

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ ТА ІНФОРМАТИЗАЦІЇ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОКОМУНІКАЦІЙ

НАВЧАЛЬНИЙ ПОСІБНИК
ДЛЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ З ДИСЦИПЛІНИ
“ТЕОРІЯ ЕЛЕКТРИЧНОГО ЗВ’ЯЗКУ”

БЕРКМАН Л.Н., КОМАРОВА Л.О., ЧУМАК Н.С.
СИСТЕМИ ЕЛЕКТРОЗВ’ЯЗКУ ТА СИГНАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ ТА ІНФОРМАТИЗАЦІЇ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОКОМУНІКАЦІЙ

НАВЧАЛЬНИЙ ПОСІБНИК
ДЛЯ САМОСТІЙНОЇ РОБОТИ СТУДЕНТІВ З ДИСЦИПЛІНИ
“ТЕОРІЯ ЕЛЕКТРИЧНОГО ЗВ’ЯЗКУ”

БЕРКМАН Л.Н., КОМАРОВА Л.О., ЧУМАК Н.С.
СИСТЕМИ ЕЛЕКТРОЗВ’ЯЗКУ ТА СИГНАЛИ

Затверджено вченою радою
факультету Інформаційних технологій
навчально-наукового інституту
Телекомунікацій та інформатизації
Державного університету телекомунікацій
як навчальний посібник для студентів вищих
навчальних закладів за напрямом
05903 «Телекомунікації»
(Протокол №5 від 26 січня 2015 р.)

УДК 621.391.13
В 685
ББК 32.811

Рецензенти: проф., д.т.н. Ю.Г. Савченко
проф., д.т.н. В.В. Поповський

Навчальний посібник призначений для самостійної роботи студентів вищих навчальних закладів з навчальної дисципліни “Теорія електричного зв’язку” (ТЕЗ) - циклу дисциплін професійної та практичної підготовки за напрямом 05903 “ТЕЛЕКОМУНІКАЦІЇ”.

Навчальна дисципліна “Теорія електричного зв’язку” вивчається протягом двох семестрів. Матеріал відповідає програмі дисципліни, сформовано відповідно чотири складових навчального посібника:

Беркман Л.Н., Комарова Л.О., Чумак Н.С. Системи електрозв’язку та сигнали.

Беркман Л.Н., Варфоломєєва О.Г., Грушевська В.П. Типові сигнали та завади в електрозв’язку.

Беркман Л.Н., Комарова Л.О., Чумак О.І. Основні поняття та теореми теорії інформації.

Жураковський Б.Ю., Варфоломєєва О.Г., Твердохліб М.Г. Основи теорії потенційної завадостійкості.

Беркман Л.Н., Комарова Л.О., Чумак Н.С. Системи електрозв’язку та сигнали. Навчальний посібник підготовлено для самостійної роботи студентів вищих навчальних закладів – Київ: ДУТ ННІТІ, 2015. – 132 с.

У посібнику розглянуто основи сучасної теорії електрозв’язку з акцентом на фізичне тлумачення процесів, які відбуваються під час передавання повідомлень та сигналів у системах зв’язку. Загальні поняття про системи електрозв’язку, їх класифікація та характеристики.

Наведено математичні моделі дискретних та неперервних каналів, перетворення випадкових сигналів у типових лінійних та нелінійних ланках каналів. Представлено метод простору станів, форми розв’язання рівнянь стану, методи розрахунків характеристик випадкових сигналів на виході каналів та аналітичний підхід до обчислення матриці переходу.

Викладено основні поняття теорії ймовірності. Розглянуто випадкові процеси, закони розподілу випадкових величин.

Контрольні запитання і задачі допоможуть студенту в підготовці до перевірки знань.

Навчальний посібник призначений для студентів, які навчаються за спеціальностями з напрямку “Телекомунікації”, а також може бути корисний для аспірантів, викладачів навчальних закладів відповідних спеціальностей, фахівців, які обслуговують телекомунікаційні мережі зв’язку.

ЗМІСТ

С.

ПЕРЕДМОВА	5
Розділ 1. ЗАГАЛЬНІ ПОНЯТТЯ ПРО СИСТЕМИ ЕЛЕКТРОЗВ'ЯЗКУ ТА СИГНАЛИ	9
Контрольні запитання для самооцінки рівня знань	25
Післямова до розділу 1	25
Розділ 2. МАТЕМАТИЧНІ МОДЕЛІ ДЕТЕРМІНОВАНИХ СИГНАЛІВ	28
2.1. Види детермінованих сигналів	28
2.2. Часове зображення сигналів. Дискретизація інформації	32
2.3. Квантування за рівнем.....	43
2.4. Геометричні зображення сигналів	46
2.5. Розклад сигналів в узагальнений ряд Фур'є	54
2.6. Аналітичний сигнал	56
Контрольні запитання для самооцінки рівня знань.....	59
Післямова до розділу 2	59
Розділ 3. КАНАЛИ ЗВ'ЯЗКУ	61
3.1. Загальні відомості про канали зв'язку.....	61
3.2. Математичні моделі каналів зв'язку.....	62
3.3. Метод простору станів	66
3.4. Форми розв'язання рівнянь стану.....	81
3.5. Аналітичний підхід до обчислення матриці переходу.....	83
Контрольні запитання для самооцінки рівня знань.....	91
Післямова до розділу 3.....	91
Розділ 4. ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ І ВИПАДКОВІ ПРОЦЕСИ	93
4.1. Предмет теорії ймовірностей	93
4.2. Основні поняття теорії ймовірності.....	95
4.3. Закони розподілу випадкових величин.....	101
4.4. Числові характеристики випадкових величин	105
4.5. Закон розподілу щільності ймовірності випадкових величин.....	109
4.6. Випадкові процеси.....	117
Контрольні запитання для самооцінки рівня знань	128
Післямова до розділу 4.....	129
Задачі для перевірки знань	129
Список літератури, рекомендованої для поглибленого вивчення дисципліни	130

ПЕРЕДМОВА

Навчальні посібники призначено для вивчення дисципліни “Теорія електричного зв’язку” (ТЕЗ) – циклу дисциплін професійної та практичної підготовки для студентів вищих навчальних закладів усіх форм навчання за напрямом 05903 “ТЕЛЕКОМУНІКАЦІЇ”.

Навчальні посібники відповідають програмі дисципліни ТЕЗ.

Посібники призначено для студентів, які вперше знайомляться з процесами передавання сигналів в інформаційних системах. Творчий колектив авторів намагалися максимально використовувати термінологію доступну і зрозумілу студентам 2 курсу. Робота з навчальними посібниками не виключає використання інших підручників та посібників, список яких подано в кожному наприкінці.

Предметом навчальної дисципліни є:

- основи сучасної теорії електрозв’язку з акцентом на фізичне тлумачення процесів, які відбуваються під час передавання повідомлень та сигналів у системах зв’язку; математичний опис основних фізичних процесів передавання сигналів та методи забезпечення граничних характеристик систем зв’язку як за достовірністю, так і за швидкістю передачі інформації; загальні принципи модуляції; методи цифрової модуляції; теорема Котельникова;

- процеси передавання сигналів каналами зв’язку при наявності завад з математичної точки зору. Методи ефективного кодування. Теорема Шеннона для каналу з завадами;

- оптимальний прийом сигналів. Принципи побудови багатоканальних модемів, багатопозиційні сигнали і їх застосування у високошвидкісних модемах;

- принципи побудови телекомунікаційних мереж; цифрові методи передачі неперервних повідомлень. Основи теорії лінійного розділу сигналів. Пропускна здатність багатоканальних систем. Архітектура сучасних телекомунікаційних мереж.

Метою вивчення навчальної дисципліни є:

- з’ясування фундаментальних понять інформаційної інфраструктури для спеціалістів з телекомунікацій;

- опанування основними термінами, категоріями, базовими знаннями із сучасної теорії електричного зв’язку, використання і оцінювання у своїй практичній діяльності математичних моделей процесів (у тому числі сигналів, каналів зв’язку) для розв’язання виробничих, проектних та наукових задач з телекомунікацій;

- здатність свідомого застосування правил, методів, принципів, законів у конкретних ситуаціях, можливість своєчасно адаптуватися до зростаючого потоку інформації, проблем розвитку галузі зв’язку та новітніх науково-технічних досягнень в галузі телекомунікацій;

- сформувати у випускників активну позицію (за вимогами до сучасних спеціалістів), спрямовану на практичну реалізацію важливих завдань - інформатизації держави та входження до глобальної інфраструктури.

Завданнями навчальної дисципліни є формування наступних умінь:

- уміти характеризувати системи електрозв’язку, типові сигнали та завади, характеризувати та оцінювати основні характеристики сигналів, аналізувати переваги і недоліки конкретних видів модуляції, проводити розрахунки характеристик випадкових сигналів на виході каналів, чітко характеризувати процес дискретизації сигналів за теоремою Котельникова;

- уміти розуміти сутність та оцінювати значення, здійснювати класифікацію каналів електрозв’язку, характеризувати математичні моделі дискретних та неперервних каналів, диференціювати, інтегрувати, уніфікувати, оцінювати зміст теорії передачі інформації каналами телекомунікаційних мереж, проводити аналітичні експерименти із використанням математичних моделей, методик визначення параметрів окремих частин та системи передавання інформації в цілому;

- уміти узгоджувати параметри джерела та каналу зв'язку, застосовувати завадостійкі коди та методи оптимальної демодуляції, реалізувати алгоритми оптимального прийому багатопозиційних сигналів багатоканальних модемів;

- уміти аналізувати процес завадостійкості систем електрозв'язку з різними видами модуляції та методами приймання, застосовувати теорію лінійного розподілу сигналів, проводити аналіз архітектури сучасних телекомунікаційних мереж, окремих технічних рішень на мережах зв'язку і прогнозувати очікувані результати.

Теми за робочою програмою в навчальних посібниках

Беркман Л.Н., Комарова Л.О., Чумак Н.С. СИСТЕМИ ЕЛЕКТРОЗВ'ЯЗКУ ТА СИГНАЛИ

Тема 1. Загальні поняття про системи електрозв'язку та сигнали

Вступ. Мета та задачі курсу ТЕЗ. Загальні поняття про системи електрозв'язку. Досягнення сучасної теорії та техніки зв'язку.

Класифікація, узагальнені структурні схеми. Поняття сигналу, основні характеристики первинних сигналів. Елементи простого кодування.

Випадкові процеси. Числові характеристики випадкових величин. Закони розподілу випадкових величин.

Спектральна щільність потужності та її зв'язок із функцією кореляції. Низькочастотний та смуговий Гауссівський шум.

Тема 2. Характеристики каналів електрозв'язку

Канали електрозв'язку, їх класифікація та характеристики. Математичні моделі дискретних та неперервних каналів.

Ідеальний канал без завад. Канал з адитивним гауссовим шумом. Канал з невизначеною фазою сигналу. Однопроменевий гауссів канал із загальними завмираннями. Канал з міжсимвольною інтерференцією (МСІ) і адитивним шумом.

Задачі для перевірки знань

Беркман Л.Н., Варфоломеєва О.Г., Грушевська В.П. ТИПОВІ СИГНАЛИ ТА ЗАВАДИ В ЕЛЕКТРОЗВ'ЯЗКУ

Тема 3. Основні характеристики сигналів Теорема Котельникова

Математичний опис сигналів та завад. Класифікація, енергетичні та кореляційні характеристики сигналів.

Подання сигналів в ортогональному базисі. Приклади ортогональних базисів. Спектральний аналіз сигналів. Амплітудний, фазовий, комплексний та енергетичний спектри сигналів. Геометричне зображення сигналів.

Теорема відліків. Динамічне уявлення сигналів.

Методи цифрової модуляції. Теорема Котельникова.

Перетворення аналогових сигналів у дискретну та цифрову форму. Відновлення сигналів. ІКМ. Похибки квантування.

Тема 4. Загальні принципи модуляції.

Модульовані сигнали. Амплітудна модуляція (АМ) гармонічного переносника. Часове та спектральне зображення АМ сигналу. Модуляція гармонічним та складним сигналами. БМ та односмугова модуляції. Принципи формування АМ, БМ, та ОМ сигналів. Синхронне детектування. Детектор обвідної.

Кутова модуляція гармонічного переносника. Часове та спектральне зображення сигналів кутових модуляцій. Принципи формування та детектування сигналів кутових модуляцій.

Шумоподібні сигнали та їхні спектри. Основні типи широкосмугових систем зв'язку (ШПС), їх основні структурні схеми.

Задачі для перевірки знань

**Беркман Л.Н., Комарова Л.О., Чумак О.І.
ОСНОВНІ ПОНЯТТЯ ТА ТЕОРЕМИ ТЕОРІЇ ІНФОРМАЦІЇ**

Тема 5. ТЕОРІЯ ПЕРЕДАЧІ ІНФОРМАЦІЇ КАНАЛАМИ ТЕЛЕКОМУНІКАЦІЙНИХ МЕРЕЖ

Кількісна міра інформації, її властивості. Ентропія джерела дискретних повідомлень та її властивості. Інформаційні характеристики джерел дискретних повідомлень. Передача повідомлень каналами з шумами. Взаємна інформація та її властивості.

Швидкість передачі інформації і пропускна здатність дискретного каналу. Методи ефективного кодування. Теорема Шеннона для каналу з завадами.

Інформаційні характеристики джерел неперервних повідомлень. Пропускна здатність неперервного каналу. Епсілон-ентропія, продуктивність, надмірність.

Тема 6. ТЕОРІЯ КОДУВАННЯ

Призначення і класифікація кодів. Лінійні двійкові блочні коди. Основні параметри кодів, їх коригуюча здатність. Принципи завадостійкого кодування.

Декодування з виявленням та виправленням помилок. Систематичні коректувальні коди. Породжуюча та перевірна матриці. Синдромне декодування. Коди Хемінга.

Тема 7. ЦИКЛІЧНІ КОДИ

Циклічні коди. Кодування та декодування циклічними кодами. Застосування твірних матриць, твірних поліномів. Приклади реалізації коректуючих та декодуючих пристроїв.

Задачі для перевірки знань

**Жураковський Б.Ю., Варфоломеева О.Г., Твердохліб М.Г.
ОСНОВИ ТЕОРІЇ ПОТЕНЦІЙНОЇ ЗАВАДОСТІЙКОСТІ**

Тема 8. Аналіз завадостійкості систем електровз'язку з різними видами модуляції та методами приймання

Оптимальний прийом сигналів. Критерії якості і правила прийому дискретних повідомлень. Методи обробки дискретних сигналів: накопичення, інтегрування, фільтрація, кореляційний та автокореляційний прийом.. Алгоритми оптимального прийому, їх реалізація.

Принципи побудови багатоканальних модемів. Багатопозиційні сигнали і їх застосування у високошвидкісних модемах. Алгоритми прийому багатопозиційних сигналів.

Тема 9. ЦИФРОВІ МЕТОДИ ПРИЙОМУ НЕПЕРЕРВНИХ ПОВІДОМЛЕНЬ

Критерії завадостійкості прийому неперервних повідомлень.

Неоптимальні методи приймання дискретних сигналів та їх порівняння з оптимальними методами. Аналіз завадостійкості систем електрозв'язку з різними видами модуляції та методами приймання. Критерії оптимальної демодуляції аналогових сигналів. Оптимальна лінійна фільтрація неперервних сигналів. Фільтр Колмогорова-Вінера. Оптимальна лінійна фільтрація неперервних повідомлень. Фільтр Калмана.

Тема 10. ЦИФРОВІ МЕТОДИ ПЕРЕДАЧІ НЕПЕРЕРВНИХ ПОВІДОМЛЕНЬ

Сучасні системи багатоканального зв'язку. Принципи побудови телекомунікаційних мереж. Цифрові методи передачі неперервних повідомлень. Пропускна здатність багатоканальних систем. Системи зі зворотним зв'язком.

Архітектура сучасних телекомунікаційних мереж.

Задачі для перевірки знань

В додатку А даються методичні вказівки для виконання курсової роботи.

У кожному навчальному посібнику після розділу запропоновано студентам контрольні запитання для самооцінки засвоєного рівня знань. Наведені типові задачі і приклади побудовані так, щоб студенти, які навчаються за вказаними спеціальностями, здобули чіткі уявлення і знання про загальні системи електрозв'язку, типові сигнали та завади, могли проводити розрахунки характеристик випадкових сигналів на виході каналів, аналізувати архітектури сучасних телекомунікаційних мереж, окремих технічних рішень на мережах зв'язку і прогнозувати очікувані результати. Якщо при самооцінці засвоєння рівня знань відповіді Вас не задовольняють, то необхідно повторити опрацювання матеріалу або прийти на консультацію до викладача.

Послідовність розділів відповідає послідовності вивчення матеріалів за навчальною програмою дисципліни.

ЗАГАЛЬНІ ПОНЯТТЯ ПРО СИСТЕМИ ЕЛЕКТРОЗВ'ЯЗКУ ТА СИГНАЛИ

Загальні відомості про теорію інформації та передавання сигналів. Під *інформацією* розуміють будь-які відомості про події, факти, явища природи, гіпотези і т. д. Передавання інформації за допомогою мови, листів, газет, книжок, телеграфу, телефону, радіо, кіно, телебачення, телекомунікаційних мереж, систем стільникового зв'язку, іншими способами є однією з найважливіших сторін повсякденного життя людини.

Процеси передавання, перетворення і використання інформації для цілеспрямованих дій, чи, як ми домовимося називати їх надалі, – *інформаційні процеси*, можуть відбуватися не тільки у разі взаємодії людини з оточуючим середовищем. Накопичення досвіду (інформації) дає можливість людині створювати машини, що допомагають ефективно впливати на природу, керувати промисловістю, сільським господарством, транспортом тощо і отримувати нову інформацію. Тому обмін інформацією може відбуватися між людиною та створеними нею машинами, і по суті, навіть між машинами.

У технічних пристроях інформація завжди передається в формі сигналів. Слово “сигнал” походить від латинського “*signum*”, що означає знак. Сигнали – це знаки, про зміст яких заздалегідь домовилися. Наприклад, зелене світло світлофора означає сигнал “Шлях вільний”, червоне світло – сигнал “Стій!” і т. д. Вид сигналу являє собою продукт умовної угоди і не має безпосереднього відношення до тієї інформації, що передається з його допомогою.

У всіх інформаційних процесах сигнали, що несуть інформацію, завжди втілені в будь-яку матеріальну форму. Інформація передається по телеграфу, телефону і радіо у вигляді електричних сигналів; вона записується в книжках, статтях і на магнітній стрічці; інформація міститься в показаннях приладів, формулах, таблицях і т. ін. Одне і те саме повідомлення може передаватися різними сигналами: за допомогою тексту, мови людини, фотографії, електричних коливань і т. д. Однак, якою би там не була фізична природа сигналів, що використовуються для передавання інформації, у разі ідеального передавання (за відсутності завад) завжди існує взаємно однозначна відповідність між переданим повідомленням і сигналом, який несе це повідомлення.

Характерною особливістю процесу передавання інформації є *незмінність інформації*, хоча сигнал, який її несе, може зазнавати значних змін. Наприклад, передана по радіо інформація спочатку існує у вигляді тексту, який перетворюється диктором у звуки мови, котрі, в свою чергу, перетворюються в електричні коливання; передані по радіоканалу зв'язку електричні коливання відтворюються на виході радіоприймача у вигляді звуків мови. Незважаючи на багаторазові перетворення сигналів, які несуть інформацію, в остаточному підсумку на виході радіоприймача ми одержуємо початково передану інформацію.

Таким чином, передавання повідомлень (а отже, й інформації) на відстань здійснюється за допомогою *матеріального носія* (паперу, магнітної стрічки і т. д.) чи *фізичного процесу* (звукових або електромагнітних хвиль, струму тощо). *Фізичний процес* (несуча), що відображає передане повідомлення, називається **сигналом**.

Сигналом може бути будь-який фізичний процес, що змінюється відповідно до повідомлення, яке переноситься. В сучасних системах управління і зв'язку найчастіше застосовують *електричні сигнали*. Фізичною величиною, що визначає такий сигнал, є *сила струму* чи *напруга*. Сигнали формуються шляхом зміни тих чи інших параметрів фізичного носія відповідно до переданого повідомлення. Процес зміни параметрів носія прийнято називати *модуляцією*.

Сигнал передає (розгортає) повідомлення в часі. Отже, він завжди є *функцією часу*, навіть коли повідомлення (наприклад, нерухоме зображення) таким не є. Якщо сигнал являє собою функцію $x(t)$, що набуває тільки визначених дискретних значень x (наприклад, 1 і 0), то його називають *дискретним* чи *дискретним за рівнем* (амплітудою). Так само і повідомлення, що набуває тільки деяких визначених рівнів, називають дискретним. Сигнал (чи повідомлення), який набуває будь-яких рівнів у деякому інтервалі, називається *неперервним* чи *аналоговим*.

У деяких випадках повідомлення або сигнал задають не на всій осі часу, а у визначені моменти t . Такі повідомлення (сигнали) називають *дискретними за часом* на відміну від *неперервних за часом*, заданих на всій осі t . Наприклад, мова – це повідомлення неперервне як за рівнем, так і за часом, а датчик температури, що видає її значення кожні 5 хв, є джерелом повідомлень, неперервних за величиною, але дискретних за часом. Сигнал із скінченим числом дискретних рівнів часто називають *цифровим*, оскільки рівні можна пронумерувати цифрами зі скінченим числом розрядів. Не слід думати, що дискретні повідомлення обов'язково перетворюються в дискретні сигнали, а неперервні повідомлення – у неперервні сигнали. Найчастіше саме неперервні сигнали використовують для передавання дискретних повідомлень як їх переносники, несучі. Дискретні ж сигнали можуть використовуватися для передавання неперервних повідомлень (після їх дискретизації).

Вирішення проблем ефективності та надійності систем зв'язку в теорії інформації ґрунтується на вивченні *статистичних властивостей* переданих повідомлень і *аналізі загальних закономірностей* передавання повідомлень по каналах зв'язку за наявності завад. Розроблені в даний час у теорії інформації методи дозволяють істотно підвищити надійність і ефективність систем зв'язку. Особливо широке застосування ці методи знаходять у радіоелектроніці й автоматичці.

За функціональною ознакою можна виділити *основні класи об'єктів інформаційної техніки*: мережі та системи електрозв'язку (телеграфні, телефонні, телевізійні мережі та системи, системи радіомовлення, системи передавання даних і т. ін.); інформаційно-вимірювальні системи й комплекси (радіолокаційні, радіонавігаційні, радіотехнічні системи автоматичного літаководіння, радіотелеметричні тощо); системи перетворення інформації (цифрові й аналогові обчислювальні машини, аналого-цифрові та цифро-аналогові перетворювачі і т. ін.); підсистеми інформаційного забезпечення автоматизованих систем управління; інформаційно-пошукові системи; системи збереження інформації; мережі та системи експериментального спостереження чи дослідження (астрономічні, фізичні, медичні тощо).

Найбільшого розвитку методи теорії інформації й передавання сигналів одержали для аналізу й синтезу систем електрозв'язку, тому що саме з задач аналізу найпростіших систем електрозв'язку і почалася теорія інформації. Цей напрямок теорії інформації та передавання сигналів називають *статистичною теорією зв'язку*. В даний час методи теорії інформації й передавання сигналів усе успішніше застосовують для аналізу й синтезу інших об'єктів інформаційної техніки.

Повідомлення за допомогою спеціальних пристроїв (датчиків), як правило, перетворюється в електричну величину $b(t)$ – *первинний сигнал*. У разі передавання мови таке перетворення виконує мікрофон, у разі передавання зображення – телевізійна камера. У більшості випадків первинний сигнал є *низькочастотним коливанням*, що відображає передане повідомлення. (Варто звернути увагу на умовність цього терміну: первинний телевізійний сигнал, наприклад, займає смуга частот в діапазоні від 0 до 6 МГц).

У деяких випадках (наприклад, при звичайному міському телефонному зв'язку) первинний сигнал безпосередньо передають по лінії. Для передавання на великі відстані (по кабелю чи радіоканалу) первинний сигнал перетворюють у *високочастотний*.

Якби передане повідомлення було детермінованим, тобто заздалегідь відомим, то передавання його не мало би сенсу, оскільки не містило б інформації. Повідомлення слід розглядати як *випадкові події* (чи випадкові величини, випадкові функції). Іншими словами, має існувати деяка множина варіантів повідомлення (наприклад, множина різних значень температури, виданих датчиком), з яких реалізується з визначеною ймовірністю одне. Тому і *сигнал є випадковою функцією*. Детермінований же сигнал не може бути носієм інформації. Його можна використовувати лише для випробування системи зв'язку або окремих її елементів.

Випадковий характер повідомлень, сигналів, а також завад зумовив найважливіше значення теорії ймовірностей у побудові теорії зв'язку. Як буде показано далі, ймовірнісні властивості сигналів і повідомлень, а також середовища, в якому передається сигнал, дозволяють визначити кількість переданої інформації та її втрати.

Передавання інформації завжди пов'язане з енергетичними процесами, тобто процесами передавання та перетворення енергії, однак при цьому об'єктом передавання є не енергія, а повідомлення. Саме цей аспект передавання інформації є загальним як для живих організмів, так і для технічних пристроїв.

Теорія інформації – наука про загальні закони передавання інформації – як самостійна галузь знання виникла наприкінці 40-х років ХХ сторіччя. Особливість цієї теорії насамперед полягає в тому, що в ній з єдиної точки зору розглядаються закони передавання інформації в різних за принципом дії та фізичною природою системах (живі організми, машини і т. д.). Тому теорія інформації відразу привернула до себе увагу фахівців самих різних профілів: радіотехніків, математиків, біологів, лінгвістів, соціологів і фахівців ряду інших галузей науки й техніки.

Зародження і розвиток теорії інформації як науки стали можливими після того, як була визначена *кількісна міра інформації*, що дозволила поставити вирішення всіх основних проблем передавання інформації на строгу математичну основу. Першими працями, в яких були зроблені спроби кількісного вимірювання інформації, стали праці Р. Хартлі (1927 р.) і В. О. Котельникова (1937 р.). Наукові основи теорії інформації були закладені в працях К. Шеннона, опублікованих у 1948 р., і розвинуті в працях Д. А. Агеєва, О. Я. Хінчина, В. О. Котельникова, А. М. Колмогорова, С. Гольдмана, Д. Міддлтона, Л. Бріллоена, Н. Вінера та ряду інших учених.

Основними проблемами теорії інформації є ефективність і надійність систем передавання інформації, тобто систем зв'язку. *Проблема ефективності систем зв'язку* полягає в забезпеченні передавання найбільшої кількості інформації по каналах зв'язку найбільш економічним шляхом – при мінімальних витратах технічних засобів у системі зв'язку.

Проблема надійності систем зв'язку полягає в забезпеченні найбільшої вірогідності переданої інформації в системі зв'язку при впливі на неї завад, тобто в забезпеченні найбільшої завадостійкості системи зв'язку.

Структура системи зв'язку. Сукупність елементів, що забезпечують передавання повідомлень від джерела повідомлень, називають *системою зв'язку*. Узагальнена структурна схема системи зв'язку, що відображає процес передавання інформації від джерела повідомлень до приймача повідомлень, показана на рис. 1.1.

До складу системи зв'язку входять *перетворювач повідомлення в сигнал*, *лінія зв'язку*, по якій відбувається передавання сигналу, і *перетворювач сигналу в повідомлення*, що надходить у приймач повідомлень.

Реальні системи зв'язку завжди працюють в умовах впливу завад, що можуть мати як природне, так і штучне походження. Тому до системи зв'язку входять *джерела завад*, які на узагальненій структурній схемі системи зв'язку об'єднані в одне еквівалентне джерело завад.

Процес перетворення повідомлення в сигнал називається кодуванням.

Лінія зв'язку являє собою фізичне середовище, в якому передається (поширюється) сигнал.

Процес перетворення сигналу в повідомлення приймальною стороною системи зв'язку називають декодуванням.

Джерело повідомлень і приймач повідомлень не входять безпосередньо до системи зв'язку, проте властивості їх істотно впливають на властивості та параметри системи, а тому джерело повідомлень і приймач повідомлень розглядаються разом із системою зв'язку.

На практиці мають місце випадки, коли по одній лінії зв'язку одночасно передаються сигнали від кількох джерел повідомлень до кількох приймачів повідомлень. У таких лініях зв'язку сигнали від різних джерел повідомлень розділяються між собою і неначе йдуть окремими каналами зв'язку. Тому такі лінії зв'язку називають *багатоканальними*. Прикладами багатоканальних ліній зв'язку є лінії телевізійного мовлення, в яких по одній лінії зв'язку передаються сигнали зображення і сигнали звукового супроводу, або проводові лінії зв'язку з ущільненням, в яких по одній парі проводів одночасно передається кілька телефонних розмов.

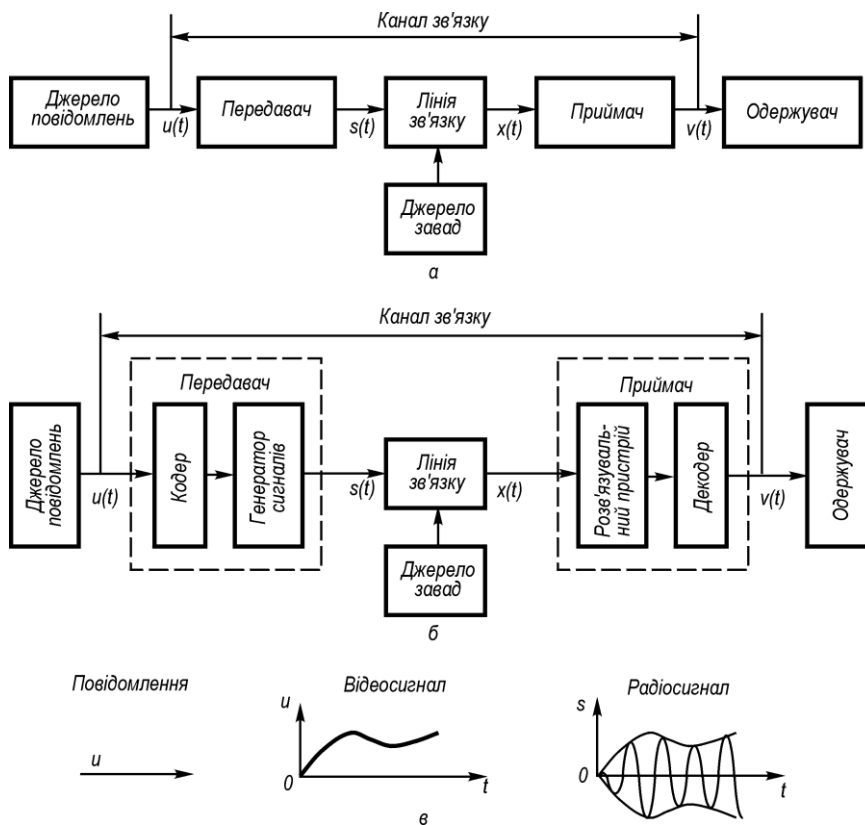


Рис. 1.1. Функціональні схеми системи зв'язку (а, б) і перетворення повідомлення в сигнал (в)

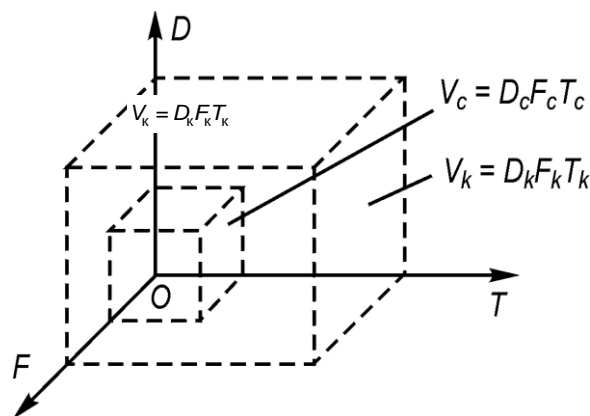


Рис. 1.2. Геометричне зображення об'ємів сигналу і каналу

Досліджуючи передавання сигналів по багатоканальних лініях зв'язку, слід враховувати не тільки умови передавання сигналів по кожному з каналів, але й взаємний вплив між каналами зв'язку.

Основні характеристики сигналів. Повідомлення, передані системою зв'язку, можуть бути величинами, які або неперервно змінюються в часі, або мають дискретний характер. Відповідно до цього сигнали, що відображають повідомлення, можуть мати також дискретний чи неперервний характер.

Дискретні сигнали є послідовністю символів, що розрізняються між собою і обрані з деякої скінченної множини елементарних символів A_1, \dots, A_n . Передбачається, що кожний із

символів A_i має певну тривалість у часі (τ_i секунд). Найпростіший приклад дискретних сигналів – телеграфні сигнали, що застосовуються для передавання чисел чи тексту.

Неперервні сигнали являють собою неперервні функції часу, два різних значення яких можуть відрізнятися одне від одного як завгодно мало. Приклади неперервних сигналів – телефонні сигнали, сигнали зображення в телебаченні.

За характером сигналів, що передаються, системи зв'язку поділяються на *дискретні системи зв'язку* і *неперервні системи зв'язку*.

Основними параметрами сигналу, що визначають його властивості як переносника інформації в системах зв'язку, є тривалість сигналу (T_c), смуга частот, займана спектром сигналу (F_c), і динамічний діапазон (D_c). Динамічний діапазон D_c – це енергетична міра сигналу, яка визначається логарифмом відношення середньої потужності сигналу до середньої потужності завад на вході лінії зв'язку:

$$D_c = \log \frac{P_c}{P_3}.$$

Добуток основних параметрів сигналу

$$V_c = T_c F_c D_c \quad (1.1)$$

називають *об'ємом сигналу*. Об'єм сигналу є узагальненою характеристикою сигналу як носія інформації. На практиці іноді використовується геометричне зображення об'єму сигналу у вигляді прямокутного паралелепіпеда в тривимірному просторі з координатами T , F і D (рис. 1.2).

За аналогією з сигналами канал зв'язку, застосовуваний для передавання сигналів, характеризують наступними основними параметрами: часом роботи каналу зв'язку T_k , смугою частот пропускання каналу зв'язку F_k і динамічним діапазоном D_k (перевищенням середньої потужності сигналу над середньою потужністю завад на виході каналу зв'язку):

$$D_k = \log \frac{P_c}{P_3}.$$

Добуток основних параметрів каналу зв'язку

$$V_k = T_k F_k D_k \quad (1.2)$$

називають *об'ємом (ємністю) каналу зв'язку*.

Основна умова узгодження каналу зв'язку з сигналом, виконання якої забезпечує можливість неспотвореного передавання сигналу, така:

$$T_k \geq T_c; \quad F_k \geq F_c; \quad D_k \geq D_c. \quad (1.3)$$

Повідомлення має бути перетворено в сигнал так, щоб об'єм сигналу в тривимірному просторі T , F , D відповідав об'єму каналу в тому самому тривимірному просторі, тобто: час передавання сигналу має збігатися з часом роботи каналу зв'язку; смуга частот сигналу має збігатися зі смугою частот каналу зв'язку; динамічні діапазони на вході і виході каналу зв'язку мають дорівнювати один одному. Тільки в цьому випадку канал зв'язку може забезпечити неспотворене передавання сигналу.

Якщо об'єм сигналу менший за об'єм каналу зв'язку або дорівнює йому, то завжди можна здійснити таке перетворення сигналу, при якому виконуватимуться умови (1.3). Отже, умова можливості узгодження сигналу і каналу зв'язку

$$V_k \geq V_c. \quad (1.4)$$

Відношення

$$r_k = \frac{V_k}{V_c} \quad (1.5)$$

називається *резервом ємності каналу зв'язку*. Резерв ємності каналу зв'язку характеризує можливість підвищення надійності каналу зв'язку.

Дискретні сигнали. Для опису дискретних сигналів і повідомлень у теорії інформації широко використовуються поняття абстрактного алфавіту та слів в абстрактному алфавіті.

Абстрактний алфавіт – це будь-яка скінченна сукупність об'єктів, які називають *літерами даного алфавіту*. Літерами абстрактного алфавіту можна вважати, наприклад, літери певної мови, цифри, ієрогліфи, малюнки і т. д. При бажанні можна ввести абстрактний алфавіт, за літери якого правитимуть цілі слова. Важливо лише, щоб алфавіт був скінченним, тобто складався зі скінченного числа літер. Будь-який абстрактний алфавіт можна розширити, включивши до нього нові додаткові літери.

Словом в абстрактному алфавіті називається будь-яка скінченна упорядкована послідовність літер. Наприклад, в алфавіті $A(x, y)$, що складається з двох літер (x і y), словами вважатимуться будь-які послідовності, що складаються з цих літер: $x, xy, yy, xxyx, \dots$. В алфавіті $A(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)$, що складається з десяти літер, словом є будь-яке ціле число. Число літер у слові називається *довжиною слова*. Наприклад, слово xuu має довжину, рівну трьом, слово 13572 – довжину, рівну п'яти, і т. д.

Кожний із символів, які входять у дискретне повідомлення, можна розглядати як літеру деякого абстрактного алфавіту, а саме повідомлення – як слово чи сукупність слів у цьому алфавіті. Прикладами дискретних повідомлень є слова тексту або команди, закодовані числами. Інтервали між словами теж можна вважати за літеру алфавіту, а повідомлення у вигляді слів у цьому розширеному алфавіті та довжину слова – за *довжиною повідомлення*.

Дискретні сигнали аналогічно з дискретними повідомленнями також можуть розглядатися як слова в деякому абстрактному алфавіті. Абстрактний алфавіт сигналів у загальному випадку може не збігатися з алфавітом, використовуваним для зображення повідомлень.

Оскільки сигнали є відображенням повідомлень, то існує однозначна відповідність між словами з алфавіту повідомлень і словами з алфавіту сигналів. Таке однозначне відображення слів може бути задане за допомогою алфавітного оператора.

Алфавітним оператором (чи *алфавітним відображенням*) називається будь-яка відповідність (функція), що зіставляє слова у тому чи іншому алфавіті зі словами в тому ж самому або деякому іншому фіксованому алфавіті. Перший алфавіт називається при цьому вхідним, а другий – вихідним алфавітом даного оператора. Сукупність усіх слів, на яких алфавітний оператор визначений, називається *областю визначення алфавітного оператора*. Алфавітний оператор може бути визначений на всіх словах чи на частині слів вхідного алфавіту.

Дискретні сигнали, передані по лініях зв'язку, є алфавітним відображенням повідомлень. Зв'язок між повідомленнями і сигналами задається відповідним алфавітним оператором. Алфавітний оператор визначає алгоритм перетворення повідомлень у сигнали, тобто сукупність правил, за якими здійснюється це перетворення.

Кодування дискретних повідомлень. Процес перетворення повідомлень у сигнали в теорії інформації називається *кодувальним відображенням* або *кодуванням*, а алгоритм, що визначає правила, за якими здійснюється це перетворення, – *кодом*.

Алфавітний оператор, що визначає код перетворення повідомлень у сигнали, може бути заданий у вигляді таблиці або у вигляді правил, які визначають порядок перетворення повідомлень у сигнали.

За *табличним способом завдання алфавітного оператора* складається таблиця, в ліву частину якої вписуються всі слова з області визначення даного алфавітного оператора, а в праву частину – слова, що утворюються внаслідок застосування даного алфавітного оператора до кожного слова лівої частини таблиці.

Правила, за допомогою яких визначається порядок перетворення повідомлень у сигнали, можуть бути задані у вигляді формули чи інструкції, котра визначає порядок одержання сигналу з повідомлення.

Найпростішим видом кодування є *кодування за літерами*, коли кожній літері алфавіту повідомлення $A(x_1, \dots, x_n)$ відповідає деяка скінченна послідовність літер (слово) в алфавіті сигналу $B(y_1, \dots, y_m)$, яка називається кодом відповідної літери. Різним літерам алфавіту мають відповідати різні коди.

При передаванні повідомлень по системах зв'язку обов'язковою вимогою є вимога *зворотності кодування*, котра полягає в тім, щоб за кодом будь-якого слова можна було б відновити спочатку закодоване слово. Тільки у разі виконання цієї вимоги можна за сигналом відновити передане повідомлення.

Для того щоб визначити умови зворотності кодування слів за літерами, введемо поняття про *початковий відрізок слова*, заданого в певному алфавіті. Домовимося називати слово p початковим відрізком слова q , якщо слово q можна подати у вигляді $q = pl$, де l – будь-яке слово, в тому числі й слово нульової довжини. Наприклад, слово $p = xy$ є початковим відрізком слова $q = xyux$, тому що слово q можна подати у вигляді $q = pl$, де $l = ux$.

Для того щоб кодування слів за літерами в алфавіті $A(x_1, \dots, x_n)$ було зворотним, необхідними і достатніми є такі умови:

1. Коди різних літер вихідного алфавіту мають відрізнятися між собою.
2. Код кожної з літер вихідного алфавіту не повинен збігатися з жодним із початкових відрізків кодів інших літер цього алфавіту.

Друга умова зворотності кодування завжди виконується, якщо коди всіх літер вихідного алфавіту мають однакову довжину. В цьому випадку в коді слова q досить виділити початковий відрізок s_1 , що має довжину, рівну довжині коду літери, і за ним визначити першу літеру слова p . Аналогічно можуть бути знайдені й всі інші літери слова p . Кодування за літерами, при якому коди всіх літер вихідного алфавіту мають рівну довжину, називається *нормальним* або *рівномірним кодуванням*.

Кодування дозволяє перетворювати слова, подані в будь-якому довільному алфавіті, на слова в обраному стандартному алфавіті. Найчастіше за такий стандартний алфавіт береться *двійковий алфавіт*, що складається з двох літер, які звичайно ототожнюються з цифрами 0 і 1.

Покажемо принципову можливість такого перетворення. Нехай $A(x_1, \dots, x_n)$ – довільний алфавіт, що містить n літер, а $B(y_1, \dots, y_m)$ – стандартний алфавіт, що містить m літер ($m > 1$). У цьому разі завжди можна вибрати ціле додатне число k так, щоб задовольнялася нерівність

$$m^k > n. \quad (1.6)$$

Оскільки число різних слів довжиною k в алфавіті, що складається з m літер, дорівнює m^k , то нерівність (В.6) показує, що можна закодувати всі літери алфавіту $A(x_1, \dots, x_n)$ словами довжиною k і алфавіту $B(y_1, \dots, y_m)$ так, що коди різних літер відрізнятимуться між собою. Будь-яке таке кодування буде нормальним, а отже, зворотним.

Найпростішими прикладами кодування за літерами повідомлень є коди Бодо і Морзе, що застосовуються при телеграфуванні.

Алфавіт сигналів коду Бодо складається з двох літер (0 і 1). Код кожної літери повідомлення складається з п'яти літер алфавіту сигналу, що дозволяє передати 32 різні літери ($2^5 = 32$). Коди деяких літер українського алфавіту для цього випадку наведені нижче:

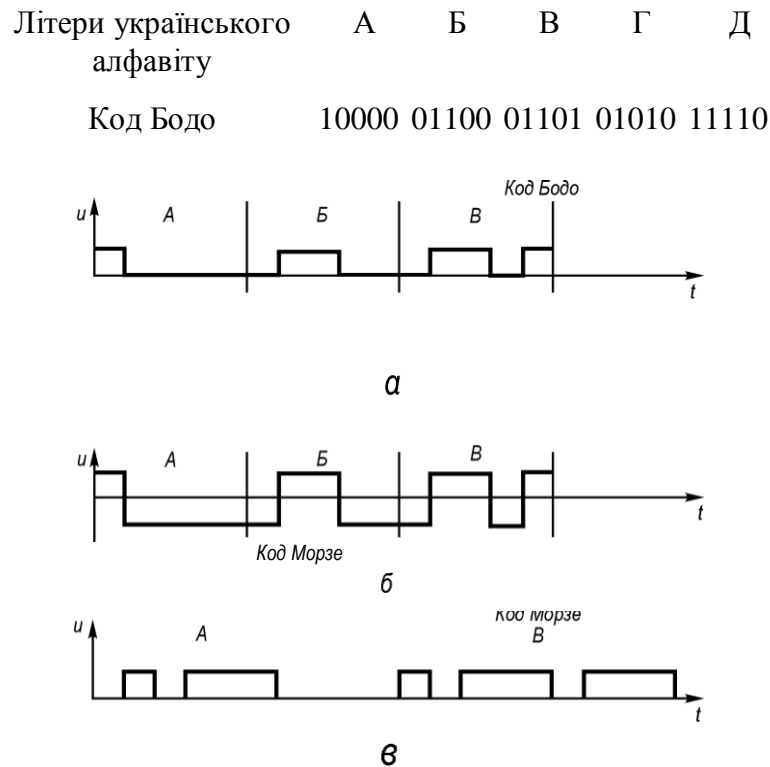


Рис. 1.3. Варіанти побудови коду Бодо (а, б) та коду Морзе (в)

Код Бодо є рівномірним кодом. При побудові сигналів у кодї Бодо для представлення одиниці використовується послілка струму в лінію зв'язку, а для представлення нуля – відсутність послілки струму (рис. 1.3, а). Можливі й інші варіанти (рис. 1.3, б).

У кодї Морзе алфавіт сигналу складається з чотирьох літер: точки, тире, паузи між літерами і паузи між словами. Цей код є нерівномірним, тобто коди букв вихідного алфавіту мають у ньому різну довжину. Наприклад, слово “КІЇВ” у цьому кодї має такий вигляд:

К	И	Ї	В
-•-	••	•	•--

При побудові сигналів у кодї Морзе точки відповідає коротка послілка струму і рівна їй за тривалістю пауза, тире – послілка струму потроєної тривалості і коротка пауза, інтервалу між літерами – пауза подвоєної тривалості, інтервалу між словами – пауза збільшеної в чотири рази тривалості (рис. 1.3, в).

Для передавання великих літер українського алфавіту (їх 32) необхідно передати числа від 0 до 31. Щоб передати будь-яке число, записане в десятковій формі, потрібно передати десять цифр – від 0 до 9. Практично для цього потрібно десять сигналів, які відповідають різним цифрам. Систему передавання дискретних повідомлень можна істотно спростити, скориставшись при кодуванні *двійковою системою числення*.

Щоб з'ясувати зміст двійкової системи, звернемося до десяткової системи, основою числення якої є число 10. Будь-яке ціле число K можна подати у вигляді

$$K = a_n \cdot 10^n + \dots + a_2 \cdot 10^2 + a_1 \cdot 10^1 + a_0 \cdot 10^0, \quad (1.7)$$

де a_0, a_1, \dots, a_n – коефіцієнти, що набувають значень від 0 до 9. Так, число 265 можна записати як $2 \cdot 10^2 + 6 \cdot 10^1 + 5 \cdot 10^0$. Очевидно, замість числа 10 можна взяти будь-яке інше ціле число m і подати число K так:

$$K = a_n m^n + \dots + a_2 m^2 + a_1 m^1 + a_0 m^0, \quad (1.8)$$

де a_0, a_1, \dots, a_n – коефіцієнти, що набувають значень від 0 до $m-1$.

Задаючи величину m , можна побудувати будь-яку систему числення.

При $m = 2$ одержимо двійкову систему, в якій числа записуються за допомогою двох цифр – 0 і 1. Наприклад, число 13 у двійковій системі записується 1101, що відповідає виразу $1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$. Арифметичні дії в двійковій системі дуже прості. Так, додавання здійснюється за наступними правилами: $0+0=0$; $0+1=1$; $1+0=1$; $1+1=10$. Розрізняють ще порозрядне додавання без перенесення в старший розряд, або “додавання за модулем 2”. Правила цього додавання наступні: $0 \oplus 0 = 0$; $0 \oplus 1 = 1$; $1 \oplus 0 = 1$; $1 \oplus 1 = 0$.

Якщо перетворити послідовність елементів повідомлення в послідовність двійкових чисел, то для передавання останніх по каналу зв'язку достатньо лише два різних сигнали. Наприклад, символи 0 і 1 можуть передаватися коливаннями з різними частотами чи імпульсами струму різної полярності. Завдяки своїй простоті двійкова система числення широко застосовується для кодування дискретних повідомлень.

Кодування – це процес перетворення елементів повідомлення у відповідні їм числа (кодові символи). Кожному елементу повідомлення присвоюється певна сукупність кодових символів, яка називається *ковою комбінацією*. Сукупність кодових комбінацій, що відображають дискретні повідомлення, утворює *код*. Правило кодування може бути виражене кодовою таблицею, в якій наводяться алфавіт повідомлень, що кодуються, і відповідні їм кодові комбінації. Множина можливих кодових символів називається *ковим алфавітом*, а їх кількість m – основою коду. У загальному випадку при основі коду m правила кодування K елементів повідомлення зводяться до правил запису K різних чисел у m -ковій системі числення. Число розрядів n , що утворюють кодову комбінацію, називається *розрядністю коду* чи *довжиною кодової комбінації*. У залежності від системи числення, використовуваної при кодуванні, розрізняють двійкові і m -кові (недвійкові) коди.

Коди, в яких усі комбінації мають однакову довжину, називають *рівномірними*. Для рівномірного коду число можливих комбінацій дорівнює m^n . Прикладом такого коду є п'ятизначний код Бодо, що містить п'ять двійкових елементів ($m=2, n=5$). Число можливих кодових комбінацій ($2^5=32$) є достатнім для кодування всіх літер українського алфавіту, проте недостатнім для передавання повідомлення, яке містить букви інших алфавітів, цифри, різні умовні знаки (точка, кома, додавання, множення і т. ін.). Тому на даний час використовується “Міжнародний код №2” (МТК-2), побудований за реєстровим принципом, відповідно до якого та сама п'ятиелементна кодова комбінація може застосовуватися до трьох разів у залежності від положення регістра: український, латинський, цифровий. Загальне число різних знаків при цьому дорівнює 84, що є достатнім для кодування телеграми. Для передавання даних рекомендовано семиелементний код МТК-5. Коди МТК-2 і МТК-5 є первинними (простими). Основними *параметрами кодів* є: основа коду m , довжина кодової комбінації n , відстань між кодовими комбінаціями d_{ij} і вага кодової комбінації ω . Відстань d_{ij} характеризує розбіжність між двома кодовими комбінаціями і визначається за Хеммінгом числом незбіжних розрядів у них, тобто числом одиниць у сумі двох комбінацій за модулем 2. Число ненульових елементів у кодовій комбінації визначає її вагу ω . Застосування рівномірних кодів спрощує побудову автоматичних літеродрукувальних пристроїв і не потребує передавання розділових символів між кодовими комбінаціями.

Нерівномірні коди характерні тим, що в них кодові комбінації відрізняються одна від одної не тільки взаємним розташуванням символів, але й їхньою кількістю, тому різні комбінації мають різну тривалість. Такі коди потребують або спеціальних розділових знаків, що вказують кінець однієї й початок іншої кодової комбінації, або ж такої побудови, щоб жодна кодова комбінація не була початком іншої. Коди, які задовольняють цій умові, називаються *незвідними* (префіксними). Зазначимо, що рівномірний код також є незвідним. Будувати код зручно у вигляді графа (кодового дерева), з кожного вузла якого виходить кількість гілок, яка дорівнює основі коду (для двійкового коду, наприклад, крок наверх означає 0, крок униз – 1).

Типовим прикладом нерівномірних кодів є код Морзе, в якому символи 0 і 1 використовуються тільки в двох сполученнях – як одиночні (1 і 0) або як потрійні (111 і 000). Сигнал, що відповідає одній одиниці, називається точкою, трьома одиницям – тире. Символ “0” використовується як знак, що відокремлює точку від тире, точку від точки і тире від тире. Сукупність “000” використовується як розділовий знак між кодовими комбінаціями.

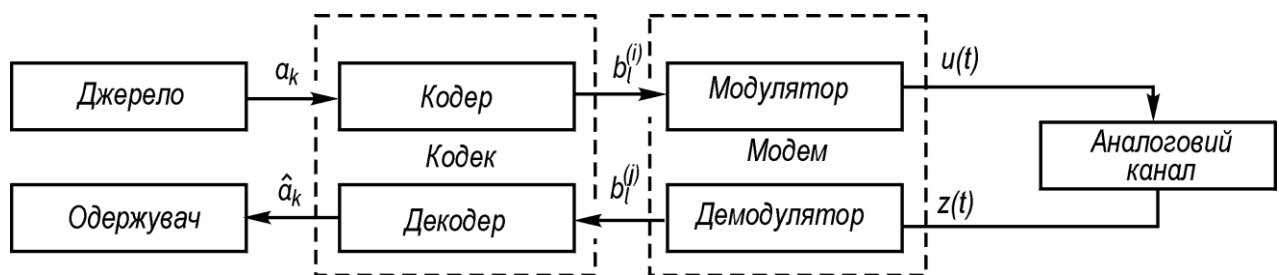


Рис. 1.4. Структурна схема системи передавання дискретних повідомлень

За ознакою завадозахищеності коди поділяють на дві групи. Коди, в яких усі можливі кодові комбінації використовуються для передавання інформації, називаються *простими* або *кодами без надлишковості* (примітивними). В простих рівномірних кодах перетворення одного символу комбінації на іншій, наприклад 1 на 0 чи 0 на 1, призводить до появи нової дозволеної комбінації, тобто до помилки. *Коректувальні коди* будуються так, що для передавання повідомлення використовуються не всі кодові комбінації, а лише деяка їх частина (дозволені кодові комбінації). Тим самим створюється можливість виявлення і виправлення помилки у разі неправильного відтворення деякого числа символів. Коректувальні властивості кодів досягаються введенням у кодові комбінації додаткових (надлишкових) символів.

Декодування полягає у відновленні повідомлення за прийнятими кодовими символами. Пристрої, що здійснюють кодування і декодування, називають відповідно *кодером* і *декодером*. Як правило, це логічні пристрої. На рис. В.4 зображена структурна схема системи передавання дискретних повідомлень, а на рис. В.5 пояснюється процес перетворення дискретного повідомлення в сигнал. Передане повідомлення позначено a_k , кодоване повідомлення (чи первинний цифровий сигнал) – $b_{ll}(t)$, його компоненти – $b_l^{(i)}$ (l – номер послідовно переданого символу, i – номер позиції коду, $i = \overline{0, m-1}$). Сигнал, що надходить у лінію зв'язку, позначено $u(t)$, прийняте коливання – $z(t)$, відновлена послідовність кодових символів – $\hat{b}_{ll}(t)$ (її компоненти – $\hat{b}_l^{(i)}$) і декодоване (відновлене) повідомлення – \hat{a}_k . Позначення прийнятих сигналів, кодових символів і відновленого повідомлення інші, ніж позначення переданих. Цим підкреслюється те, що через дію завад прийнятий сигнал відрізняється від переданого, а відновлене повідомлення може не збігатися з первинним.

У сучасних системах передавання дискретних повідомлень розрізняють дві групи самостійних пристроїв: кодеки і модеми. *Кодек* – це пристрій, що перетворює повідомлення в код (*кодер*) і код у повідомлення (*декодер*), а *модем* – пристрій, що перетворює код у сигнал (*модулятор*) і сигнал у код (*демодулятор*). Канальні пристрої (смугові підсилювачі передавача і приймача, коректори і т. ін.) разом з лінією зв'язку утворюють неперервний канал, а останній разом з модемом – дискретний канал.

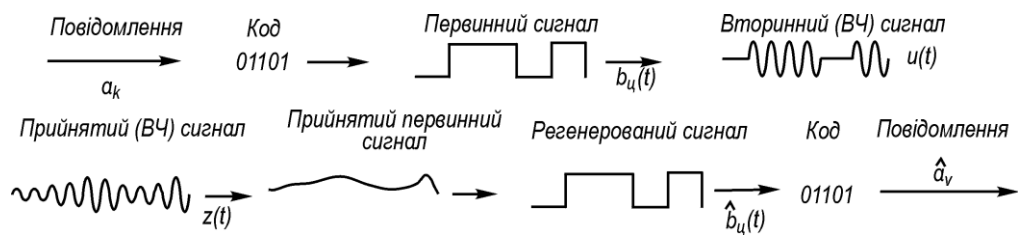


Рис. 1.5. Процес перетворення дискретного повідомлення в сигнал

Слід мати на увазі, що в системах радіозв'язку після передавача за допомогою передавальних антен утворюється просторово-часовий сигнал $u(t, r)$ (електромагнітна хвиля), що залежить не тільки від часу t , але й від просторових координат точки спостереження $r(x, y, z)$. Крім декартових, можна ввести й інші просторові координати, наприклад полярні.

Сигнал, який залежить від багатьох координат, називають полем. У місце прийому (на вихід радіоканалу) надходить поле (просторово-часовий сигнал) $z(t, r) = s(t, r) + n(t, r)$. Найчастіше воно спочатку за допомогою приймальної антени перетворюється в чисто часовий сигнал $z(t)$, який надалі підлягає чисто часовій обробці. Питання формування й обробки просторово-часових сигналів розглядаються в спеціальних курсах, а в цьому підручнику вважатимемо, що пристрої перетворення часовий сигнал – поле на передаванні і поле – часовий сигнал на прийманні знаходяться всередині заданої лінії зв'язку.

Для передавання неперервного повідомлення (рис. 1.6) його спочатку перетворюють у неперервний первинний електричний сигнал $b(t)$, а потім, як правило, за допомогою модулятора формують каналний сигнал $u(t)$, який і посиляють у лінію зв'язку. Прийняте коливання $z(t)$ піддається зворотним перетворенням, внаслідок яких виділяється первинний сигнал \hat{b} , за яким і відновлюється з тією чи іншою точністю повідомлення \hat{a} .

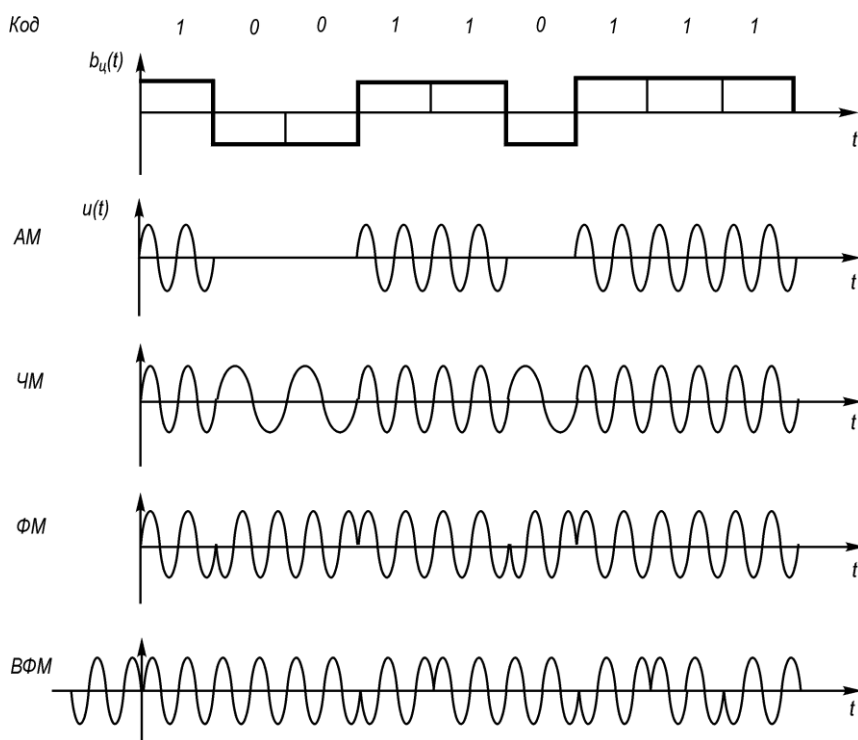


Рис. 1.6. Форми сигналів при двійковому коді для різних видів дискретної модуляції

Загальний принцип модуляції полягає в зміні одного чи кількох параметрів несучого коливання (переносника) $f(t, \alpha, \beta, \dots)$ відповідно до переданого повідомлення. Так, якщо

переносником є гармонічне коливання $f(t) = U \cos(\omega_0 t + \varphi)$, то можна створити три види модуляції: *амплітудну, частотну і фазову*.

Якщо переносником є періодична послідовність імпульсів $f(t) = U_0 \sum_{l=-\infty}^{\infty} v(t - lT - t_0)$, то при заданій формі імпульсів $v(t)$ можна створити чотири основних види імпульсної модуляції: *амплітудно-імпульсну, широтно-імпульсну, часово-імпульсну і частотно-імпульсну*. Застосування радіоімпульсів дозволяє одержати ще два види модуляції: *за частотою і за фазою високочастотного заповнення*.

При дискретній (цифровій) модуляції закодоване повідомлення у вигляді послідовності кодових символів b_l перетворюється в послідовність елементів (посилок) сигналу $u_l(t)$ шляхом впливу кодових символів на переносник $f(t)$. За допомогою модуляції один з параметрів переносника змінюється за законом, який визначається кодом. У разі безпосереднього передавання переносником може бути *постійний струм*, сила якого і напрямок змінюються. Зазвичай за переносник, як і в неперервній модуляції, використовується *змінний струм* (гармонічне коливання). У цьому випадку можна одержати амплітудну, частотну і фазову модуляції.

На рис. 1.6 наведені форми сигналу при двійковому коді для різних видів дискретної (або цифрової) модуляції (або маніпуляції). Тут проілюстрований вид модуляції, коли закон передаваного сигналу цифровий (первинний), а сигнал-переносник (несуче коливання) аналоговий (неперервний). Таке перетворення і називається *дискретною модуляцією* (або *маніпуляцією*). При амплітудній модуляції (АМ) символу 1 відповідає передавання несучого коливання протягом часу T (посилка), символу 0 – відсутність коливання (пауза). При частотній модуляції (ЧМ) передавання несучого коливання з частотою f_1 відповідає символу 1, а передавання коливання з частотою f_0 – символу 0. При двійковій фазовій модуляції змінюється фаза несучого коливання на π у разі кожного переходу від 1 до 0 і від 0 до 1.

Нарешті, на практиці застосовують систему відносної фазової модуляції (ВФМ). На відміну від фазової, при відносній фазовій модуляції фазу сигналів відраховують не від деякого еталона, а від фази попереднього елемента сигналу. Наприклад, символ 0 передається відрізком синусоїди з початковою фазою попереднього елемента сигналу, а символ 1 – таким самим відрізком з початковою фазою, що відрізняється від початкової фази попереднього елемента сигналу на π . Передача починається з посилки, яка служить опорним сигналом для порівняння фази наступного елемента.

У більш загальному випадку дискретну модуляцію доцільно розглядати як перетворення кодових символів 0, 1, ..., $m-1$ у відрізки сигналу $u_i(t)$, де $i=0, 1, \dots, m-1$ – переданий символ. При цьому вид сигналу $u_i(t)$, у принципі, може бути довільний. На практиці його вибирають так, щоб задовольнити вимоги, висунуті до системи зв'язку (зокрема, за швидкістю передавання і за займаною смугою частот), і щоб сигнали добре розрізнялися з урахуванням впливу завад.

Тривалість посилки первинного сигналу $b_{\Pi}(t)$ при дискретному передаванні визначає швидкість передавання посилок (*технічну швидкість* чи *швидкість модуляції*). Ця швидкість v виражається числом посилок, переданих за одиницю часу, і вимірюється в бодах. 1 бод – це швидкість, з якою протягом 1с передається одна посилка. Якщо тривалість посилки T виражена в секундах, то швидкість модуляції в бодах $v=1/T$. Якщо смугу частот обмежити третьою гармонікою, то ширина спектра первинного сигналу $F=1,5v$ Гц.

Демодуляція і декодування. Передане повідомлення в приймачі звичайно відновлюється в такій послідовності. Спочатку сигнал демодулюється. У системах передавання неперервних повідомлень внаслідок демодуляції відновлюється первинний сигнал, який відображає передане повідомлення. Цей сигнал потім надходить на відтворюючий чи записуючий пристрій. У радіомовленні таким пристроєм може бути гучномовець або магнітофон. У системах передавання дискретних повідомлень, як правило, в результаті демодуляції послідовність елементів сигналу перетворюється в послідовність кодових символів. Потім за ними відновлюються повідомлення, що видаються одержувачу. Останнє перетворення називається *декодуванням*.

Не слід думати, що демодуляція і декодування – це прості операції, які є зворотними до модуляції і кодування. Через різні спотворення і вплив завад отриманий сигнал може істотно відрізнятись від переданого. Тому завжди можна висловити ряд припущень (гіпотез) щодо того, яке повідомлення передавалося. Задачею приймального пристрою є прийняття рішення про те, яке з можливих повідомлень дійсно передавалося джерелом. Для цього прийнятий сигнал підлягає аналізу з урахуванням усіх відомостей про джерело (наприклад, про ймовірності, з якими джерело посилає те чи інше повідомлення), про застосований код і метод модуляції, а також про властивості каналу. За результатами аналізу, як правило, можна визначити умовні (апостеріорні) ймовірності можливих гіпотез і на цій підставі прийняти рішення, яке й надходить до одержувача. Та частина приймального пристрою, що здійснює аналіз отриманого сигналу і приймає рішення про передане повідомлення, називається *розв'язувальною схемою*.

У системах передавання неперервних повідомлень при аналоговій модуляції розв'язувальна схема визначає за спотвореним каналним отриманим (вторинним) сигналом найбільш імовірний переданий первинний сигнал і відновлює його. Тут розв'язувальною схемою є демодулятор. У системах передавання дискретних повідомлень розв'язувальна схема найчастіше складається з двох частин: демодулятора і декодера.

Іноді в процесі передавання дискретних повідомлень операції демодуляції і декодування виконує один пристрій, який перетворює послідовність елементів отриманого сигналу відразу в послідовність символів (літер) повідомлення. Такий метод приймання називають *спільною демодуляцією–декодуванням* чи *прийманням в цілому* на відміну від поелементного приймання з двома частинами розв'язувальної схеми. У першому випадку аналізується цілком відрізок сигналу, що відповідає кодовій комбінації, і на підставі того чи іншого критерію відновлюється переданий елемент повідомлення (літера). В другому випадку спочатку аналізуються окремі елементи сигналу, що відповідають кодовим символам, а потім відновлена кодова комбінація декодується, тобто перетворюється в елемент (літеру) повідомлення.

У деяких випадках роль розв'язувальної схеми виконує повністю чи частково людина. Так, приймаючи телеграфні сигнали на слух, оператор вирішує, який сигнал (точка чи тире) був переданий. Він же здійснює й операцію декодування. У приймачах дискретних повідомлень, призначених для запису інформації, всі зазначені операції виконуються автоматично. В найпростішому випадку перша частина розв'язувальної схеми являє собою граничний пристрій – реле або тригер, що працюють за принципом “Так” чи “Ні”. Якщо прийнятий елемент сигналу вищий за поріг (наприклад, за допустиму напругу), видається один символ коду (наприклад, 1), якщо нижчий – інший (0). У деяких випадках застосовують розв'язувальні схеми з двома порогами. Якщо рівень сигналу знаходиться між двома порогами, рішення не приймається – замість сумнівного елемента сигналу видається спеціальний *символ стирання*, котрий полегшує можливість правильного декодування прийнятої кодової комбінації.

Для прийняття рішення про те, яке повідомлення передавалося, необхідно проаналізувати отриманий сигнал, піддавши його різним перетворенням, які називають *обробкою сигналу*. Однією з задач теорії зв'язку є пошук правил оптимальної обробки сигналу, застосування яких дає найбільш вірогідне рішення про передане повідомлення. Ці правила залежать від властивостей каналу і методів передавання (кодування і модуляції). Іноді оптимальні правила обробки виявляються складними, і для спрощення апаратури використовують іншу, не оптимальну обробку.

Якість прийому й обробки сигналів істотно залежить від точності синхронізації переданих і прийнятих сигналів. Розрізняють синхронізацію *тактову* (визначення границь одиничних елементів сигналу), *циклову* (правильний поділ кодових комбінацій), *несучих частот* тощо. Похибки синхронізації призводять до зниження вірогідності прийому повідомлень, а в ряді випадків – до неправильного прийому всього повідомлення чи його частини.

Найпростішим методом, що дозволяє на прийомі відокремити одну кодову комбінацію від іншої, є *стартстопний режим* передавання, коли на початку і в кінці кожної комбінації передається спеціальний сигнал (“старт” і “стоп”). Такий метод передавання належить до асинхронних, тому що передавання будь-якої кодової комбінації може починатися в будь-який

момент часу після закінчення попередньої комбінації. При синхронних методах передавання елементи сигналу передаються неперервно через однакові проміжки часу. Поділ кодових комбінацій здійснюється в цьому випадку за допомогою циклової синхронізації.

Цифрове кодування неперервних повідомлень. На даний час широко застосовуються цифрові системи передавання, в яких неперервні повідомлення передаються дискретними сигналами. Перетворення неперервного повідомлення в цифрову форму здійснюється за допомогою операцій дискретизації та квантування.

Дискретизація за часом виконується шляхом взяття значень (відліків) первинного сигналу $b(t)$ у визначені дискретні моменти t , внаслідок чого неперервну функцію $k(t)$ замінюють сукупністю миттєвих значень $b(k)$ чи $b(t_k)$.

Дискретизація за часом дозволяє перетворити неперервні повідомлення в дискретний (у часі) сигнал, який після квантування перетворюється в цифровий.

Операція квантування зводиться до того, що замість даного миттєвого значення (рівня) переданого повідомлення $b(t_k)$ передають найближчі значення за встановленою цифровою шкалою дискретних рівнів $b_{\text{кв}}(t_k)$. Дискретні значення за шкалою рівнів найчастіше вибираються рівномірно: $b_{\text{кв}}^{(l)} = l\Delta b$, де Δb – крок квантування, $l = 0, 1, \dots, L-1$. Зрозуміло, що квантування вносить похибку, оскільки дійсне значення $b(t_k)$ замінюють округленим значенням $b_{\text{кв}}(t_k)$. Ця похибка $\varepsilon = b(t_k) - b_{\text{кв}}(t_k)$ не перевищує половини кроку квантування Δb і може бути зведена до допустимого рівня. Похибка ε є випадковою функцією і проявляється на виході як додатковий шум (шум квантування), накладений на передане повідомлення.

Перевагою цифрових способів передавання є можливість застосування кодів як для підвищення завадостійкості, так і для скорочення надлишковості джерела. На цей час найбільше застосування знаходить система з *імпульсно-кодовою модуляцією* (ІКМ). У цій системі неперервне повідомлення спочатку піддають дискретизації за часом і квантуванню за рівнем, а потім отриману послідовність L рівнів (цифр) кодуєть (як правило, двійковим кодом). При цьому кожному рівню присвоюється кодова комбінація, що складається із n символів 1 і 0. Отримана послідовність двійкових символів передається по каналу зв'язку одним із методів дискретної модуляції. Звичайно використовується частотна або фазова модуляція.

Завади та спотворення в каналі. У реальному каналі сигнали у процесі передавання спотворюються, а тому повідомлення відтворюються з деякими помилками. Причиною таких помилок є як спотворення, внесені самим каналом, так і завади, що впливають на сигнал. Частотні й часові характеристики каналу визначають *лінійні спотворення*. Крім того, канал може вносити і *нелінійні спотворення*, зумовлені нелінійністю тих чи інших ланок каналу. Якщо лінійні та нелінійні спотворення зумовлені відомими характеристиками каналу, то вони, принаймні в принципі, можуть бути усунуті належною корекцією. Слід відрізнити спотворення від завад. Завади мають випадковий характер і заздалегідь не відомі, а тому не можуть бути цілком усунуті.

Завадою називається будь-який випадковий вплив на сигнал, що погіршує вірогідність відтворення переданих повідомлень.

Завади дуже різноманітні як за своїм походженням, так і за фізичними властивостями. У радіоканалах часто зустрічаються *атмосферні завади*, зумовлені електричними процесами в атмосфері, насамперед грозовими розрядами. Енергія цих завад зосереджена головним чином в області довгих і середніх хвиль. Сильні завади створюються також промисловими установками. Це так звані *індустріальні завади*, що виникають через різкі зміни струму в електричних колах електроприладів. До індустріальних належать завади від електротранспорту, електричних двигунів, медичних установок, систем запалювання двигунів тощо. Розповсюдженим видом завад є *завади від сторонніх радіостанцій і каналів*. Вони зумовлені порушенням регламенту розподілу робочих частот, недостатньою стабільністю частот і поганою фільтрацією гармонік сигналу, а також нелінійними процесами в каналах, що призводять до перехресних спотворень.

У провідних каналах зв'язку основним видом завад є *імпульсні шуми та переривання зв'язку*. Поява імпульсних завад часто пов'язана з автоматичною комутацією і перехресними наведеннями. Переривання зв'язку – це явище, при якому сигнал у лінії різко зникає чи загасає.

Практично в будь-якому діапазоні частот мають місце *внутрішні шуми апаратури*, зумовлені хаотичним рухом носіїв заряду в підсилювальних приладах, резисторах й інших елементах апаратури. Ці завади особливо позначаються на радіозв'язку в діапазоні ультракоротких хвиль, де інші завади невеликі. В цьому діапазоні мають значення і *космічні завади*, спричинені електромагнітними процесами, що відбуваються на Сонці, зорях та інших неземних об'єктах.

У загальному вигляді вплив завади $n(t)$ на корисний сигнал $s(t)$ можна виразити через оператор

$$z(t) = L s(t), n(t) . \quad (1.9)$$

В окремому випадку, коли оператор є сумою:

$$z(t) = s(t) + n(t), \quad (1.10)$$

завада називається *адитивною*. Якщо ж оператор може бути зображений у вигляді добутку:

$$z(t) = k(t)s(t), \quad (1.11)$$

то заваду називають *мультиплікативною*. Тут $k(t)$ – випадковий процес.

У реальних каналах, як правило, мають місце і адитивні, і мультиплікативні завади, тому

$$z(t) = k(t)s(t) + n(t). \quad (1.12)$$

Серед адитивних завад різного походження виділяють зосереджені по спектру (*вузькосмугові*) завади, зосереджені в часі (*імпульсні*) завади і *флуктуаційні* завади, не обмежені в часі та спектрі. Флуктуаційна завада (флуктуаційний шум) являє собою випадковий процес з нормальним розподілом (гауссів процес). Вона найбільш вивчена і найцікавіша як з теоретичної, так і з практичної точки зору. Цей вид завад має місце практично в усіх реальних каналах. У діапазоні оптичних частот істотне значення має *квантовий шум*, спричинений дискретною природою сигналу.

Мультиплікативні завади зумовлені випадковими змінами параметрів каналу зв'язку. Зокрема, ці завади проявляються в *зміні рівня сигналу*.

Зазначимо, що між сигналом і завадою немає принципової різниці. Більш того, вони існують в єдності, хоча й протилежні за своєю дією. Так, випромінювання радіопередавача є корисним сигналом для приймача, якому призначене це випромінювання, і завадою для всіх інших приймачів. Електромагнітне випромінювання зірок є однією з причин космічного шуму в діапазоні надвисоких частот, і тому є завадою для систем радіозв'язку. З іншого боку, це випромінювання є корисним сигналом, за яким визначають деякі фізико-хімічні властивості зорі.

Основні характеристики системи зв'язку. Оцінюючи роботу системи зв'язку, необхідно насамперед врахувати, яку точність передавання повідомлення забезпечує система і з якою швидкістю передається інформація. Перше визначає якість передавання, друге – кількість.

У реальній системі зв'язку *якість передавання* залежить від ступеня спотворень прийнятого повідомлення. Ці спотворення залежать від властивостей і технічного стану системи, а також від інтенсивності та характеру завад. У правильно спроектованій і технічно справній системі зв'язку незворотні спотворення повідомлень зумовлені лише впливом завад, тому якість передавання цілком визначається завадостійкістю системи. Під *завадостійкістю* звичайно розуміють *здатність системи протистояти шкідливому впливу завад на передавання повідомлень*. Оскільки вплив завад проявляється в тому, що прийняте повідомлення відрізняється від переданого, то кількісно завадостійкість можна охарактеризувати ступенем відповідності прийнятого повідомлення переданому. Назвемо цю величину *вірогідністю*. Кількісну міру вірогідності

доводиться вибирати по-різному, залежно від характеру повідомлення і вимог одержувача. Нехай повідомлення є дискретною послідовністю елементів з деякої скінченної множини. Вплив завади на передавання такого повідомлення проявляється в тому, що замість фактично переданого елемента може бути прийнятий інший; така подія називається *помилкою*. За кількісну міру вірогідності можна взяти ймовірність помилки p або будь-яку монотонну функцію цієї ймовірності.

У разі передавання неперервних повідомлень ступенем відповідності прийнятого повідомлення $\hat{B}(t)$ [з реалізацією $\hat{b}(t)$] переданому $B(t)$ [з реалізацією $b(t)$] може бути величина, що характеризує “відстань” між $b(t)$ і $\hat{b}(t)$. За таку величину часто беруть критерій *квадратичного відхилення*:

$$\overline{E^2(t)} = \overline{[\hat{B}(t) - B(t)]^2}, \quad (1.13)$$

де риска зверху означає усереднення за ансамблем випадкових величин.

Кількісну міру надійності можна також визначити як імовірність того, що відхилення E не перевищить деякої наперед заданої величини ε_0 :

$$Q = P[|E| \leq \varepsilon_0]. \quad (1.14)$$

Як буде показано далі, вірогідність передавання залежить від *відношення середніх потужностей сигналу і завади*. При даній інтенсивності завади ймовірність помилки тим менша, чим сильніше різняться між собою сигнали, які відповідають різним повідомленням. Завдання полягає в тому, щоб вибрати для передавання сигнали з великою розбіжністю.

Нарешті, вірогідність передавання залежить і від *способу приймання*. Треба вибрати такий спосіб, який щонайкраще реалізує розбіжність між сигналами при даному відношенні середніх потужностей сигналу і завади.

Звернемо увагу на істотну різницю між аналоговими і дискретними системами передавання повідомлень. В аналогових системах усякий, навіть як завгодно малий вплив на сигнал, що спотворює модульований параметр, завжди призводить до внесення відповідної похибки в повідомлення. Тому *абсолютно точне відновлення переданого повідомлення неможливе*. У дискретних системах помилка при передаванні повідомлень виникає тільки тоді, коли сигнал орієнтується неправильно, а це відбувається лише при спотвореннях, що *перевищують деякий поріг*.

У теорії завадостійкості, розробленій В. О. Котельниковим, показано, що при обраному критерії та заданій множині сигналів, прийнятих за наявності адитивного білого гауссового шуму (БГШ), існує *гранична (потенційна) завадостійкість*, яка при жодному способі приймання не може бути перевищена. Приймальний пристрій, що реалізує потенційну завадостійкість, називається оптимальним (найкращим) за даним критерієм.

Поряд з вірогідністю найважливішим показником роботи системи зв'язку є *швидкість передавання інформації*. У системах передавання дискретних повідомлень швидкість v (в бодах) вимірюється числом переданих символів за одиницю часу. *Кількість переданої інформації* прийнято вимірювати в бітах (двійкових одиницях). Як буде показано максимальна кількість інформації, яку можна передати двійковим символом, дорівнює 1 біту. Там же буде показано, що в разі використання не двійкових, а m -их символів максимальна кількість інформації, яку можна передати, дорівнює $\log_2 m$ біт. Тому дискретне джерело може забезпечити *максимальну швидкість передавання інформації (максимальну продуктивність)*, біт/с,

$$R_1 = \frac{\log_2 m}{T}, \quad (1.15)$$

де T – тривалість посліжки; m – основа коду. При $m=2$ та $R_i = 1/T$ біт/с швидкість передавання інформації R_i чисельно дорівнює технічній швидкості v . При $m>2$ швидкість передавання інформації R_i може бути більшою за v . Однак нерідко в дискретних системах зв'язку $R_i < v$, коли для передавання інформації використовуються не всі посліжки, наприклад, частина з них служить для синхронізації або для виявлення і виправлення помилок (при використанні коректувального коду).

До основних характеристик системи зв'язку належать також *середня швидкість передавання інформації* по заданому каналу в одиницю часу з заданою надійністю, та *максимально можлива (гранична) швидкість передавання*, що називається **пропускною здатністю каналу C'** . Це фундаментальне поняття визначає потенційні можливості системи зв'язку, яку використовує даний канал. У реальній системі середня швидкість передавання інформації завжди менша за пропускну здатність каналу C' .

У теорії інформації К. Шенноном доведена теорема, відповідно до якої для джерела без надлишковості при $R_i < C'$ можна знайти такий спосіб кодування–декодування, який забезпечує передавання повідомлень по каналу з завадами із як завгодно малою похибкою.

Універсальним показником **ефективності** системи зв'язку є коефіцієнт η , що характеризує використання системою пропускної здатності каналу: $\eta = R_i / C'$ (*інформаційна ефективність*).

Своєчасність передавання повідомлень визначається *допустимою затримкою*, зумовленою перетворенням повідомлень і сигналів, а також кінцевим часом поширення сигналу по каналу зв'язку. Вона залежить, по-перше, від характеру і довжини каналу, по-друге, від тривалості обробки сигналу в передавальному і приймальному пристроях. Швидкість передавання і затримка є незалежними характеристиками, практично не пов'язаними одна з одною.

Існують і багато інших параметрів, що характеризують з різних точок зору якість системи зв'язку: скритність зв'язку, надійність системи, габаритні розміри і маса апаратури, вартість устаткування, експлуатаційні витрати тощо. Ці характеристики в курсі “Теорія електричного зв'язку” не розглядаються. Їм присвячені окремі розділи інших спеціальних курсів.

Контрольні запитання для самооцінки рівня знань

1. Дайте означення інформації та інформаційних процесів.
2. Виділіть основні класи об'єктів інформаційної техніки за функціональною ознакою.
3. Зобразіть узагальнену структурну схему системи зв'язку, що відображає процес передавання інформації від джерела повідомлень до приймача повідомлень.
4. Дайте означення сигналу та назвіть основні параметри сигналу, що визначають його властивості як переносника інформації в системах зв'язку.
5. Як поділяються системи зв'язку за характером сигналів, що передаються.
6. У чому полягає суть процесу кодування?
7. Що розуміють під розв'язувальною схемою при прийманні дискретних повідомлень?

Післямова до розділу 1

Ви завершили вивчення загальних відомостей теорії інформації та передавання сигналів.

Тепер ви вже знаєте, що:

Детермінований сигнал – сигнал, майбутні значення якого можна визначити для будь-якого моменту часу.

Детерміновані сигнали поділяються на періодичні і неперіодичні.

Періодичний детермінований сигнал – сигнал, однойменні значення якого повторюються через фіксований проміжок часу.

Неперіодичний детермінований сигнал – сигнал, однойменні значення якого не мають повторюваності в часі. Прикладом детермінованого неперіодичного сигналу є цифровий сигнал (рис. 2.2), оскільки наперед відомо, що він є послідовністю прямокутних імпульсів, які не мають період повторення.

Випадковий сигнал – сигнал, майбутні значення якого не можна точно визначити (побачити).

Поняття які необхідно знати для сприйняття матеріалу **розділу 2:**

1) часова форма – представляє зміну напруги сигналу в часі;

2) часова форма – представляє значення напруги сигналу на різних частотах.

Часове представлення сигналу спостерігають отримують (отримують) за допомогою осцилографа і називають формою сигналу (рис. 1.1).

Аргументом функції у часовому представленні є час t .

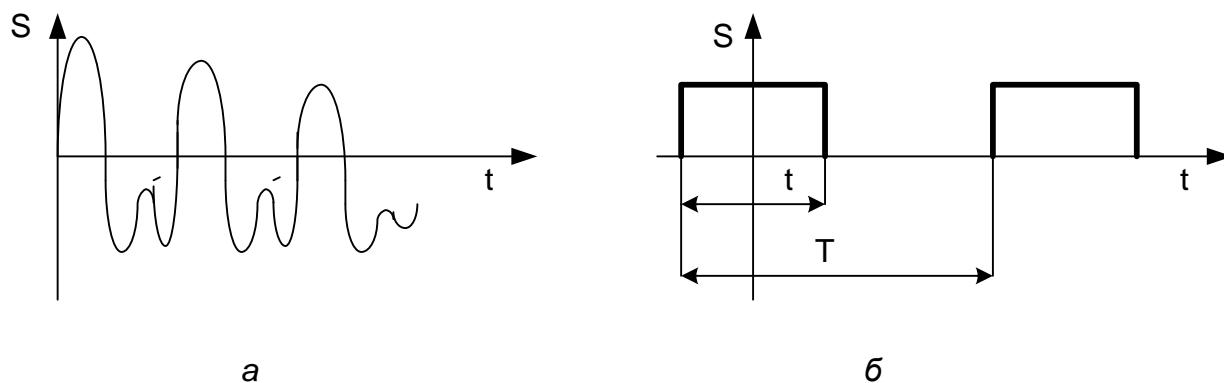


Рис.1.1 Сигнали: *a* - неперервний випадковий сигнал;
б – детермінований періодичний сигнал

Частотне представлення сигналу отримують за допомогою селективного вольтметра і називають **спектральною характеристикою сигналу**. У частотному представленні сигналу аргументом є частота $\omega = 2\pi f$.

Кроки побудови математичної моделі сигналу:

1. Вибираємо інтервал часу спостереження сигналу.
2. Знаходимо математичну функцію, яка описує форму цього сигналу. Якщо це не вдається, то знаходимо декілька функцій, за допомогою яких можна описати цей сигнал на менших інтервалах часу. Коли ставиться задача побудови математичної моделі сигналу, необхідно провести детальне вивчення сигналу і виявити його властивості.

Терміни, які необхідно знати:

Інваріантний – незмінний за певних перетворень.

Інваріанти – (незмінний)- величини, співвідношення, які не змінні, узагальнене поняття, що виявляється в конкретних змінних величинах, поняття нижнього роду – варіантах. Умовою існування є множинність варіантів, які не змінюються від тих чи інших перетворень змінних величин, пов'язаних з ними, наприклад коли змінюють систему координат.

Когерентний - (взаємозв'язаний) хвилі характерні, здатні до інтерференції, тобто такі, що в них не змінюється в часі різниці фаз.

Інтерференція - взаємне посилення або ослаблення хвиль (електромагнітних) у разі накладення їх одна на одну.

Апріорі – (з попереднього) – незалежно від досвіду. Переносно - без перевірки, наперед.

Апріорний – той що передує йому, той що не ґрунтується на досвіді.

Кореляція – співвідношення, відповідність, взаємозв'язок предметів або понять. В математичній статистиці - залежність між явищами або величинами, що не має чіткого функціонального характеру.

Оптимальний – найкращий з можливих варіантів чогось, найбільш відповідних даному завданню, умовам.

Квазі ... (ніби, майже, немовби) у складних словах означає “ніби”, несправжній, фальшивий.

Апостеріорний – набутий з досвіду, той що ґрунтується на досвіді.

Адитивна – (додаткова) - отриманий шляхом додавання. Теорія чисел - частина арифметики, що вивчає закони, за якими числа можна скласти з доданків того чи іншого виду.

Адаптивна – (адаптація) - пристосовуваність; здатність організмів до адаптації.

Розділ 2

МАТЕМАТИЧНІ МОДЕЛІ ДЕТЕРМІНОВАНИХ СИГНАЛІВ

З точки зору розв'язування задач електричного зв'язку математичні моделі, що відображають основні властивості повідомлень і завад, є фундаментом теорії електрозв'язку. Оскільки сигнали і завади перш за все є електричними комбінаціями, що змінюються в часі, то їх базовою математичною моделлю є деяка функція часу $S(t)$. За своїми фізичними та математичними властивостями процеси $x(t)$ діляться на детерміновані та випадкові. В даному розділі розглядаються детерміновані сигнали.

2.1. Види ДЕТЕРМІНОВАНИХ СИГНАЛІВ

Детермінованими (регулярними) називають такі процеси $S(t)$, перебіг яких в часі можна повністю визначити наперед [$S(t) = at$, $S(t) = at^2$, $S(t) = A \cos(\omega t + \varphi)$]. Інакше кажучи, для будь-якого наперед заданого моменту часу t можна однозначно визначити значення функції $S(t)$. Розрізняють такі види детермінованих сигналів: прості; складні; еталонні (або пробні); періодичні; неперіодичні.

Простим називають сигнал, що відповідає одній елементарній послідовності. Наприклад, при передаванні дискретної інформації простому сигналу відповідає сигнал кодового символу в комбінації (рис. 2.1, а).

Сигнал, що є сукупністю елементарних посилок, називають складним або складеним (рис. 2.1, б).

В теорії зв'язку вводиться поняття *базис сигналу* B :

$$B = 2TF,$$

де T – тривалість сигналу; F – смуга частот, що займається сигналом.

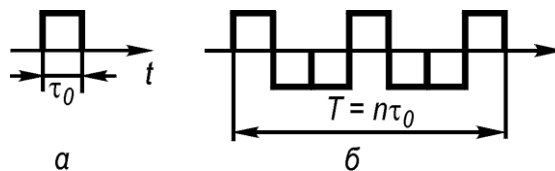


Рис. 2.1. Простий (а) та складний (б) сигнали

Для простих сигналів $B \approx 1$, а для складних $B \gg 1$. Враховуючи цю обставину, прості сигнали називають *вузькосмуговими*, а складні – *широкосмуговими*.

Еталонні (або пробні) сигнали використовуються в спеціальних вимірювальних системах, для синхронізації тощо. Такі сигнали можуть реалізовуватись за допомогою наступних функцій:

1. Гармонічний сигнал (рис. 2.2, а):

$$S(t) = A \sin(\omega t + \varphi), \quad -\infty < t < \infty,$$

де A , ω , φ – амплітуда, частота і фаза сигналу відповідно.

2. Одиначна функція [функція вмикання (рис. 2.2, б)]:

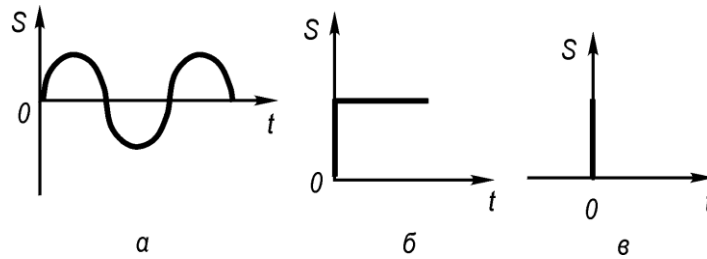


Рис. 2.2. Еталонні сигнали:
a – гармонічний; *б* – одинична функція; *в* – дельта-функція

$$f(t) = \begin{cases} 0 & \text{при } t < 0; \\ f_1 & \text{при } t > 0. \end{cases}$$

3. Одиничний імпульс [дельта-функція (рис. 2.2, в)]:

$$\delta(t) = \begin{cases} 0 & \text{при } t < 0; \\ \infty & \text{при } t = 0; \\ 0 & \text{при } t > 0. \end{cases}$$

Періодичним називається сигнал, який повторюється через певні проміжки часу T :

$$S(t) = S(t + nT), \quad -\infty < t < \infty,$$

де n – будь-яке ціле число.

Будь-яка періодична функція, що задовольняє умови Діріхле, може бути зображена у вигляді ряду Фур'є. Умови визначаються наступним чином. Функція $S(t)$ має бути скрізь однозначною, скінченною, кусково-неперервною [тобто інтервал, на якому функція визначена, може бути розбитий на скінченне число інтервалів, в кожному з яких $S(t)$ неперервна і монотонна], повинна мати скінченне число максимумів і мінімумів, у будь-якій точці розриву існує $S(t-0)$ і $S(t+0)$; значення $S(t)$ в точці розриву дорівнює $S(t) = \frac{S(t-0) + S(t+0)}{2}$. При цьому ряд Фур'є записується так:

$$S(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k \cos(k\omega_1 t - \varphi_k)$$

або

$$S(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos k\omega_1 t + b_k \sin k\omega_1 t,$$

де $a_0 = \frac{2}{T} \int_{-T/2}^{T/2} S(t) dt$; $a_k = A_k \cos \varphi_k$; $b_k = A_k \sin \varphi_k$; $A_k = \sqrt{a_k^2 + b_k^2}$; $\varphi_k = \arctg \frac{b_k}{a_k}$.

Комплексна форма запису ряду Фур'є:

$$S(t) = \frac{1}{2} \sum_{k=-\infty}^{\infty} \dot{A}_k e^{jk\omega_1 t},$$

де $\dot{A}_k = A_k e^{-j\varphi_k} = a_k + jb_k$, $A_k = |\dot{A}_k|$. Використовуючи обернене перетворення Фур'є, знаходимо коефіцієнти ряду Фур'є:

$$\dot{A}_k = \frac{2}{T} \int_{-T/2}^{T/2} S(t) e^{-jk\omega_1 t} dt.$$

Якщо $S(t) = S(-t)$ (функція парна), то $b_k = 0$, $A_k = a_k$; якщо $S(t) = -S(-t)$ (функція непарна), то $a_k = 0$, $A_k = b_k$.

Спектр періодичного сигналу зображається графічно у вигляді спектральних ліній, довжини яких пропорційні амплітудам (фазам) відповідних частотних складових. Спектр періодичного сигналу – дискретний (лінійчатий). При цьому відстань між сусідніми лініями постійна і дорівнює частоті першої гармоніки (рис. 2.3).

Неперервна крива, що з'єднує кінці ліній спектра, називається *обвідною спектра*:

$$S(j\omega) = \frac{2}{T} \int_{-T/2}^{T/2} S(t) e^{-j\omega t} dt = S(t) e^{i\varphi(\omega)},$$

де $S(\omega), \varphi(\omega)$ – обвідні спектрів амплітуд і фаз відповідно.

Середня потужність сигналу, що виділяється на навантаженні з опором 1 Ом,

$$P_{cp} = \overline{S^2(t)} = \frac{1}{T} \int_{-T/2}^{T/2} S^2(t) dt = \frac{a_0^2}{2} + \frac{1}{2} \sum_{k=1}^{\infty} A_k^2.$$

Наведена формула називається *рівністю Парсеваля*, згідно з якою середня потужність періодичного сигналу дорівнює сумі середніх потужностей всіх частотних складових його спектра.

Реальні канали зв'язку, через які проходить сигнал, мають обмежену смугу пропускання, тому сигнал $S(t)$ можна подати у вигляді

$$S(t) \approx \sum_{k=l_1}^{l_2} (a_k \cos k\omega_1 t + b_k \sin k\omega_1 t). \quad (2.1)$$

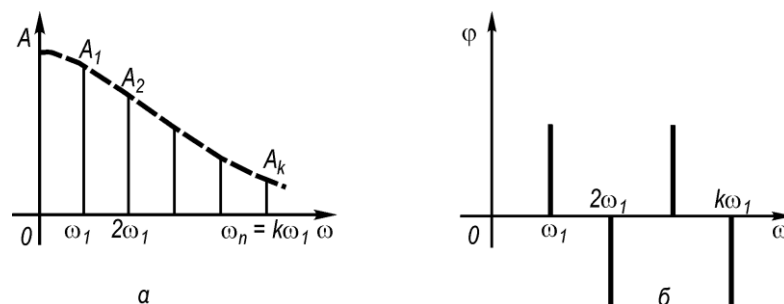


Рис. 2.3. Спектр амплітуд (а) і фаз (б) періодичного сигналу

Число членів ряду (2.1) l – ціле число – залежить від того, які спектральні складові входять до смуги $F = (l_2 - l_1)/T$ частот, що займається каналом. Апроксимація сигналу $S(t)$ рядом (2.1) дає найменше середньоквадратичне відхилення від точного значення $S(t)$, якщо a_k і b_k є коефіцієнтами ряду Фур'є.

Неперіодичними називають сигнали, які задовольняють умову $S(t) \neq S(t + nT)$ на інтервалі $-\infty < t < \infty$. Такий сигнал зображується функцією на скінченному чи напівнескінченному інтервалах $t_1 < t < t_2, t_1 < t < \infty$.

Для неперіодичного сигналу існує пряме –

$$S(j\omega) = \int_{-\infty}^{\infty} S(t)e^{-j\omega t} dt$$

та обернене перетворення Фур'є –

$$S(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega)e^{j\omega t} d\omega.$$

Величина $S(j\omega)$ виражає *спектральну щільність сигналу* і називається *спектральною характеристикою* або комплексним спектром, а модуль $|S(j\omega)|$ – *спектром сигналу*:

$$S(j\omega) = \pi \frac{dA}{d\omega}; \quad |S(j\omega)| = S(\omega) \text{ – спектр сигналу.}$$

Спектральну характеристику як комплексну величину можна подати у вигляді

$$S(j\omega) = A(\omega) + jB(\omega) = S(\omega)e^{-j\varphi(\omega)},$$

де $A(\omega) = \int_{-\infty}^{\infty} S(t) \cos \omega t dt; \quad B(\omega) = \int_{-\infty}^{\infty} S(t) \sin \omega t dt; \quad S(\omega) = \sqrt{A^2(\omega) + B^2(\omega)}; \quad \varphi(\omega) = \arctg B(\omega)/A(\omega).$

Енергія неперіодичного сигналу

$$E = \int_{-\infty}^{\infty} S^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega)S(-j\omega) d\omega = \frac{1}{\pi} \int_{-\infty}^{\infty} [S(\omega)]^2 d\omega.$$

Якщо спектр обмежений смугою частот $F = f_2 - f_1$, можна наближено записати

$$\frac{1}{\pi} \int_{f_1}^{f_2} [S(\omega)]^2 d\omega \approx \frac{1}{\pi} \int_{-\infty}^{\infty} [S(\omega)]^2 d\omega. \quad (2.2)$$

Ширину спектра сигналу слід вибрати так, щоб забезпечити виконання цієї рівності з заданою точністю. Це, однак, не є єдиним критерієм вибору ширини спектра. Другим критерієм є допустимий ступінь спотворення форми сигналу.

Практично під шириною спектра розуміємо область частот, у межах якої зосереджено понад 90 % енергії сигналу.

Зазначимо, що існує однозначна залежність між тривалістю імпульсу та шириною спектра сигналу. Чим ширший спектр, тим менша тривалість імпульсу, і навпаки. Добуток ширини спектра імпульсного сигналу і тривалості імпульсу є величина стала: $F\tau = \text{const}$; $F\tau = 1$ – для прямокутного імпульсу, $F\tau = 1,5$ – для косинусоїдального, $F\tau = 2$ – для трикутного.

На рис. 2.4 наведено спектри деяких неперіодичних сигналів.

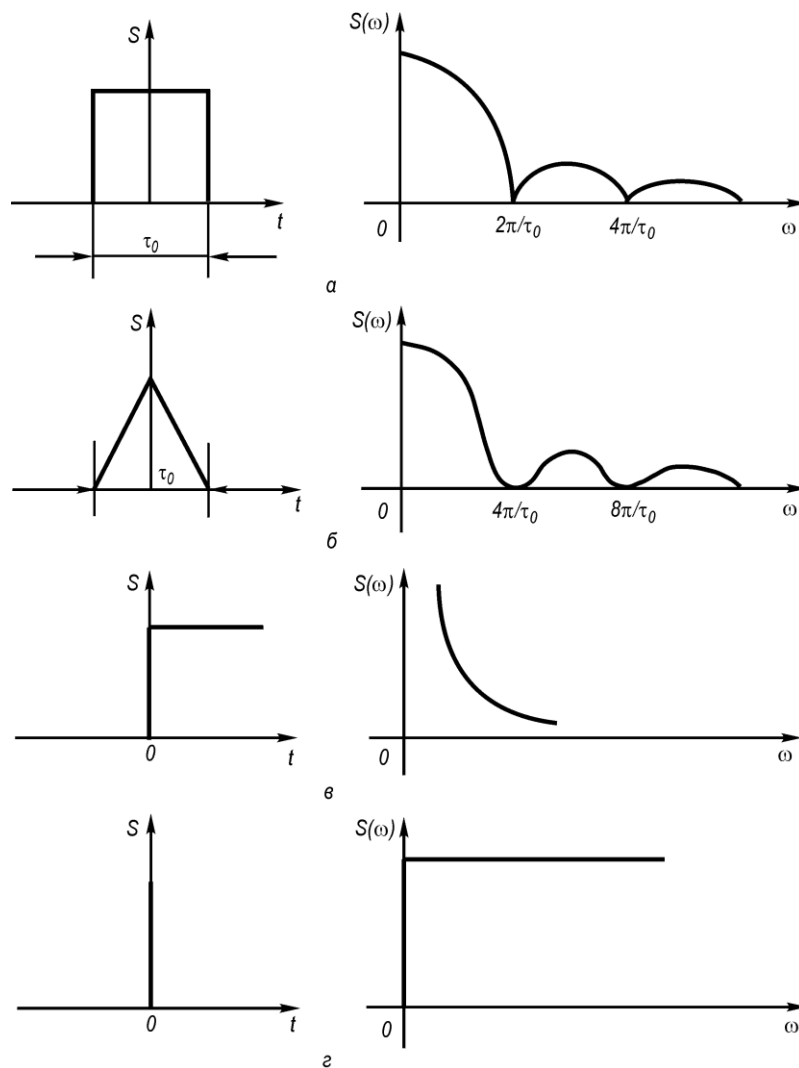


Рис. 2.4. Спектри деяких неперіодичних сигналів

2.2. Часове зображення сигналів. Дискретизація інформації

Дискретизація – це перетворення неперервних повідомлень в дискретні. При цьому використовується дискретизація за часом та рівнем. Перехід від аналогового зображення до дискретного у багатьох випадках дає значні переваги при передаванні, оброблянні та зберіганні інформації.

Заміна неперервної функції дискретною дає можливість всі результати, отримані для дискретних повідомлень, застосовувати до неперервних повідомлень.

Суть **дискретизації за часом** полягає в тому, що деякій неперервній функції ставиться у відповідність інша функція, створена шляхом переривання вихідної функції (рис. 2.5). Природно, що така заміна допустима лише в тих випадках, коли дискретизована функція повністю відображає вихідну функцію. Інтуїція підказує, що заміна неперервної функції дискретною можлива тільки за певних припущень.

Яким має бути інтервал Δt між окремими відліками? При малому інтервалі між відліками їх кількість буде великою і точність наступного відновлення функції також буде високою. Якщо ж інтервал між відліками взяти великим, то кількість відліків зменшиться, однак похибка відновлення неперервного повідомлення може виявитися більшою за допустиму. *Оптимальним* слід вважати такий інтервал між відліками, при якому вихідна функція із заданою точністю представляється мінімальним числом відлікових значень. У цьому випадку всі відліки будуть істотними для відновлення вихідної функції. При більшому числі відліків матиме місце *надлишковість інформації*.

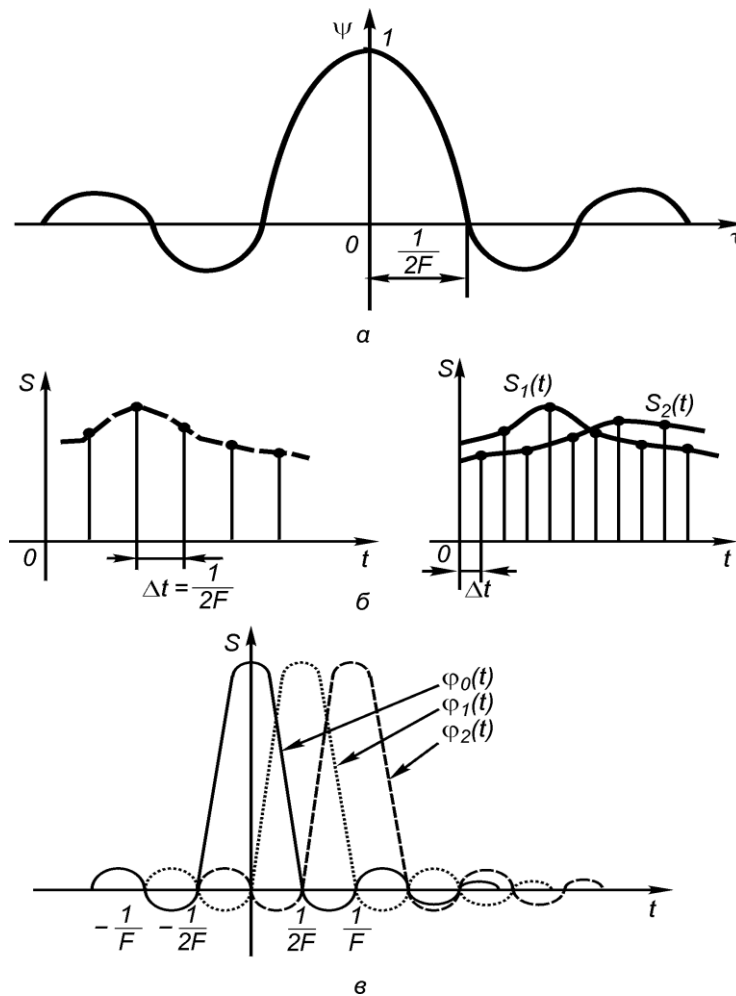


Рис. 2.5. Функції відліків (а, в) і дискретизація сигналу за часом (б)

Спосіб дискретизації, згідно з яким вибираються відліки або коефіцієнти розкладу, доцільно оцінювати за величиною похибки відновлення вихідної функції. Розрізняють три види критеріїв відліків:

1. *Частотний критерій Котельникова*, за яким інтервали між відліками вибираються виходячи з ширини спектра повідомлення, що дискретизується.
2. *Кореляційний критерій відліків Желєзнова*, за яким інтервал дискретизації вибирається рівним часу кореляції повідомлення, що передається.
3. *Квантовий критерій відліків Темникова*, запропонований для детермінованих функцій, який встановлює залежність інтервалів між відліками від ступеня квантування за рівнем і крутизною функції.

Оцінювати точність відновлення неперервного повідомлення можна за найбільшим, середньоквадратичним або інтегральним відхиленням.

Теорема В. О. Котельникова. Реальні сигнали завжди мають скінченну тривалість і обмежену смугу частот. Граничні частоти спектра визначаються головним чином властивостями системи передачі і самим отримувачем. Так, при передаванні дискретних повідомлень смуга частот визначається швидкістю передавання, при передаванні телевізійного зображення – прийнятим стандартом чіткості (кількістю рядків і т. ін.).

Для функцій з обмеженим спектром В. О. Котельников довів чудову теорему, що лежить в основі дискретизації неперервних сигналів:

функція $S(t)$, яка не містить частот, вищих від F , повністю визначається послідовністю своїх значень в моменти, які відстають один від одного на $\Delta t \leq \frac{1}{2F} = \frac{\pi}{\omega}$.

Фізичним підтвердженням цього є відомий факт про те, що сигнал $S(t)$ не може істотно змінитися за час, менший, ніж половина періоду його найвищої частоти, тобто $1/(2F)$.

Якщо спектр неперервного сигналу обмежений, не обов'язково передавати всі миттєві значення цієї функції; достатньо окремих її значень.

Доведення. Нехай функція $s(t)$ має обмежений спектр. Тоді

$$S(j\omega) = \begin{cases} \int_{-\infty}^{\infty} S(t)e^{-j\omega t} dt & \text{при } |\omega| \leq 2\pi F; \\ 0 & \text{при } |\omega| > 2\pi F. \end{cases}$$

На скінченному інтервалі $(-2\pi F, 2\pi F)$ функцію $S(j\omega)$ можна зобразити у вигляді ряду Фур'є за аргументом ω :

$$S(j\omega) = \frac{1}{2} \sum_{k=-\infty}^{\infty} A_k e^{jk \frac{\omega}{2F}}.$$

Коефіцієнти Фур'є у цьому випадку запишуться у вигляді

$$A_k = \frac{2}{4\pi F} \int_{-2\pi F}^{2\pi F} S(j\omega) e^{-j \frac{k\omega}{2F}} d\omega.$$

Оскільки $S(j\omega)$ є перетворенням Фур'є для $s(t)$, то $s(t)$ також є перетворенням Фур'є для $S(j\omega)$:

$$S(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(j\omega) e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{-2\pi F}^{2\pi F} S(j\omega) e^{j\omega t} d\omega.$$

Взявши $t = \frac{k}{2F}$, дістанемо

$$S\left(\frac{k}{2F}\right) = \frac{1}{2\pi} \int_{-2\pi F}^{2\pi F} S(j\omega) e^{j \frac{k\omega}{2F}} d\omega.$$

Порівнюючи отриманий вираз із виразом для A_k , бачимо, що

$$A_k = \frac{1}{F} S\left(-\frac{k}{2F}\right) = 2\Delta t S(-k\Delta t).$$

Отже,

$$S(j\omega) = \Delta t \sum_{k=-\infty}^{\infty} S(k\Delta t) e^{-j \frac{k\omega}{2F}}.$$

Оскільки $S(t)$ однозначно визначається своїм спектром $S(j\omega)$, то з отриманого співвідношення випливає, що $S(t)$, як і $S(j\omega)$, однозначно визначаються відліками $S(k\Delta t)$. Це і доводить теорему В. О. Котельникова.

Виразивши функцію $S(t)$ через її спектр $S(j\omega)$, отримаємо

$$S(t) = \frac{1}{2\pi} \int_{-2\pi F}^{2\pi F} S(j\omega) e^{j\omega t} d\omega = \frac{\Delta t}{2\pi} \int_{-2\pi F}^{2\pi F} e^{j\omega t} d\omega \sum_{-\infty}^{\infty} S(k\Delta t) e^{j\omega \Delta t};$$

змінивши порядок дії та зінтегрувавши, дістанемо

$$S(t) = \frac{\Delta t}{2\pi} \sum_{k=-\infty}^{\infty} S(k\Delta t) \int_{-2\pi F}^{2\pi F} e^{-j\omega(t-k\Delta t)} d\omega = \sum_{k=-\infty}^{\infty} S(k\Delta t) \frac{\sin 2\pi F(t-k\Delta t)}{2\pi F(t-k\Delta t)}. \quad (2.3)$$

Вираз (2.3) є розкладом неперервної функції $S(t)$ в ряд за ортогональними функціями вигляду $\frac{\sin x}{x}$. Величини $S(k\Delta t)$ називаються *відліками функції* $S(t)$ у дискретні моменти часу $k\Delta t$.

Множник $\frac{\sin 2\pi F(t-k\Delta t)}{2\pi F(t-k\Delta t)}$ називається *функцією відліків*.

Позначивши $\tau = t - k\Delta t$, отримаємо

$$\psi(\tau) = \frac{\sin 2\pi F\tau}{2\pi F\tau}, \quad \lim_{\tau \rightarrow 0} \psi(\tau) = 1.$$

Графік функції $\psi(\tau)$ зображений на рис. 2.5, а. Функція відліків набуває найбільшого значення, що дорівнює 1 в моменти $t = k\Delta t$ ($\tau = 0$), і перетворюється в нуль у моменти часу $t = (k \pm m)\Delta t$, де $m = 1, 2, \dots$. Ширина головної пелюстки функції відліків на нульовому рівні $-\frac{1}{F}$. Спектр функції відліків є рівномірним у смузі $(-F, F)$ і дорівнює нулеві поза цією смугою. Модуль спектра $S(\omega) = \frac{1}{2F}$.

Енергія сигналу через відлікові значення визначається так:

$$E = \int_{-\infty}^{\infty} S^2(t) dt = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} S^2(k\Delta t).$$

Якщо смуга частот сигналу $S(t)$ необмежено розширюється, то Δt необмежено зменшується, і в границі при $F \rightarrow \infty$ функція відліків прямує до дельта-функції $\delta(t - t_k)$, а ряд Котельникова перетворюється в інтеграл

$$S(t) = \int_{-\infty}^{\infty} S(t) \delta(t - t_k) dt. \quad (2.4)$$

Згортка дельта-функції з будь-якою функцією $S(t)$ дає рівність

$$\int_{-\infty}^{\infty} f(t) \delta(t - t_k) dt = f(t_k) \int_{-\infty}^{\infty} \delta(t - t_k) dt = f(t_k). \quad (2.5)$$

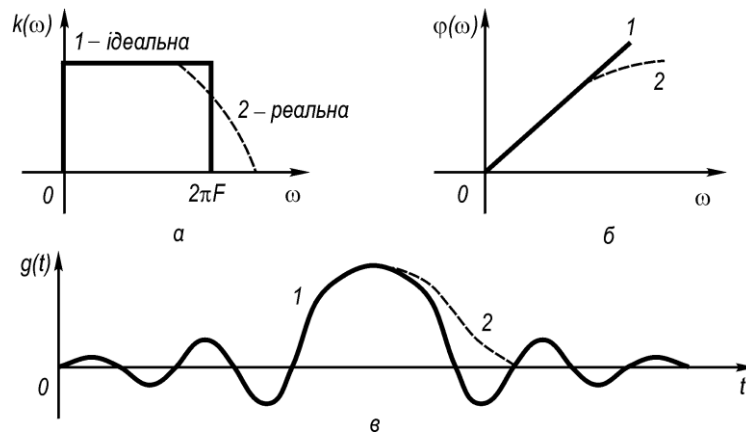


Рис. 2.6. Характеристика фільтра нижніх частот:
a – частотна; *б* – фазова; *в* – імпульсна; 1 – ідеальний фільтр;
 2 – реальний фільтр

Із порівняння виразів (2.4) і (2.5) випливає, що інтеграл, який визначається виразом (2.4), не зміниться із заміною $S(t)$ на $S(t_k)$. Тоді (2.4) можна записати у вигляді інтеграла Дюамеля:

$$S(t) = \int_0^T S(\tau) \delta(t - \tau) d\tau.$$

На рис. 2.5, *б* наведений неперервний сигнал, що дискретизується за часом, а на рис. 2.5, *в* – функції відліків для деяких сусідніх відліків.

Характеристику вигляду $\frac{\sin x}{x}$ має ідеальний фільтр нижніх частот (рис. 2.6) при передаванні на вхід одиничного імпульсу. Реальні характеристики залежності амплітуди від частоти фільтра $k(\omega)$ і фази від частоти $\varphi(\omega)$ відрізняються від ідеальних, що призводить до відхилення форми реальної функції від ідеальної і, як наслідок, до появи додаткових похибок відновлення функції $S(t)$.

Теорема Котельникова дозволяє однаково підходити до розгляду передавання будь-якого сигналу (дискретного чи неперервного) як до передавання чисел і лежить в основі всіх видів імпульсної модуляції. Згідно з нею, необхідна частота слідування імпульсів F_i , яку називають *тактовою частотою*, має визначатися з умови $F_i \geq 2F_m$, де F_m – верхня гранична частота спектра повідомлення, що передається.

Використовуючи теорему Котельникова, можна здійснювати часове ущільнення каналів, тобто передавати кілька сигналів одночасно. Два сигнали: $S_1(t)$ і $S_2(t)$, які передаються одночасно, зображені на рис. 2.5, *б*.

Теорема Котельникова може бути узагальнена і для функцій, спектр яких не починається з нульової частоти. У формулюванні автора ця теорема проголошує: *будь-яку функцію $S(t)$, що складається з частот від f_1 до f_2 , можна передавати з будь-якою точністю за допомогою чисел, які передаються одне за одним через інтервали $\Delta t = \frac{1}{F}$, де $F = f_2 - f_1$* . При цьому в кожній точці визначаються дві величини: амплітуда і фаза.

Коли тривалість сигналу дорівнює T , то кількість відліків на осі часу дорівнює $\frac{T}{\Delta t} = TF$, а загальна кількість відліків (амплітуд і фаз) $N = 2TF$.

Функцію $S(t)$, обмежену частотами f_1 і f_2 , можна зобразити рядом

$$S(t) = \sum_{k=-\infty}^{\infty} S\left(\frac{k}{F}\right) \frac{\sin \pi F \left(t - \frac{k}{F}\right)}{\pi F \left(t - \frac{k}{F}\right)} \cos \left[\omega_0 \left(t - \frac{k}{F}\right) + \varphi\left(\frac{k}{F}\right) \right],$$

де $S\left(\frac{k}{F}\right)$ і $\varphi\left(\frac{k}{F}\right)$ – відлікові значення відповідно амплітуди і фази; $\omega_0 = 2\pi \frac{f_1 + f_2}{2}$ – середнє значення колової частоти спектра сигналу.

Функція відліків в цьому випадку є модульованим коливанням із несучою, рівною ω_0 , і обвідною, яка визначається функцією вигляду $\frac{\sin x}{x}$ (рис. 2.7).

Оскільки, як зазначалося раніше, існує однозначний зв'язок між тривалістю імпульсу та шириною спектра, то поки розглядалися функції з обмеженим спектром, питання про тривалість не ставилося. Проте відомо, що реальні повідомлення мають скінченну тривалість. За теорією перетворення Фур'є, часове зображення функції з обмеженим спектром розтягується на нескінченно великий інтервал часу. В той же час скінченна тривалість повідомлення визначає нескінченно протяжний спектр. Це свідчить про неможливість існування функції скінченної тривалості з обмеженим спектром.

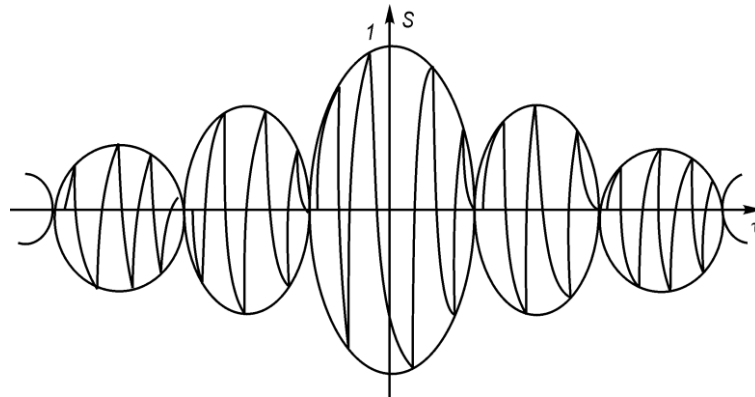


Рис. 2.7. Функція відліків для неперервного сигналу в смузі частот, обмеженій частотами f_1 та f_2

Отже, принципово теорема Котельникова є неслухною для реальних повідомлень. Але, враховуючи відмічені особливості реальних повідомлень [зосередженість майже всієї енергії в скінченних інтервалах часу і в смузі частот], можна з достатнім ступенем точності використати її для зображення реальних повідомлень. У цьому випадку при тривалості повідомлення T_{Π} та верхній частоті F_B число відліків m буде скінченним:

$$m = \frac{T_{\Pi}}{\Delta t} = 2F_B T_{\Pi}, \quad (2.6)$$

а вираз (2.3) набере вигляду

$$S(t) \approx \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} S(k\Delta t) \frac{\sin \omega_B (t - k\Delta t)}{\omega_B (t - k\Delta t)}. \quad (2.7)$$

При скінченному числі відліків сума (1.7) збігатиметься з миттєвими значеннями $S(t)$ не на всьому інтервалі існування повідомлення T_{Π} , а тільки у відлікових точках. У проміжках між цими

значеннями $S(t)$ відрізняється від суми скінченного числа членів ряду, внаслідок чого виникає похибка. Ця похибка мінімальна в середині інтервалу T_n і зростає до його країв (рис. 2.8).

Визначити середньоквадратичну похибку відновлення повідомлення скінченної тривалості рядом Котельникова зі скінченним числом членів ряду важко. Проте границі похибки, що виникає при відновленні реального повідомлення із необмеженим спектром рядом Котельникова (2.3), визначені. Позначивши відновлене повідомлення $S'(t)$, запишемо вираз для середньоквадратичної похибки δ :

$$\sqrt{\frac{\Delta E}{E}} \leq \delta = \frac{\sqrt{\int_{-\infty}^{\infty} [S'(t) - S(t)]^2 dt}}{\sqrt{\int_{-\infty}^{\infty} S^2(t) dt}} \leq \sqrt{\frac{\Delta E}{E}} (3 + Q), \quad (2.8)$$

де $E = \int_{-\infty}^{\infty} S^2(t) dt$ – повна енергія повідомлення; ΔE – енергія повідомлення, що міститься в смузі частот, вищих за $F_b = \frac{1}{2\Delta t}$; Q – коефіцієнт, що враховує швидкість зменшення модуля спектральної щільності на частотах, вищих за F_b [1]:

$$Q = 4 \sum_{r=1}^{\infty} \left[\frac{\int_{(2r+1)\omega_b}^{\infty} [S(\omega)]^2 d\omega}{\int_{\omega_b}^{\infty} [S(\omega)]^2 d\omega} \right]^{1/2}, \quad (2.9)$$

де $r = 1, 2, \dots$ – крок, що залежить від частоти.

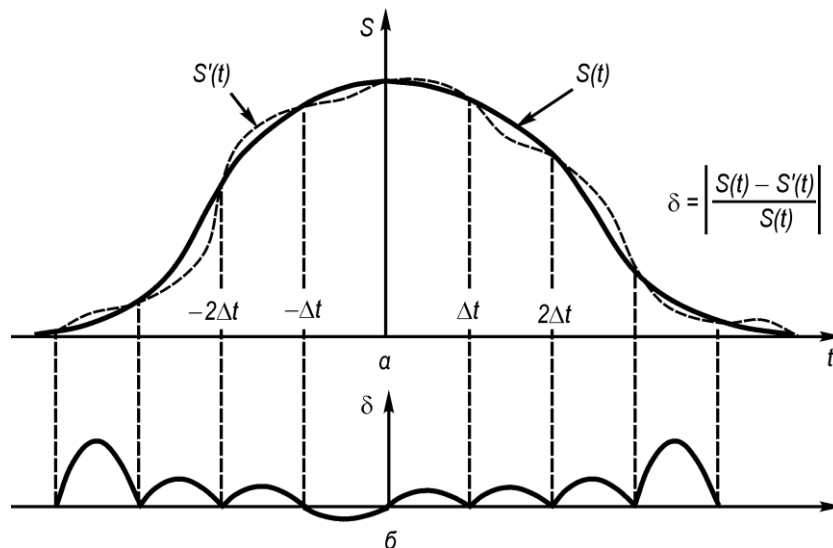


Рис. 2.8. До оцінки похибки відновлення неперервної функції

Якщо модуль спектральної щільності спадає достатньо швидко, користуються більш простою оцінкою:

$$\sqrt{\frac{\Delta E}{E}} \leq \delta \leq \sqrt{\frac{3\Delta E}{E}}, \quad (2.10)$$

за допомогою якої можна визначити F_B так, щоб похибка відтворення неперервного повідомлення не перевищувала заданої величини.

В практичних розрахунках для визначення відносної похибки відтворення повідомлень можна користуватися виразом

$$\delta^2 \approx \frac{\int_0^{\omega_B} [S(\omega)]^2 d\omega}{\int_0^{\infty} [S(\omega)]^2 d\omega}, \quad (2.11)$$

який дає змогу за заданою величиною δ і відомими спектральними характеристиками повідомлення знайти верхню граничну частоту спектра ω_0 та інтервал дискретизації

$$\Delta t = \frac{\pi}{\omega_B}. \quad (2.12)$$

Принцип дискретизації Желєзнова. М. А. Желєзновим розв'язана задача визначення максимального інтервалу між відліками випадкового повідомлення. Моделі повідомлення характеризуються скінченною тривалістю повідомлення T_n і суцільним спектром, що відрізняється від нуля по всій частотній осі, тобто $S(\omega) \neq 0$ при $-\infty \leq \omega \leq \infty$.

Вводиться припущення обмеженості інтервалу кореляції (вважається, що функція кореляції дорівнює нулю поза інтервалом τ_0) і малого розміру інтервалу кореляції в порівнянні з тривалістю повідомлення ($T_n \gg \tau_0$). Розглядаються повідомлення, що є стаціонарними і нестаціонарними функціями часу. Введені припущення не вступають у протиріччя з природою реальних повідомлень, оскільки внаслідок скінченної тривалості їх значення в будь-який момент залежать тільки від деякого відрізка попереднього повідомлення зі скінченною тривалістю. Тому інтервал кореляції реальних повідомлень є обмеженою величиною. Це обмеження, що записується у вигляді

$$B_S(\tau) = \begin{cases} B_S(\tau) & \text{при } \tau \leq \tau_0; \\ 0 & \text{при } \tau > \tau_0, \end{cases} \quad (2.13)$$

наведено на рис. 2.9, а.

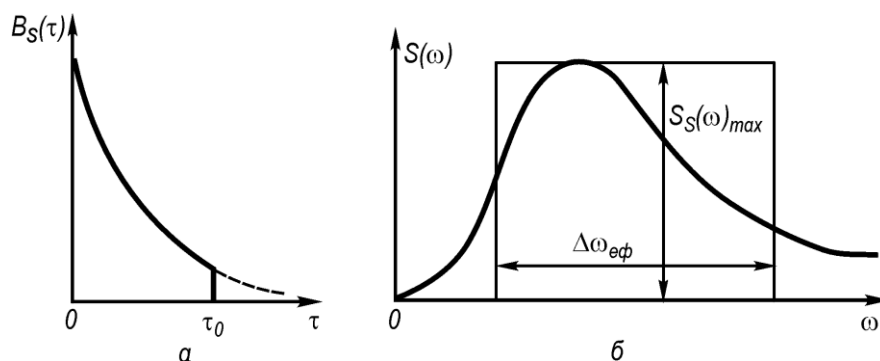


Рис. 2.9. До обмеження інтервалу кореляції (а) та визначення ефективної смуги частот (б)

М. А. Железновим доведено, що такі повідомлення можуть бути завбачені системою лінійного прогнозування із середньоквадратичною

похибкою, котра як завгодно мало відрізняється від нуля, тільки на проміжку часу, рівному інтервалу кореляції τ_0 . Таким чином, дискретизацію слід проводити з інтервалом, який не перевищує τ_0 , оскільки лише в цьому випадку можливе безпомилкове відновлення вихідного повідомлення. Число некорельованих відліків при цьому

$$N = \frac{T_{\Pi}}{\tau_0}. \quad (2.14)$$

Для визначення інтервалу кореляції τ_0 використовується поняття *ефективної смуги частот повідомлення* $\Delta\omega_{\text{еф}} = 2\pi\Delta f_{\text{еф}}$; інтервал кореляції

$$\tau_0 = \frac{1}{2\Delta f_{\text{еф}}}, \quad (2.15)$$

де $\Delta f_{\text{еф}}$ – смуга частот, у якій зосереджено до 90 % потужності сигналу.

Величина $\Delta\omega_{\text{еф}}$ визначається як основа прямокутника з висотою, яка дорівнює максимальному значенню спектральної щільності $S_S(\omega)$, і площею, яка дорівнює площі під кривою спектральної щільності повідомлення (рис. 2.9, б):

$$\Delta\omega_{\text{еф}} = \frac{1}{S_S(\omega)_{\max}} \int_0^{\infty} S_S(\omega) d\omega. \quad (2.16)$$

Для нестационарних функцій використовується поняття поточного інтервалу кореляції, який є функцією часу: $\tau_0 = \tau_0(t)$. У цьому випадку відлікові значення неперервного повідомлення розташовуються на осі часу нерівномірно. Метод дискретизації М. А. Железнова менш розроблений, ніж метод В. О. Котельникова, і тому поки що не поширений.

Дискретизація неперервних повідомлень з використанням критерію найбільшого відхилення. Вибір кроку дискретизації з використанням даного критерію проводиться в припущенні, що вихідне повідомлення відновлюється за допомогою полінома степеня n . На деякому відрізку часу t_0, t_n для рівновіддалених відліків відновлене повідомлення $S'(t)$ може бути подане виразом

$$S'(t) = \sum_{k=0}^n S(t_k) \frac{(t-t_0)(t-t_1)\cdots(t-t_{k-1})(t-t_{k+1})\cdots(t-t_n)}{(t_k-t_0)(t_k-t_1)\cdots(t_k-t_{k-1})(t_k-t_{k+1})\cdots(t_k-t_n)}, \quad (2.17)$$

або в скороченому вигляді

$$S'(t) = \sum_{k=0}^n \left[S(t_k) \prod_{\substack{i=0 \\ i \neq k}}^n \frac{t-t_i}{t_k-t_i} \right]. \quad (2.18)$$

Для відновлення функції $S'(t)$ за допомогою полінома степеня n необхідно мати $n+1$ відлік. Похибка відновлення вихідного повідомлення у цьому разі визначиться залишковим членом:

$$R(t) = S(t) - S'(t) = \frac{S^{(n+1)}(\xi)}{(n+1)!} \prod_{k=0}^n (t - t_k), \quad (2.19)$$

де $S^{(n+1)}(\xi)$ – значення $(n+1)$ -ї похідної повідомлення $x(t)$, взятої в деякій точці ξ , що лежить всередині інтервалу $t_n - t_0$. Оскільки розташування точки ξ невідоме, для оцінки використовують модуль максимального значення похідної M_{n+1} на заданому інтервалі. Тоді

$$R(t) = \frac{M_{n+1}}{(n+1)!} \prod_{k=0}^n (t - t_k). \quad (2.20)$$

Вводячи обмеження

$$R(t) \leq \varepsilon_0, \quad (2.21)$$

де ε_0 – допустима похибка дискретизації за часом, знаходять крок дискретизації або довжину інтервалу $t_n - t_0$, на якому можна визначити $n+1$ відлікове значення неперервного повідомлення. При цьому інтервал $t_n - t_0$ є деякою функцією похибки ε_0 , міри відновлення повідомлення і т. д.

Цікавим є визначення кроку дискретизації при використанні твірних поліномів нульового та першого степенів.

Нульовому степеню твірного полінома відповідає східчаста апроксимація неперервного повідомлення. У цьому випадку (1.21) набирає вигляду

$$R(t) = \frac{M_1}{1!} (t_0 - t_1) = M_1 \Delta t \leq \varepsilon_0,$$

Звідки

$$\Delta t \leq \frac{\varepsilon_0}{M_1}. \quad (2.22)$$

При $n=1$ (лінійна апроксимація) залишковий член

$$R(t) = \frac{M_2}{2!} \prod_{k=0}^1 (t - t_k) \leq \varepsilon_0.$$

Максимізуючи добуток $\prod_{k=0}^1 (t - t_k) = (t - t_0)(t - t_1)$, отримаємо $\frac{M_2 \Delta t^2}{8} \leq \varepsilon_0$, звідки

$$\Delta t \leq \sqrt{\frac{8\varepsilon_0}{M_2}}. \quad (2.23)$$

Відновлювання неперервного повідомлення з використанням поліномів нульового і першого степенів наведено на рис. 2.10, а і б.

Кроки дискретизації для поліномів більш високих степенів такі:

Степінь полінома	0	1	2	3	4
Крок дискретизації	$\frac{\varepsilon_0}{M_1}$	$\sqrt{\frac{8\varepsilon_0}{M_2}}$	$\sqrt[3]{\frac{15,53\varepsilon_0}{M_3}}$	$\sqrt[4]{\frac{24\varepsilon_0}{M_4}}$	$\sqrt[5]{\frac{33\varepsilon_0}{M_5}}$

Слід зазначити, що для деяких видів повідомлень частота дискретизації, визначена за цим методом, виявляється значно меншою від частоти, знайденої за теоремою Котельникова.

Наприклад, для відновлення повідомлення, що описується експоненціальною функцією $e^{-\frac{t}{\tau}}$ при $\tau=1$ з похибкою, рівною 0,5 %, за теоремою Котельникова необхідна частота дискретизації більша за 8 кГц, тоді як за критерієм найбільшого відхилення при $n=1$ частота дискретизації становитиме 5 Гц. При похибці 0,05 % частоти становитимуть відповідно понад 700 кГц і 15,9 Гц. Різниця, зумовлена використанням різних методів відновлення повідомлень, є дуже істотною.

Розглянуті методи не вичерпують всіх можливих методів відновлення неперервних повідомлень на основі їх дискретних значень. З цією метою можуть використовуватись різноманітні фільтри, характеристики яких істотно впливають на вибір частоти дискретизації. Основними вихідними даними і в цьому випадку є ширина спектра повідомлення та допустима похибка відновлення неперервного повідомлення. Значення частоти дискретизації у разі використання деяких видів фільтрів (фільтр Баттерворта, оптимальний фільтр, РС-фільтр).

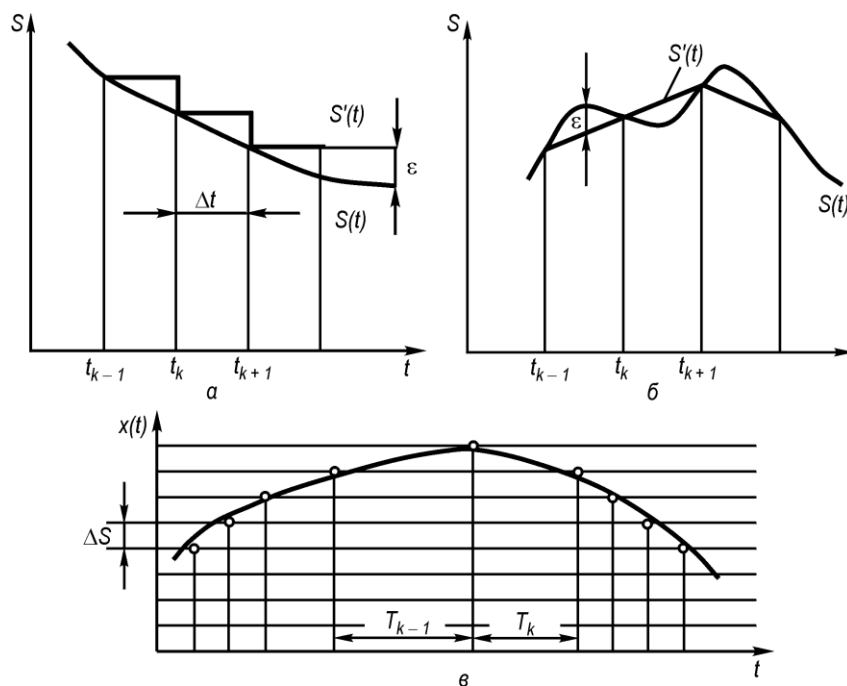


Рис. 2.10. Відновлювання неперервних сигналів (а, б) та адаптивна дискретизація (в)

Поняття про адаптивну дискретизацію. Для розглянутих раніше методів дискретизації характерна постійність інтервалів між відліковими значеннями функції, що дискретизується. В той же час із (2.22) і (2.23) випливає, що у разі фіксації швидкості змінення або прискорення неперервного повідомлення з'являється можливість такої зміни кроку дискретизації у відповідності до зміни параметрів повідомлення, яке дискретизується, що похибка ніколи не перевищуватиме встановленої величини ε_0 . При адаптивній (тобто пристосувальній до умов

каналу) дискретизації (для $n=0$) відлікові значення $S(t_k)$ неперервного повідомлення беруться в моменти часу, коли

$$|S(t) - S(t_{k-1})| = \varepsilon_0. \quad (2.24)$$

В цей момент здійснюється також відлік розміру інтервалу дискретизації T_k (рис. 2.10, в).

У разі використання твірного полінома першого степеня в пристрої дискретизації на кожному інтервалі генерується наближаюча пряма і безперервно визначається різниця між поточним значенням неперервного повідомлення і наближаючою функцією. Коли різниця сягає величини ε_0 , береться черговий відлік і визначається величина інтервалу дискретизації T_k .

Вказані методи дискретизації належать до методів адаптивної дискретизації з некрatними інтервалами, оскільки величини інтервалів дискретизації можуть набувати довільних значень. Це зумовлює необхідність фіксації всіх моментів часу відліку t_k .

При адаптивній дискретизації з кратними інтервалами відлікові значення беруться в моменти часу, що віддалені один від одного на ціле число Δt , тобто $t_k = t_{k-1} + i\Delta t$, де $i=1, 2, 3, \dots$, з використанням співвідношень, аналогічних (2.24).

Адаптивна дискретизація зменшує число відлікових значень неперервного повідомлення без збільшення похибки його відновлення, тобто знижує до певної міри надлишковість інформації. Подальше зниження надлишковості може бути досягнуто зміною степеня твірної функції.

Слід зазначити, що реалізація алгоритмів адаптивної дискретизації складніша, ніж реалізація алгоритмів рівномірної дискретизації.

2.3. Квантування за рівнем

Відліки неперервних повідомлень, взяті в дискретні моменти, характеризуються тим, що рівні повідомлень* можуть набувати нескінченну множину значень, якщо точність їх визначення не обмежена. Оскільки будь-яке повідомлення надходить до споживача разом із завадою (похибку вимірювання також можна трактувати як заваду), то нема сенсу передавати повідомлення точніше, ніж це дозволяє рівень завад. Доцільно за рівнем супровідних завад і необхідної точності повідомлення, що має бути отриманим, обмежити число рівнів повідомлення. *Заміна неперервної шкали рівнів повідомлення дискретною шкалою рівнів називається квантуванням за рівнем.* Якщо миттєве значення рівня повідомлення знаходиться всередині інтервалу, розташованого між двома дозволеними дискретними рівнями, то замість нього передається значення, що відповідає ближчому дозволеному рівню. При кількості рівнів квантування m будь-яке повідомлення міститиме не більш, ніж m значень, які можна розрізнити.

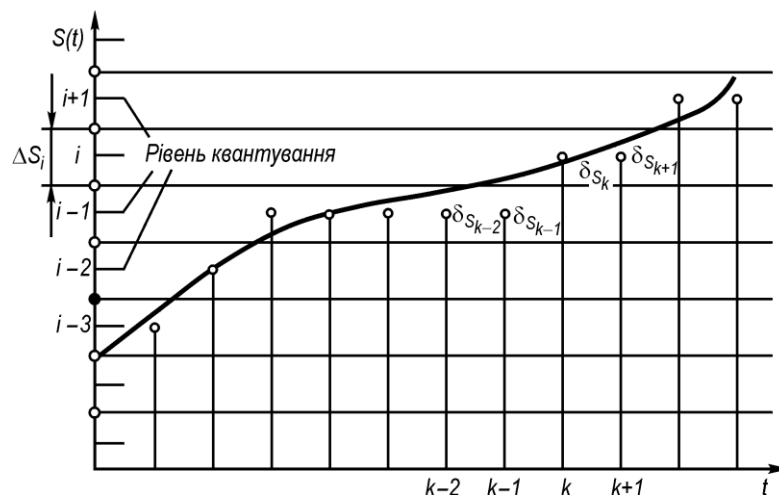


Рис. 2.11. Квантування за рівнем

Квантування за рівнем може бути рівномірним чи нерівномірним. При *рівномірному квантуванні* динамічний діапазон зміни повідомлення ділиться на $m-1$ рівні частини (тобто інтервали між всіма рівнями квантування однакові), а при *нерівномірному* – на $m-1$ нерівні частини. Таким чином отримуємо шкалу рівнів квантування.

Щоб мати можливість чітко визначити належність істинних миттєвих значень повідомлення певному рівню, динамічний діапазон повідомлення $S(t)$ ділиться на інтервали, які називають *кроками квантування*. Рівні квантування при цьому розташовуються всередині кроків квантування. Крок квантування позначають ΔS ; для i -го рівня крок квантування буде ΔS_i . Якщо значення амплітуди сигналу в деякий момент часу [звичайно цей момент збігається з моментом взяття відліку за часом $S(t_k)$] знаходиться всередині i -го кроку квантування ΔS_i , то амплітуда сигналу замінюється значенням i -го рівня $S_i(t_k)$. Оптимальним щодо точності відтворення квантованого сигналу є розташування рівня квантування всередині кроку квантування. Процедура квантування за рівнем подана на рис. 2.11.

Квантоване повідомлення $S_i(t_k)$ є сумою істинного повідомлення $S(t_k)$ і похибки δ_{S_i} . Таким чином, квантування можна розглядати як проходження повідомлення через систему, на яку діє завада δ_{S_i} . Цю заваду називають *шумом квантування*. Оскільки саме повідомлення є випадковим процесом, шум квантування – також випадкова функція часу. Оцінімо статистичні характеристики шуму квантування. Виникаюча в момент t_k помилка визначається виразом

$$\delta_{S_i} = S(t_k) - S_i(t_k). \quad (2.25)$$

Повідомлення $S(t)$ у момент часу t_k може набувати одного з можливих значень в інтервалі від S_{\min} до S_{\max} з деякою ймовірністю, що визначається законом розподілу випадкової величини $S(t)$. Ймовірність появи рівня S_i визначиться ймовірністю попадання випадкової величини $S(t)$ в інтервал від S_a до S_b в околі рівня S_i . При цьому похибка δ_{S_i} (рис. 2.11) може змінюватися в межах від $S_i - S_a$ до $S_b - S_i$, змінюючи знак. Ймовірність появи похибки δ_{S_i} визначається ймовірністю появи відповідного значення повідомлення $S(t)$ в момент t_k .

Математичне сподівання похибки –

$$M(\delta_{S_i}) = \int_{S_a}^{S_b} (S - S_i) p(S) dS, \quad (2.26)$$

а середньоквадратичне значення –

$$M(\delta_{S_i}^2) = \int_{S_a}^{S_b} (S - S_i)^2 p(S) dS, \quad (2.27)$$

де $p(S)$ – щільність ймовірності.

Вважаючи динамічний діапазон зміни повідомлення набагато більшим від величини кроку квантування, візьмемо функцію щільності ймовірності $p(S)$ на інтервалі інтегрування сталою і рівною своєму значенню для деякого середнього рівня $p(S_{\text{cp}})$. Тоді

$$M(\delta_{S_i}) = p(S_{\text{cp}}) \int_{S_a}^{S_b} (S - S_i) dS = \frac{1}{2} p(S_{\text{cp}}) [(S_b - S_i)^2 - (S_a - S_i)^2], \quad (2.28)$$

$$M(\delta_{S_i^2}) = p(S_{\text{cp}}) \int_{S_a}^{S_b} (S - S_i)^2 dS = \frac{1}{3} p(S_{\text{cp}}) [(S_b - S_i)^3 - (S_a - S_i)^3], \quad (2.29)$$

де $S_{\text{cp}} = \frac{S_b - S_a}{2}$.

Визначимо оптимальне положення рівня S_i відносно S_a і S_b , при якому середньоквадратична похибка буде мінімальною. Зрівнявши з нулем похідну $\frac{\partial [M(\delta_{S_i^2})]}{\partial S_i}$, дістанемо

$$S_b - S_i^2 = S_a - S_i^2.$$

Положення рівня S_i , яке відповідає мінімуму похибки,

$$S_i = \frac{S_b + S_a}{2} = S_{\text{cp}}. \quad (2.30)$$

Неважко помітити, що при цьому (2.28) перетворюється в нуль, тобто $M(S_i) = 0$.

Отриманий результат свідчить про те, що для забезпечення мінімальної дисперсії рівень квантування S_i слід розташовувати посередині кроку квантування. В цьому випадку $S_a = S_i - \frac{\Delta S_i}{2}$, $S_b = S_i + \frac{\Delta S_i}{2}$, а дисперсія знаходиться із (2.29):

$$M(\delta_{S_i^2}) = \frac{1}{12} p(S_i) \Delta S_i^3. \quad (2.31)$$

Дисперсію шуму квантування D отримаємо, додаючи значення $M(\delta_{S_i^2})$ за всіма рівнями:

$$D = \sum_{i=1}^m M(\delta_{S_i^2}) = \frac{1}{12} \sum_{i=1}^m p(S_i) \Delta S_i^3. \quad (2.32)$$

У разі рівномірного квантування, коли $\Delta S_1 = \Delta S_2 = \dots = \Delta S_m = \Delta S$,

$$D = \frac{\Delta S^2}{12} \sum_{i=1}^m p(S_i) \Delta S. \quad (2.33)$$

Вважаючи, що значення повідомлення обмежуються динамічним діапазоном, можна записати $\sum_{i=1}^m p(S_i) \Delta S = 1$. Остаточно для рівномірного квантування

$$D = \frac{\Delta S^2}{12}. \quad (2.34)$$

Згідно з (2.30), максимальна різниця (при рівномірному квантуванні) між істинним значенням повідомлення і квантованим значенням не перевищує $\frac{\Delta S}{2}$. Середньоквадратична похибка квантування при цьому становить $\sigma_{cp} = \sqrt{D} = \frac{\Delta S}{2\sqrt{3}}$, тобто менша за максимальну похибку в $\sqrt{3}$ разів.

Із (2.31), (2.33) випливає, що значення дисперсії шуму квантування і середньоквадратичної похибки залежать як від виду функції розподілу щільності ймовірності, так і від характеру зміни шуму квантування в межах динамічного діапазону повідомлення (2.32). Із аналізу (2.32) випливає, що для зменшення середньоквадратичної похибки доцільно для менш імовірних значень збільшувати крок квантування, а для більш імовірних – зменшувати. Квантування за рівнем, яке забезпечує мінімум дисперсії шуму квантування, називають *оптимальним квантуванням за рівнем*. Умовою оптимальності квантування є постійність добутку $\Delta S_i \sqrt[3]{p(S_i)}$ в межах динамічного діапазону. Функціональний зв'язок між функцією повідомлення y і змінною S , що утворює в інтервалі від 0 до S_{max} оптимальну шкалу квантування, визначається так:

$$S(y) = \frac{S_{max}}{S_1} \int_0^y \frac{dy}{\sqrt[3]{p(y)}}, \quad (2.35)$$

де $S_1 = \int_0^{S_{max}} \frac{dy}{\sqrt[3]{p(y)}}$; $p(y)$ – функція щільності ймовірності повідомлення $y(t)$.

Обчислення за (2.35) показали, що оптимальне квантування неперервних функцій з нормальним законом розподілу дає змогу зменшити дисперсію шуму квантування в порівнянні з рівномірним законом в окремих випадках більше, ніж учетверо. Аналогічно існує можливість зменшення дисперсії шуму квантування і для других законів розподілу.

Зупинимося на способах оптимального квантування повідомлень. Їх два: безпосередній і з попереднім перетворенням динамічного діапазону. При *безпосередньому способі* пристрій квантування на підставі апріорно відомої функції розподілу $p(S)$ вибирає величину кроку квантування ΔS_i відповідно до значень повідомлення, яке квантується. Такий пристрій є достатньо складним за своєю структурою і операціями, які він виконує. При квантуванні з *перетворенням динамічного діапазону* повідомлення подається на нелінійний амплітудний перетворювач, що забезпечує перетворення закону розподілу ймовірностей в рівноймовірний (або близький до нього). Пристрій квантування оперує з однаковою величиною кроку квантування для всіх рівнів. Амплітудна характеристика перетворювача визначається виглядом функції $p(y)$. Часто на практиці використовується логарифмічна характеристика нелінійного перетворення. Таке перетворення дозволяє зменшити шум квантування без збільшення числа рівнів. Зазначимо, що при стисканні динамічного діапазону повідомлення спотворюються. Для усунення спотворення повідомлення піддають зворотному нелінійному перетворенню згідно з прийнятим раніше законом стискання діапазону.

2.4. Геометричні зображення сигналів

Аналогія між сигналами і векторами. Будь-яка задача легше сприймається, якщо її можна пов'язати з яким-небудь відомим явищем. У математичному описі сигнали зручно розглядати як вектори чи точки в деякому функціональному просторі – *просторі сигналів*.

Електричні сигнали складної форми за своєю фізичною природою далеко не завжди схожі на звичні нам уявлення векторів як спрямлених відрізків. Проте практичний інтерес має узагальнення операцій над векторами на сигнали (функції), котрі описують різні коливання. Річ у тому, що серед різних математичних прийомів, які використовуються при дослідженні електричного кола і сигналів, найбільш широко застосовується зображення довільної функції у вигляді суми більш

простих (елементарних) функцій. Такий підхід лежить в основі принципу незалежності дії (суперпозиції) при вивченні перетворень сигналів у лінійних електричних колах. Наочні геометричні зображення, пов'язані з відображенням функцій як векторів простору сигналів, допомагають часто усвідомити фізичний зміст процесів формування, передавання і розподілу сигналів, синтезу оптимальних сигналів і пристроїв обробки сигналів у разі наявності завад.

Задача розкладення сигналу складної форми на найпростіші складові схожа з задачею розкладення звичайного вектора s тривимірного простору на його складові по координатному базису одиничних ортогональних векторів $\mathbf{i}, \mathbf{j}, \mathbf{k}$ (рис. 2.12, а):

$$\mathbf{s} = s_1\mathbf{i} + s_2\mathbf{j} + s_3\mathbf{k}. \quad (2.36)$$

Складовими вектора s за базисом $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ будуть вектори $s_1\mathbf{i}, s_2\mathbf{j}, s_3\mathbf{k}$. Коефіцієнти s_1, s_2, s_3 є проєкціями вектора s на координатні осі $\mathbf{i}, \mathbf{j}, \mathbf{k}$ і називаються *координатами* вектора s . Інакше кажучи, вектор s у тривимірному просторі цілком визначається сукупністю його координат: $\mathbf{s} = s_1, s_2, s_3$.

Щоб перейти до узагальнення поняття вектора тривимірного простору для випадку n -вимірного простору, звернемося до прикладу. Деяке наближене уявлення про функцію (сигнал) $s(t)$ можна скласти, відображаючи її послідовністю прямокутних імпульсів, що мають на інтервалах $i\Delta$ значення $s(i\Delta)$ (рис. 2.12, б). Якщо тепер умовно зобразити функцію $s(t)$ на інтервалі $0; T$ вектором, то для його визначення буде потрібно $n = T/\Delta$ координат $s_i = s(i\Delta)$. Це означає, що функцію $s(t)$ за аналогією з (2.36) можна подати у вигляді суми:

$$s(t) = \sum_{i=0}^{n-1} s_i \Psi_i(t), \quad (2.37)$$

де $\Psi_i(t) = \Psi(t - i\Delta)$ – елементарні базисні функції;

$$\Psi(t) = \begin{cases} 1, & t \in [0; \Delta]; \\ 0, & t \notin [0; \Delta]; \end{cases} \quad \frac{1}{\Delta} \int_0^{\Delta} \Psi^2(t) dt = 1.$$

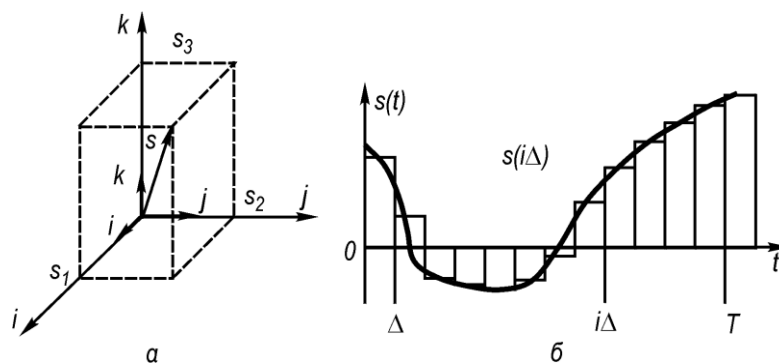


Рис. 2.12. Зображення вектора s у тривимірній ортогональній системі координат (а) та відображення неперервної функції послідовністю прямокутних імпульсів (б)

Вектор s , що відповідає функції $s(t)$, у n -вимірному просторі одиничних ортів Ψ_i цілком визначатиметься його координатами: $\mathbf{s} = s_0, s_1, \dots, s_{n-1}$. Таким чином, сигнал $s(t)$ довільної форми подається сумою n найпростіших елементарних сигналів, у даному випадку – у вигляді імпульсів прямокутної форми. Слово *простір* використовується для того, щоб надати множині

сигналів геометричного змісту і тим самим наочності. Найбільш простим і в той же час фізично досить змістовним є трактування сигналів як елементів *нормованого лінійного метричного простору*.

Основні означення в функціональних просторах Евкліда, Гільберта, Хеммінга. *Лінійним* чи *векторним* називається простір сигналів, для елементів якого виконуються правила додавання і множення на будь-яке число з деякої множини λ , що називають множиною скалярів. Додавання векторів проводиться по координатно; отже, сумою векторів \mathbf{s} [функції $s(t)$] і \mathbf{y} [функції $y(t)$] називається вектор $\mathbf{s} + \mathbf{y} = s_0 + y_0, s_1 + y_1, \dots, s_n + y_n$, що належить даному простору, а добуток вектора \mathbf{s} на число λ дає вектор $\lambda\mathbf{s} = \lambda x_0, \lambda x_1, \dots, \lambda x_{n-1}$, що також належить даному простору. В лінійному просторі існує нульовий елемент, такий, що $\mathbf{s} + 0 = \mathbf{s}$, і кожному елементу \mathbf{s} відповідає протилежний елемент $-\mathbf{s}$, тобто $\mathbf{s} + (-\mathbf{s}) = 0$. Вектор, утворений підсумовуванням n лінійно незалежних (базисних) векторів Ψ_i зі скалярними коефіцієнтами s_i , називається їх *лінійною комбінацією*: $\mathbf{s} = \sum_{i=0}^{n-1} s_i \Psi_i$.

Множина векторів Ψ_i називається *лінійно незалежною (базисом)*, якщо умова

$$\sum_{i=0}^{n-1} s_i \Psi_i = 0 \quad (2.38)$$

виконується лише тоді, коли всі $s_i = 0$. Інакше кажучи, лінійно незалежною називається множина Ψ_i , для якої жодна з її компонент не може бути утворена лінійною комбінацією інших. Розмірність лінійного простору визначається числом будь-яких лінійно незалежних базисних векторів Ψ_i , що утворюють цей простір. Лінійно незалежні вектори Ψ_i можна розглядати як координатні осі простору.

Метричним називається лінійний простір, у якому визначена відстань між елементами (векторами) простору (*метрика*), тобто кожній парі елементів, наприклад \mathbf{s} і \mathbf{y} , можуть бути поставлені у відповідність деяке дійсне невід'ємне число $d(\mathbf{s}, \mathbf{y})$ і спосіб, за допомогою якого знаходиться це число. Відстань задовольняє наступним правилам:

- 1) $d(\mathbf{s}, \mathbf{y}) = 0$, якщо $\mathbf{s} = \mathbf{y}$;
- 2) $d(\mathbf{s}, \mathbf{y}) = d(\mathbf{y}, \mathbf{s})$;
- 3) $d(\mathbf{s}, \mathbf{y}) \leq d(\mathbf{s}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$,

де $\mathbf{s}, \mathbf{y}, \mathbf{z}$ – елементи (точки) простору.

Зміст перших двох умов очевидний. Третю умову називають *нерівністю трикутника*: довжина сторони трикутника менша за суму довжин двох інших сторін або дорівнює їй.

Серед лінійних метричних просторів важливе місце займають *нормовані простори*. Цей вид просторів визначається заданням норми $\|\mathbf{s}\|$, що задовольняє наступним аксіомам:

$$\begin{aligned} \|\mathbf{s}\| &\geq 0; \\ \|\lambda\mathbf{s}\| &= |\lambda| \cdot \|\mathbf{s}\|; \\ \|\mathbf{s} + \mathbf{y}\| &\leq \|\mathbf{s}\| + \|\mathbf{y}\|. \end{aligned}$$

Перша аксіома встановлює, що норма є позитивне дійсне число, рівне нулю тільки для нульового вектора; друга аксіома – що λ є будь-яке число (скаляр); третя аксіома – аксіома трикутника.

Почнемо з перерахування термінів і означень, що використовуються в описанні n -вимірною евклідового простору R_n . Будь-який вектор \mathbf{s} у цьому просторі визначається сукупністю його координат: $\mathbf{s} = s_0, s_1, \dots, s_{n-1}$.

Сукупність n лінійно незалежних векторів утворює n -вимірний евклідовий простір, що позначається R_n . Простір R_n можна визначити як множину точок, представлених кінцями векторів, для яких норма

$$\|\mathbf{s}\| = \sqrt{\sum_{i=0}^{n-1} s_i^2}.$$

Як бачимо, норма є узагальненням довжини вектора в дво- і тривимірному просторі. Відстань між двома векторами \mathbf{s} і \mathbf{y} визначається як норма різниці векторів:

$$d(\mathbf{s}, \mathbf{y}) = \|\mathbf{s} - \mathbf{y}\| = \sqrt{\sum_{i=0}^{n-1} (s_i - y_i)^2}.$$

Для евклідового простору R_n можна ввести поняття скалярного добутку двох векторів \mathbf{s} і \mathbf{y} :

$$\mathbf{s}, \mathbf{y} = \sum_{i=0}^{n-1} s_i y_i = \|\mathbf{s}\| \cdot \|\mathbf{y}\| \cos \varphi, \quad (2.39)$$

де φ – кут між двома векторами. Для проєкцій \mathbf{s} на \mathbf{y} і, навпаки, \mathbf{y} на \mathbf{s} , маємо

$$\|\mathbf{s}\| \cos \varphi = \frac{\mathbf{s}, \mathbf{y}}{\|\mathbf{y}\|}, \quad \|\mathbf{y}\| \cos \varphi = \frac{\mathbf{s}, \mathbf{y}}{\|\mathbf{s}\|}.$$

Координати вектора являють собою проєкції вектора на координатні осі, аналогічно (2.36). Зі співвідношення (2.39) випливає очевидна нерівність

$$\|\mathbf{s}, \mathbf{y}\| \leq \|\mathbf{s}\| \cdot \|\mathbf{y}\|, \quad (2.40)$$

відома в літературі як нерівність Буняковського–Шварца. Знак рівності має місце лише тоді, коли

$$\mathbf{y} = k\mathbf{s}, \quad (2.41)$$

тобто коли вектори \mathbf{s} і \mathbf{y} колінеарні (k – скаляр). Для відповідних сигналів $s(t)$ і $y(t)$ це означає, що вони збігаються за формою: $y(t) = k s(t)$.

Квадрат норми вектора \mathbf{s} можна знайти як скалярний добуток вектора на самого себе:

$$\|\mathbf{s}\|^2 = \mathbf{s}, \mathbf{s}.$$

При $n \rightarrow \infty$ простір R_n переходить у нескінченно вимірний простір Гільберта, що позначається L_2 . Гільбертовим простором є, зокрема, простір усіх неперервних комплексних

функцій аргументу t , що задані на інтервалі $\left(-\frac{T}{2}; \frac{T}{2}\right)$, в якому скалярний добуток визначений співвідношенням

$$\langle \mathbf{s}, \mathbf{y} \rangle = \int_{-\frac{T}{2}}^{\frac{T}{2}} \dot{s}(t) y^*(t) dt; \quad (2.42)$$

квадрат норми

$$\|\mathbf{s}\|^2 = \int_{-\frac{T}{2}}^{\frac{T}{2}} \dot{s}(t) s^*(t) dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} |s(t)|^2 dt, \quad (2.43)$$

де \dot{s} – комплексне значення сигналу; y^* та s^* – спряжені з $\dot{s}(t)$ функції (їхні фази відрізняються на 90°).

Норма (2.43) має не тільки геометричний, але і виразний фізичний зміст. Так, якщо сигнал $s(t)$ – дійсний електричний струм в одиничному опорі 1 Ом, то квадрат норми

$$\|\mathbf{s}\|^2 = \int_{-\frac{T}{2}}^{\frac{T}{2}} s^2(t) dt = E_x$$

визначає енергію сигналу. Елементи гільбертового простору L_2 характеризуються квадратом, що інтегрується, тобто якщо елементи цього простору – дійсні сигнали $s(t)$, визначені на інтервалі $\left(-\frac{T}{2}; \frac{T}{2}\right)$, то виконується умова

$$E_x = \int_{-\frac{T}{2}}^{\frac{T}{2}} s^2(t) dt < \infty. \quad (2.44)$$

Гільбертовий простір позначається при цьому $L_2(T)$. При $T \rightarrow \infty$ одержуємо простір $L_2(\infty)$. Для деяких сигналів (функцій) простору $L_2(\infty)$ умова (2.44) при $T \rightarrow \infty$ може не виконуватися, але виконується умова

$$P_x = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s^2(t) dt < \infty. \quad (2.45)$$

У цьому випадку можна замість (2.42) ввести скалярний добуток з розмірністю потужності (для струмів і напруг на одиничному опорі):

$$(\mathbf{s}, \mathbf{y})_P = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \dot{s}(t) s^*(t) dt. \quad (2.46)$$

Квадрат норми вектора \mathbf{s} у цьому випадку

$$\|\mathbf{s}\|_P^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \|\dot{s}(t)\|^2 dt = P_s. \quad (2.47)$$

При виконанні умови (2.45) у просторі $L_2(\infty)$ визначені й співвідношення (2.46) і (2.47) при $T \rightarrow \infty$.

Далі, говорячи про квадратичні інтегровні функції в просторі $L_2(\infty)$, маємо на увазі виконання умов (2.44) чи (2.45) при $T \rightarrow \infty$. Квадрат відстані між двома векторами в просторі $L_2(T)$ визначається співвідношеннями

$$d^2(\mathbf{s}, \mathbf{y}) = \|\mathbf{s} - \mathbf{y}\|^2 = \int_{-\frac{T}{2}}^{\frac{T}{2}} (s(t) - y(t))^2 dt \quad (2.48)$$

або

$$d^2(\mathbf{s}, \mathbf{y}) = \|\mathbf{s} - \mathbf{y}\|_P^2 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} (s(t) - y(t))^2 dt. \quad (2.49)$$

Формулу (2.49) можна використати і при $T \rightarrow \infty$ для сигналів зі скінченною середньою потужністю.

Простір L_2 є природним узагальненням простору R_n , одержаним шляхом переходу від дискретизованої функції до функції неперервного аргументу. В курсі “Теорія електричного зв’язку” простір L_2 має особливе значення, оскільки дозволяє застосувати загальні геометричні зображення до повідомлень, сигналів і завад, визначених як функції неперервного аргументу. Взевши в (2.37) $n \rightarrow \infty$, одержуємо зображення неперервної функції $s(t)$ в просторі Гільберта:

$$s(t) = \sum_{i=0}^{\infty} s_i \Psi_i(t). \quad (2.50)$$

Простір з нескінченним числом вимірів являє собою природне узагальнення n -вимірного простору, що одержується шляхом граничного переходу від дискретної послідовності до функції неперервного аргументу. Зазначимо, що норми векторів сигналів дорівнюють кореням з їх енергії, а скалярний добуток є мірою кореляції цих сигналів.

Сигнали скінченної тривалості T обмежені смугою F , геометрично зображуються різними векторами в n -вимірному просторі. Розбіжність між двома сигналами s_1 та s_2 виражається через відстань між векторами, що зображують їх (рис. 2.13). Відстань залежить від довжин векторів і кута між ними, а косинус останнього є *коефіцієнтом взаємної кореляції сигналів*. Повна відсутність кореляції (тобто рівність нулю коефіцієнта кореляції) виражається через ортогональність векторів $\left(\alpha = \frac{\pi}{2}\right)$.

Завада, обмежена тією ж смугою, що й сигнал, також визначається вектором у n -вимірному просторі. Цей вектор w додається до вектора сигналу (рис. 2.13). На відміну від вектора сигналу, який є апіорно визначеним, вектор завади може мати будь-яку величину і напрямок (вектор є випадковим). У результаті при накладанні завади на сигнал навколо кінця вектора сигналу утвориться “хмара”, змінна щільність якої виражає ймовірність попадання результуючого вектора $x = s + w$ (вектора прийнятого сигналу) в даний елемент множини сигналів. Для флуктуаційної завади ця “хмара” має сферичну форму з ефективним радіусом $r = \sqrt{2TFP_3}$.

Повідомлення (відеосигнал) $u(t)$, яке не містить коливання з частотами, вищими за F_m , так само, як і сигнал, може бути зображене вектором у m -вимірному просторі, де $m = 2TF_m$. Сукупність можливих повідомлень визначає цей простір (простір повідомлень). На рис. 2.14 наведена двовимірна модель цього простору з двома різними повідомленнями u_1 і u_2 . У результаті при накладанні завади на сигнал навколо кінця вектора сигналу утвориться “хмара”, змінна щільність якої виражає ймовірність попадання результуючого вектора $x = s + w$ (вектора прийнятого сигналу) в даний елемент множини сигналів. Для флуктуаційної завади ця “хмара” має сферичну форму з ефективним радіусом $r = \sqrt{2TFP_3}$.

Повідомлення (відеосигнал) $u(t)$, яке не містить коливання з частотами, вищими за F_m , так само, як і сигнал, може бути зображене вектором у m -вимірному просторі, де $m = 2TF_m$. Сукупність можливих повідомлень визначає цей простір (простір повідомлень). На рис. 2.14 наведена двовимірна модель цього простору з двома різними повідомленнями u_1 і u_2 .

Повідомлення $u(t)$ при передаванні перетворюється в сигнал $s(t)$ з використанням деякого переносника $f(t)$. Математично цю операцію формування сигналу можна зобразити у вигляді

$$s = M(u, f), \quad (2.51)$$

де M – оператор, у загальному випадку нелінійний. Геометрично формування сигналу може бути зображене як перетворення простору повідомлень U у простір сигналів S : вектори u_1 і u_2 перетворюються у вектори s_1 і s_2 . Вимірність простору повідомлень m у загальному випадку не дорівнює вимірності простору сигналів n . При односмуговій передачі $m = n$. У випадку амплітудної модуляції сигнали мають вдвічі більше число координат, ніж повідомлення: $n = 2m$, а при частотній модуляції простір сигналів має значно більше число вимірів, ніж простір повідомлень.

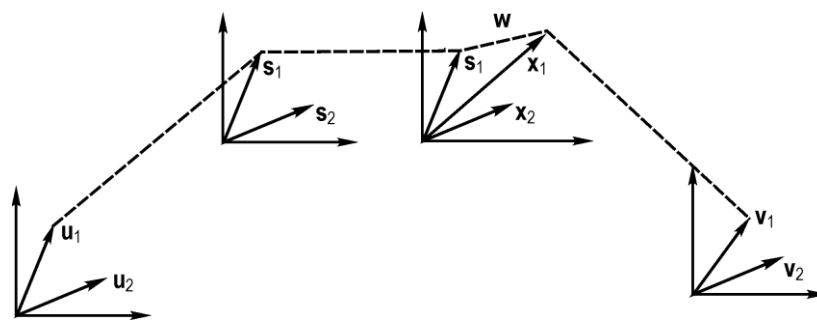


Рис. 2.14. Геометрична модель системи зв'язку: u_1, u_2 – сигнали джерела; s_1, s_2 – сигнали, що передаються по каналу зв'язку; w – завада; x_1, x_2 – сигнали, що є сумою корисного сигналу та завади ($x = s + w$); v_1, v_2 – сигнали отримувача

При накладенні завади на сигнал утворюється область невизначеності, в яку попадають прийняті сигнали $x = s + w$. Взаємодію сигналу і завади можна виразити через оператор Ψ :

$$x = \Psi(s, w). \quad (2.52)$$

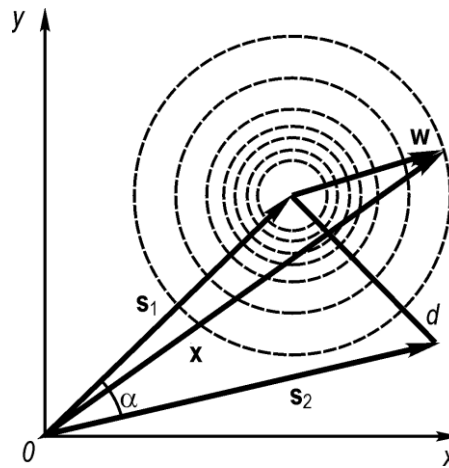


Рис. 2.13. Векторна діаграма сигналу та завади

Оператор Ψ перетворює простір сигналів S у простір прийнятих сигналів X : вектори s_1 і s_2 переходять в x_1 і x_2 .

Приймач за прийнятими сигналами x відтворює передане повідомлення:

$$v = \Phi(x), \quad (2.53)$$

тобто перетворює простір сигналів X у простір прийнятих повідомлень V .

Якщо завади відсутні, то прийнятий сигнал перетвориться в повідомлення оберненим оператором: $v = M^{-1}(s)$. У цьому випадку прийняте повідомлення тотожне переданому: $v = M^{-1}(s) = M^{-1}M = u$.

При наявності завад повідомлення на приймачі відтворюються з деякою помилкою: замість повідомлення u_1 може бути відтворене повідомлення u_2 , і навпаки. Помилка матиме місце, коли результуючий вектор x виявиться ближчим до кінця вектора того сигналу s , що у даний момент не передається. Можна побудувати приймач, що відтворює повідомлення $v = u_1$ кожного разу, коли кінець вектора v_1 ближчий до кінця вектора s_1 , ніж до кінця вектора s_2 , і навпаки. Такий приймач за Котельниковим називається ідеальним чи *оптимальним*. Очевидно, помилка при оптимальному прийманні буде тим меншою, чим більша відстань d між сусідніми сигналами. Ця відстань, у свою чергу, залежить від відстані r між сусідніми повідомленнями та від вигляду оператора M , що визначає спосіб формування сигналу (зокрема, спосіб модуляції).

Оператори і функціонали. У задачах перетворення повідомлень і сигналів нам будуть потрібні деякі узагальнення функціональних залежностей.

Величина y називається *функцією* незалежної змінної x , якщо кожному значенню x (з множини його можливих значень) відповідає певне значення y . Інакше кажучи, функціональна залежність $y = f(x)$ встановлює відповідність між деякою множиною чисел x і множиною чисел y або, те ж саме, функція встановлює залежність одного числа від іншого.

Більш загальним є поняття функціонала. *Функціонал* установлює відповідність між множиною чисел, з одного боку, і деякою множиною функцій – з іншого. Можна сказати, що функціонал Φ встановлює залежність числа від функції: $y = \Phi(f(x))$. Прикладом функціонала є визначений інтеграл, величина якого (при незмінних границях) залежить від вигляду підінтегральної функції.

Функціональний оператор L установлює відповідність між двома множинами функцій, тобто залежність функції від функції: $y(t) = L[x(t)]$. Оскільки функції можуть бути зображені векторами, а множина функцій визначається як векторний простір, то дію оператора можна описати у геометричних термінах як перетворення простору X векторів x у простір Y векторів y . Обернене перетворення Y у X позначають L^{-1} . У задачах перетворення повідомлень і сигналів використовуються поряд з лінійними операторами також нелінійні і параметричні.

Векторне зображення цифрових сигналів у просторі Хеммінга. Якщо функція $x(t)$ на кожному інтервалі $i\Delta$ може набувати одного з m можливих значень $x_i^{(k)}$ ($k = \overline{0, m-1}$), то на відрізку тривалістю T вона буде цілком визначена значеннями $x_i^{(k)}$, або сукупністю коефіцієнтів $x_0^{(k)}, x_1^{(l)}, \dots, x_{n-1}^{(v)}$ ($k, l, v \in \overline{0, m-1}$), яку називають n -набором. Зокрема, при $m=2$ коефіцієнт $x_i^{(k)}$ набуває значення 0 чи 1, а n -набір є просто кодовою комбінацією n -значного двійкового $m=2$ коду, що відображає символ (букву, цифру) переданого повідомлення. Двійкові n -набори відображаються векторами (точками) в кількості 2^n у просторі Хеммінга. Скалярний добуток у цьому просторі зручно задати функцією

$$(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=0}^{n-1} x_k y_k \right]^2,$$

де \sum – сума в звичайному значенні. Звідси норма двійкового вектора

$$\|\mathbf{x}\| = \sum_{k=0}^{n-1} x_k^2.$$

Як бачимо, норма двійкового вектора визначається кількістю одиниць, що містяться в ньому. Цю норму називають також *вагою вектора* (кодової комбінації) і позначають ω . Відстань у просторі Хеммінга

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sum_{k=0}^{n-1} |x_k - y_k| = \sum_{k=0}^{n-1} |x_k \oplus y_k|,$$

де знак \oplus означає операцію підсумовування за модулем 2 (mod 2): $0 \oplus 0 = 0$, $0 \oplus 1 = 1$, $1 \oplus 0 = 1$, $1 \oplus 1 = 0$. Наведемо приклад підсумовування за mod 2 двох векторів:

$$\begin{array}{r} x = 1001011 \\ \oplus y = 0101101 \\ \hline x \oplus y = 1100110 \end{array}.$$

У просторі Хеммінга відстань між двійковими векторами визначається за числом позицій у кодовій комбінації, вектори яких \mathbf{x} і \mathbf{y} мають символи, що відрізняються. У розглянутому прикладі $d(\mathbf{x}, \mathbf{y}) = 4$ одиницям. У більш загальному випадку, якщо число значень, які розрізняються, дорівнює m , використовується різниця за модулем m .

2.5. Розклад сигналів в узагальнений ряд Фур'є

Важливим поняттям у просторах Евкліда, Гільберта і Хеммінга є *ортогональність векторів*. Два вектори \mathbf{x} і \mathbf{y} ортогональні, якщо

$$(\mathbf{x}, \mathbf{y}) = 0. \tag{2.54}$$

Легко показати, що коли вектори Ψ_i і Ψ_j при $j \neq i$ взаємно ортогональні, то вони також лінійно незалежні. Тому сукупність ортогональних векторів можна використовувати як базис лінійних просторів.

Зобразимо неперервну (в часі і за рівнем) квадратичну інтегровну функцію $x(t)$ у $L_2(T)$ -просторі через довільну ортонормовану систему (тобто ортогональні осі нормовані) базисних $\Psi_i(t)$ функцій, для яких

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} \Psi_i(t)\Psi_j(t)dt = \begin{cases} 0, & j \neq i, \\ 1, & j = i. \end{cases} \quad (2.55)$$

Тоді в цій системі сигнал $x(t)$ можна подати у вигляді

$$x(t) = \sum_{i=0}^{\infty} C_i \Psi_i(t), \quad (2.56)$$

де C_i – коефіцієнти (координати) розкладу в ортонормованому базисі $\Psi_i(t)$. Вираз (2.56) називають *узагальненим рядом Фур'є*. Для визначення коефіцієнтів C_i знайдемо скалярний добуток

$$(\mathbf{x}, \Psi_j) = \sum_{i=0}^{\infty} C_i \int_{-\frac{T}{2}}^{\frac{T}{2}} \Psi_i(t)\Psi_j(t)dt.$$

Із урахуванням (2.50)

$$C_i = (\mathbf{x}, \Psi_j) = \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)\Psi_j(t)dt. \quad (2.57)$$

Таким чином, коефіцієнти узагальненого ряду Фур'є C_i є проєкціями вектора \mathbf{x} на ортогональні осі (одичні орти) Ψ_i . Взввши до уваги (2.55) і (2.56), можна одержати

$$E_x = \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t)dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} \left[\sum_{i=0}^{\infty} C_i \Psi_i(t) \right]^2 dt = \sum_{i=0}^{\infty} C_i^2. \quad (2.58)$$

Рівність (2.58) є окремим випадком рівності Парсеваля. З урахуванням ортонормованого базису $\Psi_i(t)$ легко бачити, що скалярний добуток і норму в просторі R_n можна знаходити за формулами (2.42) і (2.43).

Подамо тепер приблизно функцію $x(t)$ розкладу в ряд за ортонормованими базисними функціями $\Psi_i(t)$:

$$x_{\text{e}}(t) = \sum_{i=0}^{\infty} \gamma_i \Psi_i(t) \quad (2.59)$$

і визначимо коефіцієнти γ_i так, щоб мінімізувати середньоквадратичну похибку (СКП):

$$\begin{aligned}\overline{\varepsilon^2(t)} &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) - x_{\varepsilon}(t) \quad ^2 dt = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \left[x(t) - \sum_{i=0}^{n-1} \gamma_i \Psi_i(t) \right]^2 dt = \\ &= \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt - \frac{2}{T} \sum_{i=0}^{n-1} \gamma_i \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) \Psi_i(t) dt + \frac{1}{T} \sum_{i=0}^{n-1} \gamma_i^2,\end{aligned}$$

або

$$\overline{\varepsilon^2(t)} = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt + \frac{1}{T} \sum_{i=0}^{n-1} (\gamma_i - C_i)^2 - \frac{1}{T} \sum_{i=0}^{n-1} C_i^2. \quad (2.60)$$

Похибка $\overline{\varepsilon^2(t)}$ набуває мінімального значення при $\gamma_i = C_i$, тобто коли коефіцієнти розкладу є коефіцієнтами узагальненого ряду Фур'є. Позначивши $\overline{\varepsilon^2(t)} = \Delta_n$, можна записати, виходячи з (2.60),

$$\Delta_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt - \frac{1}{T} \sum_{i=0}^{n-1} C_i^2 \geq 0 \quad (1.61)$$

або

$$\int_{-\frac{T}{2}}^{\frac{T}{2}} x^2(t) dt \geq \sum_{i=0}^{n-1} C_i^2.$$

Нерівність (2.61) називають *нерівністю Бесселя*: зі зростанням n величина Δ_n зменшується. Якщо при $n \rightarrow \infty$ СКП прямує до нуля, то систему базисних функцій $\Psi_i(t)$ називають повною. Маючи на увазі, що при $n \rightarrow \infty$ є слушним вираз (2.58), можна стверджувати, що в просторі Гільберта система базисних функцій $\Psi_i(t)$ є повною. Ця система функцій є також замкнутою, тому що для будь-якої функції $x(t)$ з $L_2(T)$ нерівність (2.61) переходить при $n \rightarrow \infty$ в рівність.

2.6. Аналітичний сигнал

Реальні процеси описуються дійсними функціями часу, проте, розв'язуючи велику кількість задач, зручно виражати сигнал або у вигляді суми елементарних сигналів, кожний з яких є комплексною функцією часу, або як комплексну функцію:

$$\hat{s}(t) = s(t) + i\hat{s}(t) = U(t)e^{i\psi(t)}, \quad (2.62)$$

де $U(t)$ і $\psi(t)$ – обвідна і фаза сигналу відповідно.

Дійсний сигнал у цьому випадку

$$s(t) = \operatorname{Re} U(t)e^{i\psi(t)} = U(t) \cos \Psi(t). \quad (2.63)$$

Таке зображення особливо зручне для вузькосмугових сигналів.

Сигнал $s(t)$ називається *аналітичним*, якщо $s(t)$ і $\hat{s}(t)$ складають пари перетворень Гільберта:

$$\left. \begin{aligned} \hat{s}(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau; \\ s(t) &= -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\hat{s}(\tau)}{t-\tau} d\tau. \end{aligned} \right\} \quad (2.64)$$

Інтеграли в цих співвідношеннях є інтегралами Коші. Функція $\hat{s}(t)$ називається спряженою з функцією $s(t)$ за Гільбертом. При такому виборі $s(t)$ і $\hat{s}(t)$ обвідна і фаза сигналу визначаються однозначно:

$$U(t) = \sqrt{[s(t)^2 + \hat{s}(t)^2]}, \quad (2.65)$$

$$\psi(t) = \operatorname{arctg} \frac{\hat{s}(t)}{s(t)}. \quad (2.66)$$

Якщо ефективна ширина спектра сигналу $s(t)$ мала в порівнянні з його середньою частотою ω_0 , то $U(t)$ і $\psi(t)$ змінюються повільно в порівнянні з функцією $s(t)$. Легко показати, що функції $s(t) = \cos \omega t$ відповідає спряжена функція $\hat{s}(t) = \sin \omega t$, а функції $s(t) = \sin \omega t$ відповідає $\hat{s}(t) = -\cos \omega t$. Якщо вихідний сигнал зображений рядом Фур'є $s(t) = \sum_k (a_k \cos \omega_k t + b_k \sin \omega_k t)$, то спряжена до нього функція буде $\hat{s}(t) = \sum_k (a_k \sin \omega_k t - b_k \cos \omega_k t)$. Таким чином, найпростішому сигналові у вигляді гармонічного коливання $s(t) = A \cos \omega t$ відповідає аналітичний сигнал $s(t) = A \cos \omega t + iA \sin \omega t = A e^{i\omega t}$. Якщо сигнал $s(t)$ може бути зображений у вигляді інтеграла Фур'є $s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(i\omega) e^{i\omega t} d\omega$, то частотний спектр сигналу

$$F[\hat{s}(t)] = [-i \operatorname{sgn}(\omega) S(i\omega)], \quad (2.67)$$

де

$$\operatorname{sgn}(\omega) = \begin{cases} +1, & \omega > 0, \\ 0, & \omega = 0, \\ -1, & \omega < 0. \end{cases}$$

Отже, перетворення сигналу $s(t)$ за Гільбертом можна розглядати як результат проходження $s(t)$ через електричне коло, що зсуває фазу всіх частотних і перехідних компонент спектра на кут -90° . Частотна і перехідна характеристики цього ланцюга такі:

$$K(i\omega) = -i \operatorname{sgn}(\omega), \quad h(t) = \frac{1}{\pi t}.$$

Підстановка виразу (2.67) у перетворення Фур'є функції (2.62) показує, що спектр $S(i\omega)$ сигналу $s(t)$ є “однобічним”:

$$S(i\omega) = \begin{cases} 2S(i\omega), & \omega > 0, \\ S(0), & \omega = 0, \\ 0, & \omega < 0. \end{cases} \quad (2.68)$$

Це дуже корисна властивість аналітичного сигналу, оскільки в системах зв'язку найважливішим є економічне використання смуги каналу зв'язку.

Для періодичних функцій $s(t)$ спряжена за Гільбертом функція $\hat{s}(t)$ також є періодичною з тим же самим періодом. Ці функції на відрізку, що дорівнює їх періоду T , взаємно ортогональні, тобто $\int_0^T s(t)\hat{s}(t) dt = 0$.

Сигнали $s_i(f)$ і $s_j(f)$, для яких умова ортогональності зберігається із заміною одного з них на спряжений за Гільбертом, тобто коли виконуються умови

$$\left. \begin{aligned} (\mathbf{s}_i, \mathbf{s}_j) &= \frac{1}{T} \int_{-T/2}^{T/2} s_i(t)s_j(t)dt = 0, \\ (\mathbf{s}_i, \hat{\mathbf{s}}_j) &= \frac{1}{T} \int_{-T/2}^{T/2} s_i(t)\hat{s}_j(t)dt = 0, \quad i \neq j, \end{aligned} \right\} \quad (2.69)$$

називаються *ортогональними в підсиленому сенсі*. Можна показати, що такі сигнали також задовольняють й умову

$$(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{T} \int_{-T/2}^{T/2} s_i(t)s_j^*(t)dt = 0, \quad i \neq j, \quad (2.70)$$

де “зірочка” позначає комплексно-спряжену величину.

Введення поняття “аналітичний сигнал” дозволяє зобразити кожен функцію часу (із несуттєвими обмеженнями) у комплексному вигляді та однозначно визначити обвідну і фазу сигналу. В аналітичній формі (2.67) можуть бути подані як детерміновані, так і випадкові сигнали. Корисність такого зображення полягає в можливості окремого розгляду обвідної і фази процесу (чи сигналу завади). Так, при дослідженні випадкового процесу можна замість вивчення миттєвих значень обмежитися вивченням обвідної та фази процесу, а в багатьох випадках – або обвідної, або фази.

У загальному випадку спектри і кореляційні функції процесів $x(t)$ і $\hat{x}(t)$ однакові: $G_x(\omega) = G_{\hat{x}}(\omega)$, $B_x(\tau) = B_{\hat{x}}(\tau)$.

Взаємний енергетичний спектр дорівнює $G_{\hat{x}x}(\omega) = iG_x(\omega)$, а взаємна кореляційна функція

$B_{\hat{x}x}(\tau) = -B_{x\hat{x}}(\tau) = \frac{1}{\pi} \int_0^{\infty} G_x(\omega) \sin \omega \tau d\omega$. Між законом розподілу випадкового процесу $x(t)$, з одного боку,

і законами розподілу обвідної U і фази ψ , з іншого, існує жорсткий зв'язок: знаючи закон розподілу $p(x)$, можна знайти закони розподілу $p(U)$ і $p(\psi)$. Порівняно легко цей зв'язок встановлюється у випадку гауссових процесів.

Контрольні запитання для самооцінки рівня знань

1. Дайте означення детермінованого сигналу.
2. Які види детермінованих сигналів Ви знаєте?
3. Назвіть види еталонних сигналів.
4. Які види детермінованих сигналів можна розкласти в ряд Фур'є?
5. Який спектр має періодичний сигнал?
6. Запишіть ряд Фур'є в звичайній та комплексній формі і визначте значення коефіцієнтів ряду.
7. Напишіть пряме та обернене перетворення Фур'є для неперіодичного сигналу.
8. Як визначається середня потужність періодичного сигналу?
9. Що розуміється під практичною шириною спектра сигналу?
10. Які дискретизації сигналів за часом Ви знаєте?
11. Сформулюйте теорему В. О. Котельникова.
12. Запишіть ряд В. О. Котельникова.
13. В чому сутність теореми В. О. Котельникова для сигналів, спектр яких не розпочинається з нульової частоти?
14. У чому особливості кореляційного критерію відліків М. А. Желєзнова?
15. Як оцінюється похибка дискретизації, коли не враховуються частоти, що лежать вище граничної частоти спектра сигналу?
16. Які особливості має квантування критерію відліків Темникова?
17. У чому полягають особливості адаптивної дискретизації сигналів за часом?
18. У чому полягає суть квантування за рівнем сигналів?
19. Як визначається похибка квантування?
20. Дайте означення функціональних просторів за Евклідом, Гільбертом і Хеммінгом.
21. Дайте означення нормованого простору.
22. Як визначається відстань між двома векторами?
23. Що таке "хмара" завад при геометричному зображенні сигналу?
24. У чому полягають особливості зображення цифрових сигналів у просторі Хеммінга?
25. Запишіть узагальнений ряд Фур'є.
26. Дайте означення аналітичного сигналу.

Післямова до розділу 2

Вивчення матеріалу другого розділу навчального посібника вимагало освоєння спеціальної термінології.

Тепер ви вже знаєте, що:

- для побудови матеріальних моделей сигналів розглядають два їхні представлення: часове і частотне;
- математичні моделі часового і частотного представлень сигналів пов'язані між собою парою перетворень Фур'є ;
- класифікацію сигналів;
- для часового представлення неперервного випадкового сигналу можна формувати як пряму, так і ймовірну математичну модель;
- частотне представлення сигналу називають спектром сигналу;
- спектральні характеристики детермінованих періодичних і детермінованих і неперіодичних сигналів мають характерну відмінність;
- частотне представлення детермінованих неперіодичних і неперервних випадкових сигналів є однотипними.

Ви вивчили чудову теорему В. О. Котельникова, що лежить в основі дискретизації неперервних сигналів: функція $S(t)$, яка не містить частот, вищих від F , повністю визначається послідовністю своїх значень в моменти, які відстають один від одного на $\Delta t \leq \frac{1}{2F} = \frac{\pi}{\omega}$.

Пам'ятаємо, що поява сигналу обумовлена появою відповідного повідомлення, котре містить певну кількість інформації.

Отримані Вами знання є частиною «теорії сигналів», які необхідно розширювати і поглиблювати.

У наступному 3 розділі посібника подається спектр питань, які стосуються сукупності засобів, призначених для передавання сигналів (повідомлень) між різними точками (пунктами) системи зв'язку - каналів зв'язку.

Розділ 3

КАНАЛИ ЗВ'ЯЗКУ

3.1 Загальні відомості про канали зв'язку

Канал зв'язку – це сукупність засобів, призначених для передавання сигналів (повідомлень) між різними точками (пунктами) системи зв'язку. Під засобом розуміють і *технічні пристрої*, і *лінію зв'язку* – фізичне середовище, в якому поширюється сигнал між пунктами зв'язку. Канал зв'язку можна уявити як послідовне з'єднання пристроїв (блоків), що виконують різні функції в загальній системі зв'язку.

Класифікація каналів зв'язку можлива за різними ознаками. *За призначенням систем* канали зв'язку поділяють на телеграфні, фототелеграфні, телефонні, звукового мовлення, передавання даних, телевізійні, телеметричні, змішані і т. ін. Залежно від того, поширюються сигнали між пунктами зв'язку у *вільному просторі* чи по *напрямних лініях*, виділяють канали радіозв'язку (зокрема, космічні канали) і проводового зв'язку (повітряні, кабельні, волоконно-оптичні лінії зв'язку, хвилевідні надвисокочастотні тракти тощо).

Розрізняють канали чисто часові (із зосередженими параметрами), в яких сигнали на вході і виході описуються функціями одного скалярного параметра (часу t), і просторово-часові канали (з розподіленими параметрами), в яких сигнали на вході і (чи) виході описуються функціями не одного, а кількох скалярних параметрів (наприклад, часу t і просторових координат x, y, z). Такі сигнали називають *полями*.

Більш істотною є класифікація каналів електричного зв'язку *за діапазоном використовуваних ними частот*. Так, на сучасних симетричних кабельних лініях зв'язку застосовують сигнали, що займають смуги частот у діапазоні, обмеженому зверху частотою в кілька сотень кілогерців. Додаткові заходи щодо збільшення симетрії кабельних пар дозволяють збільшити верхню межу використовуваного діапазону частот до тисячі кілогерців. Коаксіальні кабелі, що є основою мереж магістрального далекого зв'язку, пропускають у даний час діапазон частот до сотень мегагерців. На повітряних провідних лініях використовують частоти не вищі за 150 кГц, оскільки на більших частотах у цих лініях сильно позначається дія адитивних завад і різко зростає загасання в лінії.

Радіозв'язок здійснюється за допомогою електромагнітних хвиль, що поширюються в частково обмеженому (наприклад, землею й іоносферою) просторі. В даний час у радіозв'язку застосовують частоти приблизно від $3 \cdot 10^3$ до $3 \cdot 10^{12}$ Гц. Цей діапазон поділяють відповідно до десяткової класифікації (табл. 3.1). У таблиці в дужках зазначені нестандартні, але використовувані назви діапазонів хвиль.

Таблиця 3.1. Класифікація діапазону частот

Найменування хвиль	Діапазон хвиль	Найменування частот	Діапазон частот
Декакілометрові (наддовгі)	100...10 км	Дуже низькі	3...30 кГц
Кілометрові (довгі)	10...1 км	Низькі	30...300 кГц
Гектаметрові (середні)	1000...100 м	Середні	300...3000 кГц
Декаметрові (короткі)	100...10 м	Високі	3...30 МГц
Метрові (ультракороткі)	10...1 м	Дуже високі	30...300 МГц
Дециметрові	100...10 см	Ультрависокі	300...3000 МГц

Сантиметрові	10...1 см	Надвисокі	3...30 ГГц
Міліметрові	10...1 мм	Надзвичайно високі	30...300 ГГц
Дециміліметрові	1...0,1 мм	Гіпервисокі	300...3000 ГГц

Діапазон дециміліметрових хвиль уже впритул підходить до діапазону інфрачервоних хвиль. В даний час завдяки створенню і широкому впровадженню квантових генераторів (лазерів) освоєно і діапазон світлових хвиль (оптичний діапазон). Практично у волоконно-оптичних лініях зв'язку використовуються частоти порядку 10^{14} Гц (довжини хвиль 1,55; 1,35; 0,85 мкм). Для сучасного етапу розвитку техніки зв'язку характерна тенденція до переходу на все більш високі частоти. Це пояснюється низкою чинників, зокрема необхідністю підвищувати швидкість передавання повідомлень, можливістю одержати гостронапрявлене випромінювання при невеликих розмірах випромінювачів, меншою інтенсивністю атмосферних і багатьох видів промислових завод у більш високочастотних діапазонах, можливістю застосування заводостійких широкосмугових систем модуляції тощо.

У теорії електричного зв'язку дуже важливою є класифікація каналів зв'язку *за характером сигналів на вході і виході каналу*. Розрізняють канали: дискретні (за рівнями), на вході і виході яких сигнали дискретні; неперервні (за рівнями), на вході і виході яких сигнали неперервні (наприклад, канал, заданий між виходом модулятора і входом демодулятора в будь-якій системі зв'язку); дискретні з боку входу і неперервні з боку виходу, чи навпаки. Такі канали називаються дискретно-неперервними, або напівнеперервними.

Усякий дискретний або напівнеперервний канал містить у собі неперервний канал. Слід пам'ятати, що дискретність і неперервність каналу не пов'язані з характером переданих повідомлень. Можна передати дискретні повідомлення по неперервному каналу і неперервні повідомлення – по дискретному.

За видом статичних характеристик елементів канали поділяють на два види: лінійні, що описуються лінійними диференціальними рівняннями; нелінійні канали, але описувані лінійними рівняннями при малих відхиленнях від рівноваги. Диференціальні рівняння таких каналів є нелінійні, але їх можна лінеаризувати.

Канали, що містять елементи з істотно нелінійними статичними характеристиками (наприклад, зона нечутливості, насичення), досліджуються методами, придатними для нелінійних систем.

3.2. МАТЕМАТИЧНІ МОДЕЛІ КАНАЛІВ ЗВ'ЯЗКУ

Для того щоб дати математичний опис каналу, необхідно та достатньо вказати множину сигналів, що можуть бути подані на його вхід, і для будь-якого вхідного сигналу з цієї множини задати випадковий процес (сигнал) на його виході, тобто задати в тій чи іншій формі розподіл імовірностей.

Точний математичний опис будь-якого реального каналу звичайно є дуже складним. Замість цього використовують спрощені математичні моделі, що дозволяють виявити всі найважливіші закономірності реального каналу, якщо при побудові моделі враховані найбільш істотні особливості каналу і відкинуті другорядні деталі, що мало впливають на процес зв'язку.

Розглянемо найпростіші і широко використовувані математичні моделі каналів, почавши з неперервних каналів, оскільки вони багато в чому визначають і характер дискретних каналів.

Ідеальний канал без завод – це лінійний ланцюг з постійною передавальною функцією, звичайно зосередженою в обмеженій смузі частот. Допустимими є будь-які вхідні сигнали, спектр яких лежить у визначеній смузі частот F , що мають обмежену середню потужність P (або пікову потужність $P_{\text{пик}}$). Оскільки ці обмеження характерні для всіх неперервних каналів, то надалі про них не говоритиметься. Зауважимо, що коли потужність сигналу не обмежувати, то множина допустимих сигналів утворить векторний простір – скінченновимірний (при певних обмеженнях на тривалість і ширину спектра) або нескінченновимірний (при більш слабких обмеженнях). В ідеальному каналі вихідний сигнал при заданому вхідному детермінований. Цю модель іноді

використовують для опису кабельних каналів. Однак, строго кажучи, вона непридатна для реальних каналів, у яких неминуче присутні, хоча б і дуже слабкі, адитивні завади.

Канал з адитивним гауссовим шумом. Сигнал на виході такого каналу

$$z(t) = Ku(t - \tau) + N(t), \quad (3.1)$$

де $u(t)$ – вхідний сигнал; K – коефіцієнт підсилення або послаблення сигналу (K і t – сталі величини); $N(t)$ – гауссів адитивний шум з нульовим математичним сподіванням і заданою кореляційною функцією. Найчастіше розглядається білий шум або квазібілий – з рівномірною спектральною щільністю в смузі спектра сигналу $u(t)$.

Запізнювання τ відповідає зміні початку відліку часу на виході каналу. Дещо ця модель ускладниться, якщо коефіцієнт передачі K і запізнювання τ вважати відомими функціями часу:

$$z(t) = K(t)u[t - \tau(t)] + N(t). \quad (3.2)$$

Така модель задовільно описує багато провідних каналів, радіоканали при зв'язку в межах прямої видимості, а також радіоканали з повільними загальними завмираннями, при яких можна надійно завбачати значення K і τ .

Канал з невизначеною фазою сигналу відрізняється від попереднього тим, що в ньому запізнювання є випадковою величиною. Для вузькосмугових сигналів вираз (3.2) при сталому K і випадкових $\tau(t)$ можна подати так:

$$z(t) = K[u(t) \cos \theta_k + \tilde{u}(t) \sin \theta_k] + N(t), \quad (3.3)$$

де $\tilde{u}(t)$ – перетворення Гільберта від $u(t)$; $\theta_k = \omega_0 \tau$ – випадкова початкова фаза.

Розподіл імовірностей θ_k вважається заданим, найчастіше рівномірним на інтервалі від 0 до 2π . Ця модель задовільно описує ті ж канали, що і попередня, якщо фаза сигналу в них флюктує. Така флюктуація спричинюється невеликими змінами довжини каналу, властивостей середовища, в якому проходить сигнал, а також фазовою нестабільністю опорних генераторів.

Однопроменевий гауссів канал із загальними завмираннями (флюктуаціями амплітуд і фаз сигналу) також описується формулою (3.3), але і множник K , і фаза θ_k вважаються випадковими процесами. Іншими словами, випадковими будуть квадратурні компоненти

$$X = K \cos \theta_k; \quad Y = K \sin \theta_k. \quad (3.4)$$

При зміні у часі квадратурних компонентів $X(t)$, $Y(t)$ прийняте колювання

$$\begin{aligned} z(t) &= x(t)u(t) + y(t)\tilde{u}(t) + N(t) = \\ &= K(t)[u(t) \cos \theta_k(t) + \tilde{u}(t) \sin \theta_k(t)] + N(t). \end{aligned} \quad (3.5)$$

Як зазначалося, одновимірний розподіл коефіцієнта передачі $K(t)$ може бути релеївським чи узагальненим релеївським. Такі канали називають відповідно *каналами з релеївськими* чи *узагальненими релеївськими завмираннями*. У рамках загальної гауссової моделі каналу $K(t)$ має чотирипараметричний розподіл. Моделлю однопроменевого каналу з завмираннями досить добре описуються багато каналів радіозв'язку в різних діапазонах хвиль.

Канал з міжсимвольною інтерференцією (МСІ) і адитивним шумом. Ця модель є окремим випадком (3.1), коли $G(t, \tau)$ від t не залежить (чи змінюється дуже повільно), а тому розсіювання за частотою практично не спостерігається.

Міжсимвольна інтерференція пояснюється розсіюванням сигналу в часі при його проходженні по каналу зв'язку. Вона полягає в тому, що на виході каналу сигнал деформується через наявність відкликів каналу на відрізки вхідного сигналу, які належать до досить віддалених моментів часу. При передаванні дискретних повідомлень це призводить до того, що у разі приймання одного символу на вхід приймального пристрою впливають також відклики на більш ранні (а іноді й більш пізні) символи, котрі діють як завади.

Міжсимвольна інтерференція безпосередньо спричинюється нелінійністю фазочастотної характеристики каналу й обмеженістю його смуги пропускання. У радіоканалах причиною МСІ найчастіше є багатопроменеве поширення радіохвиль.

Нехай синхронно з тактовим інтервалом T передається послідовність елементарних сигналів, що відповідають ланцюжку символів

$$b_{-Q}, b_{-(Q-1)}, \dots, b_{-2}, b_{-1}, b_0, b_1, b_2, \dots, b_{Q-1}, b_Q, \quad (3.6)$$

причому кожний із символів послідовності вибирають з можливого для даного коду набору $0, 1, \dots, m-1$ (m – основа коду).

Позначимо відклик лінійного каналу на елементарний сигнал, що відповідає символу b_r , через $s_r(t)$, $0 \leq t \leq Q+1 T$, де $Q = \left\lceil \frac{\Delta\tau}{T} \right\rceil$ – відносна пам'ять каналу, яка визначається цілою частиною від ділення часу розсіювання каналу $\Delta\tau$ (тривалості перехідного процесу в каналі) на T . При використанні двійкових протилежних сигналів і постійних параметрах каналу $s_r(t) = a_r s_0(t)$, де $s_0(t)$ – відклик каналу на елементарний сигнал, що відповідає символу 1, $a_r = \pm 1$. Тоді прийняте коливання $z(t)$ у місці приймання на інтервалі аналізу $T_a = D+1 T^*$ при пошуку розв'язку щодо символу b_0 можна записати так:

$$z(t) = \sum_{r=-Q}^D s_r(t-rT) + n(t) = s_0(t) + g_{M,i}(t) + n(t), \quad (3.7)$$

де $s_0(t)$ – сигнал, обумовлений аналізованим символом b_0 ; $g_{M,i}(t) = g_{\text{зал}}(t) + g_{\text{наст}}(t) = \sum_{r=-Q, r \neq 0}^D s_r(t-rT)$ – сигнал міжсимвольної інтерференції, обумовлений символами, переданими до і після аналізованого символу; $n(t)$ – адитивний шум у каналі; $g_{\text{зал}}(t) = \sum_{r=-Q}^D s_r(t-rT)$ – залишковий сигнал

МСІ, обумовлений символами, переданими до аналізованого; $g_{\text{наст}}(t) = \sum_{r=1}^D s_r(t-rT)$ – сигнал МСІ,

обумовлений символами, переданими після аналізованого. Чим більша швидкість передавання символів $1/T$ у кожному частотному каналі при заданій його смузі пропускання, тим більшу кількість сусідніх з аналізованим символів визначає сигнал $g_{M,i}(t)$. У деяких випадках у моделі (3.7) можна вважати, що елементарні сигнали на прийомі $s_r(t)$ і передачі $u_r(t)$ пов'язані детермінованим (як правило, лінійним) відношенням. Тоді при незначному рівні шумів $n(t)$ у каналі можна здійснити його корекцію, тобто перейти до моделі каналу, що не спотворює. Проте при значних рівнях шумів у каналі з МСІ граничну якість можна забезпечити лише оптимальним прийманням. У разі випадкових змін параметрів каналу функції $s_r(t)$ або $G(t, \tau)$ стають випадковими і модель (3.7) ускладнюється.

Моделі дискретного каналу. Як відомо, дискретний канал завжди містить неперервний канал, а також модем. Останній можна розглядати як пристрій, що перетворює неперервний канал у дискретний, а тому математичну модель дискретного каналу можна вивести з моделей неперервного каналу і модему. Такий підхід часто є плідним, але він призводить до ускладнення моделей. Розглянемо прості моделі дискретного каналу, побудовані без урахування властивостей неперервного каналу і модему. Однак слід пам'ятати: проектуючи систему зв'язку, можна варіювати в досить широких межах модель дискретного каналу при заданій моделі неперервного каналу зміною модему.

Модель дискретного каналу містить задання множини можливих сигналів на його вході і розподіл умовних імовірностей вихідного сигналу при заданому вхідному. Тут вхідним і вихідним сигналами є послідовності кодових символів. Тому для визначення можливих вхідних сигналів достатньо указати число m різних символів (основу коду), а також тривалість T передавання кожного символу. Вважатимемо значення T однаковим для всіх символів (ця умова виконується в більшості сучасних каналів). Величина $\nu = 1/T$ визначає кількість символів, переданих в одиницю часу, тобто *технічну швидкість*, і вимірюється в бодах. Кожен символ, що надійшов на вхід каналу, викликає появу одного символу на виході, тому технічна швидкість на вході і виході каналу однакова. В реальних каналах це не завжди так, оскільки з порушенням тактової

* При поелементному прийманні величина D визначає затримку (виражену в числі символів) прийняття рішення про переданий символ. Із збільшенням D якість зв'язку при оптимальному прийманні покращується. Звичайно вибирають $D \leq Q$.

синхронізації модему число символів на виході каналу може бути більшим чи меншим, ніж на ході. У даному курсі ця обставина не враховується і синхронізація вважається ідеальною. Методи забезпечення синхронізації вивчаються в спеціальних курсах.

У загальному випадку для будь-якого n має бути указана ймовірність того, що у разі подання на вхід каналу будь-якої послідовності \mathbf{b}^n кодів символів на виході з'явиться деяка реалізація випадкової послідовності $\hat{\mathbf{V}}^n$. Кодові символи позначимо символами від 0 до $m-1$, що дозволить робити над ними арифметичні операції. При цьому всі n -послідовності (вектори), число яких дорівнює m^n , утворять m^n -вимірний скінченний векторний простір, якщо “додавання” розуміти як підсумовування розрядами за модулем m .

Уведемо ще одне означення. Будемо називати *вектором помилки* різницю розрядів (зрозуміло, за модулем m) між прийнятим і переданим векторами. Це означає, що проходження дискретного сигналу через канал можна розглядати як додавання вхідного вектора з вектором помилки. Вектор помилки грає в дискретному каналі приблизно ту ж саму роль, що і завада в неперервному каналі. Отже, для будь-якої моделі дискретного каналу можна записати, користуючись підсумовуванням у векторному просторі (розрядами, за модулем m):

$$\hat{\mathbf{V}}^{[n]} = \mathbf{V}^{[n]} + \mathbf{E}^{[n]}, \quad (3.8)$$

де $\hat{\mathbf{V}}^{[n]}$ і $\mathbf{V}^{[n]}$ – випадкові послідовності з n символів на вході і виході каналу; $\mathbf{E}^{[n]}$ – випадковий вектор помилки, що у загальному випадку залежить від $\mathbf{V}^{[n]}$. Різні моделі відрізняються розподілом ймовірностей вектора $\mathbf{E}^{[n]}$. Зміст вектора помилки особливо простий у випадку двійкових каналів ($m=2$), коли його компоненти набувають значень 0 і 1. Всяка одиниця у векторі помилок означає, що у відповідному місці переданої послідовності символ прийнятий помилково, а всякий нуль означає безпомилкове приймання символу. Число ненульових символів у векторі помилок називається його *вагою*. Образно кажучи, модем, що здійснює перехід від неперервного каналу до дискретного, перетворить завади і спотворення неперервного каналу в течію помилок. Розглянемо найбільш важливі і досить прості моделі дискретних каналів.

Симетричний канал без пам'яті – це дискретний канал, у якому кожен переданий кодовий символ може бути прийнятий помилково з фіксованою ймовірністю p і правильно з ймовірністю $1-p$, причому в разі помилки замість переданого символу b може бути з рівною ймовірністю прийнятий будь-який інший символ.

Отже, ймовірність того, що прийнято символ \hat{b}_j , якщо був переданий символ b_i , дорівнює

$$p(\hat{b}_j/b_i) = \begin{cases} p/(m-1) & \text{при } i \neq j, \\ 1-p & \text{при } i = j. \end{cases} \quad (3.9)$$

Термін “без пам'яті” означає, що ймовірність помилкового приймання символу не залежить від передісторії, тобто від того, які символи передавалися до нього і як вони були прийняті. Надалі, для скорочення, замість “ймовірність помилкового приймання символу” говоритимемо “ймовірність помилки”.

Очевидно, що ймовірність будь-якого n -вимірного вектора помилки в такому каналі

$$p(\mathbf{E}^{[n]}) = [p/(m-1)]^l (1-p)^{n-l}, \quad (3.10)$$

де l – число ненульових символів у векторі помилки (вага вектора помилки). У двійковому каналі вага вектора збігається з його нормою.

Ймовірність того, що відбулося l будь-яких помилок, розташованих як завгодно на послідовності завдовжки n , визначається формулою Бернуллі

$$p(l) = C_n^l \left(\frac{p}{m-1} \right)^l (1-p)^{n-l}, \quad (3.11)$$

де $C_n^l = \frac{n!}{l!(n-l)!}$ – біномний коефіцієнт, який дорівнює числу різних сполучень l помилок у

блоці завдовжки n .

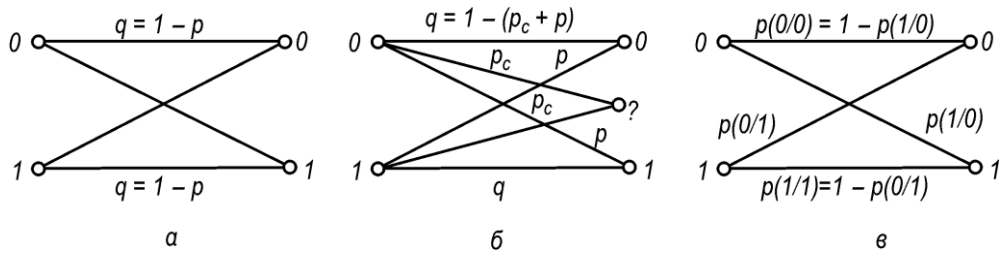


Рис.3.1. Переходи ймовірності в двійковому каналі:

a – симетричному; b – симетричному зі стиранням; c – несиметричному

Цю модель називають також *біномним каналом*. Вона задовільно описує канал, що виникає при певному виборі модему, якщо в неперервному каналі відсутні завмирання, а адитивний шум білий (чи, принаймні, квазібілий). Імовірності переходів у двійковому симетричному каналі схематично показані у вигляді графа на рис. 3.1, a .

Симетричний канал без пам'яті зі стиранням помилкового символу відрізняється від попереднього тим, що алфавіт на виході каналу містить додатковий $m+1$ -й символ, який позначається знаком “?”. Цей символ з'являється тоді, коли перша розв'язувальна схема (демодулятор) не може надійно розпізнати переданий символ. Імовірність такої відмови від рішення чи стирання символу p_c у даній моделі постійна і не залежить від переданого символу. Процедурою стирання вдається значно знизити ймовірність помилки, іноді її навіть вважають нульовою. На рис. 3.1, b схематично показані ймовірності переходів у такій моделі.

Несиметричний канал без пам'яті характеризується, як і попередні моделі, тим, що помилки виникають у ньому незалежно одна від одної, однак ймовірності помилок залежать від того, який символ передається. Так, у двійковому несиметричному каналі ймовірність $p(1/0)$ приймання символу “1” при передаванні символу “0” не дорівнює ймовірності $p(1/0)$ приймання “0” при передаванні “1” (рис. 3.1, c). У цій моделі ймовірність вектора помилки залежить від того, яка послідовність символів передається.

Марковський канал являє собою найпростішу модель дискретного каналу з пам'яттю, в якій ймовірність помилки утворює простий ланцюг Маркова, тобто залежить від того, правильно чи помилково прийнято попередній символ, але не залежить від того, який символ передається. Такий канал, наприклад, виникає, якщо в неперервному каналі з гауссовим шумом використовується відносна фазова модуляція.

Канал з адитивним дискретним шумом є узагальненням моделей симетричних каналів. У такій моделі ймовірність вектора помилки E^n не залежить від переданої послідовності, вважається заданою і не визначається його вагою. В багатьох каналах із двох векторів з однаковою вагою більш імовірним є такий, в якому одиниці розташовані близько одна до одної, тобто існує тенденція до групування помилок.

3.3. Метод простору станів

Розглянемо інший підхід до побудови математичних моделей каналів. Вище співвідношення між вхідним і вихідним сигналами задавалися інтегральними перетвореннями (наприклад, інтегралом Дюамеля). При цьому для пошуку вихідного сигналу потрібно знати крім характеристик ланцюга (каналу) також вхідний сигнал, що діяв на нього на всьому проміжку його існування, до поточного моменту t .

У багатьох випадках більш гнучким є такий опис, при якому вся передісторія до деякого фіксованого моменту часу t_0 замінюється заданням деякого початкового стану ланцюга. Знаючи

характеристики ланцюга, початковий стан і сигнал, що діє тільки на проміжку від t_0 до t_1 , можна послідовно визначити як сигнал на виході, так і новий стан ланцюга в будь-який момент часу $t > t_0$.

Подібний підхід відомий з теорії диференціальних рівнянь, у якій шукана функція визначається як самим рівнянням, так і початковими умовами, число яких дорівнює порядку рівняння. Викладений в даному параграфі метод змінних стану ілюструється на прикладах ланцюгів, описуваних за допомогою лінійних диференціальних рівнянь.

Основні поняття. Всі змінні величини, що характеризують динамічну систему (рис. 3.2, а), доцільно розділити на три групи: 1) вхідні змінні чи вхідні впливи u_i ($i=1, 2, \dots, m$), які є сигналами, що генеруються системами, зовнішніми щодо досліджуваної системи, і впливають на її поведінку; 2) вихідні y_j ($j=1, 2, \dots, m$) змінні, які характеризують реакцію системи і дозволяють описати деякі сторони поведінки системи, що представляють інтерес для дослідження; 3) змінні (координати) стану чи проміжні змінні x_k ($k=1, 2, \dots, n$), що характеризують динаміку досліджуваної системи.

Величини u_i , y_j , x_k вважаються функціями часу $u_i(t)$, $y_j(t)$, $x_k(t)$ і відповідають значенням величин u_i , y_j , x_k у момент часу t . Множина всіх значень, що може набути вектор входу $\mathbf{U}(t_0, t) = [\mathbf{u}_1(t_0, t), \mathbf{u}_2(t_0, t), \dots, \mathbf{u}_m(t_0, t)]$ в момент часу t , утворить *простір входу* системи. Множина всіх значень, що може набути вектор виходу $\mathbf{Y}(t) = [y_1(t), y_2(t), \dots, y_m(t)]$ в момент часу t , утворить *простір виходу*. Множина всіх значень, що може набути вектор стану $\mathbf{X}(t) = [x_1(t), x_2(t), \dots, x_n(t)]$ в момент часу t , утворить *простір станів* системи. В будь-який момент часу t стан системи визначається вектором $\mathbf{X}(t)$, що є функцією вектора початкового стану $\mathbf{X}(t_0)$ і вектора входу $\mathbf{U}(t_0, t)$, тобто

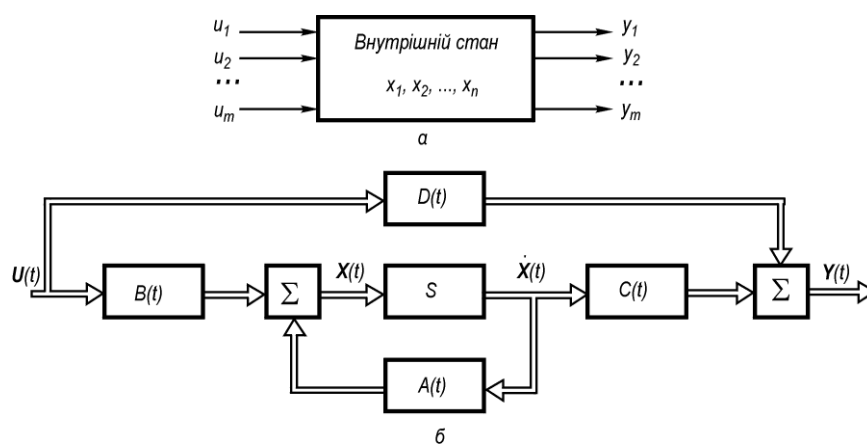


Рис. 3.2. Величини, що характеризують довільну динамічну систему (а), і узагальнена схема динамічної системи, описувана лінійними рівняннями стану (б)

$$\mathbf{X}(t) = F[\mathbf{X}(t_0); \mathbf{U}(t_0, t)]. \quad (3.12)$$

Вектор виходу $\mathbf{Y}(t)$ також є функцією $\mathbf{X}(t_0)$ і $\mathbf{U}(t_0, t)$:

$$\mathbf{Y}(t) = G[\mathbf{X}(t_0); \mathbf{U}(t_0, t)]. \quad (3.13)$$

Рівняння (3.12) і (3.13) називають *рівняннями стану* системи. Для систем, описуваних лінійними диференціальними рівняннями, рівняння стану (3.12) і (3.13) зводяться до наступних:

$$\dot{\mathbf{X}}(t) = A(t)\mathbf{X}(t) + B(t)\mathbf{U}(t_0, t); \quad (3.14)$$

$$Y(t) = C(t)X(t) + D(t)U(t_0, t), \quad (3.15)$$

де A – матриця коефіцієнтів розмірності $n \times n$; B – матриця керування (виходу) розмірності $n \times m$; C – матриця виходу розмірності $p \times n$; D – матриця обходу системи розмірності $p \times m$.

На рис. 3.2, б зображена узагальнена схема, динаміка якої описується рівняннями (3.14), (3.15) [32].

Якщо система стаціонарна, то її динаміка описується рівняннями стану, матриці яких мають незмінні в часі елементи, тобто

$$\dot{X}(t) = AX(t) + BU(t_0, t); \quad (3.16)$$

$$Y(t) = CX(t) + DU(t_0, t). \quad (3.17)$$

Матриця коефіцієнтів A визначає структуру системи, параметри елементів і їхні взаємні зв'язки. Динамічні властивості системи в основному визначаються цією матрицею. Матриця керування B показує зв'язок керуючих (збурюючих) впливів $U(t_0, t)$ із змінними стану $X(t)$. За відсутності вхідних впливів, тобто при досліджуванні вільного руху системи, матриця керування B дорівнює нулю. Матриця C формує вихідні змінні $Y(t)$ зі змінних стану $X(t)$. І нарешті, матриця обходу системи D показує безпосередню дію вхідних впливів $U(t_0, t)$ на вихідні координати без перетворення вектора $U(t_0, t)$ динамічною системою.

Слід зазначити, що рівняння стану (3.16) і (3.17) – це матричний запис системи лінійних диференціальних рівнянь першого порядку з постійними коефіцієнтами, що описують динаміку відповідної системи і взаємні зв'язки між змінними стану, вхідними та вихідними величинами.

Рівняння (3.16) і (3.17) називають *стандартною формою* запису рівнянь динаміки лінійних керованих систем з постійними параметрами, що мають довільну структуру і довільне число входів і виходів. Стандартна форма відрізняється компактністю і зручністю перетворення. Процедура розв'язання рівнянь стану в остаточному підсумку зводиться до матричних перетворень над A , B , C і D , що дуже зручно для програмування на цифрових ЕОМ. Рівняння (3.16) і (3.17) є вихідною інформацією при дослідженні і проектуванні систем керування методом простору станів.

До такого вигляду можна звести форми запису рівнянь динаміки передавальні функції, диференціальні рівняння високого порядку, матричні передавальні функції багатовимірних систем і т. д.), застосовувані в класичних методах дослідження систем.

Одержання рівнянь стану є початковим етапом дослідження систем у сучасній теорії керування.

Методи складання рівнянь стану динамічних систем. Щоб знайти рівняння стану (3.16) і (3.17), необхідно динаміку системи зобразити системою диференціальних рівнянь першого порядку. Як ілюстрацію розглянемо систему третього порядку, описувану рівнянням

$$\frac{d^3 y}{dt^3} + a_1 \frac{d^2 y}{dt^2} + a_2 \frac{dy}{dt} + a_3 y = b_0 u. \quad (3.18)$$

Для запису цього рівняння у векторно-матричній формі покладемо

$$y = x_1; \quad \frac{dy}{dt} = \frac{dx_1}{dt} = x_2; \quad \frac{d^2 y}{dt^2} = \frac{dx_2}{dt} = x_3; \quad (3.19)$$

тоді замість рівняння (3.18) з урахуванням (3.19) одержимо:

$$\left. \begin{aligned} \frac{dx_1}{dt} &= 0 \cdot x_1 + x_2 + 0 \cdot x_3 + 0 \cdot u; \\ \frac{dx_2}{dt} &= 0 \cdot x_1 + 0 \cdot x_2 + x_3 + 0 \cdot u; \\ \frac{dx_3}{dt} &= -a_3 x_1 - a_2 x_2 - a_1 x_3 + b_0 u; \end{aligned} \right\} \quad (3.20)$$

$$y = x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot u. \quad (3.21)$$

Система рівнянь (3.20) разом з рівнянням (3.21) у матричній формі запишеться так:

$$\frac{dX(t)}{dt} = \dot{X}(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix} X(t) + \begin{bmatrix} 0 \\ 0 \\ b_0 \end{bmatrix} U(t); \quad (3.22)$$

$$Y(t) = [1 \ 0 \ 0]X(t) + [0]U(t). \quad (3.23)$$

Рівняння (3.22) і (3.23) дають конкретний вигляд матриць A , B , C і D рівнянь (3.16) і (3.17).

Наведений метод може з успіхом застосовуватися для систем з одним входом і виходом, але в багатоконтурних системах з кількома входами і виходами зазначена процедура реалізується не так просто. Тому існують інші способи одержання рівнянь (3.16) і (3.17) стану досліджуваної системи.

Найрозповсюдженішим з них є зображення системи у вигляді схеми системи у змінних стану. Ця схема складається з інтеграторів, підсилювачів і підсумовуючих пристроїв. Звичайно виходи інтеграторів вибираються як координати (змінні) стану системи. Схема у змінних стану дає наочну фізичну інтерпретацію координат системи й описує їхній взаємний зв'язок. Схеми неперервних систем у змінних стану збігаються зі схемами моделювання цих систем на аналогових обчислювальних машинах. Існує багато різновидів схем моделювання для однієї і тієї ж системи, звідси і неоднозначність опису системи керування рівняннями стану.

Спочатку розглянемо методи побудови схем у змінних стану для систем з одним входом і виходом, динаміка яких описується диференціальним рівнянням

$$\frac{d^n y(t)}{dt^n} + a_1 \frac{d^{n-1} y(t)}{dt^{n-1}} + \dots + a_n y(t) = b_0 \frac{d^m u(t)}{dt^m} + b_1 \frac{d^{m-1} u(t)}{dt^{m-1}} + \dots + b_m u(t) \quad (3.24)$$

при початкових умовах $y(0) = y_0$, $y^{(1)}(0) = y_0^{(1)}$, ..., $y^{(n-1)}(0) = y_0^{(n-1)}$; $m < n$. Далі наведені приклади, застосовувані при побудові схем у змінних стану для систем більш складної структури, динаміка яких не може бути описана одним рівнянням вигляду (3.24). Кінцевим результатом побудови схем у змінних стану буде одержання рівнянь стану (3.16) і (3.17).

Метод комбінування похідних. Розглянемо спочатку простий випадок, коли початкові умови рівняння (3.24) нульові. При цьому (3.24) в операторній формі запишеться так:

$$A(p)Y(p) = B(p)U(p), \quad (3.25)$$

$$\text{де } A(p) = p^n + \sum_{i=0}^{n-1} a_{n-1} p^i; \quad B(p) = \sum_{i=0}^m b_{m-i} p^i.$$

Подамо рівняння (3.25) у такому вигляді:

$$\frac{Y(p)}{B(p)} = \frac{U(p)}{A(p)},$$

і замінимо його двома наступними:

$$\frac{Y(p)}{B(p)} = X(p); \quad \frac{U(p)}{A(p)} = X(p),$$

де $X(p)$ – зображення допоміжної змінної $x(t)$.

У диференціальній формі останні два рівняння наберуть вигляду

$$\frac{d^n x(t)}{dt^n} + a_1 \frac{d^{n-1} x(t)}{dt^{n-1}} + \dots + a_n x(t) = u(t); \quad (3.26)$$

$$b_0 \frac{d^m x(t)}{dt^m} + b_1 \frac{d^{m-1} x(t)}{dt^{m-1}} + \dots + b_m x(t) = y(t). \quad (3.27)$$

Отже, замість розв'язування рівняння (3.24) можна розв'язувати рівняння (3.26) і (3.27). Складемо схему розв'язання (3.26) щодо старшої похідної:

$$\frac{d^n x(t)}{dt^n} = -a_1 \frac{d^{n-1} x(t)}{dt^{n-1}} - a_2 \frac{d^{n-2} x(t)}{dt^{n-2}} - \dots - a_n x(t) + u(t). \quad (3.28)$$

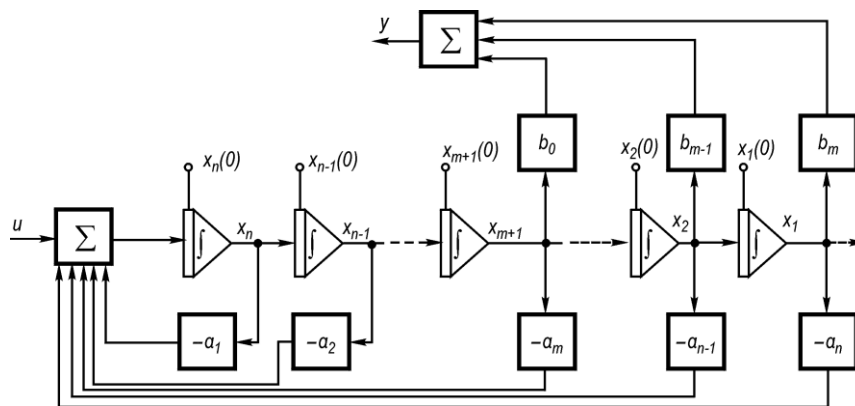


Рис. 3.3. Схема у змінних стану для методу комбінування похідних

Припустимо, що n -на похідна функції $x(t)$ відома, тоді, подаючи її на ланцюжок послідовно з'єднаних інтеграторів, на виході кожного наступного інтегратора матимемо похідні від функції $x(t)$ більш низьких порядків. Якщо кожен з цих похідних помножити на відповідний коефіцієнт a_i і взяти їх від'ємну суму разом з функцією $u(t)$, то одержимо праву частину рівняння (3.28), тобто n -ну похідну функції $x(t)$, яка раніше вважалася відомою. Функція $y(t)$, що є розв'язком рівняння (3.24), – лінійна комбінація похідних від $x(t)$, узятих з коефіцієнтами b_i відповідно до рівняння (3.27). Схема розв'язання рівнянь (3.26) і (3.27) наведена на рис. 3.3.

При ненульових початкових умовах рівняння (1.24) їх необхідно перевести в початкові умови (3.26). Припускаючи, що рівняння (1.26) і (1.27) еквівалентні (1.24), при ненульових початкових умовах запишемо їх в операторній формі:

$$\left. \begin{aligned} A(p)Y(p) + Y_0(p, a) &= B(p)U(p); \\ A(p)X(p) + X_0(p, a) &= U(p); \\ B(p)X(p) + X_0(p, b) &= Y(p), \end{aligned} \right\} \quad (3.29)$$

де $Y_0(p, a)$, $X_0(p, a)$, $X_0(p, b)$ – поліноми початкових умов.

Підставляючи $U(p)$ і $Y(p)$ з останніх двох рівнянь (3.29) у перше, одержуємо:

$$A(p) [Y(p)X(p) + X_0(p, b)] + Y_0(p, a) = Y(p) [A(p)X(p) + X_0(p, a)].$$

З цього рівняння випливає тотожна рівність, за якою визначаються початкові умови рівняння (3.26) за відомими початковими умовами (3.24):

$$Y_0(p, a) = B(p)X_0(p, a) - A(p)X_0(p, b). \quad (3.30)$$

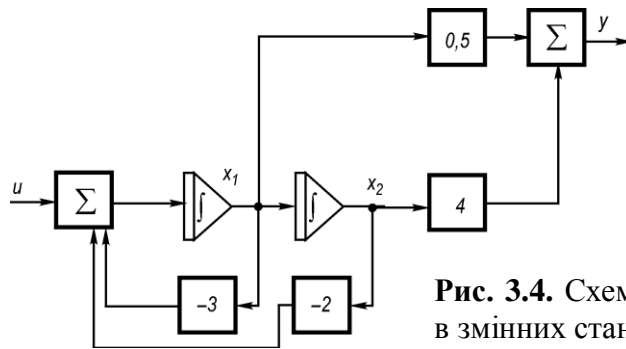


Рис. 3.4. Схема в змінних стану до прикладу

Метод послідовного інтегрування. Запишемо рівн.1.1 операторній формі при нульових початкових умовах у вигляді

$$p^n Y(p) + \sum_{i=0}^{n-1} a_{n-i} p^i Y(p) = \sum_{i=0}^m b_{m-i} p^i U(p),$$

звідки

$$Y(p) = \sum_{i=0}^{n-1} \frac{1}{p^{n-i}} [b_{m-i} U(p) - a_{n-i} y Y(p)]; \quad b_{m-i} = 0 \text{ при } i > m. \quad (3.33)$$

Складемо ланцюжок з n послідовно з'єднаних інтеграторів. Сигнал на виході крайнього правого інтегратора візьмемо за y . Виконання рівності (3.33) забезпечиться в тому випадку, якщо ввести зворотний зв'язок, тобто на вхід крайнього лівого інтегратора подати різницю $b_m u - a_n y$, на вхід наступного – різницю $b_{m-1} u - a_{n-1} y$ і вихід попереднього інтегратора, і т. д. Схема розв'язання рівняння (3.33) тоді набере вигляду, показаного на рис. 3.5.

Якщо знову вибрати за змінні стану x_k виходи інтеграторів схеми рис.3.5, то одержимо систему рівнянь

$$\begin{cases} \dot{x}_1(t) = x_2(t) - a_1 x_1(t); \\ \dot{x}_2(t) = x_3(t) - a_2 x_1(t); \\ \dots\dots\dots \\ \dot{x}_{n-m}(t) = x_{n-m+1}(t) - a_m(t) x_1(t) + b_0 u(t); \\ \dots\dots\dots \\ \dot{x}_n(t) = -a_n x_1(t) + b_m u(t); \end{cases} \quad (3.34)$$

$$y(t) = x_1(t). \quad (3.35)$$

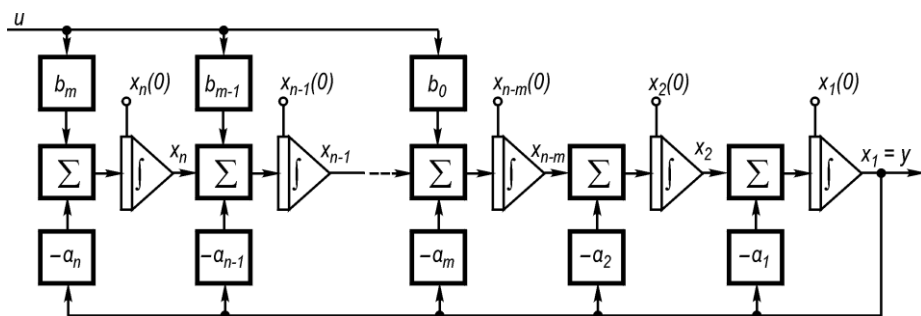


Рис. 3.5. Схема в змінних стану для методу послідовного інтегрування

При ненульових початкових умовах рівняння (3.24) початкові умови інтеграторів схеми визначаються співвідношенням [32]

$$x_i(0) = \sum_{k=0}^i a_k y_0^{(i-k)}.$$

Записуючи цю систему рівнянь у формі рівнянь (3.16) і (3.17), знаходимо матриці A , B , C , D :

$$A = \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 & 0 \\ -a_2 & 0 & 1 & \cdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_m & 0 & 0 & \cdots & 1 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -a_{n-1} & 0 & 0 & \cdots & 0 & 1 \\ -a_n & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}; \quad B = \begin{bmatrix} 0 \\ 0 \\ \dots \\ b_0 \\ \dots \\ b_m \end{bmatrix}; \quad C = 1 \ 0 \ \dots \ 0; \quad D = 0.$$

Як і для методу комбінування похідних, коефіцієнти рівняння (3.24) є одночасно і коефіцієнтами схеми у змінних стану. Тому, знаючи загальну структуру схеми в змінних стану, можна безпосередньо за видом рівняння (3.24) побудувати відповідну схему в змінних стану і знайти рівняння стану у вигляді (3.16) і (3.17).

Приклад 3.2. Скласти рівняння стану системи, описуваної рівнянням

$$\begin{aligned} \frac{d^3 y(t)}{dt^3} + \frac{d^2 y(t)}{dt^2} + 2 \frac{dy(t)}{dt} + 3y(t) = \\ = 0,2 \frac{d^3 u(t)}{dt^3} + 0,4 \frac{d^2 u(t)}{dt^2} + 1,5 \frac{du(t)}{dt} + 4u(t). \end{aligned}$$

Розв'язання. Схема в змінних стану, побудована методом безпосереднього інтегрування, показана на рис. 1.6. Зі схеми одержуємо систему рівнянь першого порядку

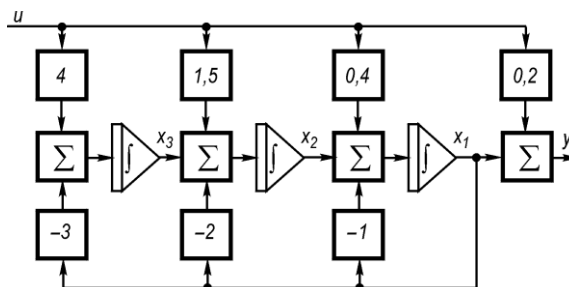


Рис. 3.6. Схема в змінних стану до прикладу 1.2

$$\begin{aligned} \dot{x}_1 &= -x_1 + x_2 + 0,4u; \\ \dot{x}_2 &= -2x_1 + x_3 + 1,5u; \\ \dot{x}_3 &= -3x_1 + 4u; \\ y &= x_1 + 0,2u. \end{aligned}$$

У матричній формі ця система рівнянь виглядає так:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ -2 & 0 & 1 \\ -3 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0,4 \\ 1,5 \\ 4 \end{bmatrix} u;$$

$$Y = 1 \ 0 \ 0 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + 0,2 u.$$

З останніх рівнянь ясно видно матриці A , B , C , D .

У прикладі 3.2 матриця D відмінна від нуля. Це має місце в тих випадках, коли $m = n$, тобто максимальний ступінь похідної правої частини диференціального рівняння дорівнює максимальному ступеню похідної лівої частини рівняння. Якщо $m < n$, то матриця D завжди буде

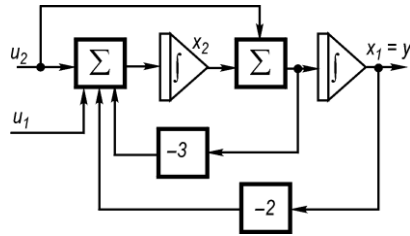


Рис. 3.8. Схема системи в змінних стану до прикладу 3.3

нульовою.
Випадок $m > n$
відповідає
фізично
нереалізованій
системі, тому
становить тільки
чисто

теоретичний інтерес і тут не розглядається.

Метод розкладу передавальної функції на елементарні дроби є дуже перспективним для аналізу систем методом простору станів і полягає в побудові схеми у змінних стану шляхом розкладу передавальної функції на елементарні дроби. Для цього рівняння (3.24) зображується у вигляді передавальної функції

$$\frac{Y(p)}{U(p)} = \frac{b_0 p^m + b_1 p^{m-1} + \dots + b_m}{p^n + a_1 p^{n-1} + \dots + a_n} \quad (3.36)$$

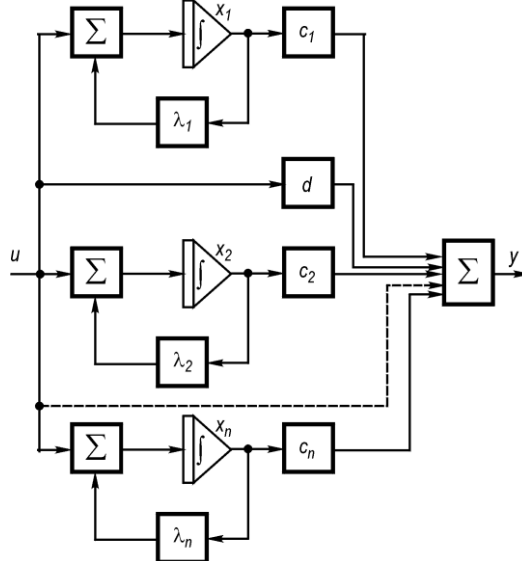
і розкладається на елементарні дроби:

$$\frac{Y(p)}{U(p)} = d + \frac{c_1}{(p - \lambda_1)} + \frac{c_2}{(p - \lambda_2)} + \dots + \frac{c_n}{(p - \lambda_n)},$$

Звідки

$$Y(p) = d \cdot U(p) + \left[\frac{c_1}{(p - \lambda_1)} + \frac{c_2}{(p - \lambda_2)} + \dots + \frac{c_n}{(p - \lambda_n)} \right] U(p). \quad (3.37)$$

Рис. 3.7. Схема в змінних стану для методу розкладу передавальної функції на елементарні дроби
Схема у змінних стану рівняння (3.37) показана на рис. 3.7. Рівняння стану мають вигляд



$$\dot{X} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} u; \quad (3.38)$$

$$Y = c_1 \ c_2 \ \dots \ c_n \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + [d]u. \quad (3.39)$$

Як видно з виразу (3.38), матриця A є діагональною.

За наявності кратних полюсів передавальної функції (3.36) матриця A матиме канонічну форму Жордана. Нижче буде показано, що обчислювальна процедура значно спрощується, якщо матриця A діагональна чи має канонічну форму Жордана. Основні труднощі одержання рівнянь стану (3.38) і (1.39) полягають у пошуку полюсів передавальної функції (3.36).

Нормальна форма рівнянь стану. Метод розкладу на елементарні дроби корисний для виведення рівнянь стану системи з одним входом і виходом. У цьому випадку матриці B і C є векторами, а D , y і u – скалярами.

Щодо багатозв'язних систем, метод розкладу на елементарні дроби стає громіздким. Тому бажано мати можливість перетворювати рівняння стану *стандартної форми* до вигляду, де матриця A діагональна або, що є більш загальним, – до вигляду, де матриця A – матриця Жордана.

Розглянемо систему з постійними параметрами, обумовлену рівняннями стану в *стандартній формі* (3.16) і (3.17). Вважається, що характеристичні числа A різні. При введенні лінійного перетворення $X(t) = Mq(t)$, де M – модальна матриця, рівняння (3.16) і (3.17) мають вигляд

$$M\dot{q}(t) = AMq(t) + BU(t); \quad (3.40)$$

$$Y(t) = CMq(t) + DU(t). \quad (3.41)$$

Множення рівняння (3.40) на M^{-1} дає

$$\dot{q}(t) = M^{-1}AMq(t) + M^{-1}BU(t). \quad (3.42)$$

Оскільки M – модальна матриця, то перетворення $M^{-1}AM$ дає діагональну матрицю Λ . Головними діагональними елементами Λ є характеристичні числа λ_i , $i=1, 2, \dots, n$. Отже,

$$\dot{q}(t) = \ddot{E}q(t) + B_q U(t), \quad (3.43)$$

$$Y(t) = Cq(t) + DU(t). \quad (3.44)$$

Рівняння (3.43) і (3.44) відомі як *нормальна форма* рівнянь стану. При цьому диференціальні рівняння розв'язані щодо змінних стану q_i , $i=1, 2, \dots, n$, тобто мають вигляд $\dot{q}_i = \lambda_i q_i + f_i$, де f_i – збурююча функція, що діє на i -ту змінну стану. Зазначена процедура застосовна, якщо відома *стандартна форма* рівнянь стану. За своїм змістом вона узагальнює метод розкладу передавальної функції на елементарні дроби.

У випадку кратних характеристичних чисел матриця Λ замінюється недиагональною жордановою матрицею.

Приклад 3.3. Одержати рівняння стану в *нормальній формі* для системи, схема у змінних стану якої наведена на рис. 3.8.

Розв'язання. Рівняння стану в *стандартній формі* мають вигляд

$$X = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}; \quad Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Знайдемо модальну матрицю M . Характеристичні числа A визначають з рівняння $|\lambda I - A| = 0$, де I – одинична матриця.

Після підстановки матриці A з останнього рівняння

$$\begin{vmatrix} \lambda & -1 \\ 2 & \lambda + 3 \end{vmatrix} = 0, \quad \lambda(\lambda + 3) + 2 = 0.$$

Розв'язками цього рівняння є $\lambda_1 = -1$; $\lambda_2 = -2$.

Оскільки характеристичні числа матриці A різні, то стовпці модальної матриці пропорційні довільному стовпцю $Adj \lambda_i I - A$. Спочатку знайдемо приєднану матрицю $Adj \lambda_1 I - A$:

$$Adj \lambda_1 I - A = \begin{bmatrix} \lambda + 3 & 1 \\ -2 & \lambda \end{bmatrix}.$$

Підставивши в цей вираз послідовно λ_1 і λ_2 , одержимо:

$$Adj \lambda_1 I - A = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix}; \quad Adj \lambda_2 I - A = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}.$$

Модальну матрицю виберемо у вигляді

$$M = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}.$$

Обернена матриця

$$M^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}.$$

Знайдемо матриці \dot{E} , B_q і C_q :

$$\dot{E} = M^{-1} A M = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}; \quad B_q = M^{-1} B = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}; \quad C_q = C M = [1 \quad 1].$$

Рівняння стану в нормальній формі

$$\dot{q} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} q + \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}; \quad Y = [1 \quad 1] q.$$

Матриця коефіцієнтів рівнянь стану в *нормальній формі* є діагональною. Елементи головної діагоналі дорівнюють характеристичним числам A рівняння стану в *стандартній формі*.

Перетворення неоднорідних рівнянь стану в однорідні. Побудувавши схему в змінних стану, можна одержати рівняння стану у вигляді (3.16). Це система неоднорідних диференціальних рівнянь першого порядку. До їх розв'язку входять дві складові – вільна і вимушена. Перша залежить від динаміки системи і початкового значення вектора змінних стану X_{t_0} . Друга залежить від матриці A та вектора вхідних впливів $U_{t_0, t}$.

Вільна складова визначається розв'язанням однорідного рівняння

$$\dot{X}(t) = A X(t), \tag{3.45}$$

якщо вектор вхідного впливу $U_{t_0, t}$ у виразі (3.16) припускається нульовим.

Має сенс спробувати перетворити систему неоднорідних рівнянь (3.16) в однорідну (3.45) з метою одержати рівняння стану системи у разі наявності зовнішніх впливів $U t$, що містять тільки змінні стану. При цьому значно спрощується розв'язання рівнянь стану.

Перетворення системи неоднорідних рівнянь в однорідну можна здійснити в тому випадку, коли вектор зовнішніх впливів $U t$ поданий у вигляді розв'язку лінійного однорідного диференціального рівняння (системи рівнянь) з постійними параметрами. Схема у змінних стану для формування вектора $U t$ матиме свої змінні стану $z t$ з початковим значенням $z 0$.

Розглянемо методику побудови схем у змінних стану для формування деяких вхідних впливів $u t$.

Побудова схем у змінних стану для типових вхідних впливів. Поліноміальний вхідний вплив описується функцією

$$u(t) = r_0 + r_1 t + r_2 t^2 + \dots + r_{n-1} t^{n-1} \text{ при } t > 0. \quad (3.46)$$

Позначимо

$$z_1(t) = u(t) \quad (3.47)$$

і здиференціюємо рівняння (3.47) за t . Тоді

$$\dot{z}_1(t) = r_1 + 2r_2 t + \dots + (n-1)r_{n-1} t^{n-2} = z_2(t).$$

Аналогічно

$$\dot{z}_2(t) = 2r_2 + 6r_3 t + \dots + (n-1)(n-2)r_{n-1} t^{n-3} = z_3(t).$$

Продовжуємо цей процес до тих пір, поки чергова похідна не дорівнюватиме нулю. Одержимо наступну систему рівнянь:

$$\begin{aligned} \dot{z}_1(t) &= z_2(t); \\ \dot{z}_2(t) &= z_3(t); \\ &\dots\dots\dots \\ \dot{z}_{n-1}(t) &= z_n(t); \\ \dot{z}_n(t) &= 0. \end{aligned} \quad (3.48)$$

Величини $z_i t \quad i=1, 2, \dots, n$ називаються змінними (координатами) стану входу. Система диференціальних рівнянь (3.48) показує, що поліноміальний вплив можна одержати за допомогою ланцюжка послідовно з'єднаних інтеграторів (рис.3.9, а). Початкові умови на інтеграторах дорівнюють

$$\begin{aligned} z_1(0) &= r_0; \\ z_2(0) &= r_1; \\ &\dots\dots\dots \\ z_i(0) &= (i-1)!r_{n-i}; \\ &\dots\dots\dots \\ z_n(0) &= (n-1)!r_{n-1}. \end{aligned} \quad (3.49)$$

Виходом цієї системи є заданий вхідний вплив, обумовлений змінною стану $z_1 t$. Матриці A_u і C_u рівнянь стану знаходяться з (3.48) і (3.47):

$$A_u = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}; \quad C_u = [1 \ 0 \ \dots \ 0]. \quad (3.50)$$

Приклад 3.4. Скласти схему в змінних стану для формування лінійного впливу $u(t) = r_0 + r_1 t$ і знайти відповідні рівняння стану.

Розв'язання. Систему рівнянь, розв'язком якої буде лінійна функція, одержимо з (3.48):

$$\begin{aligned} \dot{z}_1 &= z_2; \\ \dot{z}_2 &= 0 \end{aligned}$$

при початкових умовах $z_1(0) = r_0$, $z_2(0) = r_1$. Вихідний сигнал $u(t) = z_1$.

Схема в змінних стану показана на рис. 1.9, б. У матричному вигляді рівняння стану схеми (рис. 3.9, б) такі:

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_1(0) \\ z_2(0) \end{bmatrix}, \quad u(t) = [1 \ 0] \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

звідки

$$A_u = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad C_u = [1 \ 0].$$

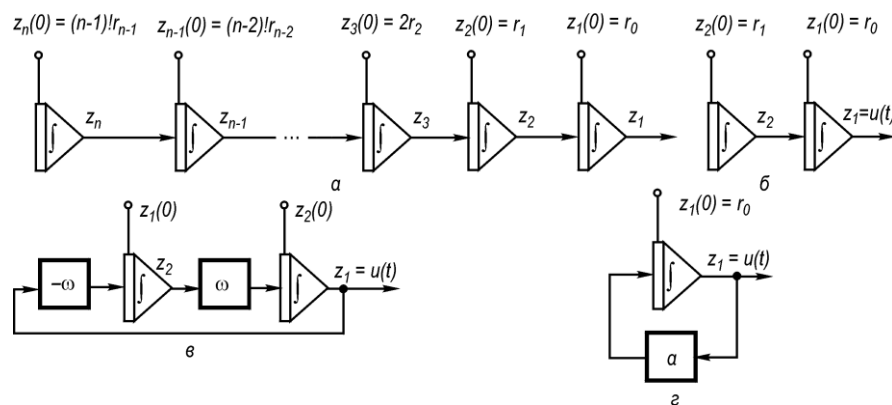


Рис. 3.9. Схеми в змінних стану для входних впливів:

a – поліноміального; *б* – лінійного; *в* – гармонічного; *г* – експоненціального

У загальному випадку гармонічний вплив зображується так:

$$u(t) = a \cos \omega t + b \sin \omega t. \quad (3.51)$$

Розв'язком диференціального рівняння

$$\ddot{z}(t) + \omega^2 z(t) = 0 \quad (3.52)$$

є гармонічний сигнал. Це рівняння може бути записане у вигляді двох рівнянь першого порядку:

$$\begin{aligned} \dot{z}_1 &= \omega \cdot z_2; \\ \dot{z}_2 &= -\omega \cdot z_1 \end{aligned} \quad (3.53)$$

з початковими умовами $z_1(0) = a$ і $z_2(0) = b$.

Схема в змінних стану, яка впливає з (3.53), показана на рис.3.9, в. Сигнал $u(t)$ визначається за формулою

$$u(t) = z_1(0) \cos \omega t + z_2 \sin \omega t. \quad (3.54)$$

Змінюючи початкові умови $z_1(0)$ і $z_2(0)$, можна одержувати гармонічні сигнали різних фаз й

Рис. 3.11. Розширена схема в змінних стану до прикладу 3.6

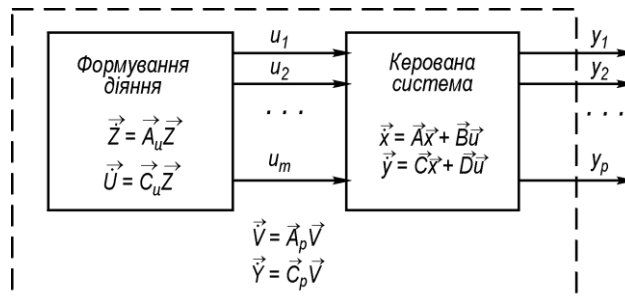
амплітуд.

Експоненціальний вхідний вплив описується функцією

$$u(t) = r_0 e^{at} \text{ при } t \geq 0. \quad (3.55)$$

Нехай $u(t) = z_1(0)$. Тоді

$$\dot{z}_1(t) = ar_0 e^{at} = au(t) = az_1(t). \quad (3.56)$$



Початкове значення $z_1(0) = r_0$. Схема в змінних стану системи, що використовує це останнє рівняння, показана на рис.3.9, г. Змінна $z(t)$ є впливом експоненціального вигляду.

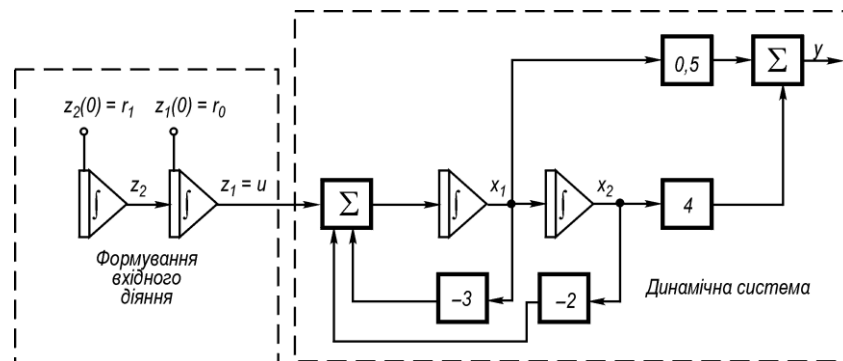


Рис. 3.10. До перетворення неоднорідного рівняння в однорідне

Методика перетворення неоднорідних рівнянь стану в однорідні (рис.3.10). Для перетворення системи неоднорідних рівнянь (3.16) в однорідну (1.34) необхідно вектор $Z(t)$ розглядати разом зі змінними стану системи $X(t)$, тобто ввести вектор $Z(t)$ до координат системи. Допустимо, що вхідний вектор $U(t)$ є розв'язком рівнянь стану (рис.3.11)

$$\dot{Z}(t) = A_u Z(t); \quad (3.57)$$

$$\dot{U}(t) = C_u Z(t), \quad (3.58)$$

отриманих шляхом побудови схеми в змінних стану диференціального рівняння, розв'язком якого є вектор $U(t)$. Вектор $U(t)$ буде вхідним для досліджуваної системи з рівняннями стану вигляду (3.16) і (3.17). Підставимо рівняння (3.47) у (3.16) і (3.17):

$$\dot{X}(t) = AX(t) + BC_u Z(t); \quad (3.59)$$

$$\dot{Y}(t) = CX(t) + DC_u Z(t). \quad (3.60)$$

Об'єднаємо в одну систему рівняння (3.59) і (3.60):

$$\begin{aligned} \dot{X}(t) &= AX(t) + BC_u Z(t); \\ \dot{Z}(t) &= CX(t) + A_u Z(t). \end{aligned} \quad (3.61)$$

Позначимо розширений вектор, що включає як змінні стану $x_k(t)$ керованої системи, так і змінні стану $z_i(t)$ схеми в змінних стану вхідного впливу $U(t)$, через $V(t)$, тобто

$$V(t) = \begin{bmatrix} X(t) \\ Z(t) \end{bmatrix}. \quad (3.62)$$

Рівняння (3.60) і (3.61) можна записати так:

$$\dot{V}(t) = \begin{bmatrix} \dot{X}(t) \\ \dot{Z}(t) \end{bmatrix} = \begin{bmatrix} A & BC_u \\ C & A_u \end{bmatrix} V(t); \quad (3.63)$$

$$\dot{Y}(t) = [C \quad DC_u] V(t), \quad (3.64)$$

звідки одержуємо

$$\dot{V}(t) = A_p V(t); \quad (3.65)$$

$$\dot{Y}(t) = C_p V(t). \quad (3.66)$$

Матриця A_p називається *розширеною матрицею коефіцієнтів*, C_p – *розширеною матрицею виходу*, $V(t)$ – *розширеним вектором стану*.

Застосування методики перетворення неоднорідної системи рівнянь стану в однорідну проілюструємо на прикладі.

Приклад 3.5. Знайти матриці A_p і C_p для системи прикладу 3.1 при лінійному вхідному впливі. Розв'язання. З прикладу 5.1

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}; \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \quad C = [4 \quad 0,5]; \quad D = [0].$$

Для лінійного впливу $A_u = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad C_u = [1 \quad 0].$

Підставляючи A, B, C, D, A_u, C_u у (5.65), одержуємо

$$A_p = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -2 & -3 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}; C_p = [4 \ 0,5 \ 0 \ 0]. \quad (3.67)$$

На практиці часто користуються наступним прикладом перетворення неоднорідних рівнянь (3.16) і (3.17) в однорідні. Складається розширена схема в змінних стану, що включає схеми в змінних стану вхідного впливу і власне динамічної системи. З цієї схеми безпосередньо одержують розширені матриці A_p і C_p .

Приклад 3.6. Для умов прикладу 1.5 знайти матриці A_p і C_p за розширеною схемою в змінних стану.

Розв'язання. Розширена схема в змінних стану показана на рис.3.11. З неї одержуємо

$$\begin{aligned} \dot{x}_1 &= x_2; \\ \dot{x}_2 &= -2x_1 - 3x_2 + z_1; \\ \dot{z}_1 &= z_2; \\ \dot{z}_2 &= 0; \\ y &= 4x_1 + 0,5x_2. \end{aligned}$$

У матричній формі цю систему рівнянь можна записати, позначивши $V = [x_1 \ x_2 \ z_1 \ z_2]^T$, так:

$$\dot{V} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -2 & -3 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} V; \quad y = 4 \ 0,5 \ 0 \ 0 V.$$

Результати прикладів 3.5 і 3.6 збігаються.

3.4. ФОРМИ РОЗВ'ЯЗАННЯ РІВНЯНЬ СТАНУ

Розглянуті методи побудови схем у змінних стану дозволяють одержати математичний опис динаміки лінійної стаціонарної системи у вигляді векторно-матричної системи диференціальних рівнянь першого порядку. Розглянемо розв'язання цих рівнянь.

Форма розв'язання однорідних рівнянь стану. Якщо на систему не подаються зовнішні діяння, то $u \ t = 0$ і рівняння (3.16) стає однорідним:

$$\dot{X}(t) = AX(t). \quad (3.68)$$

Розв'язок цього рівняння описує динаміку системи при ненульових початкових умовах (вільний рух), коли зовнішні сили дорівнюють нулю. Припустимо, що рух починається в момент $t = t_0$ з початкового стану $X \ t_0$. Розв'язок рівняння (3.68) можна дістати за аналогією з розв'язком скалярного рівняння у вигляді

$$X(t) = e^{A(t-t_0)} X(t_0). \quad (3.69)$$

Дійсно, якщо підставити (3.69) у (3.68), попередньо взявши похідну від (1.69), то одержимо тотожність. Отже, (1.69) є розв'язком однорідного матричного рівняння (1.68). Якщо позначити

$$\Phi(t - t_0) = e^{A(t-t_0)}, \quad (3.70)$$

то рівняння (3.69) можна записати так:

$$X(t) = \Phi(t - t_0)X(t_0). \quad (3.71)$$

Вихідний вектор системи матиме вигляд

$$Y(t) = C\Phi(t - t_0)X(t_0). \quad (3.72)$$

Матриця $\Phi(t - t_0) = e^{A(t-t_0)}$ називається *матрицею переходу (перехідною матрицею)* стану системи. Матриця переходу $\Phi(t - t_0)$ описує рух кінця вектора стану $X(t)$ у просторі стану з початкового положення $X(t_0)$, а, отже, й зміну (перехід) стану системи. Вектор $X(t)$ дає можливість контролювати всі координати (змінні) стану системи, тому розв'язок (3.71) несе в собі повну інформацію щодо динаміки системи. Необхідно зазначити, що обсяг обчислень при визначенні матриці переходу $\Phi(t - t_0)$ стану системи звичайно більший, ніж при розв'язанні лінійного диференціального рівняння щодо залежної змінної. Однак отримувана в результаті обчислення матриці переходу додаткова інформація дозволяє провести більш глибокий аналіз динаміки досліджуваної системи.

Форма розв'язання неоднорідних рівнянь стану. Розв'язок рівняння (3.16) шукатимемо в формі, аналогічній (3.71). Покладемо

$$X(t) = \Phi(t - t_0)C_1(t), \quad (3.73)$$

де $C_1(t)$ – вектор, що залежить від часу і заміняє вектор початкового стану $X(t_0)$ у рівнянні (3.71), якщо зовнішні впливи відсутні.

Диференціюючи (3.73) за t , одержуємо

$$\dot{X}(t) = \dot{\Phi}(t - t_0)C_1(t) + \Phi(t - t_0)\dot{C}_1(t). \quad (3.74)$$

З урахуванням того, що $\Phi(t - t_0) = e^{A(t-t_0)}$, а $\dot{\Phi}(t - t_0) = Ae^{A(t-t_0)}$, рівняння (1.74) запишемо у такому вигляді:

$$\dot{X}(t) = A(t)X(t) + \Phi(t - t_0)\dot{C}_1(t). \quad (3.75)$$

Якщо (3.73) є розв'язком рівняння (3.16), то величини в правих частинах рівнянь (3.16) і (3.75) мають бути однаковими.

Тому

$$BU(t) = \Phi(t - t_0)\dot{C}_1(t). \quad (3.76)$$

Розв'яжемо це рівняння відносно $\dot{C}_1(t)$:

$$C_1(t) = \int_{t_0}^t \Phi^{-1}(\tau - t_0)BU(\tau)d\tau + C_2. \quad (3.77)$$

Підставивши (3.77) у (3.73), матимемо

$$X(t) = \Phi(t - t_0)C_2 + \Phi(t - t_0) \int_{t_0}^t \Phi^{-1}(\tau - t_0)BU(\tau)d\tau. \quad (3.78)$$

Оскільки

$$\Phi(t - t_0)\Phi^{-1}(\tau - t_0) = \Phi(t - \tau), \quad (3.79)$$

вираз (3.78) набере вигляду

$$X(t) = \Phi(t-t_0)C_2 + \int_{t_0}^t \Phi(t-\tau)BU(\tau)d\tau. \quad (3.80)$$

Сталу інтегрування C_2 можна визначити, взявши в рівнянні (1.80) $t = t_0$. Тоді $X(t_0) = \Phi(0)C_2$. Оскільки $\Phi(0) = 1$, то

$$C_2 = X(t_0). \quad (3.81)$$

Отже,

$$X(t) = \Phi(t-t_0)X(t_0) + \int_{t_0}^t \Phi(t-\tau)BU(\tau)d\tau. \quad (3.82)$$

Вектор виходу

$$Y(t) = C\Phi(t-t_0)X(t_0) + \int_{t_0}^t C\Phi(t-\tau)BU(\tau)d\tau + DU(t). \quad (3.83)$$

Перший доданок у (3.83) – складова вихідного вектора за рахунок ненульових початкових умов $X(t_0)$. Вона збігається з розв'язком однорідного рівняння (3.68). Другий доданок характеризує реакцію системи на зовнішній вплив $U(t_0, t)$, а третій являє собою безпосередній вплив вектора входу $U(t_0, t)$ на вихід $Y(t)$.

3.5. АНАЛІТИЧНИЙ ПІДХІД ДО ОБЧИСЛЕННЯ МАТРИЦІ ПЕРЕХОДУ

Форми розв'язку однорідної (3.61) і неоднорідної (3.72) систем рівнянь стану містять матрицю переходу $\Phi(t-t_0)$. Тому одною з центральних задач дослідження систем є визначення матриці переходу. Існує багато методів визначення $\Phi(t-t_0)$, основними з яких є метод розкладу $\Phi(t-t_0)$ у нескінченний ряд, метод комплексної площини, метод Сильвестра і метод Келі–Гамільтона.

Нижче розглядаються й будуються аналітичні методи обчислення матриці переходу $\Phi(t-t_0)$, застосовувані в практиці аналізу і синтезу систем керування. Вони ілюструються прикладами.

Метод розкладу матриці переходу в нескінченний ряд. Перехідна матриця $\Phi(t) = e^{At}$ розкладається в нескінченний ряд

$$\Phi(t) = 1 + At + \frac{A^2 t^2}{2!} + \frac{A^3 t^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{A^k t^k}{k!}. \quad (3.84)$$

Метод розкладу найбільш трудомісткий, якщо елементи матриці A^k не “зникають” при невеликих k . Після виконання підсумовування слід знайти в замкненому вигляді кожен елемент матриці $\Phi(t)$. Це непросте завдання, і якщо порядок $\Phi(t)$ недостатньо низький, то вона може бути і нерозв'язною. Проте метод розкладу в нескінченний ряд дуже зручний при обчисленні $\Phi(t)$ на цифрових обчислювальних машинах, що буде показано далі.

Приклад 3.7. Знайти матрицю переходу $\Phi(t) = e^{At}$ для матриці

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

Розв'язання. Знаходимо степеневі множники:

$$A^2 = \begin{bmatrix} 1 & -4 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 4 \end{bmatrix}; \quad A^3 = \begin{bmatrix} -1 & 13 & 0 \\ 0 & -27 & 0 \\ 0 & 0 & 8 \end{bmatrix}; \quad \dots; \quad A^k = \begin{bmatrix} (-1)^k & a_k & 0 \\ 0 & (-3)^k & 0 \\ 0 & 0 & (-2)^k \end{bmatrix},$$

де a_k визначають за рекурентною формулою

$$a_k = (-1)^{k-1} - 3a_{k-1} \quad \text{при } a_0 = 1.$$

Перехідна матриця

$$\Phi(t) = e^{At} = \sum_{k=0}^{\infty} A^k \frac{t^k}{k!} = \begin{bmatrix} \sum_{k=0}^{\infty} (-1)^k \frac{t^k}{k!} & \sum_{k=0}^{\infty} a^k \frac{t^k}{k!} & 0 \\ 0 & \sum_{k=0}^{\infty} (-3)^k \frac{t^k}{k!} & 0 \\ 0 & 0 & \sum_{k=0}^{\infty} (-2)^k \frac{t^k}{k!} \end{bmatrix}.$$

Якщо згорнути нескінченні ряди усередині матриці Φt , одержимо

$$\Phi(t) = \begin{bmatrix} e^{-t} & 0,5(e^{-t} - e^{-3t}) & 0 \\ 0 & e^{-3t} & 0 \\ 0 & 0 & e^{-2t} \end{bmatrix}.$$

Метод комплексної площини. Виконаємо перетворення Лапласа над обома частинами рівняння (3.68):

$$pX(p) - X(0) = AX(p), \quad (3.85)$$

звідки

$$X(p) = [pI - A]^{-1} X(0). \quad (3.86)$$

Застосувавши до рівняння (3.86) обернене перетворення Лапласа, одержимо

$$X(t) = L^{-1} [pI - A]^{-1} X(0). \quad (3.87)$$

Порівнявши (3.71) і (3.87), дійдемо висновку, що

$$\Phi(t) = e^{At} = L^{-1} [pI - A]^{-1}. \quad (3.88)$$

Основні труднощі цього методу полягають у пошуку матриці, оберненої до $pI - A$.

Приклад 3.8. Знайти матрицю переходу $\Phi t = e^{At}$, якщо

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}.$$

Розв'язання. Знайдемо матрицю $pI - A$:

$$pI - A = \begin{bmatrix} p & 0 \\ 0 & p \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} p & -1 \\ 2 & p+3 \end{bmatrix}.$$

Матриця, обернена до $pI - A$, має вигляд

$$pI - A^{-1} = \frac{1}{p(p+3)+2} \begin{bmatrix} p & -1 \\ 2 & p+3 \end{bmatrix} = \begin{bmatrix} \frac{p+3}{(p+1)(p+2)} & \frac{1}{(p+1)(p+2)} \\ \frac{2}{(p+1)(p+2)} & \frac{p}{(p+1)(p+2)} \end{bmatrix}.$$

За таблицею перетворення Лапласа [10], знаходимо

$$\Phi(t) = L^{-1} pI - A^{-1} = \begin{bmatrix} 2e^{-t} & -e^{-2t} & e^{-t} & -e^{-2t} \\ 2e^{-2t} & -2e^{-t} & 2e^{-2t} & -e^{-t} \end{bmatrix}.$$

Метод комплексної площини дає можливість зрозуміти фізичний зміст матриці переходу $\Phi(t)$. Довільний i -тий член $x_i(t)$ рівняння (1.71) при $t_0 = 0$ можна записати як суму:

$$x_i(t) = \sum_{j=1}^n \varphi_{ij}(t) x_j(0), \quad (3.89)$$

де $\varphi_{ij}(t)$ – елемент матриці $\Phi(t)$, розташований на перетині i -го рядка і j -го стовпця.

Елемент $\varphi_{ij}(t)$ можна визначити, поклавши, що $x_j(0) = 1$ і початкові умови для інших змінних стану дорівнюють нулю. Тоді $x_j(0) = \varphi_{ij}(t)$. Стосовно схем у змінних стану це еквівалентно подачі одиничної початкової умови на вхід інтегратора j і виходу сигналу з інтегратора i . Однак одинична початкова умова на виході інтегратора рівносильна подачі одиничного миттєвого імпульсу на вхід цього інтегратора. Отже, $\varphi_{ij}(t)$ є реакцією на виході інтегратора i при подачі одиничного миттєвого імпульсу на вхід інтегратора j , якщо початкові умови інших інтеграторів дорівнюють нулю. Тому $\varphi_{ij}(t)$ можна інтерпретувати як імпульсну перехідну функцію від входу суматора на вході інтегратора j до виходу інтегратора i . Ці імпульсні перехідні функції утворюють $\Phi(t) = L^{-1} pI - A^{-1}$.

Метод Сильвестра. Цей метод дозволяє матричний багаточлен довільного степеня (у тому числі й нескінченного) зобразити як багаточлен з найвищим степенем $n - 1$.

Оскільки e^{At} розкладається в нескінченний ряд, що збігається по At , то теорему Сильвестра можна безпосередньо використовувати для e^{At} , а не для нескінченного ряду по At . Таким чином,

$$e^{At} = \sum_{i=1}^n e^{\lambda_i t} t^{\frac{j \neq i}{i \neq j}} \frac{\prod_{j \neq i} (A - \lambda_j I)}{\prod_{i \neq j} (\lambda_i - \lambda_j)}, \quad (3.90)$$

де λ_i – характеристичні числа A .

Приклад 3.9. Знайти $\Phi(t) = e^{At}$, якщо

$$A = \begin{bmatrix} 1 & -3 \\ 1 & -1 \end{bmatrix}.$$

Розв'язання. Характеристичне рівняння знаходимо за формулою $\lambda I - A = 0$:

$$\begin{vmatrix} \lambda - 1 & 3 \\ -1 & \lambda + 1 \end{vmatrix} = 0.$$

Воно може бути зведене до вигляду

$$\lambda^2 + 2 = 0.$$

Звідси знаходимо характеристичні числа $\lambda_1 = j\sqrt{2}$; $\lambda_2 = -j\sqrt{2}$.

Підставивши в формулу (5.90) λ_1, λ_2 і A , одержимо

$$e^{At} = \frac{e^{j\sqrt{2}t}}{2\sqrt{2}j} \begin{bmatrix} 1+j\sqrt{2} & -3 \\ 1 & -1+j\sqrt{2} \end{bmatrix} - \frac{e^{j\sqrt{2}t}}{2\sqrt{2}j} \begin{bmatrix} 1-j\sqrt{2} & -3 \\ 1 & -1-j\sqrt{2} \end{bmatrix}.$$

Після перетворень матимемо

$$\Phi(t) = e^{At} = \begin{bmatrix} \cos\sqrt{2}t + \frac{1}{\sqrt{2}}\sin\sqrt{2}t & -\frac{3}{\sqrt{2}}\sin\sqrt{2}t \\ \frac{1}{\sqrt{2}}\sin\sqrt{2}t & \cos\sqrt{2}t - \frac{1}{\sqrt{2}}\sin\sqrt{2}t \end{bmatrix}$$

Метод Келі–Гамільтона. Нехай потрібно знайти значення матричного багаточлена $N(A)$, степінь якого більший за порядок n матриці A . Розділимо $N(A)$ на характеристичний багаточлен $P(A)$ матриці A :

$$\frac{N(A)}{P(A)} = Q(A) + \frac{R(A)}{P(A)}, \quad (3.91)$$

де $Q(A)$ – ціла частина результату поділу; $R(A)$ – залишковий член у вигляді багаточлена відносно A , степінь якого менша за n :

$$R(A) = a_1 + a_2A + a_3A^2 + \dots + a_nA^{n-1}.$$

Помноживши вираз (3.91) на $P(A)$, одержимо

$$N(A) = Q(A)P(A) + R(A). \quad (3.92)$$

Відповідно до теореми Келі–Гамільтона $P(A) = 0$, тому

$$N(A) = R(A). \quad (3.93)$$

Таким чином, матричний багаточлен $N(A)$ довільного степеня дорівнює матричному багаточлену $R(A)$, степінь якого менший за n . Слід зазначити, що за (1.90) маємо багаточлен $R(A)$ тільки у разі різних характеристичних чисел λ_i матриці A . Вираз (1.93) є слушним і для функції $\Phi(A)$, аналітичної в будь-якій області, тому що $\Phi(A)$ може бути зображена в ній у вигляді нескінченного збіжного за A ряду. Зокрема, цій умові задовольняє функція $\Phi(t) = e^{At}$. Тому матрицю переходу $\Phi(t) = e^{At}$ можна подати матричним багаточленом степеня $n-1$:

$$\Phi(t) = e^{At} = a_1 + a_2A + a_3A^2 + \dots + a_nA^{n-1}, \quad (3.94)$$

де невідомі коефіцієнти a_k $k=1, 2, \dots, n$ залежать від t .

Коефіцієнти a_k можна знайти, використовуючи властивість характеристичного багаточлена $P(\lambda)$. Якщо матрицю A в (5.92) замінити на λ , буде слушним наступне співвідношення:

$$N(\lambda) = Q(\lambda)P(\lambda) + R(\lambda). \quad (3.95)$$

При $\lambda = \lambda_i$ замість співвідношення (3.95), з урахуванням того, що $P(\lambda_i) = 0$ і $N(\lambda_i) = e^{\lambda_i t}$, можна записати n рівнянь вигляду

$$\begin{vmatrix} \lambda+1 & -1 & 0 \\ 0 & \lambda+3 & 0 \\ 0 & 0 & \lambda+2 \end{vmatrix} = (\lambda+1)(\lambda+2)(\lambda+3) = 0.$$

Корені цього рівняння: $\lambda_1 = -1$, $\lambda_2 = -2$, $\lambda_3 = -3$. Система рівнянь (3.96) матиме вигляд

$$\begin{aligned} e^{-t} &= a_1 - a_2 + a_3; \\ e^{-2t} &= a_1 - 2a_2 + 4a_3; \\ e^{-3t} &= a_1 - 3a_2 + 9a_3. \end{aligned}$$

Розв'язок цієї системи в матричному вигляді

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -2 & 4 \\ 1 & -3 & 9 \end{bmatrix}^{-1} \begin{bmatrix} e^{-t} \\ e^{-2t} \\ e^{-3t} \end{bmatrix}.$$

Знаходимо необхідну обернену матрицю:

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & -2 & 4 \\ 1 & -3 & 9 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & -3 & 1 \\ 2,5 & -4 & 1,5 \\ 0,5 & -1 & 0,5 \end{bmatrix},$$

тоді

$$\begin{aligned} a_1 &= 3e^{-t} - 3e^{-2t} + e^{-3t}; \\ a_2 &= 2,5e^{-t} - 4e^{-2t} + 1,5e^{-3t}; \\ a_3 &= 0,5e^{-t} - e^{-2t} + 0,5e^{-3t}. \end{aligned}$$

Підставляючи в рівняння $\Phi(t) = e^{At} = a_1I + a_2A + a_3A^2$ значення a_1 , a_2 , a_3 , A^2 , одержимо

$$\Phi(t) = \begin{bmatrix} e^{-t} & 0,5(e^{-t} - e^{-3t}) & 0 \\ 0 & e^{-3t} & 0 \\ 0 & 0 & e^{-2t} \end{bmatrix}.$$

Ця матриця збігається з матрицею, отриманою в прикладі 3.7.

Приклад 3.11. Визначити e^{At} , якщо

$$A = \begin{bmatrix} 0 & 1 & 3 \\ 6 & 0 & 2 \\ -5 & 2 & 4 \end{bmatrix}.$$

Розв'язання. З виразу $|\lambda I - A| = \lambda^3 - 4\lambda^2 - 5\lambda - 2 = (\lambda - 1)^2(\lambda - 2)$ випливає, що характеристичне рівняння містить два корені $\lambda_{1,2} = 1$ і один корінь $\lambda_3 = 2$. Система рівнянь (1.96) набере вигляду

$$\begin{aligned} e^t &= a_1 + a_2 + a_3; \\ e^{2t} &= a_1 + 2a_2 + 4a_3. \end{aligned}$$

Третє рівняння знайдемо з умови (3.86):

$$te^t = a_2 + 2a_3.$$

Розв'язок системи цих трьох рівнянь

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 0 & 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} e^t \\ e^{2t} \\ te^t \end{bmatrix}.$$

Обернена матриця знаходиться за відомим правилом:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 0 & 1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & -2 \\ 2 & -2 & 3 \\ -1 & 1 & 1 \end{bmatrix}.$$

Тому

$$\begin{aligned} a_1 &= e^{2t} - 2te^t; \\ a_2 &= 2e^t - 2e^{2t} + 3te^t; \\ a_3 &= -e^t + e^{2t} - te^t. \end{aligned}$$

Тоді

$$e^{At} = a_1 I + a_2 A + a_3 A^2 = a_1 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 & 3 \\ 6 & 0 & 2 \\ -5 & 2 & 4 \end{bmatrix} + a_3 \begin{bmatrix} -9 & 6 & 14 \\ -10 & 10 & 26 \\ -8 & 3 & 5 \end{bmatrix},$$

або

$$e^{At} = \begin{bmatrix} a_1 - 9a_3 & a_2 + 6a_3 & 3a_2 + 14a_3 \\ 6a_2 - 10a_3 & a_1 + 10a_3 & 2a_2 - 26a_3 \\ -5a_2 - 8a_3 & 2a_2 + 3a_3 & a_1 + 4a_2 + 5a_3 \end{bmatrix}.$$

Підставивши в цей вираз a_1 , a_2 і a_3 , остаточно одержуємо

$$e^{At} = \begin{bmatrix} 9e^t - 8e^{2t} + 7te^t & -4e^t + 4e^{2t} - 3te^t & -8e^t + 8e^{2t} - 5te^t \\ 22e^t - 22e^{2t} + 28te^t & -10e^t + 11e^{2t} - 12te^t & -22e^t + 22e^{2t} - 20te^t \\ -2e^t + 2e^{2t} - 7te^t & e^t + e^{2t} - 3te^t & 3e^t - 2e^{2t} + 5e^t \end{bmatrix}.$$

Порівняння методів обчислення матриці переходу. Кожний з розглянутих методів обчислення матриці переходу має і переваги, і недоліки. Тому доцільно провести порівняльну оцінку для визначення області їхнього застосування.

Метод розкладу в нескінченний ряд здається найбільш простим, проте існують труднощі пошуку суми нескінченного ряду при визначенні в аналітичному зображенні елементів матриці переходу. Цей метод доцільно застосовувати при порівняно невеликих порядках системи (матриці A) і в тих випадках, коли більшість елементів матриці A нульові.

Метод комплексної площини вимагає визначення оберненої матриці, елементами якої є функції від комплексної змінної p . Крім того, необхідно знаходити, користуючись таблицями, обернені перетворення Лапласа від дробово-раціональних функцій. Ці операції практично здійсненні при розв'язанні простих задач.

Метод Сильвестра і метод Келі–Гамільтона майже ідентичні щодо труднощів і операцій, які треба виконувати для обчислення матриці переходу. Обидва методи потребують визначення характеристичних чисел A , що є непростою задачею.

Отже, кожен метод має свої труднощі, і вибір одного з них для розв’язання практичної задачі ґрунтується в значній мірі на інтуїції дослідника. Слід зазначити, що практичне використання аналітичних методів визначення Φt можливе тільки при розв’язанні відносно простих задач. Із зростанням порядку системи обчислювальні труднощі одержання Φt зростають катастрофічно, і звичайно аналітичний підхід стає неприйнятним. Звідси виникає необхідність застосування цифрових обчислювальних машин і розробки алгоритмів обчислення Φt . Однак існує один випадок, коли порядок системи не обмежує можливостей аналітичного підходу до дослідження систем керування. Якщо матриця A діагональна, тобто диференціальні рівняння розв’язані щодо змінних стану, то матриця переходу Φt може бути отримана без яких-небудь обчислень. Для діагональної матриці

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad (3.99)$$

матриця переходу буде також діагональною:

$$\Phi(t) = e^{At} = \begin{bmatrix} e^{\lambda_1 t} & 0 & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & 0 & \dots & 0 \\ 0 & 0 & e^{\lambda_3 t} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & e^{\lambda_n t} \end{bmatrix} \quad (3.100)$$

Дійсно, якщо розкласти e^{At} у нескінченний ряд

$$e^{At} = 1 + At + \frac{A^2 t^2}{2!} + \dots + \frac{A^k t^k}{k!} + \dots, \quad (3.101)$$

то кожен доданок у правій частині ряду (3.101) буде матрицею діагонального вигляду, оскільки

$$A^k = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_n \end{bmatrix}^k = \begin{bmatrix} \lambda_1^k & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^k & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^k & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_n^k \end{bmatrix}, \quad (3.102)$$

що можна показати безпосереднім множенням. Сума діагональних матриць у виразі (3.101) дає діагональну матрицю e^{At} . Вигляд цієї матриці можна одержати безпосереднім підсумовуванням елементів матричного ряду (3.101).

Як показано вище, матриця A буде діагональною або в результаті побудови схеми в змінних стану шляхом розкладання передавальної функції на елементарні дроби, або шляхом перетворення рівнянь стану до *нормальної форми*. Перший шлях придатний для систем з одним входом і одним виходом; другий шлях має труднощі, пов’язані з одержанням характеристичних векторів і діагоналізацією матриць. Одночасно не слід забувати, що процедура зведення рівнянь

стану до нормальної форми набагато ускладнюється у разі наявності кратних коренів, і вигляд матриці $\Phi(t)$ у цьому випадку відрізняється від матриці (3.100). Безпосереднє застосування схеми в змінних стану, що дає рівняння в нормальній формі, можливе і при кратних характеристичних числах, однак нові змінні стану втрачають фізичний зміст, і тільки шляхом перетворень їх можна пов'язати з реальними змінними системи, доступними для спостереження і виміру.

Особливості застосування методу простору станів у зв'язку. Важливою особливістю методу змінних станів є можливість безпосереднього моделювання систем, описуваних рівняннями стану, за допомогою аналогового чи цифрового обчислювального пристрою.

Застосування методу змінних стану обумовлюється можливістю конструктивного опису випадкових процесів. Воно полягає в тому, що випадковий процес Y із заданими ймовірнісними характеристиками подається як вихід деякої динамічної системи, збудованої іншим випадковим процесом з більш простою ймовірнісною структурою $U(t)$. Звичайно в якості породного використовують гауссів процес $U(t)$ типу білого шуму з нульовим середнім і кореляційною функцією

$$BU(t, t + \tau) = M U(t)U'(t + \tau) = Q\delta(\tau),$$

де Q – симетрична невід'ємно визначена матриця, а штрих позначає транспонування.

Метод змінних стану з успіхом застосовують і для опису стохастичних ланцюгів (каналів) з випадково змінними параметрами. Для цього деякі елементи системних функцій (матриць) A, B, C слід розглядати як випадкові функції. Крім того, в рівнянні моделі каналу треба враховувати компоненту адитивних завад.

Метод змінних стану дає універсальний підхід до моделювання (у рамках дуже широкої марковської моделі) каналів передавання інформації (систем зв'язку) для всіляких повідомлень, способів кодування і модуляції (як лінійної, так і нелінійної), ліній зв'язку з детермінованими і випадковими параметрами, розсіюванням сигналів, адитивними шумами (як гауссовими, так і негауссовими). Більш істотною є та обставина, що, зображуючи спостережувані (аналізовані в місці приймання) випадкові марковські процеси за допомогою диференціальних рівнянь, можна розв'язати обернену задачу, тобто одержати диференціальні рівняння для оцінки повідомлень, які описуються цими процесами. Такі оцінки за допомогою аналогової чи цифрової техніки значно простіші, ніж оцінки, що впливають з інтегральних рівнянь.

Контрольні запитання для самооцінки рівня знань

1. Сформулюйте означення дискретно-неперервних і аналогових каналів зв'язку.
2. У чому полягає різниця між лінійними і нелінійними каналами зв'язку?
3. Дайте означення ідеального каналу зв'язку без завад.
4. У чому полягають особливості каналу з адитивним гауссовим шумом?
5. Які особливості має канал з невизначеною фазою сигналу?
6. Охарактеризуйте однопроменевий гауссів канал із загальними завмираннями (флуктуаціями амплітуд і фаз сигналів).
7. У чому полягають особливості каналу з міжсимвольною інтерференцією (МСІ) і адитивним шумом?
8. Які моделі дискретного каналу зв'язку Ви знаєте?
9. Які особливості має симетричний канал без пам'яті?
10. Дайте означення симетричного дискретного каналу без пам'яті зі стиранням.
11. Які особливості має несиметричний дискретний канал зв'язку без пам'яті?
12. Дайте означення марковського дискретного каналу.
13. Які особливості має дискретний канал з адитивним дискретним шумом?
14. У чому полягає метод простору станів?
15. Які методи складання рівнянь стану динамічних систем Ви знаєте?
16. У чому полягає метод комбінування похідних?
17. У чому полягає метод послідовного інтегрування?

18. Поясніть метод розкладу передавальної функції на елементарні дроби вже складених рівнянь стану.
19. Як записується нормальна форма рівнянь стану?
20. Як здійснюється перетворення безпосередніх рівнянь стану в однорідні?
21. Як будуються схеми в змінних стану для типових вхідних впливів: поліноміального, гармонічного, експоненціального?
22. Які форми розв'язання рівнянь стану Ви знаєте?
23. Які існують аналітичні підходи до обчислення матриці?
24. У чому полягають особливості методу Сильвестра?
25. У чому полягають особливості методу Келі–Гамільтона?

Післямова до розділу 3

Ви завершили вивчення загальних відомостей про канали зв'язку.

Тепер ви вже знаєте:

- що канал зв'язку – це сукупність засобів, призначених для передавання сигналів (повідомлень) між різними точками (пунктами) системи зв'язку. Канал зв'язку можна уявити як послідовне з'єднання пристроїв (блоків), що виконують різні функції в загальній системі зв'язку;
- що під засобом розуміють і *технічні пристрої*, і *лінію зв'язку* – фізичне середовище, в якому поширюється сигнал між пунктами зв'язку;
- *за призначенням систем* канали зв'язку поділяють на телеграфні, фототелеграфні, телефонні, звукового мовлення, передавання даних, телевізійні, телеметричні, змішані і т. ін. Залежно від того, поширюються сигнали між пунктами зв'язку у *вільному просторі* чи по *напрямних лініях*, виділяють канали радіозв'язку (зокрема, космічні канали) і провідного зв'язку (повітряні, кабельні, волоконно-оптичні лінії зв'язку, хвилевідні надвисокочастотні тракти тощо);
- що розрізняють канали чисто часові (із зосередженими параметрами), в яких сигнали на вході і виході описуються функціями одного скалярного параметра (часу t), і просторово-часові канали (з розподіленими параметрами), в яких сигнали на вході і (чи) виході описуються функціями не одного, а кількох скалярних параметрів (наприклад, часу t і просторових координат x, y, z), такі сигнали називають *полями*;
- про математичні моделі каналів зв'язку;
- про методи простору станів.

Поняття, які необхідно знати і розуміти для сприйняття матеріалу **розділу 4**:

Теорія ймовірностей – це математична наука, що вивчає закономірності випадкових явищ (подій). Основні поняття теорії ймовірності.

Розділ 4 розкриває інформаційну сторону про теорію ймовірностей, закони розподілу випадкових величин, особливості випадкових процесів.

ЕЛЕМЕНТИ ТЕОРІЇ ЙМОВІРНОСТЕЙ І ВИПАДКОВІ ПРОЦЕСИ

4.1. ПРЕДМЕТ ТЕОРІЇ ЙМОВІРНОСТЕЙ

Теорія ймовірностей – математична наука, що вивчає закономірності випадкових явищ (подій).

Подія називається випадковою в даних умовах, якщо при виконанні цих умов (як кажуть, при випробуванні) вона може відбутися, а може не відбутися.

Приклад 4.1. Тіло кілька разів зважується на аналітичних вагах; результати повторних зважувань трохи відрізняються одне від одного. Ці розходження обумовлені впливом багатьох другорядних факторів, що супроводжують операцію зважування, таких як положення тіла на чашці ваг, випадкові вібрації апаратури, похибки відліку показань приладу і т. д.

Приклад 4.2. Літак виконує політ на заданій висоті; теоретично він летить горизонтально, рівномірно і прямиoliniйно. Фактично політ супроводжується відхиленням центра маси літака від теоретичної траєкторії і коливаннями літака навколо центра маси. Ці відхилення і коливання є випадковими і пов'язані з турбулентністю атмосфери; раз у раз вони не повторюються.

У наведених прикладах підкреслені однакові результати ряду випробувань (дослідів), основні умови яких залишаються незмінними. Ці випадкові варіації завжди супроводжуються наявністю другорядних факторів, що впливають на результат випробування. Вони змінюються кожного разу і вносять випадкові розбіжності в результати випробувань.

Очевидно, що в природі немає жодного фізичного явища без елементів випадковості. Як би точно і детально не були зафіксовані умови досліду, неможливо досягти того, щоб при його повторенні результати цілком збіглися.

Проте в ряді практичних задач цими випадковими елементами можна знехтувати, замінивши реальне явище на його спрощену схему – “модель” – і припустивши, що в даних умовах досліду явище протікає цілком визначено. Застосовуючи такий підхід для розв'язання задачі, насамперед виділяють основні умови, що враховуються, і з'ясовують, на які параметри задачі вони впливають; потім за допомогою того чи іншого математичного апарату явища описують (наприклад, диференціальними рівняннями). Отже, саме таким чином виявляється основна закономірність, властива даному явищу, за якою можна передбачити результат досліду при його заданих умовах. З розвитком науки кількість факторів, що враховуються, зростає; явище досліджується докладніше; науковий прогноз стає точнішим.

Приклад 4.3. Деякий технічний пристрій, наприклад система автоматичного управління, розв'язує певну задачу в умовах, коли на систему неперервно діють випадкові завади. Вплив завад призводить до похибки в розв'язку задачі, можливо, більшої за допустиму. Виникають питання: як часто з'являтимуться такі похибки? які заходи слід вжити для того, щоб практично виключити їх?

Щоб відповісти на це, потрібно дослідити природу і структуру випадкових збурень, що діють на систему, вивчити реакцію системи на них, з'ясувати вплив конструктивних параметрів системи на вид цієї реакції.

Подібні задачі, яких у фізиці й техніці надзвичайно багато, потребують вивчення не тільки основних, головних закономірностей, що визначають явище в загальних рисах, але й аналізу випадкових збурень і спотворень, пов'язаних з наявністю другорядних факторів, що надають результату досліду при заданих умовах невизначеності.

Які ж існують шляхи і методи дослідження випадкових явищ?

З теоретичної точки зору фактори, які умовно називають *випадковими*, нічим не відрізняються від інших – *основних*. Можна необмежено підвищувати точність розв'язку кожної задачі з огляду на групи факторів від самих істотних до найнезначніших. Однак така спроба однаково ретельного аналізу впливу на явище всіх факторів призвела б тільки до того, що розв'язання задачі через непомірну громіздкість і складність стало б практично нездійсненним, а до того ж не мало б ніякої пізнавальної цінності.

Очевидно, має існувати принципова різниця в методах урахування тих факторів, що головним чином визначають перебіг події, і вторинних, другорядних факторів, що спотворюють або “збурюють” його. Елемент невизначеності, складності, багатопричинності, властивий випадковим явищам, вимагає створення спеціальних методів для їх вивчення.

Саме такі методи й розробляються в теорії ймовірностей; її предметом є специфічні закономірності, що спостерігаються у випадкових явищах.

Практика показує, що серед великої кількості однорідних випадкових явищ в них спостерігаються цілком визначені стійкі закономірності.

Наприклад, якщо багато разів підкидати монету, частота появи герба (відношення кількості гербів до загальної кількості підкидань) поступово стабілізується, наближаючись до $\frac{1}{2}$. Така сама властивість “стійкості частоти” спостерігається і в багаторазовому повторенні будь-якого іншого експерименту, наслідок якого ми уявляємо собі заздалегідь невизначеним, випадковим. Так, при збільшенні числа пострілів частота влучання в ціль теж стабілізується, наближаючись до деякого постійного числа.

Приклад 4.4. У посудині знаходиться газ, що складається з дуже великого числа молекул. Кожна молекула за секунду безліч разів зіштовхується з іншими молекулами, багаторазово змінює швидкість і напрямок руху; траєкторія кожної молекули випадкова. Відомо, що тиск газу на стінку посудини обумовлений сукупністю ударів молекул об неї. Здавалося б, якщо траєкторія кожної окремої молекули випадкова, то й тиск на стінку посудини має змінюватися випадково, неконтрольовано; проте це не так. Якщо число молекул велике, то тиск газу практично не залежить від траєкторій окремих молекул і підпорядковується цілком визначеній і простій закономірності. Випадкові особливості, властиві руху кожної окремої молекули, у масі взаємно компенсуються; внаслідок цього, незважаючи на складність і заплутаність окремого випадкового явища, ми одержуємо закономірність, слухну для великої кількості випадкових явищ того самого типу. Підкреслимо, що саме *масовість* випадкових явищ забезпечує виконання цієї закономірності; при обмеженій кількості молекул починаються випадкові відхилення від закономірності – *флуктуації*.

Подібні специфічні закономірності, що називаються статистичними, існують завжди, коли ми маємо справу з масою однорідних випадкових явищ, і практично не залежать від індивідуальних особливостей окремих випадкових явищ. Ці особливості в масі, образно кажучи, взаємно нівелюються, і середній результат великої кількості випадкових явищ практично вже не є випадковим. Багаторазово підтверджена випробуванням стійкість масових випадкових явищ і є базою для застосування ймовірнісних (статистичних) методів дослідження. Методи теорії ймовірностей за своєю природою пристосовані тільки для дослідження масових випадкових явищ; вони не дають можливості передбачити результат окремого випадкового явища, але дозволяють передбачити середній сумарний результат маси однорідних випадкових явищ або аналогічних дослідів. Отже, чим більша кількість однорідних випадкових явищ, тим чіткіше проявляються властиві їм специфічні закони, тим з більшою впевненістю і точністю можна здійснювати науковий прогноз.

Мета застосування ймовірнісних методів дослідження – уникнення занадто складного (і найчастіше практично неможливого) вивчення окремого явища, залежного від великої кількості факторів, і звернення до законів, що керують масами випадкових явищ. Вивчення цих законів дозволяє не тільки здійснювати науковий прогноз випадкових явищ, але й допомагає цілеспрямовано впливати на них, обмежуючи сферу дії випадковості.

Ймовірнісний (або статистичний) метод у науці не протиставляє себе класичним, звичайним методам точних наук, а доповнює їх, даючи змогу глибше аналізувати явища з урахуванням елементів випадковості.

Широке і плідне застосування статистичних методів у всіх галузях знань цілком природне, тому що при поглибленому вивченні будь-якого кола явищ неминуче настає етап, коли потрібно не тільки виявити основні закономірності, але й проаналізувати можливі відхилення від них. В

одних науках, через специфіку предмета й історичних умов, впровадження статистичних методів спостерігається раніше, в інших – пізніше. На сьогодні вже цілі розділи сучасної фізики (зокрема, ядерна фізика) базуються на методах теорії ймовірностей. Усе ширше застосовуються ймовірнісні методи в сучасній електротехніці і радіотехніці, метеорології й астрономії, теорії автоматичного управління і машинній математиці.

Математичні закони теорії ймовірностей є відображенням реальних статистичних законів, що об'єктивно існують у масових випадкових явищах природи. Для вивчення цих явищ теорія ймовірностей застосовує метод, який є одним з розділів математики.

4.2. ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ЙМОВІРНОСТІ

Кожна наука, що розвиває загальну теорію певного кола явищ, містить ряд основних понять, на яких вона базується: у геометрії – це поняття крапки, прямої, лінії; у механіці – сили, маси, швидкості, прискорення і т. д.

Основні поняття існують і в теорії ймовірностей [8]. Перше з таких понять – подія.

Під подією в теорії ймовірностей розуміють усякий факт, що в результаті досліду може відбутися чи не відбутися. Наведемо кілька прикладів подій: A – поява герба при киданні монети, B – поява гербів при триразовому киданні монети; C – влучання в ціль при пострілі.

Кожна з таких подій має той чи інший ступінь можливості.

Щоб кількісно порівняти між собою події за ступенем їхньої можливості, очевидно, потрібно з кожною подією пов'язати якесь число, котре має бути тим більшим, чим більш можлива подія. Таке число називається *ймовірністю події*. Отже, ймовірність події – кількісна характеристика, міра можливості цієї події.

Зазначимо, що вводячи поняття ймовірності події, ми пов'язуємо з ним певний практичний зміст, а саме: на підставі досліду ми вважаємо більш імовірними ті події, що відбуваються частіше; менш імовірними – події, що відбуваються рідше; малоімовірними – ті, котрі майже ніколи не відбуваються. Таким чином, поняття ймовірності події пов'язане з *дослідним*, практичним поняттям частоти події.

Щоб порівняти між собою різні події за ступенем їхньої можливості, ми повинні мати одиницю виміру. За таку одиницю виміру природно взяти *ймовірність вірогідної події*, тобто такої, що в результаті досліду неодмінно має відбутися. Приклад вірогідної події – випадання не більше 6 очок при киданні одного грального кубика.

Якщо приписати вірогідній події ймовірність, рівну одиниці, то всі інші події (можливі, але не вірогідні) будуть характеризуватися ймовірностями, меншими за одиницю, тобто складатимуть якусь частку одиниці.

Протилежністю вірогідної події є *неможлива подія*, тобто така, що в певних умовах досліду не може виникнути. Приклад неможливої події – поява 12 очок при киданні одного грального кубика. Природно приписати неможливій події ймовірність, яка дорівнює нулю.

Отже, діапазон виміру ймовірностей будь-яких подій – від 0 до 1.

Кілька подій утворюють *повну групу подій*, якщо в результаті досліду неодмінно має з'явитися хоча б одна з них. Прикладами подій, що утворюють повну групу, є влучання і промах при пострілі; поява 1, 2, 3, 4, 5, 6 очок при киданні грального кубика; випадання герба і цифри при киданні монети.

Кілька подій називають *несумісними*, якщо при виконанні експерименту ніякі дві з них не можуть виникнути разом. Приклади несумісних подій: одночасне випадання герба і цифри при киданні монети; влучання і промах при одному пострілі; поява 1, 3, 4 очок при одному киданні грального кубика.

Кілька подій у даному досліді називаються *рівноможливими*, якщо за умовами симетрії є підстава вважати, що жодна з цих подій не є об'єктивно більш можливою, чим інша. Приклади рівноможливих подій: випадання герба і випадання цифри при киданні монети; поява 1, 3, 4, 5 очок на кинутому гральному кубіку; поява карти бубнової, чирвової, трєфової масті при вийманні карти з колоди.

Існують події, що мають усі три властивості, тобто вони утворюють повну групу, несумісні та рівноможливі. Такі події називаються *випадками*.

Якщо який-небудь дослід за своєю структурою має симетрію можливих закінчень, то випадки являють собою вичерпну систему рівноможливих і вилучаючих одне одного закінчень дослідів. Про такий дослід говорять, що він зводиться до *схеми випадків*.

Схема випадків переважно має місце в штучно організованих дослідах, яким заздалегідь і свідомо забезпечена однакова можливість їх закінчень (як, наприклад, в азартних іграх). Для таких дослідів можливий безпосередній підрахунок імовірностей, заснований на оцінці частки так званих “сприятливих” випадків у загальному числі випадків.

Випадок називається *сприятливим* для деякої події, якщо його відбуття спричиняє появу даної події. Наприклад, при киданні грального кубика можливі шість випадків: поява 1, 2, 3, 4, 5, 6 очок. Із них появи парного числа очок (події A) сприятливі три випадки – 2, 4, 6, і не сприятливі три інші.

Якщо дослід звести до схеми випадків, то ймовірність події A можна оцінити за відносною часткою сприятливих випадків. Імовірність події A обчислюється відношенням числа випадків m , сприятливих для події A , до загального числа випадків n :

$$P(A) = \frac{m}{n}. \quad (4.1)$$

Оскільки число сприятливих випадків завжди лежить між 0 і n (0 – для неможливої, n – для вірогідної події), то ймовірність події, обчислена за формулою (4.1), завжди є раціональним дробом:

$$0 \leq P(A) \leq 1. \quad (4.2)$$

Формула (4.1), або “класична” формула, довгий час фігурувала як означення ймовірності. Нині ж при означенні ймовірності звичайно виходять з інших принципів, безпосередньо пов’язуючи поняття ймовірності з емпіричним поняттям частоти; формула (4.1) зберігається лише для безпосереднього обчислення ймовірностей і застосовується тоді і тільки тоді, коли дослід зводиться до схеми випадків, тобто має симетрію можливих закінчень.

Приклад 4.5. В урні знаходяться 2 білі і 3 чорні кулі. З урни навмання виймають одну кулю. Потрібно знайти ймовірність того, що ця куля буде білою.

Розв’язання. Позначимо через A подію, що полягає в появі білої кулі. Загальне число випадків $n = 5$; число випадків, сприятливих події A , $m = 2$, отже $P(A) = 2/5$.

Приклад 4.6. В урні знаходяться a білих і b чорних куль. З урни виймають дві кулі. Знайти ймовірність того, що обидві кулі будуть білими.

Розв’язання. Позначимо через B подію, що складається з появи білих куль. Підрахуємо число можливих випадків n і число випадків m , сприятливих події B :

$$n = C_{a+b}^2; \quad m = C_a^2;$$

отже,

$$P(B) = C_a^2 / C_{a+b}^2.$$

Імовірності великого класу подій не можна обчислити за формулою (4.1), оскільки вони не зводяться до схеми випадків. У такому разі застосовують інші способи визначення ймовірностей, і всі вони ґрунтуються на досліді, експерименті. Щоб скласти уявлення про ці способи, необхідно усвідомити поняття частоти події і специфіку органічного зв’язку між ймовірністю і частотою.

Якщо виконано серію з n дослідів, у кожному з яких могла з'явитися чи не з'явитися деяка подія A , то частотою події A у даній серії дослідів називається відношення числа дослідів, у яких з'явилася подія A , до загального числа зроблених дослідів.

Частоту події часто називають його *статистичною ймовірністю* (на відміну від раніше введеної *математичної ймовірності*).

Домовимося частоту (статистичну ймовірність) події A позначати через $P^*(A)$. Частота події обчислюється на підставі результатів досліду за формулою

$$P^*(A) = \frac{m}{n}, \quad (4.3)$$

де m – число появ події A ; n – загальне число дослідів.

При невеликому числі дослідів частота події носить значною мірою випадковий характер і може помітно змінюватися від однієї групи дослідів до іншої. Однак із збільшенням числа дослідів частота події все більше втрачає свою випадковість: випадкові обставини, властиві кожному окремому досліду, у масі взаємно гасяться з незначними коливаннями навколо деякої середньої, сталої величини.

Властивість “стійкості частоти”, багаторазово перевірена експериментально і підтверджена всім досвідом практичної діяльності людства, є однією з найбільш характерних закономірностей, що спостерігаються у випадкових явищах.

Математичне формулювання цієї закономірності вперше дав Я. Бернуллі у теоремі, що являє собою найпростішу форму закону великих чисел. Він довів, що *при необмеженому збільшенні числа однорідних незалежних дослідів з практичною вірогідністю можна стверджувати, що частота події буде як завгодно мало відрізнятися від її ймовірності в окремому досліді*.

Ймовірність неможливої події, яка дорівнює нулю, і ймовірність вірогідної події, яка дорівнює одиниці, займають крайні положення за шкалою ймовірностей. Але на практиці часто доводиться мати справу не з неможливими і вірогідними подіями, а з *практично неможливими і практично вірогідними* подіями.

Практично неможливою називається подія, ймовірність якої не точно дорівнює нулю, а дуже близька до нуля.

Практично вірогідною називається подія, ймовірність якої не точно дорівнює одиниці, а дуже близька до одиниці.

Якщо подія A при даному випробуванні практично неможлива, то протилежна їй подія \bar{A} , що полягає у невиконанні події A , є практично вірогідною. Тому, з погляду теорії ймовірностей, не має значення, які події розглядати: практично неможливі чи практично вірогідні, тому що вони завжди супроводжують одна одну.

Практично неможливі і практично вірогідні події відіграють важливу роль у теорії ймовірностей; на них ґрунтується все практичне застосування цієї науки.

Випадкові величини. Одним з найважливіших основних понять теорії ймовірностей є поняття про випадкову величину.

Випадковою називається величина, що внаслідок досліду може набути того чи іншого значення, причому невідомо заздалегідь, якого саме.

Прикладами випадкових величин є число влучень при трьох пострілах; число викликів, що надійшли на телефонну станцію за добу. У даному разі випадкові величини можуть набувати окремих, ізольованих значень, які можна заздалегідь перелічити.

Такі випадкові величини, що набувають тільки відокремлених одне від одного значень, які можна заздалегідь перелічити, називаються *перервними* чи *дискретними* випадковими величинами.

Існують випадкові величини іншого типу, наприклад, абсциса точки влучення при пострілі; похибка зважування тіла на аналітичних вагах; швидкість літального апарата в момент досягнення заданої висоти; вага навмання взятого зерна пшениці. Можливі значення таких випадкових величин, не відокремлених одне від одного, заповнюють декотрий проміжок, який іноді має різко

означені границі, але частіше – невизначені, розплинні. Такі випадкові величини, можливі значення яких неперервно заповнюють деякий проміжок, називаються *неперервними* випадковими величинами.

Виходячи з будь-яких даних подій, можна визначити дві нові події: їх суму (об'єднання) та добуток (суміщення).

Подію B називають сумою (об'єднанням) подій A_1, A_2, \dots, A_n , якщо вона має вигляд: “відбувається або A_1 , або A_2, \dots , або A_n ”.

Подію C називають добутком (суміщенням) подій A_1, A_2, \dots, A_n , якщо вона має вигляд: “відбувається і A_1 , і A_2, \dots , і A_n ”.

Позначимо суму імовірностей двох подій A і B через $P(A+B)$, а добуток імовірностей цих подій – через $P(AB)$.

Теорема додавання ймовірностей. Імовірність суми (об'єднання) двох *несумісних* подій A і B дорівнює сумі ймовірностей цих подій:

$$P(A+B) = P(A) + P(B) . \quad (4.4)$$

Узагальнюючи теорему додавання на довільне число несумісних подій, одержимо

$$P\left(\sum_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j) . \quad (4.5)$$

З теореми додавання ймовірностей випливають два наслідки.

Наслідок 1. Якщо події A_1, A_2, \dots, A_n утворюють повну групу несумісних подій, то сума їх імовірностей дорівнює одиниці:

$$\sum_{j=1}^n P(A_j) = 1 . \quad (4.6)$$

Наслідок 2. Сума ймовірностей протилежних подій дорівнює одиниці:

$$P(A) + P(\bar{A}) = 1 , \quad (4.7)$$

де протилежними подіями називаються дві несумісні події, що утворюють повну групу, наприклад: A – влучення при пострілі; \bar{A} – промах при пострілі; B – випадання герба при киданні монети; \bar{B} – випадання цифри при киданні монети.

Цей наслідок є окремим випадком наслідку 1. Він виокремлений через його велике значення в практичному застосуванні теорії ймовірностей: дуже часто легше обчислити ймовірність протилежної події \bar{A} , ніж імовірність прямої події A . У цих випадках обчислюють $P(\bar{A})$ і знаходять $P(A) = 1 - P(\bar{A})$.

Як уже зазначалося, теорема додавання ймовірностей (4.4) слухна тільки для несумісних подій. У випадку, коли події A і B *сумісні*, ймовірність суми їх виражається формулою

$$P(A+B) = P(A) + P(B) - P(AB) . \quad (4.8)$$

Аналогічно ймовірність суми *трьох сумісних подій* обчислюється за формулою

$$P(A+B+C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) . \quad (4.9)$$

Методом повної індукції можна довести загальну формулу для ймовірності суми будь-якого числа сумісних подій:

$$P\left(\sum_{j=1}^n A_j\right) = \sum_j P(A_j) - \sum_{i,j} P(A_i A_j) + \sum_{i,j,k} P(A_i A_j A_k) - \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n), \quad (4.10)$$

де суми поширюються на різні значення індексів $i; i, j; i, j, k$ і т. д.

Формула (4.10) виражає ймовірність суми будь-якого числа подій через ймовірності добутоків цих подій, узятих по одній, по дві, по три і т. д.

Аналогічні формули можна записати для добутку подій. З формул (4.8) і (4.9) випливає, що

$$P(AB) = P(A) + P(B) - P(A + B); \quad (4.11)$$

$$P(ABC) = P(A) + P(B) + P(C) - P(A + B) - P(A + C) - P(B + C) + P(A + B + C). \quad (4.12)$$

Загальна формула, що виражає ймовірність добутку довільного числа подій через ймовірності сум цих подій, узятих по одній, по дві, по три і т. д., має вигляд

$$P(A_1 A_2 \dots A_n) = \sum_i P(A_i) - \sum_{i,j} P(A_i + A_j) + \sum_{i,j,k} P(A_i + A_j + A_k) - \dots + (-1)^{n-1} P(A_1 + \dots + A_n) \quad (4.13)$$

Формули (4.10) і (4.13) знаходять практичне застосування при перетворенні виразів, що містять ймовірності сум і добутоків подій. У залежності від специфіки задачі в деяких випадках зручніше користуватися тільки сумами, а в інших – тільки добутками подій: для перетворення одних в інші і служать подібні формули.

Умовна ймовірність. Подія A називається *незалежною від події B* , якщо ймовірність події A не залежить від того, відбулася подія B чи ні.

Подія A називається *залежною від події B* , якщо ймовірність події A змінюється в залежності від того, відбулася подія B чи ні.

Ймовірність події A , обчислена за умови, що мала місце інша подія B , називається *умовною ймовірністю події A* і позначається

$$P(A/B).$$

Умову незалежності події A від B можна записати так:

$$P(A/B) = P(A),$$

а умову залежності –

$$P(A/B) \neq P(A).$$

Теорема множення ймовірностей. Ймовірність добутоків двох подій дорівнює добутку ймовірності одної з них на умовну ймовірність іншої, обчислену за умови, що перша мала місце:

$$P AB = P A P B/A \quad (4.14)$$

Теорема множення ймовірностей може бути узагальнена на випадок довільного числа подій і формулюється так.

Ймовірність добутку декількох подій дорівнює добутку ймовірностей цих подій, причому ймовірність кожної наступної по порядку події обчислюється за умови, що всі попередні мали місце:

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2 / A_1) P(A_3 / A_1 A_2) \dots P(A_n / A_1 A_2 \dots A_{n-1}). \quad (4.15)$$

У випадку незалежних подій теорема спрощується і набуває вигляду:

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2) \dots P(A_n),$$

тобто ймовірність добутку незалежних подій дорівнює добутку ймовірностей цих подій. Застосовуючи знак добутку, теорему можна записати так:

$$P\left(\prod_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i). \quad (4.16)$$

Формула повної ймовірності. Із теореми додавання і множення ймовірностей випливає формула повної ймовірності. Нехай потрібно визначити ймовірність деякої події A , що може відбутися разом з однією з подій H_1, H_2, \dots, H_n , які утворюють повну групу несумісних подій. Будемо ці події називати *гіпотезами*.

У цьому випадку

$$P(A) = \sum_{i=1}^n P(H_i) P(A / H_i), \quad (4.17)$$

тобто ймовірність події A обчислюється як сума добутків ймовірності кожної гіпотези на ймовірність події при цій гіпотезі.

Формула (4.17) називається *формулою повної ймовірності*.

Теорема гіпотез (формула Бейєса). Наслідком теореми множення і формули повної ймовірності є теорема гіпотез або формула Бейєса.

Поставимо наступну задачу. Маємо повну групу несумісних гіпотез H_1, H_2, \dots, H_n . Ймовірності цих гіпотез до проведення досліду відомі: $P(H_1), P(H_2), \dots, P(H_n)$. Проведемо дослід, в якому спостерігатимемо появу деякої події A . Питання: як потрібно змінити ймовірності гіпотез у зв'язку з появою цієї події?

Іншими словами, мова йде про те, щоб знайти умовну ймовірність $P(H_i / A)$ для кожної гіпотези.

З теореми множення маємо

$$P(AH_i) = P(A)P(H_i / A) = P(H_i)P(A / H_i), \quad i = \overline{1, n},$$

Звідки

$$P(H_i / A) = \frac{P(H_i)P(A / H_i)}{P(A)}, \quad i = \overline{1, n}.$$

Підставивши в цю формулу вираз $P(A)$ з формули повної ймовірності (4.17), маємо

$$P(H_i / A) = \frac{P(H_i)P(A/H_i)}{\sum_{i=1}^n P(H_i)P(A/H_i)}, \quad i = \overline{1, n}. \quad (4.18)$$

Формула (4.18) називається *формулою Бейсса* або *теоремою гіпотез*.

Приклад 4.7. Прилад може збиратися з високоякісних деталей і з деталей звичайної якості; відомо, що близько 40% приладів збираються з високоякісних деталей. Якщо прилад зібраний з високоякісних деталей, його надійність (ймовірність безвідмовної роботи) за час τ дорівнює 0,95, а якщо з деталей звичайної якості, – 0,7. Прилад випробувався протягом часу t і працював безвідмовно. Знайти ймовірність того, що він зібраний з високоякісних деталей.

Розв'язання. За умовами задачі можливі дві гіпотези: H_1 – прилад зібраний з високоякісних деталей; H_2 – прилад зібраний з деталей звичайної якості. Ймовірності цих гіпотез до випробування приладу $P H_1 = 0,4$; $P H_2 = 0,6$. Через A позначимо подію, котра полягає в тому, що прилад безвідмовно працював час t .

Умовні ймовірності цієї події при гіпотезах H_1 і H_2 дорівнюють за умовами задачі

$$P A/H_1 = 0,95; \quad P A/H_2 = 0,7.$$

За формулою (4.18) знаходимо ймовірність гіпотези H_1 після випробування:

$$P(H_1) = \frac{0,4 \cdot 0,95}{0,4 \cdot 0,95 + 0,6 \cdot 0,7} = 0,475.$$

4.3. ЗАКОНИ РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

Нагадаємо, що випадковою називається величина, яка внаслідок досліду може набути того чи іншого значення, невідомо заздалегідь якого саме. Ми домовилися також розрізняти випадкові величини перервного (дискретного) і неперервного типу. Можливі значення *перервних* величин можна наперед назвати (наприклад, число появи герба при трьох киданнях монети – можливі значення 0, 1, 2, 3; частота появи герба в тому ж досліді – можливі значення 0, $\frac{1}{3}$, $\frac{2}{3}$, 1; число елементів, що відмовили, у приладі, складеному з п'яти елементів, – можливі значення 0, 1, 2, 3, 4, 5).

Можливі значення неперервних величин не можуть бути заздалегідь перераховані, тому що вони неперервно заповнюють деякий проміжок. Приклади неперервних випадкових величин – похибка вимірювача висоти; час безвідмовної роботи напівпровідникового елемента в приладі.

Надалі випадкові величини позначатимемо великими буквами, а їхні можливі значення – відповідними малими буквами. Наприклад, X – число влучень при трьох пострілах; можливі значення: $x_1 = 0$; $x_2 = 1$; $x_3 = 2$; $x_4 = 3$.

Розглянемо дискретну випадкову величину X з можливими значеннями x_1, x_2, \dots, x_n . Кожне з цих значень можливе, але не вірогідне, і величина X може набути будь-якого одного з них з деякою ймовірністю, тобто відбудеться одна подія з повної групи несумісних подій:

$$\left. \begin{array}{l} X = x_1, \\ X = x_2, \\ \dots\dots\dots \\ X = x_n. \end{array} \right\} \quad (4.19)$$

Позначимо ймовірності цих подій буквами P з відповідними індексами:

$$P(X = x_1) = P_1; P(X = x_2) = P_2; \dots; P(X = x_n) = P_n.$$

Оскільки несумісні події (4.19) утворюють повну групу, то $\sum_{i=1}^n P_i = 1$, тобто

сума ймовірностей усіх можливих значень випадкової величини дорівнює одиниці. Ця сумарна ймовірність якимось чином розподілена між окремими значеннями. Випадкова величина буде цілком описана з імовірнісної точки зору, якщо ми задамо цей розподіл, тобто в точності вкажемо, якою ймовірністю володіє кожна з подій (4.19).

Законом розподілу випадкової величини називається всяке співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини і відповідними їхніми ймовірностями.

Про випадкову величину говоритимемо, що вона підпорядковується даному закону розподілу.

Встановимо форму, в якій може бути заданий закон розподілу *перервної* (дискретної) випадкової величини X . Найпростішою формою є *таблиця*, в якій наведені можливі значення випадкової величини і відповідні їм ймовірності:

x_i	x_1	x_2	...	x_n
P_i	P_1	P_2	...	P_n

Таку таблицю називають *рядом розподілу випадкової величини X* .

Щоб надати ряду розподілу більш наочного вигляду, часто застосовують його графічне зображення: вздовж осі абсцис відкладають можливі значення випадкової величини, а вздовж осі ординат – ймовірності цих значень; отримані точки з'єднують відрізками прямих. Така фігура називається *багатокутником розподілу* (рис. 4.1, а). Багатокутник розподілу, так само як і ряд розподілу, цілком характеризує випадкову величину; він є однією з форм закону розподілу.

Отже, ряд (або багатокутник розподілу) дає вичерпну характеристику дискретної випадкової величини. *Неперервна* ж випадкова величина має незліченну кількість можливих значень; тому для неперервної випадкової величини не існує ряду розподілу в тому розумінні, в якому він існує для дискретної випадкової величини. Проте різні області можливих значень неперервної величини не є однаково ймовірними, і для неї теж існує свій “розподіл ймовірностей”.

Для кількісної характеристики цього розподілу ймовірностей зручно скористатися не ймовірністю події $X = x$, а ймовірністю події $X < x$, де x – деяка поточна змінна. Ймовірність цієї події, мабуть, залежить від x , тобто є деякою функцією від x . Ця функція називається **функцією розподілу випадкової величини X** і позначається $F(x)$:

$$F(x) = P(X < x). \quad (4.20)$$

Функцію розподілу $F(x)$ іноді називають також *інтегральною функцією* чи *інтегральним законом розподілу*.

Функція розподілу – найуніверсальніша характеристика випадкової величини, оскільки вона існує для всіх випадкових величин: як перервних, так і неперервних, та цілком характеризує випадкову величину з імовірнісної точки зору, тобто є однією з форм закону розподілу.

Сформулюємо деякі загальні властивості функції розподілу.

1. Функція розподілу $F(x)$ є неспадною функцією свого аргументу, тобто при $x_2 > x_1$ $F(x_2) \geq F(x_1)$.

2. На мінус нескінченності функція розподілу дорівнює нулю:

$$F(-\infty) = 0.$$

3. На плюс нескінченності функція дорівнює одиниці:

$$F(+\infty) = 1.$$

4. Імовірність того, що випадкова величина знаходиться в заданих межах (рис. 4.1, б), дорівнює

$$P(x_1 < X < x_2) = F(x_2) - F(x_1).$$

На практиці звичайно функція розподілу неперервної випадкової величини є неперервною в усіх точках функцією (рис. 4.1, б), а функція розподілу будь-якої перервної випадкової величини завжди є розривна східчаста функція, стрибки якої відбуваються в точках, що відповідають можливим значенням випадкової величини; значення стрибків дорівнюють ймовірностям значень випадкової величини. Сума всіх стрибків функції $F(x)$ дорівнює одиниці (рис. 4.1, в).

Щільність розподілу. Візьмемо неперервну випадкову величину X з функцією розподілу $F(x)$, яка є неперервною і диференційовною. Обчислимо ймовірність попадання X на ділянку від x до $x + \Delta x$, тобто збільшення функції розподілу на цій ділянці:

$$P(x < X < x + \Delta x) = F(x + \Delta x) - F(x).$$

Розглянемо відношення знайденої ймовірності до довжини ділянки, тобто середню ймовірність, що припадає на одиницю довжини ділянки і наближає Δx до нуля. У межі одержимо похідну від функції розподілу:

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = f'(x). \quad (4.21)$$

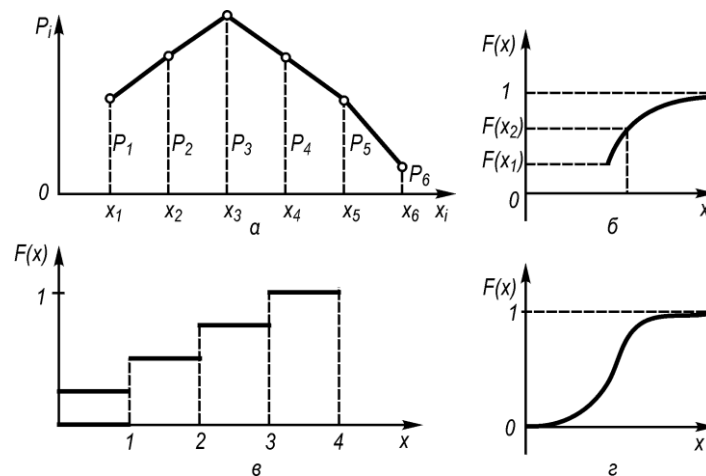


Рис. 4.1. Закони розподілу випадкових величин:

- a* – багатокутник розподілу; *б* – функція розподілу для неперервної випадкової величини; *в* – функція розподілу для дискретної випадкової величини; *г* – неперервна функція

Уведемо позначення:

$$f(x) = F'(x). \quad (4.22)$$

Функція $f(x)$ – похідна функції розподілу – характеризує начебто щільність, з якою розподіляються значення випадкової величини в даній точці. Ця функція називається *щільністю розподілу* (інакше – *щільністю ймовірності*) неперервної випадкової величини X . Іноді функцію $f(x)$ називають також *диференціальною функцією розподілу* чи *диференціальним законом розподілу* величини X .

Формула (4.22) подає щільність розподілу через функцію розподілу. Виразимо ймовірність попадання величини X на відрізок від x_1 до x_2 через щільність розподілу:

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx. \quad (4.23)$$

Інтегральна функція розподілу

$$F(x) = P(X < x) = P(-\infty < X, x)$$

або

$$F(x) = \int_{-\infty}^x f(x)dx. \quad (4.24)$$

Геометрично $f(x)$ є не що інше, як площа кривої розподілу, що лежить лівіше точки x_i (рис. 4.2, а).

Вкажемо основні властивості щільності розподілу.

1. $f(x) \geq 0$ [оскільки функція розподілу $f(x)$ є неспадною, і тому $F'(x) \geq 0$].

2. $\int_{-\infty}^{\infty} f(x)dx = 1$, що випливає з (4.24), враховуючи, що $F(+\infty) = 1$.

Геометрично основні властивості щільності розподілу означають, що:

1) уся крива розподілу лежить не нижче осі абсцис;

2) повна площа, обмежена кривою розподілу і віссю абсцис, дорівнює одиниці.

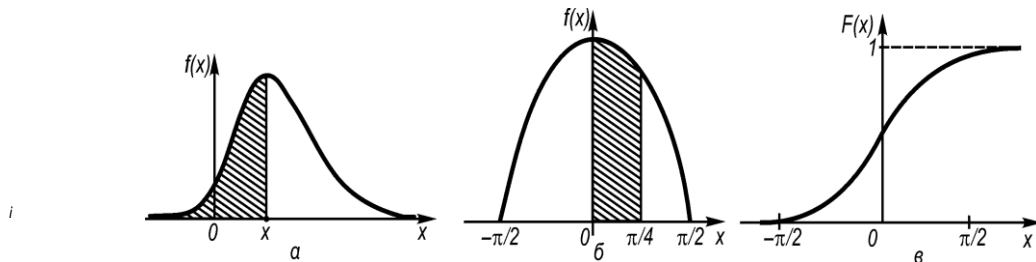


Рис. 4.2. Закони розподілу випадкових величин:
а, б – диференціальний; в – інтегральний

Функція розподілу $F(x)$, як будь-яка ймовірність, є величина безрозмірна.

Приклад 4.8. Випадкова величина X підпорядковується закону розподілу з щільністю

$$f(x) = \begin{cases} a \cos x & \text{при } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}; \\ 0 & \text{при } x < -\frac{\pi}{2} \text{ чи } x > \frac{\pi}{2}. \end{cases}$$

Потрібно: а) визначити коефіцієнт a ; б) побудувати графік щільності розподілу $f(x)$; в) знайти функцію розподілу $F(x)$ і побудувати її графік; г) знайти ймовірність попадання величини X на ділянку від 0 до $\frac{\pi}{2}$.

Розв'язання. Для визначення коефіцієнта a скористаємося другою властивістю щільності розподілу:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a \cos x dx = 2a = 1,$$

звідки $a = 1/2$.

Графік щільності розподілу $f(x)$ наведений на рис 4.2, б.

За формулою (4.24) функція розподілу

$$F(x) = \begin{cases} 0 & \text{при } x < -\frac{\pi}{2}; \\ \frac{1}{2}(\sin x + 1) & \text{при } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}; \\ 1 & \text{при } x > \frac{\pi}{2}. \end{cases}$$

Графік функції $F(x)$ зображений на рис. 4.2, в.

За формулою (4.23)

$$P\left(0 < X < \frac{\pi}{4}\right) = \frac{1}{2}\left(\sin \frac{\pi}{4} + 1\right) - \frac{1}{2}(\sin 0 + 1) = \frac{\sqrt{2}}{4}.$$

4.4. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВИХ ВЕЛИЧИН

Кожен закон розподілу являє собою деяку функцію, значення якої цілком описує випадкову величину з імовірнісної точки зору. Проте у багатьох практичних задачах немає потреби характеризувати випадкову величину цілком. Найчастіше достатньо вказати тільки окремі числові параметри деякої міри розподілу випадкової величини, що характеризують її істотні риси: наприклад, якесь середнє значення, навколо якого групується можливе значення випадкової величини; яке-небудь число, що характеризує ступінь розкиду цих значень щодо середнього, і т. д. Користуючись такими характеристиками, можна всі наявні істотні відомості про випадкову величину виразити найбільш компактно за допомогою мінімальної кількості числових параметрів. Такі характеристики, призначення яких – виразити в стислій формі найбільш істотні особливості розподілу, називаються *числовими характеристиками випадкової величини*.

У теорії ймовірностей числові характеристики й операції з ними відіграють величезну роль. За допомогою числових характеристик істотно полегшується розв'язання багатьох ймовірнісних задач. Дуже часто вдається розв'язати задачу до кінця, залишаючи осторонь закони розподілу й оперуючи одними числовими характеристиками. При цьому дуже важливою є та обставина, що коли в задачі фігурує велика кількість випадкових величин, кожна з яких впливає на числовий результат досліду, то закон розподілу цього результату значною мірою можна вважати незалежним від законів розподілу окремих випадкових величин (виникає так званий *нормальний закон розподілу*). У цих випадках для вичерпного судження про результуючий закон розподілу не потрібно знати закони розподілу окремих величин, що фігурують у задачі; для цього вистачить лише деяких числових характеристик цих величин.

У теорії ймовірностей і математичній статистиці користуються великою кількістю різних числових характеристик, що мають різне призначення. Розглянемо найбільш часто застосовувані числові характеристики.

Серед числових характеристик випадкових величин потрібно насамперед визначити ті, котрі характеризують положення випадкової величини на числовій осі, тобто вказують деяке середнє, орієнтовне значення, навколо якого групується всі можливі значення випадкової величини.

Математичним сподіванням випадкової величини називається сума добутків усіх можливих значень випадкової величини на ймовірності цих значень:

$$M[X] = m_x = \sum_{i=1}^n x_i P_i. \quad (4.25)$$

У наведеному формулюванні визначення математичного сподівання є слушним, строго кажучи, тільки для *дискретних* випадкових величин. Для *неперервної* величини X математичне сподівання виражається вже не сумою, а інтегралом:

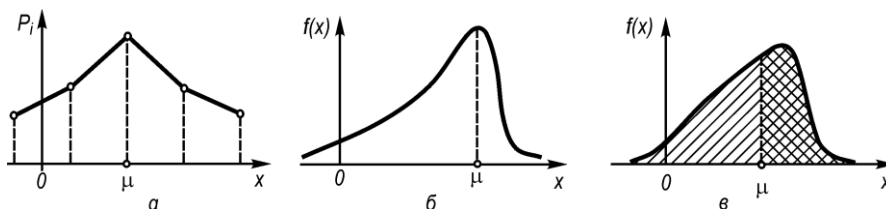


Рис. 4.3. Функція розподілу:

a – багатокутник розподілу; $б, в$ – щільність імовірності

$$M[X] = m_x = \int_{-\infty}^{\infty} xf(x)dx, \quad (4.26)$$

де $f(x)$ – щільність розподілу величини X .

Крім найважливішої з характеристик – математичного сподівання, – на практиці іноді використовують й інші характеристики, зокрема моду і медіану випадкової величини.

Модою випадкової величини називається її найбільш імовірне значення. Термін “найбільш імовірне значення” точно підходить тільки до *дискретних* величин; для *неперервної* величини модою є те значення, при якому щільність імовірності максимальна. Умовимося позначати моду буквою μ . На рис. 4.3, a і 4.3, $б$ наведено моду відповідно для перервної і неперервної випадкових величин.

У загальному випадку мода і математичне сподівання випадкової величини не збігаються. В окремому випадку, коли розподіл є симетричним і модальним (тобто має моду) та існує математичне сподівання, то воно збігається з модою і центром симетрії розподілу.

Медіаною випадкової величини X називається таке її значення Me , для якого $P(X < Me) = P(X > Me)$, тобто значення однаково ймовірне незалежно від того, чи виявиться випадкова величина меншою чи більшою за Me . Цією характеристикою користуються звичайно тільки для неперервних випадкових величин, хоча формально можна її визначити і для перервних. Геометрично медіана – це абсциса точки, в якій площа, обмежена кривою розподілу, поділяється навпіл (рис. 4.3, $в$).

Крім кількісних характеристик – середніх, типових значень випадкової величини, – вживається ще ряд характеристик, кожна з яких описує ту чи іншу властивість розподілу. Це початкові і центральні моменти, дисперсія, середньоквадратичне відхилення, середньоарифметичне відхилення.

Початковим моментом s -го порядку перервної випадкової величини X називається сума

$$a_s[X] = \sum_{i=1}^n x_i^s P_i. \quad (4.27)$$

Для *неперервної* випадкової величини X початковим моментом s -го порядку називається інтеграл

$$a_s[X] = \int_{-\infty}^{\infty} x^s f(x) dx. \quad (4.28)$$

Неважко переконатися, що введена вище основна характеристика – математичне сподівання – є початковим моментом першого порядку випадкової величини X .

Можна записати загальне визначення початкового моменту s -го порядку, слухне як для перервних, так і для неперервних величин:

$$a_s[X] = M[X^s], \quad (4.29)$$

тобто *початковим моментом s -го порядку випадкової величини X називається математичне сподівання s -го степеня цієї випадкової величини.*

Нехай є випадкова величина X з математичним сподіванням m_x . *Центрованою випадковою величиною, що відповідає величині X , називається відхилення випадкової величини X від її математичного сподівання:*

$$X^0 = X - m_x. \quad (4.30)$$

Математичне сподівання центрованої випадкової величини дорівнює нулю.

Моменти центрованої випадкової величини зветься **центральними моментами**. Вони аналогічні моментам центра ваги в механіці.

Отже, центральним моментом порядку s випадкової величини X називається математичне сподівання s -го степеня відповідної центрованої випадкової величини:

$$\mu_s[X] = M[X^s] = M[(X - m_x)^s]. \quad (4.31)$$

Для дискретної випадкової величини s -й центральний момент виражається сумою

$$\mu_s = \sum_{i=1}^n (x - m_x)^s P_i, \quad (4.32)$$

а для неперервної – інтегралом

$$\mu_s = \int_{-\infty}^{\infty} (x - m_x)^s f(x) dx. \quad (4.33)$$

Надалі в тих випадках, коли не виникає сумнівів, до якої випадкової величини відноситься даний момент, будемо для стислості замість $a_s X$ і $\mu_s X$ писати a_s і μ_s .

Очевидно, для будь-якої випадкової величини центральний момент *першого порядку* дорівнює нулю:

$$\mu_1 = M[X] = M[X - m_x] = 0, \quad (4.34)$$

тому що математичне сподівання центрованої випадкової величини завжди дорівнює нулю.

Центральний момент другого порядку називається **дисперсією випадкової величини**. Через крайню важливість цієї характеристики серед інших введемо для неї спеціальне позначення $D X$:

$$\mu_2 = D[X].$$

Відповідно до означення центрального моменту

$$D[X] = M[X^2], \quad (4.35)$$

тобто дисперсією випадкової величини X називається математичне сподівання квадрата відповідної центрованої величини.

Підставивши в (4.35) вираз величини $\overset{0}{X}$ (4.30), матимемо

$$D[X] = M[(X - m_x)^2]. \quad (4.36)$$

Для безпосереднього обчислення дисперсії відповідно дискретних і неперервних величин служать формули

$$D[X] = \sum_{i=1}^n (x_i - m_x)^2 P_i; \quad (4.37)$$

$$D[X] = \int_{-\infty}^{\infty} (x - m_x)^2 f(x) dx. \quad (4.38)$$

Слово “дисперсія” означає “розсіювання”. Дисперсія випадкової величини – це характеристика розкиду значень випадкової величини навколо її математичного сподівання.

Дисперсія випадкової величини має розмірність квадрата випадкової величини; для наочної характеристики розсіювання зручніше користуватися величиною, розмірність якої збігається з розмірністю випадкової величини. Для цього з дисперсії добувають квадратний корінь. Отримана величина називається **середньоквадратичним відхиленням** (або стандартом) випадкової величини X . Середньоквадратичне відхилення позначатимемо $\sigma[X]$:

$$\sigma[X] = \sqrt{D[X]}. \quad (4.39)$$

Крім розглянутих початкових і центральних моментів, на практиці іноді застосовуються **абсолютні моменти** (початкові і центральні), що визначаються формулами

$$\beta_s = M[|X|^s]; \quad \nu_s = M[|\overset{0}{X}|^s].$$

Очевидно, абсолютні моменти парних порядків збігаються зі звичайними моментами.

З абсолютних моментів найбільш часто застосовується абсолютний центральний момент першого порядку

$$\nu_1 = M[\overset{0}{X}] = M[|X - m_x|], \quad (4.40)$$

який називається **середньоарифметичним відхиленням**. Поряд з дисперсією і середньоквадратичним відхиленням середньоарифметичне відхилення іноді застосовується як характеристика розсіювання.

У багатьох практичних задачах повна характеристика випадкової величини – закон розподілу – або не потрібна, або не може бути отримана. У цих випадках обмежуються приблизним описом випадкової величини за допомогою числових характеристик, кожна з яких виражає певну характерну властивість розподілу.

4.5. ЗАКОН РОЗПОДІЛУ ЩІЛЬНОСТІ ЙМОВІРНОСТІ ВИПАДКОВИХ ВЕЛИЧИН

Закон рівномірної щільності. У деяких задачах зустрічаються неперервні випадкові величини, про які заздалегідь відомо, що їхні можливі значення лежать у межах деякого визначеного інтервалу; крім того, відомо, що в межах цього інтервалу всі значення випадкової величини однаково ймовірні (точніше, мають одну й ту саму щільність ймовірності). Про такі випадкові величини говорять, що вони розподіляються за *законом рівномірної щільності*.

Наведемо приклад таких величин.

Приклад 4.9. Тіло зважене на точних вагах, але в розпорядженні того, хто зважує, були тільки гирки масою не менш як 1 г; результат зважування показує, що маса тіла знаходиться між k і $k+1$ грамами. Масу тіла взято рівною $k+1/2$ грамам. Допущена при цьому похибка X , очевидно, є випадковою величиною, розподіленою з рівномірною щільністю на ділянці $\left(-\frac{1}{2}, \frac{1}{2}\right)$ г.

Розглянемо випадкову величину X , що підпорядковується закону рівномірної щільності на ділянці від x_1 до x_2 (рис. 4.4, а), і складемо для неї вираз щільності розподілу $f(x)$. Щільність є сталою і дорівнює c на відрізку x_1, x_2 ; поза цим відрізком вона дорівнює нулю:

$$f(x) = \begin{cases} c & \text{при } x_1 < x < x_2; \\ 0 & \text{при } x < x_1 \text{ або } x > x_2. \end{cases}$$

Площа, обмежена кривою розподілу, дорівнює одиниці:

$$c(x_2 - x_1) = 1,$$

звідки

$$c = \frac{1}{(x_2 - x_1)}.$$

Отже, щільність розподілу

$$f(x) = \begin{cases} 1/(x_2 - x_1) & \text{при } x_1 < x < x_2; \\ 0 & \text{при } x < x_1 \text{ або } x > x_2. \end{cases} \quad (4.41)$$

Формула (4.41) і виражає закон рівномірної щільності на ділянці x_1, x_2 .

Напишемо вираз для функції розподілу $F(x)$. Функція розподілу виражається площею кривої розподілу, що лежить лівіше точки x . Отже,

$$F(x) = \begin{cases} 0 & \text{при } x < x_1, \\ (x - x_1)/(x_2 - x_1) & \text{при } x_1 < x < x_2, \\ 1 & \text{при } x > x_2. \end{cases}$$

Графік функції $F(x)$ наведено на рис. 4.4, б.

Визначимо основні числові характеристики випадкової величини X , яка підпорядковується закону рівномірної щільності на ділянці від x_1 до x_2 .

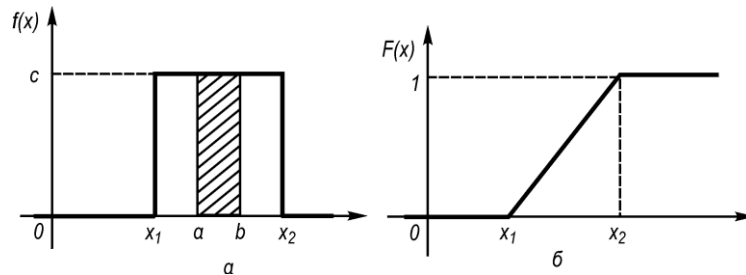


Рис. 4.4. Диференціальний (а) та інтегральний закони розподілу ймовірності (б)

Математичне сподівання величини X дорівнює

$$m_x = \int_{x_2}^{x_1} \frac{x}{x_2 - x_1} dx = \frac{x_1 + x_2}{2}. \quad (4.42)$$

Внаслідок симетричності рівномірного розподілу медіана величини X також дорівнює $x_1 + x_2 / 2$.

Моди закон рівномірної щільності не має.

За формулою (4.38) знаходимо дисперсію величини X :

$$D_x = \mu_2 = \frac{1}{x_2 - x_1} \int_{x_2}^{x_1} \left(x - \frac{x_1 + x_2}{2}\right)^2 dx = \frac{(x_1 - x_2)^2}{12}, \quad (4.43)$$

звідки середньоквадратичне відхилення

$$\sigma_x = \sqrt{D_x} = \frac{x_1 - x_2}{2\sqrt{3}}. \quad (4.44)$$

Середньоарифметичне відхилення

$$v_1 = \frac{1}{x_2 - x_1} \int_{\alpha}^{\beta} \left|x - \frac{x_1 + x_2}{2}\right| dx = \frac{2}{x_2 - x_1} \int_{\frac{\alpha+\beta}{2}}^{\beta} \left(x - \frac{x_1 + x_2}{2}\right) dx = \frac{x_2 - x_1}{4}. \quad (4.44a)$$

Нарешті, знайдемо ймовірність попадання випадкової величини X , розподіленої за законом рівномірної щільності, на ділянку (a, b) , яка є частиною ділянки (x_1, x_2) . Геометрично ця ймовірність зображена площею, заштрихованою на рис. 4.4, а. Очевидно, вона дорівнює

$$P(a < X < b) = (b - a) / (x_2 - x_1), \quad (4.45)$$

тобто відношенню довжини відрізка (a, b) до всієї довжини ділянки (x_1, x_2) , на якій задано рівномірний розподіл.

Закон розподілу Пуассона. Розглянемо перервну випадкову величину X , що може набувати тільки цілих, невід'ємних значень: $0, 1, 2, \dots, m, \dots$, причому послідовність цих значень теоретично не обмежена.

Говорять, що випадкова величина X розподілена за *законом Пуассона*, якщо ймовірність того, що вона набуває визначеного значення m , виражається формулою

$$P_m = \frac{a^m}{m!} e^{-a}, \quad m = 0, 1, \dots, \quad (4.46)$$

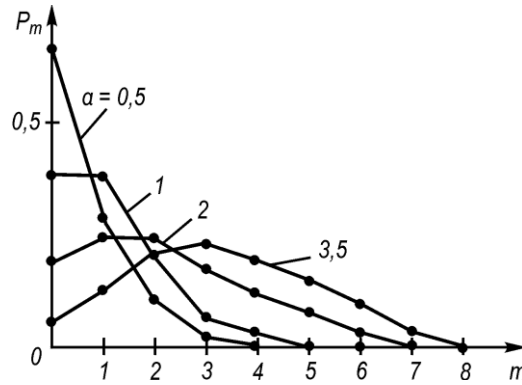


Рис. 4.5. Багатокутники розподілу

де a – деяка додатна величина, що називається *параметром* закону Пуассона. Ряд розподілу випадкової величини X за законом Пуассона має такий вигляд:

X_m	0	1	2	...	m	...
P_m	e^{-a}	$\frac{a}{1!}e^{-a}$	$\frac{a^2}{2!}e^{-a}$...	$\frac{a^m}{m!}e^{-a}$...

Послідовність імовірностей, що задається формулою (4.46), може являти собою ряд розподілу, тобто сума всіх ймовірностей P_m дорівнює одиниці:

$$\sum_{m=0}^{\infty} P_m = \sum_{m=0}^{\infty} \frac{a^m}{m!} e^{-a} = e^{-a} \sum_{m=0}^{\infty} \frac{a^m}{m!} = e^{-a} e^a = 1,$$

де $\sum_{m=0}^{\infty} \frac{a^m}{m!} = e^a$.

На рис. 4.5 наведено багатокутники розподілу за законом Пуассона випадкової величини X , що відповідають різним параметрам a .

Визначимо основні характеристики – математичне сподівання і дисперсію випадкової величини X , розподіленої за законом Пуассона. За означенням математичного сподівання

$$m_x = M[x] = \sum_{m=0}^{\infty} m P_m = \sum_{m=0}^{\infty} m \frac{a^m}{m!} e^{-a}.$$

Перший член суми $m = 0$, отже, додавання треба починати з $m = 1$:

$$\begin{aligned} m_x &= \sum_{m=1}^{\infty} m \frac{a^m}{m!} e^{-a} = a e^{-a} \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} = \\ &= a e^{-a} \sum_{m=1}^{\infty} \frac{a^{m-1}}{(m-1)!} = a e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!} = a e^{-a} e^a = a. \end{aligned} \quad (4.47)$$

Як бачимо, параметр a є математичним сподіванням випадкової величини X .

З огляду на те, що початковий момент другого порядку величини X дорівнює $a_2 = a(a+1)$, одержимо дисперсію величини

$$D_x = a_2 - m_x^2 =$$

$$= a^2 + a - a^2 = a. \quad (4.48)$$

Отже, дисперсія випадкової величини, розподіленої за законом Пуассона, дорівнює її математичному сподіванню a .

Ця властивість розподілу Пуассона часто застосовується на практиці для вирішення питання, чи правдоподібна гіпотеза про те, що випадкова величина X розподілена за законом Пуассона. Для цього дослідним шляхом визначають статистичні характеристики – математичне сподівання і дисперсію випадкової величини. Якщо їхні значення близькі, то це свідчить на користь гіпотези про пуассонівський розподіл; різка розбіжність цих характеристик свідчить проти гіпотези.

Визначимо для випадкової величини X , розподіленої за законом Пуассона, ймовірність того, що вона набуде значення не меншого за задане (k). Позначимо цю ймовірність через R_k :

$$R_k = P(X > k).$$

Очевидно, ймовірність R_k може бути обчислена як сума:

$$R_k = P_k + P_{k+1} + \dots = \sum_{m=k}^{\infty} P_m.$$

Однак значно простіше визначити її з ймовірності протилежної події:

$$R_k = 1 - (P_0 + P_1 + \dots + P_{k-1}) = 1 - \sum_{m=0}^{k-1} P_m. \quad (4.49)$$

Зокрема, ймовірність того, що величина X набуде додатного значення, виражається формулою

$$R_1 = 1 - P_0 = 1 - e^{-a}. \quad (4.50)$$

Розглянемо кілька прикладів пуассонівського розподілу величин з різних галузей практики.

Приклад 4.10. На автоматичну телефонну станцію надходять виклики із середньою щільністю K викликів у годину. Вважаючи, що число викликів на будь-якій ділянці часу розподілено за законом Пуассона, знайти ймовірність того, що за дві хвилини на станцію надійде три виклики.

Розв'язання. Середнє число викликів за дві хвилини дорівнює

$$a = 2K/60 = K/30.$$

За формулою (4.46) ймовірність надходження трьох викликів дорівнює

$$P_3 = \frac{(K/30)^3}{1 \cdot 2 \cdot 3} e^{-\frac{K}{30}}.$$

Приклад 4.11. З умови попереднього прикладу знайти ймовірність того, що за дві хвилини надійде хоча б один виклик.

Розв'язання. За формулою (4.49)

$$R_1 = 1 - e^{-a} = 1 - e^{-\frac{K}{30}}.$$

Приклад 4.12. За тих же умов знайти ймовірність того, що за дві хвилини надійде не менше трьох викликів.

Розв'язання. За формулою (4.49)

$$R_3 = 1 - (P_0 + P_1 + P_2) = 1 - e^{-\frac{K}{30}} \left[1 + \frac{K}{30} + \frac{1}{2} \left(\frac{K}{30} \right)^2 \right].$$

Нормальний закон розподілу. Нормальний закон розподілу (або закон Гаусса) грає винятково важливу роль у теорії ймовірностей. На практиці цей закон розподілу серед інших зустрічається найчастіше. Головна особливість, що виділяє нормальний закон серед інших, полягає в тому, що він є *граничним* законом, до якого наближаються інші закони розподілу при типових умовах.

Можна довести, що сума досить великої кількості незалежних (чи слабко залежних) випадкових величин, які підпорядковані яким завгодно законам розподілу, у разі дотримання деяких дуже нежорстких обмежень приблизно підпорядковується нормальному закону, і тим точніше, чим більша кількість випадкових величин підсумовується.

Нормальний закон розподілу характеризується щільністю ймовірності виду

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

де σ – середньоквадратичне відхилення величини X ; m – математичне сподівання.

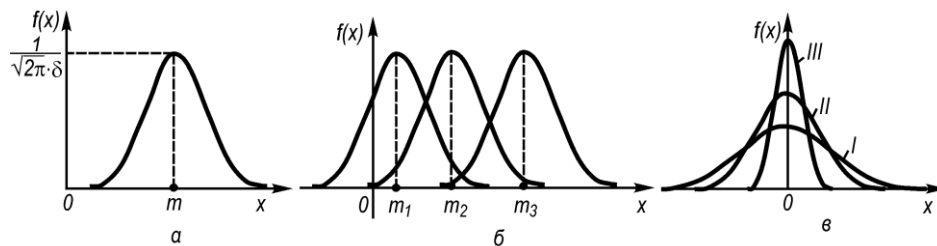


Рис. 4.6. Щільності розподілу ймовірності за нормальним законом

Крива розподілу за нормальним законом має симетричний горбоподібний вигляд (рис. 4.6, а). Максимальна ордината кривої дорівнює $\frac{1}{\sigma\sqrt{2\pi}}$ і відповідає точці $x = m$, у міру віддалення від точки m щільність розподілу зменшується, і при $x \rightarrow \pm\infty$ крива асимптотично наближається до осі абсцис.

Із зміною центра розсіювання m крива розподілу зміщується вздовж осі абсцис, не змінюючи своєї форми (рис. 4.6, б). Центр розсіювання характеризує положення розподілу на осі абсцис. Розмірність центра розсіювання – та сама, що і розмірність випадкової величини.

Параметр σ характеризує не положення, а саму форму кривої розподілу. Це є характеристика розсіювання. Найбільша ордината кривої розподілу обернено пропорційна σ ; із збільшенням σ максимальна ордината зменшується. Оскільки площа кривої розподілу завжди дорівнює одиниці, то зі збільшенням σ крива розподілу стає більш плоскою, розтягуючись уздовж осі абсцис; і навпаки, зі зменшенням σ крива розподілу витягується, одночасно стискаючись з боків, і стає голкоподібною. На рис. 4.6, в показані три нормальні криві (I, II, III) при $m = 0$; з них крива I відповідає найбільшому, а крива III – найменшому значенню σ . Зміни параметра σ рівносильні зміні масштабу кривої розподілу – збільшенню масштабу по одній осі і такому ж зменшенню по іншій.

Розмірність параметра σ , природно, збігається з розмірністю випадкової величини X .

Ймовірність попадання випадкової величини, яка підпорядковується нормальному закону, на задану ділянку. У багатьох задачах, пов'язаних з нормально розподіленими випадковими величинами, треба визначити ймовірність попадання випадкової величини X , яка підпорядковується нормальному закону з параметрами m , σ , на ділянку від α до β . Для обчислення цієї ймовірності скористаємося загальною формулою

$$P \alpha < X < \beta = P \beta - P \alpha . \quad (4.52)$$

Знайдемо функцію розподілу $F x$ випадкової величини X , розподіленої за нормальним законом з параметрами m , σ . Щільність розподілу величини X дорівнює

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2\sigma^2}} . \quad (4.53)$$

Звідси знаходимо інтегральну функцію розподілу

$$F(x) = \int_{-\infty}^x f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(x-m)^2}{2\sigma^2}} dx. \quad (4.54)$$

Зробимо в інтегралі (4.54) заміну змінної

$$\frac{x-m}{\sigma} = t,$$

після чого він набере такого вигляду:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m}{\sigma}} e^{-\frac{t^2}{2}} dx. \quad (4.55)$$

Інтеграл (4.55) не виражається через елементарні функції, але його можна обчислити через спеціальну функцію Крампа, для якої складені таблиці. Існує багато різновидів таких функцій, наприклад $\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ і т. д. Якою з цих функцій скористатися – питання смаку. Ми візьмемо за таку функцію

$$\Phi^*(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \quad (4.56)$$

Як бачимо, вона є функцією розподілу для нормально розподіленої випадкової величини з параметрами $m=0$, $\sigma=1$.

Домовимося називати функцію $\Phi^*(x)$ *нормальною функцією розподілу*. Таблиці значень функції $\Phi^*(x)$ наведені в [9].

Виразимо функцію розподілу (4.54) величини X з параметрами m і σ через нормальну функцію розподілу $\Phi^*(x)$.

Очевидно,

$$F(x) = \Phi^*\left(\frac{x-m}{\sigma}\right). \quad (4.57)$$

Тепер знайдемо ймовірність попадання випадкової величини X на ділянку від α до β . Відповідно до формули (4.52)

$$P(\alpha < X < \beta) = \Phi^*\left(\frac{\beta-m}{\sigma}\right) - \Phi^*\left(\frac{\alpha-m}{\sigma}\right). \quad (4.58)$$

Таким чином, ймовірність попадання на ділянку випадкової величини X , розподіленої за нормальним законом з будь-якими параметрами, виражається через стандартну функцію розподілу $\Phi^*(x)$, що відповідає найпростішому нормальному закону з параметрами 0, 1. Зазначимо, що аргументи функції Φ^* у формулі (4.58) мають дуже простий зміст: $\frac{\beta-m}{\sigma}$ – відстань від правого кінця ділянки β до центра розсіювання, виражена в середньоквадратичних відхиленнях; $\frac{\alpha-m}{\sigma}$ – така сама відстань для лівого кінця ділянки, причому ця відстань вважається додатною, якщо кінець розташований праворуч від центра розсіювання, і від'ємною, якщо ліворуч.

Як і усяка функція розподілу, функція $\Phi^*(x)$ має такі властивості:

1. $\Phi^*(-\infty) = 0$.
2. $\Phi^*(+\infty) = 1$.
3. $\Phi^*(x) = 0$ – неспадна функція.

Крім того, із симетричності нормального розподілу з параметрами $m=0$, $\sigma=1$ відносно початку координат випливає, що

$$\Phi^*(-x) = 1 - \Phi^*(x). \quad (4.59)$$

На практиці часто зустрічається задача обчислення ймовірності попадання нормально розподіленої випадкової величини на ділянку, симетричну щодо центра розсіювання m . Розглянемо таку ділянку, довжина якої $2l$ (рис. 4.7, а). Ймовірність попадання на цю ділянку за (4.58)

$$P(m-l < X < m+l) = \Phi^*\left(\frac{l}{\sigma}\right) - \Phi^*\left(-\frac{l}{\sigma}\right). \quad (4.60)$$

Враховуючи (4.59) і надаючи лівій частині формули (4.60) компактнішого вигляду, одержимо формулу для ймовірності попадання випадкової величини, розподіленої за нормальним законом, на ділянку, симетричну щодо центра розсіювання:

$$P(|X - m| < l) = 2\Phi^*\left(\frac{l}{\sigma}\right) - 1. \quad (5.61)$$

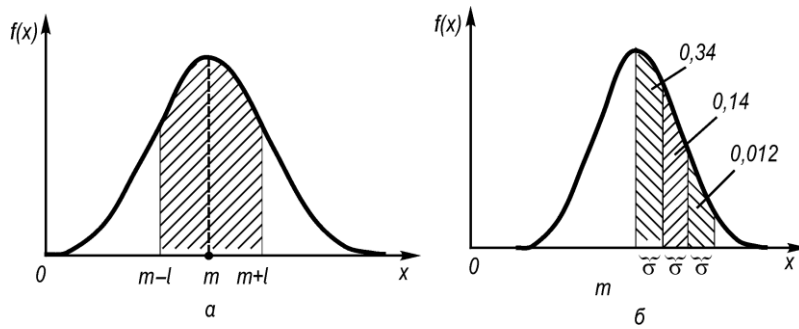


Рис. 4.7. Криві нормального розподілу щільності ймовірності:

a – для обчислення ймовірності попадання нормально розподіленої випадкової величини на ділянку, симетричну до центра розсіювання m ; *б* – до визначення правила трьох сигма

Розв'яжемо наступну задачу з визначення ймовірності попадання випадкової величини у визначений інтервал. Відкладемо від центра розсіювання m послідовних відрізків завдовжки σ (рис. 4.7, б) і обчислимо ймовірність попадання випадкової величини X у кожний з них. Оскільки крива нормального закону симетрична, достатньо відкласти такі відрізки тільки в один бік. За формулою (4.58), підставивши взяті з довідкових таблиць значення функції Φ^* , знаходимо:

$$\left. \begin{aligned} P(m < X < m + \sigma) &= \Phi^*(1) - \Phi^*(0) = 0,8413 - 0,5000 \approx 0,341, \\ P(m + \sigma < X < m + 2\sigma) &= \Phi^*(2) - \Phi^*(1) \approx 0,136, \\ P(m + 2\sigma < X < m + 3\sigma) &= \Phi^*(3) - \Phi^*(2) \approx 0,012, \\ P(m + 3\sigma < X < m + 4\sigma) &= \Phi^*(4) - \Phi^*(3) \approx 0,001. \end{aligned} \right\} \quad (4.62)$$

Як видно з (4.62), ймовірності попадання на кожний з наступних відрізків (п'ятий, шостий і т. д.) з точністю до 0,001 дорівнюють нулю.

Округливши ймовірності попадання у відрізки до 0,01 (до 1%), одержимо три числа: 0,34; 0,14; 0,01. Сума їх дорівнює приблизно 0,5. Це означає, що для нормально розподіленої випадкової величини всі розсіювання (з точністю до соті) укладаються на ділянці $m \pm 3\sigma$ [2].

Таким чином, знаючи середньоквадратичне відхилення і математичне сподівання випадкової величини, можна орієнтовно вказати інтервал її практично можливих значень. Наведений спосіб оцінки діапазону можливих значень випадкової величини відомий у математичній статистиці як правило трьох сигма. З правила трьох сигма випливає також спосіб орієнтовного визначення середньоквадратичного відхилення випадкової величини: беруть максимальне практично можливе відхилення від середнього і поділяють його на три. Зрозуміло, цей спосіб може бути рекомендований, тільки якщо немає інших, більш точних способів визначення σ .

Розподіл хі-квадрат. Нехай $X_i, i = \overline{1, n}$ – нормальні нормовані незалежні величини, тобто їх математичне сподівання дорівнює нулю, середньоквадратичне відхилення – одиниці і кожна з них розподілена за нормальним законом. Тоді сума квадратів цих величин

$$\chi^2 = \sum_{i=1}^n X_i^2$$

розподілена за законом χ^2 з $k = n$ степенями вільності.

Якщо величини X_i пов'язані одним лінійним співвідношенням, наприклад $\sum_{i=1}^n X_i = n\bar{X}$, то число степеня вільності $k = n - 1$.

Диференційна функція розподілу χ^2 має вигляд

$$f(x) = \begin{cases} 0, & x \leq 0; \\ \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} e^{-\frac{x}{2}} x^{\frac{k}{2}-1}, & x > 0, \end{cases}$$

де $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ – гамма-функція; $\Gamma(n+1) = n!$

Розподіл χ^2 визначається числом степенів вільності k . Коли k зростає, розподіл χ^2 прямує до нормального розподілу дуже повільно.

Розподіл Стьюдента. Нехай X – нормальна нормована випадкова величина, а y – незалежна від X величина, яка розподілена за законом хі-квадрат з k -м степенем вільності. Тоді величина

$$T = \frac{X}{\sqrt{y/k}}$$

має розподіл, який називається t -розподілом або розподілом Стьюдента з k степенями вільності.

Із зростанням k розподіл Стьюдента швидко наближається до нормального розподілу.

4.6. ВИПАДКОВІ ПРОЦЕСИ

Закони розподілу випадкового процесу. У реальних умовах при передаванні повідомлень сигнал у точці приймання заздалегідь невідомий і тому не може бути описаний детермінованою (з відомою ймовірністю) функцією часу. Те ж саме можна сказати і про завади, поява яких у каналі зв'язку може бути обумовлена всілякими і найчастіше невідомими для нас причинами.

Отже, реальні сигнали і завади являють собою випадкові процеси. Випадковий процес описується випадковою функцією, значення якої при будь-якому значенні аргументу є випадковими величинами. Аргументом функції може бути величина будь-якої фізичної природи. Для електричних сигналів такою величиною звичайно є час.

При незмінних умовах досліду випадковий процес може набирати тої чи іншої конкретної форми $\xi_k(t)$ (рис. 4.8, а). Ці можливі форми випадкового процесу називаються його *реалізаціями*. Сукупність усіх можливих реалізацій $\xi_k(t)$ випадкового процесу $\xi_k(t)$ називається *ансамблем*. Реалізації $\xi_k(t)$ вже є не випадковими, а детермінованими функціями. Однак передбачити, якою буде реалізація процесу в кожному конкретному випадку, неможливо.

Якщо розглядати не кожную реалізацію окремо, а їх сукупність, то можна визначити ймовірнісні характеристики випадкового процесу. Такими характеристиками є закони розподілу, що можуть бути отримані теоретично на основі експериментальних даних.

Нехай $\xi_k(t)$ – випадковий процес. У деякий фіксований момент часу t_1 різні реалізації процесу матимуть різні значення: $\xi_1(t_1), \xi_2(t_1), \dots, \xi_n(t_1)$. Значення $\xi(t_1)$ – випадкові величини.

Одновимірний закон розподілу (інтегральний закон) визначається як ймовірність того, що випадкова величина $\xi(t_1)$ не перевищить деякого значення x_1 :

$$F_1(x_1, t_1) = P \{ \xi(t_1) \leq x_1 \} .$$

Похідна від $F_1(x_1, t_1)$ називається *одновимірною щільністю* ймовірності випадкового процесу $\xi(t)$ для $t = t_1$.

Двовимірним інтегральним законом розподілу $F_2(x_1, t_1; x_2, t_2)$ випадкового процесу $\xi(t)$ називається ймовірність того, що в момент часу t_1 функція $\xi(t)$ не перевищить деякого значення x_1 , а в момент t_2 – значення x_2 :

$$F_2(x_1, t_1; x_2, t_2) = P \xi(t_1) \leq x_1, \xi(t_2) \leq x_2 .$$

Двовимірна щільність імовірності визначається як частинна похідна (якщо вона існує) другого порядку:

$$P_2(x_1, t_1; x_2, t_2) = \frac{\partial^2 F_2(x_1, t_1; x_2, t_2)}{\partial x_1 \partial x_2} .$$

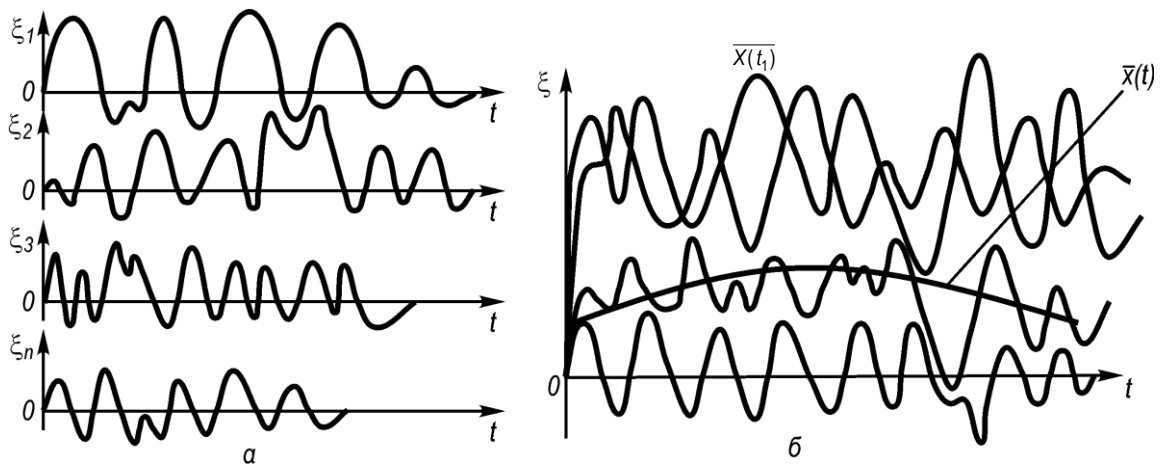


Рис. 4.8. Випадкові процеси:

a – ансамбль реалізацій; *б* – математичне сподівання сукупності реалізацій

Аналогічно визначаються тривимірний, чотиривимірний і т. д. закони розподілу ймовірності та її щільності.

Найбільш повною характеристикою випадкового процесу є n -вимірний закон розподілу, тобто розподіл $\xi(t)$ для n довільно обраних моментів часу. Функція

$$F_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = P \xi(t_1) \leq x_1; \dots; \xi(t_n) \leq x_n$$

називається n -вимірним інтегральним законом розподілу, а частинна похідна

$$\frac{\partial^n F_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n} = P(x_1, t_1; x_2, t_2; \dots; x_n, t_n)$$

називається n -вимірною щільністю ймовірності.

Випадковий процес заданий, якщо його n -вимірний розподіл відомий для будь-яких n довільно обраних моментів часу t_1, t_2, \dots, t_n . Якщо значення випадкової функції $\xi(t)$ при будь-яких значеннях t незалежні, то

$$P(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = P_1(x_1, t_1) P_2(x_2, t_2) \dots P_n(x_n, t_n).$$

Звідси випливає, що вичерпною характеристикою випадкової функції з незалежними значеннями є її одновимірний закон розподілу.

Закони розподілу є досить повними характеристиками випадкового процесу. Однак вони складні і вимагають для свого визначення оброблення великої кількості експериментального матеріалу. Крім того, такий докладний опис процесу не завжди буває потрібним. Для розв'язання багатьох практичних задач достатньо знати більш прості (хоча і менш повні) характеристики випадкового процесу, якими є середнє значення і функція кореляції випадкового процесу.

Середнє значення, або математичне сподівання випадкового процесу $X(t)$ визначається за формулою

$$\overline{X(t_1)} = \int_{-\infty}^{\infty} x_1 P_1(x_1, t_1) dx_1,$$

де риска означає усереднення за множиною, тобто деяку середню функцію, навколо якої групуються всі можливі реалізації випадкової функції (рис. 4.7, б). Аналогічно визначається середнє значення квадрата

$$\overline{X^2(t_1)} = \int_{-\infty}^{\infty} x_1^2 P_1(x_1, t_1) dx_1.$$

Дисперсія випадкового процесу визначається як математичне сподівання квадрата відхилення процесу від свого середнього значення:

$$D[X(t_1)] = \overline{[X(t_1) - \overline{X(t_1)}]^2} = \int_{-\infty}^{\infty} [x(t_1) - \overline{x(t_1)}]^2 P_1(x_1, t_1) dx_1.$$

Дисперсія є мірою розкиду значення випадкової функції навколо середнього значення. При $\overline{X(t)} = 0$ дисперсія збігається з математичним сподіванням квадрата випадкового процесу:

$$D[X(t_1)] = \overline{X^2(t_1)} = \delta_x^2.$$

Середнє значення і дисперсія характеризують перебіг випадкового процесу в окремі моменти часу.

Статистичний зв'язок між значеннями випадкового процесу $X(t_1)$ і $X(t_2)$ у момент часу t_1 і t_2 описується функцією кореляції. Ця функція визначається як математичне сподівання добутку цих значень:

$$B_x(t_1, t_2) = \overline{X(t_1)X(t_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 P_2(x_1, t_1; x_2, t_2) dx_1 dx_2.$$

Статистичний зв'язок двох випадкових процесів $X(t)$ і $Y(t)$ в моменти t_1 і t_2 описується функцією взаємної кореляції

$$B_{xy}(t_1, t_2) = \overline{X(t_1)Y(t_2)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy P_2(x, t_1; y, t_2) dx dy.$$

Якщо випадкові функції $X(t)$ і $Y(t)$ статистично незалежні, то двовимірна щільність імовірності розпадається на добуток двох одновимірних:

$$P_2(x, t_1; y, t_2) = P_1(x, t_1)P_1(y, t_2).$$

Якщо $\overline{X(t_1)} = \overline{Y(t_2)} = 0$, то $B_{xy}(t_1, t_2) = 0$.

Для деяких процесів (зокрема, з нормальним розподілом імовірностей) середнє значення і функція кореляції є вичерпними характеристиками.

Нормована функція кореляції (коефіцієнт кореляції) визначається виразом

$$P(t_1, t_2) = \frac{B_x(t_1, t_2)}{\sqrt{B_x(t_1, t_1)B_x(t_2, t_2)}}.$$

У деяких випадках шляхом оброблення однієї реалізації можна визначити середнє значення за часом випадкового процесу на інтервалі T згідно з формулою

$$\tilde{x}_T(t) = \frac{1}{T} \int_0^T x(t) dt,$$

де хвиляста риска означає усереднення за часом. Середнє значення квадрата за часом

$$\tilde{x}_T^2(t) = \frac{1}{T} \int_0^T \tilde{x}^2(t) dt.$$

При $T \rightarrow \infty$ останній вираз визначає середню потужність процесу, що виділяється на опорі 1 Ом.

Слід мати чітке уявлення про усереднення по ансамблю і усереднення за часом. Для визначення середнього по ансамблю потрібно мати у своєму розпорядженні великий набір реалізацій розглянутого процесу. Оброблення цих реалізацій дозволяє обчислити $B_x(t_1, t_2)$, $\overline{X(t_1)}$ і т. д. Але одержання цілого ансамблю реалізацій можливе за наявності ансамблю однакових систем, тобто систем, у яких відтворені одні й ті самі способи реалізації сигналу та однакові умови перебігу випадкового процесу. Постановка такого експерименту дуже складна. Звичайно експериментатор має одну установку і за даний проміжок часу (від 0 до T) одержує одну реалізацію випадкового процесу $X(t)$, за якою визначає середнє за часом.

Для деяких процесів, які ми зараз розглянемо, обидва усереднення – по ансамблю і за часом – при достатньо великому значенні T збігаються.

Стаціонарні випадкові процеси. Великий клас практично важливих випадкових процесів складають стаціонарні процеси. *Процес називається стаціонарним у вузькому розумінні, якщо n -вимірний закон розподілу при будь-якому n залежить тільки від інтервалу $t_i - t_j$ і не залежить від положення цих інтервалів на осі часу, тобто*

$$P_n(x_1, t_1; x_2, t_2; \dots; x_n, t_n) = P_n(x_1, t_1 + \tau; \dots; x_n, t_n + \tau).$$

Іншими словами, *ймовірнісні характеристики стаціонарного процесу не змінюються при зміні початку відліку часу. Очевидно, що всі одновимірні щільності ймовірності ідентичні і не залежать від часу:*

$$P_1(x_1, t_1) = P_1(x_1, t_1 + \tau) = P(x).$$

а всі двовимірні щільності ймовірності можуть залежати тільки від інтервалу $\tau = t_2 - t_1$:

$$P_2(x_1, t_1; x_2, t_2) = P(x_1, t_1 + \tau; x_2, t_2 + \tau) = P_2(x_1, t_2, \tau).$$

Випадковий процес називають стаціонарним у широкому розумінні (за О. Я. Хінчином), якщо математичне сподівання $\overline{X(t)}$ і математичне сподівання квадрата випадкової величини не залежать від часу (постійні), а кореляційна функція залежить тільки від різниці аргументів $\tau = t_2 - t_1$:

$$B_x(t_1, t_2) = B_x(t_2 - t_1) = B_x(\tau).$$

Теорія, що використовує для опису випадкових процесів $\overline{X(t)}$, $\overline{X^2(t)}$ і $B_x(\tau)$, називається **кореляційною теорією** випадкових процесів.

Якщо процес стаціонарний у вузькому розумінні, то він стаціонарний і в широкому розумінні, але не навпаки.

Для стаціонарних випадкових процесів є слушною ергодична теорема, відповідно до якої усереднення по ансамблю можна замінити усередненням за часом; тобто з імовірністю, як завгодно близькою до одиниці, можна вважати, що

$$\begin{aligned} \overline{X(t)} &= \int_{-\infty}^{\infty} xP(x)dx = \lim_{T \rightarrow 0} \frac{1}{T} \int_0^T x(t)dt = \overline{X(t)}; \\ \overline{X^2(t)} &= \int_{-\infty}^{\infty} x^2 P(x)dx = \lim_{T \rightarrow 0} \frac{1}{T} \int_0^T x^2(t)dt = \overline{X^2(t)}; \\ B_x(t) &= \overline{X(t)X(t+\tau)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 P_2(x_1, x_2, \tau) dx_1 dx_2 = \\ &= \lim_{T \rightarrow 0} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt = \overline{X(t)X(t+\tau)}; \end{aligned}$$

Властивість ергодичності має велике практичне значення і дає змогу обходитися однією реалізацією.

Найважливішою характеристикою ергодичного процесу є його кореляційна функція. Пояснимо фізичний зміст цієї функції. Нехай $x(t)$ є електрична напруга на виході приймача. Якщо в момент часу t значення $x(t)$ велике, то малоімовірно, що в момент $t+\tau$, де τ – мала величина, напруга стане дуже малою. Якщо ж узяти τ досить великим, то величина $x(t+\tau)$ може стати вже будь-якою. Іншими словами, між $x(t)$ і $x(t+\tau)$ існує залежність, що слабшає зі збільшенням τ . Мірою цієї залежності може бути середньоквадратичне відхилення $x(t)$ і $x(t+\tau)$, тобто

$$\varepsilon^2(\tau) = \overline{[x(t) - x(t+\tau)]^2} = 2[B_x(0) - B_x(\tau)].$$

Звідси випливає, що середньоквадратичне відхилення $\varepsilon^2(\tau)$ цілком визначається кореляційною функцією, яка вказує, наскільки в середньому пов'язані одне з одним два значення ергодичного процесу, розділені інтервалом часу τ .

Властивості кореляційної функції.

1. Автокореляційна функція ергодичного процесу, яка визначає статистичний зв'язок між значеннями випадкових величин в різні моменти часу, – функція парна:

$$B_x(\tau) = B_x(-\tau).$$

3. Значення кореляційної функції ергодичного процесу при $\tau=0$ збігається із середньою потужністю цього процесу:

$$B_x(\tau)|_{\tau=0} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt |_{\tau=0} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x^2(t)dt = x^2(t) = \delta^2.$$

3. Будь-яке значення кореляційної функції не може перевищувати значення цієї функції при $\tau = 0$, тобто $B_x(0) \geq B_x(\tau)$. Це випливає з нерівності

$$[x(t) - x(t+\tau)]^2 = x^2(t) - 2x(t)x(t+\tau) + x^2(t+\tau) = 2B_x(0) - 2B_x(\tau) \geq 0.$$

1. Модуль нормованої кореляційної функції не перевищує одиниці:

$$\rho_x(\tau) = \frac{B_x(\tau)}{B_x(0)} \leq 1.$$

5. Якщо автокореляційна функція процесу задовольняє умови $B_x(0) \neq 0$ при $\tau = 0$; $B_x(0) = 0$ при $|\tau| > 0$, то це означає, що між значеннями $x(t)$ і $x(t+\tau)$ не існує зв'язку. Такі процеси називаються *чисто випадковими*.

6. Якщо ергодичний процес не містить детермінованої складової, то його кореляційна функція нескінченно спадає зі збільшенням τ , тобто залежність між значеннями $x(t)$ і $x(t+\tau)$ слабшає і при $\tau \rightarrow 0$ вони стають незалежними. Якщо в процесі $x(t) = \xi(t) + x_0$ є стала (детермінована) складова, то

$$\lim_{\tau \rightarrow \infty} B_x(\tau) = \lim_{\tau \rightarrow \infty} [\xi(t) + x_0][\xi(t+\tau) + x_0] = x_0^2.$$

7. Функція автокореляції періодичного процесу є періодичною з періодом цього процесу (рис. 4.9, б)

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega_1 + \varphi_k); \quad B_x(\tau) = A_0^2 + \sum_{k=1}^{\infty} \frac{A_k^2}{2} \cos k\omega_1 \tau.$$

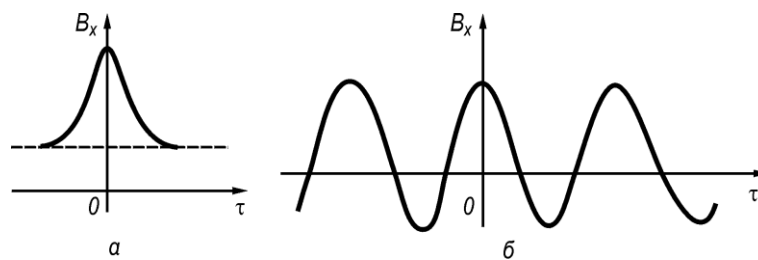


Рис. 4.9. Кореляційні функції та енергетичні спектри:
 а – кореляційна функція випадкового сигналу, що містить детерміновану складову; б – кореляційна функція періодичного сигналу

На рис. 4.9, а наведена автокореляційна функція випадкового процесу (сигналу), що містить сталу складову.

При цьому кореляційна функція не залежить від фазових кутів гармонік вихідного періодичного процесу.

Для випадкового процесу, що не містить детермінованих складових, можна вказати такий інтервал $\Delta\tau$, на якому при $\tau > \Delta\tau$ взаємний зв'язок між значеннями $x(t)$ і $x(t+\tau)$ буде несуттєвим (їх можна вважати некорельованими). Цей інтервал називається *інтервалом кореляції*. Його звичайно

визначають шириною основи прямокутника з висотою $B_x(0)$, а площа його дорівнює площі під кривою кореляційної функції:

$$\Delta\tau = \frac{1}{B(0)} \int_{-\infty}^{\infty} |B(\tau)| d\tau = \int_{-\infty}^{\infty} |\rho(\tau)| d\tau.$$

Енергетичний спектр сигналу. Спостерігаючи за перебігом випадкового процесу, ми можемо визначити лише спектр даної реалізації $x(t)$, тобто

$$S_T(j\omega) = \int_0^T x(t) e^{-j\omega t} dt.$$

Ця функція є випадковою. Щоб перейти до невідповідної функції частоти, введемо поняття *енергетичного спектра*. Енергетичний спектр стаціонарного випадкового процесу визначається як спектр його функції кореляції:

$$G(\omega) = \int_{-\infty}^{\infty} B(\tau) e^{-j\omega\tau} d\tau;$$

$$B(\tau) = \frac{1}{2T} \int_{-\infty}^{\infty} G(\omega) e^{-j\omega\tau} d\omega.$$

Оскільки $G(\omega)$, $B(\tau)$ – парні функції своїх аргументів, то ці формули можна записати в іншому вигляді:

$$G(\omega) = 2 \int_0^{\infty} B(\tau) \cos \omega\tau d\tau;$$

$$B(\tau) = \frac{1}{2T} \int_0^{\infty} G(\omega) \cos \omega\tau d\omega.$$

Фізичний зміст функції легко з'ясувати, взявши $\tau = 0$. Тоді

$$B(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) d\omega = P,$$

де P – повна потужність процесу.

Ця формула показує, що *функція $G(\omega)$ виражає спектральну щільність потужності процесу*. Потужність у смузі $\Delta\omega = \omega_2 - \omega_1$ можна визначити інтегруванням $G(\omega)$ у межах від ω_1 до ω_2 :

$$P_{\omega_1\omega_2} = \frac{1}{2\pi} \int_{\omega_1}^{\omega_2} G(\omega) d\omega.$$

Енергетичний спектр можна виразити і через поточний спектр реалізації. Відповідно до рівності Парсеваля енергія процесу $x(t)$, яка виділяється за час T , дорівнює

$$E_T = \int_0^T x^2(t) dt = \frac{1}{\pi} \int_0^{\infty} |S_T(j\omega)|^2 d\omega.$$

Середня потужність процесу визначається границею відношення E_T/T при $T \rightarrow \infty$:

$$P = \lim_{T \rightarrow \infty} \frac{E_T}{T} = \frac{1}{\pi} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T |S_T(j\omega)|^2 d\omega.$$

Оскільки $B(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) d\omega = P$, одержимо

$$G(\omega) = \lim_{T \rightarrow \infty} \frac{2 |S_T(j\omega)|^2}{T}.$$

Це співвідношення встановлює зв'язок між енергетичним спектром процесу і поточним спектром його реалізації. Енергетичний спектр характеризує поведінку реалізацій у середньому. Так, якщо $G(\omega)$ зосереджено в області низьких частот, то процес змінюється повільно в порівнянні з тим процесом, спектр якого зосереджений в області більш високих частот. Для вузькосмугового процесу $G(\omega)$ помітно відрізняється від нуля в смузі $\Delta\omega$ навколо середньої частоти ω_0 , причому $\Delta\omega \ll \omega_0$. Такий процес нагадує синусоїду з повільно змінною амплітудою і фазою.

Випадковий процес, у якого $G(\omega) = N_0/2 = \text{const}$, де N_0 – спектральна щільність потужності сигналу, називається білим шумом. Кореляційна функція білого шуму дорівнює

$$B(\tau) = \frac{N_0}{4\pi} \int_{-\infty}^{\infty} e^{j\omega\tau} d\omega = \frac{1}{2} N_0 \delta(\tau).$$

Для випадкових процесів має місце зв'язок загального характеру між шириною спектра $\Delta\omega$ та інтервалом кореляції $\Delta\tau$:

$$\Delta\omega \Delta\tau = \text{const}.$$

Геометрично $\Delta\omega$ – це ширина основи прямокутника з площею під кривою $G(\omega)$ і висотою $G(0)$.

У загальному випадку, чим ширший енергетичний спектр, тим вужча кореляційна функція, і навпаки. Якщо білий шум з рівномірним енергетичним спектром пропустити через ідеальний ФНЧ з граничними частотами $\omega_H = 0, \omega_B = 2\pi f_B$, то на виході одержимо шум з обмеженим спектром, ширина якого

$$\Delta\omega = \omega_B = 2\pi f_B.$$

Функція кореляції

$$B(\tau) = \frac{1}{\pi} \int_0^{\omega_B} \frac{N_0}{2} \cos \omega\tau d\omega = N_0 \frac{\sin \omega_B \tau}{\omega_B \tau}.$$

Тут інтервал кореляції $\Delta\tau = 1/(2f_B)$.

Енергетичний спектр випадкового процесу на виході лінійної системи з коефіцієнтом передачі $K(j\omega)$ визначається виразом

$$G_{\text{вих}}(\omega) = |K(j\omega)|^2 G_{\text{вх}}(\omega),$$

де $G_{\text{вх}}(\omega)$ – енергетичний спектр процесу на вході лінійної системи.

Кореляційна функція на виході лінійної системи

$$\begin{aligned} B_{\text{вих}}(\tau) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\text{вих}}(\omega) e^{j\omega\tau} d\omega = \frac{1}{\pi} \int_0^{\infty} G_{\text{вих}}(\omega) \cos \omega\tau d\omega = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} |K(j\omega)|^2 G_{\text{вх}}(\omega) \cos \omega\tau d\omega. \end{aligned}$$

Потужність шуму на виході лінійної системи при $G_{\text{вх}}(\omega) = N_0/2$ дорівнює

$$P_{\text{ш}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{\text{вих}}(\omega) d\omega = \frac{N_0}{4\pi} \int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega = N_0 \frac{\Delta\omega_e}{2\pi} |K(j\omega)|_{\text{max}}^2,$$

де $\Delta\omega_e = 2\pi\Delta f_e = \frac{\int_{-\infty}^{\infty} |K(j\omega)|^2 d\omega}{|K(j\omega)|_{\text{max}}^2}$ – еквівалентна смуга пропускання лінійного пристрою.

Статистичні характеристики флуктуаційної завади. Серед адитивних завад особливе місце займає флуктуаційна завада, що є процесом з нормальним розподілом. Така завада найбільш вивчена, становить великий інтерес як з теоретичної, так і практичної точок зору, має місце майже в усіх каналах зв'язку. Сума великого числа будь-яких завад від різних джерел також носить характер флуктуаційної завади. Багато завад при проходженні через приймальний пристрій часто набувають властивостей флуктуаційних.

Енергетичну структуру флуктуаційної завади можна уявити собі як послідовність нескінченно коротких імпульсів, що мають випадкову амплітуду і прямують один за одним через випадковий проміжок часу. При цьому імпульс з'являється настільки часто, що перехідні явища в приймачі від окремих імпульсів накладаються, створюючи неперервний випадковий процес.

З фізичної точки зору випадкові завади породжуються різного роду флуктуаціями, тобто випадковими відхиленнями тих чи інших фізичних величин від їхніх середніх значень. Так, джерелом шуму в електричних колах може бути флуктуація струму, зумовлена дискретною природою носія заряду (електронів, іонів). Дискретна природа електричного струму проявляється в електронних лампах у вигляді дробового ефекту. Найбільш розповсюдженою причиною шуму є флуктуація, зумовлена тепловим рухом. Тривалість імпульсів, які складають флуктуаційну заваду, дуже мала, тому спектральна щільність є сталою аж до дуже високих частот. Типовим прикладом флуктуаційних завад є внутрішні шуми приймача; флуктуаційний характер мають космічні завади, а також деякі види атмосферних та індустріальних завад.

Як зазначалося вище, флуктуаційна завада підпорядковується нормальному закону розподілу. Відомо, що одновимірною щільністю ймовірності гауссового (тобто нормального) процесу визначається так:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_3} e^{-\frac{(x-m)^2}{2\sigma_3^2}},$$

де m – середнє значення процесу; σ_3^2 – його дисперсія.

Для флуктуаційної завади додатні та від'ємні значення x зустрічаються однаково часто, тому $m=0$. Дисперсія σ_3^2 в цьому разі дорівнює середній потужності P_3 , а ефективне значення завади $u_{3, \text{эф}} = \sqrt{P_3} = \sigma_3$. Тоді

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_3} e^{-\frac{x^2}{2\sigma_3^2}}.$$

Відповідно інтегральна функція розподілу

$$F(u_0) = P(u < u_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_0} e^{-\frac{u^2}{2}} du = \frac{1}{2}[1 + \Phi(u_0)],$$

де $u = \frac{x}{\sigma_3}$ – відносне значення завади; $\Phi(u_0) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt$.

Функція Φu називається інтегралом імовірності або функцією Крампа.

Функція Φu є непарною: $\Phi -u = -\Phi u$; крім того, $\Phi \infty = 1$ та $\Phi 0 = 0$.

Знаючи закон розподілу ймовірностей, легко знайти ймовірність того, що рівень завади набуде значень, які лежать у заданих межах, наприклад, між u_1 і u_2 :

$$P(u_1 < u < u_2) = \int_{u_1}^{u_2} P(u) du \quad \text{або} \quad P(u_1 < u < u_2) = \frac{1}{2}[\Phi(u_2) - \Phi(u_1)].$$

Імовірність того, що завада перевищить деякий граничний рівень u_0 , можна визначити безпосередньо з попередньої формули, якщо в останній взяти $u_2 = \infty$ і $u_1 = u_0$, де u_0 – значення амплітуди сигналу:

$$P(u > u_0) = \frac{1}{2}[\Phi(\infty) - \Phi(u_0)] = \frac{1}{2}[1 - \Phi(u_0)].$$

Розрахунки показують, що ймовірність перевищення завадою граничного значення швидко спадає зі збільшенням u_0 . Так, ймовірність перевищення при $u_0 = 1$ дорівнює 0,16, при $u_0 = 3$ ця ймовірність дорівнює 0,0013, а при $u_0 = 4$ вона дорівнює $3,5 \cdot 10^{-5}$. Звідси випливає, що коливання завади практично не перевищують її триразового ефективного значення. Пік-фактор завади, що є відношенням максимального значення до ефективного, звичайно лежить у межах 3,4–4,6. Тому флуктуаційну заваду часто називають гладкою на відміну від імпульсних завад, для яких це відношення звичайно набагато більше.

Флуктуаційна завада називається білим шумом, якщо її значення x_i статистично незалежні. Енергетичний спектр такої завади є рівномірним. Слід мати на увазі, що поняття про білий шум є ідеалізацією, тому що енергетичний спектр реальних процесів завжди спадає з ростом частоти, а інтервал кореляції має скінченне значення $\Delta\tau \neq 0$. Ця ідеалізація застосовна в тих випадках, коли інтервал кореляції завади набагато менший за постійну часу системи, на яку вона впливає, чи, інакше, коли в межах смуги пропускання системи спектральну щільність завади можна приблизно вважати сталою.

Флуктуаційний випадковий процес називається вузькосмуговим, коли ширина спектра процесу відносно мала в порівнянні з середньою частотою цього спектра. Такого роду процеси мають місце на виході схем, що працюють на високих і проміжних частотах.

Якщо спостерігати реалізацію вузькосмугового процесу на екрані осцилографа, то вона виглядатиме як синусоїда з повільно змінною амплітудою (обвідною) і фазою.

Поняття про марковські випадкові процеси. Розглянемо послідовність дослідів $1, 2, 3, \dots, \infty$, у кожному з яких можуть відбутися події $A_1, A_2, \dots, A_n, \dots$. Для будь-якого дослідів допускається залежність настання події від результату попереднього дослідів і тільки від нього. Іншими словами, якщо в попередньому досліді відбулася подія A_j , то ймовірність настання події A_k у наступному досліді дорівнює P_{jk} . Для повного визначення такої схеми крім ймовірності P_{jk} мають бути задані ймовірності настання подій у початковому досліді, який ми позначимо a_k . Описана схема називається **ланцюгом Маркова**.

В аналізі ланцюгів Маркова вводиться спеціальна термінологія. Можливі наслідки кожного дослідів називаються *станами системи*, що нумеруються або $1, 2, \dots, j, \dots$, або $0, 1, \dots, j, \dots$ і позначаються E_j . Вважається, що дослідів відбуваються через рівні інтервали часу, тому номер дослідів служить часовим параметром. За цією термінологією говорять, що в n -й момент часу на n -му кроці система знаходиться в j -му стані; P_{jk} – ймовірність безпосереднього переходу зі стану j у стан k за один крок. Числа P_{jk} називаються *перехідними ймовірностями* і можуть скласти матрицю перехідних ймовірностей P з очевидною властивістю: для кожного j $\sum_k P_{jk} = 1$.

Поряд з ймовірностями безпосереднього переходу важливими є ймовірності переходу зі стану j у стан k точно за n кроків, що позначаються $P_{jk}^{(n)}$.

Уведемо тепер класифікацію стану.

Стан E_k досяжний зі стану E_j , якщо існує таке $n \geq 0$, що $P_{jk}^{(n)} > 0$. Іншими словами, у E_k можна потрапити, відправляючись з E_j .

Множина станів називається замкненою, якщо ніякий стан поза цією множиною не досяжний з жодного стану усередині цієї множини. Іншими словами, якщо система попадає в деякий стан усередині замкненої множини, то при подальшому розвитку процесу вона ніколи не вийде з цієї множини.

Ланцюг Маркова називається незвідним, якщо не існує замкнених множин, відмінних від множини всіх станів. Розбивши марковський ланцюг на незвідні ланцюги, кожен з них можна аналізувати незалежно від інших.

Позначимо через $f_{jk}^{(n)}$ ймовірність того, що в процесі, який починається зі стану E_j , перший перехід в стан E_k відбудеться точно на n -му кроці. Оскільки для різних n такі події несумісні, то $f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}$ означає ймовірність того, що система, виходячи з E_j , хоча б на одному кроці пройде через E_k . Тому f_{jk} – це ймовірність повернення в E_j . Якщо n – номер кроку, на якому система, вийшовши з E_j , уперше повертається в E_j , то n розглядається як випадкова величина з математичним сподіванням $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$ і середнім часом повернення.

Стан E_j називається оборотним, якщо $f_{jk} = 1$, і необоротним, якщо $f_{jk} < 1$. Якщо $\mu_j < \infty$, то оборотний стан називається *нульовим*, а якщо $\mu_j > \infty$, – *ергодичним*.

Теорема теорії ймовірностей [9] доводить неможливість існування марковських ланцюгів зі станами різних типів.

У практиці найбільш істотним є випадок ергодичних ланцюгів, які існують незалежно від початкового стану ймовірності того, що ланцюг знаходиться в стані E_j . Задачею застосування теорії є знаходження цих ймовірностей. Більш точною є наступна теорема: *у ще не зведеному ергодичному ланцюзі для $P_{jk}^{(n)}$ існує границя $\pi_k = \lim_{n \rightarrow \infty} P_{jk}^{(n)}$, що не залежить від j .* Ця границя

називається *стаціонарною ймовірністю того, що система знаходиться в стані E_k* . Ймовірності того, що $\pi_k > 0$, можуть бути знайдені з системи рівнянь

$$\begin{cases} \pi_k = \sum_i \pi_i P_{ik}; \\ \sum_k \pi_k = 1. \end{cases}$$

Узагальненням розглянутого ланцюга є ланцюг, для якого система перебуває в кожному стані деякий випадковий час, а потім переходить в інший стан з ймовірністю P_{jk} . Доведено, що коли час перебування в кожному стані має показниковий розподіл з параметром λ_j , який залежить, можливо, від номера стану, то отриманий випадковий процес знову є марковським.

Класифікація станів залишається тою самою, а стаціонарні ймовірності можуть бути знайдені з системи рівнянь

$$\begin{cases} \pi_k \lambda_k = \sum_i \pi_i \lambda_k P_{ik}; \\ \sum_k \pi_k = 1. \end{cases}$$

Останнє рівняння в цій системі називається *умовою нормування*, а перше – *рівнянням глобального балансу*.

Контрольні запитання для самооцінки рівня знань

1. Дайте визначення теорії ймовірності.
2. Які явища називаються випадковими?
3. Що таке ймовірність події і частота події?
4. Дайте означення неможливої і вірогідної події.
5. Дайте означення повної групи подій.
6. Чому дорівнює ймовірність повної групи несумісних подій?
7. Що таке практично неможлива і практично вірогідна подія?
8. Сформулюйте теореми додавання і множення ймовірностей.
9. Напишіть формулу повної ймовірності.
10. Сформулюйте теорему гіпотез і запишіть формулу Бейеса.
11. Які закони розподілу випадкових величин Ви знаєте?
12. Назвіть властивості інтегрального закону розподілу ймовірностей.
13. Які властивості має диференціальний закон розподілу випадкових величин?
14. Які числові характеристики випадкових величин Ви знаєте?
15. Напишіть вирази для математичного сподівання дискретної і неперервної випадкових величин.
16. Що таке моменти і дисперсія випадкових величин?
17. Як визначається середньоквадратичне відхилення?
18. Які особливості має закон рівномірної щільності?
19. Напишіть вирази для закону розподілу Пуассона.
20. Як залежить від математичного сподівання і дисперсії крива нормального закону?
21. Сформулюйте правило трьох сигма.
22. Дайте означення випадкового процесу.
23. Який випадковий процес є стаціонарним?
24. Як визначаються автокореляційна і кореляційна функції?

25. Перелічіть властивості кореляційної функції.
26. Дайте означення інтервалу кореляції і ширини енергетичного спектра.
27. Що таке реалізація випадкового процесу?
28. Як називається сукупність усіх можливих реалізацій випадкового процесу?
29. Як визначаються одно-, дво-, n -вимірні функції розподілу та щільність імовірності?
30. Як визначається середнє значення випадкового процесу за множиною?
31. Напишіть вирази для середнього значення квадрата випадкового процесу, дисперсії.
32. Запишіть вирази для нормованої функції кореляції.
33. Дайте означення стаціонарності у вузькому і широкому розумінні.
34. У чому полягає ергодична теорема?
35. Як визначається енергетичний спектр? Як пов'язані енергетичний спектр і кореляційна функція?
36. Як пов'язаний енергетичний спектр зі спектром реалізації?
37. Як позначається інтервал кореляції і ширина енергетичного спектра?
38. Які статистичні характеристики флуктуаційної завади Ви знаєте?
39. Дайте означення білого шуму, вузькосмугового випадкового процесу.
40. Як визначити енергетичний спектр і кореляційну функцію на виході лінійної системи?
41. Дайте означення марковського випадкового процесу.
42. У чому особливості випадкового процесу Маркова?
43. Який ланцюг Маркова є незвідним?
44. Що таке стаціонарна ймовірність?
45. Які рівняння називаються умовами нормування; глобального балансу?

Післямова до розділу 4

Тепер ви вже знаєте:

- про теорему додавання і множення ймовірностей;
- теорему гіпотез і формулу Бейеса;
- вирази для математичного сподівання дискретної і неперервної випадкових величин;
- що таке моменти і дисперсія випадкових величин;
- вирази для закону розподілу Пуассона;
- як визначаються автокореляційна і кореляційна функції, їх властивості;
- як визначаються одно-, дво-, n -вимірні функції розподілу та щільність імовірності;
- як визначається середнє значення випадкового процесу за множиною;
- зміст марковського випадкового процесу.

А тепер переходимо до опрацювання тем наступного функціональної частини, висвітлених у навчальному посібнику **Беркман Л.Н., Варфоломєєва О.Г., Грушевська В.П.** “Типові сигнали та завади в електров'язку”.

ЗАДАЧІ ЧАСТИНИ 1

1. Задано таблицю трьох кодів:

Номер коду	Повідомлення							
	а	б	в	г	д	е	ж	з
1	000	001	010	011	100	101	110	111
2	0	1	00	01	10	11	110	111
3	00	01	100	101	1100	1101	1110	1111

Необхідно: 1) визначити вид коду; 2) визначити, які з цих кодів не потребують розділових знаків; чому? 3) закодувати слова: вега, база; 4) розшифрувати послідовності кодових комбінацій:

а) коду 1: 111000001000010000; 001101100000; б) коду 3: 0100101001110; 01110110100; 10000111100.

2. Стартстопний телеграфний апарат передає одну літеру сімома послідовними: однією стартовою (20 мс), п'ятьма кодовими (20 мс кожна) і однією стоповою (30 мс). Визначити швидкість модуляції в бодах і максимальну швидкість передавання інформації R_i в бітах за секунду.

3. Практично вірогідність передавання дискретних повідомлень визначають коефіцієнтом помилок – відносним числом помилково прийнятих елементів повідомлення. Обчисліть цей коефіцієнт для випадку передавання літер кодом МТК-2 зі швидкістю 50 бод, якщо за 2 год неперервного передавання було зафіксовано 10 помилково прийнятих літер.

4. Запишіть кодові слова (комбінації), що відповідають числу 45, при основі коду $m = 2, 3, 4, 8, 10$. Як змінюється число розрядів і число використовуваних цифр у слові зі зміною основи коду?

5. Яке число розрядів має бути в рівномірному коді, призначеному для первинного кодування 32-літерного алфавіту, при основі коду $m = 2, 8, 16, 32$?

6. Текст із 180 літер передається телефонним каналом протягом 30 с. Той самий текст за той же час передається телеграфом кодом МТК-2. Порівняйте об'єми телефонного і телеграфного сигналів (при однакових динамічних діапазонах).

7. Обчислити енергію та норму $|A|$ експоненційного сигналу

$$U(t) = A e^{-\alpha t} \cdot 1(t)$$

№ варіанту	1	2	3	4	5	6	7	8	9	0
A (B)	1	2	2,4	2,2	1,6	1,8	3	2,8	3,2	3,4

Список літератури, рекомендованої для поглибленого вивчення дисципліни

1. *Абакумов В. Г.* Электронные промышленные устройства: Учеб. пособие для вузов. – К.: Вища шк., 1998. – 376 с.
2. *Банкет В. Л., Дорофеев В. М.* Цифровые методы в спутниковой связи. – М.: Радио и связь, 1988. – 448 с.
3. *Барковський В. В., Барковська Н. В., Лопатін О. К.* Математика для економістів: Теорія ймовірностей та математична статистика. – К.: Нац. акад. управління, 1997. – 225 с.
4. *Баскаков С. И.* Радиотехнические цепи и сигналы: Учеб. для вузов. – М.: Высш. шк., 1998. – 448 с.
5. *Варакин Л. Е.* Системы связи с шумоподобными сигналами. – М.: Радио и связь, 1985. – 384с.
6. *Варакин Л. Е.* Теория систем сигналов. – М.: Сов. радио, 1978. – 304 с.
7. *Варакин Л. Е.* Теория сложных сигналов. – М.: Сов. радио, 1970. – 375 с.
8. *Вентцель Е. С.* Теория вероятностей. – М.: Наука, 1969. – 576 с.
9. *Витерби А. Д., Омура Д. К.* Принципы цифровой связи и кодирования. – М.: Радио и связь, 1982. – 526 с.
10. *Волочій Б.Ю.* Передавання сигналів у інформаційних системах. Ч.1.: Нав. Посібник. – Львів: Видавництво Національного університету «Львівська політехніка», 2005. -196 с.

11. *Джури Э.* Импульсные системы автоматического регулирования / Пер. с англ.; Под ред. Я. З. Цыпкина. – М.: Мир, 1963. – 453 с.
12. *Диксон Р. К.* Широкополосные системы / Пер. с англ.; Под ред. В. И. Жуковского. – М.: Связь, 1979. – 303 с.
13. *Железнов Н. А.* Принцип дискретизации стохастических сигналов с неорганичным спектром // Радиотехника и электроника. – 1978. – № 1. – С. 42–44.
14. *Зайцев Г. Ф., Стеклов В. К., Брицький О. І.* Теорія автоматичного управління. – К.: Техніка, 2002. – 688 с.
15. *Зюко А. Г.* Помехоустойчивость и эффективность систем связи. – М.: Связь, 1973. – 359 с.
16. *Зюко А. Г., Коробов Ю. Ф.* Теория передачи сигналов: Учеб. для вузов. – М.: Связь, 1972. – 282 с.
17. *Игнатов В. А.* Теория информации и передачи сигналов: Учеб. для вузов. – М.: Сов. радио, 1979. – 280 с.
18. *Кларк Дж., Кейн Дж.* Кодирование и исправление ошибок в системах цифровой связи / Пер. с англ.; Под ред. Б. С. Цибакова. – М.: Радио и связь, 1987. – 392 с.
19. *Кловский Д. Д.* Теория передачи сигналов. – М.: Связь, 1972. – 282 с.
20. *Кловский Д. Д.* Теория передачи сигналов в задачах. – М.: Связь, 1978. – 252 с.
21. *Кловский Д. Д., Шилкин В. А.* Теория электрической связи: Сб. задач и упражнений. – М.: Радио и связь, 1990. – 280 с.
22. *Коновалов Г. Ф.* Радиоавтоматика: Учеб. для вузов. – М.: Высш. шк., 1990. – 335 с.
23. *Котельников В. А.* Теория потенциальной помехоустойчивости. – Л.: Госэнергоиздат, 1956. – 152 с.
24. *Маркюс Ж.* Дискретизация и квантование. – М.: Энергия, 1969. – 286 с.
25. *Назаров М. В., Кувшинов Б. И., Попов О. В.* Теория передачи сигналов. – М.: Связь, 1970. – 368 с.
26. *Панфілов І. П., Дирда В. Ю., Капацін А. В.* Теорія електричного зв'язку. – К.: Техніка, 1998. – 322 с.
27. *Первачев С. В.* Радиоавтоматика. – М.: Радио и связь, 1982. – 403 с.
28. *Передача дискретных сообщений: Учеб. для вузов / В. П. Шувалов, Н. В. Захарченко, В. О. Шварцман и др.; Под ред. В. П. Шувалова.* – М.: Радио и связь, 1990. – 464 с.
29. *Поляков П. Ф.* Широкополосные аналоговые системы связи со сложными сигналами. – М.: Радио и связь, 1981. – 153 с.
30. *Пугачев В. С.* Статистические методы в технической кибернетике. – М.: Сов. радио, 1971. – 192 с.
31. *Радиоавтоматика: Учеб. для вузов / Г. Ф. Зайцев, Г. Н. Арсеньев и др.* – М.: ООО “ДВК”, 2004. – Т. 1. – 523 с.
32. *Системы фазовой синхронизации / С. Н. Скляренко, В. К. Стеклов, Р. В. Уваров, В. М. Чмиль.* – К.: Техніка, 1994. – 160 с.
33. *Скляренко С. М., Стеклов В. К., Беркман Л. Н.* Поштовый зв'язок. – К.: Техніка, 2004. – 904 с.

34. *Стратонович Р. Л.* К теории оптимальной нелинейной фильтрации случайных функций // Теория вероятностей и ее применение. – 1959. – № 2. – С. 44–49.
35. *Солодов А. В.* Теория информации и ее применение к задачам автоматического управления и контроля. – М.: Наука, 1967. – 436 с.
36. *Стилкер Дж.* Цифровая спутниковая связь / Пер. с англ. В. В. Маркова. – М.: Связь, 1979. – 592 с.
37. *Статистическая динамика радиотехнических следящих систем / С. В. Первачев, А. А. Валуев, В. М. Чиликин.* – М.: Сов. радио, 1973. – 488 с.
38. *Стеклов В. К., Беркман Л. Н.* Проектування телекомунікаційних мереж. – К.: Техніка, 2002. – 792 с.
39. *Стеклов В. К., Беркман Л. Н.* Теорія електричного зв'язку. . – К.: Техніка, 2006. – 552 с.
40. *Теория передачи сигналов: Учеб. для вузов / А. Г. Зюко, Д. Д. Кловский, М. В. Назаров, Л. М. Финк.* – М.: Связь, 1986. – 304 с.
41. *Теория электрической связи: Учеб. для вузов / А. Г. Зюко, Д. Д. Кловский, В. И. Коржик, М. В. Назаров.* – М.: Радио и связь, 1998. – 432 с.
42. *Финк Л. М.* Теория передачи дискретных сообщений. – М.: Сов. радио, 1970. – 727 с.
43. *Філіпський Ю. К.* Випадкові сигнали в радіотехніці. – К.: Вища шк., 1986. – 126 с.
44. *Хлытчиев С. М., Тарасова Н. П., Лившиц В. М.* Теоретические основы почтовой связи: Учеб. для вузов. – М.: Радио и связь, 1990. – 380 с.
45. *Шеннон К.* Работы по теории информации и кибернетике / Пер. с англ.; Под ред. Н. А. Железнова. – М.: Мир, 1963. – 829 с.