множення ймовірностей (див. розділ 3), спільну густину ймовірностей спостережень та інформаційних параметрів можна представити у вигляді добутку

$$p(\mathbf{u}, \alpha) = p(\alpha / \mathbf{u}) \cdot p(\mathbf{u}),$$

де $p(\alpha / \mathbf{u})$ - апостеріорна густина розподілу вектору параметрів, а $p(\mathbf{u})$ - безумовна густина розподілу спостережень. Тоді обчислення повного ризику (10.25) можна представити як двоетапну процедуру. На першому етапі знаходиться умовний ризик

$$R(\hat{\alpha} / \mathbf{u}) = E_A\{C(\alpha, \hat{\alpha}) / \mathbf{u}\} = \int_A C(\alpha, \hat{\alpha}) \cdot p(\alpha / \mathbf{u}) d\alpha, \quad (10.26)$$

який є математичним очікуванням функції втрат по ансамблю інформаційних параметрів $\alpha \in A$ при вже відомих спостереженнях. Тому ризик (10.26) прийнято

² Цей оператор може також позначатися кутовими дужками, тобто, вирази $E\{X\}$ та $\langle X \rangle$ далі будуть розумітися як еквівалентні.

називати *апостеріорним*. Математичне очікування апостеріорного ризику по ансамблю спостережень дає безумовний середній ризик³, тобто,

$$R = E_{\mathcal{U}}\{R(\hat{\mathbf{a}}/\mathbf{u})\} = \int_{\mathcal{U}} R(\hat{\mathbf{a}}/\mathbf{u})p(\mathbf{u})d\mathbf{u} \quad (10.27)$$

Тут $E_{\mathcal{U}}\{\cdot\}$, $E_A\{\cdot/\mathbf{u}\}$ позначають оператори статистичного осереднення у просторах спостережень та інформаційних параметрів відповідно.

Далі зробимо цілком природне припущення, що втрати не можуть перетворюватися у виграш, тобто, $C(\mathbf{a}, \hat{\mathbf{a}}) \geq 0$. У силу своїх загальних властивостей розподіли ймовірностей $p(\mathbf{a}/\mathbf{u})$ та $p(\mathbf{u})$ також не можуть бути від'ємними. Таким чином, апостеріорний ризик (10.26) при будь-яких спостереженнях є невід'ємним. Тому мінімум безумовного ризику (10.27) досягається в тому випадку, коли для кожного вектору спостережень $\mathbf{u}$ апостеріорний ризик $R(\hat{\mathbf{a}}/\mathbf{u})$ буде найменшим із можливих. Отже, мінімізація середнього ризику (10.27) може бути зведена до пошуку мінімуму по $\hat{\mathbf{a}}$ функціоналу (10.26) для кожного вектору спостережень $\mathbf{u}$.

Для вирішення цієї задачі необхідна подальша деталізація властивостей функції втрат $C(\mathbf{a}, \hat{\mathbf{a}})$. Розглянемо спочатку конкретний приклад, який знаходить найбільш широке застосування в прикладних задачах, а саме, квадратичну функцію втрат. У загальному випадку оцінювання M -мірного вектору параметрів ця функція має вигляд відомої в лінійній алгебрі [9] квадратичної форми:

$$\begin{aligned} C(\mathbf{a}, \hat{\mathbf{a}}) &= \sum_{i=1}^M \sum_{k=1}^M (\alpha_i - \hat{\alpha}_i) \cdot Q_{ik} \cdot (\alpha_k - \hat{\alpha}_k) = \\ &= (\mathbf{a} - \hat{\mathbf{a}})^T \cdot \mathbf{Q} \cdot (\mathbf{a} - \hat{\mathbf{a}}). \end{aligned} \quad (10.28)$$

Зауважимо, до речі, що в цьому та наступних виразах вектора розглядаються як матриці-стовпчики відповідної розмірності, а всі операції у векторно-матричних співвідношеннях трактуються за правилами матричної алгебри. Зазначимо також, що оператор $(\cdot)^T$ позначає транспонування матриці, тобто, заміну рядків стовпчиками. Наприклад, якщо $\mathbf{a}$ - матриця-стовпчик, то $\mathbf{a}^T$ - відповідна йому матриця-рядок.

³ Мірність інтегралів у виразах (10.25) - (10.27) визначається мірністю відповідних змінних інтегрування, тобто, векторів $\mathbf{u}$ та $\mathbf{a}$.

Ядро квадратичної форми $\mathbf{Q} = [Q_{ik}]_1^M$ у виразі (10.28) - деяка симетрична матриця коефіцієнтів, на яку накладається лише одне обмеження, а саме, умова так званої **позитивної визначеності**. Суть цієї умови полягає в тому, що квадратична форма, сформована на базі цієї матриці, для будь-якого M - мірного вектору $\boldsymbol{\beta} \neq 0$, повинна задовольняти нерівності

$$\boldsymbol{\beta}^T \cdot \mathbf{Q} \cdot \boldsymbol{\beta} > 0 \quad (10.29)$$

Необхідні співвідношення для мінімуму цільової функції (10.26) можуть бути знайдені з умови рівності нулю її першої похідної по вектору параметрів $\hat{\mathbf{a}}$. Зокрема,

$$\frac{\partial R(\hat{\mathbf{a}} / \mathbf{u})}{\partial \hat{a}_i} = \int_A \frac{\partial}{\partial \hat{a}_i} C(\mathbf{a}, \hat{\mathbf{a}}) \cdot p(\mathbf{a} / \mathbf{u}) d\mathbf{a}$$

$$i = 1, \dots, M.$$

У свою чергу, беручи до уваги симетричність ядра квадратичної форми (10.28), неважко бачити, що

$$\frac{\partial}{\partial \hat{a}_i} C(\mathbf{a}, \hat{\mathbf{a}}) = -2 \sum_{k=1}^M Q_{ik} \cdot (\alpha_k - \hat{a}_k), \quad i = 1, \dots, M..$$

Отже,

$$\frac{\partial R(\hat{\mathbf{a}} / \mathbf{u})}{\partial \hat{a}_i} = -2 \cdot E_A \left\{ \sum_{k=1}^M Q_{ik} \cdot (\alpha_k - \hat{a}_k) \right\}.$$

$$i = 1, \dots, M.$$

Цей вираз більш компактно та наочно можна представити у векторно-матричних позначеннях, а саме:

$$\frac{\partial R(\hat{\mathbf{a}} / \mathbf{u})}{\partial \hat{\mathbf{a}}} = -2\mathbf{Q} \cdot \left(\int_A (\mathbf{a} - \hat{\mathbf{a}}) \cdot p(\mathbf{a} / \mathbf{u}) d\mathbf{a} \right). \quad (10.30)$$

У даному співвідношенні похідну від скалярної функції ризику по векторній змінній слід розуміти як вектор тієї ж розмірності, що й вектор $\hat{\mathbf{a}}$. При цьому k -та позиція цього вектору відповідає похідній по $\hat{a}_k$. Прирівнюючи вираз (10.30) нулю та вирішуючи одержане рівняння стосовно вектору $\hat{\mathbf{a}}$, знаходимо оптимальну баєсову оцінку інформаційних параметрів:

$$\hat{\mathbf{a}}_{MLS} = \int_{\mathbf{A}} \mathbf{a} \cdot p(\mathbf{a}/\mathbf{u}) d\mathbf{a} = E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\}. \quad (10.31)$$

Нижнє позначення оцінки $\hat{\mathbf{a}}_{MLS}$ містить посилання на застосований критерій оптимальності, а саме, мінімум середньоквадратичної похибки вимірювання (від англomовної назви **Mean Least Square error**). Достатньою умовою для того, щоб це рішення відповідало саме точці мінімуму, є позитивна визначеність так званої матриці Гессе [9] цільової функції, яка формується з похідних другого порядку від неї. Оскільки для функції втрат (10.28) матриця Гессе

$$\frac{\partial^2}{\partial \hat{a}_i \partial \hat{a}_k} C(\mathbf{a}, \hat{\mathbf{a}}) = Q_{ik}$$

співпадає з позитивно визначеною матрицею $\mathbf{Q}$, то достатні умови оптимальності оцінки (10.31) виконуються. Таким чином, для будь-якої позитивно визначеної матриці коефіцієнтів $\mathbf{Q}$ оптимальною за критерієм MLS оцінкою вектору $\mathbf{a}$ є його апостеріорне математичне очікування $E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\}$, тобто таке, що обчислюється після одержання спостережень $\mathbf{u}$.

Цей результат має в теоретичній статистиці фундаментальне значення тому що, по-перше, як у теоретичних так й прикладних дослідженнях критерій MLS вживається найчастіше, по-друге, оцінка (10.31) є оптимальною не тільки за критерієм MLS, а й для значно більш широкого класу умов [3], а саме:

1. Функція втрат є симетричною та не спадаючою функцією похибки вимірювання тобто,

$$\begin{cases} C(\mathbf{a}, \hat{\mathbf{a}}) = C(\boldsymbol{\varepsilon}) = C(-\boldsymbol{\varepsilon}), & \boldsymbol{\varepsilon} = \mathbf{a} - \hat{\mathbf{a}} \\ C(\boldsymbol{\varepsilon}_1) \geq C(\boldsymbol{\varepsilon}_2), & \|\boldsymbol{\varepsilon}_1\| > \|\boldsymbol{\varepsilon}_2\| \end{cases} \quad (10.32)$$

2. Апостеріорний розподіл $p(\mathbf{a}/\mathbf{u})$ є симетричною функцією відносно умовного очікування $E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\}$, тобто,

$$p(\mathbf{a} - E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\}/\mathbf{u}) = p(E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\} - \mathbf{a}/\mathbf{u}). \quad (10.33)$$

Отже, ці умови є достатніми для оптимальності оцінки апостеріорного середнього $E_{\mathbf{A}}\{\mathbf{a}/\mathbf{u}\}$ для широкого класу функцій втрат, які не обов'язково співпадають з квадратичною формою (10.28). Якщо ж доповнити умову (10.33) вимогою

унімодальності апостеріорного розподілу (тобто, відповідна густина ймовірностей має один максимум [9]), то з'являється додаткова можливість для подальшого спрощення структури оптимальних баєсових оцінок. Дійсно, у випадку унімодальних симетричних розподілів математичне очікування співпадає з точкою максимуму розподілу. Тому оптимальна баєсова оцінка (10.31) буде також оптимальною й за критерієм максимуму апостеріорного розподілу (MAP), тобто:

$$\hat{\alpha}_{MLS} = \hat{\alpha}_{MAP} = \arg \left\{ \max_{\alpha} p(\alpha / \mathbf{u}) \right\}. \quad (10.34)$$

Подальший крок – це перехід, як і у випадку перевірки гіпотез, до критерію максимальної правдоподібності (ML – Maximum Likelihood). Дійсно, згідно з формулою Баєса апостеріорний розподіл параметра α може бути записано наступним чином:

$$p(\alpha / \mathbf{u}) = \frac{p(\mathbf{u} / \alpha) \cdot p(\alpha)}{p(\mathbf{u})} \quad (10.35)$$

де $p(\mathbf{u} / \alpha)$ - функція правдоподібності параметрів α , $p(\alpha)$ - апіорна густина розподілу параметрів. Якщо апіорний розподіл $p(\alpha)$ є рівномірним, то точки максимуму функцій $p(\alpha / \mathbf{u})$ та $p(\mathbf{u} / \alpha)$ по параметру α , вочевидь, співпадають. Таким чином, у цьому окремому випадку критерій MAP (10.34) є повністю еквівалентним критерію ML:

$$\hat{\alpha}_{ML} = \arg \left\{ \max_{\alpha} p(\mathbf{u} / \alpha) \right\}. \quad (10.36)$$

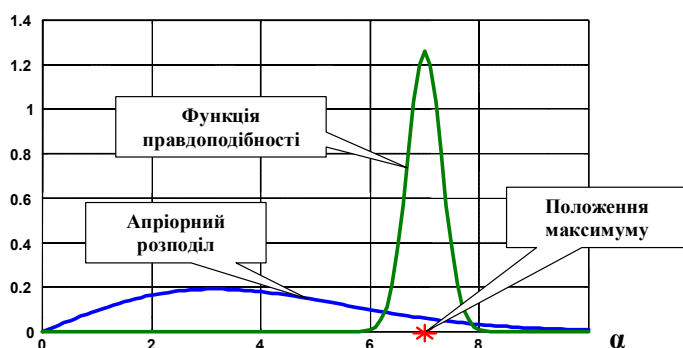


Рис. 10.1. Ілюстрація умов застосування критерію ML

Але, як і випадку перевірки гіпотез, цей критерій може бути застосовано й при довільному апіорному розподілі $p(\alpha)$, якщо функція правдоподібності $p(\mathbf{u} / \alpha)$ значно гостріша за апіорний розподіл $p(\alpha)$, як це, наприклад, показано на рис.10.1. У цьому випадку положення максимумів апостеріорного розподілу

$p(\alpha/\mathbf{u})$ та функції правдоподібності $p(\mathbf{u}/\alpha)$ по параметрах α практично співпадають навіть при нерівномірності апіорного розподілу $p(\alpha)$. Така ситуація відповідає надійним статистичним випробуванням, коли статистична похибка вимірювань значно менша за апіорне розсіювання параметра, що вимірюється.

Отже, з наведеного огляду випливає, що для широкого діапазону умов оптимальною баєсовою оцінкою векторного параметра α є його апостеріорне математичне очікування $E_A\{\alpha/\mathbf{u}\}$. Деяким звуженням оптимального підходу є оцінка (10.34) за критерієм MAP, яка, втім, також належить до класу баєсових. Окремим випадком критерію MAP є оцінка (10.36) за критерієм ML, якому відповідає рівномірний на всій множині A апіорний розподіл параметра α . Цей же метод є оптимальним для випадків надійних статистичних випробувань, коли інформація про апіорний розподіл параметрів, що вимірюються, взагалі не потрібна. Зважаючи на цю обставину, часто застосування методу ML трактується, як особливий випадок статистичних випробувань, при яких інформаційні параметри вважаються хоч наперед і невідомими, але не випадковими. З тієї ж причини оцінки ML часто також називають умовними, щоб підкреслити той факт, що значення вимірюваних параметрів є незмінними. Оцінки ж баєсового типу, в яких використовується апіорний розподіл вимірюваних параметрів, відносять до безумовних. Звернемо також увагу на тісну аналогію поміж теорією оцінювання та теорією перевірки гіпотез. Фактично останню можна розглядати як окремий випадок оцінювання параметра, що приймає дискретний набір значень.

Приклад 10.2. Завершимо розгляд прикладом задачі на побудову оцінок. Нехай спостереження $\mathbf{u} = (u_1, u_2, \dots, u_N)$ мають вигляд:

$$u_k = w_k + \alpha, \quad k = 1, 2, \dots, N,$$

де $w_1, w_2, \dots, w_N$ - відліки статистично незалежних випадкових величин, які мають гавсівський розподіл імовірностей з нульовим математичним очікуванням та дисперсією σ_w^2 . Треба знайти оптимальне правило оцінювання величини α . З умов задачі витікає, що N - мірна функція правдоподібності спостережень має вигляд:

$$p(\mathbf{u}/\alpha) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{1}{2\sigma_w^2} \sum_{k=1}^N (u_k - \alpha)^2\right\}. \quad (10.37)$$

Функція (10.37) досягає максимуму по параметру α в точці, яка відповідає нульовому значенню показника експоненти. Звідси можна відразу записати оптимальну оцінку параметра α за критерієм ML

$$\hat{\alpha}_{ML} = \frac{1}{N} \sum_{k=1}^N u_k, \quad (10.38)$$

котра, як легко бачити, співпадає з вибіркоvim середнім. Неважко показати, що дисперсія та математичне очікування оцінки (10.38) мають вигляд:

$$E\{\hat{\alpha}_{ML}\} = \alpha, \quad \text{var}\{\hat{\alpha}_{ML}\} = E\{(\hat{\alpha}_{ML} - \alpha)^2\} = \sigma_w^2 / N = d_{ML}^2.$$

Тепер припустимо, що параметр α є випадковою гавсівською величиною, тобто, має розподіл ймовірностей

$$p(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{-\frac{(\alpha - m_\alpha)^2}{2\sigma_\alpha^2}\right\}, \quad (10.39)$$

де m_α , σ_α^2 - відповідно його математичне очікування та дисперсія. Використовуючи формулу (10.35), знаходимо апостеріорний розподіл ймовірностей параметра α :

$$p(\alpha/\mathbf{u}) = \frac{1}{\sqrt{2\pi}d_A} \exp\left\{-\frac{1}{2d_A^2}(\alpha - \hat{\alpha}_{MAP}(\mathbf{u}))^2\right\},$$

$$\text{де} \quad \hat{\alpha}_{MAP}(\mathbf{u}) = \frac{1}{1 + \gamma^2} \cdot \alpha_{ML} + \frac{\gamma^2}{1 + \gamma^2} \cdot m_\alpha \quad (10.40)$$

- апостеріорне математичне очікування, яке, таким чином, є оптимальною баєсовою оцінкою параметра α за критеріями MLS та MAP, $\hat{\alpha}_{ML}$ - вибіркoве середнє (10.38), $\gamma^2 = d_{ML}^2 / \sigma_\alpha^2$ - відношення дисперсії оцінки ML до апріорної дисперсії параметру α , $d_A^2 = d_{ML}^2 / (1 + \gamma^2)$ - апостеріорна дисперсія.

Порівняння виразів (10.38) та (10.40) дозволяє побачити різницю між баєсовими оцінками та оцінками ML, а саме, баєсова оцінка (10.40) формується як зважена сума оцінки максимальної правдоподібності $\hat{a}_{ML}$ та апіорного математичного очікування m_a . Вагові коефіцієнти виразу (10.40) залежать від співвідношення $\gamma^2 = d_{ML}^2 / \sigma_a^2$ між дисперсією оцінки ML та апіорною дисперсією інформаційного параметру.

У випадку надійних статистичних випробувань $d_{ML}^2 \ll \sigma_a^2$, ($\gamma^2 \approx 0$) оцінки (10.38) та (10.40) дуже близькі. В іншому граничному випадку, коли $d_{ML}^2 \gg \sigma_a^2$, ($\gamma^2 \rightarrow \infty$), баєсова оцінка практично тотожна апіорному математичному очікуванню параметра m_a , тобто, вимірювання в цьому разі фактично не потрібно.

На завершення розгляду звернемо увагу на ту важливу обставину, що дисперсія $d_{ML}^2 = \sigma_w^2 / N$ оцінки ML при $N \rightarrow \infty$ збігається в даному прикладі до нуля. Отже, необхідні умови надійності даного статистичного експерименту можна, вочевидь, забезпечити відповідним вибором параметра N . Статистичні оцінки, надійність яких завжди може бути забезпечення шляхом збільшення об'єму статистичних даних, називають *слухними* або *конзистентними*.

10.5. Похибки статистичного оцінювання.

Через вплив завад та інших неврахованих факторів оцінки $\hat{\alpha}$ в кожному окремо взятому статистичному експерименті звичайно не співпадають із фактичним значенням вектору інформаційних параметрів α . Отже, похибка оцінювання є випадковою величиною, для опису якої доцільно використовувати різноманітні ймовірнісні характеристики. Однією з таких характеристик є так зване **зміщення (зсув) оцінки** (estimate's bias), яке визначають як різницю між умовним математичним очікуванням оцінки $E_{\mathcal{A}}\{\hat{\alpha} / \alpha\}$ та фактичним значенням параметра, тобто:

$$\mathbf{b}(\alpha) = E_{\mathcal{A}}\{\hat{\alpha} / \alpha\} - \alpha. \quad (10.41)$$

Якщо зміщення оцінки відсутнє, тобто, вектор $\mathbf{b}(\alpha) \equiv 0$, то таку оцінку називають **незміщеною**. Зокрема, до класу незміщених відноситься оцінка (10.38), яка

розглядалася вище в прикладі 10.2. Наявність зміщення оцінки може бути наслідком недосконалості обраної схеми статистичних випробувань. Тому характеристику (10.41) часто також називають систематичною похибкою оцінювання.

Більш загальною характеристикою оцінки є її середньоквадратичне розсіювання відносно фактичного значення параметрів, а саме:

$$d_{\mathbf{a}}^2 = E_{\mathcal{A}} \{ |\hat{\mathbf{a}} - \mathbf{a}|^2 / \mathbf{a} \}. \quad (10.42)$$

Неважко бачити, що цю характеристику можна представити як суму двох додатків:

$$d_{\mathbf{a}}^2 = E_{\mathcal{A}} \{ |\hat{\mathbf{a}} - E_{\mathcal{A}} \{ \hat{\mathbf{a}} / \mathbf{a} \}|^2 / \mathbf{a} \} + |\mathbf{b}(\mathbf{a})|^2.$$

Перший доданок є параметром розсіювання оцінки відносно свого умовного математичного очікування, тобто, її дисперсія. Цей доданок часто називають статистичною похибкою оцінки. Другий доданок – квадрат модуля вектора зміщення оцінки, є мірою її систематичної похибки. У розгорнутому запису даний вираз має вигляд:

$$d_{\mathbf{a}}^2 = \sum_{k=1}^M \text{var} \{ \hat{a}_k / \mathbf{a} \} + \sum_{k=1}^M b_k^2(\mathbf{a}), \quad (10.43)$$

де $\text{var} \{ \hat{a}_k / \mathbf{a} \} = E_{\mathcal{A}} \{ |\hat{a}_k - E_{\mathcal{A}} \{ \hat{a}_k / \mathbf{a} \}|^2 / \mathbf{a} \}$ - дисперсія оцінки k -ого інформаційного параметра. Для незміщених оцінок систематична похибка відсутня, отже у виразі (10.43) залишається перший доданок. Таким чином, для незміщених оцінок джерелом похибок є лише статистичні фактори, які не можуть бути враховані заздалегідь. Але звідси не витікає, що незміщені оцінки мають в усіх випадках беззаперечну перевагу. Усе визначається умовами проведення статистичних випробувань: іноді буває доцільніше допустити деяке зміщення оцінки, якщо за рахунок цього з'являється можливість суттєво зменшити статистичну похибку. Така ситуація виникає, наприклад, при вирішенні так званих зворотних задач, зокрема, томографії деяких середовищ при наявності рефракції хвиль.

Ще однією ймовірнісною характеристикою, яка знаходить застосування у випадку оцінювання векторних параметрів, є кореляційна матриця похибок:

$$\mathbf{K} = \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1M} \\ K_{21} & K_{22} & \cdots & K_{2M} \\ \cdots & \cdots & \cdots & \cdots \\ K_{M1} & K_{M2} & \cdots & K_{MM} \end{bmatrix},$$

де

$$K_{ij} = E\{(\alpha_i - \hat{\alpha}_i) \cdot (\alpha_j - \hat{\alpha}_j)\}$$

- кореляція похибок вимірювання i -го та j -го інформаційних параметрів $(i, j = 1, \dots, M)$. У більш компактних векторно-матричних позначеннях дана кореляційна матриця може бути записана наступним чином:

$$\mathbf{K} = E\{(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \cdot (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T\}, \quad (10.44)$$

де, як і вище, вважається, що операції над векторами виконуються за правилами матричної алгебри.

Кореляційна матриця (10.44) відноситься до характеристик безумовних оцінок, тобто, осереднення тут виконується по всьому ансамблю як спостережень, так і параметрів. Але в прикладних задачах частіше цікавляться не такою усередненою характеристикою, а умовними похибками оцінювання при фіксованих значеннях параметрів $\boldsymbol{\alpha}$. Представимо оператор безумовного математичного осереднення $E\{\cdot\}$ як послідовність оператора умовного осереднення в просторі спостережень $E_{\mathcal{Q}}\{\cdot / \boldsymbol{\alpha}\}$ при заданому значенні $\boldsymbol{\alpha}$ та оператора осереднення в просторі параметрів $E_A\{\cdot\}$. Тоді кореляційна матриця похибок (10.41) може бути представлена у вигляді

$$\mathbf{K} = E_A\{\mathbf{K}_{\boldsymbol{\alpha}}\}.$$

де

$$\mathbf{K}_{\boldsymbol{\alpha}} = E_{\mathcal{Q}}\{(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \cdot (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})^T / \boldsymbol{\alpha}\} \quad (10.45)$$

- умовна кореляційна матриця похибок при фіксованому значенні $\boldsymbol{\alpha}$. Неважко бачити, що слід⁴ цієї матриці співпадає з параметром розсіювання оцінки (10.42), тобто:

⁴ Нагадаємо, що слідом матриці називають суму її діагональних елементів.

$$Tr[\mathbf{K}_a] = E_{\mathcal{U}} \{ |\hat{\mathbf{a}} - \mathbf{a}|^2 / \mathbf{a} \} = d_a^2.$$

Похибки статистичного оцінювання задовольняють дуже важливому співвідношенню, яке називають нерівністю Рао-Крамера [3]. Це співвідношення можна одержати з наступних міркувань. Розглянемо спочатку задачу вимірювання скалярного (тобто, лише одного) інформаційного параметру α . Будемо вважати, що оцінка $\hat{\alpha}$ цього параметра є незміщеною, тобто,

$$E_{\mathcal{U}} \{ \hat{\alpha} / \alpha \} = \int_{\mathcal{U}} \hat{\alpha}(\mathbf{u}) p(\mathbf{u} / \alpha) d\mathbf{u} = \alpha. \quad (10.46)$$

Припустимо, що ця задача задовольняє наступним умовам регулярності [3]:

- область інтегрування $\mathcal{U}$ у виразі (10.46) від значень параметра α не залежить;
- перша похідна від функції правдоподібності $p(\mathbf{u} / \alpha)$ по α існує.

При даних припущеннях похідна від обох частин рівності (10.46) по інформаційному параметру α приводить до наступного співвідношення:

$$\int_{\mathcal{U}} \hat{\alpha}(\mathbf{u}) \cdot \frac{\partial p(\mathbf{u} / \alpha)}{\partial \alpha} d\mathbf{u} = 1.$$

Після тотожних перетворень виразу під знаком інтегралу остаточно маємо:

$$\int_{\mathcal{U}} \hat{\alpha}(\mathbf{u}) \cdot \left(\frac{\partial \ln(p(\mathbf{u} / \alpha))}{\partial \alpha} \right) \cdot p(\mathbf{u} / \alpha) \cdot d\mathbf{u} = 1. \quad (10.47)$$

Функція правдоподібності $p(\mathbf{u} / \alpha)$, як і будь-яка густина розподілу ймовірностей, задовольняє умові нормування по аргументу $\mathbf{u}$:

$$\int_{\mathcal{U}} p(\mathbf{u} / \alpha) d\mathbf{u} = 1$$

При виконанні вказаних вище умов регулярності похідна від обох частин цієї рівності по параметру α приводить до співвідношення:

$$\int_{\mathcal{U}} \left(\frac{\partial \ln(p(\mathbf{u} / \alpha))}{\partial \alpha} \right) \cdot p(\mathbf{u} / \alpha) \cdot d\mathbf{u} = 0. \quad (10.48)$$

Помножимо обидві частини рівності (10.48) на умовне математичне очікування оцінки $E_{\mathcal{U}}\{\hat{\alpha}/\alpha\}$ та віднімемо його від співвідношення (10.47). У результаті одержимо наступний вираз:

$$\int_{\mathcal{U}} (\hat{\alpha}(\mathbf{u}) - E_{\mathcal{U}}\{\hat{\alpha}/\alpha\}) \cdot \left(\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right) \cdot p(\mathbf{u}/\alpha) \cdot d\mathbf{u} = 1,$$

або

$$E_{\mathcal{U}} \left\{ (\hat{\alpha}(\mathbf{u}) - E_{\mathcal{U}}\{\hat{\alpha}/\alpha\}) \cdot \left(\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right) / \alpha \right\} = 1. \quad (10.49)$$

Далі сформуємо матрицю-стовпець

$$\mathbf{x} = \begin{bmatrix} \frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \\ \hat{\alpha} - E_{\mathcal{U}}\{\hat{\alpha}/\alpha\} \end{bmatrix}.$$

Запишемо умовну коваріаційну матрицю вектора $\mathbf{x}$. Беручи до уваги співвідношення (10.49), маємо:

$$\mathbf{C}_{\alpha} = E_{\mathcal{U}} \{ \mathbf{x}(\mathbf{u}) \cdot \mathbf{x}^T(\mathbf{u}) / \alpha \} = \begin{bmatrix} D_{\alpha} & 1 \\ 1 & \text{var}\{\hat{\alpha}/\alpha\} \end{bmatrix}, \quad (10.50)$$

де

$$D_{\alpha} = E_{\mathcal{U}} \left\{ \left(\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right)^2 / \alpha \right\}.$$

Матриця (10.50), як і будь-яка коваріаційна матриця, є невід'ємно визначеною. Тому її детермінант

$$\det(\mathbf{C}_{\alpha}) = D_{\alpha} \cdot \text{var}\{\hat{\alpha}/\alpha\} - 1 \geq 0.$$

Звідси приходимо до нерівності щодо умовної дисперсії оцінки скалярного параметра

$$\text{var}\{\hat{\alpha}/\alpha\} \geq \frac{1}{E_{\mathcal{U}} \left\{ \left(\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right)^2 / \alpha \right\}}, \quad (10.51)$$

яка і є найпростішою формою нерівності Рао-Крамера. Отже, умовна дисперсія будь-якої незміщеної оцінки скалярного параметра не може бути менша за праву

частину нерівності (10.51). Оцінки, для яких співвідношення (10.51) переходить у строгу рівність, називають *ефективними*.

Нерівність (10.51) можна також записати й в іншому вигляді, який часто буває зручнішим для виконання практичних розрахунків. З цією метою повторно візьмемо похідну по α від співвідношення (10.48):

$$\begin{aligned} \frac{\partial}{\partial \alpha} \left(\int_{\mathcal{U}} \left[\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right] \cdot p(\mathbf{u}/\alpha) \cdot d\mathbf{u} \right) &= \int_{\mathcal{U}} \left[\frac{\partial^2 \ln(p(\mathbf{u}/\alpha))}{\partial \alpha^2} \right] \cdot p(\mathbf{u}/\alpha) \cdot d\mathbf{u} + \\ &+ \int_{\mathcal{U}} \left[\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right]^2 \cdot p(\mathbf{u}/\alpha) \cdot d\mathbf{u} = 0. \end{aligned}$$

Звідси витікає, що

$$E_{\mathcal{U}} \left\{ \left[\frac{\partial \ln(p(\mathbf{u}/\alpha))}{\partial \alpha} \right]^2 / \alpha \right\} = E_{\mathcal{U}} \left\{ - \frac{\partial^2 \ln(p(\mathbf{u}/\alpha))}{\partial \alpha^2} / \alpha \right\} \quad (10.52)$$

Зі співвідношень (10.51) та (10.52) маємо альтернативну форму нерівності Рао-Крамера:

$$\text{var}\{\hat{\alpha}/\alpha\} \geq \frac{1}{E_{\mathcal{U}} \left\{ - \frac{\partial^2 \ln(p(\mathbf{u}/\alpha))}{\partial \alpha^2} / \alpha \right\}}. \quad (10.53)$$

Приклад 10.3. Проілюструємо застосування формули (10.53) для задачі оцінювання математичного очікування вибірки, яка розглядалася в прикладі 10.2. Безпосередньо з формули (10.37) маємо:

$$\ln p(\mathbf{u}/\alpha) = -\ln(\sqrt{2\pi}\sigma_w) - \frac{1}{2\sigma_w^2} \sum_{k=1}^N (u_k - \alpha)^2.$$

Після подвійного диференціювання цього виразу по α знаходимо:

$$- \frac{\partial^2 \ln p(\mathbf{u}/\alpha)}{\partial \alpha^2} = \frac{N}{\sigma_w^2}.$$

Оскільки права частина цього виразу не залежить від спостережень, то необхідності в усередненні по простору $\mathcal{U}$ немає. Отже, з нерівності (10.53) випливає, що σ_w^2/N є нижньою границею умовної статистичної похибки оцінювання параметра α . Неважко помітити, що в прикладі 10.2 саме таке значення має статистична похибка

d_{ML}^2 оцінки ML. Таким чином, у даному випадку метод ML забезпечує теоретично максимально можливу точність вимірювання. Цей збіг не є випадковим. У курсах теоретичної статистики доводиться [3], що оцінки ML, якщо вони існують, завжди є, принаймні, асимптотично незміщеними та ефективними, тобто, вони забезпечують похибку на рівні нижньої границі співвідношення (10.54) при необмеженому зростанні об'єму спостережень. Саме ця обставина й пояснюють те велике значення, яке має метод максимальної правдоподібності в теоретичних та прикладних задачах.

На завершення розгляду наведемо нерівність Рао-Крамера для випадку оцінювання векторних параметрів. Подробиці залишаємо для самостійної роботи, оскільки хід доказу є повністю аналогічним наведеному вище. Отже, дисперсія незміщеної оцінки $\hat{\mathbf{a}}$ вектору інформаційних параметрів $\mathbf{a}$ задовольняє наступній нерівності:

$$d_{\mathbf{a}}^2 \geq \text{Tr}[\mathbf{D}_{\mathbf{a}}^{-1}], \quad (10.54)$$

де $\mathbf{D}_{\mathbf{a}}$ - квадратна матриця M -го порядку, яка визначається співвідношенням

$$\mathbf{D}_{\mathbf{a}} = E_{\mathcal{U}}\{\mathbf{L}_{\mathbf{a}}(\mathbf{u}) \cdot \mathbf{L}_{\mathbf{a}}^T(\mathbf{u}) / \mathbf{a}\}, \quad (10.55)$$

де, у свою чергу, $\mathbf{L}_{\mathbf{a}}(\mathbf{u})$ - вектор (матриця-стовпчик) з компонентами

$$(\mathbf{L}_{\mathbf{a}}(\mathbf{u}))_k = \frac{\partial \ln(p(\mathbf{u} / \mathbf{a}))}{\partial \alpha_k}, \quad k = 1, \dots, M. \quad (10.56)$$

Матрицю (10.55) прийнято називати інформаційною матрицею Фішера [17]. Походження цієї назви ми пояснимо трохи пізніше, а поки що зауважимо, що, як й у випадку скалярного параметру, для визначення матриці (10.55) можливе й альтернативне визначення

$$(\mathbf{D}_{\mathbf{a}})_{ik} = E_{\mathcal{U}}\left\{-\frac{\partial^2 \ln p(\mathbf{u} / \mathbf{a})}{\partial \alpha_i \partial \alpha_k} / \mathbf{a}\right\}, \quad (10.57)$$

яке повністю еквівалентне визначенню (10.55). Отже, у випадку оцінювання векторного параметра теоретична границя статистичної похибки визначається властивостями матриці Фішера (10.55) або (10.57). Але чому до цієї назви часто додають прикметник «інформаційна»? Щоб це стало зрозуміло, згадаємо, що

відповідно до матричної алгебри слід довільної неособливої квадратної матриці $\mathbf{Q}$ порядку M може обчислюватися не тільки підсумовуванням її діагональних елементів, але й через так звані власні числа матриці $\lambda_1(\mathbf{Q}), \lambda_2(\mathbf{Q}), \dots, \lambda_M(\mathbf{Q})$, що знаходяться в результаті рішення рівняння

$$\det(\mathbf{Q} - \lambda \cdot \mathbf{I}) = 0.$$

Якщо ці числа знайдені, то слід матриці $\mathbf{Q}$ та зворотної до неї матриці $\mathbf{Q}^{-1}$ визначаються наступними співвідношеннями:

$$Tr[\mathbf{Q}] = \sum_{k=1}^M \lambda_k(\mathbf{Q}), \quad Tr[\mathbf{Q}^{-1}] = \sum_{k=1}^M \frac{1}{\lambda_k(\mathbf{Q})}, \quad (10.58)$$

Таким чином, на підставі співвідношень (10.54) та (10.58) одержуємо наступну нерівність для умовної середньоквадратичної похибки незміщених оцінок:

$$d_{\mathbf{a}}^2 \geq \sum_{k=1}^M \frac{1}{\lambda_k(\mathbf{D}_{\mathbf{a}})}, \quad (10.59)$$

де $\lambda_1(\mathbf{D}_{\mathbf{a}}), \lambda_2(\mathbf{D}_{\mathbf{a}}), \dots, \lambda_M(\mathbf{D}_{\mathbf{a}})$ - власні числа матриці Фішера, які знаходяться із рівняння

$$\det(\mathbf{D}_{\mathbf{a}} - \lambda \cdot \mathbf{I}) = 0. \quad (10.60)$$

Сума величин у правій частині нерівності (10.59), зворотних до власних чисел матриці Фішера, дає нижню границю квадрату умовної середньоквадратичної похибки незміщеної оцінки параметрів $\mathbf{a}$. Як й у випадку скалярного параметра, для ефективних оцінок нерівність (10.59) переходить у рівність.

З нерівності (10.59) випливає, що значення похибки вимірювання, яке можна принципово досягти, тим менше, чим більші власні числа $\lambda_1, \lambda_2, \dots, \lambda_M$ матриці Фішера. Оскільки власні числа визначають спектральну норму матриці, то, таким чином, норма матриці Фішера є мірилом успішності або інформаційності того чи іншого статистичного експерименту. Саме тому цю матрицю часто й називають інформаційною [3, 17].

Границі Рао-Крамера дають мінімально можливі похибки для оптимальних методів оцінювання, які в повній мірі часто не можуть бути реалізовані практично. Але на етапі проектування вимірювальної системи цей аспект і не є головним:

важливо, наприклад, раціонально вибрати взаємне розташування випромінюючих та приймальних пристроїв, характеристики антен, вид сигналу та його параметри, об'єм статистичних спостережень, тощо. Про доцільність використання тієї чи іншої схеми вимірювання можна, зокрема, судити на основі властивостей інформаційної матриці Фішера, яка й визначає статистичну компоненту похибок. Кількісною мірою властивостей цієї матриці є так зване число обумовленості $\mathcal{N} = \lambda_{\max} / \lambda_{\min}$, де $\lambda_{\max}$, $\lambda_{\min}$ - відповідно максимальне та мінімальне власні числа матриці. Якщо число $\mathcal{N}$ відносно невелике або помірно велике, то умови для вимірювання сприятливі або, принаймні, задовільні. Якщо ж $\mathcal{N}$ дуже велике (наприклад, 10^5 та більше), задовільну точність одержати досить складно, а часто й практично неможливо. Подібні ситуації можуть виникати в задачах, де спостереження та інформаційні параметри пов'язані якоюсь складною нелінійною залежністю, внаслідок чого матриця Фішера виявляється близькою до сингулярної. У таких випадках говорять, що задача погано обумовлена (або некоректно поставлена) та потребує так званої регуляризації [17]. Якщо ж і регуляризація не дає результату, то це може означати взагалі практичну неможливість вирішення такої задачі. Тому при плануванні статистичних експериментів попередній аналіз властивостей матриці Фішера має важливе значення.

Контрольні запитання та завдання до розділу 10. Частина II.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. У чому принципова відмінність задач статистичного оцінювання параметрів від задач перевірки гіпотез? Чим у цих випадках відрізняється обчислення байєсового ризику?
2. Що таке апостеріорний ризик? Чому він є невід'ємним?
3. За яких умов оптимальною байєсовою оцінкою є апостеріорне середнє? Що таке критерії MLS, MAP, ML? Які з цих критеріїв відносяться до байєсових?
4. Що таке зміщення оцінки? Внаслідок яких причин воно виникає? Наведіть приклади зміщених та незміщених оцінок.
5. Як визначається дисперсія оцінки? За яких умов вона співпадає з параметром розсіювання (10.42)?
6. Що таке нерівність Рао-Крамера? Яке воно має теоретичне та практичне значення? Що таке ефективні оцінки? Наведіть приклади.
7. Що таке матриця Фішера? Чому її називають інформаційною? Яке значення мають властивості цієї матриці для планування задач статистичного оцінювання?
8. Діючи за аналогією з випадком оцінювання скалярного параметру довести нерівність Рао-Крамера (10.54) для оцінювання векторних параметрів.
9. Довести еквівалентність виразів (10.55) та (10.57).

Розділ 11. Функціонал відношення правдоподібності спостережень для завдань виявлення сигналів.

У попередніх розділах були розглянуті фізичні постановки та основні теоретичні засади, на які спирається статистична теорія приймання сигналів. Тепер настав час поєднати розглянуті вище окремі «компоненти» та перейти безпосередньо до практичного вирішення деяких завдань. З методичних міркувань такий розгляд доцільно розпочати зі задачі виявлення сигналів, яка в загальній проблематиці приймання та обробки сигналів посідає одне з чільних місць. Це дозволить нам одержати загальні вирази для структури оптимального оброблення полів сигналів, які потім знадобляться й для вирішення інших завдань.

11.1. Статистична постановка завдань приймання полів сигналів.

Практично у будь-якій системі прийому ще до етапу виділення якоїсь інформації треба, як правило, зафіксувати сам факт наявності корисного сигналу. В локаційних системах (радіотехнічних, гідроакустичних, тощо) завдання виявлення взагалі є головними. В інших системах, зокрема, телекомунікаційних, такі завдання відіграють хоч й допоміжну, але важливу роль. Крім того, не зважаючи на удавану простоту задачі виявлення, її рішення досить часто пов'язане з виконанням складної та багатоетапної обробки, ядро котрої виявляється потім корисним й для вирішення багатьох інших задач, наприклад, вимірювання параметрів сигналу.

Отже, почнемо наш розгляд із рівняння спостережень (тобто, моделі вхідного сигналу), яке стосовно завдання виявлення сигналу та відповідно до матеріалів розділу 9, має наступний вигляд:

$$u(t, \mathbf{r}) = \theta \cdot s(t, \mathbf{r}) + z(t, \mathbf{r}) + w(t, \mathbf{r}), \quad t \in (t_0, t_0 + T), \quad \mathbf{r} \in \mathbf{D}_A.$$

Тут θ - формальний параметр ситуації, що дорівнює нулю ($\theta = 0$) при відсутності сигналу та одиниці ($\theta = 1$) при його наявності, t_0 - момент можливого початку приходу сигналу, T - тривалість сигналу, $\mathbf{D}_A$ - область

простору, яка охоплюється датчиками приймальної антени. Нагадаємо, що у цьому співвідношенні просторово-часові функції (тобто, поля) $s(t, \mathbf{r})$, $z(t, \mathbf{r})$, $w(t, \mathbf{r})$ позначають корисний сигнал, зовнішні завади та внутрішні шуми приймального пристрою відповідно.

Наряду з розгорнутими будемо надалі використовувати й векторно-матричні форми запису. Зокрема, рівняння спостережень в цих позначеннях виглядає наступним чином:

$$\mathbf{u} = \theta \cdot \mathbf{s} + \mathbf{z} + \mathbf{w}, \quad (11.1)$$

де $\mathbf{s}, \mathbf{z}, \mathbf{w}$ – вектори сигналу, зовнішніх завад та внутрішніх шумів приймальної системи відповідно. Вираз (11.1) має досить загальний вигляд, який дозволяє робити попередній аналіз завдань й без деталізації внутрішньої структури векторів. Але при переході до практичних задач треба, як правило, встановлювати певний зв'язок поміж елементами векторів, що складають рівняння (11.1), та значеннями відповідних фізичних полів. Щодо цього тут і надалі будемо дотримуватися наступної *угоди*, а саме, вважаємо, що усі ці вектори мають блочну структуру, в якій перший блок із N елементів відповідає першому датчику антени для N моментів часу, другий блок із N елементів – другому датчику і т.д. Усього кожен такий вектор містить $N_0 = M \cdot N$ елементів, де M - загальна кількість датчиків. При цьому поміж каналами антени можлива, взагалі кажучи, деяка затримка відліків у часі, тобто, сітки моментів часу для однойменних блоків різних каналів можуть поміж собою не співпадати. Тому додаткові уточнення щодо моделі формування спостережень дамо згодом.

З рівняння спостережень (11.1) легко бачити, що постановка задачі виявлення може бути сформульована у формі перевірки двох гіпотез:

$$\begin{aligned} H_0 : & \text{параметр } \theta = 0 \text{ (сигнал відсутній);} \\ H_1 : & \text{параметр } \theta = 1 \text{ (сигнал присутній).} \end{aligned} \quad (11.2)$$

Задача перевірки двох гіпотез була розглянута в §10.3, де було показано, що оптимальне правило статистичного рішення має вигляд:

$$\lambda(\mathbf{u}) \underset{H_0}{\overset{H_1}{>}} \Pi_0, \quad (11.3)$$

$$\text{де} \quad \lambda(\mathbf{u}) = \frac{p(\mathbf{u} / H_1)}{p(\mathbf{u} / H_0)} \quad (11.4)$$

- відношення правдоподібності гіпотез, а Π_0 - порогова константа, яка залежить від прийнятого критерію ефективності. Отже, оптимальне правило виявлення сигналу полягає в обчисленні функціоналу (11.4) та порівнянні одержаного значення із порогом. Якщо це значення перевищує поріг, то приймається рішення на користь гіпотези про присутність сигналу, в протилежному випадку – про його відсутність.

Для обчислення функціоналу (11.4) необхідно задати статистичні характеристики спостережень для обох гіпотез. Це завдання детально обговорювалось у розділі 9. Зокрема, було встановлено, що розподіл зовнішніх завад та внутрішніх шумів для переважної більшості практичних задач можна вважати гавсівським. Що ж стосується моделей поля сигналу, то вони можуть бути як феноменологічними¹, так й сформованими на основі рішення динамічних рівнянь для відповідних середовищ їх передачі (радіоефір, пружне механічне середовище, хвилеводи, тощо). У цих моделях полів сигналу можуть бути присутні деякі випадкові фактори, що залежать від фізичного змісту задачі та наявної апіорної інформації. З цієї точки зору прийнято виділяти 3 класи задач виявлення, які розглядаються нижче.

¹ Феноменологічними називають такі моделі, які базуються на загальних фізичних уявленнях та практичному досвіді без строгого математичного обґрунтування.

11.2. Приймання полів наперед відомих (детермінованих) сигналів.

В задачах даного класу вектор сигналу $\mathbf{s}$ у моделі спостережень (11.1) вважається точно відомим, а невідомим є лише «значення» параметру ситуації θ . Може виникнути слушне запитання: наскільки припущення про наявність повної апіорної інформації щодо вигляду сигналу на приймальній апертурі реалістичне та корисне? Дійсно, знайти практичну задачу, в якій вигляд сигналу на вході приймальної системи можна було б вважати наперед відомим, дуже складно, якщо взагалі можливо. Але подібні моделі безумовно мають велике теоретичне значення, оскільки, по-перше, дозволяють встановити оцінки потенціальної ефективності систем виявлення у найбільш сприятливих умовах, по-друге, структури обробки наперед відомого сигналу часто дають «підказку» для побудови на їх основі більш реалістичних версій з адаптацією під невідомі параметри сигналу.

При точно відомому (або детермінованому) сигналі розподіл вектора спостережень (11.1) повністю визначається статистичними характеристиками шумів та завад. Вектора $\mathbf{w}$ та $\mathbf{z}$ за прийнятими вище припущеннями є гавсівськими, а сигнальний вектор $\mathbf{s}$ є точно відомим. Отже, вектор спостережень $\mathbf{u}$ також має нормальний розподіл ймовірностей. Залишилось лише визначити параметри цього розподілу. Як зазначалось вище, математичні очікування шумів та завад вважаються нульовими. Тоді з моделі спостережень (11.1) витікає, що математичне очікування вектору $\mathbf{u}$ визначається лише сигнальним вектором, тобто,

$$E\{\mathbf{u} / \theta\} = \theta \cdot \mathbf{s},$$

а кореляційна функція спостережень має вигляд

$$\begin{aligned} K_u(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2) &= E\{(u(t_1, \mathbf{r}_1) - \theta \cdot s(t_1, \mathbf{r}_1)) \cdot (u(t_2, \mathbf{r}_2) - \theta \cdot s(t_2, \mathbf{r}_2))\} = \\ &= g_w \cdot \delta(t_1 - t_2) \cdot \delta(\mathbf{r}_1 - \mathbf{r}_2) + K_z(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2), \end{aligned}$$

де $K_z(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2)$ - кореляційна функція поля завад $z(t, \mathbf{r})$. Нагадаємо, що перший доданок цього виразу відповідає АКФ власних шумів приймальної системи, які традиційно вважаються некорельованими як у часі, так й у

просторі. Вказану обставину відображає добуток дельта-функцій з часовими та просторовими аргументами, параметр g_w позначає СЦП власних шумів.

Одержані статистичні характеристики вектору $\mathbf{u}$ дають можливість записати функцію та відношення правдоподібності спостережень. Для цього скористаємося загальною формулою багатовимірного гавсівського розподілу (3.28), яку представимо у векторно-матричній формі. Отже, беручи до уваги прийняті вище статистичні характеристики шумів та завад, маємо умовний N_0 - вимірний розподіл вектору спостережень $\mathbf{u}$ для фіксованої гіпотези стосовно параметру ситуації θ :

$$p(\mathbf{u} / \theta) = \frac{1}{(2\pi)^{N/2} \sqrt{\det(g_w \mathbf{E}_{N_0} + \mathbf{K}_z)}} \times \exp \left\{ -\frac{1}{2} (\mathbf{u} - \theta \cdot \mathbf{s})^T \cdot (g_w \mathbf{E}_{N_0} + \mathbf{K}_z)^{-1} \cdot (\mathbf{u} - \theta \cdot \mathbf{s}) \right\}, \quad \theta = 0, 1. \quad (11.5)$$

Тут $\mathbf{E}_{N_0}$ - одинична матриця порядку N_0 , діагональні елементи якої дорівнюють одиниці, а усі інші елементи є нульовими, $\mathbf{K}_z = E\{\mathbf{z} \cdot \mathbf{z}^T\}$ - коваріаційна матриця вектору завад². Зазначимо, що ядро квадратичної форми у показнику експоненти виразу (11.5) є матрицею, зворотною до коваріаційної матриці суміші шумів та завад $g_w \mathbf{E}_N + \mathbf{K}_z$. Завдяки зробленому у розділі 9 припущенню про постійну наявність у складі спостережень білого шуму ця матриця завжди є позитивно визначеною, що гарантує коректність операції її обернення.

Зауважимо також, що внаслідок прийнятої вище угоди щодо порядку слідування елементів векторів спостережень матрицю $\mathbf{K}_z$ можна представити у наступному блочному вигляді:

$$\mathbf{K}_z = \begin{bmatrix} \mathbf{K}_z^{(1,1)} & \mathbf{K}_z^{(1,2)} & \dots & \mathbf{K}_z^{(1,M)} \\ \mathbf{K}_z^{(2,1)} & \mathbf{K}_z^{(2,2)} & \dots & \mathbf{K}_z^{(2,M)} \\ \dots & \dots & \dots & \dots \\ \mathbf{K}_z^{(M,1)} & \mathbf{K}_z^{(M,2)} & \dots & \mathbf{K}_z^{(M,M)} \end{bmatrix},$$

² Нагадаємо, що математичне очікування завад вважається нульовим.

де $\mathbf{K}_z^{(i,j)}$ - матриця кореляцій $N \times N$ елементів i -го та j -го блоків $(i, j = 1, \dots, M)$ вектору завад $\mathbf{z}$. Підставляючи розподіл (11.5) до виразу (11.4), знаходимо відношення правдоподібності гіпотез даної задачі:

$$\lambda(\mathbf{u}) = \frac{p(\mathbf{u}/\theta=1)}{p(\mathbf{u}/\theta=0)} = \exp\left\{V_0(\mathbf{u}) - \frac{1}{2}V(\mathbf{s})\right\}, \quad (11.6)$$

$$\text{де} \quad V_0(\mathbf{u}) = (\mathbf{b}, \mathbf{u}) = \mathbf{b}^T \cdot \mathbf{u} \quad (11.7)$$

- достатня статистика задачі, яка знаходиться як скалярний добуток³ вектора оптимальних вагових коефіцієнтів $\mathbf{b}$ на вектор спостережень $\mathbf{u}$. У свою чергу, вектор оптимальних коефіцієнтів $\mathbf{b}$ є рішенням матричного рівняння

$$(\mathbf{g}_w \mathbf{E}_N + \mathbf{K}_z) \cdot \mathbf{b} = \mathbf{s}. \quad (11.8)$$

Значення функціоналу правдоподібності (11.6) для прийнятої реалізації спостережень $\mathbf{u}$ порівнюється з порогом й за результатами порівняння приймається рішення про наявність або відсутність сигналу.

Як вже неодноразово згадувалося, правило прийняття рішення не змінює будь-яке монотонно зростаюче нелінійне перетворення тестової статистики. Тому беручи від виразу (11.6) натуральний логарифм, одержуємо більш зручну еквівалентну форму оптимального правила виявлення точно відомого сигналу:

$$V_0(\mathbf{u}) = (\mathbf{b}, \mathbf{u}) \underset{H_0}{\overset{H_1}{>}} \Pi, \quad (11.9)$$

де Π - деякий еквівалентний поріг виявлення, який увібрав до себе усі величини, що не залежать від вектору спостережень $\mathbf{u}$.

Припустимо, що зовнішні завади відсутні ($\mathbf{K}_z \equiv 0$), тобто, фактором, що заважає, є лише білий шум. Тоді з виразу (11.8) витікає, що вектор оптимальних коефіцієнтів з точністю до постійного множника дорівнює

³ Тут скалярний добуток деяких векторів $(\mathbf{x}, \mathbf{y})$ та векторно-матрична операція $\mathbf{x}^T \cdot \mathbf{y}$ розглядаються як тотожні вирази.

вектору сигналу, а саме, $\mathbf{b} = \mathbf{s} / g_w$. У цьому разі оптимальна обробка (11.7) зводиться до обчислення скалярного добутку вектору спостережень на вектор очікуваного сигналу, тобто, $V_0(\mathbf{u}) = (\mathbf{s} / g_w, \mathbf{u})$. Для кращої фізичної наочності цей скалярний добуток доцільно записати у розгорнутому (по координатному) вигляді. З цією метою припустимо, що затримок при взятті відліків у різних каналах антени не має, тобто, моменти дискретизації спостережень у часі від номеру каналу не залежать⁴. Тоді беручи до уваги прийняту угоду щодо формування блоків спостережень, для виходу оптимальної обробки маємо наступний розгорнутий запис:

$$V_0(\mathbf{u}) = (\mathbf{s} / g_w, \mathbf{u}) = \frac{1}{g_w} \sum_{i=1}^N \sum_{m=1}^M s(t_i, \mathbf{r}_m) \cdot u(t_i, \mathbf{r}_m) \quad (11.10)$$

Даний вираз відповідає обчисленню кореляції (тобто, схожості) поміж полем спостережень на приймальній апертурі та очікуваним сигналом. Обробку такого типу прийнято називати **кореляційним прийманням** [3].

У загальному ж випадку спостереження корелюються не з самим сигналом, а з оптимальними коефіцієнтами $b_0(t_i, \mathbf{r}_m)$, вектор яких знаходиться зі співвідношення (11.8) і, таким чином, залежить від кореляційних властивостей завад. Більш детально особливості оптимальної обробки для загального випадку корельованих завад ми розглянемо у наступних розділах посібника. Поки ж зазначимо, що область застосування кореляційної обробки (11.10) виходить далеко за межі моделі білого шуму. Часто її застосовують й у випадку зовнішніх завад, які слабо корельовані у просторі та мають у робочій смузі частот приблизно рівномірну СЦП. Важливою рисою такої обробки є врахування особливостей поля сигналу при його передачі через середовище зі складними умовами поширення, зокрема, багатопроменевим. Тому обробку (11.10) часто також називають **узгодженою з полем сигналу** [7,25].

⁴ Умови, за яких можливе дане припущення, будуть обговорені у наступному розділі.

Більш детально питання побудови узгодженої обробки будуть розглянуті у розділі 13, а поки зауважимо, що функціонал оптимальної обробки (11.7) та його окремий випадок (11.10) записані у формі, яка передбачає використання дискретних відліків сигналу в часі та просторі. Цей так би мовити «дискретний» підхід в сучасній теорії та практиці приймання та обробки сигналів є домінуючим, тобто, дані та параметри обробки найчастіше розглядаються як вектори скінченої розмірності. Відповідно, прийнято широко використовувати й векторно-матричні позначення та операції. Загалом такий підхід є виправданим, оскільки він максимально відповідає принципам роботи сучасних апаратно-програмних засобів цифрової обробки сигналів. Але, наразі з тим, повністю відкидати традиційний «аналоговий» підхід, в якому дані та параметри обробки є неперервними величинами, недоцільно. По-перше, традиційний підхід має беззаперечні переваги в плані фізичної наочності та є більш інтуїтивно зрозумілим. По-друге, в рамках аналогових представлень можна значно спростити деякі перетворення алгоритмів обробки у порівнянні з векторно-матричними операціями. Особливо це стосується так званих «крайових» ефектів. По-третє, у більшості випадків зберігається подібність аналогових та дискретних операцій, яка дозволяє достатньо легко переходити від одного представлення до іншого та навпаки. Але при цьому завжди доречно перевіряти можливість та адекватність подібних переходів, оскільки певні нюанси для кожної форми представлення все ж таки мають місце.

Зокрема, для векторів нескінченної мірності, як відомо [6], не існує поняття густини розподілу ймовірностей. Тому функціонал правдоподібності $\lambda(\mathbf{u})$ у цьому випадку обчислюють як відношення так званих імовірнісних мір $dP(\mathbf{u} / \theta = 1) / dP(\mathbf{u} / \theta = 0)$, яке прийнято також називати похідною Радона-Нікодіма [6]. Ці обчислення пов'язані з певними математичними тонкощами, але у кінцевому підсумку знову ж таки приходять до загального виразу (11.6) з тією лише різницею, що у визначенні скалярного добутку векторів суми

слід замінити відповідними інтегралами. Це стосується, взагалі кажучи, як часової, так і просторових координат. Але зручнішим для нашого розгляду буде, так би мовити, «змішаний» варіант, а саме, просторову апертуру залишаємо дискретною, а неперервними будемо вважати лише відліки сигналів у часі (нагадаємо, що антени, які складаються з дискретних датчиків, прийнято називати антенними решітками⁵ – АР).

Отже, у цьому «змішаному варіанті» алгоритм оптимального виявлення детермінованих сигналів (11.9) приймає вигляд

$$V_0(\mathbf{u}) = (\mathbf{b}, \mathbf{u}) = \sum_{m=1}^M \int_{\mathbf{T}_m} b(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt \underset{H_0}{\overset{H_1}{>}} \Pi, \quad (11.11)$$

де $\mathbf{r}_m$ - координати m -го датчика антени, $t \in \mathbf{T}_m$ - область існування сигналу на m -ому датчику. При цьому вагові функції $b(t, \mathbf{r}_m)$ знаходяться в результаті рішення системи інтегральних рівнянь, яка є неперервним аналогом матричного рівняння (11.8), а саме:

$$g_w \cdot b(t, \mathbf{r}_m) + \sum_{l=1}^M \int_{\mathbf{T}_l} K_z(t, t'; \mathbf{r}_m, \mathbf{r}_l) \cdot b(t', \mathbf{r}_l) dt' = s(t, \mathbf{r}_m); \quad (11.12)$$

$$m = 1, \dots, M.$$

В окремому випадку приймання сигналу на тлі білого шуму з цих рівнянь витікає, що оптимальна вагова функція обробки з точністю до масштабу співпадає з полем сигналу, тобто:

$$b(t, \mathbf{r}_m) = \frac{1}{g_w} s(t, \mathbf{r}_m).$$

Тоді ліва частина виразу (11.11) переходить у неперервний аналог узгодженої кореляційної обробки (11.10):

$$V_0(\mathbf{u}) = \frac{1}{g_w} (\mathbf{s}, \mathbf{u}) = \frac{1}{g_w} \sum_{m=1}^M \int_{\mathbf{T}_m} s(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt d\mathbf{r}, \quad (11.13)$$

⁵ Іноді замість слова решітки вживається термін «ґратки».

Співвідношення (11.11) – (11.13) для неперервних часових апертур можуть бути застосовані при виконання певних математичних умов [25]. Але з фізичної точки зору для виконання цих умов достатньо, по-перше, обмеженості енергії реальних сигналів та завад, по-друге, постійної наявності у складі спостережень білого шуму. Тому у подальшому будемо, в разі необхідності, користуватися як дискретними, так й неперервними версіями алгоритмів обробки без додаткових застережень.

11.3. Приймання полів квазідетермінованих сигналів.

В задачах цього класу вважається, що поле сигналу $s(t, \mathbf{r})$ відоме лише з точністю до деякої сукупності параметрів $\boldsymbol{\eta}$. Спільний апіорний розподіл цих параметрів $p_{\eta}(\boldsymbol{\eta})$ може бути задано або він може залишатися невідомим. Така ситуація є типовою, наприклад, для локаційних систем. В процесі випромінювання зондувальних сигналів, їх розповсюдження у зовнішньому середовищі, відбиття від об'єктів локації, повернення до приймальної антени сигнал проходить досить складні перетворення, в результаті яких можуть змінюватись наперед невідомим чином його амплітуда, затримка, початкова фаза, несуча частота та форма сигналу.

Теоретична статистика пропонує [3, 6] для таких задач дві головних технології їх вирішення: узагальнений метод максимальної правдоподібності та узагальнений баєсовий підхід. У першому випадку рекомендується проводити вимірювання невідомих параметрів сигналу, оцінка яких $\hat{\boldsymbol{\eta}}$ надалі використовується для обчислення відношення правдоподібності $\lambda(\mathbf{u} / \hat{\boldsymbol{\eta}})$. Зауважимо, що цей метод у технічних застосуваннях частіше називають адаптацією приймальної системи.

При баєсовому підході відношення правдоподібності знаходиться шляхом усереднення умовного відношення правдоподібності $\lambda(\mathbf{u} / \boldsymbol{\eta})$ по невідомим параметрам $\boldsymbol{\eta}$ із густиною $p_{\eta}(\boldsymbol{\eta})$. Можливе й комбінування цих

двох підходів, тобто, для частини невідомих параметрів використовується адаптація, а для інших – статистичне усереднення.

Зі сказаного витікає, що для будь-якого з цих методів початковий крок синтезу алгоритмів виявлення сигналів полягає в тому, що необхідно визначити умовну функцію правдоподібності (умовний розподіл) спостережень $p(\mathbf{u}/\boldsymbol{\eta})$ при фіксованому векторі $\boldsymbol{\eta}$. Лише потім може бути застосована адаптація або баєсовий підхід. У свою чергу, вирішення цієї початкової задачі потребує визначення функціональної залежності сигналу від невідомих параметрів, тобто, $\mathbf{s} = \mathbf{s}(\boldsymbol{\eta})$. Це завдання, взагалі кажучи, не має універсального рішення, оскільки різноманіття практичних ситуацій надто велике. Тому обмежимо поки що наш розгляд деякими простими моделями, які прийнято вважати канонічними. Усі ці моделі виходять з того, що сигнал, який випромінюється джерелом, є квазігармонійним, тобто, має загальний вигляд:

$$s_{\text{дж.}}(t) = A(t) \cdot \cos(\omega_0 t + \Phi(t)), \quad (11.14)$$

де $A(t)$ та $\Phi(t)$ – закони амплітудної та фазової модуляції сигналу відповідно, ω_0 – центральна частота. У найпростішому випадку дія середовища передачі на сигнал зводиться до його затримки у часі, зміни амплітуди, частоти та фази сигналу. Отже, сигнал по виходу каналу передачі, тобто, на вході приймальної АР, буде мати вигляд

$$\begin{aligned} s_{\text{ch}}(t/\varepsilon_{\text{ch}}, \tau_{\text{ch}}, \omega_{\text{ch}}, \varphi_{\text{ch}}) = \\ = \varepsilon_{\text{ch}} \cdot A(t - \tau_{\text{ch}}) \cdot \cos(\omega_{\text{ch}}(t - \tau_{\text{ch}}) + \Phi(t - \tau_{\text{ch}}) + \varphi_{\text{ch}}), \end{aligned} \quad (11.15)$$

де ε_{ch} – коефіцієнт передачі каналу за амплітудою, τ_{ch} – затримка сигналу в каналі, ω_{ch} – центральна частота, φ_{ch} – фаза прийнятого сигналу. Розглянемо декілька моделей, які базуються на певних припущеннях щодо вказаних параметрів сигналу.

1. Канал із випадковою фазою сигналу.

Нехай серед параметрів сигналу на вході приймального пристрою випадковою є лише його початкова фаза φ_{ch} , а всі інші параметри наперед відомі. Тоді сигнал (11.15) можна представити у вигляді

$$s_{ch}(t, \mathbf{r}_m / \varphi_{ch}) = \operatorname{Re}\{\dot{s}_0(t, \mathbf{r}_m) \cdot \exp(j\varphi_{ch})\}, \quad (11.16)$$

де

$$\dot{s}_0(t, \mathbf{r}_m) = \varepsilon_{ch} \cdot A(t - \tau_{ch}, \mathbf{r}_m) \cdot \exp\{j(\omega_{ch}(t - \tau_{ch}) + \Phi(t - \tau_{ch}, \mathbf{r}_m))\} \quad (11.17)$$

— комплексне⁶ (аналітичне) представлення прийнятого сигналу, яке відповідає наперед відомим значенням його параметрів при нульовій початковій фазі. Відповідно до наведеної вище загальної методики на першому етапі вирішення задачі треба записати умовну функцію правдоподібності спостережень для деякого фіксованого значення невідомого параметра φ_{ch} , тобто, $p(\mathbf{u} / \varphi_{ch})$. Це завдання принципово нічим не відрізняється від розглянутого у попередньому параграфі випадку приймання детермінованого сигналу. Єдина відмінність полягає в тому, що у §11.2 була записана не функція, а відношення правдоподібності $\lambda(\mathbf{u})$. Але для задач з невідомими параметрами необхідно, взагалі кажучи, мати саме функцію правдоподібності $p(\mathbf{u} / \varphi_{ch})$. Однак, ця відмінність у даному разі не має суттєвого значення, оскільки відношення та функція правдоподібності відрізняються, вочевидь, лише нормуванням, яке для прийнятої моделі каналу від невідомого параметру φ_{ch} не залежить. Тому цілком припустимо використовувати як функцію правдоподібності $p(\mathbf{u} / \varphi_{ch})$, так й умовне відношення правдоподібності

$$\lambda(\mathbf{u} / \varphi_{ch}) = \frac{p(\mathbf{u} / \theta = 1, \varphi_{ch})}{p(\mathbf{u} / \theta = 0)}.$$

⁶ Про комплексні аналітичні сигнали дивись §1.3.

Отже, якщо діяти таким же чином, як й у попередньому параграфі, то можна одержати наступний вираз для умовного функціоналу відношення правдоподібності $\lambda(\mathbf{u} / \varphi_{ch})$:

$$\lambda(\mathbf{u} / \varphi_{ch}) = \exp \left\{ \operatorname{Re} [\exp(j\varphi_{ch}) \cdot \dot{V}_0(\mathbf{u})] - \frac{1}{4} \cdot Q_0 \right\}, \quad (11.18)$$

де $\dot{V}_0(\mathbf{u}) = (\dot{\mathbf{b}}, \mathbf{u})$ – функціонал лінійної обробки, $Q_0 = (\dot{\mathbf{b}}, \dot{\mathbf{s}}_0)$ – центрувальна константа, $\dot{\mathbf{b}}$ – вектор оптимальних коефіцієнтів, який є рішенням матричного рівняння

$$(g_w \mathbf{E}_N + \mathbf{K}_z) \cdot \dot{\mathbf{b}} = \dot{\mathbf{s}}_0. \quad (11.19)$$

Діючи таким же чином, як й вище, перейдемо від дискретної до дискретно-аналогової версії алгоритму лінійної обробки, а саме:

$$V_0(\mathbf{u}) = (\dot{\mathbf{b}}, \mathbf{u}) = \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt. \quad (11.20)$$

Тут $\dot{b}(t, \mathbf{r}_m)$ – оптимальні вагові функції обробки, які є рішенням системи інтегральних рівнянь

$$g_w \cdot \dot{b}(t, \mathbf{r}_m) + \sum_{l=1}^M \int_{\mathbf{T}_l} K_z(t, t'; \mathbf{r}_m, \mathbf{r}_l) \cdot \dot{b}(t', \mathbf{r}_l) dt' = \dot{s}_0(t, \mathbf{r}_m); \quad (11.21)$$

$$m = 1, \dots, M.$$

а центрувальна константа у розгорнутому вигляді має вигляд

$$Q_0 = \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot \dot{s}_0^*(t, \mathbf{r}_m) dt. \quad (11.22)$$

Зауважимо також, що при даному переході врахована прийнята вище угода стосовно структури векторів та прив'язки їх елементів до значень фізичних полів. З порівняння виразів (11.18) - (11.22) з відповідними формулами попереднього параграфу видно, що головна, по суті, відмінність поміж ними полягає в тому, що замість дійсних вагових функцій в даному разі застосовуються їх комплексні аналоги. Зокрема, при відсутності

зовнішніх завад з рівняння (11.21) маємо наступне рішення щодо оптимальної вагової функції:

$$\dot{b}(t, \mathbf{r}_m) = \frac{1}{g_w} \dot{s}_0(t, \mathbf{r}_m).$$

Звідси витікає, що оптимальною в цьому разі є обробка кореляційного типу з ваговою функцією, що відповідає комплексному сигналу (11.17), тобто:

$$\dot{V}_0(\mathbf{u}) = \frac{1}{g_w} \sum_{m=1}^M \int_{\mathbf{T}_m} u(t, \mathbf{r}_m) \cdot \dot{s}_0(t, \mathbf{r}_m) dt \quad (11.23)$$

Тепер перейдемо до наступного кроку, а саме, врахування випадковості параметру φ_{ch} . З двох основних напрямків вирішення цієї задачі для даної моделі каналу найчастіше прийнято застосовувати баєсовий підхід, тобто, статистичне усереднення умовного відношення правдоподібності $\lambda(\mathbf{u} / \varphi_{ch})$ по випадковому параметру φ_{ch} . При цьому для початкової фази сигналу звичайно приймається рівномірний в межах періоду розподіл [3], тобто, $p(\varphi_{ch}) = 1/2\pi$, де $\varphi_{ch} \in (0, 2\pi)$. Таким чином, в рамках баєсового підходу безумовне відношення правдоподібності гіпотез має вигляд:

$$\lambda(\mathbf{u}) = \frac{1}{2\pi} \int_0^{2\pi} \lambda(\mathbf{u} / \varphi) d\varphi = \exp\left\{-\frac{1}{4} \cdot Q_0\right\} \cdot \frac{1}{2\pi} \int_0^{2\pi} \exp\{\operatorname{Re}(\exp(j\varphi) \cdot \dot{V}_0(\mathbf{u}))\} d\varphi.$$

Відомо табличний інтеграл [3]:

$$\frac{1}{2\pi} \int_0^{2\pi} \exp\{x \cos \psi \pm y \sin \psi\} d\psi = I_0(\sqrt{x^2 + y^2}), \quad (11.24)$$

де $I_0(z)$ – модифікована функція Бесселя нульового порядку від уявного аргументу. Використовуючи цю формулу, для безумовного відношення правдоподібності гіпотез остаточно знаходимо:

$$\lambda(\mathbf{u}) = \exp\left\{-\frac{1}{4} \cdot Q_0\right\} \cdot I_0(|\dot{V}_0(\mathbf{u})|).$$

У цій формулі від прийнятого сигналу залежить лише аргумент функції Бесселя. Неважко пересвідчитися, що $\lambda(\mathbf{u})$ є монотонно зростаючою

функцією від невід'ємної статистики $|\dot{V}_0(\mathbf{u})| \geq 0$. Отже, оптимальне правило виявлення сигналу у каналі з випадковою початковою фазою в дискретно-аналоговій версії може бути записане у вигляді нерівності:

$$|\dot{V}_0(\mathbf{u})| = \left| \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt \right| \begin{matrix} \geq \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \text{ П.} \quad (11.25)$$

Тестова статистика правила (11.25), на відміну від правила (11.9), передбачає два етапи обробки: лінійний та нелінійний. На першому етапі обчислюється кореляція поля спостережень $u(t, \mathbf{r})$ з оптимальною комплексною функцією $\dot{b}(t, \mathbf{r})$, яка є рішенням системи рівнянь (11.21). На другому етапі знаходиться модуль виходу корелятора, який порівнюється з порогом. Саме операція взяття модуля й забезпечує нечутливість алгоритму виявлення (11.25) до фази сигналу. Клас алгоритмів оброблення, в яких не використовується фазова інформація, називають *некогерентним* прийманням сигналів [3]. Відповідно, методи оброблення, які потребують інформації про фазу сигналу, прийнято називати *когерентним* прийманням. Прикладом когерентних методів є алгоритм (11.11) та його окремий випадок (11.13), які розглядалися вище для задачі виявлення точно відомого сигналу.

2. Канал із випадковою амплітудою та фазою сигналу.

У даній моделі каналу випадковими вважаються вже два параметри: коефіцієнт передачі каналу за амплітудою ε_{ch} та початкова фаза сигналу φ_{ch} . По аналогії з виразом (11.16) поле сигналу на вході приймального пристрою представимо наступним чином:

$$s_{ch}(t, \mathbf{r} / \varepsilon_{ch}, \varphi_{ch}) = s_{AR}(t / \dot{\eta}_{ch}) = \text{Re} \{ \dot{s}_0(t, \mathbf{r}) \cdot \dot{\eta}_{ch}^* \}, \quad (11.26)$$

де $\dot{\eta}_{ch} = \varepsilon_{ch} \cdot \exp(-j \cdot \varphi_{ch})$ - комплексний коефіцієнт передачі каналу, $\dot{s}_0(t, \mathbf{r})$ - поле сигналу, що відповідає деякому стандартному (тобто, відомому) значенню $\dot{\eta}_0$. Далі, діючи за тією ж методикою, що й вище,

запишемо умовний функціонал відношення правдоподібності. В результаті одержимо:

$$\lambda(\mathbf{u} / \dot{\eta}) = \exp \left\{ \operatorname{Re} \left\{ \dot{\eta}_{ch}^* \cdot \dot{V}_0(\mathbf{u}) \right\} - \frac{1}{4} \cdot |\dot{\eta}_{ch}|^2 \cdot Q_0 \right\}. \quad (11.27)$$

Тут величини $\dot{V}_0(\mathbf{u})$ та Q_0 мають той же зміст, що й у попередньому випадку, тобто, вони визначаються співвідношеннями (11.20), (11.23). Тепер, у відповідності до баєсового підходу, вираз (11.27) треба усереднити по параметру $\dot{\eta}_{ch}$:

$$\lambda(\mathbf{u}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left\{ \operatorname{Re} \left(\dot{\eta}^* \cdot \dot{V}_0(\mathbf{u}) \right) - \frac{1}{4} \cdot |\dot{\eta}|^2 \cdot Q_0 \right\} \cdot p_{ch}(\dot{\eta}) d\dot{\eta}, \quad (11.28)$$

де $p_{ch}(\dot{\eta})$ апіорний розподіл коефіцієнта передачі каналу. Обговоримо міркування щодо вибору цього розподілу. Однією з найпоширеніших причин випадкових флуктуацій коефіцієнту передачі каналу є розсіювання поля сигналу на статистичних неоднорідностях середовища. В результаті розсіювання поле сигналу в області прийому формується підсумовуванням великої кількості сигнальних компонент із випадковими фазами та амплітудами. Отже, відповідно до центральної граничної теореми теорії ймовірностей, яка вже неодноразово згадувалася вище, сумарне поле розсіяного сигналу на приймальній апертурі має тенденцію до нормалізації. Крім того, наряду з розсіяним полем може мати місце й пряме проходження сигналу. Виходячи з вказаних міркувань, часто приймають гіпотезу про нормальний розподіл комплексного коефіцієнту передачі каналу $\dot{\eta}$, а саме:

$$p_{ch}(\dot{\eta}) = \frac{1}{2\pi\sigma_{\xi}^2} \exp \left\{ -\frac{|\dot{\eta} - \dot{a}|^2}{2\sigma_{\xi}^2} \right\}, \quad (11.29)$$

де $\dot{a} = \langle \dot{\eta} \rangle$ - математичне очікування⁷, $\xi = \dot{\eta} - \langle \dot{\eta} \rangle$ - випадкові амплітудно-фазові флуктуації коефіцієнту передачі. При цьому вважається, що флуктуації квадратурних компонент ξ статистично незалежні та мають

⁷ Фактично мається на увазі прогнозне значення параметрів сигналу.

однакову дисперсію σ_ξ^2 . Фізично параметр $\dot{a}$ визначає рівень сталої, або, як її прийнято називати [16], когерентної компоненти поля сигналу, а дисперсія σ_ξ^2 - інтенсивність флуктуацій випадкової (некогерентної) компоненти. Якщо флуктуації відсутні, тобто, $\sigma_\xi^2 = 0$, то одержуємо раніше розглянутий випадок виявлення наперед відомого сигналу.

При практичному застосуванні вказаного розподілу звичайно вдаються до певних нормувань щодо параметрів $\dot{a}$ та σ_ξ^2 . Зокрема, якщо виходити з того, що середня потужність прийнятого сигналу в умовах флуктуацій повинна зберігатися такою ж, як й при їх відсутності, то необхідно забезпечити виконання умови⁸ $\langle |\dot{\eta}|^2 \rangle = 1$. Цій вимозі задовольняє наступне визначення параметрів $\dot{a}$ та σ_ξ^2 :

$$\dot{a} = \sqrt{\gamma_{coh}} \cdot \exp(-j\varphi_0); \quad \sigma_\xi^2 = \frac{1 - \gamma_{coh}}{2}, \quad (11.30)$$

де φ_0 - відомий (прогнозований) зсув фази сигналу в каналі,

$$\gamma_{coh} = \frac{|\dot{a}|^2}{|\dot{a}|^2 + 2\sigma_\xi^2} \quad (11.31)$$

- відносний за потужністю рівень когерентної компоненти поля сигналу, причому $0 \leq \gamma_{coh} \leq 1$. Далі для стислості будемо називати коефіцієнт γ_{coh} просто параметром когерентності поля сигналу. Зауважимо, що досить часто в якості параметру каналу даного типу прийнято використовувати так званий [15, 19] K_{ch} - фактор:

$$K_{ch} = \frac{|\dot{a}|^2}{2\sigma_\xi^2},$$

⁸ В правій частині цієї умови може бути не одиниця, а деяка інша константа, якщо це буде зручніше з якихось міркувань.

що відповідає за змістом відношенню інтенсивностей когерентної та некогерентної компонент сигналу. Незавжди бачити, що параметр когерентності γ_{coh} та K_{ch} – фактор пов’язані поміж собою співвідношеннями

$$\gamma_{coh} = \frac{K_{ch}}{K_{ch} + 1}, \quad K_{ch} = \frac{\gamma_{coh}}{1 - \gamma_{coh}},$$

тобто, ці два параметри еквівалентні. Зокрема, повністю некогерентному сигналу відповідає фактор $K_{ch} = 0$, що дає значення параметру $\gamma_{coh} = 0$. У протилежному випадку повністю когерентного сигналу фактор $K_{ch} \rightarrow \infty$, що приводить до значення параметру $\gamma_{coh} = 1$.

Можна також задати статистичні характеристики каналу й через окремі розподіли фази $\varphi_{кан}$ та модулю коефіцієнту передачі η . Для вказаної моделі флуктуацій, як відомо [3], розподіл фази є рівномірним на $(0, 2\pi)$, а параметр η підлягає розподілу Релея-Райса або, як його ще називають [3], узагальненому розподілу Релея:

$$p_{\eta}(\eta) = \frac{\eta}{\sigma_{\xi}^2} \exp\left\{-\frac{\eta^2 + a^2}{2\sigma_{\xi}^2}\right\} \cdot I_0\left(\frac{\eta \cdot a}{\sigma_{\xi}^2}\right), \quad (11.32)$$

де $a = |\dot{a}| = \sqrt{\gamma_{coh}}$. З урахуванням цієї обставини дану модель середовища передачі прийнято називати райсівським каналом [3]. У граничному випадку, коли відсутня когерентна компонента ($a = 0$), розподіл (11.32) переходить у відомий закон Релея:

$$p_{\eta}(\eta) = \frac{\eta}{\sigma_{\xi}^2} \exp\left\{-\frac{\eta^2}{2\sigma_{\xi}^2}\right\}. \quad (11.33)$$

Відповідну модель середовища передачі прийнято називати релеєвським каналом [3]. Фізично ця модель каналу означає, що відсутнє пряме проходження сигналу, а поле сигналу формується лише розсіяними компонентами. Після підстановки розподілу (11.29) у вираз (11.28) та виконання інтегрування приходимо до наступного результату:

$$\lambda(\mathbf{u}) = \frac{1}{1 + \sigma_{\xi}^2 \cdot Q_0 / 2} \cdot \exp \left\{ \frac{2 \cdot \operatorname{Re}(\dot{a}^* \cdot \dot{V}_0(\mathbf{u})) + \sigma_{\xi}^2 \cdot |\dot{V}_0(\mathbf{u})|^2 - |\dot{a}|^2 \cdot Q_0 / 2}{2 \cdot (1 + \sigma_{\xi}^2 \cdot Q_0 / 2)} \right\}.$$

Звідси, з урахуванням співвідношень (11.30), витікає наступне оптимальне правило виявлення сигналу в каналі даного типу:

$$\sqrt{\gamma_{coh}} \cdot \operatorname{Re}\{\exp(j \cdot \varphi_0) \cdot \dot{V}_0(\mathbf{u})\} + \frac{(1 - \gamma_{coh})}{4} \cdot |\dot{V}_0(\mathbf{u})|^2 \stackrel{H_1}{\underset{H_0}{>}} \text{П.} \quad (11.34)$$

Тестова статистика оптимального правила (11.34) складається з двох додатків. Перший з них відповідає обробці когерентної компоненти сигналу та фактично співпадає з оптимальним алгоритмом (11.9) для ситуації прийому точно відомого сигналу. Різниця з виразом (11.9) у цій частині має суто технічний характер, що стосується способу врахування наперед відомої частини фазового зсуву у каналі φ_0 . Другий же додаток виразу (11.34) є схожим із алгоритмом некогерентного приймання (11.25), оптимальним для каналу з невідомою початковою фазою сигналу. Ці дві частини алгоритму обробки (11.34) підсумовуються з ваговими коефіцієнтами, які визначаються параметром когерентності каналу γ_{coh} . Зокрема, у випадку релеєвського каналу, тобто, при відсутності когерентної компоненти ($\gamma_{coh} = 0$), алгоритм (11.34) є еквівалентним⁹ правилу (11.25). В іншому граничному випадку, коли відсутня некогерентна компонента ($\gamma_{coh} = 1$), алгоритм (11.34) відповідає повністю когерентній обробці сигналу. Що ж стосується проміжних значень параметру $0 < \gamma_{coh} < 1$, то при зростанні відношення сигнал/завада $q_0^2 \rightarrow \infty$ основний вклад у функціонал (11.34) вносить саме другий додаток. Дійсно, як буде показано далі, вихід лінійної частини обробки (11.20) є безрозмірною величиною, середній квадрат модуля якої $\langle |\dot{V}_0(\mathbf{u})|^2 \rangle$ має порядок відношення сигнал/завада q_0^2 . Таким чином, алгоритм

⁹ Зведення у квадрат тестової статистики не змінює, вочевидь, відношення порядку величин.

виявлення (11.25), одержаний раніше для моделі каналу з випадковою початковою фазою, є асимптотично оптимальним по параметру q_0^2 й для каналу з флуктуаціями не тільки фази, але й амплітуди сигналу. Виявляється, що цей результат зберігається в силі й для більш загальних моделей каналів.

11.4. Приймання полів випадкових шумових сигналів.

В системах пасивної локації, радіоастрономії, технічної діагностики та деяких інших сигналами, що підлягають виявленню, є власне випромінювання деяких фізичних або технічних об'єктів (повітряні, наземні, надводні та підводні судна, деякі транспортні засоби, об'єкти технічної інфраструктури та багато інших). За фізичною природою випромінювання об'єктів також може бути різним: електромагнітним, оптичним, інфрачервоним, акустичним, тощо. Але спільною рисою сигналів, які несе таке випромінювання, є те, що їх форма не може бути попередньо прогнозована, тобто, ці сигнали є повністю випадковими.

Також в багатьох практичних задачах такі сигнали подібні до шумових процесів, отже, для їх опису доцільно застосовувати певні статистичні характеристики. Найбільш повною з таких характеристик є апіорний розподіл ймовірностей сигналу $p_s(\mathbf{s})$ (або імовірнісна міра $dP_s(\mathbf{s})$ у неперервному випадку). Відразу ж виникає питання про те, як задати вигляд такого розподілу. З міркувань, які було викладено вище, для поля випадкових сигналів в багатьох випадках може бути прийнято нормальний закон розподілу ймовірностей. Отже, в рамках цієї гіпотези для умовного N_0 - вимірного розподілу вектору спостережень $\mathbf{u}$ можна записати:

$$p(\mathbf{u}/\theta) = \frac{1}{(2\pi)^{N_0/2} \sqrt{\det(\mathbf{g}_w \mathbf{E}_{N_0} + \mathbf{K}_z + \theta \cdot \mathbf{K}_s)}} \times \exp\left\{-\frac{1}{2} \mathbf{u}^T \cdot (\mathbf{g}_w \mathbf{E}_{N_0} + \mathbf{K}_z + \theta \cdot \mathbf{K}_s)^{-1} \cdot \mathbf{u}\right\} \quad (11.35)$$

де $\mathbf{K}_s = E\{\mathbf{s} \cdot \mathbf{s}^T\}$ - кореляційна матриця сигналу. Використовуючи цей розподіл, знаходимо відношення правдоподібності гіпотез $\theta = 1$ та $\theta = 0$:

$$\lambda(\mathbf{u}) = \frac{p(\mathbf{u} / \theta = 1)}{p(\mathbf{u} / \theta = 0)} = \sqrt{\frac{\det(g_w \mathbf{E}_N + \mathbf{K}_z)}{\det(g_w \mathbf{E}_N + \mathbf{K}_z + \mathbf{K}_s)}} \cdot \exp\left\{\frac{1}{2} F_0(\mathbf{u})\right\},$$

$$\text{де} \quad F_0(\mathbf{u}) = \mathbf{u}^T \cdot \mathbf{B}_0 \cdot \mathbf{u} \quad (11.36)$$

- квадратична форма, ядро якої має вигляд:

$$\mathbf{B}_0 = (g_w \mathbf{E}_N + \mathbf{K}_z)^{-1} - (g_w \mathbf{E}_N + \mathbf{K}_z + \mathbf{K}_s)^{-1}. \quad (11.37)$$

Неважко бачити, що у виразі для функціоналу правдоподібності $\lambda(\mathbf{u})$ від спостережень залежить лише показник експоненти. В силу загальних властивостей ковариційних матриць, про які вже йшла мова вище, квадратична форма (11.37) є невід'ємно визначеною. Тому функціонал $\lambda(\mathbf{u})$ є монотонно зростаючою функцією цієї статистики. Звідси витікає оптимальне правило виявлення випадкового сигналу:

$$F_0(\mathbf{u}) = \mathbf{u}^T \cdot \mathbf{B}_0 \cdot \mathbf{u} \underset{H_0}{\overset{H_1}{>}} \Pi, \quad (11.38)$$

де Π - поріг виявлення, який найчастіше знаходиться у відповідності до критерію Неймана-Пірсона. Отже, оптимальне правило виявлення випадкових сигналів (11.38) передбачає обчислення квадратичної форми (11.36), ядром якої є матриця (11.37). Вказане ядро значно спрощується в окремому випадку, коли зовнішні завади відсутні, тобто, виявлення ведеться лише на тлі білого шуму. Дійсно, скористаємось матричною тотожністю

$$(g\mathbf{E} + \mathbf{K})^{-1} = \frac{1}{g} [\mathbf{E} - \mathbf{K} \cdot (g\mathbf{E} + \mathbf{K})^{-1}], \quad (11.39)$$

де $\mathbf{E}$ - одинична, а $\mathbf{K}$ - довільна матриці однакового порядку¹⁰. Беручи до уваги тотожність (11.39), матричну форму (11.37) при $\mathbf{K}_z \rightarrow 0$ можна привести до вигляду:

¹⁰ Тотожність (11.39) легко довести, помноживши обидві її частини справа на матрицю $g\mathbf{E} + \mathbf{K}$.

$$\mathbf{B}_0 = \frac{1}{g_w} \mathbf{K}_s \cdot (g_w \mathbf{E}_N + \mathbf{K}_s)^{-1}. \quad (11.40)$$

Запишемо тестову статистику правила (11.38) у розгорнутому вигляді. Почнемо з випадку дискретних спостережень. Відповідно до прийнятої вище угоди щодо блочної структури векторів та матриць, квадратичну форму (11.36) можна записати в наступному еквівалентному вигляді:

$$F_0(\mathbf{u}) = \sum_{i=1}^M \sum_{k=1}^M \mathbf{u}^{(i)} \cdot \mathbf{B}_0^{(i,k)} \cdot \mathbf{u}^{(k)}, \quad (11.41)$$

де $\mathbf{u}^{(i)}$ - блок вектору спостережень, який відповідає i -ому датчику антени ($i = 1, \dots, M$), $\mathbf{B}_0^{(i,k)}$ - ik -ий блок матриці вагових коефіцієнтів. Далі візьмемо до уваги, що елементами блоку спостережень $\mathbf{u}^{(i)}$ є відліки поля на приймальній апертурі $u(t_j, \mathbf{r}_i)$, де $j = 1, \dots, N$, N - кількість відліків у часі. Тоді кожному з додатків виразу (11.41) у розгорнутому вигляді відповідає двократна сума по часовій координаті. Отже, вираз (11.41) можна представити у наступному розгорнутому вигляді:

$$F_0(\mathbf{u}) = \sum_{i=1}^M \sum_{k=1}^M \sum_{j=1}^N \sum_{l=1}^N u(t_j, \mathbf{r}_i) \cdot b_0(t_j, t_l, \mathbf{r}_i, \mathbf{r}_k) \cdot u(t_l, \mathbf{r}_k). \quad (11.42)$$

Тут через $b_0(t_j, t_l, \mathbf{r}_i, \mathbf{r}_k)$ позначено jl -ий елемент ik -ого блоку матриці вагових коефіцієнтів $\mathbf{B}_0^{(i,k)}$. Звідси вже не важко записати «змішану» дискретно-аналогову версію тестової статистики (11.36), яка має вигляд

$$F_0(\mathbf{u}) = \sum_{i=1}^M \sum_{k=1}^M \int_{\mathbf{T}_m} \int_{\mathbf{T}_p} u(t, \mathbf{r}_i) \cdot b_0(t, t'; \mathbf{r}_i, \mathbf{r}_k) \cdot u(t', \mathbf{r}_k) dt' dt. \quad (11.43)$$

Ядро $b_0(t, t'; \mathbf{r}_i, \mathbf{r}_k)$ інтегрального оператора (11.43) є неперервним аналогом матричного співвідношення (11.37), яке в ситуації відсутності зовнішніх завад має вигляд (11.40). Помножимо зліва обидві частини співвідношення (11.40) на кореляційну матрицю суміші шумів та сигналу $(g_w \mathbf{E}_N + \mathbf{K}_s)$. В результаті одержимо наступне матричне рівняння щодо ядра квадратичної форми функціоналу обробки (11.36):

$$\mathbf{B}_0 \cdot (g_w \mathbf{E}_N + \mathbf{K}_s) = \mathbf{K}_s / g_w.$$

У свою чергу, цьому рівнянню у дискретно-аналоговій версії обробки відповідає система інтегральних рівнянь наступного вигляду:

$$g_w \cdot b_0(t, t'; \mathbf{r}_i, \mathbf{r}_k) + \sum_{p=1}^M \int_{\mathbf{T}_p} K_s(t, t''; \mathbf{r}_i, \mathbf{r}_p) \cdot b_0(t'', t'; \mathbf{r}_p, \mathbf{r}_k) dt'' = \frac{1}{g_w} K_s(t, t'; \mathbf{r}_i, \mathbf{r}_k). \quad (11.44)$$

Подібним же чином можна одержати інтегральні співвідношення й для інших версій та інтерпретацій оптимальної обробки.

Контрольні запитання та завдання до розділу 11.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Що таке детермінований сигнал? Як формулюється завдання його виявлення? Яку теоретичну та практичну цінність має вирішення цього завдання? Одержати самостійно вираз для відношення правдоподібності (11.6).

2. За яких умов вираз (11.7) переходить у вираз (11.10)? Що таке кореляційний прийом? Що таке прийом, узгоджений з полем сигналу?

3. З якими складностями пов'язана оптимізація обробки сигналів у випадку неперервних спостережень? Перейти самостійно від виразу (11.7) до виразу (11.11) та одержати рівняння (11.12).

4. Що таке квазидетерміновані сигнали? Які методи оптимізації оброблення таких сигналів пропонує теоретична статистика?

5. Що таке канал із випадковою фазою сигналу? Чому для цієї моделі каналу замість умовної функції правдоподібності $p(\mathbf{u}/\varphi_{ch})$ можна використовувати умовне відношення правдоподібності $\lambda(\mathbf{u}/\varphi_{ch})$?

6. Одержати самостійно вираз (11.18) та рівняння (11.19). Пояснити перехід до неперервної версії лінійної обробки (11.20).

7. В чому відмінність правила виявлення сигналу з випадковою початковою фазою (11.25) від правила виявлення детермінованих сигналів (11.9)? Як цю відмінність можна пояснити фізично?

8. Що таке канал із випадковою фазою та амплітудою сигналу? Одержати самостійно вираз для умовного відношення правдоподібності (11.27).

9. Чому для каналу із випадковою фазою та амплітудою сигналу найчастіше приймається гавсівський розподіл комплексного коефіцієнту передачі (11.29)? Пояснити фізичний зміст параметрів цього розподілу.

10. Що таке райсівський та релеєвський канали? В чому поміж ними різниця? Що таке параметр когерентності та K -фактор каналу?

11. Пояснити правило виявлення сигналу (11.34) та його окремі випадки в залежності від параметру когерентності каналу.

12. Що таке шумові сигнали? В яких практичних ситуаціях виникає завдання їх виявлення? Наведіть приклади із різних технічних галузей..

13. Самостійно одержати вирази (11.36), (11.37), (11.40)

14. Самостійно здійснити перехід від векторно-матричної форми (11.36) до розгорнутого запису (11.43).

15. Самостійно одержати рівняння (11.44). Якій моделі спостережень воно відповідає?

Розділ 12. Кореляційні характеристики випадкових полів.

У розділі 4 ми розглядали основні положення теорії випадкових процесів. Нагадаємо, що так прийнято називати випадкові функції часу, вигляд яких наперед невідомий. Випадкові процеси можуть бути представлені лише окремими реалізаціями, одержаними в результаті тих чи інших спостережень (дослідів). Повна сукупність таких реалізацій (можливо, нескінченна) утворюють так званий статистичний ансамбль. Отже, предметом теорії випадкових процесів є вивчення властивостей цього ансамблю на базі його імовірнісних характеристик (розподілів, статистичних моментів, тощо). У розділі 4 було показано, що теорія випадкових процесів є зручним «інструментом» для опису характеристик завад та деяких видів сигналів, зокрема, шумових. Тепер розглянемо узагальнення цієї теорії, яке стосується так званих випадкових полів, що вже є функціями не тільки часу, але й просторових координат. Таке узагальнення є необхідним для розгляду питань приймання та оброблення сигналів у часі та просторі, що й є одним з головних завдань даної частини нашого Посібника.

12.1. Загальні відомості.

Отже, випадковим полем називають [16] називають деяку випадкову функцію $U(t, \mathbf{r})$, яка залежить від часу t та вектору просторових координат¹ $\mathbf{r}$. Як й у теорії випадкових процесів, для спостереження доступна не безпосередньо функція $U(t, \mathbf{r})$, а лише окремі її реалізації зі статистичного ансамблю $\{u(t, \mathbf{r})\}$. Найбільш повною та вичерпною характеристикою випадкового поля є спільний розподіл ймовірностей його значень $u_1, u_2, \dots, u_N$ для N довільно обраних наборів моментів часу та точок спостереження, тобто, $u_i = U(t_i, \mathbf{r}_i)$, $i = 1, \dots, N$. Але в рамках гавсівської моделі розподілу ймовірностей для повного опису випадкового поля достатньо² задати лише

¹ Вигляд вектору $\mathbf{r}$ визначається, звісно, обраною системою координат.

² Дивись розділи 3 та 4.

два його статистичних момента, а саме, математичне очікування (сподівання) поля

$$m(t, \mathbf{r}) = \langle u(t, \mathbf{r}) \rangle, \quad (12.1)$$

та кореляцію його значень для двох довільно обраних моментів часу t_1, t_2 та точок прийому $\mathbf{r}_1, \mathbf{r}_2$, тобто,

$$K_U(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) = \langle U(t_1, \mathbf{r}_1) \cdot U(t_2, \mathbf{r}_2) \rangle, \quad t_1, t_2 \in (0, T), \mathbf{r}_1, \mathbf{r}_2 \in \mathbf{D}_A. \quad (12.2)$$

Тут $(0, T)$ - інтервал спостереження поля у часі, $\mathbf{D}_A$ - область простору, яку займає приймальна антена (приймальна апертура). Як й вище, кутові дужки позначають усереднення по статистичному ансамблю поля, тобто, по усім можливим його реалізаціях. Отже, всі подальші кроки з опису характеристик гавсівських випадкових полів зводяться до розгляду властивостей моментів (12.1) та (12.2). Найбільш принциповим з таких кроків є припущення про характер залежності статистичних характеристик (12.1) та (12.2) від часу та просторових координат.

Нагадаємо, що у теорії випадкових процесів центральне місце займає припущення про стаціонарність процесу, яке у **вузькому** (тобто, строгому) сенсі означає інваріантність усіх його статистичних характеристик щодо вибору початкового моменту відліку часу t_0 . Але у практичних застосуваннях найчастіше використовують так зване **широке визначення**, в рамках якого інваріантність відносно t_0 вимагається лише стосовно двох перших моментів процесу: математичного очікування та кореляційного моменту. Така інваріантність є еквівалентною до припущення про незалежність математичного очікування процесу від часу, а щодо кореляційного моменту – про його залежність лише від різниці моментів часу t_1 та t_2 . Практична поширеність широкого визначення у порівнянні з його «вузьким колегою» пояснюється просто: якщо розподіл процесу є гавсівським, то, як вже було сказано вище, усі багатовимірні статистичні

характеристики процесу визначаються саме його першими двома моментами. Отже, для гавсівських процесів «широке» та «вузьке» визначення є повністю еквівалентними. У той же час, прийняття гіпотези про стаціонарність процесу породжує низку важливих як теоретичних так й практичних наслідків, про що вже йшла мова у розділі 4.

Подібні властивості випадкових процесів у повній мірі переносяться й на випадкові поля, тобто, якщо випадкове поле є стаціонарним у широкому сенсі, то мають виконуватися певні співвідношення інваріантності, а саме:

$$\begin{aligned} m(t, \mathbf{r}) &= m(\mathbf{r}) \\ K_U(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) &= K_U(t_1 - t_2, \mathbf{r}_1, \mathbf{r}_2). \end{aligned} \quad (12.3)$$

Але принципова відмінність випадкових полів від випадкових процесів полягає в тому, що їх характеристики змінюються не лише у часі, але й у просторі. Отже, цілком логічно виглядає доповнення гіпотези стаціонарності поля припущенням про можливу інваріантність його статистичних характеристик щодо вибору початкової точки відліку координат у просторі $\mathbf{r}_0$. Таке поле вже називають стаціонарним та статистично однорідним у широкому сенсі, що еквівалентно наступним умовам:

$$\begin{aligned} m(t, \mathbf{r}) &= m, \\ K_U(t_1, t_2; \mathbf{r}_1, \mathbf{r}_2) &= K_U(t_1 - t_2; \mathbf{r}_1 - \mathbf{r}_2). \end{aligned} \quad (12.4)$$

Таким чином, ми визначили кілька можливих класів випадкових полів. У найбільш загальному випадку, коли гіпотези стаціонарності та однорідності не виконуються, маємо поле **нестационарне та статистично неоднорідне**. Математичне очікування такого поля (12.1) залежить від 4 аргументів, а саме, часу t та вектору просторових координат $\mathbf{r}$, який у фізичних просторах визначається 3 компонентами. У свою чергу, кореляційний момент (12.2) такого поля залежить окремо вже від 8 аргументів: двох моментів часу t_1, t_2 та двох векторів просторових координат $\mathbf{r}_1, \mathbf{r}_2$.

Якщо ж тепер прийняти гіпотезу стаціонарності (12.3), то маємо випадок *стаціонарного але статистично неоднорідного* поля. Внаслідок умови (12.3) кількість окремих аргументів математичного очікування скорочується до трьох, а кореляційного моменту – до 7. Нарешті, у найбільш сприятливому випадку, коли виконуються умови (12.4), маємо поле *стаціонарне та статистично однорідне*. У цьому випадку математичне очікування поля взагалі є константою, яка не змінюється у часі та просторі, а кореляційний момент визначається різницею часових $t_1 - t_2$ та просторових $\mathbf{r}_1 - \mathbf{r}_2$ координат, тобто, фактично лише 4 аргументами.

Допитливий читач міг помітити, що ми оминули ще один випадок, а саме, поля *нестационарні але статистично однорідні*. Але це не випадково: виявляється, що такі поля у більшості прикладних задач фізично не можуть існувати. Цій обставині можна дати наступне пояснення. Річ у тім, що поля, які існують у фізичному середовищі, підлягають звичайно деяким динамічним рівнянням, зокрема, хвильовому. Внаслідок зв'язків, які накладають такі рівняння, зміни поля у часі та просторі є взаємозалежними. Саме цим й пояснюється та обставина, що нестационарність фізичного поля обов'язково призводить й до його неоднорідності (але, що цікаво, не навпаки!).

На останок цього параграфу зробимо ще одне принципове зауваження, про яке вже йшлося у розділах 4 та 9. З фізичної точки зору стаціонарні випадкові поля, також як й випадкові процеси, існувати не можуть оскільки стаціонарність означає поміж іншого нескінченну протяжність поля у часі, що фізично неможливе. Абсолютно аналогічно математичне припущення про статистичну однорідність випадкових полів також є фізично неможливим внаслідок обмеженості будь-яких фізичних полів у просторі. Тому практично може мати місце лише так звана *квазістаціонарність та квазіоднорідність*, тобто, приблизне виконання відповідних математичних припущень лише на обмежених просторово-часових інтервалах, розміри яких є достатніми для

того, щоб їх екстраполювати на нескінченний простір без суттєвих втрат коректності для тих чи інших висновків при вирішенні конкретних практичних завдань.

12.2. Поле випадкового локального джерела.

Перш чим перейти до загально теоретичного розгляду, зупинимось на деяких простих прикладах, які є корисними для розуміння й деяких загальних закономірностей. Почнемо з моделі поля, що створює зосереджене (локальне) джерело на приймальній апертурі, розташованій у його дальній зоні. У цьому наближенні, по аналогії з виразом (9.11), можна записати наступну модель для поля на приймальній апертурі:

$$u(t, \mathbf{r}) = u_0(t + (\mathbf{n}, \mathbf{r})/c_0), \quad (12.5)$$

$$t \in (0, T), \mathbf{r} \in \mathbf{D}_A,$$

де $u_0(t)$ - сигнал джерела в умовному центрі приймальної апертури, $\mathbf{n}$ - одиничний вектор напрямку на джерело, $\mathbf{r}$ - вектор координат деякої точки спостереження відносно умовного центру, c_0 - швидкість поширення сигналу в середовищі, $(\mathbf{a}, \mathbf{b})$ - скалярний добуток векторів. Будемо вважати кутове положення джерела $\mathbf{n}$ незмінним, а сигнал джерела $u_0(t)$ - гавсівським стаціонарним процесом. Тоді легко бачити, що поле (12.5) як функція часу у будь-якій точці спостереження також є гавсівським стаціонарним процесом. Але залежність поля від просторових координат є цілком детермінованою та визначається законом поширення однорідної³ плоскої хвилі, основним параметром якої є орт $\mathbf{n}$.

Математичне очікування поля (12.5) без втрат загальності можна вважати нульовим. Отже, для повного опису такого поля достатньо записати лише його кореляційний момент (12.2):

$$K_U(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) = \left\langle u_0(t_1 + (\mathbf{n}, \mathbf{r}_1)/c_0) \cdot u_0(t_2 + (\mathbf{n}, \mathbf{r}_2)/c_0) \right\rangle. \quad (12.6)$$

³ Нагадаємо, що однорідною хвилею називають таку, амплітуда якої не змінюється вздовж її фронту. Якщо ж це не так, то хвилю називають неоднорідною. Зокрема, неоднорідні плоскі хвилі можуть існувати поблизу джерел, границь середовища та відбиваючих об'єктів.

Беручи до уваги припущення про стаціонарність джерела цей вираз можна представити у наступному вигляді:

$$K_U(t_1, t_2, \mathbf{r}_1, \mathbf{r}_2) = B_0(t_1 - t_2 + (\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_2)/c_0), \quad (12.7)$$

$$\text{де} \quad B_0(\tau) = \langle u_0(t) \cdot u_0(t + \tau) \rangle \quad (12.8)$$

- кореляційна функція стаціонарного сигналу джерела в центрі області приймання. Із співставлення виразу (12.7) з умовами (12.4) можна зробити важливий висновок: випадкове поле (12.5) є не тільки стаціонарним, але й статистично однорідним. Дійсно, кореляційний момент (12.7) залежить не окремо від координат $\mathbf{r}_1$ та $\mathbf{r}_2$, а лише від їх різниці, що й є ознакою статистичної однорідності поля у широкому сенсі. Отже, кореляційний момент (12.7) є просторово-часовою кореляційною функцією⁴ випадкового поля (12.5).

Але дана модель корисна не лише для ілюстрації однієї з можливих умов формування стаціонарного та статистично однорідного поля. Більш суттєвою є та обставина, вона дає можливість продемонструвати умови розділення кореляційної функції (12.7) на добуток двох множників, перший з яких залежить лише від часових аргументів t_1, t_2 , а другий – від просторових $\mathbf{r}_1, \mathbf{r}_2$. Таке розділення є принципово важливим для побудови систем просторово-часового оброблення сигналів, про що піде мова у наступному розділі.

З розгляду виразу (12.7) поки що неясно, як обґрунтувати таке розділення: моменти часу t_1, t_2 та просторові координати $\mathbf{r}_1, \mathbf{r}_2$ входять до аргументу кореляційної функції сигналу у вигляді деякої лінійної комбінації. Але якщо скористатися теоремою Вінера-Хінчина (розділ 4, формула (4.16)), то кореляційну функцію (12.7) можна представити наступним чином:

$$K_U(\tau, \mathbf{p}) = \int_{-\infty}^{\infty} g_0(f) \exp\{j2\pi f \cdot (\tau + (\mathbf{n}, \mathbf{p})/c_0)\} df, \quad (12.9)$$

⁴ Нагадаємо, що у теорії випадкових процесів поняття кореляційної функції застосовується, як правило, лише для стаціонарних процесів. Аналогічно для випадкових полів про кореляційну функцію доцільно говорити, якщо наряду зі стаціонарністю додатково виконується ще й гіпотеза статистичної однорідності.

де $\tau = t_1 - t_2$, $\mathbf{p} = \mathbf{r}_1 - \mathbf{r}_2$, а $g_0(f)$ - спектральна щільність потужності сигналу джерела у центрі приймальної апертури. Надалі буде зручніше перейти до еквівалентного комплексного (аналітичного) представлення поля (12.5)

$$\dot{u}(t, \mathbf{r}) = \dot{u}_0(t + (\mathbf{n}, \mathbf{r})/c_0), \quad t \in (0, T), \mathbf{r} \in \mathbf{D}_A, \quad (12.10)$$

де крапка над символом, як й у попередніх розділах, позначає аналітичну форму представлення відповідного сигналу (дивись §1.3). Оскільки аналітичні сигнали мають спектральні складові лише в області позитивних частот, то спектральний розклад кореляційної функції аналітичного поля (12.10) має вигляд:

$$\dot{K}_U(\tau, \mathbf{p}) = \int_0^\infty G_0(f) \exp\{j2\pi f \cdot (\tau + (\mathbf{n}, \mathbf{p})/c_0)\} df, \quad (12.11)$$

де $G_0(f) = 2g_0(f)$ - СЩП аналітичного сигналу джерела $\dot{u}_0(t)$. Далі візьмемо до уваги ту типову для радіосистем обставину, що спектри сигналів звичайно обмежені за смугою деяким інтервалом частот $|f - f_0| \leq \Delta f/2$, де f_0 - центральна частота спектру сигналу, Δf - його смуга. Це дозволяє відповідним чином обмежити границі інтегрування у виразі (12.11), тобто:

$$\dot{K}_U(\tau, \mathbf{p}) = \int_{f_0 - \Delta f/2}^{f_0 + \Delta f/2} G_0(f) \exp\{j2\pi f \cdot (\tau + (\mathbf{n}, \mathbf{p})/c_0)\} df.$$

Після заміни $F = f - f_0$ маємо:

$$\begin{aligned} \dot{K}_U(\tau, \mathbf{p}) &= \exp\{j2\pi f_0 \cdot (\tau + (\mathbf{n}, \mathbf{p})/c_0)\} \times \\ &\times \int_{-\Delta f/2}^{\Delta f/2} G_0(F + f_0) \exp\{j2\pi F \cdot \tau\} \cdot \exp\{j2\pi F \cdot (\mathbf{n}, \mathbf{p})/c_0\} dF, \end{aligned} \quad (12.12)$$

З аналізу виразу (12.12) видно, що розділення кореляційної функції аналітичного поля за часовими та просторовими аргументами можливо за умови, що у підінтегральному виразі другий експоненційний множник можна приблизно замінити одиницею, тобто:

$$\exp\{j2\pi F \cdot (\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_2) / c_0\} \cong 1, \quad (12.13)$$

$$F \in (-\Delta f / 2, \Delta f / 2), \quad \mathbf{r}_1, \mathbf{r}_2 \in \mathbf{D}_A.$$

Легко бачити, що при виконанні даної умови вираз (12.12) може бути представлено у вигляді:

$$\dot{K}_U(\tau, \mathbf{p}) \cong \dot{B}_0(\tau) \cdot \exp\{j2\pi f_0 \cdot (\mathbf{n}, \mathbf{p}) / c_0\}, \quad (12.14)$$

де

$$\dot{B}_0(\tau) = \int_{-\Delta f / 2}^{\Delta f / 2} G_0(F + f_0) \exp\{j2\pi (F + f_0) \cdot \tau\} dF$$

- кореляційна функція аналітичного сигналу $\dot{u}_0(t)$. Отже, за умови (12.13) бажане розділення просторово-часової кореляційної функції випадкового поля (12.10) досягнуто. Залишилося лише надати цій умові більш фізично прозорого вигляду. З виразу (12.13) легко бачити, що дана умова задовольняється, якщо

$$\max_{\substack{F \in (-\Delta f / 2, \Delta f / 2) \\ \mathbf{r}_1, \mathbf{r}_2 \in \mathbf{D}_A}} \{F \cdot (\mathbf{n}, \mathbf{r}_1 - \mathbf{r}_2) / c_0\} \ll 1,$$

тобто, коли показник експоненти у виразі (12.13) для усього діапазону змін частоти та координат точок спостереження не перевищує малу долю від періоду експоненти 2π . Аналіз діапазону змін правої частини цього співвідношення приводить до наступної еквівалентної умови:

$$\frac{\Delta f}{f_0} \ll \frac{1}{D_{\max} / \lambda_0}, \quad (12.15)$$

де $D_{\max}$ - максимальний розмір приймальної апертури, $\lambda_0 = c_0 / f_0$ - довжина хвилі, що відповідає центральній частоті смуги сигналу. Співвідношення (12.15) іноді називають [25] **просторовою умовою вузької смуги** сигналу, хоча ця назва й не стала загальноприйнятою. Відзначимо, що умова (12.15) схожа на звичайну умову вузької смуги сигналу, до якої ми вже неодноразово зверталися, а саме:

$$\frac{\Delta f}{f_0} \ll 1. \quad (12.16)$$

Але оскільки хвильові розмірі приймальних апертур $D_{\max}/\lambda_0$ для типових радіосистем звичайно значно більші за одиницю, то співвідношення (12.15) є, як правило, більш «жорстким» за звичайну умову (12.16). Втім, для типових радіосистем навіть посилена умова (12.15) у більшості практичних випадків звичайно виконується. У той же час можливість розділення за цієї умови просторово-часової кореляційної функції полів сигналів та завад на окремі просторовий та часовий множники має значні теоретичні та практичні наслідки, про які ми ще не раз будемо говорити на протязі даного Посібника.

12.3. Розклади кореляційної функції випадкових полів.

У розділі 4 ми говорили про винятково важливу роль теореми Вінера-Хінчина у теорії випадкових стаціонарних процесів. Математично ця теорема формулюється як пара перетворень Фур'є (4.16), що пов'язують поміж собою кореляційну функцію стаціонарного процесу та його спектральну щільність потужності. Логічно припустити, що подібні співвідношення можуть мати місце й для випадкових полів за умови їх стаціонарності та статистичної однорідності. Деякі з таких співвідношень ми й розглянемо нижче.

Звернемося знову до моделі поля у вигляді плоскої хвилі (12.5). Цілком логічним узагальненням цієї простої моделі було б збільшення кількості локальних джерел, розподілених якимсь чином у просторі. Отже, модель (12.5) можна розглядати як окремий випадок наступної моделі:

$$U(t, \mathbf{r}) = \sum_k u_k(t + (\mathbf{n}_k, \mathbf{r})/c_0), \quad t \in (0, T), \mathbf{r} \in \mathbf{D}_A, \quad (12.17)$$

де $u_k(t)$ - сигнал k -го джерела в умовному центрі приймальної апертури, $\mathbf{n}_k$ - орт напрямку приходу відповідної плоскої хвилі. Якщо джерела стаціонарні та поміж собою статистично незалежні, то по аналогії з (12.9) для просторово-часової функції поля (12.17) можна записати наступний розклад:

$$K_U(\tau, \mathbf{p}) = \sum_k \int_{-\infty}^{\infty} g_k(f) \exp\{j2\pi f \cdot (\tau + (\mathbf{n}_k, \mathbf{p})/c_0)\} df, \quad (12.18)$$

де $g_k(f)$ - СЩП k -го джерела, $(k=1,2,\dots)$. Із виразу (12.18) видно, що й для такої узагальненої моделі поле залишається стаціонарним та статистично однорідним оскільки моменти часу та просторові координати входять не окремо, а у вигляді різницевих аргументів $\tau = t_1 - t_2$, $\mathbf{p} = \mathbf{r}_1 - \mathbf{r}_2$.

Тепер зробимо ще один принциповий крок, а саме, припустимо, що джерел настільки багато, що вони щільно заповнюють усі можливі фізичні напрямки. Тоді суму в (12.18) слід замінити інтегралом по усіх можливих напрямках $\mathbf{n}$. У стандартній сферичній системі координат [9] кожен з напрямків задається двома кутами: кутом місця $\theta \in (0, \pi)$ та азимутом $\varphi \in (0, 2\pi)$. При цьому орт напрямку $\mathbf{n}$ записується у вигляді $\mathbf{n} = \{\sin \theta \cdot \cos \varphi; \sin \theta \cdot \sin \varphi; \cos \theta\}$. Отже, в результаті вираз (12.18) переходить у наступний інтегральний розклад:

$$K_U(\tau, \mathbf{p}) = \int_{-\infty}^{\infty} \int_0^{2\pi} \int_0^{\pi} g_U(f; \theta, \varphi) \exp\{j2\pi f \cdot (\tau + (\mathbf{n}, \mathbf{p})/c_0)\} \sin \theta d\theta d\varphi df. \quad (12.19)$$

Легко бачити, що вираз (12.19) є аналогом другого з пари перетворень (4.16) теореми Вінера-Хінчина для випадкових стаціонарних процесів. Але тут, на відміну від (4.16), фігурує вже не спектральна щільність потужності процесу (СЩП), а спектрально-кутова щільність потужності⁵ (СКЩП) поля $g_U(f; \theta, \varphi)$, яка відображає розподіл його потужності не лише за частотами але й за фізичними напрямками. Проте слід зазначити, що на відміну від теореми Вінера-Хінчина у даному разі зворотне до (12.19) перетворення записати вже не так просто. Річ у тім, що розклад (12.19) обмежено лише реальними напрямками, що зумовлено фізичною сутністю задачі. При цьому безпосереднє застосування до виразу (12.19) зворотного перетворення Фур'є неможливе: необхідна додаткова інформація, про що сказано в [7]. Ще одне фізичне обмеження щодо розкладу (12.19) полягає в тому, що він може бути застосований лише для тих областей фізичного простору, де практично нема

⁵ Математично СКЩП може існувати лише для стаціонарних та однорідних полів. Тут є повна аналогія з теоремою Вінера-Хінчина, яка має місце лише для стаціонарних процесів.

неоднорідних плоских хвиль. Для цього такі області мають бути віддалені від джерел, границь середовища та «непрозорих» об'єктів [25].

Але не зважаючи на вказані обмеження розклад (12.19) широко використовується для вирішення низки прикладних завдань, деякі з яких будуть розглянуті в наступних розділах Посібника. Зазначимо лише, що цей розклад у практичних застосуваннях прийнято записувати дещо в іншому вигляді [16], а саме:

$$K_U(\tau, \mathbf{p}) = \int_{-\infty}^{\infty} g_U(f) R(f, \mathbf{p}) \exp\{j2\pi f\tau\} df, \quad (12.20)$$

$$\text{де} \quad g_U(f) = \int_0^{2\pi} \int_0^{\pi} g_U(f; \theta, \varphi) \sin \theta d\theta d\varphi \quad (12.21)$$

- усереднена за усіма напрямками спектральна щільність потужності поля,

$$R(f, \mathbf{p}) = \int_0^{2\pi} \int_0^{\pi} C_U(f; \theta, \varphi) \exp\{j2\pi f \cdot (\mathbf{n}, \mathbf{p}) / c_0\} \sin \theta d\theta d\varphi, \quad (12.22)$$

- функція, яка відповідає за своїм змістом коефіцієнту просторової кореляції стаціонарного та однорідного поля на частоті f ,

$$C_U(f; \theta, \varphi) = g_U(f; \theta, \varphi) / g_U(f) \quad (12.23)$$

- нормований кутовий спектр поля (або PAS від Power Angular Spectrum). Зі співвідношень (12.20) – (12.23) неважко бачити їх повну еквівалентність до розкладу (12.19). Крім того, з визначень (12.21), (12.23) витікає умова нормування кутового спектру поля, а саме:

$$\int_0^{2\pi} \int_0^{\pi} C_U(f; \theta, \varphi) \sin \theta d\theta d\varphi = 1. \quad (12.24)$$

Отже, співвідношення (12.20), (12.22) відіграють у теорії випадкових полів фактично ту ж саму роль, що й теорема Вінера-Хінчина у теорії випадкових процесів. Розглянемо окремі випадки цих співвідношень, які зустрічаються у практичних застосуваннях та знадобляться нам у наступних розділах Посібника.

У всіх цих прикладах будемо вважати, що кутовий спектр поля не залежить від частоти, тобто, $C_U(f; \theta, \varphi) = C_U(\theta, \varphi)$. Такі поля називають **рівно забарвленими** [16]. Це означає, що «забарвлення», тобто спектральний склад поля, за усіма напрямками однаковий. Але залежність загальних рівнів потужності від напрямку зберігається, що визначається кутовим спектром поля $C_U(\theta, \varphi)$.

Першу модель, яку ми розглянемо в рамках цих припущень, це так званий **ізотропний шум**. Так називають випадкове поле, кутовий спектр якого є рівномірним для усіх фізичних напрямків, тобто, $C_U(\theta, \varphi) = \text{const}$. При цьому значення константи вибирається таким чином, щоб задовольнялася умова нормування (12.24). Отже, з урахуванням вказаної умови для кутового спектру ізотропного шуму неважко записати:

$$C_U(\theta, \varphi) = \frac{1}{4\pi}, \quad \theta \in (0, \pi), \varphi \in (0, 2\pi). \quad (12.25)$$

Після підстановки виразу (12.25) до формули (12.22) та виконання інтегрування знаходимо коефіцієнт просторової кореляції ізотропного шуму:

$$R(f, \mathbf{\rho}) = \frac{\sin(2\pi \rho / \lambda)}{2\pi \rho / \lambda}, \quad (12.26)$$

де $\rho = |\mathbf{\rho}|$ - модуль різницевого вектора просторових координат, $\lambda = c_0 / f$ - довжина хвилі на частоті f . Просторово-часова кореляційна функція такого поля може бути одержана після підстановки виразу (12.26) до формули (12.20):

$$K_U(\tau, \mathbf{\rho}) = \int_{-\infty}^{\infty} g_U(f) \frac{\sin(2\pi f \rho / c_0)}{2\pi f \rho / c_0} \exp\{j2\pi f \tau\} df.$$

Зокрема, якщо СЦП шуму «біла», тобто, $g_U(f) = g_0$, то після виконання інтегрування маємо:

$$K_U(\tau, \mathbf{\rho}) = \begin{cases} g_0 c_0 / (2\rho), & |\tau| \leq \rho / c_0; \\ 0, & |\tau| > \rho / c_0. \end{cases} \quad (12.27)$$

Отже, характерною особливістю ізотропного шуму є те, що його просторова кореляція не залежить від орієнтації різницевого вектору просторових координат $\mathbf{p}$. При додатковому припущенні про «білий» у спектральній області характер шуму його просторово-часова кореляційна функція має «ступінчастий» вигляд (12.27). Деякі результати обчислень, що ілюструють формулу (12.26), представлено на рис.12.1.

Розвитком попередньої моделі є так званий *горизонтально ізотропний шум*⁶, який характеризується тим, що його кутовий спектр не залежить від кута азимуту φ , тобто,

$$C_U(\theta, \varphi) = C_U(\theta). \quad (12.28)$$

З урахуванням цього припущення коефіцієнт просторової кореляції (12.22) приймає вигляд:

$$R(f, \mathbf{p}) = \int_0^\pi C_U(\theta) \left(\int_0^{2\pi} \exp\{j2\pi f \cdot \sin\theta \cdot (\rho_x \cos\varphi + \rho_y \sin\varphi) / c_0\} d\varphi \right) \times \\ \times \exp(j2\pi f \cdot \rho_z \cdot \cos\theta / c_0) \sin\theta d\theta.$$

Після виконання інтегрування по φ приходимо до наступного результату:

$$R(f, \mathbf{p}) = 2\pi \int_0^\pi C_U(\theta) J_0\left(\frac{2\pi}{\lambda} \cdot \sin\theta \cdot \sqrt{\rho_x^2 + \rho_y^2}\right) \cdot \exp\left(j \frac{2\pi}{\lambda} \cdot \rho_z \cdot \cos\theta\right) \sin\theta d\theta, \quad (12.29)$$

де $J_0(z)$ - функція Бесселя нульового порядку. Як бачимо з виразу (12.29), просторова кореляція такого шуму не залежить від орієнтації різницевого вектору просторових координат у горизонтальній площині, але залежить від його вертикальної компоненти ρ_z .

Тепер в рамках прийнятих припущень розглянемо модель *секторного шуму*, тобто, поля, яке є ізотропним у секторі кутів $\theta \in (\theta_1, \theta_2)$. З урахуванням умови нормування (12.24) для кутового спектру такого шуму неважко одержати наступний вираз:

⁶ Нагадаємо, що вибір координат у кожній фізичній задачі диктується міркуваннями зручності її вирішення. Тому якщо фізично ізотропність має місце, наприклад, не в горизонтальній, а в якийсь інший площині, то це означає, що систему координат треба лише повернути так, як зручно.

$$C_U(\theta, \varphi) = \begin{cases} \frac{1}{2\pi(\cos\theta_1 - \cos\theta_2)}, & \theta \in (\theta_1, \theta_2); \\ 0, & \theta \notin (\theta_1, \theta_2). \end{cases} \quad (12.30)$$

Загальний аналітичний вираз просторової кореляції поля для цієї моделі шуму одержати не вдається. Тому обмежимося лише кореляцією у вертикальній площині. Для цього підставимо (12.30) до виразу (12.29), прийняв у ньому $\rho_x = \rho_y = 0$. Після виконання інтегрування приходимо до наступного результату:

$$R(f; 0, 0, \rho_z) = \frac{\sin\left(\pi \frac{\rho_z}{\lambda} (\cos\theta_1 - \cos\theta_2)\right)}{\pi \frac{\rho_z}{\lambda} (\cos\theta_1 - \cos\theta_2)} \exp\left(j\pi \frac{\rho_z}{\lambda} (\cos\theta_1 + \cos\theta_2)\right). \quad (12.31)$$

Легко бачити, що при значенні $\theta_1 = 0, \theta_2 = \pi$ цей вираз переходить у формулу (12.26).

Останній випадок, який ми розглянемо у цьому розділі, відомий під назвою *одно направлений шум* [7]. Він відповідає ситуації, коли джерела шуму дипольного типу розташовані на деякій необмеженій площині, а приймання сигналу відбувається в області, що розташована нижче (або вище) цієї площини. Даній моделі відповідає кутовий спектр поля

$$C_U(\theta) = \begin{cases} \frac{m+1}{2\pi} \cos^m \theta, & \theta \in (0, \pi/2); \\ 0 & \theta \in (\pi/2, \pi), \end{cases} \quad (12.32)$$

де m - порядок дипольного джерела, який визначається фізичними умовами задачі. Скориставшись формулою (12.29) та прийнявши в ній $\rho_z = 0$, знаходимо коефіцієнт просторової кореляції такого поля у горизонтальній площині:

$$R(f; \rho_x, \rho_y, 0) = 2^{(m-1)/2} \cdot (m+1) \cdot \Gamma\left(\frac{m+1}{2}\right) \frac{J_{(m+1)/2}\left(\frac{2\pi}{\lambda} \sqrt{\rho_x^2 + \rho_y^2}\right)}{\left(\frac{2\pi}{\lambda} \sqrt{\rho_x^2 + \rho_y^2}\right)^{(m+1)/2}}, \quad (12.33)$$

де $\Gamma(\alpha)$ - гамма-функція, $J_p(z)$ - функція Бесселя p -го порядку.

Окремий випадок даної моделі, що відповідає значенню параметру $m=1$, має окрему назву, відому як **модель Крона-Шермана** [7]. У цьому разі з формули (12.33) витікає, що просторова кореляція поля у горизонтальній площині має вигляд

$$R(f; \rho_x, \rho_y, 0) = 2 \frac{J_1\left(\frac{2\pi}{\lambda} \sqrt{\rho_x^2 + \rho_y^2}\right)}{\frac{2\pi}{\lambda} \sqrt{\rho_x^2 + \rho_y^2}}. \quad (12.34)$$

Модель Крона-Шермана вперше була запропонована для опису такого важливого в океанології явища, як вітровий шум морської поверхні. Але рамки застосування цієї моделі, як й більш загальної моделі (12.32), завданнями лише океанології не вичерпуються. Можна знайти немало ситуацій й в галузі радіозв'язку, які непогано «вкладаються» в умови такої моделі. Для ілюстрації деякі розрахунки за формулою (12.33) показані на рис.12.1.

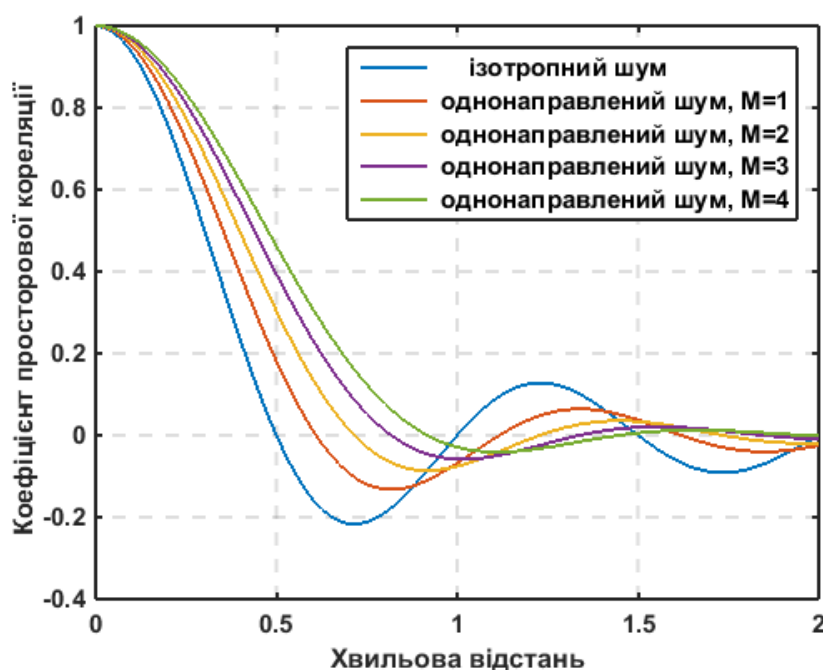


Рис.12.1. Ілюстрація розрахунків за формулами (12.26) та (12.33).

Із рисунку видно, що зростання порядку дипольних джерел M супроводжується й зростанням області підвищеної кореляції поля.

На завершення цього розділу зазначимо, що ми розглянули доволі обмежено коло моделей, в яких було відносно нескладно отримати явні аналітичні результати щодо просторової кореляції випадкових полів. Але усі ці випадки є лише ілюстрацією теоретичних ідей та деяких практичних підходів до вирішення вказаного завдання. Втім, у кожному конкретному випадку можна виконати подібні розрахунки й безпосередніми чисельними методами, що при сучасному стані цих методів та відповідних обчислювальних засобів не становить проблеми.

Контрольні запитання та завдання до розділу 12.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Що таке випадкові поля та чим вони відрізняються від випадкових процесів? Яка інформація необхідна для повного опису випадкових гавсівських полів?
2. Класифікація випадкових полів за ознаками стаціонарності та статистичної однорідності. В чому полягає різниця поміж «вузькими» та «широкими» визначеннями стаціонарності та однорідності?
3. Чи можуть існувати стаціонарні та однорідні поля фізично? Що таке квазістаціонарні та квазіоднорідні поля?
4. За яких умов просторово-часова кореляційна функція поля віддаленого джерела може бути представлена у вигляді добутку «часового» та «просторового» множників? Як можна ці умови пояснити фізично?
5. В яких фізичних умовах може бути застосовано розклад (12.19)?
6. Які параметри входять до виразів (12.20) – (12.24)? Який вони мають фізичний зміст?
7. Що таке ізотропний шум? Як фізично можна пояснити виникнення такого шуму? Одержати самостійно формули (12.25), (12.26).
8. Що таке горизонтально-ізотропний шум? Які він має особливості? Самостійно одержати формулу (12.29).
9. Самостійно одержати формули (12.30), (12.31) для секторного шуму.
10. Самостійно одержати формули (12.33), (12.34) для однонаправленого шуму. Самостійно повторити графіки рис.12.1. Як фізично пояснити зміни коефіцієнту просторової кореляції шуму зі зростанням порядку дипольного джерела M ?
11. Порівняти результати обчислень у середовищі Матлаб за формулою (12.34) та шляхом чисельного інтегрування формули (12.29).

Розділ 13. Оптимальні та квазіоптимальні алгоритми лінійного оброблення сигналів у часі та просторі

У розділі 11 було розглянуто алгоритми виявлення сигналів кількох типів. Але структури оброблення сигналів, що їм відповідають, мають надто загальний вигляд, який не пристосовано для практичної реалізації. Крім того, не з'ясовано поки що зв'язок цих структур із канонічними схемами приймання сигналів, які склалися у «традиційній» радіотехніці задовго до появи статистичної теорії. Розгляд вказаних питань й складає предмет даного розділу.

13.1. Приймання квазідетермінованих сигналів на тлі білого шуму.

Перш за все, зауважимо, що поміж процедурами приймання сигналів різних типів існує суттєва спільність. Зокрема, алгоритм виявлення квазідетермінованих сигналів передбачає виконання на першому етапі лінійної обробки (11.20), вигляд якої від типу каналу передачі не залежить. Специфіка каналу впливає лише на вибір наступних нелінійних перетворень. Таким чином, алгоритм лінійної обробки (11.20) з усіма своїми окремими випадками складає важливе ядро алгоритму приймання квазідетермінованих сигналів. Також нескладно помітити, що дійсна частина виразу (11.20) співпадає з алгоритмом оброблення детермінованих сигналів (11.11). Тому в даному разі не має необхідності окремо розглядати випадок приймання детермінованих сигналів. Більше того, нижче буде показано, що схожий алгоритм лінійного оброблення (при певних припущеннях) може знайти застосування й для завдань приймання випадкових сигналів. До цього слід лише додати, що саме лінійна частина обробки сигналів практично у будь-який приймальній системі визначає у великій мірі її завадостійкість.

Отже, з огляду на вищесказане, проаналізуємо більш детально алгоритм лінійної обробки сигналів (11.20). Поки що обмежимося кореляційною обробкою (11.23), яка є окремим випадком алгоритму (11.20) для задачі приймання сигналів на тлі білого шуму. Звернемося до деяких моделей поля

сигналу, що обговорювалися у розділі 9. Зокрема, у дальній зоні (зоні Фраунгофера) точкового джерела (див. рис.9.8) поле сигналу має вигляд плоскої хвилі (9.11). Звідси для аналітичного сигналу на виході m -ого датчика AP ($m = 1, \dots, M$) можна записати:

$$\dot{s}(t, \mathbf{r}_m) \cong A_0 \cdot \dot{s}_0(t - \tau_0 + \tau_m(\mathbf{n}_s)), \quad (13.1)$$

де $\dot{s}_0(t)$ - аналітична копія сигналу джерела, A_0 , τ_0 - амплітуда та затримка сигналу джерела в деякому умовному центрі приймальної антени (його звичайно називають фазовим),

$$\tau_m(\mathbf{n}_s) = (\mathbf{n}_s, \mathbf{r}_m) / c_0 \quad (13.2)$$

– затримка сигналу датчика відносно фазового центру, $\mathbf{r}_m$ - координати датчика, $\mathbf{n}_s$ - одиничний вектор напрямку приходу сигналу, c_0 - швидкість поширення сигналу в фізичному середовищі.

Припустимо, що в точці випромінювання сигнал існує лише на обмеженому часовому інтервалі $t \in (0, T)$. Тоді на m -ому датчику цей сигнал буде, вочевидь, існувати на часовому інтервалі $t \in (\tau_0 - \tau_m, T + \tau_0 - \tau_m)$. Беручи до уваги це припущення, після підстановки моделі поля сигналу (13.1) до алгоритму (11.23) одержимо:

$$\dot{V}_0(\mathbf{u} / \tau_0, \mathbf{n}_s) = \frac{A_0}{g_w} \sum_{m=1}^M \int_{\tau_0 - \tau_m}^{T + \tau_0 - \tau_m} u(t, \mathbf{r}_m) \cdot \dot{s}_0(t - \tau_0 + \tau_m(\mathbf{n}_s)) dt.$$

Після заміни $t + \tau_m = t'$ маємо:

$$\dot{V}_0(\mathbf{u} / \tau_0, \mathbf{n}_s) = \frac{A_0}{g_w} \sum_{m=1}^M \int_{\tau_0}^{T + \tau_0} u(t - \tau_m(\mathbf{n}_s), \mathbf{r}_m) \cdot \dot{s}_0(t - \tau_0) dt.$$

Внесемо знак суми під знак інтегралу. В результаті даний алгоритм приймає вигляд:

$$\dot{V}_0(\mathbf{u} / \tau_0, \mathbf{n}_s) = \frac{A_0}{g_w} \int_{\tau_0}^{T + \tau_0} u_{AR}(t, \mathbf{n}_s) \cdot \dot{s}_0(t - \tau_0) dt, \quad (13.3)$$

де

$$u_{AR}(t, \mathbf{n}_s) = \sum_{m=1}^M u(t - \tau_m(\mathbf{n}_s), \mathbf{r}_m). \quad (13.4)$$

З розгляду виразів (13.3) та (13.4) видно, що для прийнятої моделі поля сигналу кореляційний алгоритм (11.23) може бути представлено як каскадне поєднання двох підсистем. Першу з них складає алгоритм (13.4), який зводиться до підсумовування значень сигналів по виходу датчиків антени з компенсацією їх затримок відносно фазового центру. Одержаний в результаті виконання алгоритму (13.4) сигнал подається на вхід другої підсистеми, яка в даному разі має вигляд одно каналної кореляційної обробки (13.3).

Блок-схема оброблення сигналів у відповідності до цих алгоритмів представлена на рис.12.1. Схема складається з двох частин. Просторова обробка включає лінії затримки (Л.З.), які компенсують різницю ходу сигналів до відповідних датчиків АР та забезпечують тим самим їх синхронність перед підсумовуванням. Часова обробка полягає у кореляції прийнятого сигналу з його копією, яка зсунута у часі на очікуване значення затримки сигналу.

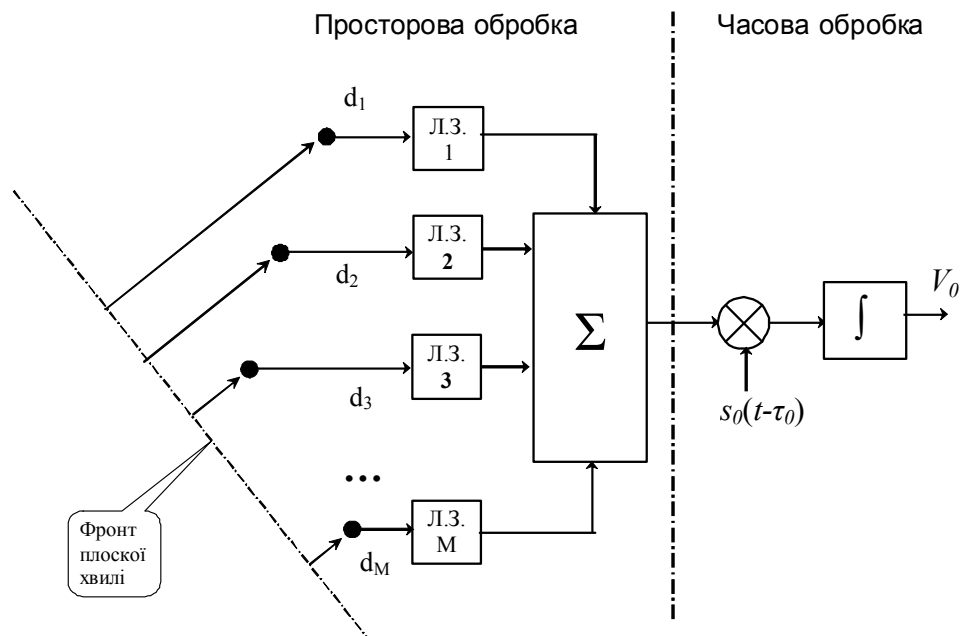


Рис.12.1. Блок-схема кореляційної обробки поля сигналу у просторі та часі.

В теорії приймання та оброблення сигналів подібне розділення лінійної обробки на дві окремі процедури має широке застосування, хоча в

загальному випадку складові алгоритму можуть виглядати й дещо інакше. При цьому суть алгоритму (13.4), який у традиційній радіотехніці називають формуванням направленості (ФН) антени, полягає у синхронному накопиченні сигналу по апертурі АР. Для виконання такого накопичення необхідно знати лише властивості сигналу як хвилі у просторі, геометрію та параметри антени. Тому алгоритм (13.4) та йому подібні прийнято називати ***просторовою обробкою сигналів*** або ***обробленням сигналів у просторі***. Наступну ж частину лінійної обробки, яку в даному разі складає алгоритм (13.3), називають ***часовою обробкою сигналів*** або ***обробленням сигналів у часі***.

Розглянута схема відображає лише ідею алгоритмів (13.3), (13.4). Практична реалізація цієї схеми сучасними засобами DSP може виглядати інакше. При виборі варіанту конкретної реалізації беруть до уваги як архітектуру заданої апаратно-програмної платформи, так й можливості економії обчислювальних затрат за рахунок перетворення алгоритмів до більш зручної математичної форми. Розглянемо деякі з найбільш відомих версій цих алгоритмів. Почнемо з просторової обробки.

Перш за все зазначимо, що алгоритм (13.4) може бути застосовано практично до будь яких сигналів: головне, щоб була відома форма фронту сигнальної хвилі у просторі. Зокрема, як буде показано нижче, подібний же алгоритм має застосування й для приймання випадкових сигналів. Також відзначимо, що алгоритм (13.4) не накладає будь-яких обмежень на ширину смуги сигналу. Але переважна більшість радіосигналів є, як відомо, вузько смуговими. Тому в практиці приймання радіосигналів замість затримок на вході ФН найчастіше використовують так зване ***фазування***, а саме,

$$u_{AR}(t, \mathbf{n}_s) = \sum_{m=1}^M \operatorname{Re}\{\dot{u}(t, \mathbf{r}_m) \cdot \exp(-j\omega_0 \tau_m(\mathbf{n}_s))\}, \quad (13.5)$$

де $\dot{u}(t, \mathbf{r}_m)$ - аналітичне представлення сигналу $u(t, \mathbf{r}_m)$ на вході ФН¹. Аналіз показує, що алгоритм (13.5) практично не поступається за ефективністю алгоритму (13.4), якщо виконується *просторова умова вузької смуги сигналу*, про яку вже йшла мова у попередньому розділі (вираз (12.15)), а саме:

$$\frac{\Delta f_s}{f_0} \ll \frac{1}{D_{\max} / \lambda_0}. \quad (13.6)$$

Тут, нагадаймо, Δf_s - смуга, f_0 - носійна частота сигналу, $D_{\max}$ - максимальний розмір просторової апертури, $\lambda_0 = c_0 / f_0$ - довжина хвилі на центральній частоті сигналу. Легко бачити, що для антен великих хвильових розмірів умова (13.6) значно жорсткіша стандартної умови вузької смуги (12.16). Втім, як вже було сказано у розділі 12, для переважної більшості радіосистем умова вузької смуги звичайно виконується як у стандартному (12.16), так й у «посиленому» варіанті (13.6).

Також слід зазначити, що в сучасних радіосистемах з метою економії обчислювальних ресурсів цифрове оброблення сигналів виконується, як правило, вже після формування їх комплексної обвідної. Отже, й цифрове формування направленості має виконуватись безпосередньо для комплексної обвідної сигналу. Відповідний алгоритм безпосередньо витікає з виразу (13.5) та має вигляд:

$$\dot{U}_{AR}(t, \mathbf{n}_s) = \sum_{m=1}^M \dot{U}(t, \mathbf{r}_m) \cdot \exp(-j\omega_0 \tau_m(\mathbf{n}_s)), \quad (13.7)$$

де $\dot{U}_{AR}(t / \mathbf{n}_s)$, $\dot{U}(t, \mathbf{r}_m)$ - комплексні обвідні сигналів $u_{AR}(t / \mathbf{n}_s)$ та $u(t, \mathbf{r}_m)$ відповідно. В принципі, повернутися від комплексної обвідної $\dot{U}_{AR}(t / \mathbf{n}_s)$ до сигналу $u_{AR}(t / \mathbf{n}_s)$ неважко (див. §1.3). Але в цьому, як правило, не має потреби, оскільки наступні етапи оброблення також звичайно виконуються безпосередньо над комплексною обвідною.

¹ Знак показника експоненти у виразі (13.5) визначається прийнятою системою координат, тому в кожному конкретному випадку це слід додатково уточнювати.

Тепер перейдемо до розгляду часової обробки, яка в даному разі представлена алгоритмом (13.3). Перш за все відзначимо, що алгоритм (13.3) шляхом заміни змінної інтегрування можна записати в наступному еквівалентному вигляді:

$$\dot{V}_0(\mathbf{u}/\tau_0, \mathbf{n}_s) = \frac{A_0}{g_w} \int_0^T u_{AR}(T + \tau_0 - t, \mathbf{n}_s) \cdot \dot{h}_0(t) dt, \quad (13.8)$$

$$\text{де} \quad \dot{h}_0(t) = s_0(T - t). \quad (13.9)$$

Вираз (13.8) відповідає алгоритму фільтрації сигналу по виходу ФН з імпульсним відгуком (13.9). Легко бачити, що частотна характеристика такого фільтру співпадає зі спряженим спектром очікуваного сигналу². Подібні фільтри ще на ранніх етапах розвитку теорії приймання сигналів одержали назву узгоджених [3]. При цьому розумілося саме узгодження характеристики фільтру зі спектром сигналу. В наш час в поняття узгодженого оброблення сигналів часто вкладають більш широкий зміст, наприклад, узгодження обробки зі структурою сигнальної хвилі у просторі, узгодження з параметрами середовища передачі, узгодження з рухом антени, тощо.

Тепер запишемо версію алгоритму (13.3), яка б відповідала обробці комплексної обвідної сигналу. Для цього скористаємося представленням сигналу на вході часової обробки через його комплексну обвідну, тобто:

$$u_{AR}(t) = \operatorname{Re}\{\dot{U}_{AR}(t) \cdot \exp(j\omega_0 t)\}.$$

Відповідне представлення копії очікуваного сигналу має вигляд

$$\dot{s}_0(t) = \dot{S}_0(t) \cdot \exp(j\omega_0 t).$$

Після підстановки цих виразів до алгоритму (13.3), маємо:

$$\dot{V}_0(\mathbf{u}/\tau_0, \mathbf{n}_s) = \frac{A_0}{g_w} \int_{\tau_0}^{T+\tau_0} \operatorname{Re}\{\dot{U}_{AR}(t, \mathbf{n}_s) \cdot \exp(j\omega_0 t)\} \cdot \dot{S}_0(t - \tau_0) \cdot \exp(j\omega_0(t - \tau_0)) dt.$$

Далі скористаємося тією очевидною обставиною, що дійсна частина довільного комплексного числа $\dot{z}$ може бути представлена у вигляді

² Для цього достатньо виконати перетворення Фур'є від обох частин рівності (13.9).

$$\operatorname{Re}(\dot{z}) = \frac{1}{2}(\dot{z} + \dot{z}^*).$$

В результаті маємо:

$$\begin{aligned} \dot{V}_0(\mathbf{u}/\tau_0, \mathbf{n}_s) = & \frac{A_0}{2g_w} \exp(-j\omega_0\tau_0) \int_{\tau_0}^{T+\tau_0} \dot{U}_{AR}(t, \mathbf{n}_s) \cdot \dot{S}_0(t-\tau_0) \cdot \exp(j2\omega_0 t) dt + \\ & + \frac{A_0}{2g_w} \exp(-j\omega_0\tau_0) \int_{\tau_0}^{T+\tau_0} \dot{U}_{AR}^*(t, \mathbf{n}_s) \cdot \dot{S}_0(t-\tau_0) dt. \end{aligned}$$

Перший з додатків цього співвідношення містить у виразі під інтегралом швидко осцилюючу комплексну гармоніку подвоєної частоти. Після інтегрування на інтервалі, що дорівнює тривалості сигналу, цим додатком можна, як правило, знехтувати. Отже, остаточно маємо алгоритм часового оброблення комплексної обвідної сигналу на виході ФН:

$$\dot{V}_0(\mathbf{u}/\tau_0, \mathbf{n}_s) = \left(\frac{A_0}{2g_w} \int_{\tau_0}^{T+\tau_0} \dot{U}_{AR}^*(t, \mathbf{n}_s) \cdot \dot{S}_0(t-\tau_0) dt \right) \cdot \exp(-j\omega_0\tau_0) \quad (13.10)$$

Якщо розглядати це співвідношення як функцію часового параметру τ_0 , то вираз у дужках є комплексною обвідною сигналу по виходу лінійної обробки, комплексна експонента – високочастотне заповнення, яке на сутність справи не впливає.

13.2. Приймання квазідетермінованих сигналів на тлі корельованих завад.

У попередньому параграфі були розглянуті оптимальні алгоритми лінійного оброблення сигналів на тлі білого шуму. З'ясувалося, що ці алгоритми практично співпадають з тими, які вже були відомі в традиційній радіотехніці задовго до появи статистичної теорії. Виникає слушне запитання: чи дає статистична теорія якісь додаткові можливості у порівнянні з традиційними підходами?

Відповідь на це питання безумовно позитивна і головна перевага статистичної теорії перш за все полягає у більш загальному та системному підході до вирішення завдань приймання та оброблення сигналів. Крім того,

статистичний підхід відкриває можливості для запозичення ідей та технологій, які у первісному вигляді призначалися для інших, іноді достатньо далеких по застосуванню прикладних галузей. У цьому параграфі ми проілюструємо деякі можливості статистичної теорії на прикладі задачі оптимізації алгоритмів оброблення сигналів на тлі корельованих завад. Врахування кореляції завад – це додаткові можливості для підвищення ефективності приймання сигналів. Треба лише цими можливостями належним чином скористатися, що, власне, статистична теорія й пропонує.

Отже, повернемося до загального алгоритму лінійної обробки (11.20). Його головна відмінність у порівнянні з розглянутим вище окремим випадком (11.23) полягає в тому, що тепер одержані на вході АР спостереження $u(t, \mathbf{r}_m)$ треба корелювати не з очікуваним сигналом, а з деякою оптимальною ваговою функцією $\dot{b}(t, \mathbf{r}_m)$, тобто,

$$V_0(\mathbf{u}) = (\dot{\mathbf{b}}, \mathbf{u}) = \sum_{m=1}^M \int_{\mathbf{r}_m} \dot{b}(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt. \quad (13.11)$$

У свою чергу, оптимальна вагова функція, як показано в розділі 11, знаходиться в результаті вирішення системи інтегральних рівнянь

$$g_w \cdot \dot{b}(t, \mathbf{r}_m) + \sum_{p=1}^M \int_{\mathbf{r}_p} K_z(t, t'; \mathbf{r}_m, \mathbf{r}_p) \cdot \dot{b}(t', \mathbf{r}_p) dt' = \dot{s}(t, \mathbf{r}_m); \quad (13.12)$$

$$m = 1, \dots, M.$$

Рішення цієї системи рівнянь пов'язано зі значними математичними труднощами, які найчастіше вдається подолати лише за рахунок деяких спрощень та наближень [3]. При цьому дуже зручним виявляється прийнятий у Посібнику підхід, що базується на змішаному дискретно-безперервному представленні алгоритмів просторово-часового оброблення сигналів та про який йшла мова у розділі 11. Зокрема, для вказаного «змішаного» представлення доцільно шукати рішення рівнянь (13.12) не в часовій, а в спектральній області. Для цього достатньо попередньо виконати перетворення Фур'є від обох частин рівняння (13.12) по змінній t . Але

застосування вказаного методу є доцільним лише за виконання певних умов [25], а саме:

1. На інтервалі спостереження сигналу має виконуватися припущення про **квазістаціонарність завад**. Це означає, що в межах інтервалу спостереження їх кореляційна функція залежить лише від різниці своїх часових аргументів, а саме:

$$K_z(t, t'; \mathbf{r}_m, \mathbf{r}_p) = K_z(t - t'; \mathbf{r}_m, \mathbf{r}_p).$$

Ця обставина, у свою чергу, дозволяє скористатися теоремою Вінера-Хінчина (див. розділ 4), відповідно до якої просторово-часова кореляційна функція по своєму часовому аргументу може бути представлена розкладом Фур'є

$$K_z(t - t'; \mathbf{r}, \mathbf{r}') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \dot{K}_z(\omega; \mathbf{r}, \mathbf{r}') \exp\{j\omega(t - t')\} d\omega, \quad (13.13)$$

де $\dot{K}_z(\omega; \mathbf{r}, \mathbf{r}')$ – просторова кореляційна функція поля завад на частоті ω . Ця функція, відповідно до тієї ж теореми, знаходиться як пряме перетворення Фур'є від кореляційної функції квазістаціонарного поля завад $K_z(\tau; \mathbf{r}, \mathbf{r}')$ по аргументу τ . Її можна також визначити за допомогою граничного співвідношення [16]:

$$\dot{K}_z(\omega; \mathbf{r}, \mathbf{r}') = \lim_{T \rightarrow \infty} \left\langle \frac{1}{T} \cdot \dot{Z}_T(\omega, \mathbf{r}) \cdot \dot{Z}_T^*(\omega, \mathbf{r}') \right\rangle \quad (13.14)$$

де $\dot{Z}_T(\omega, \mathbf{r})$ - перетворення Фур'є від поля завад $z(t, \mathbf{r})$ на часовому інтервалі тривалістю T . При $\mathbf{r}' = \mathbf{r}$ функція $\dot{K}_z(\omega; \mathbf{r}, \mathbf{r}')$ співпадає зі спектральною щільністю потужності завад, яка, взагалі кажучи, може залежати від точки приймання, тобто, $\dot{K}_z(\omega; \mathbf{r}, \mathbf{r}) = g_z(\omega, \mathbf{r})$. Для статистично однорідних полів $g_z(\omega, \mathbf{r}) \equiv g_z(\omega)$.

2. Теорема Вінера-Хінчина оперує, взагалі кажучи, з нескінченим часовим інтервалом $t \in (-\infty, \infty)$. В задачах, що розглядаються, цей інтервал

обмежено тривалістю сигналу T . Але якщо ширина ефективної смуги завад Δf_z задовольняє посиленій нерівності

$$\Delta f_z \cdot T \gg 1, \quad (13.15)$$

то з обмеженим часовим інтервалом можна [16] поводитися практично так же, як і з нескінченим.

3. Ще однією складністю при переході до спектральної форми рівнянь (13.12) є залежність границь інтегрування у часі від координат точок прийому. Для подолання цієї складності звичайно застосовується так звана модель одно миттєвих «зрізів» поля [25] (*snapshots model*). Суть моделі полягає в тому, що максимальну різницю у часі пробігу хвиль сигналу в межах просторової апертури обмежено умовою

$$\max_m \tau_m / T \ll 1. \quad (13.16)$$

При виконанні цієї умови у виразах (13.11), (13.12) можна знехтувати залежністю границь інтегрування у часі від координат точки прийому. Зауважимо, що фізично умова (13.16) означає обмеженість розмірів апертури, тобто, її компактність. Тому весь згаданий вище комплекс умов будемо в подальшому називати *квазістаціонарним наближенням для компактних просторових апертур*.

В рамках даного наближення систему рівнянь (13.12) неважко привести до еквівалентного спектрального вигляду:

$$g_w \cdot \dot{B}(\omega, \mathbf{r}_m) + \sum_{p=1}^M \dot{K}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot \dot{B}(\omega, \mathbf{r}_p) = \dot{S}(\omega, \mathbf{r}_m), \quad (13.17)$$

де $\dot{B}(\omega, \mathbf{r}_m)$, $\dot{S}(\omega, \mathbf{r}_m)$ - спектри Фур'є оптимальної вагової функції $\dot{b}(t, \mathbf{r}_m)$ та поля очікуваного сигналу $\dot{s}(t, \mathbf{r}_m)$ відповідно, які обчислюються на співпадаючих часових інтервалах тривалістю T . Отже, за рахунок вказаних спрощень від системи інтегральних рівнянь (13.12) можна перейти до системи алгебраїчних рівнянь (13.17), яку треба вирішувати окремо для кожного значення частоти ω в межах робочого діапазону приймального

пристрою. Потім від рішень $\dot{B}(\omega, \mathbf{r}_m)$ можна, в принципі, через зворотне перетворення Фур'є перейти до оптимальних вагових функцій у часовій області $\dot{b}(t, \mathbf{r}_m)$. Але в реальності це робити не обов'язково. По-перше, алгоритм оптимальної обробки (13.11) при тих самих припущеннях можна представити в еквівалентній спектральній формі:

$$V_0(\mathbf{u}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{m=1}^M \dot{B}^*(\omega, \mathbf{r}_m) \cdot \dot{U}(\omega, \mathbf{r}_m) d\omega, \quad (13.18)$$

де $\dot{U}(\omega, \mathbf{r}_m)$, $m = 1, \dots, M$ - спектри Фур'є поля $u(t, \mathbf{r}_m)$ на апертурі АР на співпадаючих часових інтервалах тривалістю T . Отже, зворотне перетворення Фур'є окремо для спектру вагової функції $\dot{B}(\omega, \mathbf{r}_m)$ в даному разі виконувати й не потрібно. По-друге, в рамках прийнятого наближення можливі й подальші спрощення за рахунок залучення фізичних міркувань щодо моделі поля сигналу. Зокрема, припустимо, як й вище, що поле сигналу є плоскою хвилею (13.1). В спектральній області цій моделі відповідає співвідношення

$$\dot{S}(\omega, \mathbf{r}_m) = \dot{S}_{ch}(\omega / \tau_0) \cdot \exp(-j\omega\tau_m(\mathbf{n}_s)), \quad (13.19)$$

$$\text{де} \quad \dot{S}_{ch}(\omega / \tau_0) = A_0 \cdot \dot{S}_0(\omega) \cdot \exp(j\omega\tau_0) \quad (13.20)$$

- спектр сигналу по виходу каналу передачі у фазовому центрі АР, $\dot{S}_0(\omega)$ - спектр сигналу джерела. Усі інші параметри ті ж самі, що й у моделі (13.1). Представимо також функцію просторової кореляції поля завад у формі добутку двох множників:

$$\dot{K}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) = g_z(\omega) \cdot \dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p). \quad (13.21)$$

Тут $g_z(\omega)$ - максимум модулю просторової кореляційної функції завад по області прийому, $\dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p)$ - відповідно нормована функція кореляції завад у просторі. У випадку, коли до поля завад може бути застосована гіпотеза квазіоднорідності, функція $g_z(\omega)$ відповідає спектральній щільності поля завад в області прийому, а нормована функція $\dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p)$ є

коефіцієнтом просторової кореляції завад, який залежить лише від різниці своїх аргументів, тобто:

$$\dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) = \dot{R}_z(\omega; \mathbf{r}_m - \mathbf{r}_p), \quad \max \left\{ \left| \dot{R}_z(\omega; \mathbf{r}_m - \mathbf{r}_p) \right| \right\} \leq 1. \quad (13.22)$$

Підставляючи співвідношення (13.19) та (13.21) до рівняння (13.17), маємо:

$$\begin{aligned} g_w \cdot \dot{B}(\omega, \mathbf{r}_m) + g_z(\omega) \sum_{p=1}^M \dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot \dot{B}(\omega, \mathbf{r}_p) = \\ = \dot{S}_{ch}(\omega / \tau_0) \cdot \exp(-j\omega\tau_m(\mathbf{n}_s)). \end{aligned} \quad (13.23)$$

Будемо шукати рішення рівняння (13.23) у вигляді

$$\dot{B}(\omega, \mathbf{r}_m) = \frac{\dot{S}_{ch}(\omega / \tau_0)}{g_w + g_z(\omega)} \cdot \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot \exp(-j\omega\tau_m(\mathbf{n}_s)), \quad (13.24)$$

де $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$ – оптимальний амплітудно-фазовий розподіл (АФР) на просторовій апертурі. Після підстановки виразу (13.24) до рівняння (13.23) маємо щодо функції $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$ наступне рівняння:

$$\begin{aligned} \frac{1}{1 + \gamma_z(\omega)} \cdot \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) + \\ + \frac{\gamma_z(\omega)}{1 + \gamma_z(\omega)} \sum_{p=1}^M \dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot \exp(j\omega(\tau_m(\mathbf{n}_s) - \tau_p(\mathbf{n}_s))) \cdot \dot{H}(\omega, \mathbf{r}_p / \mathbf{n}_s) = 1, \end{aligned} \quad (13.25)$$

де $\gamma_z(\omega) = g_z(\omega) / g_w$ - відносний рівень завад, нормований на СЦП білого шуму. Отже, рішення рівняння (13.25) дає оптимальну АФР антени $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$, а вираз (13.24) – загальну спектральну вагову функцію оптимальної обробки.

Підставимо (13.24) до алгоритму обробки в частотній області (13.18). В результаті одержимо:

$$V_0(\mathbf{u} / \mathbf{n}_s, \tau_0) = \frac{A_0}{2\pi} \int_{-\infty}^{\infty} \dot{B}_T^*(\omega) \cdot \dot{U}_{AR}(\omega, \mathbf{n}_s) \cdot \exp\{-j\omega \cdot \tau_0\} d\omega, \quad (13.26)$$

де

$$\dot{U}_{AR}(\omega, \mathbf{n}_s) = \sum_{m=1}^M \dot{H}^*(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot \dot{U}(\omega, \mathbf{r}_m) \cdot \exp\{j\omega \cdot \tau_m(\mathbf{n}_s)\}, \quad (13.27)$$

$$\dot{B}_T(\omega) = \frac{\dot{S}_0(\omega)}{g_w + g_z(\omega)} . \quad (13.28)$$

З розгляду виразів (13.26) – (13.28) неважко бачити, що оптимальне оброблення сигналів на тлі корельованих завад може бути, як й вище, представлено у вигляді каскадного поєднання двох підсистем: у першій з них виконується накопичення сигналу по просторовій апертурі за алгоритмом (13.27), на другому етапі – накопичення енергії сигналу в межах його смуги за алгоритмом (13.26). Але на відміну від випадку, розглянутому у попередньому параграфі, в даному разі перша частина обробки (13.27) потребує використання оптимального АФР, який знаходиться в результаті рішення системи рівнянь (13.25). У свою чергу, оптимальний АФР $\dot{H}_0(\omega, \mathbf{r}_m / \mathbf{n}_s)$ залежить від кореляційних властивостей завад у просторі, які змінюються з частотою. Таким чином, оптимальний АФР також залежить від частоти. Тому алгоритми обробки вигляду (13.27) називають **частотно-залежною просторовою обробкою**. Заключний же етап оброблення за алгоритмом (13.26), який не залежить від просторових властивостей сигналу та завад, відповідає **часовій обробці**. Але, на відміну від попереднього параграфу, в даному випадку оптимальна вагова функція часової обробки (13.28) враховує як спектр сигналу, так й спектральні характеристики завад. Нагадаємо також, що розглянуті алгоритми оптимальної обробки сигналу в просторі та часі одержані в рамках прийнятого вище квазістаціонарного наближення для компактних апертур. Для більш загальних умов ці алгоритми можуть виявитися й не оптимальними.

На завершення даного параграфу розглянемо практично важливий випадок приймання вузько смугових сигналів. Як вже неодноразово зазначалося вище, саме такі сигнали й мають найширше застосування в радіосистемах різноманітного призначення. Почнемо з аналізу рівняння (13.25). Залежність рішення цього рівняння від частоти визначається, головним чином, експоненційним множником, оскільки просторова

кореляція завад змінюється у будь-якому разі не швидше. Проте, цією залежністю можна знехтувати, якщо в межах робочої смуги частот $(\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2)$ показник експоненти задовольняє посиленій нерівності

$$\frac{\Delta\omega}{2} \cdot \max_{m,p} |\tau_m - \tau_p| < 2\pi.$$

У свою чергу, $\max_{m,p} |\tau_m - \tau_p| \leq 2D_{\max}/c_0$, де, як і вище, $D_{\max}$ позначає максимальний геометричний розмір антени. З урахуванням цих співвідношень легко бачити, що показник експоненти у рівнянні (13.25) від частоти практично не залежить, якщо виконується просторова умова вузької смуги сигналу (13.6), яка вже згадувалась вище. Отже, за виконання вказаної умови залежність рішення рівняння (13.25), тобто, оптимальної АФР, від частоти в межах робочої смуги приймальної системи можна знехтувати. Таким чином,

$$\begin{aligned} \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) &\cong \dot{H}(\omega_0, \mathbf{r}_m / \mathbf{n}_s) = \dot{H}_0(\mathbf{r}_m / \mathbf{n}_s), \\ \omega &\in (\omega_0 - \Delta\omega/2, \omega_0 + \Delta\omega/2). \end{aligned} \quad (13.29)$$

Додатково припустимо, що в межах тієї ж смуги частот СЩП завад практично незмінна, тобто,

$$g_z(\omega) \cong g_z(\omega_0) = g_z. \quad (13.30)$$

За умов (13.29), (13.30) алгоритми оптимальної обробки (13.26), (13.27) в часовій області можуть бути представлені у наступному наближеному вигляді:

$$\dot{V}_0(\mathbf{u} / \tau_0, \mathbf{n}_s) = \frac{A_0}{g_w + g_z} \int_{\tau_0}^{T+\tau_0} u_{AR}(t, \mathbf{n}_s) \cdot \dot{s}_0(t - \tau_0) dt, \quad (13.31)$$

$$\text{де} \quad u_{AR}(t, \mathbf{n}_s) = \sum_{m=1}^M H_0^*(\mathbf{r}_m / \mathbf{n}_s) \cdot u(t - \tau_m(\mathbf{n}_s), \mathbf{r}_m). \quad (13.32)$$

З порівняння цих співвідношень з алгоритмами оптимальної обробки на тлі білого шуму (13.3), (13.4) легко бачити, що при застосуванні прийнятих

наближень врахування корельованості завад зводиться, по-перше, до зміни нормування в алгоритмі часової обробки (13.31), по-друге, в алгоритмі просторової обробки (13.32) додаються вагові коефіцієнти, які знаходяться в результаті рішення рівняння (13.25) для носійної частоти ω_0 . В окремому випадку відсутності зовнішніх завад ($g_z = 0$) із рівняння (13.25) витікає, що АФР просторової обробки є рівномірним, тобто,

$$H_0(\mathbf{r}_m / \mathbf{n}_s) \equiv 1.$$

Для цього значення АФР алгоритми (13.31), (13.32) співпадають відповідно з алгоритмами (13.3), (13.4).

13.3. Приймання шумових сигналів.

Сигнали даного типу, як вже було сказано вище, є випадковими функціями часу. Але їх властивості у просторі не випадкові, оскільки визначаються переважно факторами середовища передачі, а не виглядом сигналу як функції часу. Наприклад, якщо джерело сигналу є близьким до точкового, тобто, має малі хвильові розміри, то поле сигналу на приймальній апертурі в зоні Фраунгофера є плоскою хвилею або сукупністю плоских хвиль. При цьому просторові властивості таких хвиль при однаковому розташуванні джерела та співпадаючих умовах передачі через середовище будуть однаковими як для наперед відомого, так й для випадкового у часі сигналу. Відповідно, слід очікувати, що просторова частина оброблення шумових сигналів при певних умовах залишиться подібною до розглянутої вище.

Будемо вважати, що й для даної задачі виконуються умови квазістаціонарного наближення для компактних апертур, які були сформульовані вище. Додатково також припустимо, що випадковий сигнал є стаціонарним шумовим процесом, смуга Δf_s та тривалість T якого задовольняють посиленій нерівності

$$\Delta f_s T \gg 1. \quad (13.33)$$

За цих умов функціонал обробки випадкових сигналів (11.43) можна наближено представити у наступній спектральній формі:

$$F_0(\mathbf{u}) \cong \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{i=1}^M \sum_{k=1}^M U^*(\omega, \mathbf{r}_i) \cdot B_0(\omega; \mathbf{r}_i, \mathbf{r}_k) \cdot U(\omega, \mathbf{r}_k) d\omega \quad (13.34)$$

де $U(\omega, \mathbf{r})$ - спектр Фур'є поля $u(t, \mathbf{r})$ на інтервалі $t \in (0, T)$, $B_0(\omega; \mathbf{r}_i, \mathbf{r}_k)$ - перетворення Фур'є по $\tau = t - t'$ ядра обробки $b_0(t, t'; \mathbf{r}_i, \mathbf{r}_k) = b_0(t - t'; \mathbf{r}_i, \mathbf{r}_k)$.

Обмежимося поки що випадком приймання сигналу на тлі білого шуму. У цьому разі, як було показано у розділі 11, ядро оптимальної обробки (11.43) є рішенням системи рівнянь (11.44). У спектральній області цим рівнянням відповідає система алгебраїчних рівнянь

$$g_w \cdot B_0(\omega; \mathbf{r}_i, \mathbf{r}_k) + \sum_{p=1}^M K_s(\omega; \mathbf{r}_i, \mathbf{r}_p) \cdot B_0(\omega; \mathbf{r}_p, \mathbf{r}_k) = \frac{1}{g_w} K_s(\omega; \mathbf{r}_i, \mathbf{r}_k), \quad (13.35)$$

де $K_s(\omega; \mathbf{r}_i, \mathbf{r}_k)$ - функція кореляції поля сигналу в просторі, структуру якої необхідно задати. Припустимо, що джерело випадкового сигналу знаходиться у дальній зоні приймальної системи. Тоді за умов компактної апертури для поля сигналу може бути прийнята та ж модель плоскої хвилі (13.1), що й вище. Єдина відмінність полягає в тому, що залежність сигналу, як функції часу є випадковою. Отже, для просторово-часової функції кореляції поля випадкового сигналу на приймальній апертурі АР можна записати³:

$$\begin{aligned} K_s(t_1, t_2; \mathbf{r}_i, \mathbf{r}_k) &= \langle s(t_1, \mathbf{r}_i) \cdot s(t_2, \mathbf{r}_k) \rangle = \\ &= \langle s(t_1 - \tau_0 + \tau_i(\mathbf{n}_s)) \cdot s(t_2 - \tau_0 + \tau_k(\mathbf{n}_s)) \rangle. \end{aligned}$$

З урахуванням припущення про квазістаціонарність сигналу цей вираз приймає вигляд:

$$K_s(t_1, t_2; \mathbf{r}_i, \mathbf{r}_k) = K_s(t_1 - t_2 + (\mathbf{n}_s, \mathbf{r}_i - \mathbf{r}_k) / c_0). \quad (13.36)$$

З цього співвідношення витікає, що поле сигналу на інтервалі спостереження є стаціонарним, а в області приймання однорідним, оскільки

³ Амплітуду сигналу внесено безпосередньо до поля сигналу.

кореляційна функція поля випадкового сигналу (13.36) залежить лише від різниці як часових, так й просторових координат.

Після виконання перетворення Фур'є від виразу (13.36) по часовому аргументу $\tau = t_1 - t_2$ знаходимо функцію кореляції поля сигналу в просторі:

$$K_s(\omega; \mathbf{r}_i, \mathbf{r}_k) = K_s(\omega; \mathbf{r}_i - \mathbf{r}_k) = g_s(\omega) \cdot \exp(j\omega \cdot (\mathbf{n}_s, \mathbf{r}_i)/c_0) \cdot \exp(-j\omega \cdot (\mathbf{n}_s, \mathbf{r}_k)/c_0), \quad (13.37)$$

$$\text{де} \quad g_s(\omega) = \lim_{T \rightarrow \infty} \left\langle \frac{1}{T} \cdot |\dot{S}_T(\omega, \mathbf{r}_0)|^2 \right\rangle \quad (13.38)$$

- спектральна щільність потужності сигналу у фазовому центрі АР, $\dot{S}_T(\omega, \mathbf{r}_0)$ - спектр Фур'є поля сигналу $s(t, \mathbf{r}_0)$ на часовому інтервалі тривалістю T .

У структурі виразу (13.37) неважко помітити розділення (факторизацію) просторової кореляції поля сигналу на добуток двох множників, кожний з яких залежить лише від одного з аргументів $\mathbf{r}_i$ та $\mathbf{r}_k$. Така факторизація просторової кореляції поля сигналу свідчить про його повну когерентність на приймальній апертурі. Отже, не дивлячись на випадковий характер поля шумового сигналу у часі, його властивості у просторі є точно відомими, якщо, звичайно, відомо напрямок приходу сигналу.

Підставимо вираз (13.37) до рівнянь (13.35). В результаті маємо:

$$\begin{aligned} & g_w \cdot B_0(\omega; \mathbf{r}_i, \mathbf{r}_k) + \\ & + g_s(\omega) \cdot \exp(j\omega \cdot (\mathbf{n}_s, \mathbf{r}_i)/c_0) \sum_{p=1}^M B_0(\omega; \mathbf{r}_p, \mathbf{r}_k) \cdot \exp(-j\omega \cdot (\mathbf{n}_s, \mathbf{r}_p)/c_0) = \\ & = \frac{g_s(\omega)}{g_w} \cdot \exp(j\omega \cdot (\mathbf{n}_s, \mathbf{r}_i)/c_0) \cdot \exp(-j\omega \cdot (\mathbf{n}_s, \mathbf{r}_k)/c_0). \end{aligned}$$

Прямою підстановкою неважко впевнитися, що рішення цього рівняння має вигляд

$$\begin{aligned} B_0(\omega; \mathbf{r}_i, \mathbf{r}_k) &= \frac{g_s(\omega)}{g_w \cdot (g_w + M \cdot g_s(\omega))} \times \\ &\times \exp(j\omega \cdot (\mathbf{n}_s, \mathbf{r}_i)/c_0) \cdot \exp(-j\omega \cdot (\mathbf{n}_s, \mathbf{r}_k)/c_0). \end{aligned} \quad (13.38)$$

Після підстановки виразу (13.38) до функціоналу (13.34), знаходимо алгоритм оптимального приймання випадкового сигналу на тлі білого шуму:

$$F_0(\mathbf{u}) \cong \frac{1}{2\pi} \int_{-\infty}^{\infty} |K_0(\omega)|^2 \cdot |U_{AR}(\omega, \mathbf{n}_s)|^2 d\omega, \quad (13.39)$$

$$\text{де} \quad U_{AR}(\omega, \mathbf{n}_s) = \sum_{k=1}^M U(\omega, \mathbf{r}_k) \cdot \exp(j\omega \cdot \tau_k(\mathbf{n}_s)/c_0) \quad (13.40)$$

- спектр сигналу на виході формувача направленості (13.4), який налаштовано на прийом сигналу з напрямку $\mathbf{n}_s$,

$$|K_0(\omega)|^2 = \frac{g_s(\omega)}{g_w \cdot (g_w + M \cdot g_s(\omega))} \quad (13.41)$$

- спектральна вагова функція. Таким чином, в разі приймання випадкового сигналу також маємо розділення його обробки на дві частини: на першому етапі виконується просторова обробка (13.40) фактично за тим же самим алгоритмом, що й у випадку приймання детермінованих або квазідетермінованих сигналів. Етап часової обробки зводиться до накопичення енергії сигналу в робочій смузі з ваговою функцією (13.41).

Скориставшись теоремою Парсеваля [3] алгоритмам оброблення випадкового сигналу в спектральній області (13.39), (13.40) можна поставити у відповідність еквівалентні алгоритми оброблення у часовій області, а саме:

$$F_0(\mathbf{u}/\mathbf{n}_s) = \int_0^T \left| \int_{-\infty}^t b_{AR}(t-t') \cdot u_{AR}(t', \mathbf{n}_s) dt' \right|^2 dt, \quad (13.42)$$

де $b_{AR}(t)$ - імпульсний відгук фільтра, якому відповідає частотна характеристика (13.41), $u_{AR}(t, \mathbf{n}_s)$ - вихідний сигнал ФН за алгоритмом (13.4). Зауважимо, що функціонал (13.42) за своєю структурою відповідає відомій традиційній схемі оброблення сигналів "фільтр – квадратичний детектор – інтегратор" [20], яку прийнято називати типовим трактом виявлення сигналів (ТТВ).

На завершення параграфу зазначимо, що структура оброблення випадкового сигналу для більш загального випадку прийому на тлі

корельованих завад виглядає схожим чином [10]. Але ця структура має дві принципові відмінності: етап просторової обробки виконується за алгоритмом частотно-залежної просторової обробки (13.27), а у часовій обробці (13.39) змінюється спектральна вагова функція, яка в цьому випадку має вигляд

$$|K_0(\omega)|^2 = \frac{g_s(\omega)}{(g_w + g_z(\omega)) \cdot \left(g_w + g_z(\omega) + g_s(\omega) \cdot \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right)}, \quad (13.43)$$

де, як й у попередньому параграфі, $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$ є рішенням рівняння (13.25).

13.4. Приймання сигналів на тлі локальних завад.

Як вже було сказано вище, лінійна частина оброблення сигналів є важливою складовою приймальної системи, яка у найбільшій мірі «пристосована» до врахування фізичних особливостей полів сигналів та завад, що, у свою чергу, сприяє підвищенню ефективності прийому сигналів у цілому. Також було показано, що лінійна частина оброблення, в незалежності від типу сигналів (детерміновані, квазидетерміновані, випадкові), при певних умовах може бути представлена у вигляді каскадного поєднання двох складових, а саме, підсистеми просторової та підсистеми часової обробки сигналів. У свою чергу, перша з цих підсистем, тобто, просторова обробка, має, по суті, уніфіковану структуру, яка практично не залежить від типу сигналу. Також зазначимо, що саме просторова частина оброблення сигналів в сучасних умовах має найбільший потенціал щодо можливості підвищення ефективності приймальної системи в цілому.

Одним з важливих прикладів оптимізації просторової обробки, який вже знайшов практичне застосування, є приймання сигналів на тлі так званих локальних завад. Це особливий вид завад, які породжуються зосередженими (або локалізованими) у просторі джерелами. За своїми властивостями такі завади є, по суті, сторонніми сигналами, які відрізняються від корисного

сигналу розташуванням джерел у просторі та своїми спектрально-часовими характеристиками. Вказана обставина веде до специфічних алгоритмів просторової обробки поля сигналу, які прийнято називати компенсацією (або придушенням) локальних завад.

Припустимо, що локальну заваду створює джерело малих хвильових розмірів, яке знаходиться у дальній зоні приймальної системи. Тоді поле завади у просторі буде мати вигляд плоскої хвилі, подібної до поля сигналу (13.1), але з іншими параметрами. Отже, по аналогії з (13.37) для функції просторової кореляції завад можна записати:

$$\dot{K}_z(\omega; \mathbf{r}_m - \mathbf{r}_p) = g_z(\omega) \cdot \dot{R}_z(\omega; \mathbf{r}_m - \mathbf{r}_p),$$

де, як й вище, $g_z(\omega)$ - СЦП завади на приймальній апертурі АР, а

$$\dot{R}_z(\omega; \mathbf{r}_m - \mathbf{r}_p) = \exp(j\omega \cdot \tau_m(\mathbf{n}_z)) \cdot \exp(-j\omega \cdot \tau_p(\mathbf{n}_z)) \quad (13.44)$$

– коефіцієнт просторової кореляції поля завад, $\tau_m(\mathbf{n}_z) = (\mathbf{n}_z, \mathbf{r}_m) / c_0$ - затримка завади на m -ому датчику ($m = 1, \dots, M$), $\mathbf{n}_z$ – одиничний вектор напрямку приходу її плоскої хвилі.

Вище було показано, що при виконанні умов квазістаціонарного наближення для компактних апертур алгоритм оптимальної просторової обробки прийнятого поля, в не залежності від вигляду сигналу в часі, визначається функціоналом (13.27), де оптимальний амплітудно-фазовий розподіл $\dot{H}(\omega, \mathbf{r})$ на апертурі АР є рішенням рівняння (13.25). Отже, рішення цього рівняння у випадку, коли коефіцієнт просторової кореляції має вигляд (13.44), дає відповідний алгоритм приймання сигналу на тлі локальної завади. При розгляді цього рішення попередньо перейдемо до векторно-матричних позначень. Почнемо з того, що представимо у цих позначеннях алгоритм просторової обробки (13.27):

$$\begin{aligned}
\dot{U}_{AR}(\omega, \mathbf{n}_s) &= \sum_{m=1}^M \dot{H}_m^*(\omega, \mathbf{n}_s) \dot{U}(\omega, \mathbf{r}_m) \exp\{j\omega \cdot \tau_m(\mathbf{n}_s)\} = \\
&= \sum_{m=1}^M \dot{L}_m^*(\omega, \mathbf{n}_s) \dot{U}(\omega, \mathbf{r}_m) = \dot{\mathbf{L}}^H(\omega, \mathbf{n}_s) \cdot \dot{\mathbf{U}}(\omega).
\end{aligned} \tag{13.45}$$

Тут $\dot{\mathbf{U}}(\omega)$ - вектор (тобто, матриця-стовпчик), який складається зі значень спектру поля $\dot{U}(\omega, \mathbf{r}_m)$ на частоті ω по всім датчикам АР, $\dot{\mathbf{L}}(\omega, \mathbf{n}_s)$ - вектор вагових коефіцієнтів просторової обробки, m – та компонента якого має вигляд

$$L_m(\omega, \mathbf{n}_s) = \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot \exp\{-j\omega \cdot \tau_m(\mathbf{n}_s)\}, \tag{13.46}$$

$(\cdot)^H$ - оператор ермітового спряження матриць, тобто, їх транспонування, яке поєднано із комплексним спряженням. Використовуючи ці позначення та співвідношення (13.25), для вектору вагових коефіцієнтів (13.46) можна записати наступне рівняння:

$$(\mathbf{E}_M + \gamma_z(\omega) \cdot \dot{\mathbf{R}}_z(\omega)) \cdot \mathbf{L}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \mathbf{v}(\omega, \mathbf{n}_s), \tag{13.47}$$

де $\mathbf{E}_M$ - одинична матриця порядку M , $\dot{\mathbf{R}}_z(\omega)$ - матриця коефіцієнтів кореляції завад $\dot{R}_z(\omega; \mathbf{r}_m, \mathbf{r}_p)$, $\dot{\mathbf{v}}(\omega, \mathbf{n}_s)$ - так званий розподільчий вектор АР (*array manifold vector* [25]), компоненти якого у випадку плоскої хвилі, що приходить з напрямку $\mathbf{n}$, мають вигляд:

$$\dot{v}_m(\omega, \mathbf{n}) = \exp(-j\omega \tau_m(\mathbf{n})). \tag{13.48}$$

В тотожності рівнянь (13.47) та (13.25) неважко впевнитися, якщо записати розгорнуту версію співвідношення (13.47) та враховувати заміну (13.46). Отже, на основі рівняння (13.47) оптимальна вагова функція просторової обробки може бути записана у вигляді матричного співвідношення

$$\mathbf{L}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot (\mathbf{E}_M + \gamma_z(\omega) \cdot \dot{\mathbf{R}}_z(\omega))^{-1} \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_s). \tag{13.49}$$

Скориставшись прийнятими позначеннями, підставимо до цього виразу матрицю коефіцієнтів просторової кореляції завад (13.44):

$$\mathbf{L}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\mathbf{E}_M + \gamma_z(\omega) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_z) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_z) \right)^{-1} \cdot \mathbf{v}(\omega, \mathbf{n}_s). \quad (13.50)$$

Далі застосуємо так звану *лему про обернення матриць* [3], яка у найпростішому випадку має вигляд тотожності

$$\left(\mathbf{E} + \mathbf{x} \cdot \mathbf{y}^H \right)^{-1} = \mathbf{E} - \frac{\mathbf{x} \cdot \mathbf{y}^H}{1 + \mathbf{y}^H \cdot \mathbf{x}}, \quad (13.51)$$

де $\mathbf{x}, \mathbf{y}$ - довільні вектори (матриці-стовпчики) однакового порядку. Скориставшись цією тотожністю, співвідношення (13.50) неважко привести до наступного вигляду:

$$\mathbf{L}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\mathbf{v}(\omega, \mathbf{n}_s) - \beta(\omega) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_z) \right), \quad (13.52)$$

$$\text{де} \quad \beta(\omega) = \frac{\gamma_z(\omega) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_z) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_s)}{1 + \gamma_z(\omega) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_z) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_z)}$$

Після підстановки рішення (13.52) до алгоритму просторової обробки (13.45) маємо:

$$\dot{U}_{AR}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\dot{\mathbf{v}}^H(\omega, \mathbf{n}_s) \cdot \dot{\mathbf{U}}(\omega) - \beta(\omega) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_z) \cdot \dot{\mathbf{U}}(\omega) \right),$$

або у розгорнутому вигляді:

$$\dot{U}_{AR}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\dot{Y}(\omega, \mathbf{n}_s) - \beta(\omega) \cdot \dot{Y}(\omega, \mathbf{n}_z) \right), \quad (13.53)$$

де

$$\dot{Y}(\omega, \mathbf{n}) = \sum_{m=1}^M \dot{U}(\omega, \mathbf{r}_m) \exp\{j\omega \cdot \tau_m(\mathbf{n})\} \quad (13.54)$$

позначає вихід первинного каналу просторової обробки, який орієнтовано у напрямку $\mathbf{n}$,

$$\beta(\omega) = \frac{\gamma_z(\omega) \cdot \sum_{m=1}^M \exp\{j\omega \cdot (\tau_m(\mathbf{n}_z) - \tau_m(\mathbf{n}_s))\}}{1 + \gamma_z(\omega) \cdot M} \quad (13.55)$$

- розгорнутий запис для вагового коефіцієнту. Вирази (13.53) - (13.55) мають дуже прозорий та фізично наочний вигляд. Справді, у відповідності з ними для забезпечення компенсації локальної завади, потрібно, перш за все, організувати два первинних канали просторової обробки, які формуються за

алгоритмом (13.54) та відрізняються лише напрямом своєї орієнтації. Перший з цих каналів, який будемо називати основним (сигнальним), забезпечує виділення корисного сигналу на тлі білого шуму за напрямком $\mathbf{n}_s$. Другий канал, який будемо називати опорним, забезпечує виділення локальної завади за напрямком $\mathbf{n}_z$. Відповідно до виразу (13.53) компенсація локальної завади відбувається шляхом віднімання виходів основного та опорного каналів з ваговим коефіцієнтом (13.55). Вказана процедура виконується для кожної частоти робочої смуги приймальної системи окремо, що є наслідком прийнятого вище квазістаціонарного наближення.

Ще більш наочний вигляд цей алгоритм приймає у випадку, коли локальна завада значно переважає рівень білого шуму, тобто, $\gamma_z \gg 1$. За цієї умови вираз для вагових коефіцієнтів (13.55) наближено приймає вигляд:

$$\beta(\omega) \cong \frac{1}{M} \sum_{m=1}^M \exp\{j\omega \cdot (\tau_m(\mathbf{n}_z) - \tau_m(\mathbf{n}_s))\}. \quad (13.56)$$

Як неважко бачити з цього співвідношення, ваговий коефіцієнт компенсатора приблизно визначається рівнем бокового пелюстка нормованої діаграми направленості (НДН) сигнального каналу АР за напрямком приходу локальної завади, що цілком відповідає й апріорним фізичним уявленням.

Розглянуту ідею компенсації однієї локальної завади неважко розповсюдити на випадок кількох таких завад. Дійсно, припустимо, що на приймальну систему діють $N_{\text{лок.}}$ локальних завад, які поміж собою статистично незалежні. Тоді для функції просторової кореляції поля завад з використанням прийнятих вище позначень можна записати:

$$\dot{K}_z(\omega; \mathbf{r}_m, \mathbf{r}_p) = \sum_{k=1}^{N_{\text{лок.}}} g_k(\omega) \cdot \dot{v}_m(\omega, \mathbf{n}_k) \cdot \dot{v}_p^H(\omega, \mathbf{n}_k),$$

де $g_k(\omega)$ - СЩП k -ої компоненти завади, $\mathbf{n}_k$ - одиничний вектор напрямку приходу відповідної плоскої хвилі, $\dot{v}_m(\omega, \mathbf{n})$ - компоненти розподільчого вектору (13.48). Даній кореляційній функції відповідає матриця коефіцієнтів просторової кореляції завад

$$\dot{\mathbf{R}}_z(\omega) = \sum_{k=1}^{N_{\text{лок.}}} v_k(\omega) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_k) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_k), \quad (13.57)$$

де

$$v_k(\omega) = \frac{g_k(\omega)}{g_1(\omega) + g_2(\omega) + \dots + g_{N_{\text{лок.}}}(\omega)} = \frac{g_k(\omega)}{g_z(\omega)} \quad (13.58)$$

- відносний рівень k -ої компоненти локальних завад, $g_z(\omega)$ - СЦП сумарного поля локальних завад. Після підстановки виразу (13.57) до рівняння (13.47) маємо:

$$\left(\mathbf{E}_M + \gamma_z(\omega) \cdot \sum_{k=1}^{N_{\text{лок.}}} v_k(\omega) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_k) \cdot \dot{\mathbf{v}}^H(\omega, \mathbf{n}_k) \right) \cdot \mathbf{L}(\omega, \mathbf{n}_s) = \quad (13.59)$$

$$= (1 + \gamma_z(\omega)) \cdot \mathbf{v}(\omega, \mathbf{n}_s).$$

По аналогії з розглянутим вище випадком придушення однієї завади будемо шукати рішення рівняння (13.59) у вигляді

$$\dot{\mathbf{L}}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\dot{\mathbf{v}}(\omega, \mathbf{n}_s) - \sum_{k=1}^{N_{\text{лок.}}} \beta_k(\omega) \cdot \dot{\mathbf{v}}(\omega, \mathbf{n}_k) \right). \quad (13.60)$$

Після підстановки (13.60) до співвідношення (13.59) неважко бачити, що у випадку інтенсивних завад ($\gamma_z \gg 1$) рівняння (13.59) задовольняється, якщо вагові коефіцієнти компенсатора $\beta_k(\omega)$, у свою чергу, задовольняють рівнянню

$$\sum_{p=1}^{N_{\text{лок.}}} \beta_p(\omega) \cdot \Gamma_{kp}(\omega) = \Gamma_{k0}(\omega), \quad (13.61)$$

$$\text{де} \quad \Gamma_{kp}(\omega) = \dot{\mathbf{v}}^H(\mathbf{n}_k, \omega) \cdot \dot{\mathbf{v}}(\mathbf{n}_p, \omega) \quad (13.62)$$

- елементи матриці кореляції первинних завадових каналів,

$$\Gamma_{k0}(\omega) = \dot{\mathbf{v}}^H(\mathbf{n}_k, \omega) \cdot \dot{\mathbf{v}}(\mathbf{n}_s, \omega) \quad (13.63)$$

- кореляція первинних завадових та сигнального каналів.

Таким чином, після підстановки рішення (13.60) до виразу (13.45) знаходимо алгоритм придушення кількох локальних завад:

$$\dot{U}_{AR}(\omega, \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(\dot{Y}(\omega, \mathbf{n}_s) - \sum_{k=1}^{N_{\text{лок}}} \beta_k(\omega) \cdot \dot{Y}(\omega, \mathbf{n}_k) \right), \quad (13.64)$$

де первинні канали просторової обробки формуються за тим же алгоритмом (13.54), що й у випадку однієї локальної завади. Отже, випадок кількох завад відрізняється, головним чином, ваговими коефіцієнтами компенсатора, які знаходяться з рівняння (13.61). Зауважимо, що рівняння (13.61) має рішення лише за умови, що кількість локальних завад не перевищує числа датчиків АР, тобто, $N_{\text{лок}} \leq M$. Отже, використання даного методу придушення завад можливо лише за цієї умови.

Також зазначимо, що до схожих результатів можна прийти в результаті вирішення оптимізаційної задачі, яка має вигляд:

$$\left\langle \left| \dot{Y}(\omega, \mathbf{n}_s) - \sum_{k=1}^{N_{\text{лок}}} \beta_k(\omega) \cdot \dot{Y}(\omega, \mathbf{n}_k) \right|^2 / \theta = 0 \right\rangle \rightarrow \min_{\beta_1, \dots, \beta_{N_{\text{лок}}}} \quad (13.65)$$

Дійсно, вирішення цієї оптимізаційної задачі за умови $\gamma_z \gg 1$ приводить до системи рівнянь, яка співпадає з (13.61). Тому задача (13.65) складає одну з основ теорії адаптивної обробки, де її прийнято називати критерієм мінімуму потужності завад по виходу системи приймання.

Контрольні запитання та завдання до розділу 13.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Як у загальному випадку виглядає лінійна частина оброблення сигналів різних типів (детермінованих, квазидетермінованих, випадкових)? Яку роль відіграє така обробка в алгоритмах приймання сигналів?
2. За яких умов може бути застосована модель поля сигналу (13.1)? Які її основні параметри?
3. За яких умов відбувається розділення лінійного оброблення сигналів на просторову та часову частини? Які функції ці частини виконують? Пояснити схему рис.12.1.
4. За яких умов можлива заміна алгоритму просторової обробки (13.4) алгоритмом (13.5)? Який з них ефективніший, а який простіший для практичної реалізації?
5. Пояснити, як з алгоритму (13.5) можна одержати алгоритм (13.7).
6. Чому фільтр (13.8) називають узгодженим? Що таке узгоджене оброблення сигналів у сучасному розумінні?
7. Самостійно повторити викладки, які приводять до виразу (13.10).
8. Які можливості відкриває статистична теорія приймання сигналів у порівнянні з традиційними підходами? Наведіть приклади.
9. Що таке квазістаціонарне наближення для компактних апертур? З яких умов воно складається? В чому полягає необхідність кожної з цих умов?
10. Самостійно одержати вирази (13.17) – (13.23).
11. Самостійно одержати вирази (13.25) – (13.28).
12. Чим відрізняються оптимальні алгоритми оброблення сигналів на тлі корельованих завад (13.26), (13.27) у порівнянні з алгоритмами (13.3), (13.4) для білого шуму.
13. Пояснити, з яких міркувань витікає просторова умова вузької смуги сигналу (13.6).
14. Самостійно одержати оптимальні алгоритми оброблення вузько смугових сигналів (13.31), (13.32).
15. Самостійно здійснити перехід від алгоритму (11.43) до алгоритму (13.34) та від рівняння (11.44) до рівняння (13.35).
16. Що таке просторова когерентність поля сигналу? Чим вона визначається та від чого залежить?

17. Що є спільного в методах приймання сигналів різних типів? Як таку спільність пояснити фізично?
18. Самостійно одержати вирази (13.38) – (13.40).
19. Якій традиційній схемі прийому відповідає алгоритм (13.42)? За яких умов його можна вважати оптимальним?
20. Що таке локальні завади? Наведіть фізичні приклади таких завад.
21. Самостійно одержати вирази (13.52) – (13.55). Пояснити фізичну суть цих алгоритмів.
22. Самостійно перевірити рішення (13.60) та одержати рівняння (13.61).
23. Записати у розгорнутому вигляді елементи матриць (13.62), (13.63).
24. Самостійно одержати рішення оптимізаційної задачі (13.65) для випадку інтенсивних локальних завад, тобто, $\gamma_z \gg 1$.

Розділ 14. Підсилення просторової обробки.

У попередньому розділі була показана можливість та доцільність розділення загального функціоналу оброблення сигналів на часову та просторову частини обробки. З цих двох частин визначальною для ефективності приймальної системи у цілому є без сумніву просторова частина обробки. Тому даний розділ присвячено розгляду одного з найважливіших параметрів завадостійкості приймальних систем, який називають підсиленням просторової обробки сигналів.

14.1. Загальні відомості.

Будемо вважати, як й вище, що виконуються умови квазістаціонарного наближення для компактних апертур. При цих умовах доцільно зосередитись, головним чином, на варіантах оброблення сигналів, орієнтованих на реалізацію обробки у спектральній області. Отже, вихідними даними для просторової обробки будемо вважати комплексні спектральні амплітуди $\dot{U}_T(\omega, \mathbf{r}_m)$, $m = 1, \dots, M$, які одержані в результаті перетворення Фур'є сигналів по виходу датчиків АР $u(t, \mathbf{r}_m)$ в межах часового інтервалу $t \in (0, T)$. Ґрунтуючись на матеріалах розділу 13 запишемо загальний алгоритм просторового оброблення сигналів у спектральній області для деякої дискретної АР :

$$\dot{U}_{AR}(\omega, T) = \sum_{m=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot \dot{U}_T(\omega, \mathbf{r}_m) . \quad (14.1)$$

Тут $\dot{L}(\omega, \mathbf{r}_m)$ - деяка комплексна вагова (апертурна) функція просторової обробки, а кругова частота ω «пробігає» набір значень в межах робочої смуги приймальної системи. Коефіцієнт підсилення (або просто підсилення) просторової обробки $\chi(\omega)$ звичайно визначають [25] наступним чином:

$$\chi(\omega) = \frac{\mu_{out}^{AR}(\omega)}{\mu_{in}^{AR}(\omega)}, \quad (14.2)$$

де

$$\mu_{out}^{AR}(\omega) = \frac{g_{out}^{AR}(\omega/\theta=1) - g_{out}^{AR}(\omega/\theta=0)}{g_{out}^{AR}(\omega/\theta=0)}; \quad (14.3)$$

- відношення сигнал/завада по виходу АР на частоті ω ,

$$\mu_{in}^{AR}(\omega) = \frac{g_{in}^{AR}(\omega/\theta=1) - g_{in}^{AR}(\omega/\theta=0)}{g_{in}^{AR}(\omega/\theta=0)}, \quad (14.4)$$

- відношення сигнал/завада по входу АР. У цих співвідношеннях $g_{out}^{AR}(\omega/\theta)$, $g_{in}^{AR}(\omega/\theta)$ - спектральні щільності потужності суміші сигналів, шумів та завад відповідно на виході та на вході АР. При цьому входом АР вважають сигнал, який формується по виходу одиночного ненаправленого (ізоτροпного) датчика, розташованого у деякому умовному центрі антени. Тут доречно зауважити, що термін підсилення широко вживається й в теорії антен. Але не зважаючи на свою схожість підсилення просторової обробки та підсилення антен, взагалі кажучи, не одне й теж. В даному посібнику ми дотримуємося визначень, які прийняті у літературі з питань оброблення сигналів [25]. В теорії антен [21] підсилення визначається схожим чином, але при цьому можуть враховуватись деякі технічні аспекти побудови антен, наприклад, параметри перетворення датчиків сигналу. Більш близький за змістом до визначення (14.2) є так званий коефіцієнт направленої дії (КНД) антени. Але в літературних джерелах [21, 25] визначення КНД (в англійському варіанті *directivity*) звичайно «прив'язано» до певної моделі поля сигналу та поля завад. Зокрема, найбільш поширеною з таких моделей є ситуація прийому плоскої хвилі сигналу на тлі ізоτροпного білого шуму. Зауважимо також, що опорним елементом у теорії антен може бути не лише ізоτροпний датчик. Наприклад, в теорії радіотехнічних систем опорним часто вважають [21] напівхвильовий диполь. Отже, у даному посібнику ми використовуємо термін підсилення лише по відношенню до просторової обробки, використовуючи іноді як синоніми до нього також терміни підсилення антени або підсилення антенної решітки. Але не слід забувати, що повна еквівалентність поміж цими поняттями не завжди має місце.

Тепер запишемо характеристики сигналів та завад, які необхідні для обчислення параметрів (14.3), (14.4). Для обчислення спектру потужності $g_{out}^{AR}(\omega/\theta)$ по виходу просторової обробки (14.1) скористаємося граничним співвідношенням (4.14), а саме:

$$g_{out}^{AR}(\omega/\theta) = \lim_{T \rightarrow \infty} \frac{1}{T} \left\langle \left| \dot{U}_{AR}(\omega, T) \right|^2 / \theta \right\rangle.$$

Якщо добуток загальної робочої смуги прийому ΔF на тривалість сигналу T задовольняє посиленій нерівності $\Delta F T \gg 1$, то при практичних розрахунках граничний перехід можна, як правило, й не виконувати. Тоді беручи до уваги вираз (14.1), наближено маємо:

$$g_{out}^{AR}(\omega/\theta) \cong \frac{1}{T} \sum_{m=1}^M \sum_{p=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot \left\langle \dot{U}_T(\omega, \mathbf{r}_m) \cdot \dot{U}_T^*(\omega, \mathbf{r}_p) / \theta \right\rangle \cdot \dot{L}(\omega, \mathbf{r}_p) \quad (14.5)$$

Відповідно до робочих моделей, прийнятих у попередніх розділах, для кореляції спектральних амплітуд сигналів по виходу датчиків АР можна записати:

$$\begin{aligned} \frac{1}{T} \left\langle \dot{U}_T(\omega, \mathbf{r}_m) \cdot \dot{U}_T^*(\omega, \mathbf{r}_p) / \theta \right\rangle \cong & g_w \delta_{mp} + g_z(\omega) \cdot R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) + \\ & + \theta \cdot g_s(\omega) \cdot v_m(\omega, \mathbf{n}_s) \cdot v_p^*(\omega, \mathbf{n}_s), \end{aligned} \quad (14.6)$$

де, як і раніше, g_w , $g_z(\omega)$ - спектральні щільності потужності (СЩП) власних шумів приймальної системи та зовнішніх завад відповідно, $R_z(\omega; \mathbf{r}_m, \mathbf{r}_p)$ - нормована кореляція зовнішніх завад у просторі, $g_s(\omega)$ - спектр потужності сигналу в умовному центрі АР, $v_m(\omega, \mathbf{n}_s)$ - компоненти розподільчого вектора АР, які для задач приймання сигналу у вигляді плоскої хвилі визначаються співвідношенням (13.48), тобто:

$$\begin{aligned} \dot{v}_m(\omega, \mathbf{n}) = \exp(-j\omega \cdot \tau_m(\mathbf{n})) = \exp(-j\omega \cdot (\mathbf{n}, \mathbf{r}_m) / c_0) \\ m = 1, \dots, M. \end{aligned} \quad (14.7)$$

Після підстановки виразу (14.6) у співвідношення (14.5) одержуємо СЦП суміші сигналу, шумів та завад по виходу АР:

$$g_{out}^{AR}(\omega/\theta) \cong g_w \sum_{m=1}^M |\dot{L}(\omega, \mathbf{r}_m)|^2 + \quad (14.8)$$

$$+ g_z(\omega) \cdot \sum_{m=1}^M \sum_{p=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot \dot{L}(\omega, \mathbf{r}_p) + \theta \cdot g_s(\omega) \cdot \left| \sum_{m=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot v_m(\omega, \mathbf{n}_s) \right|^2$$

При цих же припущеннях СЦП сигналів та завад на вході АР (тобто, по виходу опорного ненаправленого датчика) має, вочевидь, вигляд:

$$g_{in}^{AR}(\omega) = g_w + g_z(\omega) + \theta \cdot g_s(\omega) \quad (14.9)$$

Підстановка співвідношень (14.8), (14.9) у вирази (14.3), (14.4) а потім у визначення (14.2) приводить до наступної загальної формули для обчислення підсилення довільної просторової обробки (14.1):

$$\chi(\omega) = \frac{(1 + \gamma_z(\omega)) \cdot \left| \sum_{m=1}^M L^*(\omega, \mathbf{r}_m) v_m(\omega, \mathbf{n}_s) \right|^2}{\sum_{m=1}^M |L(\omega, \mathbf{r}_m)|^2 + \gamma_z(\omega) \sum_{m=1}^M \sum_{p=1}^M L^*(\omega, \mathbf{r}_m) R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) L(\omega, \mathbf{r}_p)} \quad (14.10)$$

де, як й вище, $\gamma_z(\omega) = g_z(\omega)/g_w$ - відношення завада/ шум.

Досі ми не конкретизували вигляд вагових функцій просторової обробки $L(\omega, \mathbf{r}_m)$. Але у переважній більшості практичних розрахунків звичайно вважається, що сигнал на апертурі АР є плоскою хвилею, що приходить зі заданого напрямку $\mathbf{n}_s$. Тоді, базуючись на матеріалах розділу 13, неважко бачити, що для прийнятих умов задачі вагову функцію просторової обробки можна представити у наступному вигляді:

$$\dot{L}(\omega, \mathbf{r}_m) = \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot v_m(\omega, \mathbf{n}_s), \quad (14.11)$$

де, нагадаймо, $v_m(\omega, \mathbf{n}_s)$ - розподільчий вектор (14.7), а $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$ - деякий амплітудно-фазовий розподіл (АФР) по апертурі АР, налаштованої на

приймання сигналу, що приходить з напрямку¹ $\mathbf{n}_s$. Після підстановки виразу (14.11) до формули (14.10) приходимо до наступного результату:

$$\chi(\omega/\mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left| \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2 \cdot \left(\sum_{m=1}^M \left| \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2 + \right. \quad (14.12)$$

$$\left. + \gamma_z(\omega) \sum_{m=1}^M \sum_{p=1}^M \dot{H}^*(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot v_m^*(\omega, \mathbf{n}_s) \cdot R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot v_p(\omega, \mathbf{n}_s) \dot{H}(\omega, \mathbf{r}_p / \mathbf{n}_s) \right)^{-1}.$$

Співвідношення (14.12) справедливе для довільного АФР решітки. Як показано у розділі 13, у випадку оптимальної просторової обробки розподіл $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$ має задовольняти рівнянню (13.25), з урахуванням якого вираз (14.12) значно спрощується:

$$\chi_{opt.}(\omega/\mathbf{n}_s) = \sum_{m=1}^M \dot{H}_{opt.}(\omega, \mathbf{r}_m / \mathbf{n}_s). \quad (14.13)$$

Неважко бачити, що в силу ермітової симетрії кореляційної функції завад $R_z(\omega; \mathbf{r}_m, \mathbf{r}_p)$ параметр (14.12) та його окремий випадок (14.13) є дійсними функціями частоти². Коефіцієнт підсилення, обчислений відповідно до формул (14.12), (14.13), може бути застосовано до поля сигналу у вигляді плоскої хвилі та довільних завад. На основі цих виразів неважко одержати, як окремі випадки, й деякі традиційні визначення, прийняті у теорії антен.

Припустимо, зокрема, що поле зовнішніх завад в межах області прийому є статистично однорідним (точніше, квазіоднорідним), тобто, просторова кореляція завад задовольняє умові

$$R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) = R_z(\omega; \mathbf{r}_m - \mathbf{r}_p), \quad \mathbf{r}_m, \mathbf{r}_p \in \mathbf{D}_A. \quad (14.14)$$

Отже, кореляція завад поміж двома датчиками АР залежить не від кожної з координат датчиків окремо, а лише від різниці цих координат. Це співвідношення аналогічне за змістом поняттю стаціонарності випадкових процесів у часі але з переносом на властивості поля завад у просторі (див.

¹ У багатьох практичних задачах АФР від напрямку приходу сигналу, як правило, не залежить, тобто, $\dot{H}(\omega, \mathbf{r} / \mathbf{n}_s) = \dot{H}(\omega, \mathbf{r})$. Але ми поки що розглядаємо загальний випадок.

² Втім, цей висновок можна зробити й безпосередньо з визначення підсилення відповідно до виразів (14.2) – (14.4).

розділ 12). Вище вже було зроблено припущення про квазістаціонарність поля завад на інтервалі приймання сигналу. Ця гіпотеза забезпечує доцільність застосування спектральних розкладів сигналів у кожному приймальному каналі АР. Тепер робиться припущення про їх однорідність у просторі. Ця гіпотеза дає можливість застосувати розклад функції просторової кореляції завад на сукупність «просторових гармонік», що, по суті, є «просторовим» аналогом теореми Вінера-Хінчина (4.16). Зокрема, для прийнятих умов задачі, як було показано у розділі 12, коефіцієнт просторової кореляції поля завад може бути представлено розкладом

$$\dot{R}_z(\omega, \mathbf{p}) = \int_0^{2\pi} \int_0^{\pi} C_z(\omega; \mathbf{n}) \cdot \exp\{j \cdot \omega \cdot (\mathbf{n}, \mathbf{p}) / c_0\} \sin \theta d\theta d\varphi, \quad (14.15)$$

де $\mathbf{p} = \mathbf{r}' - \mathbf{r}''$ - різницевий вектор просторових координат двох точок спостереження, $\mathbf{n} = \{\cos \varphi \cdot \sin \theta, \sin \varphi \cdot \sin \theta, \cos \theta\}$ - одиничний вектор напрямку в стандартній сферичній системі координат, θ, φ - вертикальний та горизонтальний кути відповідно, $C_z(\omega; \mathbf{n})$ - кутовий спектр потужності завад, що має задовольняти умові нормування

$$\int_0^{2\pi} \int_0^{\pi} C_z(\omega; \mathbf{n}) \sin \theta d\theta d\varphi = 1. \quad (14.16)$$

Неважко бачити, що кожний елементарний член розкладу (14.15), тобто,

$$C_z(\omega; \mathbf{n}) \cdot \exp\{j \cdot \omega \cdot (\mathbf{n}, \mathbf{p}) / c_0\} \sin \theta d\theta d\varphi$$

відповідає плоскій хвилі, що приходить із напрямку $\mathbf{n}$. Такі хвилі часто й називають «просторовими гармоніками». Амплітуда кожної з цих елементарних хвиль визначається відповідним значенням кутового спектру потужності завад. Отже, вибираючи той чи інший розподіл потужності завад по напрямках прийому, можна достатньо легко «конструювати» випадкові поля із заданими властивостями, не вступаючи у протиріччя з фізичною реальністю. Тому розклади такого типу отримали широке застосування у прикладних завданнях радіофізики, оптики, акустики.

Але повернемося до нашої задачі. Після підстановки розкладу (14.15) у співвідношення (14.12) маємо наступну еквівалентну формулу для обчислення підсилення просторової обробки:

$$\chi(\omega/\mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left| \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2 \times \\ \times \left(\sum_{m=1}^M \left| \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2 + \gamma_z(\omega) \cdot \int_0^{2\pi} \int_0^\pi C_z(\omega; \mathbf{n}) \cdot \left| \dot{D}(\omega; \mathbf{n} / \mathbf{n}_s) \right|^2 \sin \theta d\theta d\phi \right)^{-1} \quad (14.17)$$

де

$$\dot{D}(\omega; \mathbf{n} / \mathbf{n}_s) = \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot \exp \{ j \cdot \omega \cdot (\mathbf{n} - \mathbf{n}_s, \mathbf{r}_m) / c_0 \} \quad (14.18)$$

– ненормована діаграма направленості (ДН) решітки, що налаштована на приймання сигналу із напрямку $\mathbf{n}_s$ на частоті ω . Враховуючи визначення (14.18) та вираз (14.13) неважко бачити, що у випадку оптимальної обробки підсилення АР може бути представлено у вигляді

$$\chi_{opt.}(\omega/\mathbf{n}_s) = \dot{D}_{opt.}(\omega, \mathbf{n}_s / \mathbf{n}_s), \quad (14.19)$$

тобто, підсилення оптимальної просторової обробки дорівнює значенню її ненормованої ДН у напрямку на сигнал.

Розглянемо деякі окремі випадки одержаних співвідношень. Припустимо спочатку, що зовнішні завади відсутні ($\gamma_z = 0$), тобто, прийому заважають лише некорельовані поміж каналами АР власні шуми приймальної системи. Тоді з виразу (14.17) витікає, що підсилення просторової обробки на тлі власних шумів має вигляд:

$$\chi_w(\omega/\mathbf{n}_s) = \frac{\left| \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2}{\sum_{m=1}^M \left| \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2} \quad (14.20)$$

У випадку оптимальної просторової обробки функція АФР має задовольняти рівнянню (13.25), звідки при відсутності зовнішніх завад ($\gamma_z = 0$) маємо рішення $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \equiv 1$. У свою чергу, для такого АФР з

виразу (14.20) одержуємо значення підсилення $\chi_{opt.} \equiv M$. Отже, для задачі приймання сигналів на тлі власних шумів підсилення (14.20) не може перевищувати числа каналів прийому, тобто, для будь-якого АФР при $\gamma_z = 0$ підсилення $\chi_{opt.} \leq M$, причому, верхня межа цієї нерівності досягається у випадку рівномірного розподілу. Будь-який нерівномірний розподіл у цій ситуації може лише погіршити справу.

Фізичне пояснення даного результату досить просте. Дійсно, просторова обробка у даному випадку визначається традиційним алгоритмом формування направленості (13.4), тобто,

$$u_{AR}(t, \mathbf{n}_s) = \sum_{m=1}^M u(t - \tau_m(\mathbf{n}_s), \mathbf{r}_m), \quad (14.21)$$

згідно з яким дія АР полягає у вирівнюванні по затримках сигналів на виходах датчиків та їх наступному підсумовуванню. Рівень сигналу в результаті такої процедури зростає пропорційно до кількості каналів прийому M . Отже, потужність корисного сигналу по виходу АР при належному фазуванні в M^2 разів більше, чим на вході кожного з датчиків. Власні ж шуми поміж виходами датчиків статистично не корельовані та однакові за своїми статистичними характеристиками. Тому при підсумовуванні вихідних сигналів датчиків дисперсія шумів зростає лише у M разів. Таким чином, за відношенням потужності сигналу до власних шумів на виході просторової обробки (14.21) забезпечується вииграш, який дорівнює числу каналів прийому M .

Тепер припустимо, що відсутні власні шуми приймальної системи, тобто, $\gamma_z \rightarrow \infty$. Тоді з виразу (14.17) маємо:

$$\chi_z(\omega) = \frac{1}{\int_0^{2\pi} \int_0^\pi C_z(\omega; \mathbf{n}) \cdot \left| \tilde{D}(\omega; \mathbf{n}/\mathbf{n}_s) \right|^2 \sin \theta d\theta d\varphi}. \quad (14.22)$$

$$\text{де} \quad \tilde{D}(\omega; \mathbf{n}/\mathbf{n}_s) = \dot{D}(\omega; \mathbf{n}/\mathbf{n}_s) / \left| \dot{D}(\omega; \mathbf{n}_s/\mathbf{n}_s) \right|, \quad (14.23)$$

- нормована ДН, модуль якої має одиничне значення³ за напрямком приходу плоскої хвилі сигналу, тобто при $\mathbf{n} = \mathbf{n}_s$. Розглянемо, зокрема, одну з найбільш поширених для практичних розрахунків моделей так званого ізотропного шуму, якому відповідає рівномірний розподіл енергії завад за усіма можливими напрямками. З урахуванням умови нормування (14.16) цій моделі завад відповідає кутовий спектр наступного вигляду:

$$C_z(\theta, \varphi) = 1/(4\pi) \quad (14.24)$$

де $\theta \in (0, \pi)$, $\varphi \in (0, 2\pi)$. Неважко впевнитися, що після підстановки кутового спектру (14.24) до розкладу (14.15) та виконання інтегрування одержимо наступний вираз для коефіцієнту просторової кореляції завад:

$$R_z(\omega, \mathbf{p}) = \frac{\sin(2\pi \cdot \rho / \lambda)}{2\pi \cdot \rho / \lambda}, \quad (14.25)$$

де, як звичайно, $\lambda = c / f$ - довжина хвилі на частоті $\omega = 2\pi f$. Підстановка кутового спектру (14.24) у вираз (14.22) приводить до наступного співвідношення щодо підсилення просторової обробки на тлі ізотропного шуму:

$$\chi_z(\omega) = \frac{4\pi}{\int_0^{2\pi} \int_0^\pi \left| \tilde{D}(\omega; \mathbf{n} / \mathbf{n}_s) \right|^2 \sin \theta d\theta d\varphi}. \quad (14.26)$$

Останній вираз співпадає з традиційним визначенням КНД антен, яке у застосуванні до випромінюючих антен також називають коефіцієнтом осьової концентрації [21]. Як бачимо, це традиційне визначення стосується дуже обмеженого кола умов. Визначення ж підсилення просторової обробки, яке ми розглядаємо, може бути застосовано для більш загальних ситуацій, коли, по-перше, присутні як зовнішні завади, так й власні шуми антени, по-друге, кутовий спектр завади, на відміну від ізотропного шуму, може бути суттєво нерівномірним за напрямками.

³ Зауважимо, що для оптимальної просторової обробки це значення не завжди є максимальним, що буде проілюстровано далі.

Зазначимо, що підсилення просторової обробки у загальному випадку можна на підставі виразів (14.17), (14.20) та (14.22) представити у вигляді:

$$\chi(\omega) = \frac{(1 + \gamma_z) \cdot \chi_w(\omega) \cdot \chi_z(\omega)}{\gamma_z \cdot \chi_w(\omega) + \chi_z(\omega)}, \quad (14.27)$$

де $\chi_w(\omega)$, $\chi_z(\omega)$ - відповідно підсилення просторової обробки на тлі власних шумів та зовнішніх завад, якщо їх розглядати порізно, тобто, обчислювати за формулами (14.20) та (14.22). Вираз (14.27) наочно показує, що власні шуми приймальної системи є одним з тих факторів, які обмежують теоретичну ефективність просторової обробки. Дійсно, для деяких видів зовнішніх завад оптимальна обробка може теоретично забезпечити їх повне придушення. Цій ситуації відповідає значення параметру $\chi_z \rightarrow \infty$. Тоді з формули (14.27) витікає, що при $\chi_z \gg \gamma_z \cdot \chi_w$ можна наближено записати:

$$\chi(\omega) \approx (1 + \gamma_z) \cdot \chi_w(\omega). \quad (14.28)$$

Отже, загальне підсилення просторової обробки навіть при практично повному придушенні зовнішніх завад обмежується двома параметрами:

- відношенням завада/шум по входу антени γ_z ;
- значенням параметру підсилення антени щодо власних шумів χ_w .

До цього слід також додати, що результатом оптимізації просторової обробки є, як правило, відхід від рівномірного АФР $\dot{H}(\omega, \mathbf{r})$. У свою чергу, нерівномірність АФР має наслідком зменшення параметру підсилення χ_w у порівнянні з його теоретично максимальним значенням, що дорівнює кількості каналів прийому M . Отже, оптимізація просторової обробки з метою придушення зовнішніх завад може одночасно призводити до підсилення власних шумів по виходу АР.

14.2. Підсилення просторової обробки у випадку приймання сигналів на тлі ізотропних завад.

Проілюструємо наведені вище загальні співвідношення прикладами вирішення практичних задач. Одною з найбільш типових моделей поля завад, що знаходить широке застосування при аналізі ефективності радіозв'язку є згаданий вище ізотропний шум. Кутовий спектр такого шуму у відповідності до виразу (14.24) не залежить від напрямку прийому. Перш за все, відзначимо принципову різницю поміж ізотропним шумом та власними шумами антени: ізотропний шум є одним з видів зовнішніх завад, що існує в середовищі незалежно від приймальної системи та має просторову кореляційну функцію вигляду (14.25). Тому кореляція ізотропного шуму поміж датчиками залежить від їх взаємного розташування у просторі. Що ж стосується власних шумів, то вони завжди некорельовані поміж датчиками при будь-якому їх розташуванні на приймальній апертурі. Тому, на відміну від ізотропного шуму, ніяка оптимізація геометрії та параметрів АР по відношенню до власних шумів неможлива: усе визначається характеристиками датчиків поля сигналу. Але у суміші зі зовнішніми завадами власний шум може суттєво впливати на результати оптимізації просторової обробки, про що вже згадувалось вище.

Обмежимося розглядом АР лінійної геометрії. Визначальними параметрами лінійних АР є кількість приймальних елементів M та крок решітки d . При виборі цих параметрів враховуються, перш за все, міркування ефективності прийому та технічні обмеження. Зокрема, до технічних обмежень можна віднести прийнятний з конструктивних міркувань розмір загальної апертури та допустиму з точки зору апаратних затрат кількість каналів прийому. Ще одним обмеженням, що має вже не технічний, а принциповий характер, є умова відсутності бокових максимумів ДН, які б що дорівнювали основному. Для забезпечення виконання цієї умови необхідно, як відомо [21], вибирати крок АР виходячи з нерівності

$$\frac{d}{\lambda} \leq 1/2. \quad (14.29)$$

Зауважимо, що нерівність (14.29) є, по суті, аналогом відомої теореми відліків (див. розділ 5) у застосуванні до функцій просторових координат. Отже, розглянемо оптимізацію кроку решітки d по критерію максимального підсилення лінійної АР на тлі ізотропного шуму з урахуванням умови (14.29). При цьому технічним обмеженням будемо вважати кількість елементів АР. Таким чином, стосовно параметрів АР маємо наступну оптимізаційну задачу:

$$d = \arg\left(\max_d \chi(\omega)\right), \quad M = \text{const}, \quad d \leq \lambda/2.$$

На рис.14.1 представлені результати обчислень коефіцієнту підсилення χ лінійної АР для числа каналів $M=10$ та значенні відношення завада/шум $\gamma_z=10^3$. Зокрема, на рис.14.1а показана залежність параметру χ від кроку АР по відношенню до довжини хвилі d/λ , на рис. 14.1б – залежність цього ж показника від кута φ_0 орієнтації ДН. При цьому значення $\varphi_0=0^\circ$ відповідає орієнтації ДН у напрямку поздовжньої вісі АР, значення $\varphi_0=90^\circ$ – фронтальному напрямку.

Суцільні криві на цих рисунках відповідають оптимальному розподілу $\dot{H}_{opt}(\omega, \mathbf{r}/\mathbf{n}_s)$, пунктирні – рівномірному розподілу $\dot{H}(\omega, \mathbf{r}_m) \equiv 1$, тобто, традиційному алгоритму формування направленості (14.21). Для порівняння тут же штрих-пунктиром нанесені результати обчислень за формулою (14.26), яка не враховує власні шуми приймальної системи. Просторова кореляційна функція ізотропних завад обчислювалася за формулою (14.25), оптимальний розподіл $\dot{H}_{opt}(\omega, \mathbf{r}/\mathbf{n}_s)$ для кожного зі значень напрямків орієнтації ДН знаходився в результаті рішення системи рівнянь (13.25). Для обчислення підсилення АР були використані формул (14.20), (14.26) та (14.27).

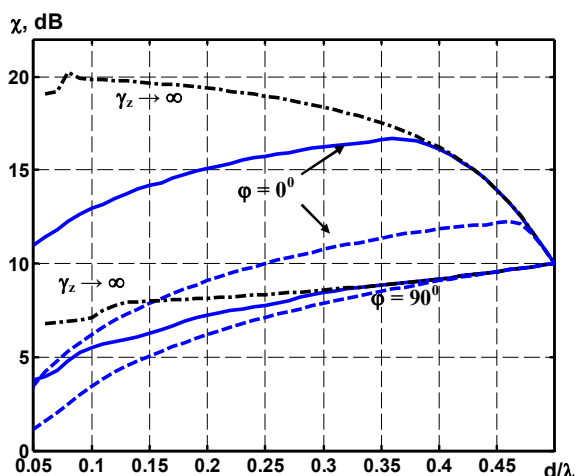


Рис.14.1а. Підсилення лінійної АР в ізотропному шумі в залежності від кроку решітки.

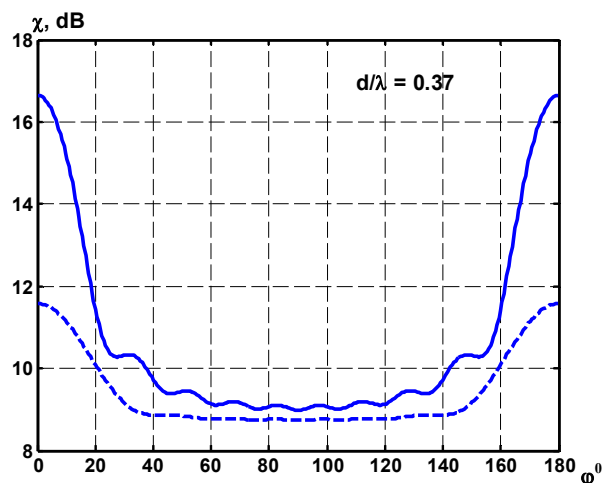


Рис.14.1б. Підсилення лінійної АР в ізотропному шумі в залежності від орієнтації ДН.

Із розгляду вказаних рисунків неважко помітити існування оптимальних значень відносного кроку решітки d/λ для оптимальної та традиційної просторової обробки за умови врахування як поля зовнішніх завад, так й власних шумів антени. Якщо ж власні шуми не брати до уваги, (штрихпунктирні криві), то ефективність оптимальної просторової обробки тим вища, чим менше значення параметру d/λ . Більш детальні обчислення показують, що оптимальні значення d/λ загалом залежать від орієнтації ДН, кількості каналів M та відношення завада/шум γ_z . Існування оптимуму по параметру d/λ обумовлюють дві протилежні тенденції: зменшення кроку решітки призводить до зростання кореляції зовнішніх завад вздовж приймальної апертури, що, у свою чергу, сприяє їх кращій компенсації. Протилежна тенденція полягає в тому, що нерівномірність оптимального розподілу $\dot{H}_{opt}(\omega, \mathbf{r})$ знижує ефективність придушення власних шумів антени. Тому зменшення параметру d/λ ефективне лише до тих пір, поки залишкові рівні зовнішніх завад та власних шумів по виходу АР приблизно не зрівняються. При подальшому зменшенні кроку решітки краща компенсація зовнішніх завад вже мало впливає на загальну ситуацію, натомість зростає рівень власних шумів по виходу АР. Наведені міркування свідчать про ту

ключову роль, яку може відігравати фактор власних шумів приймальної системи при оптимізації її параметрів.

Звернемо також увагу на залежність підсилення АР та результатів оптимізації від напрямку орієнтації ДН. Вище вже зазначалося, що підсилення просторової обробки для випадку приймання сигналу на тлі власних шумів незалежно від геометрії АР дорівнює числу її датчиків M , що для нашого чисельного прикладу складає 10 дБ. Але як видно з рис.14.1а, для фронтальних до лінійної АР напрямків приходу сигналу ($\varphi = 90^\circ$) підсилення оптимальної обробки у випадку приймання на тлі суміші власних шумів та ізотропних завад сягає рівня 10 дБ лише при відносному кроці решітки $d/\lambda = 1/2$, а для менших значень кроку решітки підсилення менше 10 дБ. У той же час, для повздовжнього напрямку ($\varphi = 0^\circ$) оптимізація просторової обробки забезпечує значення підсилення майже 17 дБ. Цей результат, вочевидь, є наслідком специфічної особливості лінійної геометрії АР, яка має найменший загальний об'єм ДН саме для поздовжнього напрямку спостереження, тобто, для випадку приймання так званої «біжучої хвилі».

Отже, наведені результати показують, що як при оптимальній, так й традиційній просторовій обробці існує оптимум по параметру відносного кроку решітки d/λ , що визначається фактором наявності власних шумів антени. Але у випадку традиційної обробки оптимальне значення цього параметру не дуже відрізняється від $1/2$. Що ж до оптимальної просторової обробки, то тут варіацій оптимального відносного кроку решітки набагато більше, хоча помітні резерви для оптимізації є лише для повздовжніх напрямків прийому лінійної АР.

14.3. Підсилення просторової обробки для випадку приймання сигналів на тлі локальних завад.

Тепер розглянемо принципово іншу модель зовнішніх завад, а саме, локальних, які, на відміну від ізотропних шумів, зосереджені за явно вираженими напрямками у просторі. Цій моделі завад відповідає кутовий спектр у вигляді суперпозиції дельта-функцій, тобто:

$$C_z(\mathbf{n}) = \sum_{k=1}^{N_{\text{лок.}}} \nu_k \delta(\mathbf{n} - \mathbf{n}_k), \quad (14.30)$$

де $\mathbf{n}_k, k = 1, \dots, N_{\text{лок.}}$ - орти напрямків приходу плоских хвиль локальних завад, $N_{\text{лок.}}$ - кількість локальних завад, ν_k - їх відносні рівні, що задовольняють умові

$$\sum_{k=1}^{N_{\text{лок.}}} \nu_k = 1. \quad (14.31)$$

Після підстановки кутового спектру (14.30) у розклад (14.15) знаходимо відповідний коефіцієнт просторової кореляції локальних завад:

$$\dot{R}_z(\omega, \mathbf{p}) = \sum_{k=1}^{N_{\text{лок.}}} \nu_k \cdot \exp\{j \cdot \omega \cdot (\mathbf{n}_k, \mathbf{p}) / c_0\}. \quad (14.32)$$

Вважаємо, як звичайно, що окрім локальних завад також присутній власний шум приймальної системи. Дана модель завад детально розглядалася у §13.4, де були отримані усі необхідні для розрахунків формули. Зокрема, показано, що структура оптимальної просторової обробки для такої моделі завад може бути представлена співвідношення (13.60), яке, по-суті, є об'єднанням з певними ваговими коефіцієнтами $N_{\text{лок.}} + 1$ каналів, що виконують формування ДН за традиційним алгоритмом (13.54) із рівномірним АФР, тобто,

$$\dot{U}_{AR}(\omega, \mathbf{n}) = \sum_{m=1}^M \dot{U}(\omega, \mathbf{r}_m) \exp\{j \omega \cdot (\mathbf{n}, \mathbf{r}_m) / c_0\}. \quad (14.33)$$

При цьому один з каналів має бути орієнтовано у напрямку приходу сигналу $\mathbf{n}_s$, а інші $N_{\text{лок.}}$ каналів - по напрямках приходу локальних завад $\mathbf{n}_k$.

Перейдемо до аналізу підсилення просторової обробки. Перш за все розглянемо обробку за традиційним алгоритмом (14.33), тобто, без врахування локальних завад. У даному разі доцільно скористатися формулою (14.12), з якої після підстановки рівномірного розподілу $\dot{H}(\omega, \mathbf{r}_m) \equiv 1/M$ та співвідношень (14.7) та (14.32) маємо:

$$\chi_{trad}(\omega / \mathbf{n}_s) = \frac{(1 + \gamma_z(\omega)) \cdot M}{1 + \gamma_z(\omega) \cdot M \sum_{k=1}^{N_{\text{лок}}} \nu_k \cdot \left| \tilde{D}_{trad}(\omega; \mathbf{n}_s - \mathbf{n}_k) \right|^2}, \quad (14.34)$$

де

$$\tilde{D}_{trad}(\omega; \mathbf{n}) = \frac{1}{M} \sum_{m=1}^M \exp\{j \cdot \omega \cdot (\mathbf{n}, \mathbf{r}_m) / c_0\} \quad (14.35)$$

- нормована ДН традиційної просторової обробки з рівномірним АФР.

У випадку, коли ДН сигнального каналу (14.35) має нулі за напрямками приходу локальних завад, тобто, за умови

$$\tilde{D}_{trad}(\omega; \mathbf{n}_s - \mathbf{n}_k) = 0, \quad k = 1, \dots, N_{\text{лок}}, \quad (14.36)$$

підсилення традиційної просторової обробки (14.34) досягає своєї верхньої границі

$$\chi_{trad}(\omega / \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot M. \quad (14.37)$$

Але для довільних напрямків приходу локальних завад умову (14.36) виконати практично, як правило, неможливо. Отже, реальне підсилення практично завжди буде менше верхньої межі (14.37), причому, величина програшу визначається рівнями «пролазу» локальних завад до сигнальної ДН, тобто, значеннями лівої частини співвідношення (14.36).

Тепер розглянемо для цієї ж ситуації оптимальну просторову обробку. Скориставшись співвідношення (13.60) можна записати наступний вираз для оптимального АФР просторової обробки даної задачі:

$$\dot{H}_{opt.}(\omega, \mathbf{r}_m / \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot \left(1 - \sum_{k=1}^{N_{\text{лок}}} \dot{\beta}_k(\omega) \cdot \dot{\nu}_m(\omega, \mathbf{n}_k) \cdot \dot{\nu}_m^*(\omega, \mathbf{n}_s) \right), \quad (14.38)$$

$$m = 1, \dots, M,$$

де $\dot{v}_m(\omega, \mathbf{n})$ - компоненти розподільчого вектору АР (14.7), $\dot{\beta}_k(\omega)$ - оптимальні вагові коефіцієнти, які знаходяться у результаті рішення системи рівнянь (13.61).

Після підстановки виразу (14.38) до формули (14.13) та з урахуванням співвідношень (14.7), (14.18) та (14.35) знаходимо наступний вираз для підсилення оптимальної просторової обробки даної задачі:

$$\chi_{opt.}(\omega / \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot M \cdot \left(1 - \sum_{k=1}^{N_{лок.}} \dot{\beta}_k(\omega) \cdot \tilde{D}_{trad.}(\omega; \mathbf{n}_s - \mathbf{n}_k) \right). \quad (14.39)$$

В окремому випадку однієї локальної завади великої інтенсивності ($\gamma_z \gg 1$), що приходить з напрямку $\mathbf{n}_{лок.} \neq \mathbf{n}_s$, ваговий коефіцієнт оптимальної обробки визначається співвідношенням (13.56), з урахуванням якого вираз (14.39) приймає вигляд

$$\chi_{opt.}(\omega / \mathbf{n}_s) = (1 + \gamma_z(\omega)) \cdot M \cdot \left(1 - \left| \tilde{D}_{trad.}(\omega; \mathbf{n}_s - \mathbf{n}_{лок.}) \right|^2 \right). \quad (14.40)$$

Останнє співвідношення ілюструє рис.14.2, на якому представлено залежність підсилення лінійної АР від напрямку приходу локальної завади для оптимальної (суцільна крива) та традиційної (пунктир) обробки.

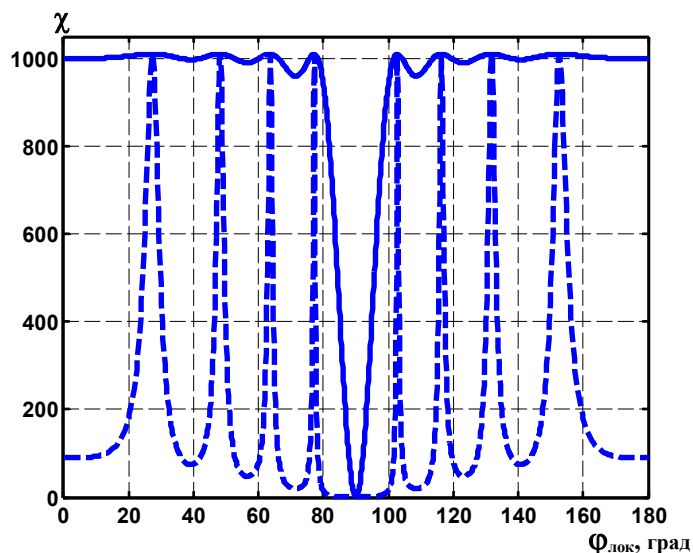


Рис.14.2. Підсилення лінійної АР в присутності однієї локальної завади в залежності від напрямку її приходу

Розрахунок виконано для числа датчиків $M=10$, відносного шагу решітки $d/\lambda=0.45$, напрямку приходу сигналу $\varphi_s=90^\circ$ та відношення завади/шум $\gamma_z=100$. З рисунку видно, що при оптимальній обробці підсилення для переважної більшості напрямків приходу завади є досить близьким до максимально можливого рівня $(1+\gamma_z)\cdot M=1010$. Суттєве погіршення ефективності оптимальної обробки спостерігається лише у відносно вузькому секторі кутів, що безпосередньо прилягають до напрямку приходу сигналу φ_s . Як витікає з виразу (14.40), зменшення ефективності оптимальної компенсації однієї локальної завади внаслідок часткового придушення сигналу визначається фактором

$$\Delta\chi_{opt.}=1-\left|\tilde{D}_{trad.}(\omega;\mathbf{n}_s-\mathbf{n}_1)\right|^2,$$

який за межами основного пелюстка ДН опорного каналу складає, як правило, долі дБ.

У випадку традиційної обробки (14.33) максимальна ефективність можлива лише для окремих напрямків приходу локальної завади, які співпадають з нулями відповідної ДН. Для переважної ж більшості напрямків оптимальна обробка значно виграє у порівнянні з традиційним формуванням направленості, причому, величина виграшу при прийнятих параметрах досягає 10-15 дБ. По мірі зростання відносної інтенсивності завади γ_z цей виграш, вочевидь, буде пропорційно збільшуватися.

Тепер розглянемо випадок приймання сигналу на тлі кількох локальних завад. Зокрема, на рис.14.3 показані нормовані ДН лінійної 10-елементної АР при використанні оптимальної (суцільна крива) та традиційної просторової обробки (пунктир). Розрахунок ДН традиційної обробки було виконано за формулою (14.35), а для оптимальної обробки – за загальною формулою (14.18) та наступним нормуванням (14.23) при використанні оптимального АФР (14.38).

Модель, для якої виконувались обчислення, включала суміш власних шумів та 5 локальних завад, напрямки приходу яких відмічено на рис.14.3 стрілками. З рисунку видно, що ДН антени з оптимальною обробкою має в напрямках приходу локальних завад чітко сформовані нулі. В той же час, бокові пелюстки ДН в напрямках, де нема завад, значно підвищені у порівнянні з ДН антени з традиційною просторовою обробкою.

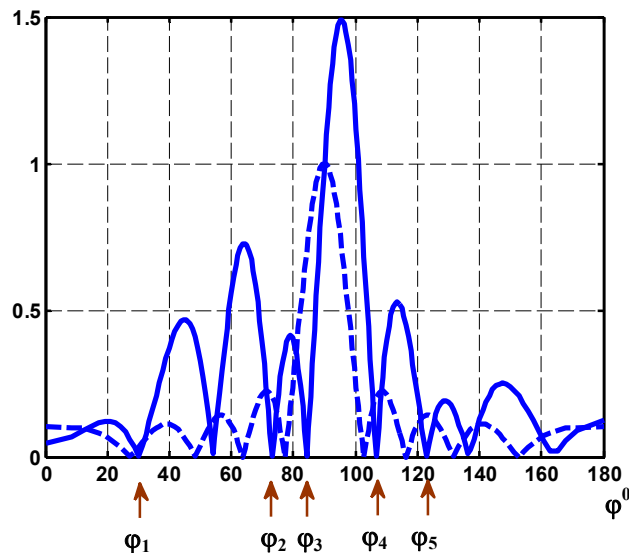


Рис.14.3. Нормовані діаграми направленості лінійної АР у присутності 5 локальних завад

Отже, вигреш оптимальної обробки у порівнянні з традиційною забезпечується за рахунок перерозподілу рівня пелюстків ДН антени. Для прийнятих у розрахунках параметрів підсилення оптимальної обробки становило 25.9дБ, у той час, як для традиційної обробки воно складало 9.25дБ, тобто, вигреш за рахунок оптимізації просторової обробки досягає у даному прикладі ~ 16.7 дБ.

Також відзначимо, що, максимум нормованої ДН оптимальної обробки не співпадає з напрямком прийому сигналу та перевищує одиничне значення. Цей ефект зумовлено попаданням однієї із локальних завад у межі основного сектору прийому. Формування нуля по напрямку приходу цієї завади й призвело до деформації головного пелюстка ДН.

Контрольні запитання та завдання до розділу 14.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Що таке підсилення просторової обробки? Чим воно відрізняється від підсилення антен?
2. Самостійно повторити усі викладки, що приводять до виразів (14.10), (14.12). Довести, що підсилення просторової обробки завжди є дійсним параметром.
3. Самостійно одержати вирази (14.17), (14.20), (14.22), (14.26).
4. Які фактори обмежують підсилення просторової обробки на тлі зовнішніх завад? Як це фізично пояснити?
5. В чому полягає принципова різниця поміж ізотропним шумом та власними шумами приймальної системи? Як цю різницю можна продемонструвати фізично?
6. Прокоментувати графіки, що представлені на рис.14.1а,б.
7. Повторити самостійно графіки, що представлені на рис.14.2 та їх пояснити.
8. Повторити самостійно графіки, що представлені на рис.14.3 та їх пояснити.

Розділ 15. Ефективність виявлення сигналів.

У попередніх розділах були розглянуті загальні структури алгоритмів виявлення сигналів та проведена їх деталізація для певних моделей полів сигналів та завад. Тепер перейдемо до розгляду основних показників ефективності виявлення сигналів. Результати аналізу таких показників мають важливе значення для оцінки доцільності використання тих чи інших алгоритмів оброблення сигналів та можуть бути використані для пошуку методів їх найбільш ефективної реалізації.

15.1. Загальні відомості.

У розділі 11 було встановлено, що практично будь-яке нерандомізоване правило виявлення сигналів має звичайно вигляд нерівності

$$F(\mathbf{u}) \underset{\hat{\theta}=0}{\overset{\hat{\theta}=1}{>}} \Pi, \quad (15.1)$$

де $F(\mathbf{u})$ – деякий функціонал обробки (тестова статистика) над вектором спостережень $\mathbf{u}$ (дискретних або безперервних), $\hat{\theta}$ – рішення про значення параметру ситуації θ , яке відповідає наявності ($\theta = 1$) або відсутності ($\theta = 0$) сигналу, Π – поріг виявлення. Як зазначалося у §10.3, відповідно до прийнятого в локаційних застосуваннях критерію Неймана-Пірсона основними показниками ефективності виявлення сигналів є ймовірність їх правильного виявлення $P_{п.в.}$ при заданій ймовірності хибних тривог $P_{х.т.}$. Нагадаємо, що рівнем хибних тривог називають ймовірність прийняття рішення $\hat{\theta} = 1$ у ситуації $\theta = 0$, тобто, при відсутності сигналу. Отже,

$$P_{х.т.} = P(F(\mathbf{u}) \geq \Pi / \theta = 0).$$

У відповідності до критерію Неймана-Пірсона це співвідношення звичайно розглядається як рівняння стосовно порогу виявлення, тобто:

$$P(F(\mathbf{u}) \geq \Pi / \theta = 0) = \alpha_0, \quad (15.2)$$

де α_0 - заданий рівень ймовірності хибних тривог. Рішення $\Pi = \Pi(\alpha_0)$ рівняння (15.2) використовується надалі у процедурі виявлення (15.1).

Аналогічним чином знаходиться ймовірність правильного виявлення сигналу $P_{\text{п.в.}}$:

$$P_{\text{п.в.}} = P(F(\mathbf{u}) \geq \Pi(\alpha_0) / \theta = 1). \quad (15.3)$$

В деяких застосуваннях використовують також й додаткові характеристики, зокрема, повні ймовірності правильних $P_{\text{прав.}}$ та помилкових $P_{\text{пом.}}$ рішень:

$$\begin{aligned} P_{\text{прав.}} &= p_1 \cdot P_{\text{п.в.}} + p_0 \cdot (1 - P_{\text{х.т.}}), \\ P_{\text{пом.}} &= p_1 \cdot (1 - P_{\text{п.в.}}) + p_0 \cdot P_{\text{х.т.}}, \end{aligned}$$

де p_θ - апіорні ймовірності ситуацій наявності ($\theta = 1$) та відсутності ($\theta = 0$) сигналу. Ці та інші ймовірності звичайно обчислюють у функціональній залежності від рівня хибних тривог α_0 та деяких параметрів завадостійкості. Найважливішими з таких параметрів є відношення сигнал/завада¹ та пов'язані з ними характеристики, які обчислюються як безпосередньо у середовищі передачі сигналу, так й на різних етапах його оброблення.

Тепер, користуючись цими загальними співвідношеннями, розглянемо ефективність виявлення сигналів в типових умовах приймання сигналів при застосуванні функціоналів обробки, які були одержані у попередніх розділах. Зазначимо, що базові параметри ефективності виявлення сигналів (15.2) і (15.3) залежать від розподілу тестової статистики $F(\mathbf{u})$ в ситуаціях $\theta = 0$ та $\theta = 1$. У свою чергу, вказаний розподіл визначається структурою функціоналу $F(\mathbf{u})$, статистичними характеристиками спостережень та типом сигналів, які підлягають виявленню. У розділі 11 було визначено три основних типів сигналів та відповідних задач їх виявлення: виявлення точно відомих сигналів, виявлення сигналів із невідомими параметрами та

¹ У світовій літературі цей параметр звичайно прийнято позначати SNR (від Signal to Noise Ratio). Це ж позначення на протязі даного посібника будемо використовувати й ми.

виявлення випадкових сигналів. Розглянемо ефективність виявлення для кожного з цих сигналів окремо.

15.2. Виявлення точно відомих сигналів.

Якщо вигляд сигналу та усі його параметри наперед відомі (нагадаємо, що такі сигнали також називають детермінованими), то, як було показано у §11.2, оптимальною статистикою для задачі виявлення є кореляція (11.11) прийнятих на приймальній апертурі реалізацій поля $u(t, \mathbf{r}_m)$ з ваговими функціями $b(t, \mathbf{r}_m)$, тобто:

$$V(\mathbf{u}) = \sum_{m=1}^M \int_{\mathbf{r}_m} b(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt, \quad (15.4)$$

де оптимальні вагові функції $b(t, \mathbf{r}_m)$ знаходяться в результаті рішення системи рівнянь (11.12). Але в даному розділі, на відміну від (11.11), будемо вважати, що вагові функції $b(t, \mathbf{r})$ можуть вибиратися довільним чином, тобто, не є обов'язково оптимальними. Як вже було зазначено у §11.2, відповідно до прийнятої моделі шумів та завад розподіл ймовірностей поля спостережень $u(t, \mathbf{r})$ є гавсівським як при відсутності, так й при наявності сигналу. Лінійні ж функціонали від гавсівських величин (див. розділ 3) також мають гавсівський розподіл. Отже, розподіл статистики (15.4) є нормальним, тобто, визначається її математичними очікуваннями $m_V(\theta) = E\{V(\mathbf{u})/\theta\}$ і дисперсіями $\sigma_V^2(\theta) = E\{V^2(\mathbf{u})/\theta\} - m_V^2(\theta)$ в ситуаціях $\theta = 0$ та $\theta = 1$. Діючи за методикою, що вже була застосована у §10.3, неважко знайти загальний вираз для ймовірності правильного виявлення сигналу при заданому рівні ймовірності хибних тривог:

$$P_{\text{п.в.}} = \Phi \left\{ \sqrt{q_V^2} - \frac{\sigma_V(\theta=0)}{\sigma_V(\theta=1)} \cdot \Phi^{-1}(1 - \alpha_0) \right\}, \quad (15.5)$$

де $\Phi\{\cdot\}$ – інтегральна функція нормального розподілу (10.15) (інтеграл Лапласа), $\Phi^{-1}\{\cdot\}$ – зворотна до нього функція,

$$q_R^2 = \frac{(m_V(\theta=1) - m_V(\theta=0))^2}{\sigma_V^2(\theta=1)} \quad (15.6)$$

– параметр завадостійкості, який прийнято називати **відношенням сигнал/завада по виходу² приймального пристрою** [3]. Походження цієї назви пояснюється тим, що будь-який позитивний приріст тестової статистики $V(\mathbf{u})$ правила виявлення (15.1) сприяє прийняттю рішення на користь гіпотези $\theta=1$. Отже, сигналом по виходу функціоналу обробки доцільно вважати різницю поміж математичними очікуваннями статистики $V(\mathbf{u})$ в ситуаціях $\theta=1$ та $\theta=0$. Прийняттю правильного рішення заважають флуктуації значень цієї статистики, які зумовлені дією завад. Оскільки чисельною мірою флуктуацій випадкової величини є її дисперсія, то це й пояснює назву параметру (15.6). Взагалі кажучи, треба відрізнити параметр (15.6) від первинного відношення сигнал/завада q_{in}^2 на вході приймача, яке визначається об'єктивним чином та не залежить від способу обробки сигналу (детальніше ця різниця буде проілюстрована далі). Повертаючись до класу функціоналів (15.4), неважко бачити, що для прийнятих у §11.2 статистичних моделей полів спостережень математичне очікування цієї статистики має вигляд

$$m_V(\theta) = \theta \cdot \sum_{m=1}^M \int_{\mathbf{T}_m} b(t, \mathbf{r}_m) \cdot s(t, \mathbf{r}_m) dt,$$

а дисперсія визначається співвідношенням

$$\sigma_V^2(\theta) = \sum_{m=1}^M \sum_{p=1}^M \int_{\mathbf{T}_m} \int_{\mathbf{T}_p} b(t, \mathbf{r}_m) \cdot E\{u(t, \mathbf{r}_m) \cdot u(t', \mathbf{r}_p) / \theta\} \cdot b(t', \mathbf{r}_p) dt dt'.$$

Модель кореляції значень поля $u(t, \mathbf{r})$ була визначена у §11.2. З урахуванням цієї моделі легко бачити, що в незалежності від значення параметру ситуації θ останній вираз приймає вигляд:

² Нижній індекс підкреслює належність даного параметру до виходу приймача (від англ. **Receiver**).

$$\sigma_V^2 = g_w \cdot \sum_{m=1}^M \int_{\mathbf{T}} b^2(t, \mathbf{r}_m) dt + \sum_{m=1}^M \sum_{p=1}^M \int_{\mathbf{T}_m} \int_{\mathbf{T}_p} b(t, \mathbf{r}_m) \cdot K_z\{t, t'; \mathbf{r}_m, \mathbf{r}_p\} \cdot b(t', \mathbf{r}_p) dt dt'.$$

Оскільки у випадку приймання детермінованих сигналів $\sigma_V^2(\theta = 0) = \sigma_V^2(\theta = 1)$, то вираз (15.5) для ймовірності правильного виявлення сигналу дещо спрощується, а саме:

$$P_{\text{п.в.}} = \Phi \left\{ \sqrt{q_R^2} - \Phi^{-1}(1 - \alpha_0) \right\}. \quad (15.7)$$

У свою чергу, з урахуванням одержаних вище співвідношень для математичних очікувань та дисперсії тестової статистики (15.4) відношення сигнал/завада (15.6) по виходу приймальної системи може бути записано у наступному вигляді:

$$q_R^2 = \left(\sum_{m=1}^M \int_{\mathbf{T}_m} b(t, \mathbf{r}_m) \cdot s(t, \mathbf{r}_m) dt \right)^2 \times \left(g_w \cdot \sum_{m=1}^M \int_{\mathbf{T}_m} b^2(t, \mathbf{r}_m) dt + \sum_{m=1}^M \sum_{p=1}^M \int_{\mathbf{T}_m} \int_{\mathbf{T}_p} b(t, \mathbf{r}_m) \cdot K_z\{t, t'; \mathbf{r}_m, \mathbf{r}_p\} \cdot b(t', \mathbf{r}_p) dt dt' \right)^{-1}. \quad (15.8)$$

Неважко бачити, що в окремому випадку оптимальної обробки, коли вагові функції $b_{opt}(t, \mathbf{r})$ задовольняють системи рівнянь (11.12), вираз (15.8) приймає вигляд

$$q_{R_{opt}}^2 = \sum_{m=1}^M \int_{\mathbf{T}_m} b_{opt}(t, \mathbf{r}_m) \cdot s(t, \mathbf{r}_m) dt. \quad (15.9)$$

Тепер припустимо, що виконуються умови квазістаціонарного наближення для компактних апертур, визначені у §13.2. В цьому наближенні, як було показано у розділі 13, доцільно використовувати еквівалентні спектральні форми запису. Зокрема, виразу (15.8) можна поставити у відповідність його спектральний аналог, що має вигляд

$$q_R^2 = \frac{1}{2\pi} \left| \sum_{m=1}^M \int_{-\infty}^{\infty} \dot{B}^*(\omega, \mathbf{r}_m) \cdot \dot{S}(\omega, \mathbf{r}_m) d\omega \right|^2 \times \left(g_w \cdot \sum_{m=1}^M \int_{-\infty}^{\infty} |\dot{B}(\omega, \mathbf{r}_m)|^2 d\omega + \sum_{m=1}^M \sum_{p=1}^M \int_{-\infty}^{\infty} \dot{B}^*(\omega, \mathbf{r}_m) \cdot \dot{K}_z\{\omega; \mathbf{r}_m, \mathbf{r}_p\} \cdot \dot{B}(\omega, \mathbf{r}_p) d\omega \right)^{-1}. \quad (15.10)$$

Тут $\dot{B}(\omega, \mathbf{r})$ - спектр Фур'є вагової функції обробки $b(t, \mathbf{r})$, $\dot{K}_z\{\omega; \mathbf{r}_m, \mathbf{r}_p\}$ - кореляція спостережень по виходу m -го та p -го датчиків АР на частоті ω , $\dot{S}(\omega, \mathbf{r})$ - спектр Фур'є поля сигналу на інтервалі спостереження. Відповідно до прийнятої вище моделі (див. вирази (13.19), (13.20)) спектр сигналу на вході АР може бути представлений у вигляді

$$\dot{S}(\omega, \mathbf{r}_m) = \frac{1}{r_s} \cdot \dot{S}_0(\omega) \cdot \exp(j\omega\tau_s) \cdot v_m(\omega, \mathbf{n}_s). \quad (15.11)$$

У цій моделі, як й раніше, $S_0(\omega)$ - спектр Фур'є сигналу біля джерела, $\tau_s = r_s / c_0$ - затримка сигналу, r_s - відстань до джерела, c_0 - швидкість поширення сигналу в фізичному середовищі, $v_m(\omega, \mathbf{n}_s)$ - компоненти (14.48) розподільчого вектору АР, $\mathbf{n}_s$ - одиничний вектор напрямку приходу плоскої хвилі сигналу.

Для поля сигналу вигляду (15.11) у розділі 13 була одержана структура оптимальної обробки, спектральна вагова функція $\dot{B}(\omega, \mathbf{r})$ якої визначається співвідношеннями (13.24), (13.25). По аналогії з ними представимо спектральну функцію обробки у вигляді

$$\dot{B}(\omega, \mathbf{r}_m) = \frac{S_0(\omega) \cdot \exp(j\omega\tau_s)}{g_w + g_z(\omega)} \cdot \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot v_m(\omega, \mathbf{n}_s), \quad (15.12)$$

де $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s), (m=1, \dots, M)$ - деякий амплітудно-фазовий розподіл (АФР) на просторовій апертурі, що налаштовано для приймання сигналу з напрямку $\mathbf{n}_s$. Як й у випадку оптимальної обробки, розглянутої у розділі 13, вибір спектральної функції у вигляді (15.12) дає можливість розділення загальної обробки сигналу на часову та просторову. При цьому, як легко бачити з виразів (15.11) та (15.12), часова частина обробки повністю узгоджена з формою сигналу джерела, але АФР просторової обробки $\dot{H}(\omega, \mathbf{r} / \mathbf{n}_s)$ у даному випадку може вибиратися довільним чином.

Після підстановки співвідношень (15.11) та (15.12) до формули (15.10) маємо наступний вираз для SNR по виходу алгоритму приймання (15.4):

$$\begin{aligned}
q_R^2 = & \nu(r_s) \cdot \frac{T}{2\pi} \cdot \left| \int_{-\infty}^{\infty} \mu_0(\omega) \cdot \sum_{m=1}^M \dot{H}^*(\omega, \mathbf{r}_m / \mathbf{n}_s) d\omega \right|^2 \times \\
& \times \left(\int_{-\infty}^{\infty} \frac{\mu(\omega)}{1 + \gamma_z(\omega)} \left(\sum_{m=1}^M |\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)|^2 + \right. \right. \\
& \left. \left. + \gamma_z(\omega) \sum_{m=1}^M \sum_{p=1}^M \dot{H}^*(\omega, \mathbf{r}_m / \mathbf{n}_s) \cdot \nu_m^*(\omega, \mathbf{n}_s) \cdot \dot{R}_z\{\omega; \mathbf{r}_m, \mathbf{r}_p\} \cdot \nu_p(\omega, \mathbf{n}_s) \cdot \dot{H}(\omega, \mathbf{r}_p / \mathbf{n}_s) \right) d\omega \right)^{-1},
\end{aligned} \tag{15.13}$$

де $\nu(r_s) = 1/r_s^2$ - втрати потужності сигналу при його поширенні у фізичному середовищі³,

$$\mu_0(\omega) = \frac{1}{T} \cdot \frac{|\dot{S}_0(\omega)|^2}{(g_w + g_z(\omega))} \tag{15.14}$$

– відношення спектральної щільності потужності сигналу біля джерела до спектральної щільності потужності шумів та завад в умовному центрі АР. Будемо у подальшому називати параметр (15.14) відношенням сигнал/завада біля джерела на частоті ω .

Знаменник виразу (15.13) можна представити через розглянутий у попередньому розділі параметр підсилення просторової обробки $\chi(\omega / \mathbf{n}_s)$. Для цього достатньо скористатися формулою (14.12). Таким чином, зі співвідношень (15.13) та (14.12) витікає наступний вираз для SNR по виходу алгоритму приймання (15.4):

$$q_R^2 = \nu(r_s) \cdot \frac{T}{2\pi} \cdot \frac{\left| \int_{-\infty}^{\infty} \mu_0(\omega) \cdot \sum_{m=1}^M \dot{H}^*(\omega, \mathbf{r}_m / \mathbf{n}_s) d\omega \right|^2}{\int_{-\infty}^{\infty} \frac{\mu_0(\omega)}{\chi(\omega)} \cdot \left| \sum_{m=1}^M \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s) \right|^2 d\omega}. \tag{15.15}$$

Нагадаємо, що вираз (15.15) одержано у припущенні, що часова частина обробки повністю узгоджена з формою сигналу джерела, а просторова частина відповідає довільному розподілу $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n}_s)$. Якщо ж й просторова

³ У даному випадку ми користуємося моделлю середовища у вигляді однорідного безмежного простору. Для іншої моделі середовища формула (15.13) залишиться такою ж, але залежність втрат потужності $\nu(r_s)$ сигналу від дистанції може бути не обов'язково зворотно квадратичною.

обробка є оптимальною, то підсилення АР визначається співвідношенням (14.13), з урахуванням якого із виразу (15.15) маємо

$$q_{R_{opt}}^2 = \nu(r_s) \cdot \frac{T}{2\pi} \cdot \int_{-\infty}^{\infty} \mu_0(\omega) \cdot \chi_{opt.}(\omega/\mathbf{n}_s) d\omega \quad (15.16)$$

Але й у загальному випадку вираз (15.15) також можна значно спростити, якщо скористатися тією обставиною, що радіосигнали звичайно займають вузьку смугу частот $\Delta\omega_s$ у порівнянні зі своєю несучою частотою ω_0 . Відповідним чином також будуть обмежені й границі інтегрування по частоті у виразі (15.15). Тоді за умови $\Delta\omega_s/\omega_0 \ll 1$ можна звичайно вважати, що в межах смуги приймальної системи розподіл $\dot{H}(\omega, \mathbf{r}_m/\mathbf{n}_s)$, підсилення просторової обробки $\chi(\omega)$ та спектр потужності завад $g_z(\omega)$ практично не змінюються, тобто:

$$\begin{aligned} \dot{H}(\omega, \mathbf{r}_m/\mathbf{n}_s) &\cong \dot{H}(\omega_0, \mathbf{r}_m/\mathbf{n}_s), \\ \chi(\omega/\mathbf{n}_s) &\cong \chi(\omega_0/\mathbf{n}_s), \\ g_z(\omega) &\cong g_z(\omega_0), \end{aligned} \quad (15.17)$$

де $|\omega - \omega_0| \leq \Delta\omega_s/2$. За умови (15.17) відповідні функції у виразі (15.15) можна винести за межі інтегралів по частоті. В результаті приходимо до значно простішого результату:

$$q_R^2 \cong q_{in}^2 \cdot \chi(\omega_0), \quad (15.18)$$

де

$$q_{in}^2 = \nu(r_s) \cdot \frac{E_s}{g_w + g_z(\omega_0)} \quad (15.19)$$

- SNR на вході приймальної системи,

$$E_s = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\dot{S}_0(\omega)|^2 d\omega = \int_0^T s_0^2(t) dt$$

- повна енергія сигналу джерела. Тут доречно нагадати, що спектральні щільності потужності шумів g_w та завад g_z , які ми використовуємо у даному посібнику, відповідають їх звичайному математичному визначенню. Отже,

дисперсія $\sigma_{\Delta f}^2$ шумів та завад у деякій смузі частот Δf знаходиться як добуток загальної СЩП шумів та завад $g_w + g_z$ на задану ширину смуги Δf з врахуванням як позитивних так й негативних частот, тобто,

$$\sigma_{\Delta f}^2 = 2\Delta f \cdot (g_w + g_z).$$

Але у прикладній літературі зі статистичної радіотехніки часто застосовується так зване «фізичне» визначення СЩП, яке враховує множник «2» в останньому виразі, тобто, дисперсія шумів та завад у цій же смузі визначається як

$$\sigma_{\Delta f}^2 = \Delta f \cdot (\tilde{g}_w + \tilde{g}_z),$$

де $\tilde{g}_w = 2 \cdot g_w$, $\tilde{g}_z = 2 \cdot g_z$. Вказана різниця не є принциповою, але при виконанні деяких практичних розрахунків її слід враховувати.

Отже, застосування припущень (15.17) дозволило привести вираз для SNR по виходу приймальної системи, що виконує оброблення сигналів за алгоритмом (15.4), до наочного та зручного для практичного користування вигляду (15.18). Зокрема, із розгляду цього виразу видно, що SNR по виходу обробки відрізняється від аналогічного параметру q_{in}^2 на вході приймача лише коефіцієнтом підсилення AP. Звідси витікає, що у разі приймання сигналу одиночним ненаправленим датчиком, якому відповідає коефіцієнт підсилення $\chi = 1$, SNR по входу та по виходу приймальної системи мають співпадати. Але слід взяти до уваги, що цей висновок є справедливим лише у тому разі, коли часова частина обробки повністю узгоджена із формою сигналу джерела. Отже, при не ідеальній практичній реалізації обробки можуть мати місце й певні втрати завадостійкості, які також звичайно мають бути враховані при обчисленні SNR по виходу приймальної системи.

На останок даного параграфу зауважимо, що хоча припущення (15.17) для переважної більшості завдань приймання радіосигналів звичайно виконуються, але для сигналів іншої фізичної природи вони можуть бути й не прийнятними. Зокрема, така ситуація нерідко виникає для деяких завдань

приймання гідроакустичних сигналів [7]. Отже, у цих випадках розрахунок параметру завадостійкості q_R^2 слід виконувати за більш загальними формулами (15.15) або (15.16).

15.3. Виявлення сигналів із невідомими параметрами.

Тепер розглянемо клас задач, у яких поле сигналу $s(t, \mathbf{r})$ відоме лише з точністю до сукупності деяких параметрів $\boldsymbol{\eta}$ (такі сигнали часто називають квазідетермінованими). У §11.3 було показано, що оброблення поля сигналів даного типу складається з послідовності двох етапів: лінійної обробки (11.20) та нелінійних процедур. При цьому статистична модель флуктуацій параметрів $\boldsymbol{\eta}$ у каналі передачі впливає лише на структуру нелінійної обробки, а функціонал лінійної обробки (11.20) від цієї моделі не залежить. У свою чергу, нелінійні процедури оброблення були конкретизовані у §11.3 для типових моделей середовища передачі сигналу, а саме, каналу з випадковою фазою сигналу, релеєвського та райсівського каналів. Розглянемо обчислення параметрів ефективності виявлення сигналів для вказаних моделей.

1. Канал із випадковою початковою фазою сигналу.

Відповідно до моделі §11.3 поле сигналу у каналі даного типу має вигляд

$$s_{ch}(t, \mathbf{r}_m / \varphi_{ch}) = \text{Re}\{\dot{s}_0(t, \mathbf{r}_m) \cdot \exp(j\varphi_{ch})\}, \quad (15.20)$$

де φ_{ch} - випадковий зсув фази сигналу в каналі, $\dot{s}_0(t, \mathbf{r}_m)$ комплексне аналітичне представлення поля сигналу, що відповідає деякому відомому значенню початкової фази, наприклад, $\varphi_{ch} = 0$. Як було зазначено вище, лінійний етап оброблення визначається функціоналом (11.20), який зводиться до обчислення кореляції поміж полем прийнятого сигналу $u(t, \mathbf{r})$ та комплексною ваговою функцією $\dot{b}(t, \mathbf{r})$. У свою чергу, вказана функція знаходиться в результаті рішення рівняння (11.21). Правило прийняття рішення, що має вигляд (11.25), передбачає обчислення модуля комплексної величини (11.20) та порівняння її з порогом. Але у подальшому зручніше

розглядати еквівалентне йому правило, яке відрізняється лише піднесенням обох частин нерівності (11.25) у квадрат, що принципово нічого не змінює. Отже, будемо аналізувати вирішуюче правило⁴

$$\left| \dot{V}(\mathbf{u}) \right|^2 = \left| \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot u(t, \mathbf{r}_m) dt \right|^2 \underset{H_0}{\overset{H_1}{>}} \text{П.} \quad (15.21)$$

Відповідно до прийнятих у розділі 11 припущень поле на приймальній апертурі $u(t, \mathbf{r})$ є сумішшю поля сигналу (15.20), поля зовнішніх завад $z(t, \mathbf{r})$ та поля власних шумів приймальної системи $w(t, \mathbf{r})$. Таким чином, вихід лінійної частини обробки правила (15.21) може бути представлено у вигляді:

$$\begin{aligned} \dot{V}(\mathbf{u}) = & \theta \cdot \int_{\mathbf{T}} \text{Re} \{ \dot{s}_0(t, \mathbf{r}_m) \cdot \exp(j \cdot \varphi_{ch}) \} \cdot \dot{b}(t, \mathbf{r}_m) dt + \\ & + \sum_{m=1}^M \int_{\mathbf{T}_m} z(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt + \sum_{m=1}^M \int_{\mathbf{T}_m} w(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt, \end{aligned} \quad (15.22)$$

де, як й вище, θ - параметр ситуації наявності або відсутності сигналу. Беручи до уваги тотожність $\text{Re}(\dot{z}) = (\dot{z} + \dot{z}^*)/2$ перший (сигнальний) член виразу (15.22) можна переписати у вигляді:

$$\begin{aligned} \dot{V}_s(\mathbf{u}) = & \theta \cdot \int_{\mathbf{T}} \text{Re} \{ \dot{s}_0(t, \mathbf{r}_m) \cdot \exp(j \cdot \varphi_{ch}) \} \cdot \dot{b}(t, \mathbf{r}_m) dt = \\ = & \frac{\theta}{2} \int_{\mathbf{T}} \dot{s}_0(t, \mathbf{r}_m) \cdot \exp(j \cdot \varphi_{ch}) \cdot \dot{b}(t, \mathbf{r}_m) dt + \frac{\theta}{2} \int_{\mathbf{T}} \dot{s}_0^*(t, \mathbf{r}_m) \cdot \exp(-j \cdot \varphi_{ch}) \cdot \dot{b}(t, \mathbf{r}_m) dt. \end{aligned}$$

Перший доданок у цьому виразі можна відкинути, оскільки він звичайно відповідає складовим подвоєної частоти. В цьому неважко переконатися для окремого випадку співпадіння вагової функції обробки із сигналом, тобто, $\dot{b}(t, \mathbf{r}_m) = \dot{s}_0(t, \mathbf{r}_m)$. Але цей же висновок є вірним й для загального випадку оптимальної обробки, коли вагова функція $\dot{b}(t, \mathbf{r}_m)$ та поле очікуваного сигналу $\dot{s}_0(t, \mathbf{r}_m)$ пов'язані через рівняння (11.21) позитивно визначеним та

⁴ При цьому позначення порогу залишаємо тим же, оскільки це на сутність задачі не впливає.

симетричним кореляційним оператором завад. Таким чином, результат лінійної обробки (15.22) можна представити у вигляді

$$\begin{aligned} \dot{V}(\mathbf{u}) = & \frac{\theta}{2} \cdot \exp(-j \cdot \varphi_{ch}) \cdot Q_0 + \sum_{m=1}^M \int_{\mathbf{T}_m} z(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt + \\ & + \sum_{m=1}^M \int_{\mathbf{T}_m} w(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt, \end{aligned} \quad (15.23)$$

$$\text{де} \quad Q_0 = \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot \dot{s}_0^*(t, \mathbf{r}_m) dt \quad (15.24)$$

- сигнальна частина функціоналу лінійної обробки (15.22) для точно відомого сигналу. Згідно до прийнятих у розділі 11 робочих моделей розподіл шумів та завад вважається гавсівським. Отже, при фіксованій фазі сигналу φ_{ch} комплексна величина (15.23) також є гавсівською з наступними параметрами:

– умовне математичне очікування

$$\dot{m}_V(\theta / \varphi_{ch}) = E\{\dot{V}(\mathbf{u}) / \theta\} = \frac{1}{2} \cdot \theta \cdot \exp(-j \cdot \varphi_{ch}) \cdot Q_0; \quad (15.25)$$

– умовна дисперсія

$$\begin{aligned} 2\sigma_V^2(\theta / \varphi_{ch}) = & E\left\{\left|\dot{V}(\mathbf{u}) - \dot{m}_V(\theta)\right|^2 / \theta\right\} = 2\sigma_V^2 = \\ = & g_w \cdot \sum_{m=1}^M \int_{\mathbf{T}} \left|\dot{b}(t, \mathbf{r}_m)\right|^2 dt + \sum_{m=1}^M \sum_{p=1}^M \int_{\mathbf{T}} \int_{\mathbf{T}} K_z(t, t'; \mathbf{r}_m, \mathbf{r}_p) \cdot b(t, \mathbf{r}_m) \cdot b(t', \mathbf{r}_p) dt dt', \end{aligned} \quad (15.26)$$

де, як й вище, g_w - спектральна щільність потужності власних шумів приймальної системи, $K_z(t, t'; \mathbf{r}, \mathbf{r}')$ - кореляційна функція зовнішніх завад. З виразу (15.26) витікає, що для даної моделі каналу дисперсія сигналу по виходу лінійної обробки (15.23) від параметру ситуації θ та набігу фази φ_{ch} в каналі не залежить. Також неважко показати, що дисперсії дійсної $V_c = \text{Re} \dot{V}(\mathbf{u})$ та уявної $V_s = \text{Im} \dot{V}(\mathbf{u})$ частин функціоналу лінійної обробки (15.23) співпадають. Тому параметр σ_V^2 у співвідношенні (15.26) позначає, як це загально прийнято [3], дисперсію шумів та завад на один канал

квадратурної обробки, причому ця дисперсія для ситуацій $\theta = 0$ та $\theta = 1$ однакова. Таким чином, спільний умовний розподіл квадратурних компонент $V_c = \text{Re}(\dot{V})$, $V_s = \text{Im}(\dot{V})$ лінійної частини обробки (15.23) при фіксованій фазі сигналу φ_{ch} має вигляд:

$$p(V_c, V_s / \theta, \varphi_{ch}) = \frac{1}{2\pi\sigma_V^2} \exp \left\{ -\frac{1}{2\sigma_V^2} \left| (V_c + j \cdot V_s) - \frac{1}{2} \cdot \theta \cdot \exp(-j \cdot \varphi_{ch}) \cdot Q_0 \right|^2 \right\}. \quad (15.27)$$

Для того, щоб одержати ймовірності перевищення порогу для вирішуючого правила (15.21), треба записати розподіл квадрату модуля статистики $\dot{V} = V_c + j \cdot V_s$. Цей розподіл неважко одержати, якщо від змінних V_c , V_s перейти до нових величин $y = V_c^2 + V_s^2$ та $\psi = \text{artg}(V_s / V_c)$. Зворотне перетворення має, вочевидь, вигляд $V_c = \sqrt{y} \cos \psi$, $V_s = \sqrt{y} \sin \psi$, а якобіан перетворення дорівнює $1/2$. Отже, з урахуванням вказаного перетворення з виразу (15.27) знаходимо спільний умовний розподіл величин y та ψ :

$$p(y, \psi / \theta, \varphi_{ch}) = \frac{1}{4\pi\sigma_V^2} \exp \left\{ -\frac{1}{2\sigma_V^2} \left| \sqrt{y} \cdot \exp(j \cdot \psi) - \frac{1}{2} \cdot \theta \cdot \exp(-j \cdot \varphi_{ch}) \cdot Q_0 \right|^2 \right\}$$

Для того, щоб знайти розподіл квадрату модуля виходу лінійної частини обробки $y = |\dot{V}|^2$, цей вираз треба згорнути по змінній ψ , розподіл якої звичайно приймається рівномірним на інтервалі $\varphi_{ch} \in (0, 2\pi)$. Отже, після розкриття в останньому виразі показника експоненти, перегруповування членів та використання формули (11.24), для вказаного розподілу одержуємо наступну формулу:

$$\begin{aligned} p(y / \theta, \varphi_{ch}) &= \frac{1}{2\pi} \int_0^{2\pi} p(y, \psi / \varphi_{ch}, \theta) d\psi = \\ &= \frac{1}{2\sigma_V^2} \exp \left\{ -\frac{1}{2\sigma_V^2} (y + \theta \cdot a^2) \right\} \cdot I_0 \left(\frac{\theta \cdot a \cdot \sqrt{y}}{\sigma_V^2} \right) = p(y / \theta) \end{aligned}, \quad (15.28)$$

де $I_0(z)$ – модифікована функція Беселя нульового порядку [9],

$$a = |\dot{m}_V(\theta = 1)| = \frac{1}{2} \cdot |Q_0|.$$

Зазначимо, що умовний, взагалі кажучи, розподіл (15.28) від фази сигналу в каналі φ_{ch} не залежить, тобто, після згортання по змінній ψ умовний розподіл став безумовним. Отже, вираз (15.28) є для нашої задачі й остаточною. Тепер проаналізуємо одержаний результат.

У випадку $\theta = 0$ з виразу (15.28) знаходимо розподіл шумів та завад по виходу приймальної системи:

$$p(y/\theta = 0) = \frac{1}{2\sigma_V^2} \exp\left\{-\frac{y}{2\sigma_V^2}\right\}.$$

Звідси вже не важко записати ймовірність хибних тривог правила (15.21):

$$P_{\text{х.т.}} = P(y \geq \Pi/\theta = 0) = \frac{1}{2\sigma_V^2} \int_{\Pi}^{\infty} \exp\left\{-\frac{y}{2\sigma_V^2}\right\} dy = \exp\left\{-\frac{\Pi}{2\sigma_V^2}\right\}. \quad (15.29)$$

Із даного співвідношення можна знайти поріг Π , який забезпечує виконання умови $P_{\text{х.т.}} = \alpha_0$, де α_0 - заданий рівень хибних тривог, а саме:

$$\Pi = -2\sigma_V^2 \cdot \ln \alpha_0. \quad (15.30)$$

У свою чергу, для ймовірності правильного виявлення сигналу в ситуації $\theta = 1$ із виразу (15.28) витікає наступна формула:

$$P_{\text{п.в.}} = P(y \geq \Pi/\theta = 1) = \frac{1}{2\sigma_V^2} \int_{\Pi}^{\infty} \exp\left\{-\frac{y+a^2}{2\sigma_V^2}\right\} \cdot I_0\left(\frac{a \cdot \sqrt{y}}{\sigma_V^2}\right) dy.$$

Після заміни $y = \sigma_V^2 \cdot x^2$ одержимо:

$$P_{\text{п.в.}} = P(y \geq \Pi/\theta = 1) = \int_{\sqrt{\Pi/\sigma_V^2}}^{\infty} x \cdot \exp\left\{-\frac{x^2 + q_R^2}{2}\right\} \cdot I_0(q_R \cdot x) dx,$$

$$\text{де} \quad q_R^2 = \frac{a^2}{\sigma_V^2} = \frac{|Q_0|^2}{4 \cdot \sigma_V^2}. \quad (15.31)$$

- параметр завадостійкості приймальної системи. З урахуванням формули (15.30) остаточно маємо:

$$P_{\text{п.в.}} = Q_{\text{марс}}(\sqrt{q_R^2}, \sqrt{-2 \cdot \ln \alpha_0}), \quad (15.32)$$

$$\text{де} \quad Q_{\text{marc}}(v, z) = \int_z^\infty x \cdot \exp\left\{-\frac{x^2 + v^2}{2}\right\} \cdot I_0(v \cdot x) dx \quad (15.33)$$

- табульований інтеграл, який звичайно називають⁵ Q -функцією Маркума (Marcum Q-function) [3].

Тепер розглянемо структуру параметру (15.31), який з урахуванням співвідношень (15.24) та (15.26) може бути записаний у наступному розгорнутому вигляді:

$$q_R^2 = \frac{1}{2} \cdot \left| \sum_{m=1}^M \int_{\mathbf{T}_m} \dot{b}(t, \mathbf{r}_m) \cdot \dot{s}_0^*(t, \mathbf{r}_m) dt \right|^2 \times \quad (15.34)$$

$$\times \left(g_w \cdot \sum_{m=1}^M \int_{\mathbf{T}_m} |\dot{b}(t, \mathbf{r}_m)|^2 dt + \sum_{m=1}^M \sum_{p=1}^M \int_{\mathbf{T}_m} \int_{\mathbf{T}_p} \dot{b}^*(t, \mathbf{r}_m) \cdot K_z\{t, t'; \mathbf{r}_m, \mathbf{r}_p\} \cdot \dot{b}(t', \mathbf{r}_p) dt dt' \right)^{-1}.$$

Порівнюючи цю формулу з виразом (15.8), який було одержано у попередньому параграфі для випадку прийому точно відомого сигналу, можна довести їх тотожність. Дійсно, базуючись на матеріалах §1.3 неважко показати, що для двох довільних аналітичних сигналів $\dot{s}_i(t) = s_i(t) + j \cdot \hat{s}_i(t)$, $i = 1, 2$ де $s_i(t)$ і $\hat{s}_i(t)$ - відповідно їх синфазні та квадратурні компоненти, мають місце наступні загальні співвідношення:

$$\int_{\mathbf{T}} s_i(t) \cdot s_k(t) dt = \int_{\mathbf{T}} \hat{s}_i(t) \cdot \hat{s}_k(t) dt, \quad (15.35)$$

$$\int_{\mathbf{T}} s_i(t) \cdot \hat{s}_k(t) dt = 0, \quad i, k = 1, 2,$$

де інтегрування виконується по усій області $\mathbf{T}$ спільного існування сигналів. Беручи до уваги ці співвідношення та симетричні властивості кореляційного оператора завад, нескладно встановити тотожність виразу (15.34) до виразу (15.8). Таким чином, методика та результати розрахунку SNR по виходу лінійної частини обробки для даної моделі каналу не відрізняються від випадку каналу із відомими параметрами. Специфіка ж моделі каналу впливає лише на структуру нелінійної обробки та, відповідно,

⁵ Інтеграл (15.33) іноді також називають функцією Релея-Райса [3].

розподіл тестової статистики правила виявлення (15.21). У свою чергу, розподіл тестової статистики визначає такі важливі показники ефективності, як ймовірність хибних тривог $P_{х.т.}$ та ймовірність правильного виявлення сигналу $P_{п.в.}$. Отже, при одному й тому ж значенні SNR по виходу лінійної частини обробки q_R^2 ефективність виявлення сигналу може бути суттєво різною в залежності від прийнятої моделі флуктуацій в каналі.

Відносне погіршення або покращання завадостійкості для тієї чи іншої моделі каналу прийнято оцінювати за величиною так званого порогового відношення сигнал/завада $q_{Rnop.}^2 = q_{Rnop.}^2(D_0, \alpha_0)$, тобто, це таке значення параметру завадостійкості q_R^2 , що забезпечує задане значення ймовірності правильного виявлення D_0 при заданому рівні хибних тривог α_0 . У даному разі ця величина знаходиться як розв'язок рівняння (15.32) відносно q_R^2 при $P_{п.в.} = D_0$ та фіксованому значенні параметру α_0 . Деякі результати таких обчислень, що відповідають значенню $D_0 = 0.9$, представлені у табл.15.1. Обчислення проводились у середовищі Матлаб з використанням функції `marcumq`.

Таблиця 15.1. Порогові SNR (дБ) для трьох моделей каналу.

Рівень хибних тривог α_0	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
Канал із точно відомими параметрами сигналу ($\gamma_{coh} = 1$).	16.23	15.61	14.88	13.98	12.81	11.15	8.18
Канал із випадковою початковою фазою сигналу.	16.76	16.21	15.54	14.77	13.78	12.43	10.21
Релеєвський канал ($\gamma_{coh} = 0$)	24.83	24.15	23.36	22.38	21.11	19.32	16.20

Як видно з таблиці, зростання (тобто, погіршення) параметру $q_{Rnop.}^2$ внаслідок відсутності інформації про початкову фазу сигналу доволі незначне. Крім того, зменшення заданого рівня хибних тривог α_0 супроводжується й помітним зменшенням цих втрат. Наприклад, для рівня

$\alpha_0 = 10^{-1}$ погіршення у пороговому SNR для каналу з випадковою фазою у порівнянні з каналом, де параметри сигналу точно відомі, складає близько 2дБ, а при пониженні рівня хибних тривог до $\alpha_0 = 10^{-7}$ втрати зменшуються вже до 0.5дБ.

2. Канал із випадковими амплітудою та фазою сигналу

Для даної моделі каналу випадковими вважаються вже два параметри: коефіцієнт передачі каналу за амплітудою ε_{ch} та початкова фаза сигналу φ_{ch} . Відповідно до матеріалів §11.3 поле сигналу на вході приймальної системи, яка працює в такому каналі, доцільно представити наступним чином:

$$s_{ch}(t, \mathbf{r} / \varepsilon_{ch}, \varphi_{ch}) = s_{AR}(t / \dot{\eta}_{ch}) = \operatorname{Re} \{ \dot{s}_0(t, \mathbf{r}) \cdot \dot{\eta}_{ch}^* \},$$

де $\dot{\eta}_{ch} = \varepsilon_{ch} \cdot \exp(-j \cdot \varphi_{ch})$ - комплексний коефіцієнт передачі каналу, $\dot{s}_0(t, \mathbf{r})$ - поле сигналу, що відповідає деякому стандартному (тобто, відомому) значенню η_{ch} , наприклад, $\eta_{ch} \equiv 1$. Як було показано у §11.3, оптимальне правило виявлення сигналу для даної моделі каналу має вигляд (11.34), ліва частина якого передбачає наявність каналів когерентного та некогерентного оброблення з наступним їх ваговим підсумовуванням. Але дане правило значного практичного застосування не знайшло через відсутність, як правило, апріорних даних про K_{ch} - фактор каналу. Тому у цьому розділі обмежимося аналізом лише того ж некогерентного правила виявлення (15.21), яке розглядалося вище для попередньої моделі каналу та яке у даному випадку не є оптимальним.

Перш за все запишемо по аналогії з виразом (15.23) вихідний ефект лінійної частини обробки, а саме,

$$\begin{aligned} \dot{V}(\mathbf{u}) = & \frac{\theta}{2} \cdot \dot{\eta}_{ch} \cdot Q_0 + \sum_{m=1}^M \int_{\mathbf{T}_m} z(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt + \\ & + \sum_{m=1}^M \int_{\mathbf{T}_m} w(t, \mathbf{r}_m) \cdot \dot{b}(t, \mathbf{r}_m) dt, \end{aligned} \quad (15.36)$$

де усі позначення ті ж самі, що й у попередньому випадку. Як бачимо, різниця у порівнянні з виразом (15.23) стосується лише сигнальної частини виходу обробки. Відповідно до прийнятих у §11.3 моделей флуктуації комплексного коефіцієнту каналу передачі $\dot{\eta}$ є гавсівськими. Оскільки шуми та завади також гавсівські, то неважко бачити, що статистика (15.36) при обох гіпотезах $\theta = 0$ та $\theta = 1$ має нормальний розподіл ймовірностей з наступними параметрами:

- математичне очікування

$$m_V(\theta) = \frac{\theta}{2} \cdot \dot{a}_{ch} \cdot \dot{Q}_0; \quad (15.37)$$

- дисперсія

$$\sigma_V^2(\theta) = \sigma_V^2 + \frac{\theta}{4} \cdot \sigma_\xi^2 \cdot |\dot{Q}_0|^2 = \sigma_V^2 \cdot (1 + \theta \cdot \sigma_\xi^2 \cdot q_R^2). \quad (15.38)$$

Тут $\sigma_V^2 = \sigma_V^2(\theta = 0)$ - дисперсія шумів та завад на один канал квадратурної обробки, яка визначається співвідношенням (15.26), q_R^2 - SNR (15.31) по виходу приймальної системи, $\dot{a}_{ch} = \langle \dot{\eta}_{ch} \rangle$ - математичне очікування коефіцієнта передачі каналу по ансамблю його реалізацій, σ_ξ^2 - дисперсія його флуктуацій $\dot{\xi} = \dot{\eta}_{ch} - \dot{a}_{ch}$. З урахуванням прийнятих у §11.3 умов нормування параметри флуктуацій коефіцієнта передачі каналу мають вигляд

$$\dot{a}_{ch} = \sqrt{\gamma_{coh}} \cdot \exp(-j\varphi_0); \quad \sigma_\xi^2 = \frac{1}{2}(1 - \gamma_{coh}), \quad (15.39)$$

де γ_{coh} - параметр когерентності сигналу (11.31). Враховуючи дані визначення, вираз (15.38) можна переписати у вигляді:

$$\sigma_V^2(\theta) = \sigma_V^2 \cdot (1 + \theta \cdot (1 - \gamma_{coh}) \cdot q_R^2 / 2) \quad (15.40)$$

Зазначимо, що на відміну від попередньої моделі каналу дисперсія флуктуацій сигналу по виходу лінійної обробки в ситуаціях $\theta = 0$ та $\theta = 1$ неоднакова. Тепер перейдемо до аналізу ймовірностей хибних тривог $P_{х.т.}$ та

правильного виявлення сигналу $P_{\text{п.в.}}$. Неважко бачити, що в ситуації $\theta = 0$ статистичні характеристики виходу лінійної обробки (15.36) ті ж самі, що й для попередньої моделі каналу. Отже, ймовірність хибних тривог та поріг виявлення визначаються відповідно співвідношеннями (15.29) та (15.30).

Розгляд формул для обчислення ймовірності правильного виявлення сигналів доцільно розділити на два випадки.

а). *Релесівський канал*, якому відповідає значення параметру когерентності $\gamma_{\text{coh}} = 0$. По аналогії з (15.29) та з врахуванням (15.40) формула для ймовірності правильного виявлення може бути записана у вигляді

$$P_{\text{п.в.}} = P(y \geq \Pi / \theta = 1) = \exp \left\{ -\frac{\Pi}{2\sigma_V^2(\theta = 1)} \right\} = \exp \left\{ -\frac{\Pi}{2\sigma_V^2 \cdot (1 + q_R^2/2)} \right\}.$$

З урахуванням співвідношення (15.30) остаточно маємо:

$$P_{\text{п.в.}} = \exp \left\{ \frac{2 \cdot \ln \alpha_0}{2 + q_R^2} \right\}. \quad (15.41)$$

З виразу (15.41) неважко знайти порогове SNR по виходу лінійної частини обробки, що відповідає заданому значенню ймовірності правильного виявлення сигналу $P_{\text{п.в.}} = D_0$, а саме:

$$q_{\text{Rпор.}}^2 = 2 \cdot \left(\frac{\ln \alpha_0}{\ln D_0} - 1 \right). \quad (15.42)$$

Результати розрахунків за формулою (15.42) представлено у табл.15.1 (нижній рядок), що була наведена вище. Як бачимо, втрати ефективності для релесівського каналу у порівнянні з каналом, де параметри сигналу точно відомі, перевищують 8 дБ. При цьому рівень хибних тривог на величину втрат впливає незначним чином, що суттєво відрізняє цей тип каналу від розглянутого вище каналу із випадковою початковою фазою.

б). *Райсівський канал*, для якого параметр когерентності $0 < \gamma_{\text{coh}} < 1$.

Викладки у цьому разі повністю аналогічні тим, що виконувалися для моделі каналу із випадковою фазою. Отже, запишемо перш за все спільний

розподіл квадратурних компонент сигналу (15.36). По аналогії з виразом (15.27) маємо:

$$p(V_c, V_s / \theta = 1) = \frac{1}{2\pi\sigma_V^2(1)} \exp \left\{ -\frac{1}{2\sigma_V^2(1)} \left| (V_c + j \cdot V)_s - \frac{1}{2} \cdot \dot{a}_{ch} \cdot \dot{Q}_0 \right|^2 \right\}.$$

Далі, як й вище, переходимо до полярних координат:

$$p(y, \psi / \theta = 1) = \frac{1}{4\pi\sigma_V^2(1)} \exp \left\{ -\frac{1}{2\sigma_V^2(1)} \left| \sqrt{y} \cdot \exp(j \cdot \psi) - \frac{1}{2} \cdot \dot{a}_{ch} \cdot \dot{Q}_0 \right|^2 \right\},$$

де $y = V_c^2 + V_s^2$ та $\psi = \text{artg}(V_s / V_c)$. Розподіл модуля статистики (15.36) знаходиться як інтеграл по фазі ψ від цього розподілу на інтервалі $(0, 2\pi)$. Після інтегрування з використанням формули (11.24) одержимо:

$$\begin{aligned} p(y / \theta = 1) &= \frac{1}{2\pi} \int_0^{2\pi} p(y, \psi / \theta = 1) d\psi = \\ &= \frac{1}{2\sigma_V^2(1)} \exp \left\{ -\frac{1}{2\sigma_V^2(1)} \left(y + \frac{1}{4} \cdot |\dot{a}_{ch}|^2 \cdot |\dot{Q}_0|^2 \right) \right\} \cdot I_0 \left(\frac{|\dot{a}_{ch}| \cdot |\dot{Q}_0| \cdot \sqrt{y}}{2\sigma_V^2(1)} \right). \end{aligned}$$

Звідси можна записати ймовірність правильного виявлення сигналу:

$$\begin{aligned} P_{\text{п.в.}} &= P(y \geq \Pi / \theta = 1) = \int_{\Pi}^{\infty} p(y / \theta = 1) dy = \\ &= \frac{1}{2\sigma_V^2(1)} \int_{\Pi}^{\infty} \exp \left\{ -\frac{1}{2\sigma_V^2(1)} \left(y + \frac{1}{4} \cdot |\dot{a}_{ch}|^2 \cdot |\dot{Q}_0|^2 \right) \right\} \cdot I_0 \left(\frac{|\dot{a}_{ch}| \cdot |\dot{Q}_0| \cdot \sqrt{y}}{2\sigma_V^2(1)} \right) dy. \end{aligned}$$

Після заміни $y = \sigma_V^2(1) \cdot x^2$ приходимо до виразу

$$P_{\text{п.в.}} = \int_{\sqrt{\Pi / \sigma_V^2(1)}}^{\infty} x \cdot \exp \left\{ -\frac{1}{2} \left(x^2 + \frac{1}{4} \cdot \frac{|\dot{a}_{ch}|^2 \cdot |\dot{Q}_0|^2}{\sigma_V^2(1)} \right) \right\} \cdot I_0 \left(\frac{|\dot{a}_{ch}| \cdot |\dot{Q}_0| \cdot x}{2\sigma_V(1)} \right) dx.$$

Скориставшись виразами (15.30), (15.31), (15.37), (15.39), остаточно маємо:

$$P_{\text{п.в.}} = Q_{\text{marc}} \left(\sqrt{\frac{\gamma_{coh} \cdot q_R^2}{1 + (1 - \gamma_{coh}) \cdot q_R^2 / 2}}, \sqrt{\frac{-2 \cdot \ln \alpha_0}{1 + (1 - \gamma_{coh}) \cdot q_R^2 / 2}} \right), \quad (15.43)$$

де, як й вище, $Q_{\text{marc}}(v, z)$ - табульована функція (15.33). Неважко бачити, що окремими випадками формули (15.43) є при $\gamma_{\text{coh}} = 0$ вираз (15.41), що відповідає релеєвському каналу, а при $\gamma_{\text{coh}} = 1$ - формула (15.32), яка була одержана вище для моделі каналу із випадковою фазою сигналу.

15.4. Виявлення шумових сигналів

Як вже зазначалося у §11.4, суттєва особливість шумових сигналів (їх також називають випадковими), полягає в тому, що добуток смуги сигналу Δf_s на його тривалість T_s є, як правило, дуже великим, тобто, $\Delta f_s \cdot T_s \gg 1$. У свою чергу, інтервал кореляції випадкових сигналів у часі $\tau_s^{\text{кор.}}$, як відомо, є зворотно пропорційним до його смуги (див. розділ 4), тобто, $\tau_s^{\text{кор.}} \sim 1/\Delta f_s$. Таким чином, при накопиченні енергії шумового сигналу на інтервалі спостереження відбувається усереднення великої кількості сигнальних відліків, поміж якими практично нема статистичного зв'язку. Відповідно до центральної граничної теореми теорії ймовірностей (див. §3.2) таке усереднення сприяє нормалізації розподілу вихідного ефекту приймальної системи, тобто, тестової статистики $F(\mathbf{u})$. Отже, для обчислення ймовірності правильного виявлення шумових сигналів можна з достатньою для практичних застосувань точністю скористатися виразом (15.5), який є загальним для будь-яких гавсівських статистик. Залишилось лише визначити параметри формули (15.5), зокрема, SNR по виходу обробки (15.6) та дисперсію тестової статистики $\sigma_F^2(\theta)$. Для цього, вочевидь, треба визначити вигляд функціоналу $F(\mathbf{u})$. Як було показано у §11.4, функціонал оптимальної обробки випадкових сигналів визначається співвідношенням (11.36). Але спершу ми розглянемо більш загальний клас алгоритмів обробки, в яких параметри просторової та часової обробки можуть бути задані довільним чином. Вибираючи в якості прототипу функціонал (11.42), розглянемо алгоритм прийому шумових сигналів наступного вигляду:

$$F(\mathbf{u}) = \int_0^T y^2(t) dt, \quad (15.44)$$

$$\text{де} \quad y(t) = \int_{-\infty}^t b(t-\tau) \cdot u_{AR}(\tau) d\tau \quad (15.45)$$

- алгоритм часової обробки, $b(\tau)$ - відповідний імпульсний відгук фільтру, $u_{AR}(t)$ - вихід просторової обробки. Нагадаємо, що таке представлення просторово-часової обробки можливе лише у рамках квазістаціонарного наближення для компактних апертур (див. §13.2).

У відповідності з робочими моделями та з урахуванням наведених вище міркувань розподіл функціоналу (15.44) при обох гіпотезах відносно параметру ситуації θ можна, як правило, вважати близьким до нормального. Тому з урахуванням матеріалів §14.1 для математичного очікування та дисперсії статистики (15.44) можна записати наступні вирази:

$$m_F(\theta) = \langle F(\mathbf{u}) / \theta \rangle = \frac{T}{2\pi} \int_{-\infty}^{\infty} |B(\omega)|^2 \cdot [g_0^{AR}(\omega) + \theta \cdot g_s^{AR}(\omega)] d\omega; \quad (15.46)$$

$$\begin{aligned} \sigma_F^2(\theta) &= \langle F^2(\mathbf{u}) / \theta \rangle - (\langle F(\mathbf{u}) / \theta \rangle)^2 = \\ &= \frac{T}{\pi} \int_{-\infty}^{\infty} |B(\omega)|^4 \cdot [g_0^{AR}(\omega) + \theta \cdot g_s^{AR}(\omega)]^2 d\omega, \end{aligned} \quad (15.47)$$

де $B(\omega)$ - перетворення Фур'є від $b(\tau)$, $g_0^{AR}(\omega) = g_{out}^{AR}(\omega / \theta = 0)$ - СЦП шумів і завад по виходу просторової обробки, яка визначається виразом (14.8), $g_s^{AR}(\omega) = g_{out}^{AR}(\omega / \theta = 1) - g_{out}^{AR}(\omega / \theta = 0)$ - відповідна СЦП сигналу. Підставляючи вирази (15.46), (15.47) до формул (15.5), (15.6), знаходимо ймовірність правильного виявлення шумового сигналу:

$$P_{п.в.} = \Phi \left\{ \sqrt{q_F^2} - \sqrt{\gamma_F^2} \cdot \Phi^{-1}(1 - \alpha_0) \right\} \quad (15.48)$$

де

$$q_F^2 = \frac{\frac{T}{4\pi} \left(\int_{-\infty}^{\infty} |B(\omega)|^2 \cdot g_0^{AR}(\omega) \cdot \mu_{out}^{AR}(\omega) d\omega \right)^2}{\int_{-\infty}^{\infty} |B(\omega)|^4 \cdot (g_0^{AR}(\omega))^2 \cdot (1 + \mu_{out}^{AR}(\omega))^2 d\omega}, \quad (15.49)$$

- відношення сигнал/завада по виходу приймальної системи,

$$\gamma_F^2 = \frac{\sigma_F^2(\theta=0)}{\sigma_F^2(\theta=1)} = \frac{\int_{-\infty}^{\infty} |B(\omega)|^4 \cdot (g_0^{AR}(\omega))^2 d\omega}{\int_{-\infty}^{\infty} |B(\omega)|^4 \cdot (g_0^{AR}(\omega))^2 \cdot (1 + \mu_{out}^{AR}(\omega))^2 d\omega} \quad (15.50)$$

- відношення дисперсій тестової статистики для гіпотез $\theta = 0$ та $\theta = 1$,

$\mu_{out}^{AR}(\omega)$ - відношення сигнал/завада по виходу просторової обробки (14.3).

Параметр $\mu_{out}^A(\omega)$ можна зв'язати з характеристиками сигналу та завад по входу приймальної системи так, як це було зроблено у §14.1. При цьому СЩП завад та сигналу по виходу просторової обробки визначаються загальним виразом (14.8), звідки можна записати:

$$g_0^{AR}(\omega) = g_w \sum_{m=1}^M |\dot{L}(\omega, \mathbf{r}_m)|^2 + g_z(\omega) \cdot \sum_{m=1}^M \sum_{p=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot R_z(\omega; \mathbf{r}_m, \mathbf{r}_p) \cdot \dot{L}(\omega, \mathbf{r}_p), \quad (15.51)$$

$$g_s^{AR}(\omega) = g_s(\omega) \cdot \left| \sum_{m=1}^M \dot{L}^*(\omega, \mathbf{r}_m) \cdot v_m(\omega, \mathbf{n}_s) \right|^2, \quad (15.52)$$

де, як й у попередніх розділах, g_w , $g_z(\omega)$ - СЩП відповідно власних шумів та зовнішніх завад по входу приймальної системи, $g_s(\omega)$ - спектральна потужність сигналу біля джерела, $R_z(\omega; \mathbf{r}, \mathbf{r}')$ - нормована функція просторової кореляції завад, $\dot{L}(\omega, \mathbf{r}_m)$ - апертурна функція довільної просторової обробки, $v_m(\omega, \mathbf{n}_s)$ - компоненти (14.7) розподільчого вектору АР. Підставляючи формули (15.51), (15.52) у вираз (14.3), маємо:

$$\mu_{out}^{AR}(\omega) = \nu(r_s) \cdot \chi(\omega) \cdot \mu_0(\omega), \quad (15.53)$$

де, як і вище, $\nu(r_s)$ - втрати на розповсюдження сигналу по трасі від джерела до умовного центру АР, r_s - нормована (відносно деякої еталонної дистанції) довжина цієї траси, $\chi(\omega)$ - коефіцієнт підсилення просторової обробки, що визначається формулою (14.10),

$$\mu_0(\omega) = \frac{g_s(\omega)}{g_w + g_z(\omega)} \quad (15.54)$$

– відношення СЩП сигналу біля джерела до СЩП шумів та завад на приймальній апертурі. В окремому випадку оптимальної просторової обробки з урахуванням співвідношення (14.11) та рівняння (13.25), параметр $\chi(\omega)$ визначається відповідно до виразу (14.13).

Неважко бачити, що параметр завадостійкості (15.49) подібний за структурою до виразу (15.10), яким визначається SNR по виходу обробки для наперед відомого сигналу. Отже, усі міркування щодо впливу на завадостійкість прийому параметрів просторової обробки, які було наведено у розділі 14, залишаються в силі й для шумових сигналів. Зокрема, для даного випадку можуть бути застосовані результати обчислень підсилення АР та оптимізації її параметрів. Але є й суттєві відмінності у порівнянні з розглянутими вище видами сигналів.

По-перше, відносна смуга шумових сигналів звичайно досить широка. Тому, на відміну від сигналів, розглянутих у §§14.2,14.3 у даному випадку не завжди можна нехтувати частотною залежністю відношення сигнал/завада (15.53) по виходу просторової обробки. По-друге, параметр (15.53) не визначає завадостійкість прийому в цілому: для цього треба ще обчислити параметри (15.49), (15.50). У свою чергу, ці показники завадостійкості залежать від параметру (15.53) досить складним та нелінійним чином. По-третє, на кінцевий результат може суттєво вплинути вибір частотної характеристики фільтру часової обробки $B(\omega)$. Розглянемо останнє питання докладніше.

У нашому аналізі вигляд фільтру часової обробки $B(\omega)$ та апертурна функція просторової обробки $\dot{L}(\omega, \mathbf{r})$ поки що не конкретизувалися. При виборі цих параметрів звичайно беруться до уваги не тільки міркування завадостійкості прийому, але й технічні аспекти, зокрема, складність практичної реалізації. З технічної точки зору найбільш складною з усіх підсистем приймального пристрою є саме просторова частина обробки. Тому на вибір апертурної функції $\dot{L}(\omega, \mathbf{r})$ існує значно більше технічних обмежень,

чим на вибір характеристики фільтру $B(\omega)$. Вказані міркування дають підстави для постановки задачі часткової оптимізації, а саме, вигляд функції $\dot{L}(\omega, \mathbf{r})$ вибирається деяким довільним чином, а оптимізації підлягає лише частотна характеристика $B(\omega)$ фільтру часової обробки. Вирішення такої задачі може бути здійснене тими ж методами, що вже використовувались у §13.3. Достатньо прийняти припущення про те, що прийом сигналу здійснюється одиночним датчиком, який є еквівалентним за своїми характеристиками усій АР, тобто, спектральні характеристики сигналів, шумів та завад по його виходу будуть визначатися співвідношеннями (15.51), (15.52). Діючи таким же чином, як й у §13.3, приходимо до наступного результату щодо вигляду квадрату модуля оптимальної частотної характеристики $B_{opt}(\omega)$:

$$|B_{opt}(\omega)|^2 = \frac{\mu_{out}^{AR}(\omega)}{g_0^A(\omega) \cdot (1 + \mu_{out}^{AR}(\omega))}. \quad (15.55)$$

Не важко бачити, що з урахуванням співвідношень (15.53) - (15.55) та рівняння (13.25) в окремому випадку, коли оптимізації підлягає й просторова обробка, цей вираз співпадає з формулою (13.43). Зазначимо також, що оптимізація часової обробки визначає саме модуль частотної характеристики фільтру, фазова ж характеристика може вибиратися довільним чином, наприклад, з міркувань зручності технічної реалізації. В цьому проявляється специфіка шумового сигналу, який є випадковим процесом. Тому всілякого роду фазові співвідношення на етапі часової обробки шумового сигналу значення не мають. Ця обставина також показує принципову різницю поміж часовою та просторовою обробками. Дійсно, у розділі 2 вже відзначалося, що властивості шумового сигналу, як функції просторових координат не є випадковими. Вони принципово не відрізняються від аналогічних властивостей сигналів інших типів. Тому побудова просторової обробки виконується єдиним чином, який не залежить від специфіки сигналу, як функції часової координати. Не слід тільки забувати, що таке розділення

принципово можливе лише в рамках квазістаціонарного наближення для компактних просторових апертур.

Після підстановки оптимальної частотної характеристики (15.55) у вирази (15.49) та (15.50) одержуємо наступний результат:

$$q_F^2 = \frac{T}{4\pi} \left(\int_{-\infty}^{\infty} \frac{(\mu_{out}^A(\omega))^2}{1 + \mu_{out}^A(\omega)} d\omega \right)^2 \left(\int_{-\infty}^{\infty} (\mu_{out}^A(\omega))^2 d\omega \right)^{-1}; \quad (15.56)$$

$$\gamma_F^2 = \left(\int_{-\infty}^{\infty} \frac{(\mu_{out}^A(\omega))^2}{(1 + \mu_{out}^A(\omega))^2} d\omega \right) \left(\int_{-\infty}^{\infty} (\mu_{out}^A(\omega))^2 d\omega \right)^{-1} \quad (15.57)$$

Не важко показати, що ефективність приймання буде тим краща, чим більше параметр $\mu_{out}^{AR}(\omega)$. Отже, оптимізація часової обробки дозволила звести задачу аналізу ефективності прийому шумового сигналу до єдиного параметру (15.53), який обчислюється по виходу просторової обробки. Таким чином, параметр $\mu_{out}^{AR}(\omega)$ фактично є універсальним показником ефективності прийому для сигналів будь-якого виду. При цьому тип сигналу впливає лише на функціональну залежність поміж параметром $\mu_{out}^{AR}(\omega)$ та ймовірністю правильного виявлення сигналу. В даному випадку ця залежність визначається формулами (15.48), (15.56), (15.57).

Зазначимо також, що для шумових сигналів визначальний вплив на ефективність прийому окрім параметру (15.53) має також добуток ефективної смуги сигналу на його тривалість $\Delta f_s \cdot T_s$. Оскільки цей добуток є, як правило, дуже великим, то порогові значення параметру $\mu_{out}^{AR}(\omega)$, при яких забезпечується надійне виявлення сигналу, є досить малими. Таким чином, у практично важливих випадках виявлення шумових сигналів часто можна вважати, що $\mu_{out}^{AR}(\omega) \ll 1$. Ця обставина дозволяє суттєво спростити наведені вище формули. Зокрема, з виразів (15.56), (15.57) в даному разі витікає, що параметр $\gamma_F^2 \approx 1$, а SNR q_F^2 можна наближено обчислювати за формулою

$$q_F^2 \cong \frac{T}{4\pi} \int_{-\infty}^{\infty} (\mu_{out}^{AR}(\omega))^2 d\omega. \quad (15.58)$$

Отже, замість формули (15.48) обчислення ймовірності правильного виявлення сигналу в цьому випадку може наближено виконуватися за формулою

$$P_{п.в.} \cong \Phi \left\{ \sqrt{q_F^2} - \Phi^{-1}(1 - \alpha_0) \right\}. \quad (15.59)$$

Нагадаємо лише, що наближені формули (15.58), (15.59) можуть бути застосовані лише за умови, що частотна характеристика фільтру часового оброблення обрана відповідно до виразу (15.55).

15.5. Приклади розрахунків ефективності виявлення сигналів.

У цьому параграфі ми розглянемо чисельні ілюстрації, які демонструють методику та результати розрахунків характеристик ефективності виявлення сигналів для деяких прикладних завдань. Почнемо з так званих **робочих характеристик приймача**, які прийнято позначати ROC (від англ. Receiver Operating Characteristic). Вперше ці характеристики (їх також називають ROC-кривими) були запропоновані та практично застосовані ще в роки Другої Світової війни для аналізу ефективності систем радіо та гідролокації.

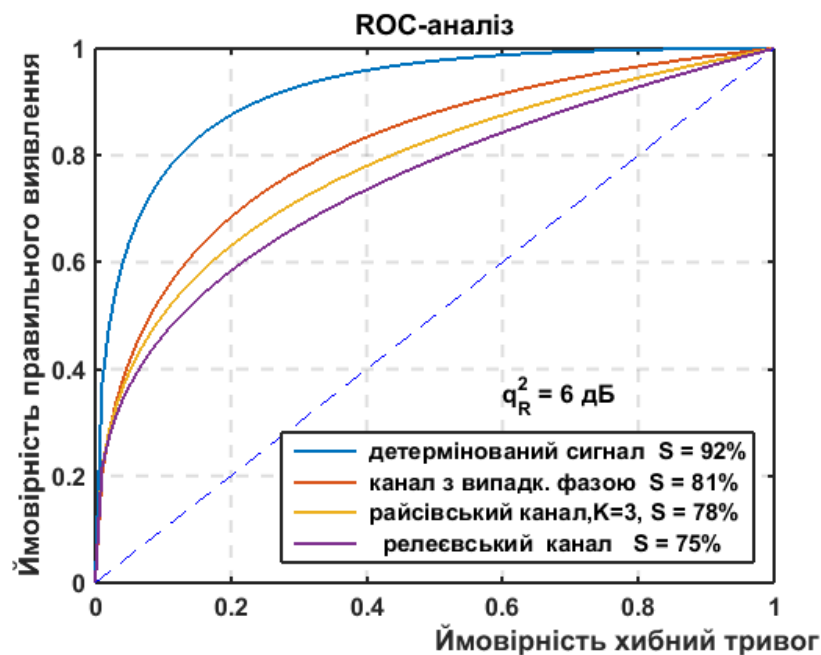


Рис.15.1. Робочі характеристики приймача для 4-х моделей каналів

Отже, кожна ROC-крива показує залежність ймовірності правильного виявлення сигналу $P_{\text{п.в.}}$ від ймовірності хибних тривог $P_{\text{х.т.}}$ при інших заданих умовах, зокрема, значенні SNR по виходу приймального пристрою q_R^2 . Відповідний приклад показано на рис.15.1.

На цьому рисунку представлено ROC-криві, які відповідають чотирьом моделям сигналів та каналів: сигналу з точно відомими характеристиками, сигналу з випадковою фазою, релеєвському та райсівському каналам. SNR по виходу лінійної частини обробки сигналів q_R^2 в усіх чотирьох випадках було задано на рівні 6 дБ. Для райсівського каналу значення К-фактора обрано на рівні $K_{ch} = 3$ (~4,8 дБ), що відповідає параметру когерентності каналу $\gamma_{coh} = 0,75$. Розрахунок виконано в обчислювальному середовищі Matlab, яке містить весь необхідний інструментарій для проведення такого числового аналізу.

Зокрема, для моделі детермінованого сигналу характеристика ROC визначається формулою (15.7). У цій формулі фігурує функція Лапласа, яка безпосередньо у середовищі Matlab не реалізована. Але цю складність легко обійти, якщо скористатися співвідношеннями (10.17) та замінити функцію Лапласа інтегралом ймовірності помилок (функцією Крампа):

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-y^2) dy. \quad (15.60)$$

В результаті формула (15.7) приводиться до вигляду:

$$P_{\text{п.в.}} = \frac{1}{2} \left(1 + \text{erf} \left(\sqrt{q_R^2/2} - \text{erf}^{-1}(1 - 2 \cdot \alpha_0) \right) \right), \quad (15.61)$$

де $z = \text{erf}^{-1}(y)$ - функція, зворотна до (15.60). Обидві ці функції (пряма та зворотна) ефективно реалізовані у середовищі Matlab, так що складнощів з обчисленнями за формулою (15.61) не виникає.

Для моделі каналу з випадковою фазою сигналу вихідною є формула (15.32), а райсівського каналу – формула (15.43). В обох цих формулах

фігурує Q-функція Маркума (15.33). Вказана функція безпосередньо реалізована у Matlab як утиліта *marcumq*. Нарешті, для моделі релеєвського каналу відповідна ROC-крива задається формулою (15.41), яка обчислюється в елементарних функціях.

Тепер розглянемо рис.15.1 детальніше. Перш за все зазначимо, що при застосуванні ROC-аналізу кращою прийнято вважати ту робочу характеристику, яка містить під собою відносно більшу частину загальної площі квадранту. Зокрема, якщо у деякому гіпотетичному випадку ROC-крива співпадала би з бісектрисою квадранту, яка показана пунктиром на рис.15.1, то відносна площа склала би 50%. Такий показник означав би, що правило рішення є статистично нікчемним та фактично еквівалентно киданню монетки. Отже, позитивним вважається лише такий результат, коли відносна площа⁶ переважає 50%.

Для даного прикладу відповідні частини⁷ площі під ROC-кривими показані на тому ж рис.15.1. Із рисунку видно, що в нашому прикладі для всіх чотирьох моделей каналу умова перевищення 50% відносної площини виконується, але з неоднаковою якістю. Зокрема, найвища ефективність виявлення сигналів за ROC-критерієм має місце для моделі детермінованого сигналу (відносна площа $S=92\%$), їй поступається модель сигналу з випадковою фазою ($S=81\%$), далі слідує модель райсівського каналу ($S=78\%$), а найгірший результат одержано для моделі релеєвського каналу ($S=75\%$). Вказані результати виглядають доволі переконливо, але для локаційних задач користі від них не дуже багато. Річ у тім, що локаційні системи працюють у досить вузькому діапазоні рівнів хибних тривог, який звичайно складає $\alpha_0 = 10^{-2} - 10^{-7}$. Тому поведінка ROC-кривих за межами цього діапазону практичного значення, як правило, не має. Більш суттєвим для локаційних застосувань є аналіз впливу енергетичних показників задачі,

⁶ Якщо ж відносна площа виявиться меншою 50%, то це буде означати, що правило рішення треба змінити на протилежне.

⁷ Для обчислення цих частин можна використовувати функцію *trapz*

моделі каналу передачі, тощо. Але для вирішення цих завдань більш зручними є деякі інші характеристики, до розгляду яких ми перейдемо нижче. Що ж стосується ROC-аналізу, то в наш час він розглядається як один з основних методів аналізу ефективності правил класифікації. При цьому цей метод досить несподівано знайшов широке застосування⁸ в галузях, які доволі далекі від локаційних застосувань. Зокрема, до них відносяться завдання медичної діагностики, соціології, контролю якості продукції, машинного навчання та деякі інші.

Повертаючись до локаційних завдань зазначимо, що для аналізу їх ефективності найчастіше використовують так звані **характеристики виявлення** (в англійській літературі – *Probability of Detection*) які показують ймовірність правильного виявлення сигналів $P_{\text{п.в.}}$ в залежності від SNR по виходу приймального пристрою при заданому рівні хибних тривог α_0 . Для розрахунку вказаних характеристик використовуються ті ж самі формули, що для ROC-аналізу. Тільки тепер змінним параметром є відношення сигнал/завада q_R^2 , а параметр α_0 залишається фіксованим. Приклади таких характеристик представлені на рис.15.2. Тут же у розшифровці маркування кривих вказані моделі каналів та значення деяких параметрів.

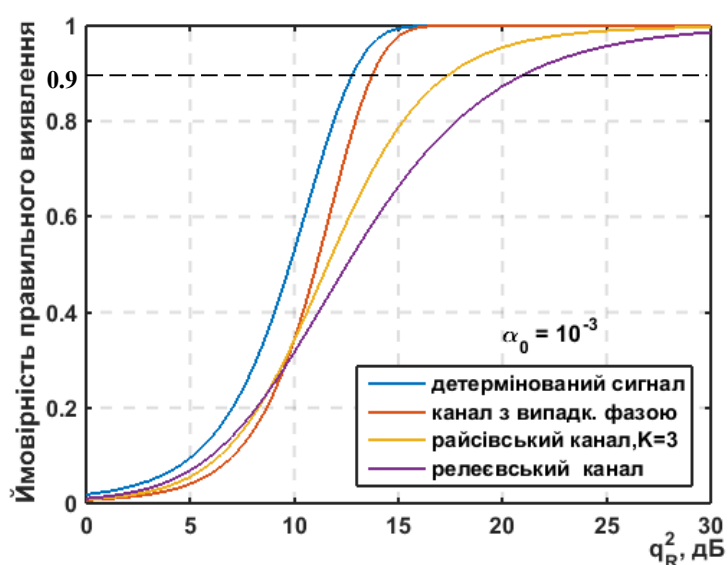


Рис.15.2. Характеристики виявлення сигналів

⁸ Варто зазначити, що при цьому назва методу не змінилася, тобто, «локаційні» корені збереглися!

Як бачимо з рисунка, головні закономірності у цьому разі такі ж самі, як й на рис.15.1. Але характеристики виявлення дозволяють оцінити зміни втрат/виграшів вже у діапазоні значень параметру q_R^2 , що зручніше для практичних застосувань. Зокрема, на кожному заданому рівні ймовірності $P_{п.в.}$ можна оцінити виграш або програш ефективності того чи іншого алгоритму приймання у термінах відповідних значень параметру q_R^2 . Наприклад, для рівня $P_{п.в.} = 0.9$ на вказаному графіку можна знайти порогові значення SNR, які вже були показані у таблиці 15.1.

У той же час обчислення порогових рівнів параметру q_R^2 доцільно все ж таки проводити окремо за відповідними формулами або обчислювальними алгоритмами, які базуються на безпосередньому оберненні аналітичних співвідношень, наведених вище. Зокрема, для моделі детермінованого сигналу можна використати формулу (15.61), з якої витікає наступний вираз для порогового значення відношення сигнал/завада:

$$q_{R_{nop}}^2 = 2 \cdot \left(\operatorname{erf}^{-1}(2D_0 - 1) + \operatorname{erf}^{-1}(1 - 2 \cdot \alpha_0) \right), \quad (15.62)$$

де, нагадаємо, D_0 - заданий рівень ймовірності правильного виявлення сигналів. Також в аналітичному вигляді, але вже за формулою (15.42), вирішується аналогічна задача й для моделі релеєвського каналу. Результати розрахунків за формулами (15.42) та (15.62) представлені на рис.15.3. Тут же дана розшифровка умов та параметрів обчислень. Неважко бачити, що деякі точки цих графіків дублюють дані таблиці 15.1. Що ж стосується більш складних моделей, для яких аналітичні розв'язки знайти не вдається, то для них можна застосувати процедури чисельного вирішення рівняння

$$f(q_R^2) = P_{п.в.}(q_R^2) - D_0 = 0 \quad (15.63)$$

Середовище Matlab містить для цього достатньо можливостей, зокрема, утиліту *fzero*. Саме таким чином побудовано представлені на рис.15.3 графіки, які відповідають каналу з випадковою фазою та райсівському каналу. Для цього були використані відповідно формули (15.32) та (15.43).

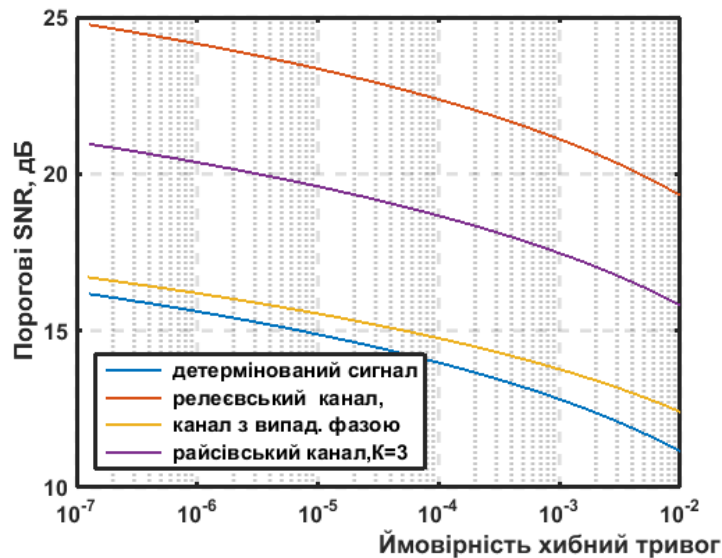


Рис.15.3 Порогові SNR для чотирьох моделей каналу

На останок розділу розглянемо ще один цікавий випадок, в якому досліджується залежність завадостійкості приймання у райсівському каналі в залежності від параметру його когерентності $\gamma_{coh} \in (0,1)$. Результати аналізу представлені на рис.15.4, де показані порогові SNR для двох алгоритмів виявлення сигналів, а саме, некогерентного правила (15.21) та оптимального правила (11.34), яке передбачає врахування як когерентної, так й некогерентної складової сигналу на вході приймального пристрою.

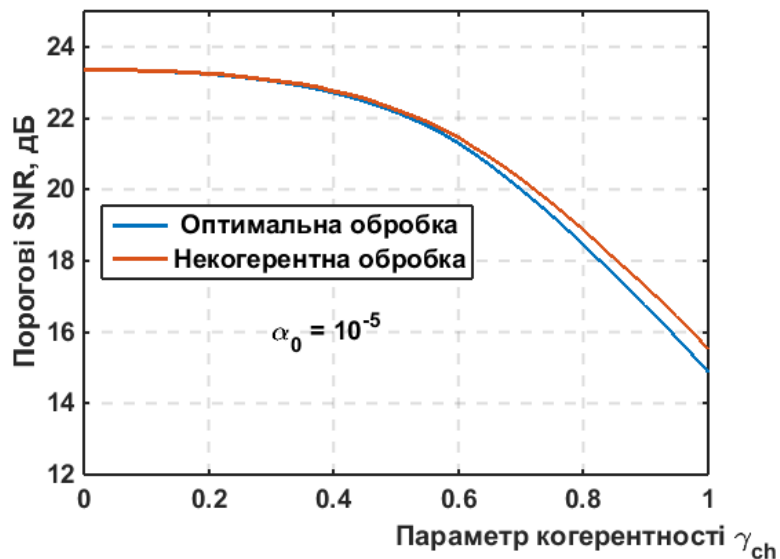


Рис.14.4. Порогові SNR в залежності від параметру когерентності райсівського каналу.

Як видно з даного рисунку, виграш за рахунок оптимізації приймання сигналів при обраних значеннях параметрів складає не більше 0.5дБ, причому, досягається цей виграш лише при значеннях параметру когерентності близьких до одиниці. В діапазоні ж значень $\gamma_{ch} \leq 0.6$ ефективність обох правил виявлення практично однакова.

Контрольні запитання та завдання до розділу 15.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. Базові показники ефективності виявлення сигналів. Як вони визначаються?
2. Самостійно одержати вираз (15.5).
3. Що таке відношення сигнал/завада по виходу приймального пристрою?
4. Що таке умови квазістаціонарного наближення для компактних апертур? Чому при цих умовах доцільно використовувати спектральні форми представлення параметрів завадостійкості?
5. Пояснити фізичну суть моделі поля сигналу (15.11) та її параметри.
6. Чому вибір спектральної функції оброблення у вигляді (15.12) дає можливість розділення загальної обробки сигналу на часову та просторову?
7. Самостійно одержати вирази (15.15), (15.16) та пояснити зміст параметрів, що до них входять.
8. Що таке часова обробка, узгоджена із сигналом? Чому вона так називається?
9. Самостійно одержати вираз (15.18) та пояснити зміст параметрів, що до нього входять. Як та чому відрізняються SNR по виходу та входу приймального пристрою?
10. Що таке «математичне» та «фізичне» визначення СЩП шумів та завад? Який поміж ними зв'язок?
11. Коли для виконання розрахунків завадостійкості можна застосовувати вираз (15.18), а коли він є неприйнятним?
12. Пояснити вигляд моделі поля сигналу (15.20) та відповідний алгоритм його оброблення.
13. Довести що статистичні характеристики квадратурних компонент по виходу лінійної обробки (15.21) співпадають. Довести співвідношення (15.35).

Вказівка: Скористатися загальними властивостями аналітичних сигналів (§1.3).

14. Самостійно одержати формулу (15.28).
15. Довести тотожність виразів (15.34) та (15.8).
16. Що таке порогове відношення сигнал/завада та для чого воно використовується?
17. Пояснити модель каналу із випадковими амплітудою та фазою сигналу та відповідні алгоритми обробки.
18. Пояснити формули (15.39), (15.40) та зміст параметрів, що до них входять.
19. Самостійно повторити викладки, що приводять до виразів (15.41) та (15.43).
20. В чому полягає особливість оброблення шумових сигналів?
21. Чому розподіл статистики (15.44) можна вважати гавсівським?
22. Самостійно одержати вирази (15.46) - (15.50).
23. Самостійно одержати вирази (15.55) – (15.57).
24. Що таке ROC-аналіз? Де та для чого він застосовується?
25. Самостійно повторити у середовищі Матлаб розрахунок ROC-кривих для $q_R^2 = 9$ дБ та параметру когерентності райсівського каналу $\gamma_{ch} = 4/5$.
26. Що таке характеристики виявлення сигналів? Як їх обчислити? Виконати розрахунок, аналогічний рис.14.2, для рівня хибних тривог 10^{-5} .
27. Як можна обчислити порогові SNR для різних моделей каналів? Шляхом розрахунків самостійно перевірити дані таблиці 15.1.

Розділ 16. Вимірювання параметрів сигналів.

Будь-яке поле сигналу (електромагнітне, акустичне, оптичне, тощо), є, як правило, джерелом важливої інформації для вирішення багатьох прикладних завдань. Наприклад, функціонування локаційних систем різноманітного призначення [8, 20, 23] базується на вимірюванні напрямків приходу хвиль у просторі, значень затримки сигналів у часі, величин доплерівського зміщення їх частоти, тощо. Ті ж параметри можуть бути корисні й для вирішення завдань медичної чи технічної діагностики. Вимірювання цих та багатьох інших параметрів може бути виконано на базі методів теорії статистичного оцінювання, які розглядалися у розділі 10. Наповнення вказаних методів фізичним змістом та конкретизація алгоритмів вимірювання тих чи інших параметрів складає великий розділ сучасної теорії та практики статистичної обробки сигналів. Докладний розгляд усіх цих алгоритмів у рамках одного розділу навряд чи можливий. Тому зосередимося на огляді базових ідей та тих питаннях, що мають безпосереднє відношення до тематичної спрямованості даного Посібника. Будемо вважати, що параметри, які підлягають вимірюванню, є квазістатичними, тобто, незмінними в межах інтервалів та областей спостереження. Це не означає, що такі параметри взагалі не можуть змінюватися. Але характерні масштаби цих змін мають суттєво переважати розміри області спостереження сигналу в просторі та часі. В іншому випадку завдання їх оцінювання стає предметом іншого розділу статистичної науки, а саме, теорії фільтрації [4], розгляд якої поки не планується.

16.1. Загальні положення.

Відповідно до теорії, що розглядались у §10.4, оптимальні процедури статистичного оцінювання визначаються, взагалі кажучи, прийнятим критерієм ефективності та доступним об'ємом апріорної інформації. Кількість та різноманітність таких процедур дуже великі. Але в їх основі лежить, як правило, вирішення оптимізаційної задачі наступного вигляду:

$$\hat{\alpha}_F = \arg \left\{ \max_{\alpha} F(\mathbf{u}/\alpha) \right\}. \quad (16.1)$$

Тут $F(\mathbf{u}/\alpha)$ - деякий функціонал (тестова статистика задачі) вектору спостережень $\mathbf{u}$, що відповідає певній гіпотезі відносно вектору невідомих параметрів α , $\hat{\alpha}_F$ - значення оцінки, яке знаходиться як точка максимуму¹ статистики $F(\mathbf{u}/\alpha)$ на множині усіх допустимих значень параметрів α .

В технічних застосуваннях параметри α називають звичайно параметрами налаштування (настроювання) системи приймання сигналів. Отже, завдання вимірювань певних інформаційних параметрів полягає у виборі таких параметрів настроювання приймальної системи, які б співпадали в межах заданої похибки з фактичними параметрами сигналу. При цьому «індикатором» узгодженості параметрів настроювання та фактичних параметрів сигналу якраз і є тестова статистика $F(\mathbf{u}/\alpha)$. Відмінність поміж різноманітними процедурами оцінювання полягає, головним чином, у виборі вигляду статистики $F(\mathbf{u}/\alpha)$ та методу вирішення оптимізаційної задачі (16.1).

При виборі вигляду статистики $F(\mathbf{u}/\alpha)$ звичайно прагнуть до максимально оптимальних рішень. Найбільш загальним класом таких рішень, як вже зазначалося у §10.4, є баєсові процедури, які мінімізують середній ризик (10.25). Зокрема, до баєсових процедур, як окремий випадок, належить метод максимізації апостеріорного розподілу $p(\alpha/\mathbf{u})$ (метод MAP – від англ. *Maximum A posteriori Probability*). При цьому замість розподілу $p(\alpha/\mathbf{u})$ може бути використана будь-яка монотонно зростаюча від нього функція, зокрема, логарифм. Отже, відповідно до методу MAP в якості оптимальної статистики звичайно вибирають функціонал обробки наступного вигляду:

$$F_{MAP}(\mathbf{u}/\alpha) = \ln p(\alpha/\mathbf{u}).$$

¹ В залежності від вибору тестової статистики можлива також постановка задачі з пошуку її мінімуму по параметрам α , але принципово це нічого не змінює.

У §10.4 також було зазначено, що у випадку надійних вимірювань більшість оптимальних процедур зводиться до методу максимальної правдоподібності ML (від англ. *Maximum Likelihood*). Згідно методу ML оцінювання деякої скінченої сукупності параметрів α зводиться до пошуку максимуму функціоналу правдоподібності спостережень $p(\mathbf{u}/\alpha)$ (або будь-якої монотонної функції від нього) по вектору α , тобто, у цьому випадку тестова статистика має вигляд

$$F_{МП}(\mathbf{u}/\alpha) = \ln p(\mathbf{u}/\alpha).$$

Зауважимо, що розподіл $p(\mathbf{u}/\alpha)$ існує лише тоді, коли спостереження $\mathbf{u}$ є вектором у деякому просторі скінченої розмірності. На фізичному рівні це означає дискретність та обмеженість часової і просторової апертур. Перехід до нескінченновимірному випадку пов'язаний з деякими математичними складнощами [3], про які вже згадувалось у розділі 11. Але у практичному плані такий перехід у багатьох випадках можна виконати значно простіше, якщо для вирішення задачі оцінювання застосовувати не функцію правдоподібності $p(\mathbf{u}/\alpha)$, а функціонал відношення правдоподібності $\lambda(\mathbf{u}/\alpha)$ чи будь-яку монотонно зростаючу функцію від нього, наприклад, логарифм. Отже, тестова статистика для методу ML звичайно має вигляд:

$$F_{ВП}(\mathbf{u}/\alpha) = \ln \lambda(\mathbf{u}/\alpha),$$

$$\text{де} \quad \lambda(\mathbf{u}/\alpha) = \frac{p(\mathbf{u}/\alpha)}{p_0(\mathbf{u})}, \quad (16.2)$$

$p_0(\mathbf{u}) = p(\mathbf{u}/\alpha_0)$ - розподіл спостережень, що відповідає деякій фіксованій гіпотезі щодо значень параметрів α_0 . Зокрема, досить часто в якості нормуючої функції у виразі (16.2) використовують просто розподіл шумів та завад, тобто, розподіл спостережень в ситуації відсутності корисного сигналу $p_0(\mathbf{u}) = p(\mathbf{u}/\theta = 0)$. Але у будь-якому разі нормуюча функція відношення правдоподібності від вектору параметрів, що підлягають оцінюванню, залежати не повинна. При виконанні цієї умови оцінки

параметрів сигналу за критерієм максимуму функції правдоподібності $p(\mathbf{u}/\boldsymbol{\alpha})$ та максимуму функціоналу відношення правдоподібності будуть повністю еквівалентні. Перевагою ж подібного нормування є те, що хоча для нескінченновимірних спостережень розподіл $p(\mathbf{u}/\boldsymbol{\alpha})$ й не існує, але існує, як правило, функціонал (16.2), який у математичній літературі називають відношенням імовірнісних мір або похідною Радона-Нікодіма [6].

У подальшому розгляді ми зосередимося головним чином на процедурах оцінювання, що базуються на використанні статистики (16.2), яка є оптимальною за критерієм ML. Але при цьому також будуть розглядатися й деякі спрощення, пов'язані з певним відходом від строгої оптимальності. В цьому зв'язку зауважимо, що навіть й при неоптимальному виборі функціоналу $F(\mathbf{u}/\boldsymbol{\alpha})$ бажано, як правило, забезпечити виконання певних базових умов [6], а саме:

- відсутність зміщення оцінки $\hat{\boldsymbol{\alpha}}_F$;
- статистичну конзистентність (слушність);
- ефективність.

Нагадаємо, що відсутність умовного зміщення деякої оцінки $\hat{\boldsymbol{\alpha}}_F$ означає виконання рівності

$$E\{\hat{\boldsymbol{\alpha}}_F / \boldsymbol{\alpha}_0\} = \boldsymbol{\alpha}_0 \quad (16.3)$$

де $E\{\cdot / \boldsymbol{\alpha}_0\}$ - символ умовного статистичного усереднення у просторі спостережень при наявності сигналу з параметрами $\boldsymbol{\alpha}_0$. Рівність (16.3) означає, по суті, що метод оцінювання не має систематичних похибок. Що ж стосується статистичних похибок, то їх поведінка визначається іншою важливою умовою – статистичною конзистентністю оцінки. Суть цієї умови полягає у забезпеченні збіжності оцінки за деяким імовірнісним критерієм до точного значення параметрів при необмеженому зростанні певного параметру надійності вимірювань. Стосовно задач сигнальної обробки таким параметром найчастіше є відношення сигнал/завада по виходу обробки q^2 ,

яке у випадках гавсівського розподілу функціоналів обробки визначається виразом (15.6). Цей параметр, як було показано у розділі 15, залежить від тривалості інтервалу спостереження, розмірів просторової апертури та відношення сигнал/завада по входу системи обробки. Отже, при збільшенні будь-якого із цих факторів для статистично конзистентних оцінок повинен мати місце граничний перехід

$$\hat{\mathbf{a}}_F \xrightarrow{q^2 \rightarrow \infty} \mathbf{a}_0 .$$

При цьому збіжність розуміється не в класичному математичному сенсі, а за деяким імовірнісним критерієм [3, 6], одним з яких є середньо квадратична похибка вимірювань, тобто:

$$E \left\{ \left| \hat{\mathbf{a}}_F - \mathbf{a}_0 \right|^2 / \mathbf{a}_0 \right\} \xrightarrow{q^2 \rightarrow \infty} 0 \quad (16.4)$$

При скінченному параметрі q^2 нижня границя похибки оцінювання визначається нерівністю Рао-Крамера (10.51). Оцінки, для яких досягається ця нижня границя, називають ефективними. Зокрема, до ефективних, як уже було сказано, належать й оцінки за методом ML, якщо вони існують [17]. Відношення нижньої границі Рао-Крамера до фактичної середньо квадратичної похибки, тобто,

$$\varepsilon_F = \frac{E \left\{ \left| \hat{\mathbf{a}}_{ML} - \mathbf{a}_0 \right|^2 / \mathbf{a}_0 \right\}}{E \left\{ \left| \hat{\mathbf{a}}_F - \mathbf{a}_0 \right|^2 / \mathbf{a}_0 \right\}} ,$$

характеризує відносну ефективність деякої довільної оцінки [17]. За визначенням цей параметр заходиться в межах (0,1). Отже, при виборі структури функціоналу $F(\mathbf{u}/\mathbf{a})$ природно прагнути до того, щоб цей показник був якомога ближче до 1.

16.2. Методи вимірювань.

Тепер перейдемо до розгляду методів вирішення задачі (16.1). Взагалі кажучи, ця задача вписується у стандартні постановки теорії максимізації функцій [22]. Однак, з огляду на статистичний характер вимірювань певна специфіка у порівнянні із універсальними оптимізаційними методами все ж таки має місце. По-перше, слід взяти до уваги, що в нашому випадку цільова функція, тобто, функціонал обробки $F(\mathbf{u}/\boldsymbol{\alpha})$, залежить від випадкового вектору спостережень $\mathbf{u}$. Тому існування рішень задачі (16.1), збіжність наближених рішень, тощо, звичайно слід розуміти на основі імовірнісних критеріїв. По-друге, імовірнісні моделі відіграють помітну роль й у побудові методів її вирішення. Наприклад, широке застосування в таких задачах знаходять марковські моделі [2]. Специфічним для задач статистичного оцінювання є також алгоритм ЕМ [17] (від англ. *Expectation – Maximization*). По-третє, слід враховувати й технічні обмеження, як, наприклад, обчислювальні затрати, необхідність забезпечення взаємодії підсистеми вимірювання з іншими підсистемами, тощо. Нарешті, не останню роль відіграють й певні традиції, які склалися у теорії та практиці статистичних вимірювань.

Основні підходи до вирішення загальної задачі (16.1) базуються, як правило, на співвідношеннях, що визначають необхідні для досягнення екстремуму цільової функції умови:

$$\nabla_{\boldsymbol{\alpha}} F(\mathbf{u}/\boldsymbol{\alpha}) = \left[\frac{\partial F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \alpha_1}, \dots, \frac{\partial F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \alpha_M} \right]_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}}^T = 0. \quad (16.5)$$

Тут $\nabla_{\boldsymbol{\alpha}} F$ - градієнт функціоналу $F(\mathbf{u}/\boldsymbol{\alpha})$ по вектору $\boldsymbol{\alpha}$, тобто, вектор, що складається з часткових похідних цільового функціоналу по компонентах вектору $\boldsymbol{\alpha}$. Отже, задача (16.1) зводиться до вирішення відносно $\boldsymbol{\alpha}$ системи рівнянь (16.5), яка, зазначимо, часто є нелінійною по $\boldsymbol{\alpha}$. Достатньою ж умовою, що визначає точку максимуму є, як відомо [6], від'ємна визначеність

матриці других похідних (матриці Гессе або гессіану) цільової функції. При цьому слід враховувати, що цільова функція $F(\mathbf{u}/\boldsymbol{\alpha})$ може мати не один, а кілька екстремумів. Тому завдання полягає в пошуку саме глобального максимуму функціоналу $F(\mathbf{u}/\boldsymbol{\alpha})$, що іноді пов'язано зі значними труднощами [22].

Більшість відомих алгоритмів оцінювання відрізняються, по суті, методом вирішення системи рівнянь (16.5). Замкнене аналітичне рішення системи рівнянь (16.5) відомо лише для окремих випадків, коли можлива лінеаризація рівнянь по $\boldsymbol{\alpha}$. Але у загальному випадку таке рішення знайти або не вдається, або воно виявляється занадто складним для практичного застосування. Тому значного поширення набули різного роду наближені методи вирішення рівнянь (16.5), які, у свою чергу, можна розділити на певні групи, основними з яких є:

1. **Методи послідовних наближень.** Рівняння (16.5) вирішують шляхом деяких рекурсивних ітерацій

$$\hat{\boldsymbol{\alpha}}^{(n+1)} = f\left\{ \mathbf{u}, \hat{\boldsymbol{\alpha}}^{(n)} \right\}, \quad (16.6)$$

на кожній з яких знаходиться чергове наближення вектору оцінки $\hat{\boldsymbol{\alpha}}^{(n+1)}$. При цьому критерієм збіжності може бути виконання на кожній ітерації наступної нерівності щодо цільового функціоналу:

$$F(\mathbf{u}/\hat{\boldsymbol{\alpha}}^{(n+1)}) > F(\mathbf{u}/\hat{\boldsymbol{\alpha}}^{(n)}).$$

Один з поширених підходів до побудови таких процедур базується на використанні квадратичної апроксимації цільової функції. Припустимо, що функціонал $F(\mathbf{u}/\boldsymbol{\alpha})$ має похідні по параметрам $\boldsymbol{\alpha}$. Тоді цей функціонал можна представити ступеневим рядом в околі деякого опорного значення вектору параметрів $\boldsymbol{\alpha}_{ref}$:

$$\begin{aligned}
F(\mathbf{u}/\boldsymbol{\alpha}) = & F(\mathbf{u}/\boldsymbol{\alpha}_{ref}) + (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{ref})^T \cdot \left. \frac{\partial F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_{ref}} + \\
& + \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{ref})^T \cdot \left. \frac{\partial^2 F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_{ref}} \cdot (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{ref}) + \dots
\end{aligned} \quad (16.7)$$

Якщо опорний вектор $\boldsymbol{\alpha}_{ref}$ достатньо близький до фактичного значення параметрів $\boldsymbol{\alpha}_0$, то в цьому розкладі звичайно обмежуються членами зі ступенями не вище другого. Після підстановки такої апроксимації до співвідношення (16.5) маємо наближене рівняння

$$\left. \frac{\partial F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_{ref}} + \left. \frac{\partial^2 F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^T} \right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_{ref}} \cdot (\boldsymbol{\alpha} - \boldsymbol{\alpha}_{ref}) \cong 0.$$

Звідси витікає ітеративний алгоритм наступного вигляду:

$$\hat{\boldsymbol{\alpha}}^{(n+1)} = \hat{\boldsymbol{\alpha}}^{(n)} - \boldsymbol{\Gamma}^{-1}(\hat{\boldsymbol{\alpha}}^{(n)}) \cdot \nabla_{\boldsymbol{\alpha}} F(\mathbf{u}/\hat{\boldsymbol{\alpha}}^{(n)}) \quad (16.8)$$

де $\boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})$ - ітеративна матриця, зворотна до матриці других похідних цільової функції, тобто,

$$[\boldsymbol{\Gamma}(\boldsymbol{\alpha})]_{ik} = \frac{\partial^2 F(\mathbf{u}/\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_k}.$$

Зокрема, у випадку оцінювання скалярного параметру α маємо:

$$\hat{\alpha}^{(n+1)} = \hat{\alpha}^{(n)} - \left(\left. \frac{d^2 F\{\mathbf{u}/\alpha\}}{d\alpha^2} \right|_{\alpha=\hat{\alpha}^{(n)}} \right)^{-1} \cdot \left. \frac{dF\{\mathbf{u}/\alpha\}}{d\alpha} \right|_{\alpha=\hat{\alpha}^{(n)}}. \quad (16.9)$$

Неважко бачити, що алгоритм (16.8) та його окремий випадок (16.9) є, по суті, версіями відомого алгоритму Ньютона для вирішення нелінійних рівнянь [9]. Загалом, вказані алгоритми відносяться до великого класу градієнтних методів найшвидшого спуску [17], які відрізняються поміж собою, головним чином, побудовою контуру зворотного зв'язку. Зокрема, замість ітеративної матриці $\boldsymbol{\Gamma}^{-1}(\boldsymbol{\alpha})$ в алгоритмі, подібному до (16.8), може використовуватись деяка скалярна функція $\gamma(\boldsymbol{\alpha})$.

Методи найскорішого спуску одержали значний розвиток у сучасній обчислювальній математиці, теорії оптимізації, теорії автоматичних та керуючих систем, тощо. Тому весь той багатий арсенал обчислювальних алгоритмів, що накопичено у вказаних галузях науки, може бути застосований (та й з успіхом застосовується) для вирішення вимірювальних задач. Але досить часто методи даного класу розглядаються в контексті дещо іншої задачі, а саме, лінійної та нелінійної фільтрації [3]. Річ у тім, що параметри $\mathbf{a}$ не є, взагалі кажучи, незмінними у часі. Тому досить часто номер кроку ітерацій n пов'язують з певним моментом часу t_n , тобто, в процесі ітерацій стає змінним й вектор спостережень $\mathbf{u} = \mathbf{u}^{(n)}$. Тоді така ітеративна процедура фактично перетворюється у фільтр, який дозволяє відстежувати зміни параметрів у часі. Особливо широкого застосування подібні алгоритми набули для задач стеження за поточною фазою сигналу (так звані системи ФАПЧ), які переважно застосовуються в системах приймання радіосигналів, сигналів зв'язку, траєкторних вимірюваннях [20], тощо.

2. Методи перебору. Ця група методів полягає в пошуку максимуму функціоналу $F(\mathbf{u}/\mathbf{a})$ шляхом прямого перебору значень вектору параметрів, який охоплює апріорну множину його існування $\mathbf{a} \in \mathbf{A}$. У технічній літературі такий перебір називають звичайно **скануванням** по тій чи іншій сукупності параметрів [12, 24]. Сканування може бути послідовним або паралельним. В останньому випадку фактично утворюються незалежні канали обробки, кожний з яких узгоджується під певні значення параметрів сигналу. Зокрема, паралельне сканування може застосовуватися для огляду можливих напрямків приходу фронту хвилі сигналу в просторі або для визначення доплерівського зміщення частоти сигналу. Послідовне ж сканування використовується, наприклад, для вимірювання часу приходу сигналу.

Методи перебору визначає надійність та хороша швидкодія. Однак, побудова багатоканальних систем пов'язана зі значними технічними затратами. Ці затрати, вочевидь, залежать від вибору кроку зміни параметрів: чим крок менше, тим більші затрати, але при цьому зростає й точність вимірювань.

3. *Методи інтерполяції*. Ця група методів базується на використанні певної апроксимації залежності функціоналу обробки $F(\mathbf{u}/\boldsymbol{\alpha})$ від вектору параметрів $\boldsymbol{\alpha}$, наприклад, лінійної або квадратичної. Як окремий випадок квадратичної апроксимації може розглядатися й наведений вище ітеративний алгоритм Ньютона. Але інтерполяційні методи відрізняються тим, що для обчислення оцінки параметрів $\hat{\boldsymbol{\alpha}}$ використовуються значення тестової статистики не для одного, а для кількох опорних значень вектору параметрів $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_K$. Таким чином, оцінка утворюється на основі деякої функціональної залежності

$$\hat{\boldsymbol{\alpha}} = f\{F(\mathbf{u}/\boldsymbol{\alpha}_1), \dots, F(\mathbf{u}/\boldsymbol{\alpha}_K); \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_K\} \quad (16.10)$$

При цьому вважається, що фактичне значення невідомих параметрів $\boldsymbol{\alpha}_0$ достовірно знаходиться в інтервалі², який перекривається набором опорних векторів $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_K$. Подібні процедури уточнення оцінок у технічних застосуваннях часто називають *дискримінаторами* [11, 24] з додаванням найменування параметру, по якому виконується вимірювання: частотний дискримінатор, кутовий дискримінатор, часовий дискримінатор, тощо. При цьому кількість опорних каналів, що використовують дискримінатори, звичайно досить невелика (типові значення $K = 2 \div 5$).

Для побудови дискримінаторів часто використовують методи скінчених різниць [11]. Розглянемо деякі приклади. При цьому обмежимося випадком вимірювань скалярного параметру, хоча більшість цих методів неважко розповсюдити й на випадок оцінювання векторних параметрів. Отже, скористаємося формулами чисельного диференціювання функцій [5]:

² При цьому, як правило, фактичне значення параметрів $\boldsymbol{\alpha}_0$ знаходиться поблизу центру цього інтервалу.

$$\left. \frac{\partial f}{\partial \alpha} \right|_{\alpha=\alpha_0} \approx (f(\alpha_0 + \Delta/2) - f(\alpha_0 - \Delta/2)) / \Delta ,$$

$$\left. \frac{\partial^2 f}{\partial \alpha^2} \right|_{\alpha=\alpha_0} \approx (f(\alpha_0 + \Delta) + f(\alpha_0 - \Delta) - 2f(\alpha_0)) / \Delta^2 ,$$

де Δ - достатньо малий приріст параметрів. Після підстановки цих формул у співвідношення (16.9) та елементарних перетворень можна прийти до наступного алгоритму уточнення оцінки:

$$\hat{\alpha} = \alpha_{ref} - \Delta \cdot \frac{F(\mathbf{u} / \alpha_{ref} + \Delta/2) - F(\mathbf{u} / \alpha_{ref} - \Delta/2)}{F(\mathbf{u} / \alpha_{ref} + \Delta) + F(\mathbf{u} / (\alpha_{ref} - \Delta)) - 2F(\mathbf{u} / \alpha_{ref})} , \quad (16.11)$$

де α_{ref} - центр попередньо локалізованої області існування фактичного значення параметру α . Як неважко бачити, для реалізації вказаного алгоритму необхідно мати 5 опорних значень функціоналу обробки для значень параметрів α_{ref} , $\alpha_{ref} \pm \Delta/2$, $\alpha_{ref} \pm \Delta$. Більш простим для практичної реалізації є так званий сумарно-різницевий метод [12], при якому використовуються лише два канали обробки. Назвемо умовно ці канали "лівим" та "правим", хоча їх фізичне розташування по параметру настроювання α може цим назвам й не відповідати. Припустимо, що "лівий" канал налаштовано для прийому сигналу з параметром $\alpha = \alpha_L$, а "правий" – для прийому сигналу з параметром $\alpha = \alpha_R$. Якщо рознесення каналів відносно невелике, то можна скористатися параболічною апроксимацією функціоналів обробки (16.7) в околі фактичного значення параметру α_0 . Беручи до уваги ту обставину, що за умовою отримання оцінки перша похідна цільової функції в точці $\alpha = \alpha_0$ обертається в нуль, ця апроксимація приймає вигляд:

$$F_L = F(\mathbf{u} / \alpha_L) \approx F_0 + \frac{(\alpha_L - \alpha_0)^2}{2} F_0'' , \quad F_R = F(\mathbf{u} / \alpha_R) \approx F_0 + \frac{(\alpha_R - \alpha_0)^2}{2} F_0'' ,$$

де F_0 , F_0'' - відповідно значення функціоналу обробки та його друга похідна при $\alpha = \alpha_0$. Далі утворимо різницю величин F_L та F_R :

$$F_R - F_L \approx \frac{F_0''}{2} [(\alpha_R - \alpha_0)^2 - (\alpha_L - \alpha_0)^2] = F_0'' \cdot \Delta \cdot (\alpha_C - \alpha_0),$$

де $\alpha_C = (\alpha_R + \alpha_L)/2$ - середнє по каналам значення параметру настроювання, $\Delta = \alpha_R - \alpha_L$ - рознесення каналів. Тепер утворимо суму цих же величин:

$$F_R + F_L \approx 2F_0 + \frac{F_0''}{2} [(\alpha_R - \alpha_0)^2 + (\alpha_L - \alpha_0)^2].$$

Сумарно-різницевий метод використовується, як правило, для таких схем вимірювань, коли фактичне значення параметру α_0 є достатньо близьким до середнього значення параметрів настроювання каналів $\alpha_C = (\alpha_L + \alpha_R)/2$. Тому в останньому виразі можна без значної похибки замінити невідоме фактичне значення параметру α_0 на фіксоване значення α_C . В результаті одержуємо:

$$F_R + F_L \approx 2F_0 + \frac{F_0''}{4} \cdot \Delta^2.$$

Тепер сформуємо частку від ділення різниці величин F_L та F_R на їх суму:

$$\frac{F_R - F_L}{F_R + F_L} \approx \frac{F_0'' \cdot \Delta}{2F_0 + F_0'' \cdot \Delta^2 / 4} \cdot (\alpha_C - \alpha_0).$$

Отже, з урахуванням вказаних припущень неважко бачити, що цей вираз приблизно лінійно залежить від розладу поміж фактичним значенням параметру α_0 та середнім значенням параметрів настроювання α_C . Звідси витікає наступний алгоритм уточнення оцінки:

$$\hat{\alpha}_0 = \alpha_C - \gamma(\alpha_C) \cdot \frac{F_R - F_L}{F_R + F_L}, \quad (16.12)$$

де $\gamma(\alpha_C)$ - масштабний коефіцієнт. Алгоритм (16.12) широко використовується в системах стеження за параметрами сигналу, наприклад, для керування напрямком орієнтації антен. Недоліком цього методу є необхідність попереднього підбору константи $\gamma(\alpha_C)$. Як витікає з

попереднього розгляду, ця константа, взагалі кажучи, може залежати й від прийнятої реалізації спостережень. Але практично константу $\gamma(\alpha_C)$ підбирають за деякими усередненими характеристиками сигналів та завад, що може слугувати джерелом додаткових похибок. Тому якщо завдання полягає саме в одноразовому уточненні оцінки параметру, а не в безперервному стеженні за ним, то більш доцільно використовувати не два, а три опорних канали. В цьому випадку до "лівого" та "правого" каналів додається ще й "центральный" канал, налаштований на значення параметру $\alpha_C = (\alpha_R + \alpha_L)/2$. Вважається, що фактичне значення параметру $\alpha_0 \in (\alpha_L, \alpha_R)$. Тоді на основі тієї ж параболічної апроксимації приходимо до алгоритму

$$\hat{\alpha}_0 = \alpha_C - \frac{\Delta}{2} \cdot \frac{F_R - F_L}{F_R + F_L - 2F_C}, \quad (16.13)$$

де $\Delta = \alpha_C - \alpha_L = \alpha_R - \alpha_C$. Неважко бачити, що алгоритм (16.13) є подібним до алгоритму (16.11), але на відміну від нього дозволяє обійтись лише трьома опорними каналами.

16.3. Сигнальна функція обробки та функція невизначеності.

Вибір методу вирішення рівнянь (16.5) – питання не стільки принципове, скільки технічне. Успішність же вимірювань визначається деякими загальними властивостями функціоналу обробки $F(\mathbf{u}/\boldsymbol{\alpha})$, зокрема, його здатністю реагувати на ступінь неузгодженості поміж параметрами настроювання $\boldsymbol{\alpha}$ та фактичними параметрами сигналу $\boldsymbol{\alpha}_0$. Цю властивість прийнято називати селективністю системи прийому. Мірою селективності може бути відносне зменшення рівня сигналу по виходу приймальної системи в залежності від розбіжності у значеннях $\boldsymbol{\alpha}$ та $\boldsymbol{\alpha}_0$. Але результати обробки залежать не тільки від параметрів сигналу та параметрів настроювання приймальної системи, але й від реалізації шумів та завад. Тому природно визначати селективність системи обробки за деякими

усередненими характеристиками. Зокрема, в якості такої характеристики доцільно використовувати залежність математичного очікування виходу обробки $F(\mathbf{u}/\mathbf{a})$ від параметрів настроювання $\mathbf{a}$ за умови наявності на вході системи сигналу з параметрами $\mathbf{a}_0$, тобто:

$$F_S(\mathbf{a}/\mathbf{a}_0) = E\{F(\mathbf{u}/\mathbf{a})/\mathbf{a}_0\} \quad (16.14)$$

Характеристика (16.14), яку в деяких джерелах [11,24] називають *сигнальною функцією обробки (СФО)*, визначає селективні можливості алгоритму приймання сигналів в ситуації надійних вимірювань. Дійсно, для будь-яких конзистентних статистик має місце [24] збіжність за деяким імовірнісним критерієм функціоналу обробки до свого математичного очікування, а саме:

$$F(\mathbf{u}/\mathbf{a})|_{\mathbf{a}_0} \xrightarrow{q^2 \rightarrow \infty} F_S(\mathbf{a}/\mathbf{a}_0),$$

де, нагадаймо, q^2 - відношення сигнал/завада по виходу обробки. Отже, дослідження поведінки сигнальної функції (16.14) при наявності розходження у значенні параметрів $\mathbf{a}$ та $\mathbf{a}_0$ дозволяє робити висновки про асимптотичні ($q^2 \gg 1$) селективні властивості функціоналу $F(\mathbf{u}/\mathbf{a})$. Головна вимога до функції (16.14) полягає в тому, що вона повинна досягати свого максимально можливого значення лише при збігу аргументів $\mathbf{a}$ з фактичними значеннями параметрів сигналу $\mathbf{a}_0$. Для усіх інших значень $\mathbf{a}$ повинна виконуватись нерівність

$$F_S(\mathbf{a}/\mathbf{a}_0) < F_S(\mathbf{a}_0/\mathbf{a}_0).$$

При цьому чим вказана нерівність сильніша, тим селективність системи краща. Отже, швидкість спадання сигнальної функції обробки при наявності неузгодженості поміж параметрами настроювання приймальної системи $\mathbf{a}$ та фактичними параметрами сигналу $\mathbf{a}_0$ може бути кількісною мірою селективності обробки. Наприклад, якщо необхідно проводити вимірювання частоти гармонійного сигналу, то система обробки повинна мати частотну селективність. Цю селективність можна визначити наступним чином:

зафіксуємо таке відхилення $\tilde{f}$ центральної частоти настроювання приймача $f = f_0 + \tilde{f}$ від фактичної частоти сигналу f_0 , при якому спостерігається спад сигнальної функції на D дБ. Тоді говорять, що приймач має частотну селективність « D дБ при розладі на $\tilde{f}$ Гц». Схожим чином визначається й селективність по кутам приходу фронту хвилі сигналу, затримці сигналу в часі та багатьом іншим параметрам.

Область високих значень СФО при малих розходженнях поміж параметрами α та α_0 прийнято називати її головним максимумом (або пелюстком). Саме шириною головного пелюстка СФО на певному рівні й характеризують найчастіше селективні можливості системи обробки. Окрім головного сигнальні функції можуть мати й додаткові (бокові) максимуми (або пелюстки), які розташовані в області значень параметрів настроювання α , значно віддалених від фактичних параметрів сигналу α_0 . Зокрема, відносний рівень бокових пелюстків СФО характеризує здатність приймальної системи до розрізнення багатьох сигналів.

Важливе значення також має розташування максимуму головного пелюстка. Якщо положення максимуму $\alpha_{\max}$ відповідає фактичним значенням параметрів сигналу α_0 , то можна стверджувати, що принаймні для надійних вимірювань систематичні похибки оцінювання відсутні. Зокрема, оптимальна за критерієм ML обробка має ту важливу особливість, що її СФО завжди цій вимозі відповідає³. В цьому можна переконатися й цілком формальним шляхом. Дійсно, на підставі виразів (16.2) та (16.14) сигнальна функція оптимальної за критерієм ML обробки може бути представлена у вигляді:

$$F_S^{MP}(\alpha/\alpha_0) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \ln \left(\frac{p(\mathbf{u}/\alpha)}{p_0(\mathbf{u})} \right) p(\mathbf{u}/\alpha_0) d\mathbf{u} \quad (16.15)$$

³ Звісно, за умови, що ML оцінка існує.

Припустимо, що задача оцінювання задовольняє вказаним у §10.5 умовам регулярності, зокрема, межі інтегрування у виразі (16.15) від параметрів $\mathbf{a}$ не залежать. Запишемо похідну від функції (16.15) по деякій компоненті вектору параметрів α_k :

$$\frac{\partial F_S^{МП}(\mathbf{a}/\mathbf{a}_0)}{\partial \alpha_k} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \frac{p(\mathbf{u}/\mathbf{a}_0)}{p(\mathbf{u}/\mathbf{a})} d\mathbf{u} .$$

В точці $\mathbf{a} = \mathbf{a}_0$ з огляду на вказані вище умови регулярності маємо:

$$\left. \frac{\partial F_S^{МП}(\mathbf{a}/\mathbf{a}_0)}{\partial \alpha_k} \right|_{\mathbf{a}=\mathbf{a}_0} = \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} d\mathbf{u} \right]_{\mathbf{a}=\mathbf{a}_0} = \frac{\partial}{\partial \alpha_k} \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\mathbf{u}/\mathbf{a}) d\mathbf{u} \right]_{\mathbf{a}=\mathbf{a}_0} .$$

Функціонал правдоподібності є густиною ймовірностей розподілу спостережень для певної гіпотези щодо значення параметрів сигналу. Отже, як й кожна густина ймовірностей, розподіл $p(\mathbf{u}/\mathbf{a})$ має задовольняти умові нормування

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\mathbf{u}/\mathbf{a}) d\mathbf{u} = 1 .$$

З урахуванням цього співвідношення неважко бачити, що

$$\left. \frac{\partial F_S^{МП}(\mathbf{a}/\mathbf{a}_0)}{\partial \alpha_k} \right|_{\mathbf{a}=\mathbf{a}_0} \equiv 0 .$$

Таким чином, в точці $\mathbf{a} = \mathbf{a}_0$ необхідні умови екстремуму СФО оптимальної за критерієм ML обробки виконуються. Залишилось лише перевірити, що точка $\mathbf{a} = \mathbf{a}_0$ є саме точкою максимуму цієї функції. Для цього запишемо компоненти відповідної матриці Гессе:

$$H_{kl}^{МП}(\mathbf{a}/\mathbf{a}_0) = \frac{\partial^2 F_S^{МП}(\mathbf{a}/\mathbf{a}_0)}{\partial \alpha_k \partial \alpha_l} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial^2 p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k \partial \alpha_l} \frac{p(\mathbf{u}/\mathbf{a}_0)}{p(\mathbf{u}/\mathbf{a})} d\mathbf{u} - \\ - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_l} \frac{p(\mathbf{u}/\mathbf{a}_0)}{p^2(\mathbf{u}/\mathbf{a})} d\mathbf{u} .$$

В точці $\mathbf{a} = \mathbf{a}_0$ маємо:

$$H_{kl}^{МП}(\mathbf{a}_0/\mathbf{a}_0) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[\frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \right]_{\mathbf{a}=\mathbf{a}_0} \left[\frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_l} \right]_{\mathbf{a}=\mathbf{a}_0} \frac{d\mathbf{u}}{p(\mathbf{u}/\mathbf{a}_0)} .$$

Далі утворимо квадратичну форму від цієї матриці з довільним вектором $\mathbf{x} \neq 0$:

$$Q_H = \sum_k \sum_l x_k \cdot H_{kl}^{МП}(\mathbf{a}_0/\mathbf{a}_0) \cdot x_l = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[\sum_n x_n \cdot \frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_n} \right]_{\mathbf{a}=\mathbf{a}_0}^2 \frac{d\mathbf{u}}{p(\mathbf{u}/\mathbf{a}_0)} .$$

Якщо задача вимірювання є коректною, то похідна від функціоналу правдоподібності по параметрах, що вимірюються, не може бути тотожно нульовою (в іншому випадку задача вимірювання взагалі б не мала сенсу). Отже, підінтегральний вираз в останній формулі завжди є позитивним, а весь вираз, відповідно, негативним. Таким чином, матриця Гессе сигнальної функції оптимальної за критерієм ML обробки в точці $\mathbf{a} = \mathbf{a}_0$ є негативно визначеною, що і є достатньою умовою максимуму.

Відзначимо ще одну важливу особливість сигнальної функції оптимальної обробки, а саме те, що її гессіан з точністю до знаку співпадає з умовною інформаційною матрицею Фішера. Дійсно, беручи до уваги тотожність

$$\left[\frac{\partial p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \right]_{\mathbf{a}=\mathbf{a}_0} = \left[\frac{\partial \ln p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \right]_{\mathbf{a}=\mathbf{a}_0} p(\mathbf{u}/\mathbf{a}_0) ,$$

та співвідношення (10.49), неважко бачити, що

$$H_{kl}^{МП}(\mathbf{a}_0/\mathbf{a}_0) = - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[\frac{\partial \ln p(\mathbf{u}/\mathbf{a})}{\partial \alpha_k} \right]_{\mathbf{a}=\mathbf{a}_0} \left[\frac{\partial \ln p(\mathbf{u}/\mathbf{a})}{\partial \alpha_l} \right]_{\mathbf{a}=\mathbf{a}_0} p(\mathbf{u}/\mathbf{a}_0) d\mathbf{u} = -(\mathbf{D}_{\mathbf{a}_0})_{kl}$$

Отже, маємо ще одну формулу для обчислення умовної інформаційної матриці вимірювальної задачі⁴:

$$(\mathbf{D}_{\mathbf{a}_0})_{ik} = - \left. \frac{\partial^2 F_S^{МП}(\mathbf{a}/\mathbf{a}_0)}{\partial \alpha_i \partial \alpha_k} \right|_{\mathbf{a}=\mathbf{a}_0} . \quad (16.16)$$

⁴ Нагадаємо, що цей висновок є вірним при виконанні згаданих вище умов регулярності.

Сигнальні функції обробки тісно пов'язані з іншою важливою мірою селективності, а саме, так званими *функціями невизначеності*⁵ (ФН). Вперше поняття функції невизначеності було запропоновано Вудвордом (Woodward), який розглядав її як нормовану кореляційну функцію сигналу по його інформаційним параметрам [23]:

$$\Phi_{s,s}(\mathbf{a}/\mathbf{a}_0) = \frac{1}{K_{norm}} \int_T s(t/\mathbf{a}) \cdot s(t/\mathbf{a}_0) dt ,$$

де K_{norm} - деяка нормуюча константа або функція, а інтегрування виконується по часовій області спільного існування сигналів $s(t/\mathbf{a})$ та $s(t/\mathbf{a}_0)$. Неважко бачити, що ця функція пропорційна сигнальній частині виходу кореляційної обробки

$$F(\mathbf{u}/\mathbf{a}) = \int_T u(t) \cdot s(t/\mathbf{a}) dt ,$$

якщо на її вхід подається адитивна суміш корисного сигналу та завад. Отже, ФН Вудворда у своєму первісному вигляді є СФО обробки кореляційного типу, що нормована деяким чином. У численних наступних роботах [3, 11, 17] ідеї Вудворда одержали подальший розвиток, але й ці роботи, як правило, виходили з кореляційного принципу прийому. Змінювалися лише вигляд опорної функції, спосіб обчислення кореляційного інтегралу, спосіб нормування, тощо. Тому в сучасній літературі найчастіше ФН визначають [3, 17] наступним чином:

$$\Phi_{b,s}(\mathbf{a}/\mathbf{a}_0) = (\mathbf{b}(\mathbf{a}), \mathbf{s}(\mathbf{a}_0)) / K_{norm} , \quad (16.17)$$

де $\mathbf{b}(\mathbf{a})$ - деякий (не обов'язково оптимальний) опорний вектор обробки, налаштованої на прийом сигналу з параметрами $\mathbf{a}$, $\mathbf{s}(\mathbf{a}_0)$ - сигнальний вектор, що відповідає його фактичним параметрам $\mathbf{a}_0$. При цьому вектори $\mathbf{b}$ та $\mathbf{s}$ можуть бути не тільки дійсними, але й комплексними. В останньому випадку розглядають, як правило, не саму функцію (16.17), а лише її обвідну.

⁵ У світовій літературі –ambiguity function, що можна також перекласти як функція неоднозначності.

Нормування у виразі (16.17) також може бути різним. Зокрема, доволі поширеним підходом є таке нормування, що забезпечує досягнення максимуму ФН в точці $\alpha = \alpha_0$. З нерівності Коші-Буняковского [9] витікає, що для цього достатньо прийняти

$$K_{norm} = (\mathbf{b}(\alpha_0), \mathbf{s}(\alpha_0)).$$

Але насправді селективні можливості обробки у загальному випадку визначаються, як було показано вище, не функцією (16.17), а сигнальною функцією обробки (16.14). В окремих випадках ці функції можуть співпадати або не співпадати поміж собою. Часто ФН входить лише як складова частина до певної сигнальної функції. Тому вибір нормування такої часткової характеристики, як функція (16.17), на суть справи не впливає. Це лише питання певних інтерпретацій або традицій, що склалися. Деякі ілюстрації з цього приводу будуть розглянуті нижче.

Слід підкреслити, що як СФО, так й ФН є саме функціями параметрів настроювання системи α при фіксованих параметрах сигналу α_0 , а не навпаки⁶. Розуміння цієї досить простої обставини дозволяє іноді уникнути принципових помилок. Так, наприклад, відповідно до традиційного визначення, діаграма направленості (ДН) антени – це нормована залежність рівня сигналу на виході антени від напрямку його приходу [21]. Отже, це традиційне визначення ДН не є, взагалі кажучи, ФН за напрямком у вказаному вище сенсі. Але для багатьох видів обробки ФН має властивості симетрії⁷, а саме,

$$\Phi_{b,s}(\alpha/\alpha_0) = \Phi_{b,s}^*(\alpha_0/\alpha). \quad (16.18)$$

Таку властивість, зокрема, має традиційний алгоритм формування направленості антен (13.4) та усі його версії. Якщо умова (16.18) виконується, то не має значення, якій з параметрів α , α_0 фіксований, а який змінюється. Тоді традиційне визначення ДН антени збігається з ФН за

⁶ Можна їх також розглядати як функції двох аргументів α та α_0 , але це принципово нічого не змінює.

⁷ У загальному випадку сигнали, а отже, й ФН, можуть бути комплексними.

напрямок приходу сигналу. Але такий збіг має місце не завжди. Наприклад, у §14.3 наводилися діаграми направленості оптимальної просторової обробки для деякої моделі середовища передачі та поля завад (рис.14.3). Ці діаграми обчислювались відповідно до традиційного визначення, а саме, фіксованим був параметр настроювання системи, а змінним – напрямок приходу сигналу. Як при цьому відзначалося максимум ДН оптимальної обробки на цих графіках не збігається з параметром настроювання системи обробки. Причина такого розходження обумовлена анізотропією поля завад, наслідком якої є нерівномірність ефективності прийому сигналу для різних напрямків. Але подібне розходження може мати місце лише за умови неузгодження алгоритму обробки з напрямком приходу сигналу. Якщо ж при зміні напрямку орієнтації ДН одночасно змінювати й вагову функцію обробки, то максимум як сигнальної функції, так й ФН завжди буде займати "правильне" положення.

16.4. Аналіз похибок вимірювань.

Будь-яке вимірювання має сенс лише тоді, коли воно здатне забезпечити прийнятну точність результату. У свою чергу, аналіз можливих похибок часто пов'язано з необхідністю виконання складних обчислень та/або трудомісткого комп'ютерного моделювання. Слід відразу зазначити, що можливості аналітичних методів оцінки похибок доволі обмежені. Дійсно, обчислення розподілу похибок оцінювання та їх числових характеристик є складною математичною проблемою, рішення якої вдається знайти лише для окремих випадків. Зокрема, до цих випадків не відносяться, як правило, сучасні методи обробки сигналів, в яких широко використовуються нелінійні та адаптивні процедури. Крім того, більш-менш достовірні аналітичні результати вдається одержати, як правило, лише для досить високих відношень сигнал/завада по виходу обробки, тобто, для надійних вимірювань. З вказаних причин значного поширення в наш час набули методи аналізу похибок на основі комп'ютерного моделювання. Проте, на

початкових етапах розробки деякої технічної системи, коли кількість варіантів її побудови може бути дуже велика, доцільно все ж таки мати можливість порівнювального аналізу ефективності тих чи інших варіантів й суто теоретичним шляхом.

Основним теоретичним методом аналізу ефективності вимірювальних схем є використання границь Рао-Крамера, які розглядалися у §10.5. Ці границі дають мінімально можливі похибки для оптимальних методів оцінювання, які у повній мірі звичайно не можуть бути реалізовані практично. Але на етапі проектування вимірювальної системи цей аспект й не є головним: важливо, наприклад, раціонально вибрати взаємне розташування випромінюючих та приймальних пристроїв, характеристики антен, вид сигналу та його параметри, об'єм статистичних спостережень, тощо. Про доцільність використання тієї чи іншої схеми вимірювання можна, зокрема, судити на основі властивостей інформаційної матриці Фішера, яка визначає статистичну компоненту похибок. В задачах, де спостереження та інформаційні параметри пов'язані якоюсь складною залежністю, нерідко виникає ситуація, коли ця матриця виявляється близькою до сингулярної (або, як прийнято говорити, погано обумовленою). Така ситуація може свідчити, зокрема, про невдало обрану схему вимірювань. Тоді слід спробувати змінити цю схему, параметри приймально-передаючих пристроїв, їх розташування, вид та параметри сигналу, тощо. Якщо ж, не дивлячись на всі зусилля, суттєвого покращання результату не відбувається, то це може означати й взагалі практичну неможливість вирішення такої задачі. Отже, суто теоретичний підхід до аналізу похибок оцінювання заслуговує певної уваги, оскільки дозволяє на початкових етапах проектування досить суттєво економити час та матеріальні ресурси. Потім, вже після вибору практичної схеми побудови системи, похибки вимірювань, які реально досягаються, можуть бути знайдені шляхом комп'ютерного моделювання та/або експериментально.

Основу теоретичного підходу до аналізу похибок вимірювання складають формули (10.54) – (10.57) та усі їх окремі випадки і модифікації, що наведені у §10.5. Згідно до цих формул теоретичний аналіз похибок зводиться до обчислення умовного зміщення оцінки (систематичної похибки) та дисперсії оцінки (статистичної похибки), яка визначається властивостями умовної інформаційної матриці Фішера (10.55). Зокрема, у випадку незміщених оцінок статистична похибка повністю визначається спектром власних чисел матриці (10.55), які знаходяться в результаті рішення рівняння (10.60). Втім, можливості теоретичного підходу можуть бути в деякій мірі поширені й на довільні способи сигнальної обробки, які частково або повністю не є оптимальними. Розглянемо один з найбільш вживаних методів розрахунку [11, 24], який базується на припущенні про те, що похибка вимірювань є малою у порівнянні зі шириною основного пелюстка сигнальної функції. Це відповідає випадку надійних вимірювань та відсутності так званих *аномальних похибок*, тобто, таких, які виникають внаслідок помилкового сприйняття в якості основного пелюстка СФО якогось її бокового викиду, що може бути зумовлено дією шумів та завад. При наявності аномальних помилок відхилення оцінки від вірного значення може бути дуже значним. Тому метод розрахунку, який тут розглядається, на аномальні похибки не поширюється.

Представимо деякий довільний функціонал обробки $F(\mathbf{u}/\alpha)$ у вигляді суми його математичного очікування та флуктуацій, що зумовлені дією випадкових факторів:

$$F(\mathbf{u}/\alpha) = F_S(\alpha/\alpha_0) + F_Z(\mathbf{u}/\alpha) \quad (16.19)$$

Тут $F_S(\alpha/\alpha_0)$ - сигнальна функція обробки (16.14),
 $F_Z(\mathbf{u}/\alpha) = F(\mathbf{u}/\alpha) - F_S(\alpha/\alpha_0)$ - флуктуаційна складова функціоналу обробки, яку можна вважати завадою по виходу системи обробки. Розкладемо⁸

⁸ Для цього СФО повинна звичайно мати похідні відповідних порядків.

детерміновану функцію $F_S(\mathbf{a}/\mathbf{a}_0)$ в ряд по ступеням параметрів $\mathbf{a}$ поблизу їх фактичного значення $\mathbf{a}_0$:

$$F_S(\mathbf{a}/\mathbf{a}_0) = F_S(\mathbf{a}_0/\mathbf{a}_0) + (\mathbf{a} - \mathbf{a}_0) \cdot \left. \frac{\partial F_S(\mathbf{a}/\mathbf{a}_0)}{\partial \mathbf{a}} \right|_{\mathbf{a}=\mathbf{a}_0} + \\ + \frac{1}{2} (\mathbf{a} - \mathbf{a}_0)^T \cdot \left. \frac{\partial^2 F_S(\mathbf{a}/\mathbf{a}_0)}{\partial \mathbf{a} \partial \mathbf{a}^T} \right|_{\mathbf{a}=\mathbf{a}_0} \cdot (\mathbf{a} - \mathbf{a}_0) + \dots$$

Припустимо, що максимум сигнальної функції $F_S(\mathbf{a}/\mathbf{a}_0)$ знаходиться в точці $\mathbf{a} = \mathbf{a}_0$. Тоді лінійний член даного розкладу дорівнює нулю і функціонал (16.19) з точністю до квадратичних членів в околі точки $\mathbf{a}_0$ може бути представлений у вигляді:

$$F(\mathbf{u}/\mathbf{a}) \cong F_S(\mathbf{a}_0/\mathbf{a}_0) + \frac{1}{2} (\mathbf{a} - \mathbf{a}_0)^T \cdot \mathbf{H}_S \cdot (\mathbf{a} - \mathbf{a}_0) + F_Z(\mathbf{u}/\mathbf{a}), \quad (16.20)$$

де $\mathbf{H}_S$ - гессіан сигнальної функції $F_S(\mathbf{a}/\mathbf{a}_0)$ в точці $\mathbf{a} = \mathbf{a}_0$. Тепер підставимо розклад (16.20) у рівняння (16.5) щодо оцінки $\hat{\mathbf{a}}$:

$$\nabla_{\mathbf{a}} F(\mathbf{u}/\mathbf{a}) \cong \mathbf{H}_S \cdot (\hat{\mathbf{a}} - \mathbf{a}_0) + \nabla_{\mathbf{a}} F_Z(\mathbf{u}/\mathbf{a}) \Big|_{\mathbf{a}=\hat{\mathbf{a}}} = 0. \quad (16.21)$$

Це рівняння, взагалі кажучи, може бути нелінійним, оскільки від шуканого значення оцінки $\hat{\mathbf{a}}$ залежать обидва доданки його лівої частини. Однак, статистичні характеристики завад по виходу обробки звичайно слабо залежать від точки обчислення вказаного градієнту. Тому прийнято вважати [24], що без значної похибки можлива наступна заміна:

$$\nabla_{\mathbf{a}} F_Z(\mathbf{u}/\mathbf{a}) \Big|_{\mathbf{a}=\hat{\mathbf{a}}} \approx \nabla_{\mathbf{a}} F_Z(\mathbf{u}/\mathbf{a}) \Big|_{\mathbf{a}=\mathbf{a}_0}.$$

Беручи до уваги цю обставину, із співвідношення (16.21) маємо наступне наближене рівняння стосовно оцінки $\hat{\mathbf{a}}$:

$$\mathbf{H}_S \cdot (\hat{\mathbf{a}} - \mathbf{a}_0) \cong \mathbf{V}_Z, \quad (16.22)$$

$$\text{де} \quad \mathbf{V}_Z = -\nabla_{\mathbf{a}} F_Z(\mathbf{u}/\mathbf{a}) \Big|_{\mathbf{a}=\mathbf{a}_0} \quad (16.23)$$

- вектор, протилежний градієнту завадової складової по параметрах α .

Тепер нескладно записати рівняння щодо кореляційної матриці помилок оцінювання

$$\mathbf{K}_\alpha = \langle (\hat{\alpha} - \alpha_0) \cdot (\hat{\alpha} - \alpha_0)^T \rangle.$$

Використовуючи співвідношення (16.22), маємо:

$$\mathbf{H}_S \cdot \mathbf{K}_\alpha \cdot \mathbf{H}_S^T \cong \mathbf{W}_Z. \quad (16.24)$$

$$\text{Тут} \quad \mathbf{W}_Z = \langle \mathbf{V}_Z \cdot \mathbf{V}_Z^T \rangle \quad (16.25)$$

- кореляційна матриця градієнту завадових складових. Рішення наближеного рівняння (16.24) має вигляд:

$$\mathbf{K}_\alpha \cong \mathbf{H}_S^{-1} \cdot \mathbf{W}_Z \cdot (\mathbf{H}_S^T)^{-1}. \quad (16.26)$$

Беручи слід матриці (16.26), знаходимо середній квадрат похибки вимірювання:

$$d_\alpha^2 = \langle |\hat{\alpha} - \alpha_0|^2 \rangle \cong \text{Tr} \left[\mathbf{H}_S^{-1} \cdot \mathbf{W}_Z \cdot (\mathbf{H}_S^T)^{-1} \right] = \text{Tr} \left[\mathbf{W}_Z \cdot (\mathbf{H}_S \cdot \mathbf{H}_S^T)^{-1} \right]. \quad (16.27)$$

В окремому випадку оцінювання скалярного параметру з виразу (16.27) витікає формула:

$$\langle |\hat{\alpha} - \alpha_0|^2 \rangle \cong \left\langle \left[\frac{dF_Z(\mathbf{u}/\alpha)}{d\alpha} \right]_{\alpha=\alpha_0}^2 \right\rangle \cdot \left[\frac{d^2 F_S(\alpha/\alpha_0)}{d\alpha^2} \right]_{\alpha=\alpha_0}^{-2}. \quad (16.28)$$

Неточність наближених формул (16.26) – (16.28) має [11] порядок $O(1/q^2)$. Отже, користуватися ними доцільно лише у випадку надійних вимірювань, коли $q^2 \gg 1$. Втім, розглянутий підхід може бути узагальнено й на випадок помірних значень q^2 . Для цього слід скористатися розкладом похибки оцінювання в ряд за ступенями параметру $1/q^2$. Зокрема, наближення такого роду розглянуто в класичних монографіях [3, 11, 24]. Також відзначимо, що у випадку оптимального оцінювання формули (16.26) – (16.28) переходять у границі Рао-Крамера, згадувані вище. Дійсно, відповідно до (16.16) для оптимальної обробки $\mathbf{H}_S^{opt} = -\mathbf{D}_\alpha$. Неважко також

довести, що й кореляційна матриця градієнтів завад $\mathbf{W}_Z^{opt} = -\mathbf{D}_\alpha$. Тоді права частина виразу (16.27) переходить у границю (10.54) для незміщених оцінок.

16.5. Оптимальні алгоритми вимірювання параметрів детермінованих сигналів.

На завершення розділу розглянемо більш детально структури оптимальних оцінок параметрів сигналів, сигнальних функцій обробки, функцій невизначеності та похибок вимірювання. При цьому обмежимося сигналами, вигляд та параметри яких (крім, звичайно, параметрів α) є точно відомими. Це дозволить встановити деякі важливі закономірності та визначити теоретичні границі ефективності процедур оцінювання. Що ж стосується сигналів інших типів, тобто, квазидетермінованих та випадкових, про які йшла мова у розділах 11, 13, 15, то для них подібний розгляд за складністю та об'ємом виходить далеко за рамки Посібника. Тому можемо лише порекомендувати читачам, які потребують більш ґрунтовної інформації з цих питань, звернутися до відповідних спеціалізованих джерел [11, 24].

У розділі 11 були розглянуті структури максимально правдоподібної обробки спостережень стосовно завдання виявлення сигналів. Ці ж результати можуть бути використані (з певними, звичайно, корективами) й для задач вимірювання параметрів сигналів, оскільки оптимальна обробка й у цьому випадку спирається на метод максимальної правдоподібності. Різниця полягає лише в тому, що для задач виявлення використовується відношення правдоподібності гіпотез про наявність або відсутність сигналу на приймальній апертурі. В задачах вимірювання факт наявності сигналу на апертурі вважається достовірно відомим, але невідомими є, принаймні, деяка сукупність параметрів сигналу α . Відповідно, оптимальним функціоналом обробки є умовне відношення правдоподібності (16.2) для фіксованої гіпотези щодо параметрів α .

Спираючись на формули (11.6) – (11.8) розділу 11 запишемо логарифм функціоналу відношення правдоподібності (16.2) для завдань прийому детермінованих сигналів:

$$\ln \lambda(\mathbf{u} / \boldsymbol{\alpha}) = (\mathbf{b}(\boldsymbol{\alpha}), \mathbf{u}) - \frac{1}{2} q^2(\boldsymbol{\alpha}). \quad (16.29)$$

Тут $(\mathbf{x}, \mathbf{y})$ позначає, як й раніше, скалярний добуток деяких векторів $\mathbf{x}$ та $\mathbf{y}$ в заданому просторі спостережень, $\mathbf{b}(\boldsymbol{\alpha})$ - ваговий вектор обробки, який у випадку оптимальної обробки знаходиться в результаті рішення рівняння (11.8), $q^2(\boldsymbol{\alpha}) = (\mathbf{b}(\boldsymbol{\alpha}), \mathbf{s}(\boldsymbol{\alpha}))$ – відношення сигнал/завада по виходу обробки при прийманні сигналу $s(t, \mathbf{r} / \boldsymbol{\alpha})$. Після підстановки функціоналу (16.29) у співвідношення (16.5) маємо систему рівнянь щодо оптимальної за критерієм ML оцінки параметрів $\boldsymbol{\alpha}$:

$$\frac{\partial \ln \lambda(\mathbf{u} / \boldsymbol{\alpha})}{\partial \alpha_i} = \left(\frac{\partial \mathbf{b}(\boldsymbol{\alpha})}{\partial \alpha_i}, \mathbf{u} \right) - \frac{1}{2} \frac{\partial q^2(\boldsymbol{\alpha})}{\partial \alpha_i} = 0, \quad i = 1, \dots, M \quad (16.30)$$

Другий доданок лівої частини цих рівнянь виникає внаслідок залежності відношення сигнал/завада по виходу оптимальної обробки від параметрів, що оцінюються. Можна виділити широкий клас задач статистичного оцінювання [11, 24], в яких показник завадостійкості $q^2(\boldsymbol{\alpha})$ від параметрів $\boldsymbol{\alpha}$ не залежить, тобто, $q^2(\boldsymbol{\alpha}) \equiv q_0^2$. Тоді рівняння (16.30) суттєво спрощуються оскільки в їх лівій частині зберігається лише перший доданок, тобто,

$$\left(\frac{\partial \mathbf{b}(\boldsymbol{\alpha})}{\partial \alpha_i}, \mathbf{u} \right) = 0, \quad i = 1, \dots, M.$$

Подібні ситуації прийнято називати оцінюванням **неенергетичних параметрів** сигналу [11, 24]. Наприклад, до неенергетичних параметрів звичайно відносять початкову фазу сигналу. Якщо ж залежністю $q^2(\boldsymbol{\alpha})$ від параметрів $\boldsymbol{\alpha}$ або його окремих компонент знехтувати неможливо, то такі параметри називають **енергетичними**. Простий приклад енергетичного параметру – амплітуда сигналу. Зазначимо, що до класифікації параметрів на

енергетичні або неенергетичні слід підходити досить обережно, оскільки параметр q^2 є відношенням сигнал/завада не у первинному полі на приймальній апертурі, а вже по виходу обробки в цілому. Тому він залежить від усієї сукупності факторів задачі прийому, зокрема, виду сигналу, моделі середовища передачі, моделі завад, способу обробки, тощо. З огляду на ці фактори можливі ситуації, коли один й той же параметр в залежності від структури та параметрів приймальної системи чи моделі середовища передачі може бути віднесений в одних випадках до енергетичних, в інших – до неенергетичних.

Проаналізуємо систему рівнянь (16.30). Враховуючи визначення параметру завадостійкості $q^2(\alpha) = (\mathbf{b}(\alpha), \mathbf{s}(\alpha))$, ці рівняння можна записати у вигляді

$$\left(\frac{\partial \mathbf{b}(\alpha)}{\partial \alpha_i}, \mathbf{u} \right) - \frac{1}{2} \left(\frac{\partial \mathbf{b}(\alpha)}{\partial \alpha_i}, \mathbf{s}(\alpha) \right) - \frac{1}{2} \left(\mathbf{b}(\alpha), \frac{\partial \mathbf{s}(\alpha)}{\partial \alpha_i} \right) = 0.$$

Дійсні вектори $\mathbf{b}$ та $\mathbf{s}$ у випадку оптимальної обробки пов'язані поміж собою через рівняння (11.8) симетричним та позитивно визначеним кореляційним оператором. Враховуючи цю обставину, неважко бачити, що

$$\left(\mathbf{b}(\alpha), \frac{\partial \mathbf{s}(\alpha)}{\partial \alpha_i} \right) = \left(\frac{\partial \mathbf{b}(\alpha)}{\partial \alpha_i}, \mathbf{s}(\alpha) \right).$$

Звідси для оптимальної за критерієм ML оцінки параметрів α детермінованого сигналу остаточно маємо:

$$\left(\frac{\partial \mathbf{b}(\alpha)}{\partial \alpha_i}, \mathbf{u} \right) - \left(\frac{\partial \mathbf{b}(\alpha)}{\partial \alpha_i}, \mathbf{s}(\alpha) \right) = 0, \quad i = 1, \dots, M. \quad (16.31)$$

Замкнені аналітичні рішення рівнянь (16.31) відомі для доволі обмеженого кола задач, коли можлива їх лінеаризація по параметрах α . Усі ці випадки докладно розглянуті в класичних монографіях [3, 11, 24], тому нема потреби на них зупинятися.

Тепер перейдемо до розгляду сигнальних функцій оптимальної обробки. Згідно з прийнятою моделлю вектор спостережень по входу приймальної системи має вигляд

$$\mathbf{u} = \mathbf{z} + \mathbf{w} + \mathbf{s}_0, \quad (16.32)$$

де $\mathbf{z}, \mathbf{w}$ - вектори зовнішніх завад та власних шумів, $\mathbf{s}_0 = \mathbf{s}(\mathbf{a}_0)$ - сигнальний вектор, що відповідає фактичному значенню вимірюваних параметрів $\mathbf{a}_0$. Підставляючи вектор (16.32) до функціоналу оптимальної за критерієм ML обробки (16.29) та виконуючи усереднення по шумам і завадам, знаходимо сигнальну функцію оптимальної обробки детермінованих сигналів:

$$F_s^{opt}(\mathbf{a} / \mathbf{a}_0) = q^2(\mathbf{a}_0) \cdot \Phi_{b,s}^{opt}(\mathbf{a} / \mathbf{a}_0) - \frac{1}{2} q^2(\mathbf{a}). \quad (16.33)$$

де

$$\Phi_{b,s}^{opt}(\mathbf{a} / \mathbf{a}_0) = (\mathbf{b}(\mathbf{a}), \mathbf{s}(\mathbf{a}_0)) / (\mathbf{b}(\mathbf{a}_0), \mathbf{s}(\mathbf{a}_0)) \quad (16.34)$$

– нормована функція невизначеності Вудворда по параметрах $\mathbf{a}$. З виразу (16.33) витікає, що різницю у вигляді СФО та відповідної ФН (16.34) у даному разі визначає центрувальний член, пропорційний параметру завадостійкості $q^2(\mathbf{a})$. Але вказане розходження поміж функціями (16.33) та (16.34) має значення лише тоді, коли параметри $\mathbf{a}$ є енергетичними. Якщо ж вимірюванню підлягають неенергетичні параметри, тобто, коли $q^2(\mathbf{a}) \equiv q_0^2$, то сигнальна функція (16.33) та традиційна ФН (16.34) фактично співпадають за виглядом, відрізняючись лише нормуванням та постійним зсувом, що, вочевидь, не має принципового значення.

Розглянемо детальніше структуру функції невизначеності (16.34), записавши її для наочності у розгорнутому (тобто, по координатному) вигляді:

$$\Phi_{b,s}^{opt}(\mathbf{a} / \mathbf{a}_0) = \frac{1}{q^2(\mathbf{a}_0)} \cdot \sum_{m=1}^M \int_{\mathbf{T}} b(t, \mathbf{r}_m / \mathbf{a}) \cdot s(t, \mathbf{r}_m / \mathbf{a}_0) dt. \quad (16.35)$$

Тут, нагадаймо, $b(t, \mathbf{r}/\mathbf{a})$ - вагова функція просторово-часової обробки, що налаштована на приймання сигналу з параметрами $\mathbf{a}$, $s(t, \mathbf{r}_m/\mathbf{a}_0)$ - поле сигналу на приймальній апертурі, яке відповідає вектору фактичних параметрів $\mathbf{a}_0$. Перш за все відзначимо, що для неенергетичних параметрів функція (16.35) задовольняє умові симетрії (16.18). Для доказу достатньо взяти до уваги ту обставину, що у випадку оптимальної обробки її опорна функція $b(t, \mathbf{r}/\mathbf{a})$ пов'язана через рівняння (13.12) з полем сигналу $s(t, \mathbf{r}/\mathbf{a})$ дійсним симетричним кореляційним оператором.

Конкретизація вигляду функції (16.35) потребує залучення додаткових припущень щодо властивостей полів сигналу та завад з числа тих, що вже використовувалися у попередніх главах. Зокрема, у §13.2 було розглянуто так зване квазістаціонарне наближення для компактних апертур, в рамках якого загальний функціонал просторово-часової обробки вдається розділити на послідовність процедур обробки у просторі та у часі. Ця ж ідея може бути застосована й для задач вимірювання параметрів сигналу. Перш за все по аналогії з (13.18) перепишемо вираз (16.35) в еквівалентному спектральному вигляді:

$$\Phi_{b,s}^{opt}(\mathbf{a}/\mathbf{a}_0) = \frac{1}{2\pi q^2(\mathbf{a}_0)} \sum_{m=1}^M \int_{-\infty}^{\infty} \dot{B}^*(\omega, \mathbf{r}_m/\mathbf{a}) \dot{S}(\omega, \mathbf{r}_m/\mathbf{a}_0) d\omega, \quad (16.36)$$

де $\dot{B}(\omega, \mathbf{r}_m/\mathbf{a})$ і $\dot{S}(\omega, \mathbf{r}_m/\mathbf{a}_0)$ - спектри Фур'є вагової функції $b(t, \mathbf{r}_m/\mathbf{a})$ та поля сигналу $s(t, \mathbf{r}_m/\mathbf{a}_0)$ відповідно. Далі по аналогії з виразами (13.19) – (13.21) для поля сигналу запишемо наступну модель:

$$\dot{S}(\omega, \mathbf{r}_m/\mathbf{a}) = \dot{S}_0(\omega/\mathbf{a}_T) \cdot \exp\{-j\omega \cdot (\mathbf{n}, \mathbf{r}_m)/c_0\}. \quad (16.37)$$

Тут $\dot{S}_0(\omega/\mathbf{a}_T)$ - спектр сигналу джерела в умовному центрі приймальної апертури, який залежить від частини компонент $\mathbf{a}_T$ загального вектору параметрів $\mathbf{a}$. До складу часткового вектору $\mathbf{a}_T$ входять параметри⁹ сигналу, які характеризують його як функцію часу. Просторові характеристики

⁹ До складу цього часткового вектору можуть входити, зокрема, амплітуда, затримка, доплерівський зсув частоти сигналу, тощо.

сигналу в рамках даної моделі представлені одиничним вектором напрямку приходу відповідної плоскої хвилі сигналу $\mathbf{n}$. Таким чином, загальний вектор параметрів має вигляд $\mathbf{a} = \{\mathbf{a}_T, \mathbf{n}\}$.

Також по аналогії з формулою (13.24) запишемо вираз для спектру відповідної вагової функції обробки:

$$\dot{B}(\omega, \mathbf{r}_m / \mathbf{a}) = \frac{\dot{S}_0(\omega / \mathbf{a}_T)}{g_w + g_z(\omega)} \cdot \dot{H}(\omega, \mathbf{r}_m / \mathbf{n}) \cdot \exp\{-j\omega \cdot (\mathbf{n}, \mathbf{r}_m) / c_0\}, \quad (16.38)$$

де, нагадаймо, $g_w, g_z(\omega)$ - СЩП шумів та завад в області прийому, $\dot{H}(\omega, \mathbf{r}_m / \mathbf{n})$ - амплітудно-фазовий розподіл (АФР) на просторовій апертурі, який у випадку оптимальної обробки є рішенням рівняння (13.25). Після підстановки виразів (16.37), (16.38) до формули (16.36) маємо:

$$\Phi_{b,s}^{opt}(\mathbf{a} / \mathbf{a}_0) = \frac{1}{2\pi \cdot q^2(\mathbf{a}_0)} \cdot \int_{-\infty}^{\infty} \frac{\dot{S}_0^*(\omega / \mathbf{a}_T) \cdot \dot{S}_0(\omega / \mathbf{a}_{T_0})}{g_w + g_z(\omega)} \cdot \dot{D}^*(\omega, \mathbf{n}_0 / \mathbf{n}) d\omega, \quad (16.39)$$

де $\mathbf{a}_T$ та $\mathbf{a}_{T_0}$ - очікувані та фактичні параметри сигналу як функції часу, $\mathbf{n}, \mathbf{n}_0$ - очікуваний та фактичний напрямок приходу плоскої хвилі сигналу, $\dot{D}(\omega, \mathbf{n}_0 / \mathbf{n})$ - ненормована ДН антени (14.18), що відповідає АФР $\dot{H}(\omega, \mathbf{r} / \mathbf{n})$.

З виразу (16.39) видно, що у загальному випадку ФН оптимальної обробки не можна розділити на складові, один з яких залежав би лише від параметрів сигналу як функції часу, а другий – від його просторових параметрів. Отже, необхідні додаткові припущення. Основним з них є умова вузької смуги сигналу у посиленій «просторовій» версії (12.15). До вказаної умови ми вже не раз зверталися, отже додаткові пояснення навряд чи потрібні. Зазначимо лише, що для радіо систем ця умова, як вже було сказано раніше, практично майже завжди виконується. У свою чергу, це дає можливість нехтувати частотною залежністю ДН антени у робочій смузі частот, тобто, можна вважати, що

$$\dot{D}(\omega, \mathbf{n}_0 / \mathbf{n}) \cong \dot{D}(\omega_0, \mathbf{n}_0 / \mathbf{n}), \text{ де } |\omega - \omega_0| \leq \Delta\omega_s / 2. \quad (16.40)$$

Тут, як й раніше, ω_0 - центральна частота, а $\Delta\omega_s$ - смуга приймальної системи. Таку ж наближену рівність можна прийняти й щодо СЦП завад $g_z(\omega)$. Отже, на підставі цих припущень у виразі (16.39) можна винести функції $\dot{D}(\omega, \mathbf{n}_0/\mathbf{n})$ та $g_z(\omega)$ з під знаку інтегралу. Крім того, врахуємо симетрію підінтегрального виразу по частоті та структуру нормуючого коефіцієнту, який має вигляд:

$$q^2(\mathbf{a}_0) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \frac{|\dot{S}_0(\omega/\mathbf{a}_{T_0})|^2}{g_w + g_z(\omega)} \cdot \dot{D}^*(\omega, \mathbf{n}_0/\mathbf{n}_0) d\omega.$$

В результаті вираз (16.39) можна наближено представити у вигляді:

$$\Phi_{b,s}^{opt}(\mathbf{a}/\mathbf{a}_0) \cong \text{Re} \left\{ \dot{\Psi}(\mathbf{a}_T/\mathbf{a}_{T_0}) \cdot \tilde{\dot{D}}^*(\omega_0, \mathbf{n}_0/\mathbf{n}) \right\}, \quad (16.41)$$

де

$$\tilde{\dot{D}}(\omega_0, \mathbf{n}_0/\mathbf{n}) = \dot{D}(\omega_0, \mathbf{n}_0/\mathbf{n}) / \dot{D}(\omega_0, \mathbf{n}_0/\mathbf{n}_0) \quad (16.42)$$

- нормована ДН просторової обробки,

$$\dot{\Psi}(\mathbf{a}_T/\mathbf{a}_{T_0}) = \frac{\int_{-\infty}^{\infty} \dot{S}_0^*(\omega/\mathbf{a}_T) \cdot \dot{S}_0(\omega/\mathbf{a}_{T_0}) d\omega}{\int_{-\infty}^{\infty} |\dot{S}_0(\omega/\mathbf{a}_{T_0})|^2 d\omega}$$

- нормована ФН по вектору «часових» параметрів сигналу. На підставі теореми Парсеваля [9] останній вираз можна також представити в еквівалентній часовій формі:

$$\dot{\Psi}(\mathbf{a}_T/\mathbf{a}_{T_0}) = \frac{\int_T \dot{s}_0^*(t/\mathbf{a}_T) \cdot \dot{s}_0(t/\mathbf{a}_{T_0}) dt}{\int_T |\dot{s}_0(t/\mathbf{a}_{T_0})|^2 dt}. \quad (16.43)$$

Отже, відповідно до співвідношення (16.41) загальна ФН обробки при застосуванні просторової умови вузької смуги сигналу розпадається під знаком $\text{Re}\{\cdot\}$ на добуток двох функцій. Перша з них, а саме, вираз (16.43), залежить лише від властивостей сигналу, як функції часу. Тому її можна

назвати *часовою функцією невизначеності* сигналу. Відповідно, функцію (16.42) логічно називати *просторовою функцією невизначеності*¹⁰, оскільки її вигляд зумовлюється лише просторовими властивостями хвильового поля на приймальній апертурі.

Конкретний вигляд функцій (16.42) та (16.43) визначається складом параметрів α , видом сигналу, параметрами приймальної системи, моделлю середовища передачі звуку, тощо. Багато прикладів та чисельних ілюстрацій з цього приводу можна знайти у відповідних підручниках та навчальних посібниках [11, 24]. Тому не має потреби на цьому детально зупинятися. Зазначимо лише, що як функція (16.42), так й функція (16.43) при співпадінні параметрів налаштування приймальної системи з фактичними параметрами сигналу, (тобто, $\alpha = \alpha_0$) приймають завдяки нормуванню одиничне значення. Але це значення є максимальним лише у випадку оцінювання неенергетичних параметрів. У загальному випадку оцінювання енергетичних параметрів максимум функцій (16.42) та (16.43) може бути й більшим за одиницю (дивись, наприклад, рис.14.3).

Тепер перейдемо до розгляду похибок оцінювання. Як вже було сказано вище, оптимальні за критерієм ML оцінки для випадку надійних вимірювань, тобто, при $q^2 \gg 1$, є асимптотично незміщеними. Отже, у відповідності з формулами §10.5 умовну дисперсію оптимальних оцінок можна знайти зі співвідношення.

$$\sigma_{opt}^2(\alpha_0) = E \left\{ |\alpha_0 - \hat{\alpha}|^2 / \alpha_0 \right\} \cong Tr \left[\mathbf{D}^{-1}(\alpha_0) \right] \quad (16.44)$$

де усереднення виконується по підпростору спостережень, що відповідають значенню параметрів $\alpha = \alpha_0$, $\mathbf{D}(\alpha)$ - умовна інформаційна матриця вимірювальної задачі. У попередніх розділах наводилось декілька тотожних формул для обчислення матриці $\mathbf{D}(\alpha)$. Зокрема, якщо скористатися формулою (10.57), то після підстановки до неї функціоналу оптимальної

¹⁰ У даному разі ця функція просто співпадає з нормованою діаграмою направленості просторової обробки, але це лише особливість обраної простої моделі. У загальному випадку можливі й більш складні функціональні залежності.

обробки (16.29) маємо наступний вираз для ik -го елементу інформаційної матриці вимірювальної задачі:

$$[\mathbf{D}(\boldsymbol{\alpha}_0)]_{ik} = E \left\{ -\frac{\partial^2}{\partial \alpha_i \partial \alpha_k} \left[(\mathbf{b}(\boldsymbol{\alpha}), \mathbf{u}) - \frac{1}{2} q^2(\boldsymbol{\alpha}) \right]_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} / \boldsymbol{\alpha}_0 \right\}. \quad (16.45)$$

Подальші дії залежать від того, чи є параметри $\boldsymbol{\alpha}$ енергетичними. Якщо ці параметри неенергетичні, то похідна від другого додатку в квадратних дужках дорівнює нулю. Звідси відразу витікає необхідний результат:

$$[\mathbf{D}(\boldsymbol{\alpha}_0)]_{ik} = E \left\{ -\left(\frac{\partial^2 \mathbf{b}(\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_k}, \mathbf{u} \right)_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} / \boldsymbol{\alpha}_0 \right\} = \left(-\frac{\partial^2 \mathbf{b}(\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_k}, \mathbf{s}(\boldsymbol{\alpha}_0) \right)_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}.$$

Порівнюючи цей вираз зі співвідношенням (16.34), неважко бачити, що

$$\mathbf{D}(\boldsymbol{\alpha}_0) = -q_0^2 \cdot \mathbf{H}^{opt}(\boldsymbol{\alpha}_0), \quad (16.46)$$

де $\mathbf{H}^{opt}(\boldsymbol{\alpha}_0)$ - матриця Гессе від функції невизначеності (16.34) з елементами

$$[\mathbf{H}^{opt}(\boldsymbol{\alpha}_0)]_{ik} = \frac{\partial^2 \Phi_{b,s}^{opt}(\boldsymbol{\alpha} / \boldsymbol{\alpha}_0)}{\partial \alpha_i \partial \alpha_k} \Big|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}. \quad (16.47)$$

Після підстановки виразу (16.46) у співвідношення (16.44) знаходимо статистичну похибку оптимального оцінювання неенергетичних параметрів:

$$\sigma_{opt}^2(\boldsymbol{\alpha}_0) \cong -q_0^{-2} \cdot \text{Tr} \left[\left(\mathbf{H}^{opt}(\boldsymbol{\alpha}_0) \right)^{-1} \right]. \quad (16.48)$$

Зокрема, у випадках оцінювання скалярних параметрів маємо:

$$\sigma_{opt}^2(\alpha_0) \cong - \left(q_0^2 \cdot \frac{d^2 \Phi_W^{opt}(\alpha / \alpha_0)}{d\alpha^2} \Big|_{\alpha=\alpha_0} \right)^{-1}. \quad (16.49)$$

Вираз (16.48) та його окремий випадок (16.49) наочно показують ті фактори, від яких залежить похибка оцінювання. Перший фактор – відношення сигнал/завада по виходу оптимальної обробки q_0^2 : чим цей показник більший, тим похибка менша. Другий фактор визначається, по суті, селективними можливостями системи обробки по параметрах сигналу, а

саме, матрицею (16.47) других похідних від функції невизначеності задачі. Дійсно, зростання норми матриці $\mathbf{H}^{opt}$ сприяє зменшенню похибки (16.48). У свою чергу, норма цієї матриці тим більша [12], чим вище фактор кривизни ФН у точці фактичних значень параметрів. Отже, зростання кривизни ФН сприяє поліпшенню селективних можливостей приймальної системи та відповідному зменшенню похибки вимірювань при однакових значеннях параметру завадостійкості q_0^2 .

Тепер розглянемо загальний випадок вимірювань енергетичних параметрів. Звертаючись знову до виразу (16.45) та беручи до уваги визначення показника q^2 , після виконання всіх необхідних обчислень маємо:

$$[\mathbf{D}(\mathbf{a}_0)]_{ik} = \left(\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \alpha_i}, \frac{\partial \mathbf{s}(\mathbf{a})}{\partial \alpha_k} \right)_{\mathbf{a}=\mathbf{a}_0}. \quad (16.50)$$

Знову ж таки порівнюючи вираз (16.50) зі співвідношенням (16.34), бачимо, що

$$[\mathbf{D}(\mathbf{a}_0)]_{ik} = \frac{\partial^2}{\partial \alpha_i \partial \alpha'_k} \left[q^2(\mathbf{a}') \cdot \Phi_{b,s}^{opt}(\mathbf{a} / \mathbf{a}') \right]_{\mathbf{a}=\mathbf{a}_0, \mathbf{a}'=\mathbf{a}_0}. \quad (16.51)$$

Отже, й у випадку оцінювання енергетичних параметрів на точність впливають ті ж фактори, що й вище: відношення сигнал/завада по виходу обробки та селективність функціоналу обробки. Тільки тепер, на відміну від попереднього випадку, селективні здатності визначаються не ФН окремо, а добутком ФН та показника $q^2(\mathbf{a})$. Зауважимо, що формула (16.51) має більш загальний характер, чим формула (16.46), оскільки вираз (16.51) є придатним для задач вимірювання будь-яких параметрів, енергетичних чи неенергетичних. Неважко впевнитися, що в окремому випадку вимірювань неенергетичних параметрів формула (16.51) переходить у вираз (16.46). Дійсно, для неенергетичних параметрів повинна виконуватись тотожність

$$q^2(\mathbf{a}) = (\mathbf{b}(\mathbf{a}), \mathbf{s}(\mathbf{a})) \equiv const.$$

Двічі диференціюючи цей вираз по параметрах $\mathbf{a}$, одержуємо наступну умову, якій повинні задовольняти сигнальна та опорна функція оптимальної обробки у випадку оцінювання неенергетичних параметрів:

$$\left(\frac{\partial^2 \mathbf{b}(\mathbf{a})}{\partial \alpha_i \partial \alpha_k}, \mathbf{s}(\mathbf{a}) \right) + \left(\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \alpha_i}, \frac{\partial \mathbf{s}(\mathbf{a})}{\partial \alpha_k} \right) + \left(\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \alpha_k}, \frac{\partial \mathbf{s}(\mathbf{a})}{\partial \alpha_i} \right) + \left(\mathbf{b}(\mathbf{a}), \frac{\partial^2 \mathbf{s}(\mathbf{a})}{\partial \alpha_i \partial \alpha_k} \right) \equiv 0.$$

Беручи до уваги, як це вже неодноразово робилося вище, симетричні властивості оператора, який пов'язує оптимальний опорний вектор з вектором сигналу, одержуємо наступний ланцюжок рівностей:

$$\begin{aligned} \left(\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \alpha_i}, \frac{\partial \mathbf{s}(\mathbf{a})}{\partial \alpha_k} \right) &= \left(\frac{\partial \mathbf{b}(\mathbf{a})}{\partial \alpha_k}, \frac{\partial \mathbf{s}(\mathbf{a})}{\partial \alpha_i} \right) = \\ &= - \left(\frac{\partial^2 \mathbf{b}(\mathbf{a})}{\partial \alpha_i \partial \alpha_k}, \mathbf{s}(\mathbf{a}) \right) = - \left(\mathbf{b}(\mathbf{a}), \frac{\partial^2 \mathbf{s}(\mathbf{a})}{\partial \alpha_i \partial \alpha_k} \right). \end{aligned} \quad (16.52)$$

З цього співвідношення легко бачити, що для випадку вимірювань неенергетичних параметрів формули (16.51) та (16.46) повністю збігаються.

16.6. Висновки.

Отже, відповідно до загальних положень, розглянутих вище, побудова деякої вимірювальної системи пов'язана з вирішенням таких основних задач:

1. Вибір функціоналу обробки (тестової статистики) $F(\mathbf{u}/\mathbf{a})$, який би забезпечував принципову можливість проведення необхідних вимірювань. Одним зі зручних та наочних методів перевірки такої можливості є обчислення сигнальних функцій (16.14). Якщо функція $F_s(\mathbf{a}/\mathbf{a}_0)$ «реагує» задовільним чином на розлад поміж параметрами настроювання системи $\mathbf{a}$ та фактичними параметрами сигналу $\mathbf{a}_0$, то є сенс у подальшому розгляді задачі. Якщо ж така залежність відсутня або занадто слабо виражена, то слід перейти до розгляду інших варіантів вибору статистики $F(\mathbf{u}/\mathbf{a})$ або взагалі інших схем приймання сигналів.
2. Побудова конкретної вимірювальної схеми на основі одного з розглянутих вище методів розв'язання системи рівнянь (16.5). Будь-яка з таких схем базується на обчислення у той чи інший спосіб градієнту

функціоналу обробки $\nabla_{\alpha} F$ по параметрах, що підлягають вимірюванню.

При цьому найбільшого практичного поширення набули наближені методи, зокрема, сканування та інтерполяції.

3. Оцінка похибок вимірювання, яка включає як теоретичний аналіз, так й математичні та/або натурні експерименти. При цьому теоретичний аналіз можливий, як правило, лише в ситуаціях великих значень відношення сигнал/завада по виходу обробки q^2 . Тому такий аналіз застосовується переважно на початкових етапах побудови систем вимірювань, коли кількість варіантів побудови занадто велика і треба обґрунтувати лише базові параметри системи. Подальші уточнення ефективності найчастіше в сучасних умовах виконуються методами комп'ютерного моделювання.

4. У випадку одержання незадовільних щодо точності вимірювань результатів усі вказані вище етапи можуть бути повторені.

Контрольні запитання та завдання до розділу 16.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. У чому полягає загальна постановка задач вимірювання параметрів сигналу? Що таке тестова статистика задачі та яким базовим вимогам вона має задовольняти?
2. У чому полягає специфіка задач вимірювань параметрів сигналів у порівнянні зі загальними методами оптимізації? Наведіть приклади.
3. Що таке методи послідовних наближень? У чому їх переваги та недоліки? Наведіть приклади.
4. Що таке методи перебору (сканування)? Їх переваги та недоліки.
5. Що таке методи інтерполяції? Їх переваги та недоліки. Самостійно одержати інтерполяційні алгоритми (16.11)-(16.13).
6. Що таке селективність вимірювальної системи? Як її можна охарактеризувати? Наведіть приклади.
7. Що таке сигнальна функція обробки? Як її вигляд впливає на ефективність вирішення завдань вимірювань?
8. На прикладі задачі вимірювання скалярного параметру самостійно довести, що оцінки ML є асимптотично незміщеними.

9. Самостійно одержати формулу (16.16).
10. Що таке функція невизначеності? Чим вона відрізняється від СФО?
11. Що складає основу теоретичного аналізу помилок вимірювання? Яку роль у цьому аналізі відіграє інформаційна матриця Фішера?
12. Самостійно одержати формулу (16.28), повторивши викладки (16.19) – (16.26) для задачі оцінювання скалярного параметру.
13. Що таке енергетичні та неенергетичні параметри сигналу? В чому полягає сенс подібної класифікації. Наведіть приклади.
14. Самостійно повторити викладки, що приводять до рівняння (16.31) щодо оптимальної МЛ оцінки параметрів. Розглянути приклади вирішення цього рівняння для задач як незалежного, так й спільного оцінювання фази та амплітуди сигналу.
15. Що таке сигнальна функція обробки (СФО)? Для чого вона використовується? Що таке функція невизначеності (ФН) та чим вона відрізняється від СФО?
16. Самостійно повторити викладки, що приводять до виразу (16.41). Що таке часова та просторова функція невизначеності? Одержати вираз та надати відповідні ілюстрації для ФН імпульсу із тональним заповненням за параметрами його затримки та зсуву частоти.
17. Самостійно повторити викладки, що приводять до виразів (16.48), (16.49). Які висновки витікають з їх розгляду?
18. Самостійно повторити викладки, що приводять до виразу (16.51). В чому полягає його принципова відмінність від виразу (16.48).
19. Самостійно показати, що для випадку оцінювання неенергетичних параметрів вираз (16.51) переходить у вираз (16.46).
20. З яких етапів складається побудова будь-якої вимірювальної системи? Прокоментувати зміст цих етапів.

Розділ 17. Базові принципи цифрової демодуляції сигналів.

У попередніх розділах ми розглянули статистичний опис завдань приймання сигналів та методи їх вирішення. Зокрема, у розділі 9 була дана постановка та класифікація таких завдань, у розділі 10 розглянуті математичні основи статистичних методів, у розділі 11 застосування цих методів для вирішення завдань виявлення сигналів, а у розділі 16 – стосовно завдань вимірювання параметрів сигналів. Тепер ми у значній мірі підготовлені для розгляду одного з найважливіших та складних класів завдань, а саме, демодуляції сигналів. Дійсно, про важливість цих завдань говорить хоча б надзвичайна поширеність та велике значення у сучасному світі телекомунікаційних систем, роботу яких у переважній більшості й забезпечують так звані модеми, тобто, поєднання модулаторів (на передавальній стороні) та демодуляторів (на приймальній стороні). У свою чергу, фізичні явища, що супроводжують передачу сигналів через середовища різних видів (радіоефір, дротові та кабельні канали, радіорелейні лінії, акустичне, оптичне середовище, тощо) надають цим завданням суттєвої складності.

В межах одного розділу розглянути весь комплекс питань, що відносяться до проблематики демодуляції сигналів, навряд чи можливо. Тому тут ми обмежимося розглядом основних принципів демодуляції сигналів, а розгляд більш складних питань, пов'язаних, у тому числі, із впливом фізичних факторів середовища передачі сигналу, продовжимо у третій частині Посібника. Для більш продуктивного засвоєння матеріалу даного розділу рекомендуємо попередньо повторити матеріали розділу 8, який присвячено розгляду методів цифрової модуляції.

17.1. Демодуляція сигналів як завдання перевірки статистичних гіпотез.

Почнемо наш розгляд із загальної схеми тракту передачі цифрової інформації каналами зв'язу, що представлена на рис.17.1. На вхід тракту подається дискретне повідомлення $\mathbf{X} = (x_1, \dots, x_N)$, тобто, послідовність деяких цифрових символів (букв, цифр, команд управління, тощо). Завданням тракту є доставка цього повідомлення споживачу інформації, причому, бажано без втрат або лише з такими втратами, які

© О.Я. Калужний

© КПІ ім. Ігоря Сікорського, 2021

споживач може визнавати допустимими. Першій блоку даної схеми виконує низку процедур, що готують цифрову інформацію до передачі через деяке фізичне середовище. До числа таких процедур, зокрема, відносяться інформаційне стиснення повідомлення, кодування та модуляція. Але у даному розділі нас буде цікавити, головним чином, цифрова модуляція (ЦМ). Інші процедури теж важливі, але для поточного розгляду вони поки що не є необхідними.

Методи ЦМ ми вже розглядали у розділі 8, тому тут обмежимося лише загальним описом його функцій. Отже, входом ЦМ є послідовність дискретних символів $\mathbf{X} = (x_1, \dots, x_N)$, а виходом – сигнал $s(t) = s(t/\mathbf{X})$, який пристосовано для передачі через деяке фізичне середовище. Цей сигнал й «несе» задане повідомлення $\mathbf{X}$.

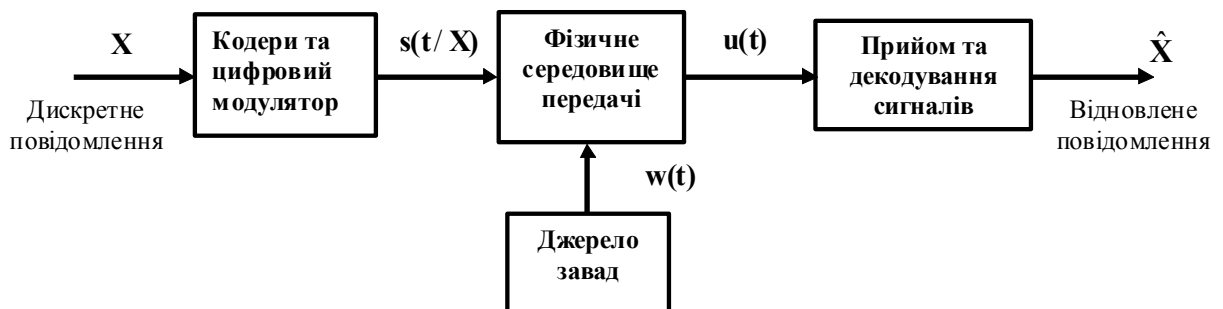


Рис.17.1. Загальна схема передачі цифрової інформації

Сформований у модуляторі сигнал $s(t/\mathbf{X})$ подається у фізичний канал передачі, де з ним можуть відбуватися перетворення, що зумовлені певними фізичними явищами, які є характерними для середовища заданого виду. Також у каналі практично завжди присутні шуми та завади, про що вже неодноразово йшла мова у попередніх розділах. Таким чином, після проходження через фізичне середовище на вході приймального пристрою будемо мати сигнал

$$u(t) = \tilde{s}(t/\mathbf{X}) + w(t), \quad (17.1)$$

де $\tilde{s}(t/\mathbf{X})$ - інформаційний сигнал, який, можливо, зазнав деяких перетворень у фізичному середовищі (їх часто називають спотвореннями), $w(t)$ - загальні шуми та завади на вході приймача. Завданням приймальних пристроїв (приймачів, декодерів, тощо) є відновлення первісного повідомлення $\mathbf{X}$. Але внаслідок спотворень сигналу та наявності у каналі шумів та завад приймальний пристрій насправді формує лише

оцінку повідомлення $\hat{X}$, яка у загальному випадку може не співпадати із повідомленням X . Тому тракт передачі звичайно намагаються побудувати так, щоб різниця поміж оцінкою та первісним повідомленням у будь-якому разі вкладалася у певні та, бажано, наперед задані границі. Цілком очевидно, що завдання такого класу відносяться до теорії оптимального (або, що теж саме, завадостійкого) приймання сигналів, основи якого ми розглядали у попередніх розділах. Але постановка таких завдань у системах передачі інформації має певну специфіку, деякі аспекти якої ми й обговоримо нижче.

Перш за все зауважимо, що внаслідок дискретності цифрової інформації загальне число повідомлень N_{Mess} на скінченному часовому інтервалі завжди обмежено. Зокрема, при передачі N символів, кожен з яких може приймати значення з алфавіту об'ємом M , усього можливо $N_{Mess} = M^N$ різних повідомлень. Отже, завдання демодуляції можна трактувати, як пошук істинної траєкторії на діаграмі переходів станів джерела, що представлена на рис.17.2. На цій діаграмі кожна позиція по горизонталі відповідає передачі чергового символу, рівні по вертикалі – певному значенню символу.

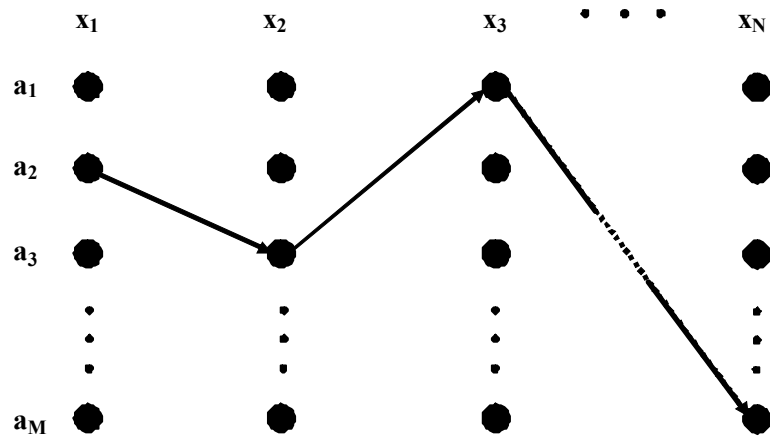


Рис.17.2. Діаграма переходів станів джерела

Нагадаємо також, що передача кожного символу відбувається на деякому часовому інтервалі фіксованої довжини T_s . При цьому m -ому значенню n -ого символу $x_n = a_{m_n}$ ($m = 1, \dots, M; n = 1, \dots, N$) ставиться у відповідність елементарний сигнал $s_{m_n}(t)$ тривалістю T_s , який формується зі скінченної множини елементарних

сигналів $\{s_m(t)\}_{m=1}^M$. Нагадаємо, що загальну кількість таких сигналів M називають позиційністю модуляції. Вигляд та параметри набору сигналів $\{s_m(t)\}_{m=1}^M$ визначається обраним методом цифрової модуляції.

Усі ці елементарні сигнали «вишиковуються» на виході модулятора у послідовність $\{s_{m_1}(t), s_{m_2}(t), \dots, s_{m_N}(t)\}$, що відповідає загальному повідомленню $\mathbf{X} = (x_1, \dots, x_N)$. Отже, відносно переданих сигналів також може бути побудована діаграма переходів, подібна до рис.17.2. Тільки у цьому разі станом слід називати передачу на поточному символному інтервалі відповідного сигналу $s_m(t)$. Відповідний приклад ми розглянемо у §17.4.

Якби спотворення сигналу та завади в каналі були відсутні, то кожному з можливих повідомлень відповідав би наперед відомий сигнал на вході приймача. Тоді завдання приймання звелось б до розпізнавання N_{Mess} детермінованих ситуацій. Але в реальності, внаслідок впливу факторів фізичного середовища та наявності шумів і завад, сигнали на вході приймача є випадковими. Тому кожному повідомленню може відповідати на вході приймального пристрою не один, а цілий ансамбль сигналів вигляду (17.1). Відповідно, рішення, які виносить приймач, також є випадковими, або, як прийнято говорити, статистичними. Отже, маємо завдання перевірки статистичних гіпотез щодо змісту повідомлення

$$H_p : \mathbf{X} = \mathbf{X}_p, \quad (p = 1, \dots, N_{Mess}).$$

У свою чергу, для його вирішення може бути застосовано математичний апарат, який розглядався у розділі 10. Нагадаємо, що відповідно до матеріалів вказаного розділу оптимальними правилами статистичних рішень є такі, що мінімізують баєсовий ризик (10.2). Тому такі правила прийнято називати баєсовими. У свою чергу, ризик (10.2) суттєво залежить від вибору матриці втрат $\|C_{ik}\|$ ($i, k = 1, \dots, N_{Mess}$), яка дозволяє, так би мовити, «матеріалізувати» ефективність застосування тих чи інших правил прийняття рішень. Специфікою систем передачі інформації є те, що матрицю втрат звичайно вибирають у вигляді (10.5). У цьому разі, як неважко бачити, баєсовий ризик (10.2) приймає вигляд

$$R = \sum_{i=1}^{N_{Mess}} \sum_{\substack{k=1 \\ k \neq i}}^{N_{Mess}} P(\mathbf{X}_i / \mathbf{X}_k) \cdot P(\mathbf{X}_k), \quad (17.2)$$

що, вочевидь, співпадає із середньою ймовірністю P_{ME} помилкових рішень щодо прийнятих повідомлень. Отже, головним критерієм оптимізації правил прийняття рішень у застосуванні до систем передачі інформації є, як правило, мінімізація середньої ймовірності помилок (17.2).

У розділі 10 було показано, що відповідне оптимальне правило має вигляд (10.7), яке у нашому випадку можна записати наступним чином:

$$\hat{\mathbf{X}} = \arg \left\{ \max_{\mathbf{X}_p} \left(P(\mathbf{X}_p / \mathbf{u}) \right) \right\}, \quad p = 1, 2, \dots, N_{Mess}, \quad (17.3)$$

де $P(\mathbf{X}_p / \mathbf{u})$ - апостеріорна ймовірність передачі p -го повідомлення. Нагадаємо, що критерій (17.3) прийнято називати критерієм максимуму апостеріорної ймовірності гіпотез або MAP (від англ. **M**aximum **A** posteriori **P**robability). Цілком очевидно, що даний критерій також є баєсовим, але лише для окремого випадку матриці втрат (10.5). Отже, оптимальне баєсове правило прийняття рішень у даному зводиться до обчислення на прийнятому масиві даних об'ємом N_{Mess} значень апостеріорних ймовірностей $P(\mathbf{X}_p / \mathbf{u})$. Рішення приймається на користь тієї гіпотези (повідомлення), якій відповідає максимальне значення $P(\mathbf{X}_p / \mathbf{u})$. У свою чергу, відповідно до формули Баєса вказана апостеріорна ймовірність може бути представлена у вигляді:

$$P(\mathbf{X}_p / \mathbf{u}) = \frac{p(\mathbf{u} / \mathbf{X}_p) \cdot P(\mathbf{X}_p)}{p(\mathbf{u})}, \quad (17.4)$$

де $p(\mathbf{u} / \mathbf{X}_p)$ - розподіл прийнятих даних у випадку, коли передається повідомлення $\mathbf{X}_p$, $P(\mathbf{X}_p)$ - апіорна ймовірність передачі даного повідомлення, $p(\mathbf{u})$ - безумовний розподіл спостережень $\mathbf{u}$. Якщо усі повідомлення є однаково ймовірними, то незалежно від значення $\mathbf{X}_p$ його апіорна ймовірність $P(\mathbf{X}_p) \equiv 1/N_{Mess}$. У цій ситуації із формули (17.4) витікає, що апостеріорна

імовірність $p(\mathbf{X}_p/\mathbf{u})$ з точністю до множників, які не залежать від повідомлення $\mathbf{X}_p$, співпадає з умовним розподілом $p(\mathbf{u}/\mathbf{X}_p)$. Нагадаємо, що цей розподіл прийнято називати функцією правдоподібності гіпотез для спостережень $\mathbf{u}$. Тоді, вочевидь, критерій MAP є тотожним до вже відомого нам критерію максимуму правдоподібності гіпотез або критерію ML (від англ. Maximum Likelihood), тобто:

$$\hat{\mathbf{X}} = \arg \left\{ \max_{\mathbf{X}_p} (P(\mathbf{u}/\mathbf{X}_p)) \right\}, \quad p = 1, 2, \dots, N_{\text{Mess}}. \quad (17.5)$$

Більше того, доведено [3, 17, 18], що критерій ML є асимптотично оптимальним (тобто, збігається до MAP) не тільки для рівномірного апіорного розподілу гіпотез, але для будь-яких інших апіорних розподілів, якщо спостереження є надійними. Стосовно завдань приймання сигналів це практично означає достатньо велике значення параметру SNR, тобто, $q_R^2 \gg 1$. Але слід все ж таки пам'ятати, що критерій ML не відноситься до класу баєсових, тобто, він не є строго оптимальним.

17.2. Проекційна (координатна) технологія цифрової демодуляції сигналів.

Отже, як вже було сказано вище, передача даних в системах із цифровою модуляцією відбувається на послідовних символьних інтервалах тривалістю $T_s = 1/V_s$, де V_s символьна швидкість, тобто, кількість символів в одиницю часу. На n -ому інтервалі ($n = 1, 2, \dots, N$) передається деяка порція інформації (група біт), якій поставлено у відповідність певний елементарний сигнал, що належить до заданого набору сигналів $\{s_m(t)\}_{m=1}^M$.

Важливою для подальшого розгляду обставиною є те, що усі можливі елементарні сигнали $s_m(t)$ можуть бути, як правило, представлені розкладами у функціональному базисі скінченної розмірності, тобто, для кожного з них можна записати:

$$s_m(t) = \sum_{k=1}^{N_s} c_{mk} f_k(t), \quad t \in (0, T_s), \quad (17.6)$$

де $\{f_k(t)\}_{k=1}^{N_s}$ - базис сигнального простору скінченної розмірності $N_s \leq M$, c_{mk} - коефіцієнти розкладу. У найпростішому випадку базис $\{f_k(t)\}_{k=1}^{N_s}$ може співпадати із множиною усіх можливих елементарних сигналів $\{s_m(t)\}_{m=1}^M$. Тоді, вочевидь, $N_s = M$, а коефіцієнти розкладу приймають значення

$$c_{mk} = \begin{cases} 0, & m \neq k; \\ 1, & m = k. \end{cases}$$

Але далі нас, в основному, буде цікавити нетривіальний випадок, коли розмірність базису $N_s < M$. При цьому будемо вважати, що базис залишається функціонально повним, тобто, скінчений ряд (17.6) не має залишку. З математичної точки зору елементами базису $\{f_k(t)\}_{k=1}^{N_s}$ може бути деяка множина лінійно незалежних функцій [9]. Але у прикладних задачах цей базис звичайно вибирають ортонормованим, тобто, його координатні функції мають задовольняти співвідношенню:

$$\int_0^{T_s} f_i(t) \cdot f_k(t) dt = \begin{cases} 0, & i \neq k; \\ 1, & i = k. \end{cases} \quad (17.7)$$

Користуючись співвідношенням (17.7), можна записати просту формулу для коефіцієнтів розкладу (17.6):

$$c_{mk} = \int_0^{T_s} s_m(t) \cdot f_k(t) dt, \quad k = 1, \dots, N_s. \quad (17.8)$$

Тепер застосуємо цей же базис для розкладу вже не самого сигналу, а його суміші зі завадами (17.1) на вході приймального пристрою. В результаті для n -ого символного інтервалу можна записати наступний ряд:

$$\tilde{u}_n(t) = \sum_{k=1}^{N_s} V_k(n) f_k(t - (n-1) \cdot T_s); \quad t \in ((n-1) \cdot T_s, n \cdot T_s), \quad (17.9)$$

де, по аналогії з (17.8), коефіцієнти розкладу обчислюються за формулою

$$V_k(n) = \int_0^{T_s} u_n(t) \cdot f_k(t) dt, \quad (17.10)$$

$u_n(t) = u(t + (n-1) \cdot T_s), t \in (0, T_s)$ - прийнята в межах n -го інтервалу суміш¹ інформаційного сигналу, шумів та завад (17.1). Неважко бачити, що розклад (17.9), на відміну від розкладу (17.6), має залишок, тобто, не є тотожним прийнятому сигналу $u_n(t)$. Але цей залишок визначається лише шумами і завадами, які, вочевидь, не містять корисної інформації. В цьому легко переконатися, якщо записати різницю поміж сигналом $u_n(t)$ та рядом (17.9):

$$u_n(t) - \tilde{u}_n(t) = w_n(t) - \sum_{k=1}^{N_s} W_k(n) e_k(t - (n-1) \cdot T_s), \quad (17.11)$$

де коефіцієнти розкладу шуму $W_k(n)$ мають вигляд аналогічний (17.10), тобто,

$$W_k(n) = \int_0^{T_s} w_n(t) \cdot f_k(t) dt, \quad (17.12)$$

$w_n(t)$ - шуми та завади на поточному інтервалі. Вказана обставина дозволяє зробити висновок, що заміна поточної реалізації сигналу на вході приймача $u_n(t)$ рядом (17.9) не призводить до будь-яких втрат корисної інформації. При цьому все інформаційне «навантаження» сигналу-носія переходить у коефіцієнти (17.10). У теоретичній статистиці у такому випадку прийнято говорити, що коефіцієнти $\{V_k(n)\}_{k=1}^{N_s}$ є достатніми статистиками² задачі. Отже, замість безпосередніх спостережень на вході приймача $u_n(t)$ можна для прийняття рішення використовувати координатні коефіцієнти $\{V_k(n)\}_{k=1}^{N_s}$. Така заміна не лише не зменшить, а навіть суттєво підвищить ефективність демодуляції, про що свідчить «шумовий» залишок (17.11), який формується внаслідок фільтраційного ефекту процедури (17.10). Оскільки коефіцієнти (17.10) з математичної точки зору є проекціями поточних спостережень $u_n(t)$ на сигнальний простір, який визначається базисом $\{f_k(t)\}_{k=1}^{N_s}$, то даний підхід у науково-технічній літературі часто називають проекційним або координатним. Цих назв ми й будемо дотримуватися далі.

¹ Далі, якщо це не викликає плутанини, будемо для стислості називати цю суміш прийнятим сигналом.

² Достатні статистики – функції (або функціонали) від масиву спостережень, що несуть усю інформацію, яка необхідна для вирішення статистичної задачі [6, 17]

Проекційний підхід є далеко не єдиним з можливих, але він, по-перше, досить загальний, по-друге, зручний для застосування сучасних модуляційних форматів. Слід також зазначити, що заміна безперервних реалізацій сигналів на вході приймального пристрою скінченою сукупністю їх проекцій (координатних коефіцієнтів) на сигнальний простір дозволяє оминати деякі математичні складності, які пов'язані з операціями у нескінченно мірних просторах, про що вже неодноразово згадувалося у попередніх розділах.

Отже, з вищесказаного витікає, що в рамках проекційного (координатного) підходу завдання демодуляції сигналів можна розділити на два окремих завдання:

- обчислення проекцій (17.10) прийнятих сумішей сигналів та завад $u_n(t)$ на сигнальний простір, який визначається базисом $\{f_k(t)\}_{k=1}^{N_s}$;
- прийняття на основі множини проекцій $\{V_k(n)\}_{k=1}^{N_s}$ ($n=1,2,\dots,N$) рішення про те, яке саме повідомлення $\mathbf{X}$ було передано.

На рис.17.3 представлена загальна структура системи демодуляції сигналів, яка, відповідно до вказаного розбиття завдань, складається з двох підсистем зі вказаними вище функціями. Зауважимо, що саме першу з цих підсистем часто й називають власне демодулятором. Другу ж підсистему частіше називають детектором [15] або пристроєм прийняття рішення. Вказаної термінології ми й будемо дотримуватися далі. Об'єднання ж підсистем будемо називати приймачем сигналів із цифровою модуляцією або просто цифровим приймачем.

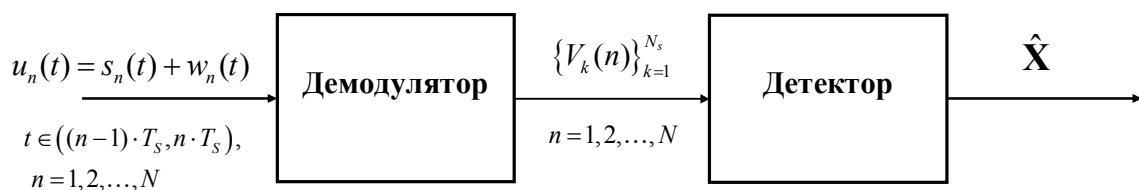


Рис.17.3. Загальна структура приймача сигналів із цифровою модуляцією.

Зауважимо, що представлене розбиття структури приймача є зручним й з точки зору застосування сучасних модуляційних форматів та методів їх демодуляції, зокрема, так званої кодової модуляції [15, 19], демодуляції в каналах із завмираннями, модуляції OFDM, тощо. Про деякі з них ми ще поговоримо пізніше.

Тепер розглянемо роботу даного приймача детальніше. Отже, функцією демодулятора у вказаній вище інтерпретації є обчислення коефіцієнтів розкладу (17.9) в обраному функціональному базисі. У відповідності з виразом (17.10) ця процедура зводиться до обчислення кореляції спостережень $u_n(t)$ на кожному із символічних інтервалів із координатними базисними функціями $\{f_k(t)\}_{k=1}^{N_s}$. Ми вже неодноразово зустрічалися зі схожими процедурами, наприклад, у розділі 11. Там же відзначалося, що пряме обчислення кореляції можна замінити еквівалентною процедурою фільтрації

$$V_k(n) = \int_0^{T_s} u_n(t) \cdot h_k(T_s - t) dt \quad (17.13)$$

де $h_k(\tau) = f_k(T_s - \tau)$ - імпульсний відгук фільтру, узгодженого із базисними функціями сигнального простору. Таким чином, основу демодулятора складають низка кореляторів (17.10) або узгоджених фільтрів (17.13). Наведемо деякі приклади.

Зокрема, широкого практичного застосування в наш час набули методи цифрової модуляції, в яких носіями інформації є амплітуда та фаза елементарних сигналів $s_m(t)$. Як було показано у розділі 8, для амплітудно-фазових методів модуляції ці елементарні сигнали можуть бути записані у наступному загальному вигляді:

$$s_m(t) = A_m \cdot g(t) \cdot \cos(\omega_0 t + \Phi_m), \quad m = 1, \dots, M, \quad (17.14)$$

де $\{A_m, \Phi_m\}$ - пара інформаційних параметрів, яка включає амплітуду та фазу елементарного сигналу. Ці параметри приймають значення із дискретного набору таких пар об'ємом M значень, який визначається заданим сигнальним сузір'ям (див. розділ 8). Також у виразі (17.14) фігурує сигнальна функція $g(t)$ тривалістю T_s , яку звичайно називають базовим імпульсом. Часто цей імпульс обирається прямокутним, але для деяких видів модуляції він може мати й більш складну форму.

Легко бачити, що будь-який з елементарних сигналів (17.14) можна представити суперпозицією усього лише двох базисних функцій $f_c(t) = g(t) \cdot \cos \omega_0 t$ та $f_s(t) = -g(t) \cdot \sin \omega_0 t$ оскільки

$$s_m(t) = A_m g(t) \cdot \cos(\omega_0 t + \Phi_m) = A_m \cos \Phi_m \cdot g(t) \cos \omega_0 t - A_m \sin \Phi_m \cdot g(t) \sin \omega_0 t.$$

Таким чином, сигнальний простір для вказаних методів модуляції є двовимірним, тобто, $N_s = 2$. Також легко бачити, що дані базисні функції є ортогональними на символьному інтервалі, оскільки

$$\int_0^{T_s} f_c(t) \cdot f_s(t) dt = 0,$$

а за умови відповідного нормування базового імпульсу $g(t)$ ще й ортонормованим. Це, зокрема, означає, що на кожному символьному інтервалі для прийняття рішення достатньо мати усього два коефіцієнти (17.10) розкладу (17.9), які мають у даному разі вигляд:

$$V_c(n) = \int_0^T u_n(t) \cdot g(t) \cdot \cos \omega_0 t dt, \quad V_s(n) = -\int_0^T u_n(t) \cdot g(t) \cdot \sin \omega_0 t dt. \quad (17.15)$$

Обчислення цих коефіцієнтів з урахуванням прийнятих припущень приводить до наступного результату:

$$V_c(n) = A_{m_n} \cos \Phi_{m_n} + W_c(n), \quad V_s(n) = A_{m_n} \sin \Phi_{m_n} + W_s(n), \quad (17.16)$$

де m_n - номер елементарного сигналу на n -ому інтервалі, $W_c(n)$ та $W_s(n)$ - шумові коефіцієнти розкладу, які визначаються виразом (17.12).

Співвідношення (17.15) фізично відповідають пропусканню прийнятого сигналу через фільтри, які узгоджені з базовим імпульсом $g(t)$ та налаштовані на центральну частоту сигналу ω_0 . Відповідна схема демодулятора сигналів з амплітудною та фазовою модуляцією представлена на рис.17.4.

Неважко впевнитися, що дана схема виконує операції оброблення сигналів у повній відповідності до виразів (17.15). Але до цієї схеми введені додаткові важливі елементи, про які ми ще досі не згадували.

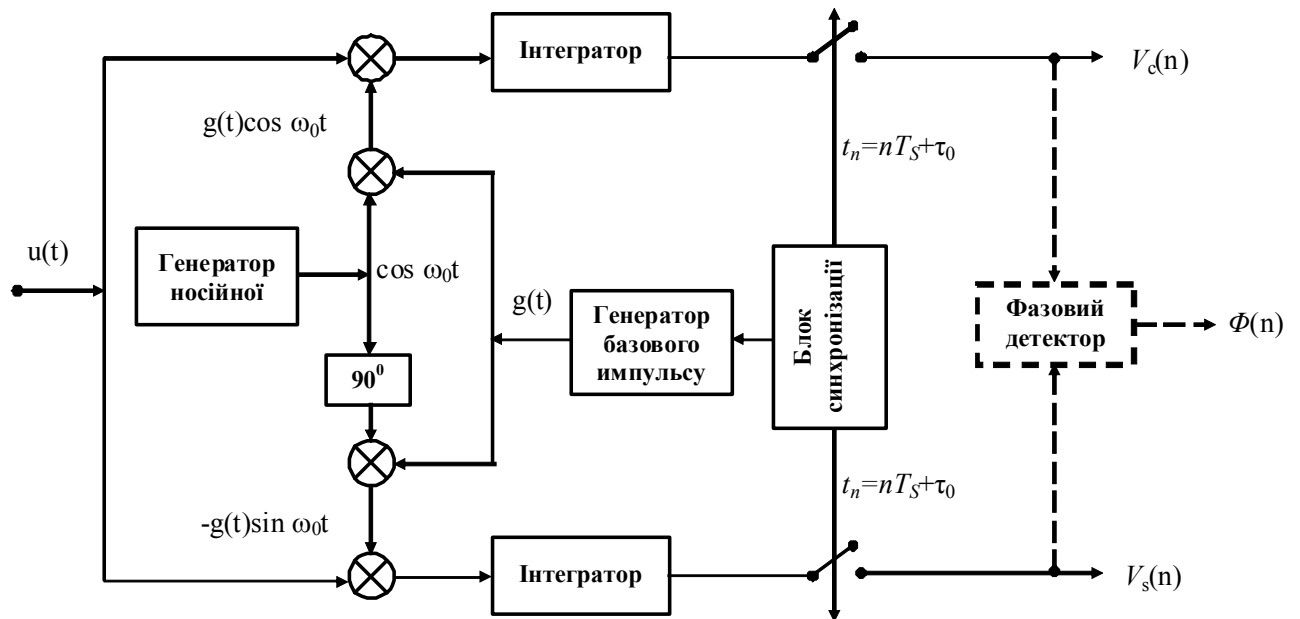


Рис.17.4. Демодулятор сигналів з амплітудною та фазовою модуляцією

Перш за все, це пристрої синхронізації. Взагалі, синхронізація є важливою та, мабуть, найскладнішою частиною будь-яких цифрових приймачів. Щодо схеми рис.17.4, то тут елементів синхронізації показано лише два³: це синхронізація носійної частоти сигналу та символна синхронізація. Синхронізація носійної є традиційним завданням для будь якого приймача радіо сигналів, як аналогових, так й цифрових. А от символна синхронізація – це характерна особливість саме цифрових пристроїв. Зокрема, у даному разі завдання символної синхронізації полягає у визначенні моментів часу закінчення кожного символного інтервалу з урахуванням невідомої затримки сигналу τ_0 у каналі та забезпечення зчитування у ці моменти виходів інтеграторів на вихід демодулятора.

Відмітимо ще одну важливу інтерпретацію схеми рис.17.4. Оскільки сигнальний простір для амплітудно-фазових методів модуляції є двовимірним, то вихід демодулятора рис.17.4 часто зручно трактувати як комплексні величини

$$\dot{V}(n) = V_c(n) + j \cdot V_s(n), \quad (17.17)$$

³ У сучасних телекомунікаційних протоколах, таких, наприклад, як DVB-S2, підсистем синхронізації іноді налічується більше десяти.

які можна відображати точками на фазовій площині. Також до цієї схеми часто вводять фазовий детектор (ФД), який обчислює фазу сигналу (17.17) на кожному символному інтервалі:

$$\Phi(n) = \arctg\left(\frac{V_s(n)}{V_c(n)}\right).$$

Схема рис.17.4 є базовою для побудови багатьох сучасних трактів приймання сигналів, причому, не тільки з амплітудно-фазовою модуляцією. Тому в різних джерелах вона відома під різними назвами. Наприклад, досить часто її називають демодулятором із розщепленням на квадратурні компоненти або просто квадратурним демодулятором. До можливих застосувань цієї схеми або подібних до неї ми ще повернемося трохи згодом.

Тепер перейдемо до розгляду другої підсистеми приймача рис.17.3, а саме, пристрою прийняття рішень або детектору. На відміну від попередньої підсистеми, яка лише виконує підготовку даних, цей пристрій у повній мірі має реалізувати статистичний підхід до вирішення завдання демодуляції у тій постановці, про яку йшла мова у §17.1. Слід лише взяти до уваги, що тепер спостереженнями є вектори⁴ проєкцій $\mathbf{V}(n)$, які складені з коефіцієнтів розкладу $V_k(n)$, $(k=1, \dots, N_s)$ прийнятого на відповідному символному інтервалі сигналу. У свою чергу, на основі прийнятих припущень та виразу (17.10) для векторів $\mathbf{V}(n)$ можна записати аналог рівняння спостережень (17.1) у просторі проєкцій:

$$\mathbf{V}(n) = \mathbf{c}_p(n) + \mathbf{W}(n), \quad n = 1, \dots, N; \quad p = 1, \dots, N_{Mess}, \quad (17.18)$$

де $\mathbf{c}_p(n)$ вектор, компонентами якого є коефіцієнти розкладу (17.8) сигнальної частини спостережень $\{c_{pk}(n)\}_{k=1}^M$, які відповідають p -ій гіпотези щодо переданого повідомлення $\mathbf{X} = \mathbf{X}_p$, $\mathbf{W}(n)$ - коефіцієнти розкладу (17.12) завадо-шумової частини спостережень. Тепер перейдемо до розгляду деяких правил прийняття рішень на основі спостережень (17.18).

⁴ Як й у попередніх розділах тут вектори розглядаються як матриці-стовпчики. Відповідно, усі операції над векторами мають виконуватися за правилами матричної алгебри.

17.3. Когерентне приймання повідомлень «у цілому», послідовна та посимвольна демодуляція.

Отже, відповідно до вказаної вище постановки, завдання детектора полягає у прийнятті рішення щодо переданого повідомлення на основі послідовності спостережень (17.18). Якщо рішення приймається спільно для усіх символів повідомлення на символьних інтервалах $n = 1, \dots, N$, які охоплюють загальний час передачі повідомлення $\mathbf{X}$, то відповідне правило часто називають демодуляцією (або прийманням) повідомлення «у цілому». Відповідно до критерію максимальної правдоподібності (17.5) для побудови такого правила необхідно записати спільний умовний розподіл $p(\mathbf{V}(1), \mathbf{V}(2), \dots, \mathbf{V}(N) / \mathbf{X})$ векторів (17.18), для кожної з можливих гіпотез щодо переданого повідомлення $\mathbf{X} = \mathbf{X}_p$, $p = 1, \dots, N_{\text{Mess}}$. Для цього перш за все слід конкретизувати модель каналу передачі. Припустимо, що маємо канал із адитивним білим гавсівським шумом (АБГШ). Відповідно до цієї моделі (див. §4.3) завадово-шумова складова прийнятого сигналу $w(t)$ розподілена за гавсівським законом із нульовим математичним очікуванням та кореляційною функцією

$$K_w(\tau) = \frac{g_w}{2} \cdot \delta(\tau),$$

де g_w - спектральна щільність потужності білого шуму в каналі, $\delta(\tau)$ - дельта-функція. Оскільки лінійне перетворення гавсівських величин від розподілу не змінює (див. розділ 3), то зі співвідношення (17.12) витікає, що вектори «шумових» коефіцієнтів $\mathbf{W}(n)$ також мають гавсівський розподіл з параметрами:

$$\langle \mathbf{W}(n) \rangle \equiv 0, \quad \langle \mathbf{W}(n) \cdot \mathbf{W}^T(m) \rangle = \frac{g_w}{2} \cdot I_{N_s} \cdot \delta_{nm}, \quad (17.19)$$

де $\mathbf{W}^T$ - вектор, транспонований до вектору $\mathbf{W}$, (тобто, відповідна матриця-рядок), I_{N_s} - одинична матриця порядку N_s , δ_{np} - символ Кронекера. Отже, компоненти $W_k(n)$ векторів шумових коефіцієнтів $\mathbf{W}(n)$ є некорельованими як поміж собою, так й поміж символьними інтервалами. У свою чергу,

некорельованість гавсівських величин означає, як відомо (див. розд.3), й їхню статистичну незалежність.

Тепер залишилося лише визначитися з властивостями сигналів по виходу каналу. Як вже було сказано вище, переданому повідомленню $\mathbf{X} = (x_1, \dots, x_N)$ цифровий модулятор ставить у взаємно однозначну відповідність послідовність сигналів $\mathbf{S} = \{s_{m_1}(t), s_{m_2}(t), \dots, s_{m_N}(t)\}$, де $s_{m_n}(t)$ - елементарний сигнал, який відповідає символу x_n . Припустимо, що спотворення сигналу в каналі відсутні. Тоді цей же сигнал без змін (за винятком, можливо, лише масштабу та часової затримки⁵) надходить на вхід приймального пристрою. Таким чином, як сам сигнал $\mathbf{S}$, так й вектори коефіцієнтів $\mathbf{c}_p(n)$ для кожної з можливих гіпотез про передане повідомлення є наперед відомими величинами. Вказані умови прийнято називати **когерентним прийманням** [3] на відміну від ситуацій, де канал змінює непередбачуваним чином як параметри, так й вигляд сигналу.

Отже, при когерентному прийманні єдиним джерелом випадковості у рівнянні спостережень (17.18) є вектори «шумових» коефіцієнтів $\mathbf{W}(n)$, які мають статистичні параметри (17.19). Беручи до уваги вказані параметри та статистичну незалежність компонент $W_k(n)$ цих векторів для функції правдоподібності векторів проекцій (17.18) можна записати⁶:

$$\begin{aligned} p(\mathbf{V}(1), \mathbf{V}(2), \dots, \mathbf{V}(N) / \mathbf{X}_p) &= \prod_{n=1}^N \frac{1}{\sqrt{\pi g_w}} \exp \left\{ \left| \mathbf{V}(n) - \mathbf{c}_p(n) \right|^2 / g_w \right\} = \\ &= \prod_{n=1}^N \frac{1}{\sqrt{\pi g_w}} \exp \left\{ \sum_{k=1}^{N_s} \left(V_k(n) - c_{pk}(n) \right)^2 / g_w \right\}. \end{aligned} \quad (17.20)$$

Відповідно до оптимального за критерієм ML правила (17.5) рішення приймається на користь тієї з гіпотез $\mathbf{X}_p$, ($p = 1, \dots, N_{Mess}$), для якої значення функції правдоподібності (17.20) є максимальним. Тому, як й у попередніх розділах, замість статистики (17.20) можна використовувати будь-яке її монотонне перетворення, що не змінює відношень порядку поміж величинами, зокрема, логарифм. З урахуванням

⁵ Поки що вважаємо, що зміну масштабу та часову затримку сигналу попередньо належним чином компенсовано.

⁶ Надалі ми будемо використовувати як векторні, так й розгорнуті записи в залежності від задачі.

цієї обставини оптимальним за критерієм ML рішенням є повідомлення $\mathbf{X}_q$, номер якого відповідає мінімуму тестової статистики

$$D_p = \sum_{n=1}^N |\mathbf{V}(n) - \mathbf{c}_p(n)|^2 = \sum_{n=1}^N \sum_{k=1}^{N_s} (V_k(n) - c_{pk}(n))^2, \quad (17.21)$$

$$p = 1, 2, \dots, N_{\text{Mess}},$$

де були відкинуті складові, що не залежать від гіпотези щодо повідомлення $\mathbf{X}_p$.

Отже, відповідно до виразу (17.21) рішення приймається на користь тієї з гіпотез, для якої забезпечується мінімальна середня **евклідова відстань** по усьому повідомленню поміж послідовностями векторів $\mathbf{V}(n)$ та $\mathbf{c}_p(n)$. Статистику (17.21) можна представити й у іншому вигляді. Для цього розкриємо у виразі (17.21) скобки та відкинемо члени, які не залежать від номеру гіпотез. В результаті одержимо:

$$Q_p = \sum_{n=1}^N \left(\mathbf{V}^T(n) \cdot \mathbf{c}_p(n) - \frac{1}{2} |\mathbf{c}_p(n)|^2 \right) = \sum_{n=1}^N \sum_{k=1}^{N_s} \left(V_k(n) \cdot c_{pk}(n) - \frac{1}{2} c_{pk}^2(n) \right), \quad (17.22)$$

$$p = 1, 2, \dots, N_{\text{Mess}}.$$

При цьому рішення приймається на користь того повідомлення, для якого тестова статистика (17.22) є максимальною. Цілком очевидно, що це правило є повністю еквівалентним попередньому, але використовує замість метрики Евкліда так звану **кореляційну метрику**, що іноді буває зручніше. Виразу (17.22) можна надати ще більш простого вигляду, якщо коефіцієнти розкладу інформаційного сигналу $c_{pk}(n)$ задовольняють умові

$$\sum_{k=1}^{N_s} c_{pk}^2(n) \equiv \text{const.}, \quad p = 1, 2, \dots, N_{\text{Mess}}, \quad (17.23)$$

тобто, повна енергія проєкцій не залежить від гіпотези щодо переданого повідомлення. Наприклад, умова (17.23) забезпечується для фазової та частотної модуляції, що буде проілюстровано нижче. У цьому випадку замість статистики (17.22) можна використовувати статистику

$$Q_p = \sum_{n=1}^N \mathbf{V}^T(n) \cdot \mathbf{c}_p(n) = \sum_{n=1}^N \sum_{k=1}^{N_s} V_k(n) \cdot c_{pk}(n), \quad (17.24)$$

$$p = 1, 2, \dots, N_{\text{Mess}}.$$

яка, з точки зору об'єму обчислень, виглядає дещо простіше, чим статистики (17.21) або (17.22).

Процедуру прийняття рішення за розглянутими вище правилами можна представити як перебір гіпотез у відповідності з діаграмою переходів станів джерела рис.17.2. Усього на даній діаграмі можливо $N_{Mess} = M^N$ повідомлень та відповідних гіпотез. Для кожного можливого повідомлення оптимальний детектор має накопичувати евклідові метрики відповідно до виразу (17.21) або кореляційні метрики (17.22) чи (17.24). В результаті одержимо N_{Mess} значень метрик, а рішення приймаємо на користь того повідомлення, якому буде відповідати мінімальна з них. Дане правило детектування звичайно називають **демодуляцією (або прийманням) повідомлень «у цілому»**.

На перший погляд дане правило виглядає доволі простим та інтуїтивно зрозумілим. Його безумовною перевагою є висока ефективність на рівні потенційно можливої. Особливо ця перевага є суттєвою, коли елементарні сигнали поміж символьними інтервалами якимось чином пов'язані поміж собою, тобто, залежні. Така залежність може бути, наприклад, наслідком застосування методів модуляції з пам'яттю (див. розділ 8) або виникати під дією фізичних факторів каналу передачі, про що ми поговоримо пізніше.

Але «каменем спотикання» для застосування оптимальних статистик (17.21), (17.22) чи (17.24) є загальна кількість переборів M^N , яка необхідна для демодуляції N символьного повідомлення. Легко бачити, що зі зростанням кількості символів N кількість переборів та відповідних обчислювальних затрат зростає дуже швидко. Тому практичне застосування цих правил якщо й можливе, то при доволі обмежених значеннях N . Тим не менш, ідея демодуляції повідомлень «у цілому» є важливим теоретичним орієнтиром, який впливає й на деякі практичні рішення. Зокрема, замість демодуляції «у цілому» на практиці іноді застосовують схожий метод, який називають **послідовна демодуляція**. Суть вказаного методу полягає в тому, що рішення щодо поточного символу приймається у «ковзаючому» вікні шириною в N_w інтервалів. При цьому є дві суттєві відмінності у порівнянні з демодуляцією «у цілому»:

- Розмір вікна N_w вибирається із врахуванням інтервалу залежності поміж символами та технічних можливостей певної апаратно-програмної платформи.
- Рішення приймається лише по одному, останньому символу. Рішення по наступному символу приймається після зсуву вікна на один інтервал.

Окремим випадком послідовної демодуляції є так звана **посимвольна демодуляція**, яка відповідає значенню ширини вікна $N_w = 1$. Це означає, що для кожного символьного інтервалу рішення приймається незалежно від попередніх символів. При цьому кількість конкуруючих гіпотез визначається лише позиційністю модуляції M , а для усього N символьного повідомлення таких гіпотез буде $N \cdot M$. Економія обчислювальних затрат у порівнянні з M^N переборами, яких потребує демодуляція «у цілому», дійсно вражає. Але відразу виникає питання: а чи допустиме таке спрощення з точки зору ефективності приймання?

Відповідь на це питання практично очевидна: помітного зниження ефективності не буде, якщо сигнали на різних символьних інтервалах ніяким чином між собою не пов'язані. У свою чергу, зв'язок поміж сигналами на різних символьних інтервалах відсутній, якщо, по-перше, застосовуються методи модуляції без пам'яті, по-друге, фізичний канал передачі пам'яті також не має. Ось для таких ідеалізованих умов вказаний метод демодуляції теоретично є близьким до оптимального. Але, не зважаючи на вказану ідеалізацію, посимвольна демодуляція має дуже широке практичне застосування, а її ефективність для багатьох практичних задач виявляється цілком прийнятною.

Запишемо оптимальне правило посимвольного детектування при когерентному прийманні сигналу. Діючи за тією ж схемою, що й вище, неважко прийти до наступного правила: на кожному поточному символьному інтервалі рішення слід приймати на користь того значення символу $x_n = a_{m_0}$, номер якого відповідає мінімуму евклідової дистанції

$$D_m = |\mathbf{V}(n) - \mathbf{c}_m|^2 = \sum_{k=1}^{N_s} (V_k(n) - c_{mk})^2, \quad (17.25)$$

$$m = 1, 2, \dots, M$$

По аналогії із виразом (17.22) можна також записати еквівалентне правило для кореляційної метрики:

$$Q_m = \mathbf{V}^T(n) \cdot \mathbf{c}_m - \frac{1}{2} |\mathbf{c}_m|^2 = \sum_{k=1}^{N_s} \left(V_k(n) \cdot c_{mk} - \frac{1}{2} c_{mk}^2 \right), \quad (17.26)$$

$$m = 1, 2, \dots, M,$$

де, на відміну від (17.25), рішення приймається за максимальним значенням статистики (17.26). Якщо ж при виконанні умови (17.23) відкинути центрувальний член, то одержимо ще більш просту тестову статистику:

$$Q_m = \mathbf{V}^T(n) \cdot \mathbf{c}_m = \sum_{k=1}^{N_s} V_k(n) \cdot c_{mk}, \quad (17.27)$$

$$m = 1, 2, \dots, M$$

Проілюструємо застосування цих правил на тому ж прикладі демодуляції сигналів з амплітудно-фазовою модуляцією, що вже розглядався вище. Як вже було сказано раніше, сигнальний простір у цьому випадку є двовимірним. Неважко бачити, що відповідні коефіцієнти розкладу сигнальної частини спостережень мають вигляд

$$c_{mc} = A_m \cdot \cos \Phi_m, \quad c_{ms} = A_m \cdot \sin \Phi_m. \quad (17.28).$$

Отже, тестова статистика (17.25)

приймає вигляд

$$Q_m = (V_c(n) - A_m \cos \Phi_m)^2 + (V_s(n) - A_m \sin \Phi_m)^2, \quad (17.29)$$

а еквівалентна йому статистика (17.26)

запишеться наступним чином:

$$Q_m = V_c(n) \cdot A_m \cos \Phi_m + V_s(n) \cdot A_m \sin \Phi_m - \frac{1}{2} A_m^2 \quad (17.30)$$

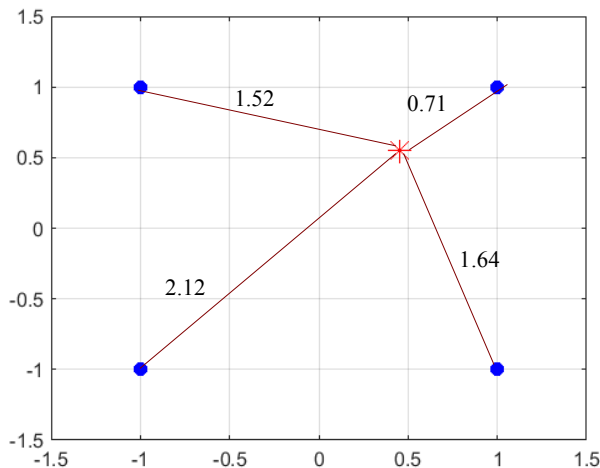


Рис.17.5. Сузір'я QPSK та проєкції прийнятого сигнал

Для ілюстрації на рис.17.5 представлено на фазовій площині деякий умовний випадок приймання сигналу із модуляцією QPSK. Кружечками позначені «штатні» точки сигнального «сузір'я» QPSK, які в даному разі є можливими гіпотезами щодо переданого символу, а зірочкою – фактичний результат приймання на деякому символному інтервалі. Там же показані відстані між поточним символом та точками «сузір'я» у метриці Евкліда.

Відповідно до правила (17.29) рішення приймається на користь тієї гіпотези (або точки «сузір'я»), до якої результат приймання є найближчим за метрикою Евкліда. У даному разі це, вочевидь, точка сигнального «сузір'я» з координатами (1,1). Втім, у даному простому прикладі можна було обійтися й без обчислення метрик: достатньо лише визначити квадрант, куди попадають проекції сигналу. Саме так для модуляції QPSK і діють практично. Цей же принцип може бути застосований й для інших методів амплітудно-фазової модуляції з тією лише різницею, що конфігурація областей для прийняття рішення стає дещо складнішою. У той же час, є немало практичних задач, в яких без прямого обчислення метрик вже не обійтися.

17.4. Демодуляція за алгоритмом Вітербі.

Алгоритм Вітербі (AB) було запропоновано у 1967 р. для ефективного декодування згорткових кодів, що передаються мережами зв'язку при наявності шуму. В наш час цей алгоритм (його ще іноді називають пошук Вітербі) становить основу багатьох інформаційних технологій і не лише у сфері свого початкового застосування. Взагалі кажучи, АВ може застосовуватися для оцінки найбільш ймовірної послідовності подій при певних загальних припущеннях щодо причинно-наслідкових зв'язків між подіями та їх можливими відображеннями. Зокрема, однією зі сфер застосування АВ можуть бути задачі послідовної демодуляції. Це, по суті, ще один можливий спосіб для практичного втілення ідеї демодуляції повідомлення «у цілому», яка розглядалася вище.

Отже, як вже було сказано у §17.1, демодуляцію сигналу можна трактувати як пошук найбільш вірогідної траєкторії на діаграмі переходу станів джерела рис.17.1.

Безпосередній пошук такого шляху для N -символьного повідомлення джерела, що має алфавіт об'ємом M , потребує M^N переборів. Для реальних значень параметрів M та N ця кількість переборів у більшості випадків значно перевищує можливості як сучасних, так й перспективних обчислювальних засобів. АВ дозволяє суттєво зменшити кількість таких переборів за рахунок використання додаткової інформації щодо властивостей джерела сигналів. Зокрема, такою інформацією може бути наявність певних періодичностей у зміні станів джерела, наявність заборонених або детермінованих переходів, тощо. Таким чином, перед застосуванням АВ треба перш за все проаналізувати наявну інформацію про характеристики джерела.

Стосовно завдання демодуляції, яке тут розглядається, джерелом сигналу є вихід каналу зв'язку. У свою чергу, цей вихід формується низькою попередніх пристроїв та систем, з яких ми обмежимося лише аналізом функцій цифрового модулятора (ЦМ). Нагадаємо, що головною функцією ЦМ є співставлення символів повідомлення $\mathbf{X} = (x_1, \dots, x_N)$ послідовності сигналів $\mathbf{S} = \{s_{m_1}(t), s_{m_2}(t), \dots, s_{m_N}(t)\}$, де $s_{m_n}(t)$ - елементарний інформаційний сигнал, який відповідає символу x_n та вибирається зі скінченної множини сигналів $\{s_m(t)\}_{m=1}^M$ об'ємом M . При когерентному прийманні фізичний канал передачі не порушує взаємно-однозначної відповідності поміж множинами $\mathbf{X}$ та $\mathbf{S}$. Отже, діаграмі переходів станів джерела рис.17.1 можна поставити у відповідність діаграму зміни станів інформаційних сигналів. Розглянемо деякі приклади. При цьому мову доцільно вести лише про ЦМ із пам'яттю, оскільки при модуляції без пам'яті, як вже було сказано вище, достатньо ефективним та простим рішенням є посимвольна демодуляція.

Одним із найбільш простих методів ЦМ з пам'яттю (див. розділ 8, ч.2) є відносна (вона же диференційна або різницева) фазова модуляція (ВФМ). У випадку бінарної ВФМ правило формування інформаційного сигналу по виходу ЦМ виглядає наступним чином: якщо на поточному інтервалі передачі на вхід ЦМ надходить символ «0», то ЦМ формує елементарний сигнал $s_0(t)$, а у випадку подачі

символу «1» формується протилежний сигнал $-s_0(t)$. Логіка роботи модулятора ВФМ можна відобразити діаграмою станів, що показана на рис.17.6.

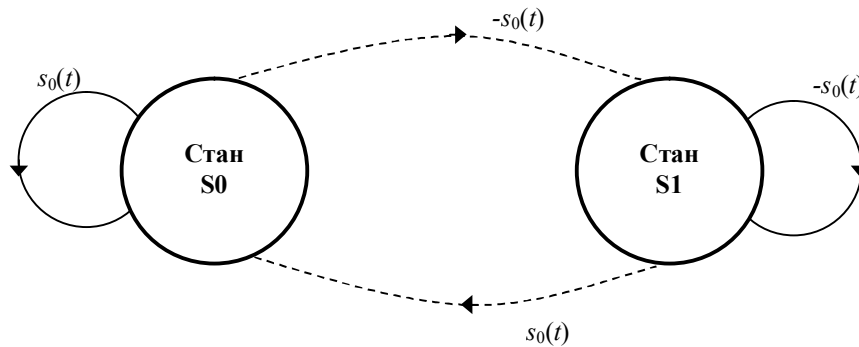


Рис.17.6. Діаграма станів модулятора ВФМ (суцільні лінії – перехід по коду «0», пунктир – по коду «1»).

Як бачимо із цієї діаграми, ЦМ може перебувати у двох станах: стан S0, при якому на вихід подається сигнал $s_0(t)$ та стан S1, коли по виходу формується сигнал $-s_0(t)$. Подача символу «0» стан ЦМ не змінює, подача «1» змінює стан на протилежний.

Тепер застосуємо проекційний підхід, що розглядався у §17.2. У даному прикладі маємо, вочевидь, лише одну базисну функцію $s_0(t)$ та одну проекцію прийнятого спостереження на сигнальний простір, яка, відповідно до виразу (17.10), має вигляд:

$$V(n) = \int_0^{T_s} u_n(t) \cdot s_0(t) dt.$$

Для сигналу без завад єдиний коефіцієнт розкладу приймає значення $c_0 = 1$ при передачі символу $x_n = 0$ та $c_1 = -1$ у випадку передачі символу $x_n = 1$. Отже, при накопиченні метрик Евкліда за правилом (17.21) до суми слід занести, в залежності від гіпотези, або дистанцію $(V(n) - 1)^2$ або $(V(n) + 1)^2$. Якщо ж застосувати кореляційну метрику, то оскільки центрувальний член $c_k^2 / 2 \equiv 1/2$ у даному прикладі від гіпотези не залежить, то до суми (17.24) заносяться відповідно $V(n)$ або $-V(n)$.

Тепер можна безпосередньо розглянути застосування АВ для демодуляції ВФМ. Головною особливістю підходу Вітербі до вирішення подібних задач є те, що

було принципово змінено опис характеристик джерела, а саме, замість діаграм станів, подібних до рис.17.6, Вітербі запропонував так звані треліс-діаграми⁷ (від англ. *trellis* – решітка), які зберігаючи усі можливості діаграми станів, дозволяли бачити процес у динаміці, тобто, зміну станів у часі. Власне, до їх числа можна віднести й діаграму рис.17.1, яку ми розглядали раніше. Але діаграма рис.17.1 демонструє лише саму можливість відстеження переходів станів джерела у часі, тоді як підхід Вітербі більш спеціалізований та конкретний. Зокрема, на рис.17.7 показана треліс-діаграма модулятора ВФМ, яка, по суті, може бути одержана розгорткою у часі діаграми станів рис.17.6.

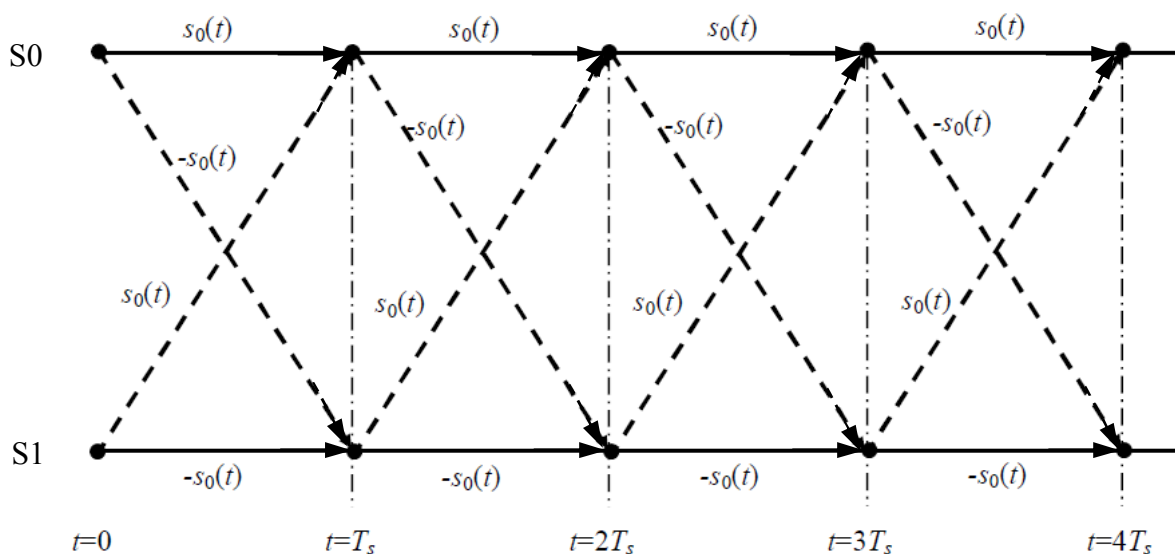


Рис. 17.7. Треліс-діаграма модулятора бінарної ВФМ(суцільні лінії – перехід по коду «0», пунктир – по коду «1»)

На діаграмі горизонтальні лінії відповідають двом можливим станам модулятора ВФМ: передача символу «0» (стан S0) та передача символу «1» (стан S1). Вузли на цих лініях відповідають закінченню поточного та початку наступного символного інтервалу. Стрілками показана зміна станів. Тут же позначено сигнали, який ЦМ при цьому передає в канал. Неважко бачити, що зберігаючи усі можливості діаграми станів, треліс-діаграма чітко проявляє періодичність переходів станів у часі, що є наслідком скінченої пам'яті пристрою. Зокрема, у даному разі структура переходів повторюється через кожні два символних інтервали.

⁷ В деяких вітчизняних академічних джерелах застосовується назви решітчаста діаграма або діаграма решітки. Але в сучасному інженерному середовищі назва треліс-діаграма вже давно стала загальноновживаною.

Саме вибір зручного та наочного відображення динаміки зміни станів дискретних автоматів зі скінченою пам'яттю у вигляді треліс-діаграм дозволив Вітербі зробити своє відкриття. Суть його полягає в тому, що після перетину кількох шляхів у деякому вузлі з огляду на періодичність треліс-діаграми сенс продовжувати далі має лише один шлях, якому відповідає найбільш сприятлива на момент перетину вузла метрика. Його Вітербі назвав таким, що вижив. Інші шляхи, що прийшли до того ж вузла з гіршими метриками, продовжувати далі не має сенсу: вони у будь-якому разі вже не можуть бути кращими за той шлях, що вижив.

Повертаючись до нашого прикладу зазначимо, що до кожного вузла на діаграмі рис. 17.6 підходять два шляхи. Отже, далі продовжувати доцільно лише один шлях із кращою метрикою. Таким чином, кількість необхідних переборів у даному прикладі скорочується у двічі. Зазначимо, що даний приклад дуже простий та розглядається лише з навчальною метою оскільки для практичної демодуляції сигналів ВМФ звичайно застосовуються простіші методи. Але при зростанні позиційності модуляції ефект від застосування АВ може бути значно більшим [15].

17.5. Некогерентна демодуляція.

Основним припущенням, яке лежить в основі розглянутих вище методів когерентної демодуляції сигналів, є те, що фізичний канал передачі на суттєві характеристики сигналу не впливає. Оскільки вказане припущення виглядає, на перший погляд, нереалістично, то може створюватися враження, що когерентні методи мають суто теоретичне значення та є досить далекими від практики. Насправді ж реальна ситуація виглядає, як то іноді кажуть, «з точністю до навпаки». Дійсно, одним з головних та найбільш видатних досягнень теорії та техніки зв'язку останніх десятиріч стало створення технологій когерентного приймання сигналів у каналах передачі із випадковими характеристиками. В наш час вказані технології, які іноді також називають **квазі-когерентним прийманням сигналу**, займають панівне положення як у теорії, так й практиці зв'язку.

Про деякі методи квазі-когерентного приймання в каналах із випадковими характеристиками ми детальніше поговоримо у наступному розділі. У даному ж параграфі зосередимося на розгляді альтернативного підходу, який прийнято

називати некогерентною демодуляцією або **некогерентним прийманням сигналів**. На противагу когерентним методам тут вже у самій початковій постановці завдання приймання сигналів робиться припущення про те, що сукупність деяких характеристик сигналу η на вході приймальної системи невідома та не піддаються прогнозуванню. Тому єдине, що залишається робити, це побудувати такі схеми приймання, які б до цих не прогнозованих характеристик сигналу були би нечутливі. Звісно, що така нечутливість приймальної системи «безкоштовно» не дається, а «платою» за неї, як правило, є погіршення ефективності приймання у порівнянні з когерентними схемами. Але не дивлячись на цей суттєвий недолік, методи некогерентного приймання зберігають певне практичне значення й донині.

При розгляді некогерентного приймання будемо спиратися на матеріали попередніх розділів, зокрема, §11.3, де вже розглядалися схожі методи, але стосовно завдання виявлення сигналів із випадковими параметрами. Зокрема, у §11.3 зазначалося, що у теоретичній статистиці застосовуються два основних підходи для забезпечення нечутливості системи оброблення сигналів до певних параметрів задачі: узагальнений баєсовий підхід та узагальнений метод максимальної правдоподібності. У першому випадку застосовується статистичне усереднення умовної функції правдоподібності (ФП) спостережень по сукупності не прогнозованих параметрів η , у другому – використовується їх оцінка $\hat{\eta}$, одержана на тій же реалізації сигналу, яка використовується для виділення корисної інформації. Але в будь-якому разі першим важливим кроком до побудови некогерентних схем приймання сигналів є визначення умовної ФП прийнятих сигналів для фіксованого припущення щодо параметрів η . Саме цією задачею, перш за все, ми й займемося нижче.

Як витікає із матеріалів §11.3, для вирішення вказаної задачі доцільно використовувати аналітичні представлення сигналів. Нагадаємо, що аналітичними у радіотехніці прийнято називати комплексні сигнали з одностороннім спектром (див. §1.3). Такі сигнали за своєю інформативністю нічим не поступаються звичайним дійсним сигналам, але є більш зручними як для теоретичного аналізу, так й для технічної реалізації алгоритмів приймання сучасними засобами обчислювальної

техніки. Отже, відтворимо ланцюжок викладок та міркувань, що вже наводилися у §17.2 та §17.3, але з відповідними корективами. Почнемо, як звичайно, із рівняння спостережень, яке, на відміну від (17.1), тепер має вигляд:

$$\dot{u}(t) = \tilde{s}(t/\mathbf{X}) + \dot{w}(t).$$

Тут усі величини мають той же фізичний зміст, що й у виразі (17.1), але математично відносяться до класу комплексних аналітичних сигнальних функцій. Належність до цього класу будемо, як й раніше, позначати крапкою над відповідними величинами. Перший додаток цього рівняння – аналітичне представлення сигналу, який пройшов через фізичний канал передачі та зазнав непередбачуваних змін. Конкретний вигляд можливих змін визначається обраною моделлю каналу передачі. У свою чергу, вибір та обґрунтування моделі каналу – це доволі складне завдання, оскільки діапазон умов функціонування сучасних систем радіозв'язку є надзвичайно широким. Тому більш ґрунтовний експертський уявлення про фізичні моделі каналів планується у третій частині нашого Посібника, а поки що, як й у §11.3, обмежимося традиційними («канонічними») моделями, в яких вплив каналу обмежується, як правило, випадковими змінами фази та амплітуди сигналу. Вказані зміни сигналу зручно визначити через випадковий комплексний коефіцієнт передачі сигналу

$$\dot{\eta}_{ch} = \varepsilon_{ch} \cdot \exp(-j \cdot \varphi_{ch}), \quad (17.31)$$

де ε_{ch} – коефіцієнт передачі за амплітудою, φ_{ch} – зсув фази у каналі. При цьому будемо вважати, що характеристика каналу (17.31) на усьому інтервалі передачі повідомлення хоч й невідома, але не змінюється у часі. Таким чином, в рамках даної моделі каналу рівняння спостережень може бути записано у вигляді:

$$\dot{u}(t) = \dot{\eta}_{ch}^* \cdot \dot{s}(t/\mathbf{X}) + \dot{w}(t), \quad (17.32)$$

де $\dot{s}(t/\mathbf{X})$ – комплексний інформаційний сигнал по виходу каналу із ідеальною передачею, який для стислості будемо далі називати опорним. У свою чергу, опорний сигнал, як й раніше, складається із послідовності точно відомих для кожної з можливих гіпотез елементарних сигналів $\dot{s}_m(t)$ з тією лише різницею, що тепер ці сигнали будуть комплексними.

Припустимо, як й у §17.2, що вказані елементарні сигнали можуть бути представлені координатним розкладом, аналогічним (17.6), тобто:

$$\dot{s}_m(t) = \sum_{k=1}^{N_s} \dot{c}_{mk} \dot{f}_k(t), \quad t \in (0, T_s), \quad (17.33)$$

де, на відміну від (17.6), базисні функції $\{\dot{f}_k(t)\}_{k=1}^{N_s}$ та коефіцієнти розкладу $\{\dot{c}_{mk}\}_{k=1}^{N_s}$ є комплексними. Також припустимо, що по аналогії з (17.7) базисні функції задовольняють умові ортонормованості, але тепер в унітарному просторі⁸:

$$\int_0^{T_s} \dot{f}_k(t) \cdot \dot{f}_i^*(t) dt = \begin{cases} 0, & i \neq k; \\ 1, & i = k. \end{cases} \quad (17.34)$$

Використовуючи дане співвідношення, можна записати наступний вираз для коефіцієнтів розкладу (17.33):

$$\dot{c}_{mk} = \int_0^{T_s} \dot{s}_m(t) \cdot \dot{f}_k^*(t) dt, \quad k = 1, \dots, N_s. \quad (17.35)$$

Далі, як й у §17.2, скористаємося базисом $\{\dot{f}_k(t)\}_{k=1}^{N_s}$ для розкладу прийнятої суміші сигналів та завад (17.32) на кожному із символічних інтервалів:

$$\tilde{u}_n(t) = \sum_{k=1}^{N_s} \dot{V}_k(n) \dot{f}_k(t - (n-1) \cdot T_s); \quad t \in ((n-1) \cdot T_s, n \cdot T_s), \quad (17.36)$$

де

$$\dot{V}_k(n) = \int_0^{T_s} \dot{u}_n(t) \cdot \dot{f}_k^*(t) dt \quad (17.37)$$

- відповідні коефіцієнти розкладу, $\dot{u}_n(t)$ - аналітичне представлення прийнятої на поточному інтервалі суміші сигналу та завад. На основі міркувань, аналогічних тим, що вже наводилися у §17.2, неважко прийти до висновку, що коефіцієнти (17.37) є достатніми статистиками задачі. Отже, для прийняття рішення на поточному символічному інтервалі слід мати лише повну сукупність проєкцій (коефіцієнтів) $\dot{V}(n) = \{\dot{V}_k(n)\}_{k=1}^{N_s}$ прийнятих спостережень на сигнальний простір.

⁸ Нагадаємо, що унітарним називають [9] простір, аналогічний евклідовому, який визначається на множині комплексних чисел. З практичної точки зору головна відмінність унітарного простору від евклідового в тому, що при виконанні операцій скалярного добутку векторів треба належним чином враховувати знаки комплексного спряження.

Безпосередньо реалізації сигналу на вході приймальної системи при цьому, вочевидь, вже не потрібні.

При розгляді прикладних аспектів даної теорії може виникнути природне запитання: а як же практично обчислювати проекції (17.37)? Адже на вході приймального пристрою, яке працює у реальному фізичному середовищі, маємо лише дійсні сигнали, а в усіх формулах цього параграфу фігурують сигнали аналітичні, тобто, комплексні? Для відповіді на це запитання слід знову звернутися до матеріалів розділу 1, зокрема, §§1.3, 1.4. Зокрема, там було роз'яснено поняття комплексної обвідної та обґрунтована проста схема (рис.1.3) її формування для сигналів із відносно вузькою смугою частот. Спираючись на вказані матеріали, представимо сигнал на вході приймального пристрою та базисні функції у наступному вигляді:

$$\dot{u}_n(t) = \dot{U}_n(t) \cdot \exp\{j\omega_0 t\}, \quad \dot{f}_k(t) = \dot{F}_k(t) \cdot \exp\{j\omega_0 t\},$$

де $\dot{U}_n(t)$ та $\dot{F}_k(t)$ - комплексні обвідні прийнятого сигналу та базисних функцій відповідно. Після підстановки цих виразів до формули (17.37) маємо:

$$\dot{V}_k(n) = \int_0^{T_s} \dot{U}_n(t) \cdot \dot{F}_k^*(t) dt. \quad (17.38)$$

Таким чином, маємо еквівалентну до виразу (17.37) формулу для обчислення проекцій прийнятого сигналу на базисний простір через комплексні обвідні. Як вже було сказано вище, комплексна обвідна прийнятої суміші сигналу та завад $\dot{U}_n(t)$ може бути одержана за допомогою схеми рис.1.3 (розділ 1). Тому залишається лише врахувати процедуру фільтрації (17.38), яка має бути узгоджена із базисними функціями. Отже, поєднуючи схему рис.1.3 з алгоритмом фільтрації (17.38) та беручи до уваги властивості комплексних обвідних сигналу, приходимо до схеми, представленої на рис.17.8.

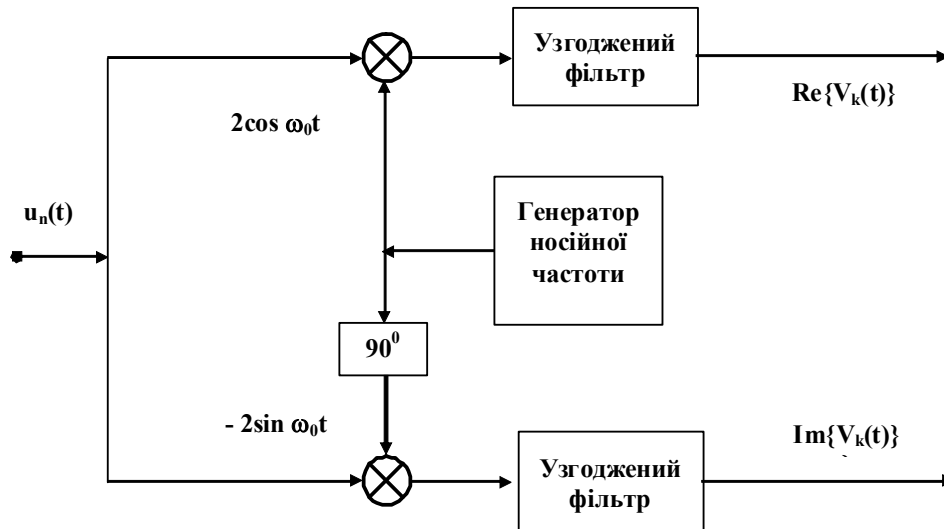


Рис.17.8. Формування проєкцій прийнятого сигналу на координатний базис

Дана схема загалом подібна до схеми рис.17.4, що вже розглядалася вище. Але головною відмінністю є те, що в схемі рис. 17.4 використовуються корелятори, а в даній схемі – фільтри низьких частот, які узгоджені з базисними функціями сигнального простору. Отже, кількість таких схем визначається розмірністю сигнального простору, яка фактично залежить від кількості підносійних частот, що використовуються модулятором. Що ж стосується узгоджених фільтрів, то у їх якості найчастіше використовуються фільтри підійнятого косинусу або подібні до нього, про що вже йшла мова у розділі 8. Нагадаємо, що їх головною особливістю є відсутність міжсимвольної інтерференції (МСІ).

Тепер повернемося до розгляду правила оптимального детектування символів, обмежившись при цьому по символною демодуляцією. Як й у випадку когерентної демодуляції, для формування відповідного правила прийняття рішення треба записати ФП щодо вектору коефіцієнтів $\dot{\mathbf{V}}(n) = \{\dot{V}_k(n)\}_{k=1}^{N_s}$ на поточному символному інтервалі, тобто, їх умовний розподіл для деякої гіпотези стосовно переданого символу $p(\dot{\mathbf{V}}(n)/x_n = a_m)$, $m = 1, \dots, M$. Але тепер необхідно враховувати випадковість не лише шумів та завад, але й коефіцієнту передачі каналу $\dot{\eta}_{ch}$. Як вже

зазначалося вище, для цього, перш за все, треба записати умовну ФП при фіксованому коефіцієнті передачі каналу $\dot{\eta}_{ch}$, тобто, $p(\dot{\mathbf{V}}(n)/x_n = a_{m_n}; \dot{\eta}_{ch})$.

З цією метою, перш за все, запишемо рівняння спостережень у просторі проєкцій. По аналогії з виразом (17.18) та із врахуванням співвідношень (17.32), (17.37) маємо:

$$\dot{\mathbf{V}}(n) = \dot{\eta}_{ch} \cdot \dot{\mathbf{c}}_m(n) + \dot{\mathbf{W}}(n), \quad (17.39)$$

де $\dot{\mathbf{c}}_m(n)$ - вектор, компонентами якого є коефіцієнти (17.35) розкладу опорного сигналу на поточному символьному інтервалі для m -ої гіпотези щодо значення поточного символу x_n , $\dot{\mathbf{W}}(n)$ - вектор коефіцієнтів розкладу шуму, компоненти якого по аналогії з (17.12) знаходяться за формулою

$$\dot{W}_k(n) = \int_0^{T_s} \dot{w}_n(t) \cdot \dot{f}_k^*(t) dt, \quad (17.40)$$

де $\dot{w}_n(t)$ - аналітичне представлення шумів та завад на поточному інтервалі. При фіксованому значенні параметру каналу $\dot{\eta}_{ch}$ єдиним джерелом випадковості у рівнянні спостережень (17.39) є шумові коефіцієнти (17.40). Отже, треба визначити їхні ймовірнісні властивості. Для цього, у свою чергу, слід згадати характеристики комплексних випадкових процесів і величин та характеристики шуму в каналі з АБГШ (див. §4.3, §4.4). Використовуючи вказані матеріали, неважко бачити, що дійсна $W_k(n)$ та уявна $\hat{W}_k(n)$ частини кожного із «шумових» коефіцієнтів

$$\dot{W}_k(n) = W_k(n) + j \cdot \hat{W}_k(n)$$

розподілені за гавсівським законом із нульовими математичними очікуваннями та однаковими дисперсіями, які дорівнюють $g_w/2$. Поміж собою величини $W_k(n)$ та $\hat{W}_k(n)$ некорельовані, що у випадку гавсівського розподілу означає й їхню статистичну незалежність. Отже, для спільного розподілу цих величин можна записати:

$$p_W(W_k(n), \hat{W}_k(n)) \triangleq p(\dot{W}_k(n)) = \frac{1}{\pi g_w} \exp \left\{ -\frac{1}{g_w} |\dot{W}_k(n)|^2 \right\}.$$

Беручи до уваги властивості шуму в каналі АБГШ та ортонормованість сигнального базису неважко бачити, що компоненти шумового вектору $\dot{\mathbf{W}}(n) = \{\dot{W}_k(n)\}_{k=1}^{N_s}$ поміж собою також статистично незалежні. Це дає можливість записати наступний вираз для їхнього спільного розподілу:

$$p_W(\dot{\mathbf{W}}(n)) = \frac{1}{(\pi g_w)^{N_s}} \exp \left\{ -\frac{1}{g_w} \sum_{k=1}^{N_s} |\dot{W}_k(n)|^2 \right\} = \frac{1}{(\pi g_w)^{N_s}} \exp \left\{ -\frac{1}{g_w} |\dot{\mathbf{W}}(n)|^2 \right\}.$$

Тепер, маючи розподіл «шумових коефіцієнтів та враховуючи рівняння спостережень (17.39), можна записати умовний розподіл вектору проєкцій на поточному символічному інтервалі:

$$\begin{aligned} p(\dot{\mathbf{V}}(n) / x_n = a_m; \dot{\eta}_{ch}) &= p_W(\dot{\mathbf{V}}(n) - \dot{\eta}_{ch} \cdot \dot{\mathbf{c}}_m(n)) = \\ &= \frac{1}{(\pi g_w)^{N_s}} \exp \left\{ -\frac{1}{g_w} |\dot{\mathbf{V}}(n) - \dot{\eta}_{ch} \cdot \dot{\mathbf{c}}_m(n)|^2 \right\}. \end{aligned} \quad (17.41)$$

Наступний крок полягає у виключенні залежності ФП від невідомого параметру $\dot{\eta}_{ch}$, для чого, як вже було сказано вище, застосовуються два основних підходи:

- узагальнений байєсовий підхід, який в даному разі передбачає статистичне усереднення виразу (17.41) по параметру $\dot{\eta}_{ch}$;
- метод узагальненої правдоподібності, в якому замість невідомого параметру $\dot{\eta}_{ch}$ використовується його оцінка $\hat{\dot{\eta}}_{ch}$, що одержана на тому ж вектору спостережень (17.39).

Перший із вказаних підходів було досить детально розглянуто у вже неодноразово згадуваному §11.3 стосовно завдання виявлення сигналів. Як можна бачити з матеріалів §11.3, суттєвим недоліком байєсового підходу є необхідність висунення певних, але, не завжди переконливих припущень, щодо ймовірнісних властивостей параметру $\dot{\eta}_{ch}$. У §11.3 було розглянуто три «канонічні» моделі каналів та для кожного із них отримано відповідні результати. До того ж, математичні викладки, які знадобилися для цього розгляду, виявилися не дуже простими та доволі громіздкими. Тому у даному параграфі ми обмежимося розглядом лише методу узагальненої правдоподібності, який виглядає як більш сучасним, так й більш простим.

Отже, для застосування методу узагальненої правдоподібності слід, перш за все, побудувати оцінку параметру $\hat{\eta}_{ch}$ по вектору спостережень (17.39). Для цього, як було показано у розділі 16, слід знайти точку максимуму ФП (17.41) по цьому параметру. Необхідною умовою максимуму є рівність нулю першої похідної виразу (17.41) або будь-якого монотонного перетворення від нього. Застосовуючи до виразу (17.41) логарифм та відкидаючи складові, що від η_{ch} не залежать, маємо наступне рівняння:

$$\frac{d}{d\eta_{ch}} \left| \dot{\mathbf{V}}(n) - \eta_{ch} \cdot \dot{\mathbf{c}}_m(n) \right|^2 = 0.$$

Після виконання диференціювання одержимо рівняння

$$-2 \cdot \dot{\eta}_{ch} \cdot \left| \dot{\mathbf{c}}_m(n) \right|^2 + 2 \cdot \dot{\mathbf{V}}^H(n) \cdot \dot{\mathbf{c}}_m(n) = 0,$$

де $(\cdot)^H$ позначає операцію ермітового спряження матриць, тобто, поєднання транспонування та комплексного спряження. Вирішуючи останнє рівняння, знаходимо оцінку невідомого параметру для кожної з можливих гіпотез щодо переданого символу:

$$\hat{\eta}_{ch} = \frac{\dot{\mathbf{V}}^H(n) \cdot \dot{\mathbf{c}}_m(n)}{\left| \dot{\mathbf{c}}_m(n) \right|^2}. \quad (17.42)$$

Після підстановки оцінки (17.42) у вираз (17.41), переходу до логарифму та відкидання неінформативних складових, приходимо до наступного правила некогерентного детектування символів, а саме: рішення слід приймати на користь того значення символу $x_n = a_{m_0}$, номер якого відповідає максимуму тестової статистики

$$Q_m = \left| \dot{U}_m(n) \right|^2 / C_m, \quad m = 1, 2, \dots, M; \quad (17.43)$$

де

$$\dot{U}_m(n) = \dot{\mathbf{V}}^H(n) \cdot \dot{\mathbf{c}}_m(n) = \sum_{k=1}^{N_s} \dot{V}_k^*(n) \cdot \dot{c}_{mk}(n) \quad (17.44)$$

- скалярний добуток вектору проекцій спостережень та вектору коефіцієнтів очікуваного сигналу, який відповідає m -ій гіпотезі щодо значення переданого символу,

$$C_m = |\dot{\mathbf{c}}_m(n)|^2 = \sum_{k=1}^{N_s} |\dot{c}_{mk}(n)|^2 \quad (17.45)$$

- нормуючі коефіцієнти.

Якщо порівняти співвідношення (17.43) – (17.45) із тестовою статистикою посимвольної когерентної демодуляції (17.26), то легко бачити, що в обох випадках основою цих правил є скалярний добуток векторів проекцій прийнятого сигналу та векторів коефіцієнтів розкладу очікуваного сигналу для відповідного набору гіпотез. Розходження поміж когерентним та некогерентним детектуванням по цій частині – лише у формі представлення сигналів: у когерентному випадку – це дійсні величини, у некогерентному – комплексні. Але найбільш значущою різницею є подальший етап обробки, а саме: в алгоритмі (17.26) скалярний добуток безпосередньо використовується для прийняття рішення, а в некогерентному алгоритмі (17.43) – лише після обчислення квадрату модулю комплексної величини (17.44). Остання операція й є тією особливістю, завдяки якій забезпечується нечутливість алгоритму некогерентної демодуляції до параметрів каналу передачі в рамках прийнятої моделі. Зокрема, саме для забезпечення виконання цієї операції й був потрібен перехід від дійсних до комплексних представлень сигналу при застосуванні некогерентної демодуляції.

Відзначимо також різницю у врахуванні залежності енергії проекцій очікуваного сигналу від гіпотези щодо переданого символу. В алгоритмі (17.26) для цього застосовується центрування виходу корелятору, а в тестовій статистиці (17.43) – нормування. Втім, ця різниця й зовсім зникає при застосуванні методів модуляції, для яких у випадку дійсних сигналів виконується умова (17.23), а при застосуванні комплексних сигналів аналогічна їй умова

$$C_m = \sum_{k=1}^{N_s} |\dot{c}_{mk}(n)|^2 \equiv const.. \quad (17.46)$$

Ми розглянули лише один з двох можливих підходів до побудови правила некогерентного детектування символів. Але із літературних джерел [3, 15, 19] можна зробити висновок, що узагальнений баєсовий підхід приводить до схожих результатів: основу алгоритмів у всіх випадках складає корелятор (17.44). Деякі відмінності мають місце лише на наступних етапах оброблення. Сподіваємося, що уважні та старанні читачі вже в змозі перевірити ці висновки самостійно.

Тепер розглянемо деякі приклади, які ілюструють можливі застосування некогерентної демодуляції. Почнемо з методів амплітудно-фазової модуляції, що вже розглядалися вище. Як було показано у §17.2, основу демодуляторів для цих методів модуляції становить схема рис.17.4, виходом якої на кожному символному інтервалі є комплексна проекція $\dot{V}(n)$, тобто, у комплексному представленні сигнальний простір для цих видів модуляції є одномірним. Відповідний коефіцієнт розкладу очікуваного сигналу з урахуванням виразів (17.28) має вигляд

$$c_m = A_m \exp\{j\Phi_m\}, m=1, \dots, M.$$

Отже, цей вираз та співвідношення (17.44), (17.45) приводять до наступного виразу щодо виходу корелятора:

$$\dot{U}_m(n) = \dot{V}^*(n) \cdot A_m \exp(j\Phi_m), \quad C_m = A_m^2. \quad (17.47)$$

Після підстановки виразів (17.47) до тестової статистики (17.43) маємо:

$$Q_m = \frac{|\dot{U}_m(n)|^2}{C_m} = \frac{|\dot{V}^*(n) \cdot A_m \exp(j\Phi_m)|^2}{A_m^2} \equiv |\dot{V}(n)|^2.$$

Як легко бачити, ця статистика не має ніякої селективності щодо завдання розрізнення гіпотез. Іншими словами, при застосуванні амплітудно-фазових методів модуляції правило некогерентної демодуляції (17.43) є непридатним. Втім, це цілком логічний наслідок відмови від врахування фазової інформації у каналі.

Але слід взяти до уваги те, що мова до сих пір йшла про методи посимвольного детектування. Якщо ж для прийняття рішення використовувати не один, а кілька символних інтервалів, то зроблений вище висновок вже не є таким однозначним. Розглянемо, наприклад, детектування символів відносної фазової модуляції (ВФМ), про яку йшла мова у §17.4. При цьому методі модуляції передача одного символу

відбувається фактично за два символних інтервали оскільки інформація «закодована» у різниці фаз сигналу поміж цими інтервалами. Тоді для некогерентного детектування символів ВФМ часто використовується [15] тестова статистика наступного вигляду:

$$Q_m(n) = \operatorname{Re} \{ \dot{V}(n-1) \cdot \dot{V}^*(n) \cdot \exp(-j\Phi_m) \} \quad (17.48)$$

де $\dot{V}(n-1), \dot{V}(n)$ - проєкції спостережень на одномірний сигнальний простір, що формуються схемою рис.17.4, для двох послідовних символних інтервалів, Φ_m - зсуви фаз сигналу поміж інтервалами, які відповідають набору можливих гіпотез щодо переданого символу. Зазначимо, що правило (17.48) та подібні йому, часто називають [15] **автокореляційним прийманням**, оскільки тут фактично в ролі опорного колювання використовується сигнал, прийнятий на попередньому інтервалі. Втім, у літературних джерелах немає єдиної думки щодо того, чи слід відносити подібні алгоритми до методів некогерентної демодуляції.

На завершення даного параграфу розглянемо ще один приклад, який ілюструє, мабуть, найбільш популярне застосування некогерентної демодуляції. Мова піде про системи передачі, які використовують маніпуляцію частотним зсувом, тобто, FSK (див. розділ 8, ч.І). Для цього виду модуляції використовуються елементарні сигнали наступного вигляду:

$$s_m(t) = A \cdot \cos(\omega_m t), \quad m = 1, \dots, M;$$

де ω_m - кругова частота сигналу, який відповідає передачі m -го символу, A - амплітуда сигналу. Даному дійсному сигналу відповідає, вочевидь, аналітичний сигнал

$$\dot{s}_m(t) = A \exp\{j\omega_m t\}, \quad m = 1, \dots, M. \quad (17.49)$$

Виберемо в якості базису сигнального простору повну сукупність таких сигналів, тобто, m -та базисна функція $\dot{f}_m(t) = \dot{s}_m(t)$. Тоді сигнальний простір для даного прикладу є M -вимірним. Припустимо також, що нормування сигналів за амплітудою та набір частот $\omega_1, \dots, \omega_M$ обрано таким чином, щоб задовольнялася умова ортонормованості базису (17.34), тобто:

$$\int_0^{T_s} \dot{s}_m(t) \cdot \dot{s}_k^*(t) dt = \begin{cases} 0, & m \neq k; \\ 1, & m = k. \end{cases}$$

Отже, проєкції спостережень на поточному символічному інтервалі слід обчислювати за формулою:

$$\dot{V}_k(n) = \int_0^{T_s} \dot{u}_n(t) \cdot \dot{s}_k^*(t) dt = \int_0^{T_s} \dot{u}_n(t) \cdot \exp\{-j\omega_k t\} dt.$$

У свою чергу, коефіцієнти розкладу очікуваного сигналу, що відповідають m -ій гіпотезі, мають, вочевидь, вигляд

$$\dot{c}_{mk} = \delta_{mk},$$

де, нагадаємо, δ_{mk} - символ Кронекера, $m, k = 1, \dots, M$. Беручи до уваги усі ці співвідношення, легко бачити, що вихід корелятора (17.44) у даному випадку співпадає із відповідною проєкцією спостережень, тобто,

$$\dot{U}_m(n) = \dot{V}^H(n) \cdot \dot{c}_m(n) = \dot{V}_m^*(n),$$

а нормувальна константа (17.45) $C_m \equiv 1$. Тестова статистика (17.43) приймає у цих умовах наступний вигляд:

$$Q_m = |\dot{V}_m(n)|^2, \quad m = 1, 2, \dots, M.$$

Для обчислення цієї статистики можна використовувати набір квадратурних демодуляторів, аналогічних тим, що були показані вище на рис.17.8. Кількість таких демодуляторів відповідає кількості M можливих частот сигналів FSK. Отже, некогерентна демодуляція у цьому випадку зводиться до відбору максимального за потужністю каналу приймання.

Зазначимо, що дана схема виявилась дуже простою та практично очевидною. Але слід відмітити, що ця простота є наслідком прийнятого вище припущення про ортогональність елементарних сигналів (17.49). Якби це було не так, то алгоритм демодуляції був би дещо складнішим.

Контрольні завдання та запитання до розділу 17.

Дати в конспекті письмові відповіді на наступні запитання:

1. Чому завдання демодуляції сигналів складають розділ теорії оптимального (завадостійкого) приймання сигналів? Які характерні особливості мають ці завдання?
2. Як формуються завдання демодуляції сигналів у термінах теорії перевірки статистичних гіпотез? Який зміст для таких завдань має баєсовий ризик? Що таке критерії MAP та ML? Який з цих критеріїв є оптимальним, а який квазіоптимальним та чому?
3. Що таке базис сигнального простору та яким вимогам він має задовольняти? Як виглядає цей базис у випадку застосування методів цифрової амплітудної та фазової модуляції?
4. В чому полягає суть проекційної (координатної) технології цифрової демодуляції сигналів? Які переваги дає ця технологія у теоретичному та практичному відношенні?
5. Чому коефіцієнти розкладу прийнятої суміші інформаційного сигналу, шумів та завад по сигнальному базису називають проекціями спостережень? Чому ці проекції в завданнях демодуляції є достатніми статистиками? Пояснити схему рис.17.3.
6. Пояснити роботу схеми рис.17.4. Яку роль у цій схемі відіграють підсистеми синхронізації? В чому полягає відмінність синхронізації аналогових та цифрових приймальних пристроїв?
7. Що таке когерентна демодуляція? Пояснити правила когерентної демодуляції (17.21) – (17.24). Чому ці правила відносять до методів приймання повідомлень «у цілому»? Які переваги та недоліки цих методів?
8. Що таке послідовна та посимвольна демодуляція? Особливості їх застосування, переваги та недоліки.
9. Представити схему та алгоритм когерентної посимвольної демодуляції сигналів QPSK. Проілюструвати її роботу засобами моделювання у середовищі Matlab-Simulink.
10. Теж саме, що й у попередньому завданні, але для модуляції PSK-8.
11. Теж саме, що й у попередньому завданні, але для модуляції КАМ-16.
12. Що таке алгоритм Вітербі? В яких завданнях він застосовується? В чому полягають його переваги? Що таке треліс-діаграма цифрового модулятора?
13. Пояснити застосування алгоритму Вітербі для демодуляції сигналів із бінарною та квадратною модуляцією ВФМ.

14. Що таке квазі-когерентна та некогерентна демодуляція сигналів? Який із цих методів в наш час застосовується частіше? В чому полягає основний принцип некогерентної демодуляції? Її переваги та недоліки.
15. Пояснити призначення та роботу схеми рис.17.8.
16. Пояснити правило некогерентної посимвольної демодуляції (17.43). З яких міркувань воно одержано? Де воно може, а де не може бути застосовано?
17. Намалювати схему некогерентної посимвольної демодуляції сигналів FSK та пояснити її роботу. Проілюструвати роботу цієї схеми засобами моделювання у середовищі Matlab-Simulink.

Розділ 18. Аналіз ефективності демодуляції сигналів в обчислювальному середовищі Matlab.

Аналіз ефективності демодуляції складає один із важливих етапів розробки систем передачі інформації. Такий аналіз пов'язано, як правило, із необхідністю проведення доволі складних обчислень та/або математичного моделювання, що потребує значних часових і трудових затрат. Виконання завдань аналізу суттєво полегшують сучасні програмні додатки, серед яких найбільш розповсюдженими та загально визнаними є засоби обчислювального середовища Matlab та асоційованого із ним пакету візуально-орієнтованого програмування Simulink. Розгляду деяких можливостей вказаних засобів й присвячено даний розділ. Вважаємо, що із основними можливостями середовища Matlab та пакету Simulink читачі вже попередньо ознайомлені¹.

18.1. Теоретичний аналіз ефективності за допомогою інструменту BERTool.

На фізичному рівні телекомунікаційних систем загальноприйнятими показниками ефективності демодуляції сигналів є середні ймовірності символних P_{SE} та бітових P_{BE} помилок. Перша з них визначається як відношення середньої кількості невірно розпізнаних символів до загального числа переданих, другий визначається аналогічно, але по відношенню до передачі біт. Ці показники, взагалі кажучи, пов'язані між собою, але єдиної універсальної формули для такого зв'язку не існує. При певних припущеннях [15] користуються наближеним теоретичним співвідношенням

$$P_{SE} \approx k_b \cdot P_{BE}, \quad (18.1)$$

де k_b - кількість біт, що припадає на один символ. Але у загальному випадку обчислення ймовірностей P_{SE}, P_{BE} все ж таки розглядається як окремі завдання. Найчастіше з цих двох показників прийнято застосовувати аналіз ймовірності бітових помилок P_{BE} (загально прийняте коротке позначення BER – від англ. Bit

¹ Більш детальну інформацію щодо цих можливостей та прийомів роботи можна знайти у відповідних довідкових файлах середовища Matlab.

Error Rate). Це дає змогу порівнювати у зіставних умовах різні за побудовою технології передачі.

Ймовірності P_{SE} та P_{BE} обчислюються, як правило, у функціональній залежності від важливого параметру завадостійкості систем передачі інформації, а саме, відношення сигнал/шум на біт q_b^2 . Цей параметр визначається відповідно до співвідношення² [19]

$$q_b^2 = E_b / N_0, \quad (18.2)$$

де E_b - енергія сигналу, що припадає на передачу одного біта, N_0 - спектральна щільність потужності шумів. Оскільки основною одиницею передачі, як правило, є елементарний символ, що несе $k_b = \log_2 M$ біт інформації, де M - позиційність модуляції, то параметр (18.2) можна пов'язати з іншим параметром, а саме, відношенням сигнал/шум на символ, який визначається у відповідності до співвідношення

$$q_s^2 = E_s / N_0, \quad (18.3)$$

де $E_s = E_b \cdot k_b$ - енергія елементарного сигналу, що відповідає передачі одного символу. З порівняння виразів (18.2) та (18.3) легко бачити, що

$$q_b^2 = q_s^2 / k_b. \quad (18.4)$$

Параметри завадостійкості (18.2) і (18.3) можна також пов'язати з традиційним для аналогової техніки відношенням сигнал/шум за потужністю

$$q_p^2 = P_s / \sigma_w^2. \quad (18.5)$$

де P_s - середня потужність сигналу на інтервалі його існування тривалістю T_s , σ_w^2 - дисперсія шуму в робочій смузі ΔF_s . Зокрема, відношення сигнал/шум на символ (18.3) може бути представлено через традиційний параметр (18.5) співвідношенням

$$q_s^2 = E_s / N_0 = P_s T_s / (\sigma_w^2 / \Delta F_s) = q_p^2 \cdot B_s, \quad (18.6)$$

² У світовій науково-технічній та навчальній літературі, у тому числі й у документації до середовища Matlab, для наведених нижче параметрів завадостійкості часто використовуються дещо інші позначення, про які буде сказано нижче. Тут застосовано такі позначення, які відповідають за стилем попереднім розділам Посібника.

де $B_s = \Delta F_s \cdot T_s$ - база сигналу³. Звідси, враховуючи визначення (18.4), неважко знайти аналогічний зв'язок й для параметру (18.2), тобто:

$$q_b^2 = q_P^2 \cdot B_s / k_b. \quad (18.7)$$

Обчислення ймовірностей P_{SE} , P_{BE} у середовищі Matlab забезпечують програмні засоби інструментального пакету **Communication Systems Toolbox** (надалі для стислості будемо називати цей пакет як **Comm**). Даний пакет містить низку програмних функцій Matlab та блоків Simulink, які забезпечують повний комплекс засобів для вирішення завдань як теоретичного аналізу, так й комп'ютерного моделювання систем передачі. Більш детально з деякими можливостями цих інструментальних пакетів ми ознайомимося пізніше, а поки що розглянемо зручний засіб оперативного аналізу ефективності систем передачі, яким є графічна оболонка BERTool. Для її виклику достатньо набрати в основному вікні Матлаб команду *bertool*, в результаті виконання якої з'явиться основне робоче вікно BERTool. Якщо у цьому вікні натиснути кнопку *Plot*, то одержимо результат, що показано на рис.18.1.

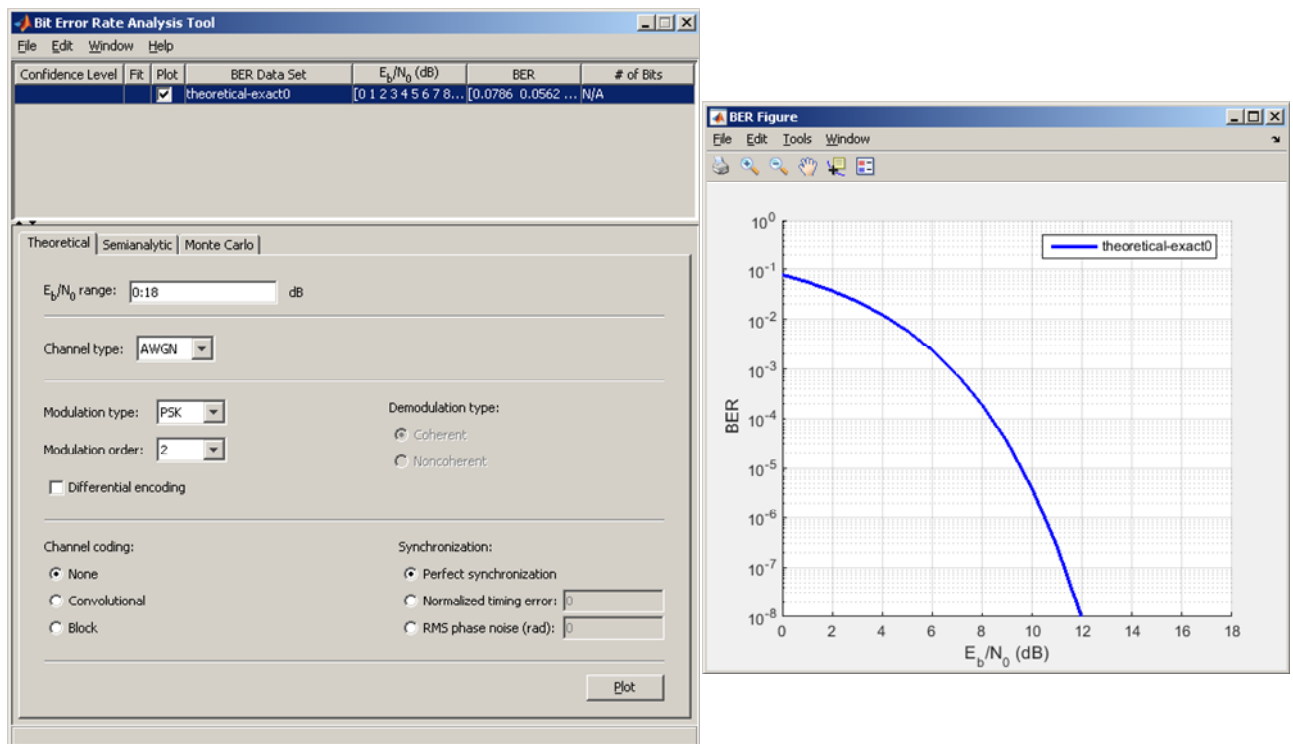


Рис.18.1. Графічна оболонка BERTool (зліва) та результати обчислень середньої ймовірності бітових помилок (графік справа) від відношення сигнал/шум на біт.

³ Для вузько смугових методів модуляції звичайно приймають $B_s \approx 1$.

На цьому рисунку зліва показано головне вікно інструменту BERTool. Це вікно поділяється на дві зони, а саме, на верхній частині показано набори даних, для яких у поточному сеансі вже були зроблені відповідні розрахунки (цю зону називають браузер даних). На нижній частині вікна визначається режим роботи та параметри обчислень. Усього передбачено три режима роботи, яким відповідають наступні вкладки вікна BERTool:

- Theoretical – режим теоретичного аналізу на основі відомих обчислювальних формул;
- Semianalytic – напів аналітичний (гібридний) режим, в якому застосовуються як відомі обчислювальні формули, так й частково математичне моделювання;
- Monte Carlo – режим повного математичного моделювання.

Особливо відзначимо перший з цих режимів, у якому реалізована переважна більшість відомих аналітичних формул для обчислення BER. Слід зазначити, що у математичному відношенні обчислення ймовірностей P_{SE} та P_{BE} є доволі складним завданням, яке має враховувати низку факторів. До них, зокрема, відносяться тип каналу передачі, вид модуляції, її позиційність, наявність завадостійкого кодування, якість виконання цифрової синхронізації, тощо. Значним досягненням розробників даного інструменту є те, що їм вдалося зібрати «до купи» переважну більшість відомих аналітичних формул для обчислення бітових та символьних помилок, які охоплюють найбільш практично важливі набори параметрів⁴. Це значно полегшує працю розробників систем передачі, які позбавлені від необхідності «копирсатися» у численних літературних джерелах, не усі з яких є широко доступними. До цього ще додається необхідність самостійно розробляти відповідні обчислювальні процедури і програми, що теж не є простою справою. Зазначимо, що у разі необхідності усі математичні подробиці можна знайти у першоджерелах⁵.

Графік, який знаходиться у правій частині рис.18.1, одержано саме у теоретичному режимі. Як видно з нижньої частини вікна інструменту BERTool, цей

⁴ Усі ці формули, до речі, наведено у довідкових файлах до BERTool

⁵ <https://www.mathworks.com/help/comm/ug/selected-bibliography-for-measurements.html#measbib002>

графік відповідає наступним вихідним даним: застосовано бінарну фазову маніпуляцію (тобто, BPSK або, що те ж саме, PSK2), демодуляція сигналу когерентна, каналне кодування відсутнє, синхронізація цифрового приймача ідеальна. Змінюючи вказані параметри, можна дослідити вплив на ефективність демодуляції тих чи інших факторів. Зокрема, на рис.18.2 показані результати порівнювальних розрахунків для декількох значень позиційності модуляції PSK.

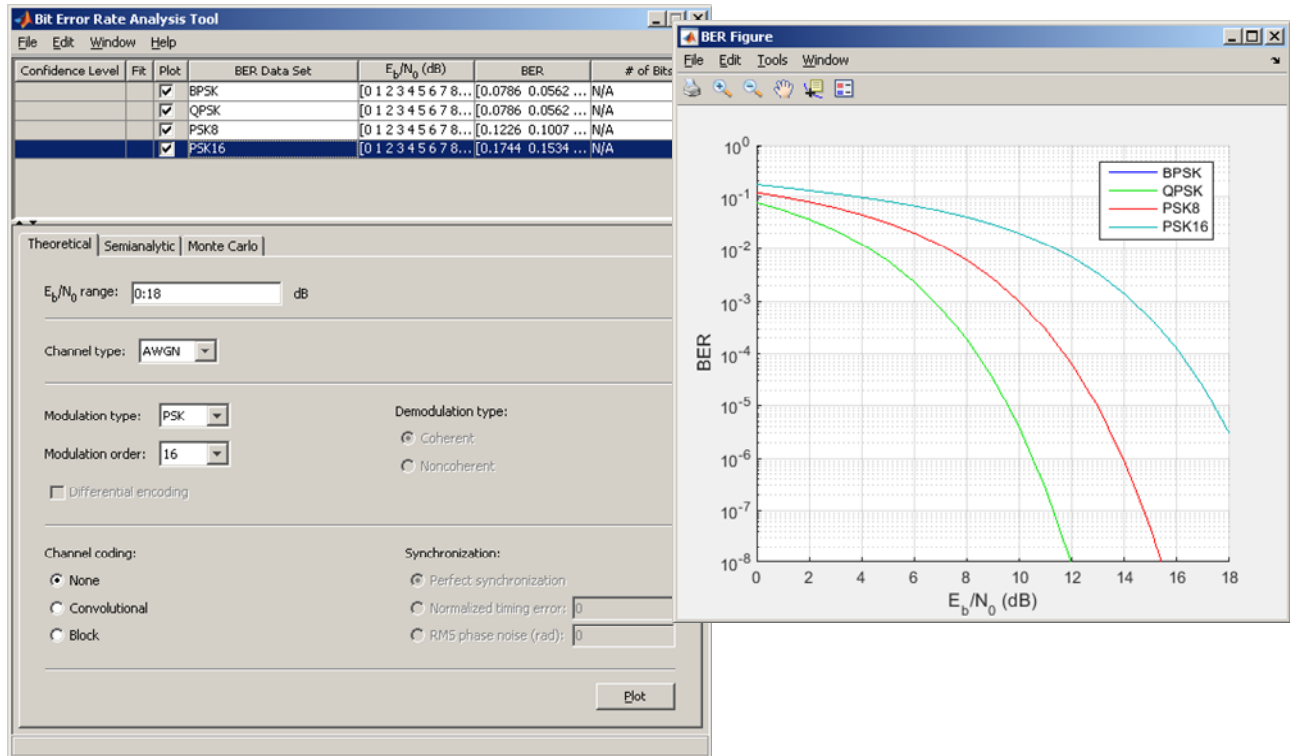


Рис.18.2. Середня імовірність бітових помилок для систем передачі із фазовою маніпуляції та позиційністю $M = 2, 4, 8, 16$ в залежності від параметру завадостійкості E_b/N_0

Як можна бачити з графіків рис.18.2, зростання позиційності модуляції при збереженні однакових бітових помилок потребує (за винятком модуляції QPSK, графік якої збігається із графіком BPSK) відповідного підвищення відношення сигнал/шум на біт, що повністю узгоджується зі загальнотеоретичними міркуваннями. Відзначимо лише кілька технічних моментів. По-перше, на цьому рисунку, на відміну від попереднього, змінено написи як на графіку BER, так й у браузері даних. Для такої заміни достатньо було лише двічі клікнути по відповідному напису у колонці Data Set, а потім вписати там потрібну назву. В результаті напис змінюється не лише у браузері даних, але й на графіку BER. По-друге, кожен з графіків можна як виключити, так й знову включити за допомогою

відповідних *checkbox* браузера даних. По-третє, через меню File передбачено можливість експорту результатів розрахунків у командне вікно Matlab для подальшого оброблення. Зокрема, у середовищі Matlab ці графіки можна побудувати не лише в залежності від q_b^2 , як це відбувається у вікні BERTool, а й для інших визначень параметрів завадостійкості, наприклад, (18.3) або (18.5).

Зазначимо, що у повній відповідності до своєї назви інструмент BERTool показує результати розрахунків лише для бітових помилок. Але це не означає, що обчислення ймовірності символьних помилок зовсім не передбачено. Безпосередньо у вікні BERTool такої можливості нема, але можна скористатися відповідними обчислювальними функціями у командному режимі. Зокрема, для прикладу, наведеного на рис.18.2, це функція *berawgn*. Якщо виконати низку команд⁶, представлених у табл.18.1, то одержимо графік, представлений на рис.18.3.

Таблиця 18.1. Теоретичний розрахунок BER для модуляції QPSK

№	Команди	Коментарі
01.	<code>EbNo = (0:18)';</code>	% Діапазон змін відношення сигнал/шум на біт, дБ
02.	<code>Ber=zeros(4,19);</code>	% Резервування пам'яті для результатів обчислення BER
03.	<code>Ser=zeros(4,19);</code>	% Те ж саме для символьних помилок
04.	<code>i=0;</code>	
05.	<code>for m=[2 4 8 16]</code>	% Цикл по позиційності модуляції PSK
06.	<code> i=i+1; M=m;</code>	% Поточна позиційність
07.	<code> [ber,ser] = berawgn(EbNo,'psk',M, 'nondiff');</code>	% Теоретичне обчислення бітових та символьних помилок
08.	<code> Ber(i,:)=ber;</code>	
09.	<code> Ser(i,:)=ser;</code>	
10.	<code>end</code>	
11.	<code>semilogy(EbNo,Ber(1,:),'b-',EbNo,Ber(2,:),'g-',EbNo,Ber(3,:),'r-',EbNo,Ber(4,:),'m-', EbNo,Ser(1,:),'b--', EbNo,Ser(2,:),'g--', EbNo,Ser(3,:),'r--', EbNo,Ser(4,:),'r--')</code>	% Побудова графіків
12.	<code>xlabel('Eb/No (dB)'), ylabel('BER, SER'),</code>	
13.	<code>legend('BPSK', 'QPSK', 'PSK8', 'PSK16'), grid</code>	

⁶ Зазначимо, що у першому рядку цього коду для параметра (18.2) замість q_b^2 використано стандартне для Matlab позначення, яке фактично повторює формулу правої частини виразу (18.2).

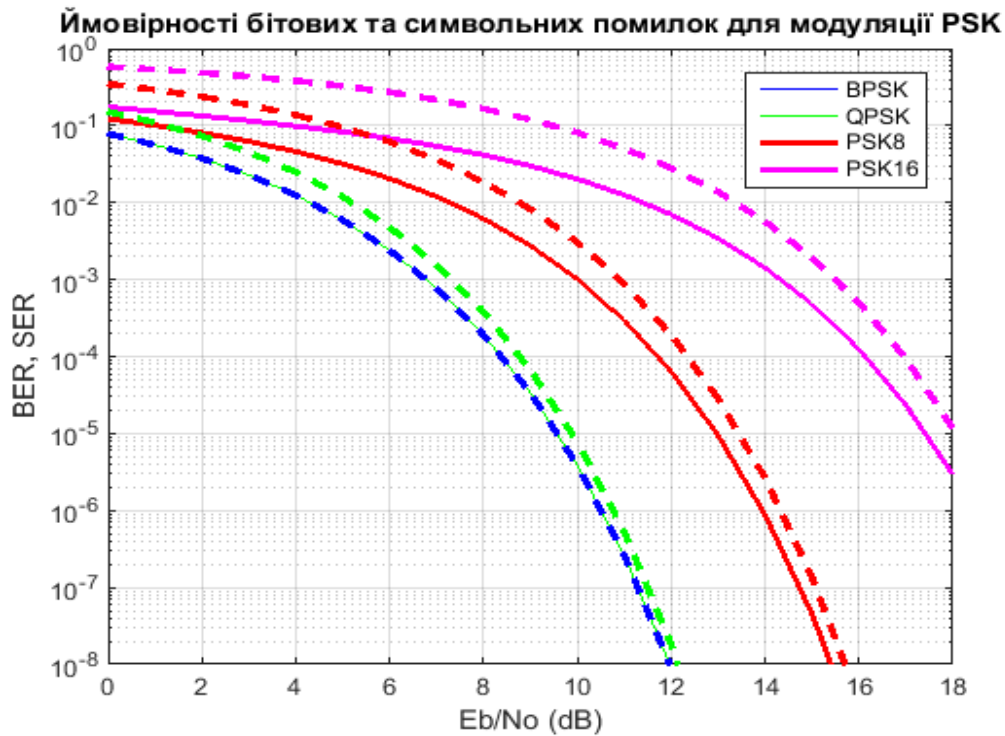


Рис.18.3. Результати розрахунку бітових (суцільні лінії) та символічних (пунктир) помилок за допомогою утиліти *berawgn*.

Із порівняння рис.18.2 та 18.3 видно, що при однаковій позиційності модуляції ймовірності бітових помилок на обох рисунках співпадають. Що ж стосується ймовірностей символічних помилок (пунктирні криві на рис.18.3), то для $M = 2$ вони цілком очікувано співпадають із BER, а для $M > 2$ ймовірності $P_{SE} > P_{BE}$. При цьому для великих значень q_b^2 співвідношення поміж P_{SE} та P_{BE} відповідає наближеній теоретичній оцінці (18.1). Помітні відмінності спостерігаються лише в області відносно малих значень $E_b N_0$.

Теоретичний режим, якій ми розглядали, залишає багато можливостей для дослідження закономірностей поведінки параметрів P_{SE} та P_{BE} в залежності від застосування різних модуляційних форматів, методу демодуляції (когерентна чи некогерентна), методів завадостійкого кодування, моделі каналу передачі. Залишаємо розгляд цих закономірностей для самостійних вправ наших читачів. Зазначимо також, що не слід втрачати із поля зору можливість застосування наряду з інструментом BERTool й командного режиму обчислень з використанням окремих утиліт, як це було проілюстровано вище. Зокрема, це може бути не тільки утиліта

berawgn, застосування якої вже розглядалося вище, а й, наприклад, функція *berfading*, яка призначена для обчислення ймовірностей P_{SE} , P_{BE} у каналах із завмираннями сигналів. До речі, у цьому випадку додатково передбачено можливість аналізу рознесеного приймання для різних варіантів об'єднання гілок рознесення.

Також дуже корисною особливістю інструменту BERTool є можливість порівнювального аналізу теоретичних результатів з результатами комп'ютерного моделювання, виконання якого передбачено у режимі *Monte Carlo*. Розглянемо це питання докладніше.

18.2. Моделювання систем передачі інформації у командному режимі.

Перш за все слід зауважити, що більшість інструментальних пакетів сучасних версій Матлаб побудовано з використанням принципів [] так званого об'єктно-орієнтованого програмування (ООП). Це означає, що будь-яка програма, створена за допомогою цих інструментальних засобів, складається із множини системних *об'єктів*, кожен з яких є екземпляром певного *класу*. Сьогодні багато мов програмування або підтримують ООП або ж є цілком об'єктно-орієнтованими (наприклад, Java, C#, C++, Python та інші). Зокрема, однією з переваг ООП є більш компактна побудова програмного забезпечення, що дозволяє замінити тисячі функцій процедурної мови кількома десятками класів. Принципи ООП у значній мірі знайшли своє відображення й у пакеті *Comm*, який забезпечує побудову та моделювання фізичного рівня телекомунікаційних систем. Але слід зазначити, що впровадження методів ООП у середовищі Матлаб розпочалося лише у другій половині нульових років. На той час вже існувало багато напрацювань для попередніх версій Матлаб, що були виконані у вигляді спеціалізованих програмних функцій. Серед цих функцій є досить багато складних, які швидко переробити під принципи ООП не вдалося або це було визнано недоцільним. Тому у сучасних версіях Матлаб застосовується змішаний підхід, при якому окрім об'єктів ООП збережено й певну частку раніше напрацьованих функцій. Зокрема, чимало таких функцій зустрічається й в інструментальному пакеті *Comm*. Розрізнити їх

нескладно: системні об'єкти даного пакету, на відміну від функцій, містять у своїй назві кореневе слово *"comm"*.

Зі складом пакету можна ознайомитися за допомогою команди *help comm*. В результаті виконання цієї команди в основному вікні Матлаб буде показано повний перелік програмних функцій та системних об'єктів, частину яких відображено на рис.18.4. Тут усі функції та об'єкти згруповані за певними категоріями відповідно до свого призначення. До таких категорій відносяться, зокрема, модуляція/демодуляція сигналів, проходження сигналів через фізичні канали передачі, моделювання та кодування джерел інформації, синхронізація приймачів, вирівнювання каналів, фільтрація та багато інших.

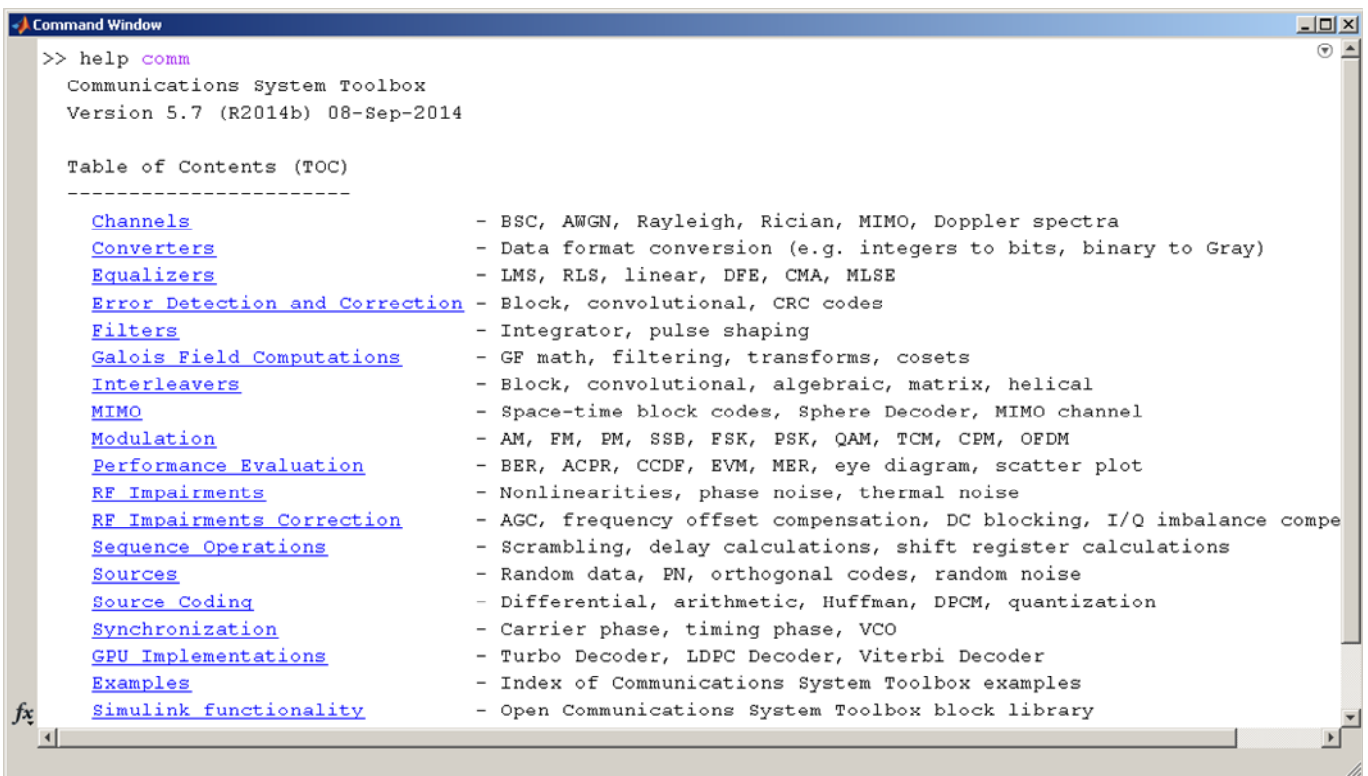


Рис.18.4. Категорії функцій та системних об'єктів інструментального пакету *Communication Systems Toolbox*.

Зміст кожної категорії можна подивитися, якщо у вказаному переліку клікнути на її назву. Наприклад, на рис.18.5 показано частину переліку функцій та об'єктів категорії *Modulation*. Якщо тепер клікнути назву окремої функції чи класу об'єктів, то одержимо коротку довідку щодо її призначення та правил використання, яка часто супроводжується й прикладами застосування. Тут же звичайно наводиться й посилання на більш повний опис функцій та об'єктів у довідковому меню *help*.

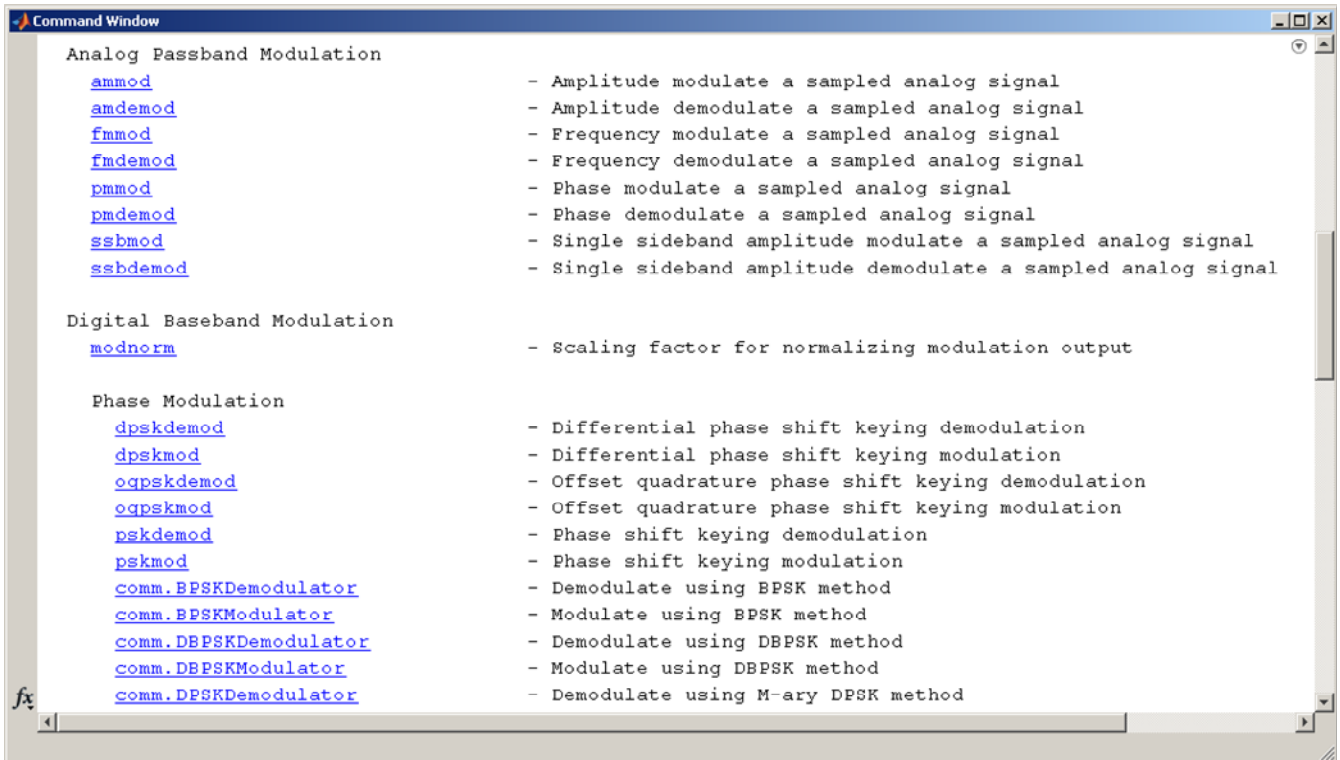


Рис.18.5. Функції та системні об'єкти, що входять до категорії Modulation.

Розглянемо деякі приклади застосування функцій та об'єктів пакету *Comm*. Зауважимо, що моделювання систем передачі включає, як правило, низку задач, зокрема, моделювання джерела інформації, його кодування, каналне кодування, модуляцію, проходження сигналу через фізичне середовище, демодуляцію, декодування, підрахунок частоти помилок, відображення результатів, тощо. Для кожної з цих задач в інструментальному пакеті *Comm* передбачено певну сукупність програмних функцій та системних блоків. Розглянемо приклад, в якому виконується оцінювання BER після проходження сигналу із багатопозиційною PSK-модуляцією через канал АБГШ. В табл.18.2 показана частина програмного коду Матлаб, що відповідає початковій ініціалізації програми моделювання.

Таблиця 18.2. Моделювання системи передачі з модуляцією M-PSK. Початкова ініціалізація

№	Команди	Коментарі
01.	<code>siglen = 1e5;</code>	% Число біт у кожному статистичному випробовуванні
02.	<code>M = 8;</code>	% Позиційність модуляції
03.	<code>kb=log2(M) ;</code>	% Число біт на символ

04.	<code>hMod = comm.PSKModulator(M);</code>	% Створення екземпляру об'єкту, що виконує M-PSK Модуляцію
05.	<code>hDemod= comm.PSKDemodulator(M);</code>	% Створення екземпляру об'єкту, що виконує M-PSK демодуляцію
06.	<code>hChan = comm.AWGNChannel('NoiseMethod', 'Signal to noise ratio (Eb/No)', 'BitsPerSymbol', kb);</code>	% Створення екземпляру об'єкту, що моделює проходження сигналу через канал із адитивним ґавсівським шумом
07.	<code>EbNomin = 0; EbNomax = 12;</code>	% Діапазон змін відношення сигнал/завада на біт
08.	<code>EbNovec = EbNomin:1:EbNomax; numEbNos = length(EbNovec);</code>	% Вектор значень відношення сигнал/завада на біт та його довжина
09.	<code>Ber = zeros(1,numEbNos);</code>	% Резервування пам'яті під теоретичне обчислення BER
10.	<code>ber_Est = zeros(1,numEbNos);</code>	% Резервування пам'яті для оцінки BER

Прокоментуємо деякі позиції. Зокрема у рядку 04 створюється об'єкт, що забезпечує модуляцію PSK. В якості аргументу вказана позиційність модуляції M . Але цей параметр є не єдиним із можливих – повний перелік аргументів, які у Матлаб називаються властивостями (Properties) об'єктів, можна знайти за відповідним *help*. У даному ж разі значення інших аргументів прийняті за замовчуванням. Для того, щоб їх подивитися, достатньо у командному вікні виконати команду *hMod*. В результаті побачимо наступний текст:

`hMod =`

System: [comm.PSKModulator](#)

Properties:

ModulationOrder: 8

PhaseOffset: 0.392699081698724

BitInput: false

SymbolMapping: 'Gray'

OutputDataType: 'double'

Перший рядок властивостей об'єкту *hMod* позначає його приналежність до класу [comm.PSKModulator](#). Параметри цього об'єкту мають у даному разі наступні значення:

- позиційність модуляції *ModulationOrder* = 8 (значення за замовчуванням $M=4$, тому в рядку 02 попередньо було встановлено $M=8$);

- фазовий поворот першої точки сигнального сузір'я $PhaseOffset = \pi/M$;
- формат бітового потоку на вході модулятора `BitInput` має значення `'false'`, що відповідає десятковій формі представлення вхідних символів;
- кодування точок сигнального сузір'я відповідає коду Грея, тобто, *SymbolMapping: 'Gray'*;
- представлення виходу модулятора має формат *OutputDataType: 'double'*, тобто, стандартне для Matlab.

Будь який з цих параметрів можна при необхідності легко змінити. Для цього при початковій ініціалізації об'єкту у переліку його аргументів треба явно вказати їх значення у вигляді пар (*'Properties', 'Value', ...*). При цьому порядок слідування пар значення не має. Можна також зробити це пізніше шляхом виконання відповідної

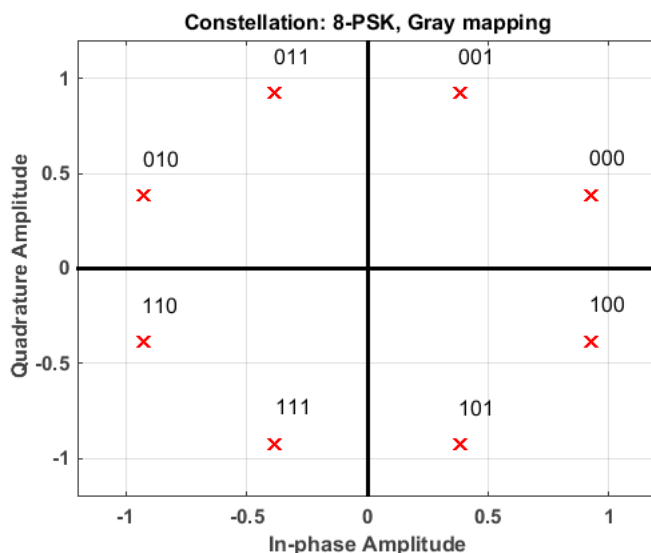


Рис.18.6. Сигнальне сузір'я, що відповідає прикладу табл.18.2

команди. Наприклад, якщо від десяткового треба перейти до бітового представлення вхідного потоку, то достатньо виконати команду `hMod.BitInput = true`. При цьому формат вхідних даних буде відповідати бінарній формі, а усі інші параметри залишаться незмінними.

Також зазначимо, що майже кожному системному об'єкту відповідає, як правило, низка допоміжних функцій (команд), перелік яких також наведено у відповідних файлах *help*. Як приклад, розглянемо застосування команди *constellation*, що показує сигнальне сузір'я, яке відповідає даному об'єкту. Формат команди у даному разі є *constellation(hMod)*, після виконання якої й побачимо «картинку», яка представлена⁷ на рис.18.6.

⁷ Попередньо формат вхідного потоку був змінений на `'true'`.

Продовжимо розгляд коду табл.18.2. У рядку 05 створюється об'єкт *hDemod*, який забезпечує демодуляцію сигналу. Все, що стосується його властивостей, визначається таким же чином, як й для об'єкту *hMod*. У наступному рядку створюється об'єкт *hChan*, який забезпечує моделювання каналу АБГШ. Властивості цього об'єкту за замовчуванням наведено нижче:

```
hChan =
System: comm.AWGNChannel
Properties:
    NoiseMethod: 'Signal to noise ratio (Eb/No)'
    EbNo: 10
    BitsPerSymbol: 2
    SignalPower: 1
    SamplesPerSymbol: 1
    RandomStream: 'Global stream'
```

Тут прокоментуємо лише опцію *NoiseMethod*, яка встановлює одне із визначень параметру відношення сигнал/завада, розглянутих на початку даного розділу. Зокрема, у цьому прикладі застосовується відношення сигнал/шум на біт (18.2), яке у Матлаб позначено *Eb/No* та яке визначено за замовчуванням. Якщо ж треба встановити відношення сигнал/шум на символ (18.3), то треба застосувати опцію '*Signal to noise ratio (Es/No)*'. Традиційному ж визначенню (18.5) відповідає опція '*Signal to noise ratio (SNR)*'. Можна також явно задати дисперсію шуму в каналі через опцію '*Variance*', що при нормованій потужності сигналу по виходу каналу також дозволяє змінювати відповідним чином відношення сигнал/шум на вході приймача.

У табл.18.3 наведено продовження коду таблиці 18.2, який відповідає власне моделюванню проходження сигналу через заданий канал. Неважко бачити, що тут виконується цикл по значеннях параметру *Eb/No*. При кожному значенні цього параметру, що заноситься до параметрів об'єкту *hChan* (рядок 13) формується інформаційний вектор джерела *msg* (рядок 14) та виконується модуляція (рядок 15). Тут слід зауважити що застосування об'єкту *hMod* до вектору *msg* виконується за допомогою команди *step*, яку в даному разі слід сприймати як дію оператора *hMod* на дані *msg*. Це не єдина форма виконання даної операції: в різних версіях Матлаб зустрічаються й інші еквівалентні процедури, але на суть справи це не впливає.

Таблиця 18.3. Моделювання системи передачі з модуляцією QPSK. Цикл по параметру EbNo

№	Команди	Коментарі
11.	<code>for jj = 1:numEbNos</code>	% Цикл по параметру сигнал/завада на біт EbNo
12.	<code>q2b = EbNovec(jj);</code>	% Чергове значення параметру EbNo
13.	<code>hChan.EbNo = q2b;</code>	% Внесення параметру EbNo до об'єкту hChan
14.	<code>msg = randi([0,M-1],symlen,1);</code>	% Генерація siglen символів джерела
15.	<code>txsig = step(hMod, msg);</code>	% Генерування сигналу по виходу модулятора
16.	<code>hChan.SignalPower = (txsig'*txsig)/length(txsig);</code>	% Нормування сигналу в каналі
17.	<code>rxsig = step(hChan, txsig);</code>	% Сигнал після проходження каналу
18.	<code>decodmsg = step(hDemod, rxsig);</code>	% Демодуляція
19.	<code>[Nb,Be] = biterr(msg,decodmsg);</code>	% Кількість та частота бітових помилок.
20.	<code>ber_Est(jj) = Be;</code>	% Середня частота бітових помилок
21.	<code>ber(jj) = berawgn(qb,'psk',M,'nondiff');</code>	% Теоретичний розрахунок BER
22.	<code>end</code>	% закінчення циклу по параметру EbNo
23.	<code>semilogy(EbNovec,ber,'b-',EbNovec,ber_Est,'*b'),grid shg</code>	% Побудова графіку

Дія каналу АБГШ враховується у рядку 17, де також застосовується команда *step*, але вже відносно об'єкту *hChan*. Ефект проходження сигналу через канал можна наочно представити так званою діаграмою розсіювання для чого

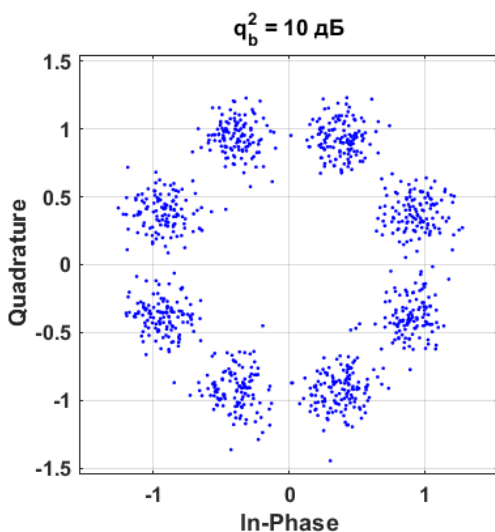


Рис.18.7. Діаграма розсіювання сигналу з модуляцією PSK-8 у каналі АБГШ.

використовується допоміжна команда *scatterplot(rxsig)*. Результат виконання цієї команди представлено на рис.18.7. Тут ми бачимо сигнальне сузір'я по виходу каналу передачі. У даному разі – це просто «зашумлення» сигнального сузір'я. Але можливі й значно складніші умови в каналах та, відповідно, більш суттєві спотворення сигналу. Тому при виконанні моделювання систем передачі даний інструмент є дуже зручним та

наочним.

Продовжимо розгляд коду табл.18.3. У рядку 18 за допомогою команди *step* та об'єкту *hDemod* виконується демодуляція сигналу на виході каналу. У рядках 19 – 21 підраховуються експериментальні та теоретичні значення BER за допомогою функції *biterr* та вже відомої нам функції *berawgn*. Після завершення перебору усіх заданих значень параметру *EbNo* виконується побудова графіків із результатами моделювання. Деякі приклади представлені на рис.18.8.

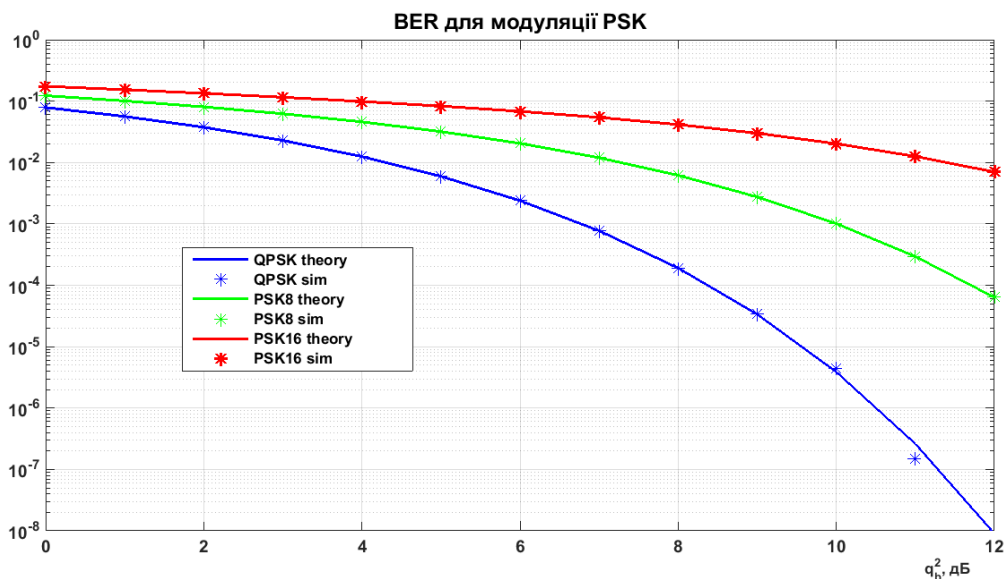


Рис.18.8. Результат виконання програмного коду табл. 18.2, 18.3 для значення позиційності $M=4, 8, 16$. Суцільні лінії відповідають результатам теоретичних розрахунків, маркерами позначено результати моделювання.

Ці графіки були одержані шляхом накладання за допомогою команди *hold* результатів трикратного виконання програмного коду табл.18.2, 18.3. При цьому перед кожним запуском задавалися відповідним чином параметр M та кольорове маркування кривих. Як видно з представлених графіків, результати моделювання добре узгоджуються з теоретичними розрахунками, що досягнуто за рахунок належного вибору довжини тестової сигнальної послідовності *siglen*.

Зазначимо також, що розглянутий вище приклад стосувався моделювання проходження сигналу з PSK модуляцією у так званій смузі *baseband*, тобто, для комплексної обвідної сигналу. Якщо ж треба дослідити проходження сигналу на носійній або проміжній частоті, то цей випадок у Матлаб називають моделюванням у смузі *passband*. Приклади можна знайти у відповідних довідкових розділах.

Тепер повернемося до питання, що було анонсовано на початку даного параграфу, а саме, виконання моделювання з використанням засобів графічної оболонки BERTool. Відразу зазначимо, що якогось принципового спрощення застосування вказаної оболонки не дає: мова може йти лише про деякі сервісні можливості. Фактично засіб BERTool відіграє при моделюванні інтегруючу роль, яка дозволяє проводити порівняння результатів моделювання із результатами, одержаними іншими методами, зокрема, теоретичним аналізом, на пів аналітичним моделюванням та моделювання засобами Simulink, про що піде мова у наступному параграфі.

Отже, для забезпечення вказаних можливостей BERTool перш за все треба розробити відповідну програмну функцію тими ж методами, що вже розглядалися вище. Але при цьому слід враховувати й додаткові вимоги, які б забезпечували можливість спільної роботи програмної моделі зі засобом BERTool. Зокрема, першим входним аргументом цієї функції має бути відношення сигнал/шум на біт E_b/N_0 , другим аргументом – кількість помилок *maxNumErrs*, достатніх для обчислення BER, третім – максимально допустима кількість бітів тестової послідовності *maxNumBits*. Засіб BERTool керує послідовними запусками цієї функції для кожного із передбачених значень параметру E_b/N_0 . Моделювання відбувається по циклах, в яких поступово збільшується об'єм статистичних випробувань. Зупинка набору статистики має відбуватися у трьох випадках:

- фактична кількість бітових помилок досягає наперед заданого значення параметру *maxNumErrs*;
- загальна кількість біт у тестовій послідовності перевищує задане значення параметру *maxNumBits*;
- аварійне переривання циклу при натисненні у вікні BERTool (вкладка *Monte Carlo*) кнопки *Stop*.

Вихідними параметрами функції на кожному кроці по параметру E_b/N_0 є значення BER (перший аргумент) та фактична довжина використаної бітової послідовності *numBits* (другий аргумент). При цьому входні параметри задаються

безпосередньо у вікні BERTool (вкладка *Monte Carlo*), а результати оцінювання BER відображаються графіком залежності BER від параметру E_b/N_0 .

У табл.18.4 представлено приклад функції, що відповідає вказаним вимогам та яка створена на основі розглянутих вище табл.18.2, 18.3.

Таблиця 18.4. Функція моделювання передачі сигналу з модуляцією PSK спільно із засобом BERTool.

№	Команди	Коментарі
01.	<code>function [ber,numBits] = MPSK_bertool(q2b,maxNumErrs,maxNumBits)</code>	% заголовок функції з обов'язковими параметрами
02.	<code>import com.mathworks.toolbox.comm.BERTool;</code>	% завантаження програмного модулю взаємодії з оболонкою BERTool
03.	<code>berVec = zeros(3,1);</code>	% резервування пам'яті для результатів підрахунку помилок та загальної кількості біт
04.	<code>M = 8;</code>	% позиційність модуляції
05.	<code>kb=log2(M);</code>	% кількість біт на символ
06.	<code>symlen = 1e3;</code>	% кількість символів на один фрейм
07.	<code>hMod = comm.PSKModulator('ModulationOrder', M, 'BitInput', true, 'PhaseOffset', 0, 'SymbolMapping', 'Gray');</code>	% Створення об'єкту модулятора
08.	<code>hDemod= comm.PSKDemodulator('ModulationOrder', hMod.ModulationOrder, 'BitOutput', hMod.BitInput, 'PhaseOffset', hMod.PhaseOffset, 'SymbolMapping', hMod.SymbolMapping, 'DecisionMethod', 'Hard decision');</code>	% Створення об'єкту демодулятора
09.	<code>hChan = comm.AWGNChannel('NoiseMethod', 'Signal to noise ratio (Eb/No)', 'EbNo', q2b, 'SignalPower', 1, 'BitsPerSymbol', kb);</code>	% Створення об'єкту каналу АБГШ
10.	<code>hErrorCalc = comm.ErrorRate;</code>	% Створення об'єкту калькулятора помилок
11.	<code>while((berVec(2) < maxNumErrs) && (berVec(3) < maxNumBits))</code>	% цикл статистичних випробувань
12.	<code>if (BERTool.getSimulationStop)</code>	% Перевірка натиснення кнопки

		Stop
13.	<code>break;</code>	% Переривання циклу випробувань
14.	<code>end</code>	
15.	<code>msg = randi([0,1], kb*symlen, 1);</code>	% формування чергового фрейму вхідних інформаційних бітів
16.	<code>txsig = step(hMod, msg);</code>	% виконання модуляції
17.	<code>hChan.SignalPower =(txsig'*txsig)/length(txsig);</code>	% нормування потужності сигналу в каналі
18.	<code>rxsig = step(hChan, txsig);</code>	% проходження сигналу через канал АБГШ
19.	<code>decodmsg = step(hDemod, rxsig);</code>	% демодуляція
20.	<code>berVec = step(hErrorCalc, msg, decodmsg);</code>	% підрахунок помилок
21.	<code>end</code>	% кінець циклу випробувань
22.	<code>ber = berVec(1);</code>	% перший вихідний аргумент
23.	<code>numBits = berVec(3);</code>	% другий вихідний аргумент

Прокоментуємо цей код. По-перше, відзначимо, що на відміну від прикладу, наведеного у табл.18.2, 18.3, код таблиці 18.4 є *m*-функцією, тобто, він має заголовок (рядок 01) зі списками вхідних та вихідних аргументів, що відповідають вказаним вище вимогам. Зазначимо лише, що позначення змінних можуть бути довільними: головне, щоб зберігалася їх відповідність з програмним кодом. Новим елементом також є рядок 02, в якому завантажується спеціальний програмний модуль взаємодії функції з оболонкою BERTool. Наступні вісім рядків коду принципово не відрізняються від кодів початкової ініціалізації табл.18.2. Відзначимо лише, що тут з'явився новий об'єкт, а саме, лічильник помилок, роль якого у табл.18.3 відігравала функція *biterr*. Перехід від цієї функції до об'єкту з класу *comm.ErrorRate* зумовлено зручністю керування ним зі сторони засобу BERTool. Також зазначимо, що у властивостях об'єктів модуляції/демодуляції використовується бінарне представлення символів, що, знов ж таки, зумовлено

специфікою застосування калькулятора помилок даного типу, про що ще буде сказано трішки пізніше.

Далі у рядках 11-21 виконується цикл з набору статистики сигналу, достатньої для обчислення оцінки BER. Порядок дій на кожній ітерації принципово співпадає з кодом рядків 14 – 18 табл.18.3 за винятком нового елементу: в рядках 12 – 14 відстежується необхідність переривання циклу при натисненні кнопки *Stop* у засобі BERTool. Це зроблено для того, щоб запобігти «зациклюванню» моделювання внаслідок якогось некоректного завдання його параметрів. Також слід відзначити певну особливість роботи калькулятора помилок у рядку 20. На відміну від функції *biterr* об'єкт класу *comm.ErrorRate ErrorCalc* може обчислювати як символьну, так й бітову помилки. Використання відповідного режиму роботи калькулятора визначається формою представлення символьних послідовностей: якщо використовується їх бінарна форма, тобто, *BitInput='true'*, то калькулятор оцінює BER, а разі десяткового представлення символів, тобто, *BitInput='false'*, калькулятор обчислює символьні помилки.

Для того, щоб запустити моделювання з використанням функції табл.18.4 треба у вікні BERTtool відкрити вкладку *Monte Carlo*, підключити вказану функцію, встановити усі необхідні параметри та натиснути кнопку *Run*. Засіб BERTtool почне виконання функції табл.18.4 для заданого діапазону змін параметру *q2b* та побудову відповідного графіку BER. Більш детально процедура виконання моделювання з використанням засобу BERTtool буде розглянута у наступному параграфі.

18.3. Моделювання систем передачі інформації у середовищі Simulink.

Тепер розглянемо інший потужний засіб аналізу систем передачі інформації, а саме, візуально-орієнтовано середовище моделювання динамічних систем Simulink. Цей засіб є невід'ємною складовою частиною середовища Matlab у тому сенсі, що Simulink без Matlab функціонувати не може (але не навпаки!). Тому часто говорять не окремо про Matlab та Simulink, а про інтегроване середовище Matlab-Simulink. У той же час, за своєю функціональністю Simulink – це окрема система, яка працює за іншими принципами чим традиційний Matlab. Ми вживаємо прикметник «традиційний» тому що середовище Simulink було введено до складу інтегрованого середовища років на 10 пізніше появи початкових версій Matlab.

Головна відмінність у порівнянні з Matlab полягає в тому, що Simulink оперує не з функціями та об'єктами, а із графічними блоками. Засобами візуального програмування ці блоки можна компоувати у певні структури та з'єднувати поміж собою деякими «електричними» або логічними інтерфейсами, наприклад, «системними шинами». Тому розробка моделей у середовищі Simulink схожа на розробку схем реальних пристроїв, що для інженерів-розробників є більш звичною справою, чим програмні коди. Інша принципова відмінність – це наявність у середовищі Simulink віртуального модельного «часу», який при моделюванні в Matlab у явному вигляді не фігурує. При цьому можна спостерігати за проходженням та зміною сигналів у часі для різних ділянок таких віртуальних «схем» за допомогою «пристроїв», які за своєю функціональністю виконують роль «тестерів», «осцилографів», «аналізаторів» і т.п. Ці віртуальні пристрої можуть підключатися до різних ділянок «схем» через різного роду з'єднання. Одержані таким чином графічні структури у документації Matlab називають S-моделями.

Як було сказано на початку розділу, ми виходимо з того, що наші читачі попереднього вже ознайомлені з базовими принципами побудови та основними правилами роботи у середовищі Simulink. Тому тут обмежимося лише деякими прикладами. Отже, для запуску Simulink треба відкрити відповідну модель. Це можна зробити кількома рівноцінними способами. Якщо це вже існуюча модель, то

вона відкривається звичайним шляхом через опцію **Open** головного вікна Matlab. Якщо ж треба створити нову модель, то слід скористатися випадającym меню **New** та вибрати у ньому опцію **Simulink Model**, в результаті виконання якої з'явиться графічний редактор нової моделі.

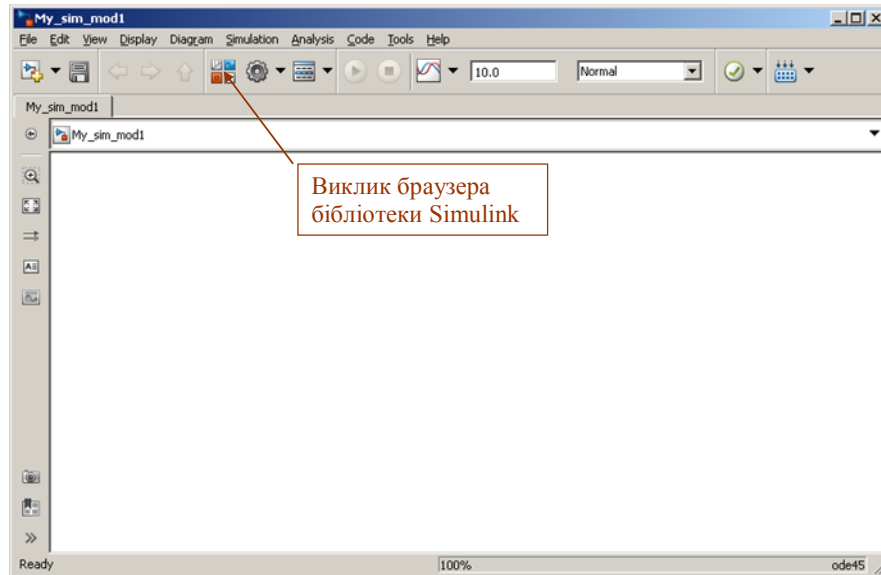


Рис.18.9. Вікно нової моделі My_sym_mod1

За умовчуванням будь-яка нова модель має ім'я **untitled**. Тому перш за все слід надати їй своє оригінальне ім'я, для чого можна скористатися опцією **Save As** меню **File** вікна моделі. Назвемо нашу першу модель **My_sim_mod1**. Отже, після виконання команди **Save As** одержимо вікно графічного редактора моделі, що й представлено на рис.18.9.

Наступний крок – це виклик браузера бібліотеки блоків Simulink, для чого слід скористатися кнопкою, що позначена у вікні рис.18.9. (Аналогічна кнопка є й у головному вікні Matlab). В результаті з'явиться вікно браузера бібліотеки, яке представлено на лівій частині рис.18.10. Як бачимо, у головному вікні показані основні розділи бібліотеки. Якщо потім у цьому вікні натиснути на певний розділ, наприклад, **Channels** із каталогу **Communications System Toolbox**, то одержимо його зміст, тобто, перелік підрозділів або безпосередньо блоків, що до нього входять (права частина рис.18.10). Ці блоки вже можна використовувати для побудови власних моделей.

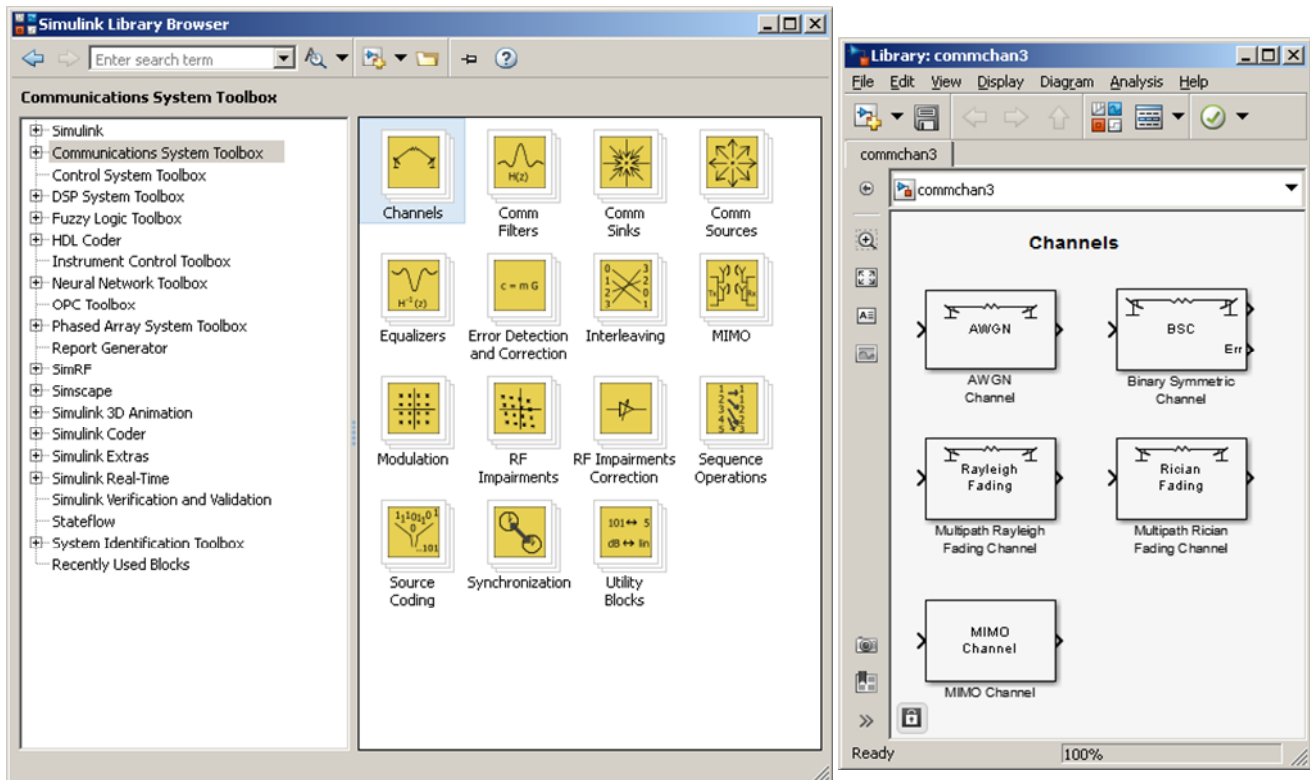


Рис.18.10. Вікно браузера бібліотеки блоків Simulink та розділу Communications System Toolbox/Channels

Проілюструємо принципи побудови S-моделей на прикладі створення моделі генератора сигналів та засобів їх візуального відтворення. Система графічного редагування моделей побудована на звичайному для Windows принципі "drag-and-drop", тобто, "перетягни та залиш". У прикладі, який розглядається, будемо діяти наступним чином. У браузері бібліотеки в розділі Simulink вибираємо вкладку Sources (Джерела) та знаходимо там блок Signal Generator, який дозволяє генерувати сигнали кількох видів: синусоїдальний, прямокутний, пилкоподібний, випадковий. Далі цей блок можна або просто перетягнути у вікно моделі при натиснутій лівій клавіші миші або, натиснувши праву клавішу, вибрати опцію Add block to model My_sim_mod1. В обох випадках копія цього блоку опиниться у вікні редактора нашої моделі. Далі аналогічним чином додаємо до моделі блок Scope з піддиректорії Simulink\Sinks (Одержувачі), блок Zero-Order Hold з розділу Simulink\Discrete і блоки спектроаналізатора Spectrum Analyser та прослуховування сигналу To Audio Device з розділу DSP System Toolbox\Sinks. За допомогою "миші" (при натиснутій лівій клавіші) треба з'єднати усі блоки в схему, що показана на рис.18.11. Також корисно

включити до складу моделі блок Model Info, який міститься у підрозділі Simulink/Model-Wide Utilities. Цей блок дозволяє робити написи різного характеру, наприклад, коментарі, які б пояснювали призначення та умови використання моделі, значення певних параметрів, тощо.

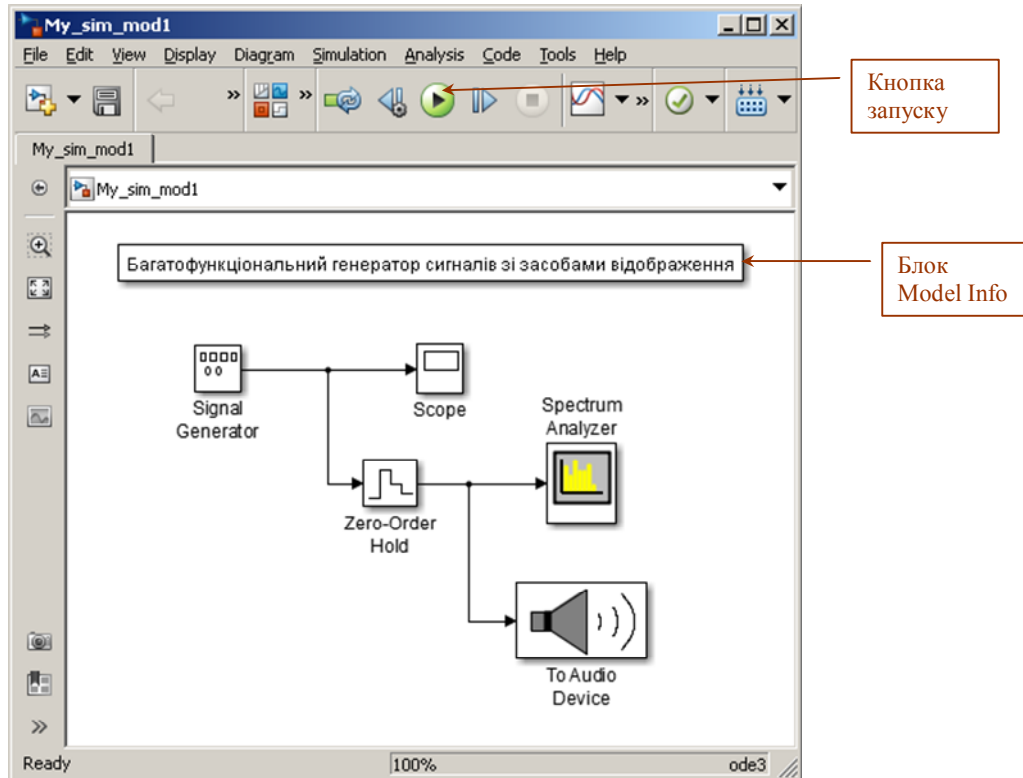


Рис.18.11. Модель My_sim_mod1

Тепер треба задати параметри окремих блоків. Для цього слід навести курсор «миші» на кожний із блоків моделі та двічі клацнути лівою клавiшею. Для відповідних блоків будуть з'являтися вікна параметрів, які показані на рис.18.12. Зокрема, у даному прикладі задана генерація синусоїдального сигналу із частотою 500 Гц. Вихідний сигнал генератора подається у трьох напрямках. По-перше, реалізацію сигналу у часі можна спостерігати у вікні Scope, по-друге, блок Spectrum Analyzer обчислює його миттєвий спектр Фур'є, по-третє, сигнал можна почути через аудіо систему PC завдяки наявності блоку To Audio Device. Між генератором сигналу та блоком Spectrum Analyzer є важливий блок Zero-Order Hold. Щоб зрозуміти його функцію, треба пояснити принципи керування модельним часом у середовищі Simulink. Відповідні параметри зосереджені у меню Simulation, вкладка Model Configuration Parameters, яка присутня у вікні кожної моделі (Рис.18.13).

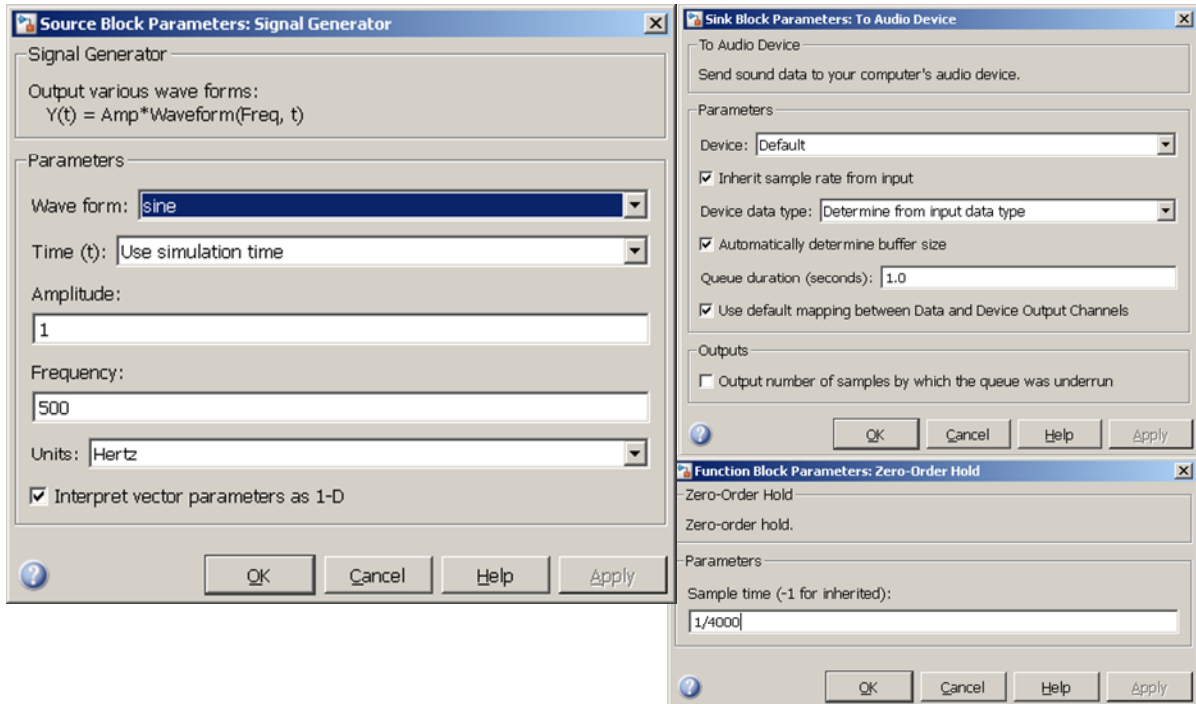


Рис.18.12. Параметри блоків Signal Generator, Zero Order Hold та To Audio Device

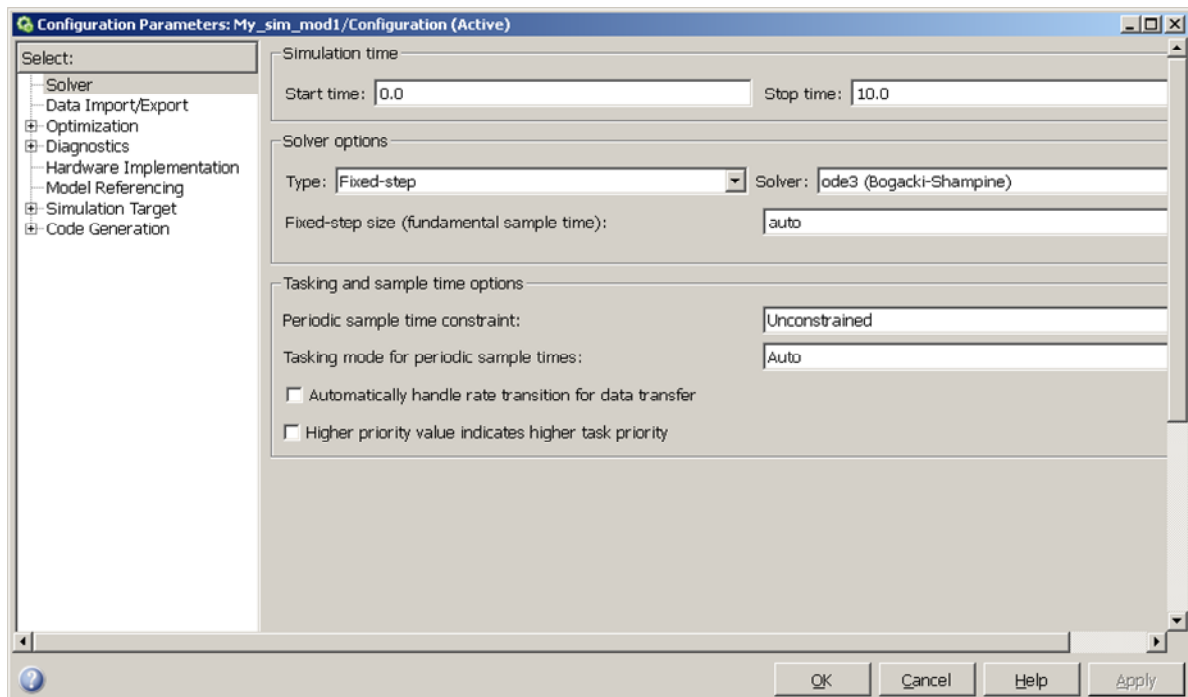


Рис.18.13. Вкладка конфігураційних параметрів моделі

Оскільки середовище Simulink призначене для моделювання систем реального часу, то у загальному випадку час вважається умовно безперервним. Більш того, багато блоків (зокрема, блок Signal Generator) можуть працювати лише у

«безперервному» часі. Перемикання між режимами безперервного та дискретного часу визначається параметрами розділу Solver цієї вкладки, а саме: для випадку безперервного часу треба задати змінний крок часу Variable-Step та одну з опцій, що визначають спосіб вирішення відповідних диференціальних (фактично, різницевих) рівнянь, скажімо, ode-45. Для того щоб час був дискретним, треба задати, як у нашому випадку, фіксований крок, тобто, Fixed-Step. Отже, якщо час моделювання у загальному випадку є умовно безперервний, то треба передбачити блоки, які б дозволяли з'єднувати між собою блоки безперервного часу із блоками, робота яких принципово можлива лише у дискретному часі⁸. Цю роль, зокрема, й виконує блок Zero-Order Hold, який можна розглядати як свого роду аналогово-цифровий перетворювач (АЦП). У даній моделі, він поєднує вихід генератора «безперервного» сигналу із входом блоку, що обчислює дискретне перетворення Фур'є. Єдиним параметром цього блоку є інтервал взяття відліків вхідного сигналу Sample time. У даному прикладі величина цього параметру встановлена 1/4000, що відповідає частоті дискретизації 4кГц (див. рис.18.12). Аналогічний параметр передбачено і в наступному за АЦП блоці спектрального аналізу. Однак, у багатьох випадках можна скористатися неявним способом. Для цього у відповідній позиції вікна параметрів наступного блоку треба задати значення Sample time = -1. При цьому блок наслідує (inherit) інтервал взяття відліків із попереднього блоку, тобто, АЦП. Також важливими параметрами моделі є тривалість її роботи у часі, яка визначається параметрами Start time та Stop time на вкладці конфігураційних параметрів рис.18.13. В даному разі вони мають значення відповідно до замовчування, а саме, 0 та 10 с. відповідно.

Тепер модель повністю готова до запуску. Для цього на панелі інструментів моделі достатньо натиснути кнопку пуск (рис.18.11). У процесі виконання моделювання можна спостерігати сигнал на виході генератора та його поточний спектр (рис.18.14). Як бачимо, спектр дійсного гармонійного сигналу у повній відповідності до теорії містить дві спектральні складові на частотах 500 та -500 Гц. Параметри аналізу і зображень на екранах пристроїв Scope та Spectrum Analyzer

⁸ У більш загальному плані слід фактично говорити про з'єднання різнотемпових блоків.

можна змінювати, наприклад, масштабувати, розтягувати, тощо. Можна також запускати дану модель із сигналами інших типів з переліку меню блоку Signal Generator.

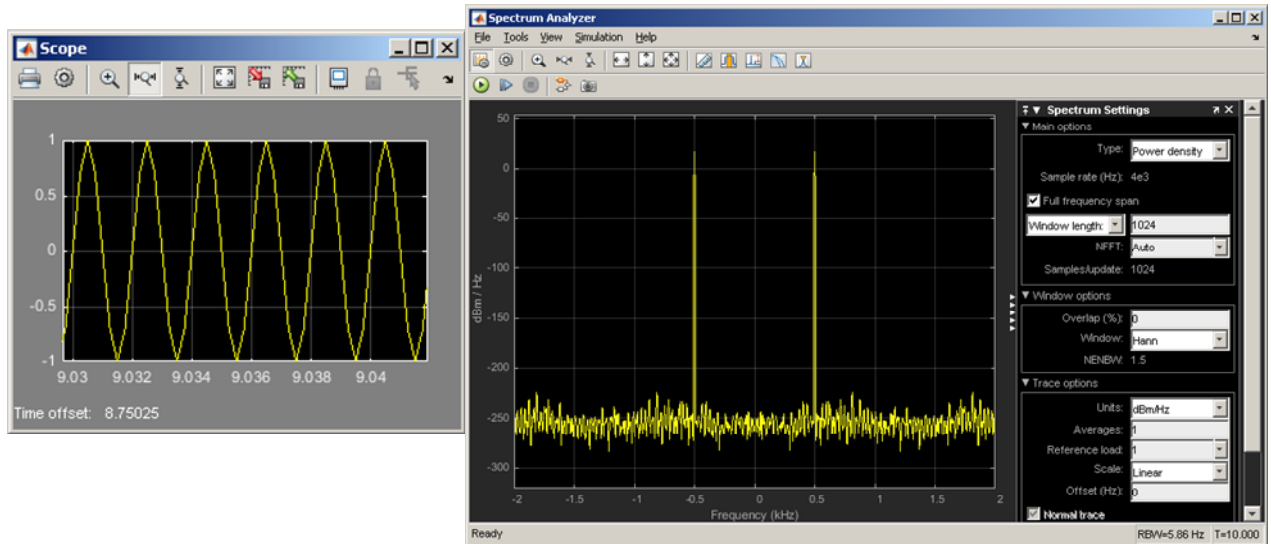


Рис.18.14. Форма та спектр сигналу у моделі рис.18.11

Повне меню налаштувань кожного з блоків з'являється, як правило, при натисненні на панелі інструментів кнопки у вигляді «шестерінки». Також для цих, як і для багатьох інших блоків, передбачена можливість обміну даними із середовищем Matlab. Більш детальну інформацію щодо роботи блоків можна одержати за кнопкою Help, яка звичайно входить до загального меню кожного з них.

Тепер перейдемо до розгляду більш складних моделей, які мають безпосереднє відношення до завдань аналізу ефективності демодуляції сигналів. Структура таких моделей в цілому побудована за тією ж схемою, що й розглянуті вище програмні коди для моделювання систем передачі у командному режимі, тобто, до її складу мають входити блоки джерела інформації, цифровий модулятор, канал передачі, демодулятор, лічильник помилок, засоби візуального контролю, допоміжні блоки.

Розглянемо, зокрема, схему, що представлена на рис.18.15. Ця модель призначена для моделювання передачі інформації сигналами із М-позиційною фазовою маніпуляцією M-PSK. Моделювання у даному разі виконується у смузі baseband, тобто, на рівні комплексної обвідної. Відповідно до рис.18.15 модель складається із наступних блоків:

- Генератор випадкових цілих чисел Random Integer Generator із розділу Communications System Toolbox/Comm Sources/Random Data Sources;
- Перетворювач символів у біти Integer To Bit Converter з розділу Communications System Toolbox/Utility Blocks;
- Блок модуляції M-PSK Modulator Baseband із розділу Communications System Toolbox /Modulation/Digital Baseband Modulation/PM;
- Блок демодуляції M-PSK Demodulator Baseband з того ж розділу;
- Канал із адитивним білим гаусівським шумом AWGN Channel з розділу Communications System Toolbox/Channels;
- Лічильник помилок Error Rate Calculation з розділу Communications System Toolbox/Comm Sinks;
- Блок побудови діаграм розсіювання Constellation Diagram з того ж розділу;
- Блоки відображення значень параметрів Display з розділу Simulink/Sinks;
- Інформаційні блоки Model Info з розділу Simulink/Model-Wide Utilities.

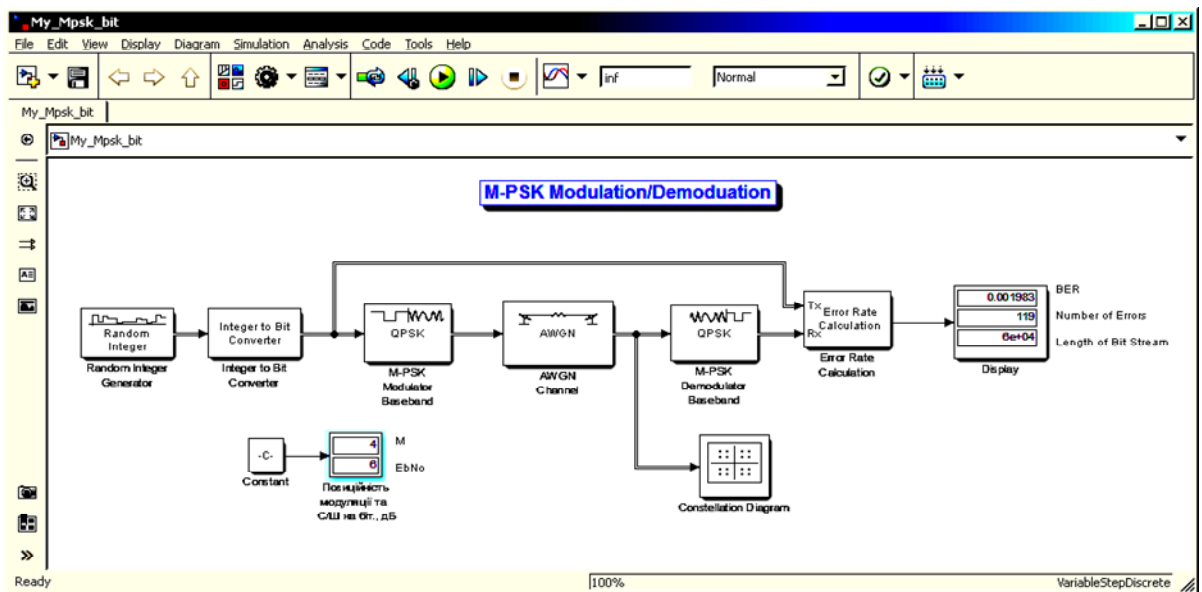


Рис. 18.15. S-модель системи передачі комплексної обвідної сигналу із модуляцією M-PSK.

Перед запуском моделі у робочому просторі Matlab треба задати її основні параметри: позиційність модуляції M та відношення сигнал/завада на один біт E_b/N_0 . Після натискання на кнопку запуску поточні значення цих параметрів відобразяться у віконцях нижнього монітору та почнеться набір статистики. По закінченні процесу набору статистики у віконцях монітору, що підключений до виходу лічильника, одержимо оцінку частоти бітових помилок, загальну їх кількість та довжину серії випробовувань.

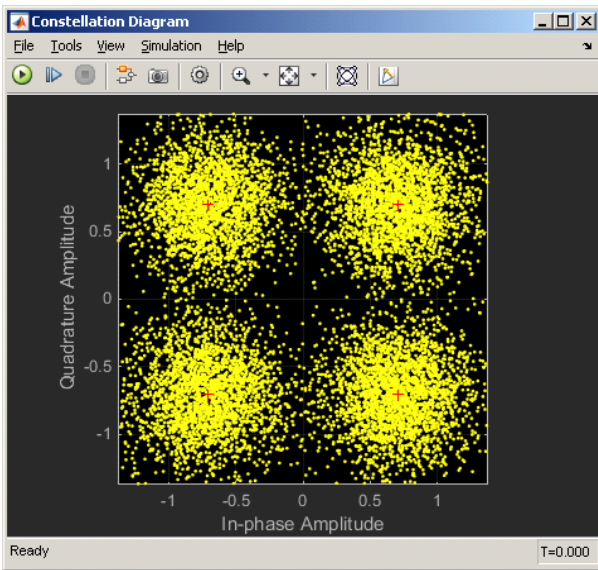


Рис.18.16. Діаграма розсіювання для модуляції QPSK.

Також для наочності виводиться діаграма розсіювання (блок Constellation Diagram), що представлена на рис.18.16. Параметри основних блоків даної моделі представлено на рис. 18.17.

На відміну від попередньої моделі рис.18.11, у даному разі модельний час встановлено безперервним, що відповідає значенню Variable Step на вкладці Solver. Але стани моделі при цьому визначено дискретними. Що ж стосується структури

моделі, то вона за своєю побудовою фактично повторює програмний код табл.18.2, 18.3, який розглядався вище. Єдина принципова різниця (якщо не брати до уваги графічний інтерфейс) полягає в тому, що тут, на відміну від коду табл.18.2, 18.3, застосовується представлення вхідного інформаційного потоку у бітовій формі. Вказана різниця зумовлена особливостями роботи лічильника помилок бібліотеки Comm, про яку вже було сказано раніше. А саме, при надходженні на його вхід символів (тобто, цілих чисел) лічильник підраховує символні помилки, а при надходженні бітового потоку – BER. У попередньому параграфі ми вже обговорювали цю особливість лічильника помилок та застосували перерахунок символних помилок у BER відповідно до наближеного співвідношення (18.1). У моделі рис.18.15 застосовано інший метод, а саме, моделювання передачі інформації безпосередньо виконується на рівні бітового потоку для чого символи по виходу інформаційного джерела перетворюються у біти. Відповідно, у параметрах налаштування усіх блоків моделі (див. рис.18.17) ця обставина належним чином врахована. Це, звісно, не означає, що таким же методом не можна було скористатися й при моделюванні у командному режимі: усі необхідні для цього утиліти у бібліотеці Comm передбачені. Просто це було зроблено для ілюстрації можливостей різних підходів.

а) Генератор випадкових чисел

Source Block Parameters: Random Integer Generator

Random Integer Generator (mask) (link)

Generate random uniformly distributed integers in the range $[0, M-1]$, where M is the M-ary number.

Parameters

M-ary number:

Initial seed:

Sample time:

☒ Frame-based outputs

Samples per frame:

Output data type:

OK Cancel Help Apply

б) Перетворювач символів у біти

Function Block Parameters: Integer to Bit Converter

Integer to Bit Converter (mask) (link)

Map a vector of integer-values inputs to a vector of bits. Block inputs must be integer values in the range $[-2^{(M-1)}, 2^{(M-1)}-1]$ when they are treated as signed and $[0, 2^M-1]$ when they are treated as unsigned. For fixed-point inputs, the stored integer value is used.

Parameters

Number of bits per integer (M):

Treat input values as:

Output bit order:

Output data type:

OK Cancel Help Apply

в) Модулятор

Function Block Parameters: M-PSK Modulator Baseband

M-PSK Modulator Baseband

Modulate the input signal using the phase shift keying method.

This block accepts a scalar or column vector input signal.

The input signal can be either bits or integers. When you set the 'Input type' parameter to 'Bit', the input width must be an integer multiple of the number of bits per symbol.

Main Data Types

Parameters

M-ary number:

Phase offset(rad):

Constellation ordering:

Input type:

View Constellation

OK Cancel Help Apply

г) Канал АБГШ

Function Block Parameters: AWGN Channel

complex. This block supports multichannel processing.

When using either of the variance modes with complex inputs, the variance values are equally divided among the real and imaginary components of the input signal.

Parameters

Input processing:

Initial seed:

Mode:

Eb/No (dB):

Number of bits per symbol:

Input signal power, referenced to 1 ohm (watts):

Symbol period (s):

OK Cancel Help Apply

д) Демодулятор

Function Block Parameters: M-PSK Demodulator Baseband

M-PSK Demodulator Baseband

Demodulate the input signal using the phase shift keying method.

This block accepts a scalar or column vector input signal.

When you set the 'Output type' parameter to 'Integer', the block always performs Hard decision demodulation.

When you set the 'Output type' parameter to 'Bit', the output width is an integer multiple of the number of bits per symbol. In this case, the 'Decision type' parameter allows you to select 'Hard decision' demodulation, 'Log-likelihood ratio' or 'Approximate log-likelihood ratio'. The output values for Log-likelihood ratio and Approximate log-likelihood ratio decision types are of the same data type as the input values.

Main Data Types

Parameters

M-ary number:

Phase offset (rad):

Constellation ordering:

Output type:

Decision type:

OK Cancel Help Apply

е) Лічильник помилок

Function Block Parameters: Error Rate Calculation

The 'Stop simulation' option stops the simulation upon detecting a target number of errors or a maximum number of symbols, whichever comes first.

Parameters

Receive delay:

Computation delay:

Computation mode:

Output data:

☐ Reset port

☒ Stop simulation

Target number of errors:

Maximum number of symbols:

OK Cancel Help Apply

Рис.18.17. Параметри моделі рис.18.15

Послідовно запускаючи дану модель при різних значеннях її параметрів, можна одержати відповідні залежності BER та порівняти їх з одержаними раніше результатами, що були наведені на рис.18.8. Неважко переконатися, що для прийнятих при моделюванні значень параметрів співпадіння результатів є досить високим. Але слід все-таки зазначити, що розглянутий процес моделювання є досить трудомістким та не дуже зручним для практичного використання. У цьому зв'язку час згадати про інструмент BERTool, який ми вже розглядали вище. Дійсно, як буде показано, цей інструмент значно полегшує виконання статистичного моделювання та наступний аналіз результатів. Його переваги особливо відчутні при моделюванні саме у середовищі Simulink. Зокрема, у цьому випадку засіб BERTool бере на себе автоматичну зміну параметру відношення сигнал/завада, повторний запуск моделі, зупинку набору статистичних даних при виконанні певних критеріїв. По завершенні моделювання його результати представляються у вигляді графіків у вікні BERTool, де їх можна порівнювати з результатами теоретичного аналізу або якогось іншого моделювання.

Розглянемо застосування інструменту BERTool до моделі рис.18.15. Перш за все модель має бути скорегована так, щоб забезпечити можливість її сумісної роботи з цим інструментом. Засоби керування випробуваннями моделі, які містяться у вкладці Monte Carlo інструменту BERTool, показано на рис.18.18. Як бачимо з цього рисунка, тут задано діапазон змін параметру відношення сигнал/шум на біт (дБ), назва моделі, яка піддається випробуванням та значення параметрів зупинки випробувань. Що ж стосується S-моделі, яка підлягає тестуванню, то вона має задовольняти наступним вимогам:

- Параметр відношення сигнал/шум на біт у блоках моделі має бути задано через формальну змінну E_b/N_0 , яка допускає можливість керування нею у робочому просторі Matlab;
- Критерії зупинки набору статистики визначаються параметрами `maxNumErrs` та `maxNumBits`. Ці параметри у вигляді формальних змінних мають входити до складу параметрів лічильника помилок. Випробування при

кожному значенні параметру E_b/N_0 тривають доти, поки загальна кількість помилок не досягне заданого значення параметру $\maxNumErrs$ або загальна довжина тестової послідовності не перевищить $\maxNumBits$ бітів.

- Слід передбачити вивід результатів підрахунку помилок при кожному значенні E_b/N_0 із виходу лічильника у робочий простір Matlab у вигляді 3-х елементного вектору під певною назвою, який містить значення BER, кількість помилок та загальну довжину відповідної тестової бітової послідовності;

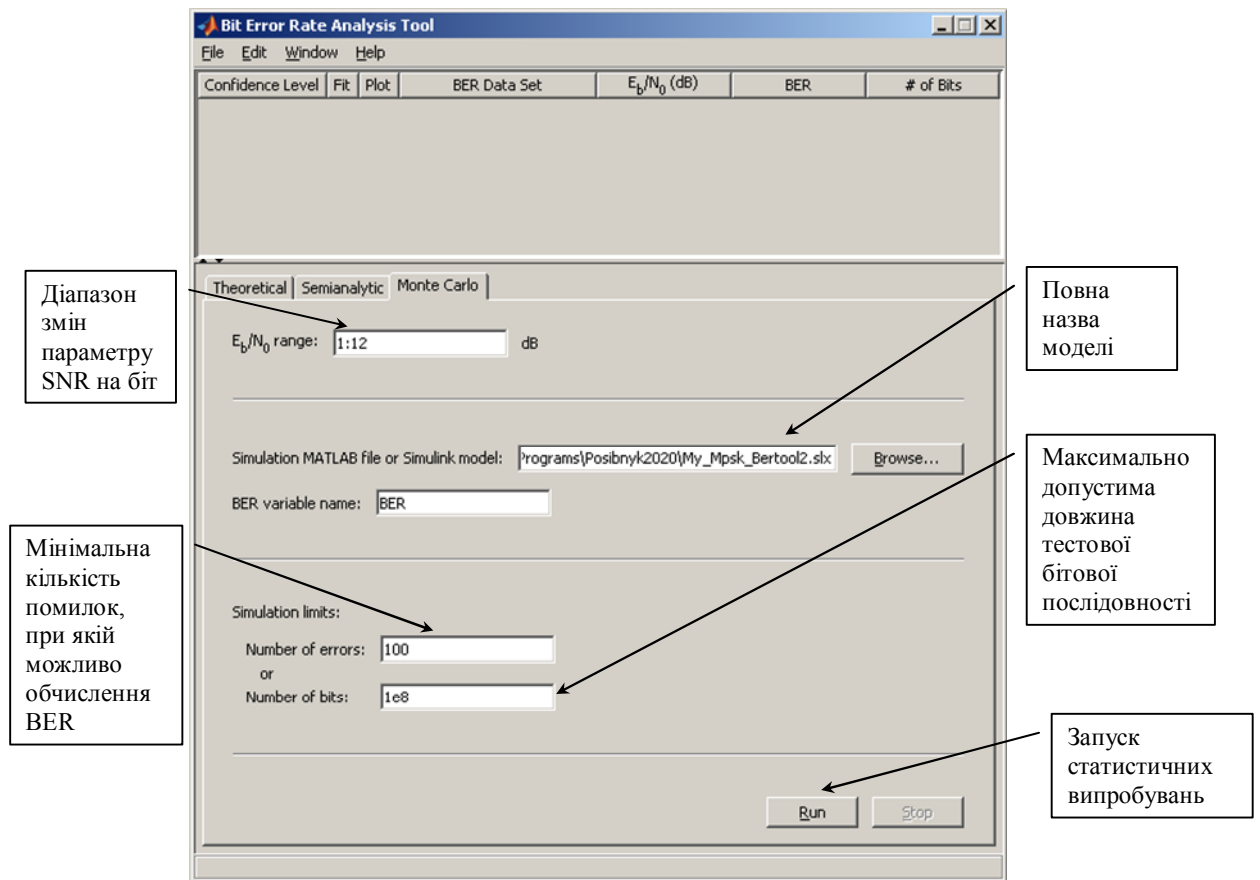


Рис.18.18. Вкладка статистичного моделювання інструменту BERTool.

- Перед початком статистичних випробувань треба задати початкові значення вказаних вище параметрів у вигляді наступної командної строки:

$E_b/N_0 = 0$; $\maxNumErrs = 100$; $\maxNumBits = 1e8$;

(Конкретні значення параметрів можуть бути, звісно, й іншими в залежності від вибору користувача). Для зручності цю командну строку можна також включити у функцію попередніх завантажень моделі у наступному вигляді:

set_param('повна назва моделі у робочому каталозі', 'preLoadFcn', 'EbNo = 0;...maxNumErrs = 100; maxNumBits = 1e8;'); Якщо потім зберегти модель, то при її наступному завантаженні вказані параметри будуть ініційовані автоматично.

На рис.18.19 показано S-модель, яка задовольняє усім вказаним вимогам та яка одержана корегуванням її прототипу, що був розглянутий на рис.18.15. У цій моделі у порівнянні з прототипом зроблено наступні зміни:

- Введено додатковий блок виводу даних у робочий простір To Workspace з каталогу⁹ DSP System Toolbox/Sinks.
- Для економії обчислювальних затрат зі складу моделі-прототипу виключено блок побудови діаграм розсіювання.

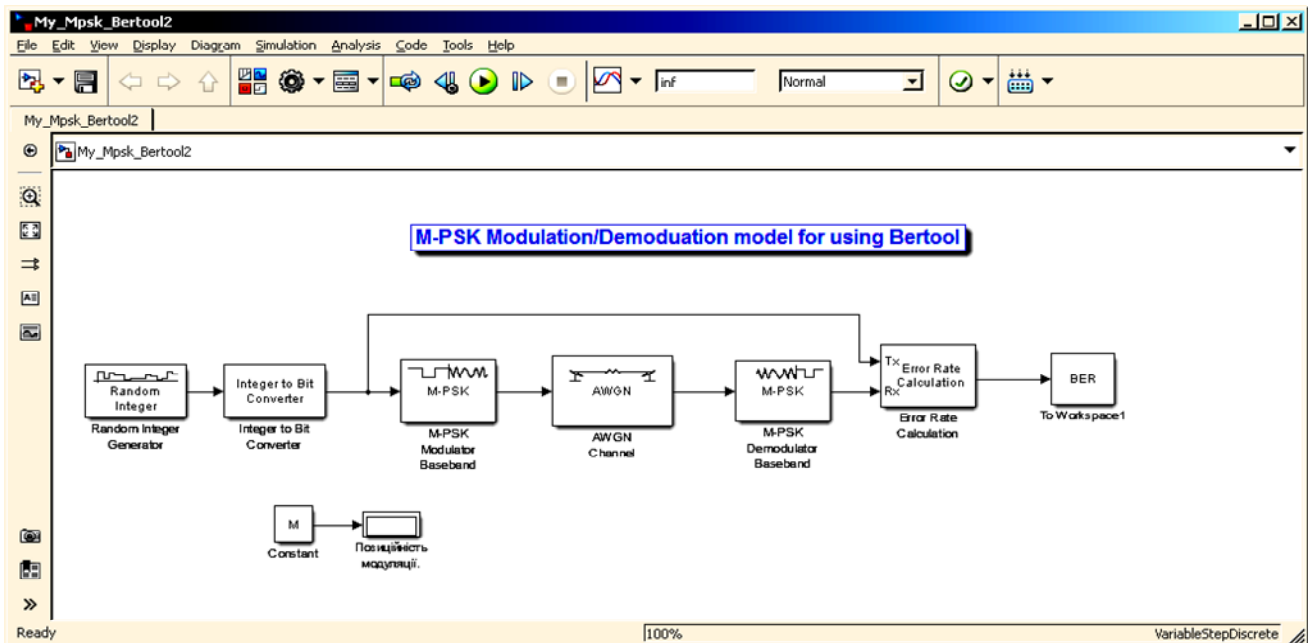


Рис. 18.19. S-модель системи передачі комплексної обвідної сигналу із модуляцією M-PSK, що пристосована для роботи зі засобом BerTool.

Параметри усіх блоків відповідають тим же значенням, що й на рис.18.17. Перед початком кожної серії випробувань у робочому просторі Matlab треба задавати позиційність модуляції M . Тепер залишається лише викликати інструмент BerTool та відкрити вкладку Monte Carlo (див. рис. 18.18). Після цього для початку моделювання достатньо натиснути на кнопку Run. У графічному вікні BerTool

⁹ Аналогічні блоки є й у деяких інших каталогах, але через певні розбіжності в інтерфейсі документація Matlab рекомендує саме цю версію блоку.

почнуть з'являтися результати статистичних випробувань для усіх заданих значень параметру E_b/N_0 . Наступний крок звичайно полягає у порівнянні результатів моделювання із даними теоретичного аналізу. Для цього слід перейти у вкладку Theoretical, задати там необхідні параметри системи передачі та натиснути кнопку Plot. В результаті для, наприклад, значення $M = 4$ одержимо графіки, що показані на рис.18.20.

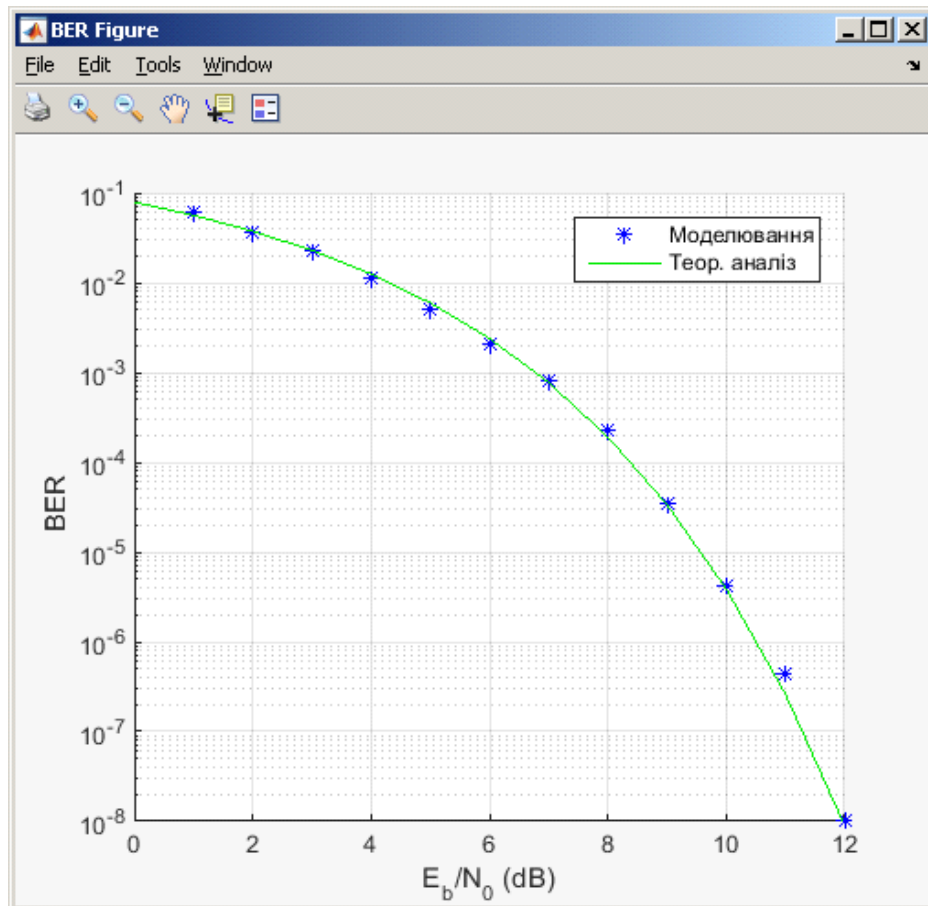


Рис.18.20. Результати статистичних випробувань моделі рис.18.19 для позиційності модуляції $M=4$.

Порівняння цих результатів із даними рис.18.8 показує їх практично повну ідентичність із результатами, що були одержані раніше шляхом моделювання у командному режимі.

Контрольні запитання та завдання до розділу 18.

Дати в конспекті письмові відповіді на запитання та навести рішення задач.

1. У чому полягає завдання аналізу ефективності демодуляції сигналів? Що таке BER та в залежності від яких параметрів завадостійкості цей показник

обчислюється? Який зв'язок існує поміж відношеннями сигнал/завада на біт, на символ та відношення сигнал/завада за потужністю?

2. За допомогою засобів теоретичного аналізу BERTool самостійно повторити графіки рис.18.2. Які закономірності витікають із аналізу цих графіків?

3. Самостійно повторити у середовищі Матлаб програмний код табл.18.1 та побудувати графіки рис.18.3. Які закономірності витікають із аналізу цих графіків?

4. За допомогою засобів теоретичного аналізу BERTool провести порівнювальне дослідження ефективності демодуляції сигналів від одного чи кількох факторів: вид модуляції, спосіб демодуляції (когерентна, некогерентна), наявність завадостійкого кодування та його параметрів, якість синхронізації, тощо.

5. Розказати на Ваш вибір про деякі з каталогів пакету *Comm*: призначення, склад, найбільш цікаві та корисні з Вашої точки зору об'єкти, функції, тощо.

6. Проаналізувати та пояснити програмний код табл.18.2, 18.3. Самостійно його повторити у середовищі Матлаб та повторити графіки рис.18.6, 18.7 та 18.8. Які закономірності витікають із аналізу цих графіків? Як впливає на результати аналізу параметр програми `symlen` та як це можна пояснити?

7. Як забезпечити спільну роботу програми моделювання та інструменту BERTool? Самостійно повторити у середовищі Матлаб програмний код табл.18.4, виконати повний цикл моделювання та порівняти одержані результати з графіками рис.18.8.

8. Що таке середовище Simulink та в яких «взаємовідносинах» воно знаходиться із середовищем Матлаб? Які воно має переваги та для яких завдань його доцільно застосовувати?

9. Повторити самостійно модель рис.18.11. Пояснити призначення окремих блоків моделі та особливості їх застосування. Якими параметрами визначається модельний час та як він регулюється? Запустити вказану модель для різних типів сигналів та зробити відповідні копії екранів осцилографа та спектроаналізатора. Пояснити параметри моделі, що визначають її налаштування та які увійшли до меню налаштувань окремих блоків.

10. Яким вимогам має задовольняти S-модель для того, щоб вона могла працювати під управлінням інструменту BERTool?

11. Самостійно повторити модель рис.18.19 та провести повний цикл її випробувань для позиційності модуляції PSK 4, 8, 16. Порівняти одержані результати з графіками рис.18.8.

Список літератури.

1. Калюжний О.Я. Математичні алгоритми приймання та оброблення сигналів. Частина І. Сигнали та системи. Електронне мережне навчальне видання: https://oleksa-site.blogspot.com/p/blog-page_27.html
2. Баруча-Рид А. Т. Элементы теории марковских процессов и их приложения. М., Физматгиз, 1969, 512 с.
3. Ван Трис Г. Теория обнаружения, оценок и модуляции. – Т. I. Пер. с англ. под ред. В.И.Тихонова. М., «Советское радио», 1972, 744 с.
4. Ван Трис Г.Л. Теория обнаружения, оценок и модуляции. – Т. III: Обработка сигналов в радио- и гидролокации / Пер. с англ. под ред. В.Т. Горяинова. – М.: Сов. радио, 1977. – 662 с.
5. Демидович Б.П., Марон И.А. Основы вычислительной математики. М. – Физматгиз, 1970, 664 с.
6. Закс Ш. Теория статистических выводов. М.: Мир, 1975. — 776 с.
7. Ильичев В.И. и др. Статистическая теория обнаружения гидро-акустических сигналов. – М.: Наука, 1992. – 415 с.
8. Караваев В.В., Сазонов В.В. Статистическая теория пассивной локации. М.: Радио и связь, 1987, 240 с.
9. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М., 1970, 720 с.
10. Кремер И.Я. и др. Пространственно-временная обработка сигналов. М.: Радио и связь, 1984, 224 с.
11. Куликов Е. И., Трифонов А. П. Оценка параметров сигналов на фоне помех. – М.: Сов. Радио, 1978. – 296 с.
12. Кукес И.С., Старик М.Е. Основы радиопеленгации. М.: Сов. Радио, 1964, 640 с.
13. Монзинго Р.А., Миллер Т.У. Адаптивные антенные решетки. Введение в теорию: Пер. с англ. – М.: Радио и связь, 1986. – 448 с.
14. Петров Б.М. Электродинамика и распространение радиоволн: Учебник для вузов. - 2-е изд., испр. -М.: Горячая линия-Телеком, 2007. – 558 с
15. Прокис Дж. Цифровая связь. Пер. с англ. под ред. Д.Д. Кловского. М., Радио и связь, 2000, 800 с.
16. Рытов С.М., Кравцов Ю.А., Татарский В.И. Введение в статистическую радиофизику. Часть 2: Случайные поля. М.: Наука, 1978. – 464 с.

17. Сейдж Э., Мелс Дж. Теория оценивания и ее применение в связи и управлении. Пер. с англ. под ред. проф. Б.Р. Левина. М., Связь, 1976, 496 с.
18. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. М., Физматгиз, 1969, 512 с.
19. Складар Б. Цифровая связь. Теоретические основы и практическое применение. Пер. с англ. – М. Издательский дом «Вильямс», 2003. – 1104 с.
20. Сосулин Ю.Г. Теоретические основы радиолокации и радионавигации. Учебн. пособие для вузов. – М. Радио и связь, 1992, 304 с.
21. Справочник по антенной технике: в 5 т. Т.1 / Л.Д. Бахрах, Л.С. Бенинсон, Е.Г. Зелкин и др. Под ред. Я.Н. Фельда, Е.Г. Зелкина. – М.: ИПРЖР, 1997. – 256 с.
22. Струченков В.И. Методы оптимизации в прикладных задачах. М.: Солон-пресс, 2009, 320 с.
23. Теоретические основы радиолокации / под ред. проф. Дулевича В.Е. Москва, Сов. Радио, 1964. - 732 с.
24. Фалькович С.Е. Оценка параметров сигнала. М., Изд-во «Советское радио», 1970, 336 с.
25. Van Trees H.L. Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory. J.Wiley & Sons, Inc., New York, 2002. – 1470 p.
26. Mengali Umberto, D'Andrea Aldo N. Synchronization Techniques for Digital Receivers. Plenum Press, New York and London, 1997. – 529 p.
27. Гордиенко В.А. Векторно-фазовые методы в акустике. – М.: ФИЗМАТЛИТ, 2007. – 480 с.