МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

И.П. Гуров

ФОРМИРОВАНИЕ И АНАЛИЗ СИГНАЛОВ В СИСТЕМАХ КОМПЬЮТЕРНОЙ ФОТОНИКИ

Учебно-методическое пособие

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 12.03.03 Фотоника и оптоинформатика
в качестве учебно-методического пособия для реализации основных
профессиональных образовательных программ высшего образования
бакалавриата

УНИВЕРСИТЕТ ИТМО

Санкт-Петербург

2018

Гуров И.П. Формирование и анализ сигналов в системах компьютерной фотоники. Учебно-методическое пособие. – СПб: Университет ИТМО, 2018. – 108 с.

Рецензент: Тропченко А.Ю., доктор технических наук, профессор.

Рассмотрены теоретические основы формирования, преобразования и регистрации сигналов в системах фотоники, математические модели сигналов и систем, интегральные преобразования сигналов, методы решения обратных задач при бесконтактном контроле объектов.

Пособие адресовано прежде всего студентам бакалавриата, обучающимся по направлению подготовки 12.03.03 "Фотоника и оптоинформатика" по учебным дисциплинам "Методы компьютерной фотоники", "Теория систем и системный анализ в оптике", и может быть полезно магистрантам, аспирантам (в части материалов разделов 3 и 4) и специалистам в области фотоники.

УНИВЕРСИТЕТ ИТМО

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

Оглавление

Введение

Информационные технологии составляют одно из основных направлений развития современного общества. В технических системах информация передается при помощи сигналов. Основным видом сигналов являются электрические сигналы, однако в последнее время всё большее распространение получают оптические сигналы. Двумерные сигналы в форме изображений представляют собой пространственное распределение интенсивности света и содержат большое количество информации.

Развитие компьютерных технологий, систем регистрации изображений и визуализации данных определяют новые возможности информационных систем с использованием многомерных оптических сигналов. Современные информационные системы основываются на принципах фотоники – области науки и технологий, которая связана с использованием светового излучения (или потока фотонов) в оптических элементах, устройствах и системах, в которых генерируются, преобразуются, распространяются и детектируются оптические сигналы, а также производится их запись или отображение. Активное развитие фотоники при интеграции с компьютерными технологиями создает качественно новые возможности информационных систем применительно к решению широкого круга задач, определяемых приоритетными направлениями развития науки и технологий.

Компьютерная фотоника – это область науки, изучающая методы и технику получения, обработки и анализа оптических сигналов. В отличие от традиционных систем, основанных главным образом на формировании и обработке одномерных сигналов, современные системы компьютерной фотоники обеспечивают получение и совместный анализ последовательности изображений. При этом рассматриваются многомерные оптические сигналы.

Системы фотоники могут быть разделены на два класса: пассивные и активные. Пассивные системы используют только естественное внешнее освещение. Таким системам присущи принципиальные недостатки, такие как существенное влияние внешних условий освещения, а также необходимость использования сложных алгоритмов обработки получаемых изображений. В активных системах используется специализированный источник освещения. К таким системам можно отнести интерферометрические системы, системы, использующие структурированную подсветку и другие. Преимущества выбора источников излучения с различной когерентностью эффективно реализованы, например, в микроскопии и оптической когерентной томографии.

Для решения задач фотоники требуется понимание физических процессов формирования оптических полей. Часто главной задачей

обработки является восстановление информации о свойствах исследуемого объекта по фотоэлектрическим сигналам, полученным при регистрации оптического поля. Для решения этой задачи привлекаются современные математические методы, составляющие основу анализа регистрируемых сигналов. Обработка зарегистрированных сигналов средствами компьютерных технологий требует использования дискретных представлений оптической информации, что накладывает некоторые ограничения на возможности использования методов обработки, основанных на непрерывных математических моделях. Широкое распространения получили методы анализа многомерных сигналов на основе линейной алгебры и матричных представлений.

Формирование и анализ сигналов в фотонике основываются на современных разработках в области высоких технологий фотоники и микроэлектроники с привлечением методов прикладной математики, особенности которых отражены в отдельных разделах настоящего учебного пособия.

Раздел 1. Принципы формирования, преобразования и регистрации оптических сигналов

1.1 Свойства световых полей и формирование изображений

Исследования объектов методами фотоники основываются на детальном анализе особенностей формирования и регистрации оптических полей. При этом на стадии первичного преобразования определяющими являются характеристики взаимодействия излучения с исследуемым объектом при различной степени когерентности излучения. Доступными для анализа и обработки являются фотоэлектрические сигналы. Обеспечение требуемой помехоустойчивости рассматривается как необходимое условие при анализе информационных параметров сигналов с требуемой точностью.

Изображение представляют собой двумерную картину распределения интенсивности оптического излучения. Изображения принято классифицировать с учетом свойств оптических полей, прежде всего, степени когерентности излучения.

Временная когерентность характеризуется временем и длиной когерентности. Время когерентности τ_c – это интервал времени, на протяжении которого возможно наблюдение явления интерференции. Длина когерентности L_c характеризует расстояние, которое проходит электромагнитное излучение в вакууме за время когерентности, а именно

$$L_c = c\tau_c, \tag{1.1.1}$$

где c – скорость света.

Длина когерентности источника излучения соответствует функции когерентности Γ, которая связана со спектральным распределением интенсивности источника излучения $I(\lambda)$ через преобразование Фурье:

$$\Gamma = \mathrm{FT}\{I(\lambda)\}. \tag{1.1.2}$$

На рис. 1.1 показаны примеры гауссова спектра источника излучения и его функция когерентности.

При этом длина когерентности и полуширина спектра связаны известным соотношением:

$$L_c \approx \frac{2\ln 2}{\pi}\frac{\lambda_0^2}{\Delta\lambda}, \tag{1.1.3}$$

где λ_0 – центральная длина волны, а $\Delta\lambda$ – ширина спектра источника излучения.

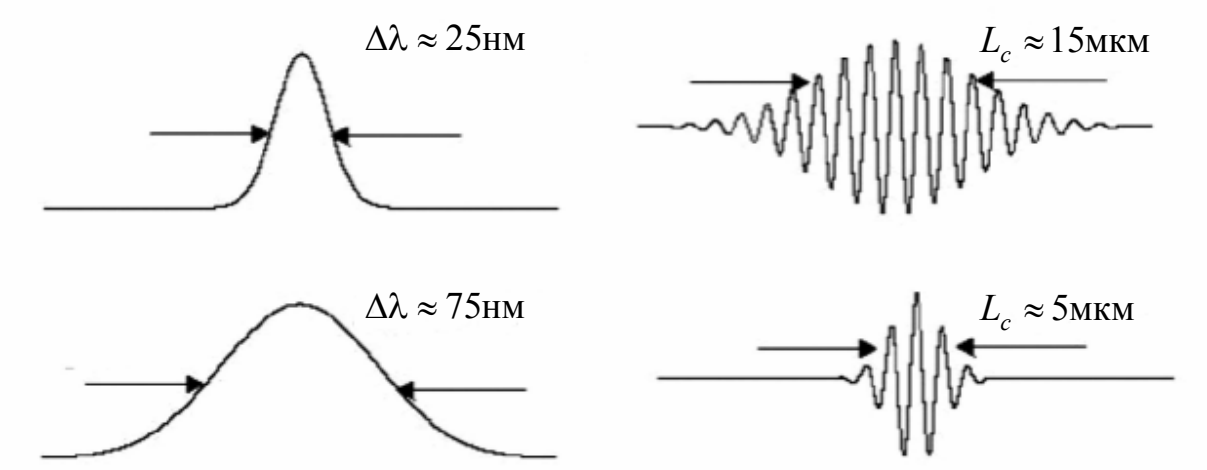


Рис. 1.1. Примеры спектров источников излучения и их функций когерентности

Когерентное излучение с весьма узкой спектральной линией $\Delta\lambda/\lambda_0 << 1$ формируется в лазерных источниках излучения. При этом длина когерентности излучения согласно (1.1.3) может достигать десятков метров.

В качестве примера некогерентного источника можно привести галогенную лампу. Ее спектр содержит все длины волн видимого диапазона, а максимум спектральной интенсивности приходится на инфракрасную область спектра. Длина когерентности излучения галогенной лампы составляет величину порядка 1 мкм. Источники с широким спектром широко используются в таких областях, как интерферометрия малой когерентности и оптическая когерентная томография, в которых разрешение по глубине тем выше, чем больше относительная ширина спектра источника излучения.

В настоящее время нашли широкое применение суперлюминесцентные диоды (СЛД), представленные широкой номенклатурой изделий для различных спектральных диапазонов. Спектр СЛД имеет форму, близкую к гауссовой, диаграмма направленности излучения источника удобно согласуется с оптическими системами. Длина когерентности СЛД составляет десятки микрометров, что обеспечивает возможность их применения для решения многих задач бесконтактного контроля объектов.

Таким образом, в настоящее время имеется широкое разнообразие источников излучения с различной степенью когерентности, и основная задача состоит в выборе модели источника с требуемыми характеристиками.

1.2 Формирование некогерентных изображений

Получение информации о наблюдаемом объекте требует понимания особенностей формирования изображения. Как отметил А. Розенфельд, автор одной из первых монографий по проблематике обработки изображений, «предмет обработки изображений имеет отношение к изображениям не просто как к произвольным функциям и матрицам, а к изображению чего-то *конкретного*, что должно отображать *реальные явления и предметы*». При анализе характеристик объектов необходимо, прежде всего, установить соответствие между объектом и изображением.

Центральная проекция

Для формирования центральной проекции можно использовать экран с точечным отверстием в плоскости изображения, как показано на рис. 1.2. Поскольку свет распространяется прямолинейно, каждой точке изображения (распределения интенсивности света на экране) соответствует направление, определяемое лучом, идущим из этой точки через отверстие. В результате возникает центральная (перспективная) проекция. Простейшим реальным примером такой системы является камера обскура.

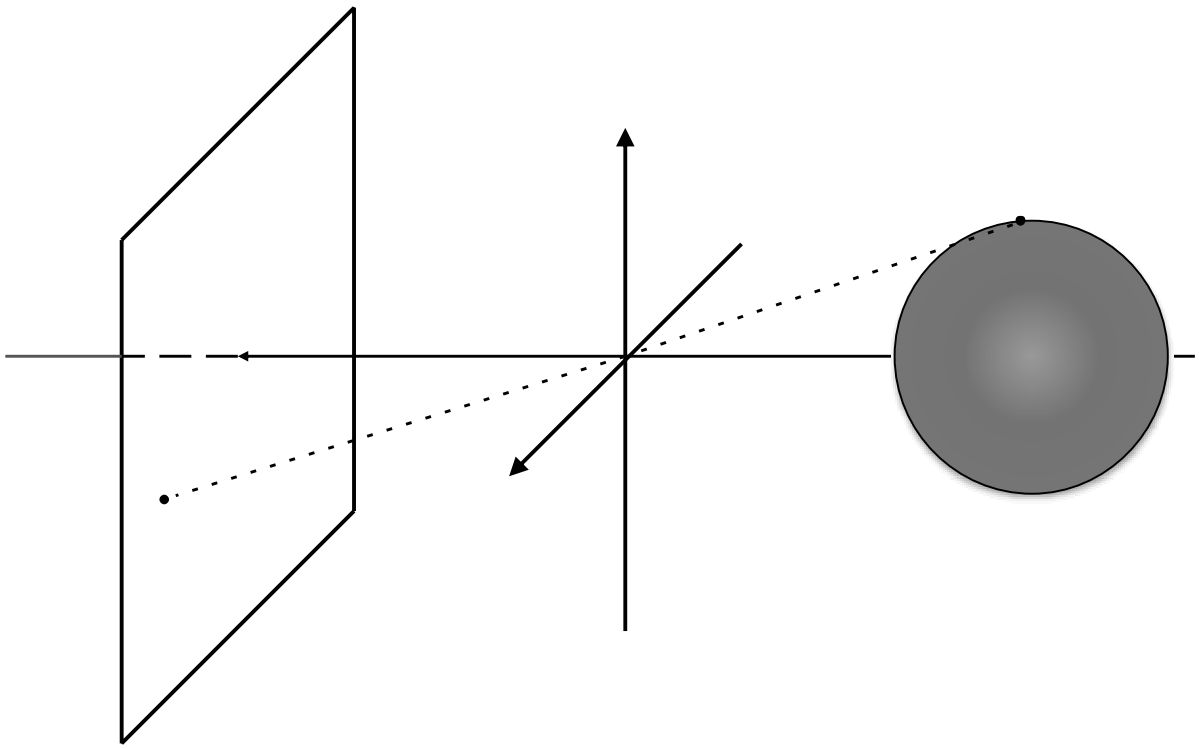


Рис. 1.2. Формирование изображения в центральной проекции

Удобно ввести систему координат, в которой плоскость *xy* на рис 1.2 параллельна плоскости изображения, а начало координат совпадает с точечным отверстием *O*. Ось *z* направлена вдоль оптической оси

перпендикулярно плоскости изображения. Определим положение образа P' некоторой точки P объекта, находящегося перед камерой.

Пусть $\mathbf{r} = (x, y, z)^{\mathrm{T}}$ – вектор, соединяющий O с P, а $\mathbf{r}' = (x', y', z')^{\mathrm{T}}$ – вектор, соединяющий O с P'. Здесь f' – расстояние от плоскости изображения до отверстия, а x' и y' координаты точки P' на этой плоскости. Два вектора $\mathbf{r}$ и $\mathbf{r}'$ коллинеарны и отличаются только отрицательным скалярным множителем. Если луч, соединяющий точки P и P', составляет угол α с оптической осью, то длина вектора $\mathbf{r}$ определяется выражением

$$|\mathbf{r}| = -z \sec\alpha = -(\mathbf{r} \cdot \hat{\mathbf{z}}) \sec\alpha, \tag{1.2.1}$$

где $\hat{\mathbf{z}}$ – единичный вектор вдоль оптической оси. В принятой на схеме системе координат величина z отрицательна для точки, расположенной перед камерой.

Длина вектора $\mathbf{r}'$ составляет $|\mathbf{r}'| = f' \sec\alpha$, следовательно $(1/f')|\mathbf{r}| = -(1/\mathbf{r} \cdot \hat{\mathbf{z}})|\mathbf{r}|$. Это выражение можно представить в виде выражений для компонентов векторов $\mathbf{r}$ и $\mathbf{r}'$:

$$x'/f' = x/z, \tag{1.2.2}$$

$$y'/f' = y/z. \tag{1.2.3}$$

Иногда для упрощения уравнений проекции координаты изображения нормализуют, разделив x' и y' на f'.

Ортогональная проекция

Рассмотрим процесс формирования изображения плоскости, которая параллельна плоскости изображения и задается уравнением $z = z_0$. В этом случае можно определить коэффициент поперечного увеличения m как отношение расстояния между двумя точками на плоскости к расстоянию между соответствующими им точками на изображении.

Рассмотрим малый отрезок $(\delta x, \delta y, 0)^{\mathrm{T}}$ на плоскости и соответствующий ему малый интервал $(\delta x', \delta y', 0)^{\mathrm{T}}$ на изображении. Тогда коэффициент увеличения можно выразить в форме

$$m = \frac{\sqrt{(\delta x')^2 + (\delta y')^2}}{\sqrt{(\delta x)^2 + (\delta y)^2}} = \frac{f'}{-z_0}, \tag{1.2.4}$$

где z_0 – расстояние от плоскости до отверстия. Коэффициент увеличения является одинаковым для всех точек плоскости. (Заметим, что $m < 1$, за исключением изображений, получаемых с помощью микроскопа.)

Изображение небольшого объекта, расположенного на среднем расстоянии $-z_0$, увеличивается в m раз при условии, что изменение z-

координаты точек его видимой поверхности невелико по сравнению с величиной $-z_0$. Площадь изображения объекта пропорциональна m^2. Коэффициент увеличения объектов, расположенных на различных расстояниях от зрительной системы, будет различным.

Пусть глубина сцены – это разброс расстояний от видимых поверхностей до камеры. Значение коэффициента увеличения приблизительно постоянно, когда глубина наблюдаемой сцены мала по сравнению со средним расстоянием от видимых поверхностей до камеры. В этом случае можно упростить уравнение проекции следующим образом:

$$x' = -mx, \quad (1.2.5)$$

$$y' = -my, \quad (1.2.6)$$

где $m = f'/(-z_0)$, а $-z_0$ – среднее значение расстояния от начала координат до изображаемой плоскости. Часто для удобства скалярный множитель m полагают равным 1 или –1. Тогда уравнения проекции принимают еще более простой вид: $x' = x$ и $y' = y$.

Ортогональная (параллельная) проекция может быть представлена лучами, распространяющимися параллельно оптической оси, как это показано на рис 1.3. Различие между центральной и ортогональной проекциями незначительно, если расстояние до объекта во много раз превышает размер объекта.

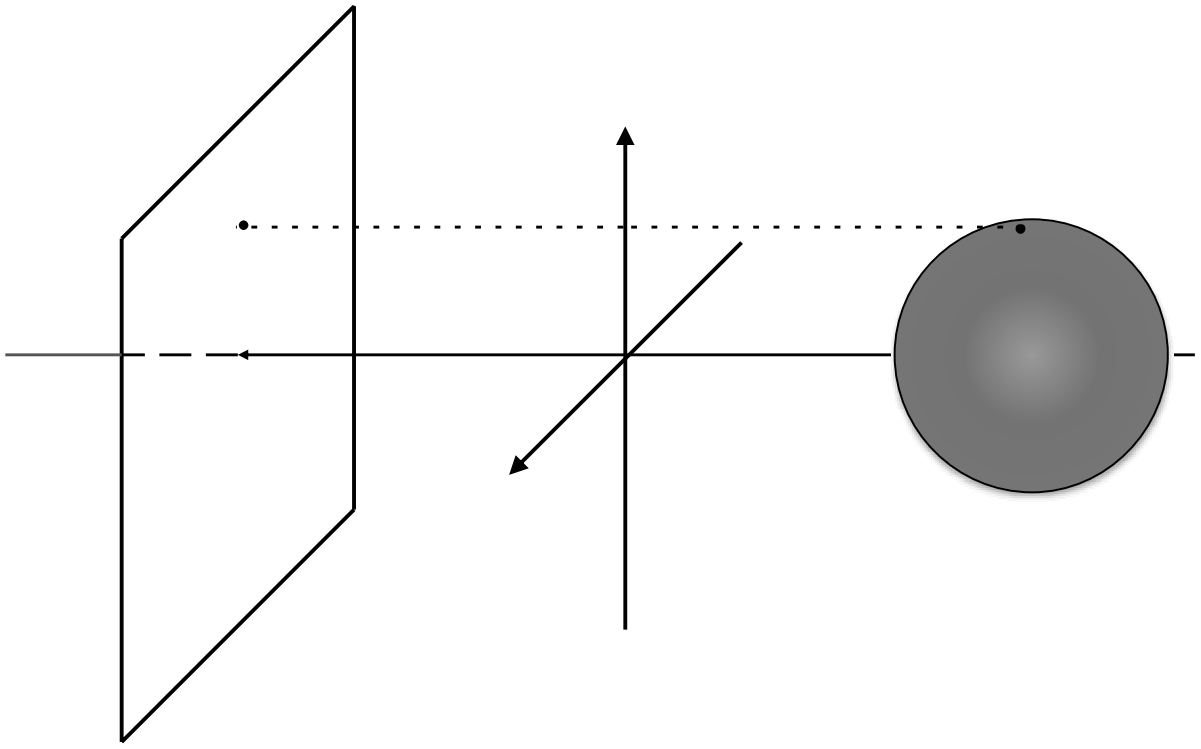


Рис. 1.3. Формирование изображения в ортогональной проекции

Углом поля зрения зрительной системы является угол раствора конуса всех возможных лучей, направленных к объектау. Этот конус имеет те же

самые форму и размер, что и конус, образованный прямыми, соединяющими центр проецирования с границей изображения (экрана).

Альтернативной характеристикой системы является числовая апертура

$$NA = f / d, \qquad (1.2.7)$$

где f – фокусное расстояние, d – диаметр зрачка объектива.

В зависимости от угла поля зрения оптические системы можно разделить на три вида: «нормальные», телескопические (длиннофокусные), широкоугольные (короткофокусные).

«Нормальные» системы характеризуются углом поля зрения примерно 25–40°. К телескопическим относятся системы с фокусным расстоянием, значительно превышающим размеры изображения, и, как следствие этого, с малым углом поля зрения. Широкоугольные системы, напротив, обладают малым фокусным расстоянием, по сравнению с размерами изображения, и поэтому широким полем зрения. Приближенно можно считать, что эффекты перспективы существенны при использовании широкоугольных систем, в то время как изображения, получаемые с помощью телескопических систем, могут быть аппроксимированы ортогональной проекцией.

Влияние характеристик оптической системы формирования изображений

При отображении крупных деталей объекта влияние несовершенства оптической системы не приводит к заметным искажениям на изображении. Однако если размер деталей не превосходит некоторой малой величины, характеристики оптической системы оказывают существенное влияние, как проиллюстрировано на рис. 1.4.

На рис. 1.4, *а* показан идеальный процесс формирования изображения: излучение с интенсивностью I от участка F объекта, попадает на участок изображения F'. Кроме того, на F' попадает свет только от F.

На рис. 1.4, *б* показано изменение контраста изображения вследствие равномерного наложения рассеянного света, исходящего от линзы неидеального качества. При изображении малых деталей объекта свет попадает не только на F', но и на соседние участки, что приводит к тому, что точка уже не отображается в виде точки.

На рис. 1.4, *в* проиллюстрирован пример формирования нерезкого оптического изображения вследствие влияния аберраций линзы или плохой фокусировки. Видно, что излучение F распределяется на площади, большей чем F', или, наоборот, что излучение, падающее на F', исходит с площади, большей, чем F. Поскольку это явление имеет место только в непосредственной близости к F', снижение контраста происходит только для малых деталей. Причем, чем меньше размеры деталей, тем сильнее снижение контраста.

На рис 1.4, *г* видно, как снижение контраста в изображении двух щелей приводит к уменьшению резкости изображения.

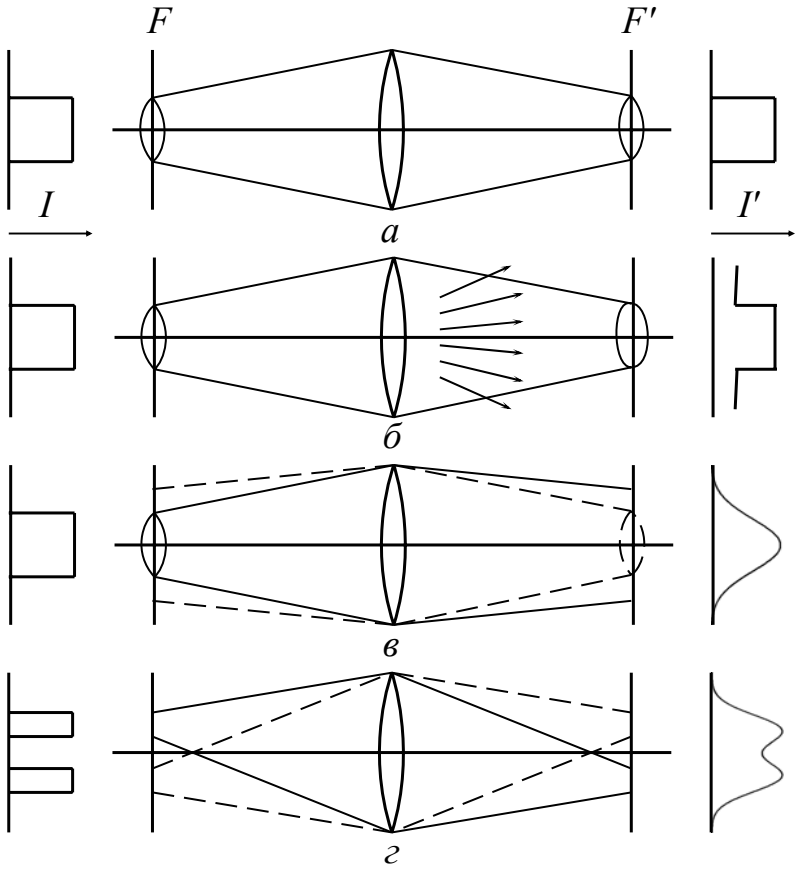


Рис. 1.4. Виды искажений в реальных системах формирования изображений

Вследствие аддитивности интенсивностей при некогерентном освещении, размытое изображение можно представить как сумму отдельных элементов, претерпевших размытие. Иными словами, размытие оптического изображения при некогерентном освещении рассматривается как линейное преобразование.

Влияние шероховатости поверхности объектов

Большинство исследуемых поверхностей в той или иной степени являются шероховатыми. Шероховатость влияет на характеристики рассеянии световых волн, поскольку падающая волна преобразуется в зеркально отраженную и диффузно отраженную составляющие (рис. 1.5). При зеркальном отражении лучи синфазны, угол падения равен углу отражения. В случае шероховатой поверхности возникает разность фаз

$$\delta_i = (4\pi d / \lambda)\cos\alpha_i . \qquad (1.2.8)$$

Если $\delta_i << 2\pi$, то плоскую поверхность можно считать зеркально отражающей. Поэтому, согласно (1.2.8), степень шероховатости следует

рассматривать по отношению к длине волны оптического излучения и в зависимости от направления распространения и рассеяния волны.

Известен критерий Рэлея, согласно которому поверхность считается шероховатой, если разность фаз $\delta_i > \pi/2$. При этом допустимые отклонения профиля поверхности определяются согласно условию

$$d < \lambda/(8\cos\alpha_i). \qquad (1.2.9)$$

При взаимодействии с негладкой поверхностью происходят изменения не только фазовых соотношений, но и мощности излучения в различных направлениях. Зеркальному отражению соответствует так называемая когерентная мощность P_c. Диффузная составляющая соответствует некогерентному полю. При этом регистрируемая мощность P_s в выбранном направлении определяется дифракцией на элементах профиля шероховатой поверхности.

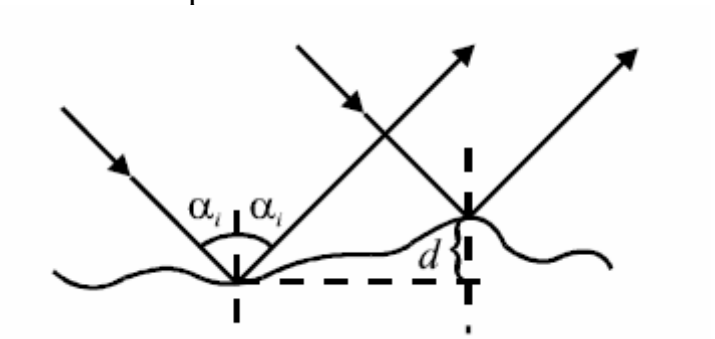


Рис. 1.5. Схема отражения световой волны от шероховатой поверхности

Рассмотрим схему освещения-наблюдения, представленную на рис. 1.6.

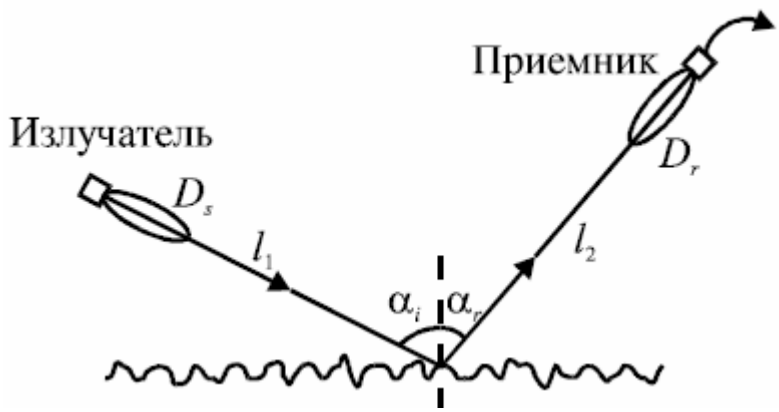


Рис. 1.6. Схема освещения-наблюдения шероховатых поверхностей

Излучатель с диаграммой направленности D_s расположен на расстоянии l_1 от освещаемой точки поверхности. Приемник излучения с диаграммой чувствительности D_r установлен на расстоянии l_2. Можно показать, что доля когерентной мощности, регистрируемой приемником излучения, определяется выражением

$$K_p = P_c / P = (\lambda / 4\pi)^2 D_s D_r |\rho|^2 / (l_1 + l_2)^2, \qquad (1.2.10)$$

где ρ – коэффициент отражения по мощности.

Диффузно рассеиваемая (некогерентная) мощность обычно изменяется в зависимости от угла α_r (рис. 1.6), как это иллюстрируется пунктирной кривой на рис. 1.7.

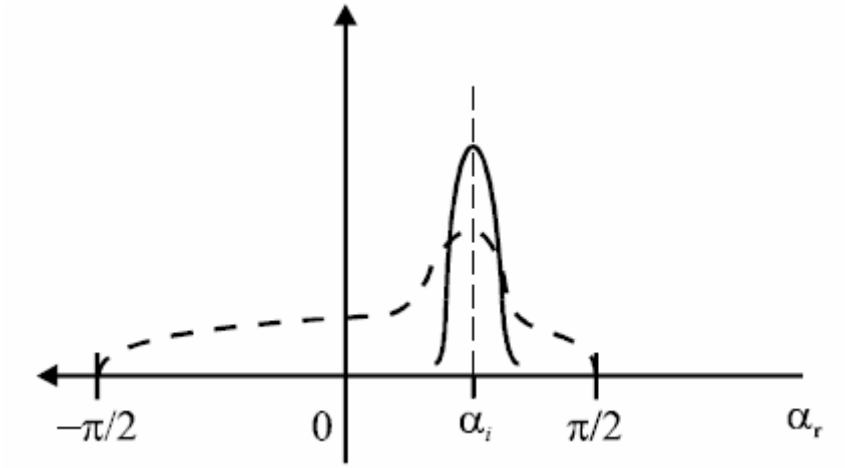


Рис. 1.7. Регистрируемая мощность излучения при различных углах наблюдения

Основной интерес представляет проблема рассеяния на сильно шероховатых поверхностях. При освещении такой поверхности результирующее поле, согласно известному принципу Гюйгенса-Френеля, формируется суммированием полей, рассеянных различными точками поверхности. Регистрируемое поле зависит от статистики отклонений рельефа поверхности, формы макроскопического рельефа (так называемой подстилающей поверхности), поляризационных характеристик излучения, локальных значений коэффициента отражения. Точный расчет рассеянного поля с учетом влияния всех перечисленных факторов представляет собой сложную проблему, решение которой получено только при использовании ряда упрощающих допущений. Вариант приближенного решения основывается на использовании метода Кирхгофа.

Пусть исследуемый объект 1 освещается источником излучения 2, и требуется рассчитать рассеянное поле в различных точках пространства $\mathbf{r}$, в том числе в плоскости наблюдения 3 (см. рис. 1.8). Падающее излучение возбуждает на поверхности и внутри объекта вторичные источники, распределение которых $\mathbf{q}(\mathbf{r})$ зависит от характеристик освещения и физико-геометрических свойств объекта и является известным. Электрический вектор рассеянного объектом поля $\mathbf{E}(\mathbf{r})$, возбужденного вторичными источниками, удовлетворяет уравнению

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{E}(\mathbf{r}) = \mathbf{q}(\mathbf{r}), \qquad (1.2.11)$$

где c – скорость распространения волны.

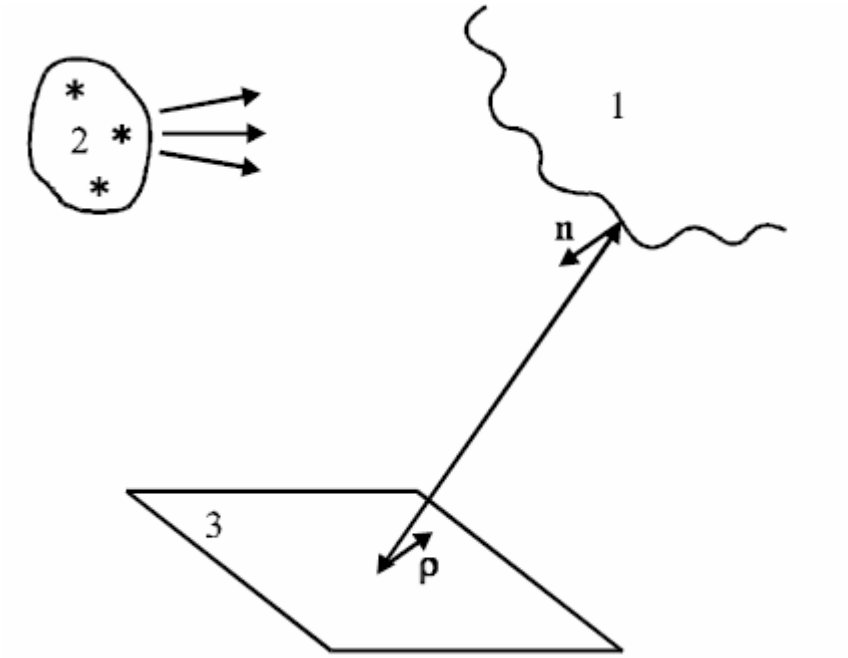


Рис. 1.8. Иллюстрация к расчету характеристик поля рассеяния

Согласно принципу Гюйгенса-Френеля, комплексная амплитуда A в точке P_0 на расстоянии $r >> \lambda$ от объекта (рис. 1.9) является суперпозицией элементарных сферических волн, а именно:

$$A(P_0) = \frac{1}{j\lambda}\iint_S A(P_1)\frac{\exp(jkr)}{r}\kappa(\alpha_r)\mathrm{d}s, \qquad (1.2.12)$$

где P_1 – точка поверхности S объекта, $k = 2\pi/\lambda$, r – расстояние между точками P_0 и P_1, α_r – угол между нормалью к поверхности и отрезком P_0P_1, $\kappa(\alpha_r)$ – так называемый коэффициент наклона или амплитудный коэффициент направленности, $0 \le \kappa(\alpha_r) \le 1$.

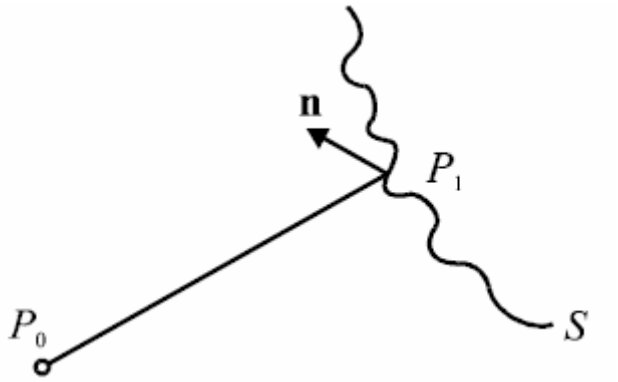


Рис. 1.9. Геометрическое расположение точек освещения-наблюдения

Операция интегрирования по объему, необходимая для решения уравнения (1.2.11), преобразована в (1.2.12) в операцию интегрирования по поверхности с использованием теоремы Грина. В приближении Кирхгофа в качестве функции Грина выбрана функция $(1/r)\exp(jkr)$, а производные поля по нормали к поверхности учитываются амплитудным коэффициентом $\kappa(\alpha_r)$ в предположении, что радиус кривизны в локальных точках профиля поверхности существенно превышает длину волны λ.

Формула Френеля-Кирхгофа (1.2.12) устанавливает связь между освещающим полем на поверхности объекта и полем, рассеянным объектом в окружающем пространстве, и, следовательно, представляет собой решение прямой задачи рассеяния. Эта формула непосредственно обобщается на случаи источников с различными спектральными характеристиками, а также применительно к динамическим объектам.

1.3 Формирование когерентных оптических полей

Рассмотрим формирование когерентных изображений на примере двух когерентных колебаний с амплитудами a_1, a_2 и начальными фазами φ_1, φ_2. Регистрируемый сигнал определяется выражением

$$\begin{aligned} s &= \mu\left|a_1\exp(j\varphi_1) + a_2\exp(j\varphi_2)\right|^2 = \\ &= \mu\left[(a_1\cos\varphi_1 + a_2\cos\varphi_2)^2 + (a_1\sin\varphi_1 + a_2\sin\varphi_2)^2\right], \end{aligned} \tag{1.3.1}$$

где μ – коэффициент преобразования. Результирующая фаза может быть определена в виде

$$\mathrm{tg}\varphi = \frac{a_1\sin\varphi_1 + a_2\sin\varphi_2}{a_1\cos\varphi_1 + a_2\cos\varphi_2}. \tag{1.3.2}$$

При фиксированных значениях a_1, a_2 сигнал (2.6) зависит от разности фаз $\delta = \varphi_2 - \varphi_1$ интерферирующих волн, имеющих оптическую разность хода Δ. При этом

$$\delta = 2\pi\Delta/\lambda = 2\pi N_I = 2\pi C_I + \varepsilon, \tag{1.3.3}$$

где λ – длина волны, N_I – порядок интерференции, $C_I = 0,\pm1,...$ – целый порядок интерференции, $\varepsilon = 2\pi\varepsilon_I$, $0 \le \varepsilon_I < 1$ – дробная часть порядка интерференции.

Из выражения (1.3.1) с учетом (1.3.3) и свойства периодичности тригонометрических функций получим

$$s = s_0 + s_m\cos\varepsilon, \tag{1.3.4}$$

где $s_0 = \mu(a_1^2 + a_2^2)$ – фоновая составляющая, $s_m = 2\mu a_1 a_2$ – амплитуда информационной составляющей.

В интерферометрической системе амплитуда и фаза предметной волны могут изменяться в зависимости от пространственных координат и времени $a_1 = a_1(x, y, t)$, $\varphi_1 = \varphi_1(x, y, t)$. При этом выражение (1.3.4) следует записывать в виде трехмерной функции

$$s(x, y, t) = s_0(x, y, t) + s_m(x, y, t)\cos\Phi(x, y, t), \qquad (1.3.5)$$

где $\Phi(x, y, t) = \varepsilon + \psi(x, y, t)$. Зависимость отдельных параметров в (1.3.5) от координат и времени, определяются изменениями амплитуды и фазы предметной волны с учетом нелинейного преобразования (1.3.1).

При фиксированной ненулевой разности хода Δ и изменении длины волны λ имеет место изменение разности фаз δ согласно (1.3.3). Порядок интерференции в таком случае зависит от длины волны: $N_I = N_I(\lambda)$. Поскольку разность фаз не влияет на значение амплитуд колебаний, выражение (1.3.5) остается справедливым и для рассматриваемого случая, если принять $\lambda = \lambda(t)$.

Модель (1.3.5) носит достаточно общий характер, и для ее практического использования необходимо принимать дополнительные допущения, определяемые видом решаемых задач. Например, если исследуемый объект является неподвижным и освещается источником излучения с фиксированной длиной волны, то в (1.3.5) можно опустить зависимость сигнала от времени. В случае плавных изменений амплитуды и фазы предметной волны фоновая составляющая $s_0(x, y)$ и амплитуда информационной составляющей $s_m(x, y)$ рассматриваются как функции, изменяющиеся медленно по сравнению с функцией $\cos\Phi(x, y)$, где $\Phi(x, y) = \varepsilon + 2\pi(u_0 x + v_0 x) + \varphi(x, y)$ – функция изменения фазы, включающая в качестве параметров пространственные частоты интерференционных полос (u_0, v_0) в направлениях (x, y) и плавные отклонения фазы $\varphi(x, y)$.

В случае шероховатой поверхности изменения амплитуды и фазы отраженной волны носят стохастический характер, и конкретизация модели (1.3.4) возможна при известных статистических характеристиках интерференционного поля, получаемых с учетом (1.3.1).

При освещении шероховатой поверхности монохроматическим излучением комплексная амплитуда в точке наблюдения P_0 с координатами (x, y, z) может быть представлена в форме суммы амплитуд $a_n(x, y, z)$ элементарных колебаний, полученных при отражении от различных точек поверхности

$$A(x, y, z) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} a_n(x, y, z) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} |a_n| \exp(j\varphi_n). \qquad (1.3.6)$$

Предположим, что значения амплитуд a_n и фаз φ_n статистически независимы, причем для сильно шероховатых поверхностей значения фазы

φ_n распределены равномерно на интервале $(-\pi, \pi)$. При этом средние значения для результирующей амплитуды поля по ансамблю реализаций профиля шероховатой поверхности будут равны

$$\langle \mathrm{Re}\, A \rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \langle |a_n| \cos\varphi_n \rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \langle |a_n| \rangle \langle \cos\varphi_n \rangle = 0, \tag{1.3.7}$$

$$\langle \mathrm{Im}\, A \rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \langle |a_n| \sin\varphi_n \rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \langle |a_n| \rangle \langle \sin\varphi_n \rangle = 0. \tag{1.3.8}$$

Дисперсии действительной и мнимой частей поля определяются как

$$\langle (\mathrm{Re}\, A)^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{N} \langle |a_n| |a_l| \rangle \langle \cos\varphi_n \cos\varphi_l \rangle = \frac{1}{N} \sum_{n=1}^{N} \langle |a_n^2| \rangle / 2, \tag{1.3.9}$$

$$\langle (\mathrm{Im}\, A)^2 \rangle = \frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{N} \langle |a_n| |a_l| \rangle \langle \sin\varphi_n \sin\varphi_l \rangle = \frac{1}{N} \sum_{n=1}^{N} \langle |a_n^2| \rangle / 2. \tag{1.3.10}$$

Действительная и мнимая части амплитуды поля не коррелированны, то есть $\langle \mathrm{Re}\, A \, \mathrm{Im}\, A \rangle = 0$. Поскольку значения фаз распределены равномерно и являются взаимно независимыми, то справедливы соотношения

$$\langle \cos\varphi_n \cos\varphi_l \rangle = \langle \sin\varphi_n \sin\varphi_l \rangle = \begin{cases} 0{,}5 & \text{при} \quad n = l, \\ 0 & \text{при} \quad n \neq l. \end{cases} \tag{1.3.11}$$

Полагая для большого числа отражающих точек поверхности $N \to \infty$, получим, что совместная плотность вероятностей действительной и мнимой частей амплитуды поля является асимптотически гауссовой с круговой симметрией

$$p_A(\mathrm{Re}\, A, \mathrm{Im}\, A) = \left(\frac{1}{2\pi\sigma^2} \right) \exp\left(- \frac{(\mathrm{Re}\, A)^2 + (\mathrm{Im}\, A)^2}{2\sigma^2} \right), \tag{1.3.12}$$

где

$$\sigma^2 = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \langle |a_n|^2 \rangle / 2. \tag{1.3.13}$$

Интенсивность поля I и результирующая фаза определяются соответственно соотношениями

$$I = (\mathrm{Re}\, A)^2 + (\mathrm{Im}\, A)^2, \tag{1.3.14}$$

$$\varphi = \mathrm{arctg}(\mathrm{Im}\, A / \mathrm{Re}\, A). \tag{1.3.15}$$

Совместная плотность вероятности интенсивности и фазы определяется из выражения (1.3.12) в форме

$$p_{I,\varphi}(I, \varphi) = p_A(\sqrt{I} \cos\varphi, \sqrt{I} \sin\varphi) |\det J|, \tag{1.3.16}$$

где $|\det J|$ – модуль определителя якобиана преобразования,

$$|\det J| = \begin{vmatrix} \dfrac{\partial \mathrm{Re}\, A}{\partial I} & \dfrac{\partial \mathrm{Re}\, A}{\partial \varphi} \\ \dfrac{\partial \mathrm{Im}\, A}{\partial I} & \dfrac{\partial \mathrm{Im}\, A}{\partial \varphi} \end{vmatrix} = 0{,}5. \tag{1.3.17}$$

Используя выражение (1.3.16) найдем:

$$p_{I,\varphi}(I,\varphi)=\left(\frac{1}{4\sigma^2}\right)\exp\left(-\frac{I}{2\sigma^2}\right), \tag{1.3.18}$$

где интенсивность поля $I \geq 0$, а фаза $-\pi \leq \varphi < \pi$.

Интегрирование (1.3.18) по фазе и интенсивности позволяет найти маргинальные плотности вероятности соответственно значений интенсивности и фазы

$$p_I(I)=\int_{-\pi}^{\pi} p_{I,\varphi}(I,\varphi)d\varphi=\left(\frac{1}{2\sigma^2}\right)\exp\left(-\frac{I}{2\sigma^2}\right),\ I \geq 0, \tag{1.3.19}$$

$$p_\varphi(\varphi)=\int_{0}^{\infty} p_{I,\varphi}(I,\varphi)dI=\frac{1}{2\pi},\ -\pi \leq \varphi < \pi. \tag{1.3.20}$$

Видно, что интенсивность рассеянного поля имеет плотность вероятности в форме отрицательного экспоненциального распределения (рис. 1.10), а фаза – равномерного.

Используя (1.3.19), можно показать, что $\left\langle I^2 \right\rangle = 2\left\langle I \right\rangle^2$, и дисперсия интенсивности равна

$$\sigma^2 = \left\langle I^2 \right\rangle - \left\langle I \right\rangle^2 = \left\langle I \right\rangle^2. \tag{1.3.21}$$

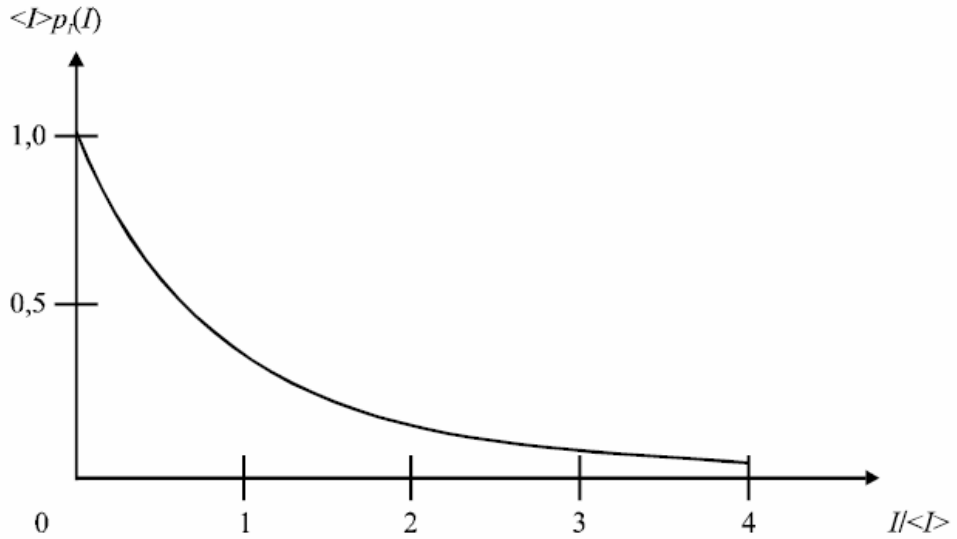


Рис. 1.10. Нормализованная функция плотности вероятности значений интенсивности рассеянного поля

Формирование оптических полей в интерферометрии и голографии

В интерферометрии и голографии используется когерентная подсветка, обеспечиваемая опорной волной, интенсивность которой I_r обычно известна. Тогда без потери общности можно принять значение фазы опорной волны φ_r равной нулю. При этих допущениях и

использовании соотношения (1.3.18) получим выражение для совместной плотности вероятности значений действительной и мнимой частей амплитуды результирующего поля

$$p_A(\mathrm{Re}\,A, \mathrm{Im}\,A) = \left(\frac{1}{2\pi\sigma^2}\right)\exp\left(-\frac{(\mathrm{Re}\,A - \sqrt{I_r})^2 + (\mathrm{Im}\,A)^2}{2\sigma^2}\right). \qquad (1.3.22)$$

Используя соотношения (2.13), (2.14) и (2.17), найдем выражение для совместной плотности вероятности интенсивности и фазы интерференционной картины в виде

$$p_{I,\varphi}(I,\varphi) = \left(\frac{1}{4\sigma^2}\right)\exp\left(-\frac{I + I_r - 2\sqrt{II_r}\cos\varphi}{2\sigma^2}\right), \qquad (1.3.23)$$

где $I,\ I_r \geq 0$, $-\pi \leq \varphi < \pi$.

Маргинальная плотность вероятности значений интенсивности интерференционной картины определяется из (1.3.23) как

$$p_I(I) = \left(\frac{1}{4\sigma^2}\right)\exp\left(-\frac{I + I_r}{2\sigma^2}\right)I_0\left(\frac{\sqrt{II_r}}{\sigma^2}\right), \qquad (1.3.24)$$

где $I_0(.)$ – модифицированная функция Бесселя нулевого порядка, причем при выводе (2.20) использовано известное соотношение

$$\int_{-\pi}^{\pi}\exp\left(-\frac{\sqrt{II_r}\cos\varphi}{\sigma^2}\right)d\varphi = 2\pi I_0\left(\frac{\sqrt{II_r}}{\sigma^2}\right). \qquad (1.3.25)$$

Следует отметить, что в (1.3.24) значение $2\sigma^2$ представляет интенсивность рассеянного поля без опорной волны.

Плотность вероятности (1.3.24) носит название плотности модифицированного распределения Райса. Эта функция показана на рис. 1.11 для различных значений I_r/σ^2.

Маргинальная плотность вероятности значений фазы может быть выражена как

$$p_\varphi(\varphi) = \left(\frac{1}{2\pi}\right)\exp\left(-\frac{I_r}{2\sigma^2}\right) + \sqrt{\frac{I_r}{2\sigma^2}}\cos\varphi \times$$
$$\times \exp\left(-\frac{I_r}{2\sigma^2}\sin^2\varphi\right)\Phi\left(\sqrt{\frac{I_r}{\sigma^2}}\cos\varphi\right), \qquad (1.3.26)$$

где принято обозначение

$$\Phi(b) = \frac{1}{2\pi}\int_{-\infty}^{b}\exp\left(-\frac{y^2}{2}\right)dy. \qquad (1.3.27)$$

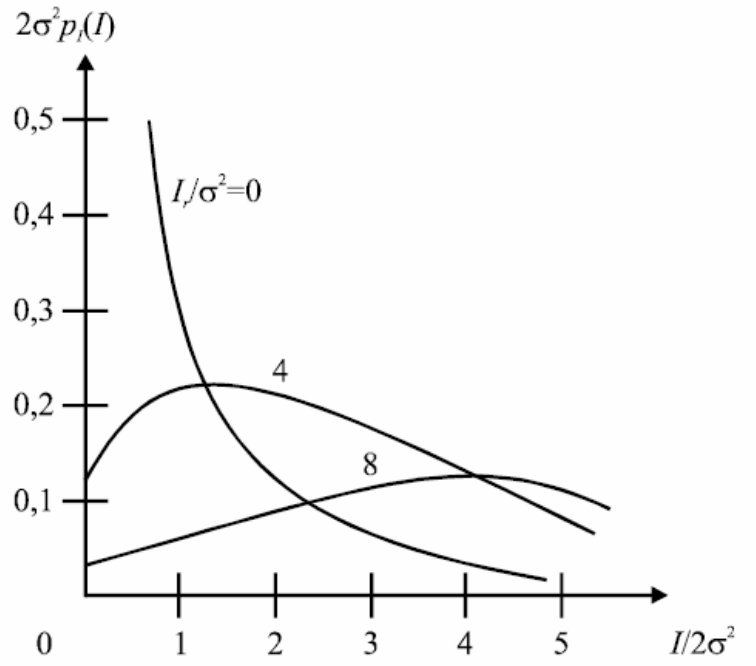


Рис. 1.11. Плотность вероятности значений интенсивности света в интерференционной картине

Плотность вероятности (1.3.26) имеет вид колоколообразной кривой (рис. 1.12), центр которой совпадает со значением фазы опорной волны.

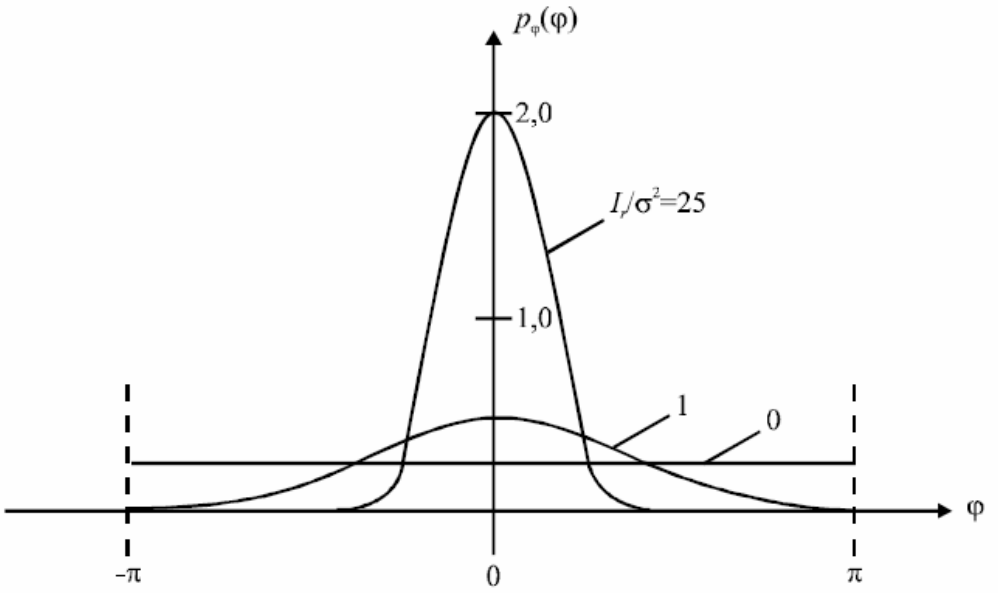


Рис. 1.12. Плотность вероятности значений интенсивности света в интерференционной картине

Из этого следует, что при известной интенсивности I_r опорной волны по виду кривой плотности вероятности (1.3.24) и по значению ширины кривой (1.3.26) можно судить о значении σ^2, связанной с

характеристиками рассеяния когерентного излучения на шероховатой поверхности.

1.4 Сравнение процессов формирования оптических полей в когерентной и некогерентной системах

Основное отличие процессов формирования изображений в некогерентной и когерентной системах состоит в том, что интенсивность излучения на выходе некогерентной системы определяется в форме свертки интенсивности (квадрата комплексной амплитуды волны) с квадратом модуля функции рассеяния точки $h(x,y)$, а именно

$$I = \left|h(x,y)\right|^2 * \left|A(x,y)\right|^2, \qquad (1.4.1)$$

тогда как для когерентной системы вначале имеет место свертка комплексной амплитуды волны с функцией рассеяния точки

$$I = \left|h(x,y) * A(x,y)\right|^2. \qquad (1.4.2)$$

При рассмотрении соотношений (1.4.1) и (1.4.2) в частотной области с учетом теоремы о свертке можно получить выражения, характеризующие преобразование некогерентных и когерентных полей соответственно

$$I_{u,v} = (H * H^*)(S * S^*), \qquad (1.4.3.)$$

$$I_{u,v} = (HS) * (HS)^*, \qquad (1.4.4)$$

где $H(u,v) = \mathrm{F}\{h(x,y)\}$, $S(u,v) = \mathrm{F}\{A(x,y)\}$ – пространственно-частотные спектры функций $h(x,y)$ и $A(x,y)$, а $\mathrm{F}\{.\}$ обозначает операцию преобразования Фурье.

Рассмотрим одномерный случай и примем, что функция $h(x)$ является действительной функцией симметричной формы с ограниченным по протяженности спектром $H(u)$ в интервале частот $[-u_m, u_m]$.

Пусть два объекта имеют одинаковый коэффициент пропускания по интенсивности

$$g(x) = \cos^2 2\pi u_0 x, \qquad (1.4.5)$$

и различные изменения фазы при амплитудном пропускании

$$t(x) = \cos 2\pi u_0 x, \qquad (1.4.6)$$

$$t(x) = \left|\cos 2\pi u_0 x\right|, \qquad (1.4.7)$$

где $u_m / 2 < u_0 < u_m$, u_m – граничная пространственная частота при когерентном освещении. В (1.4.7) фаза сдвинута на π при отрицательных значениях исходной косинусоиды. Обе функции (1.4.6) и (1.4.7) являются действительными и четными, поэтому в данных примерах можно опустить знаки комплексного сопряжения в (1.4.3) и (1.4.4).

На рис. 1.13 показаны преобразования (1.4.3) и (1.4.4) при формировании спектра интенсивности первого объекта (1.4.6). Контраст распределения интенсивности изображения для некогерентного излучения хуже, чем для когерентного. Таким образом, для исследования данного объекта предпочтительно когерентное освещение.

Для второго объекта (1.4.7) амплитуда представляет собой периодическую функцию с частотой основной гармоники $2u_0$. Поскольку $2u_0 > u_m$, в случае когерентного освещения интенсивность изображения меняться не будет, тогда как некогерентная система формирует то же самое изображение, как и для первого объекта (1.4.6). В этом случае предпочтительнее некогерентное освещение.

Из сказанного выше следует, что невозможно однозначно заранее определить, какой тип освещения предпочтительнее использовать. Результат в значительной мере зависит от структуры объекта и, в частности, от распределения фазы в точках объекта.

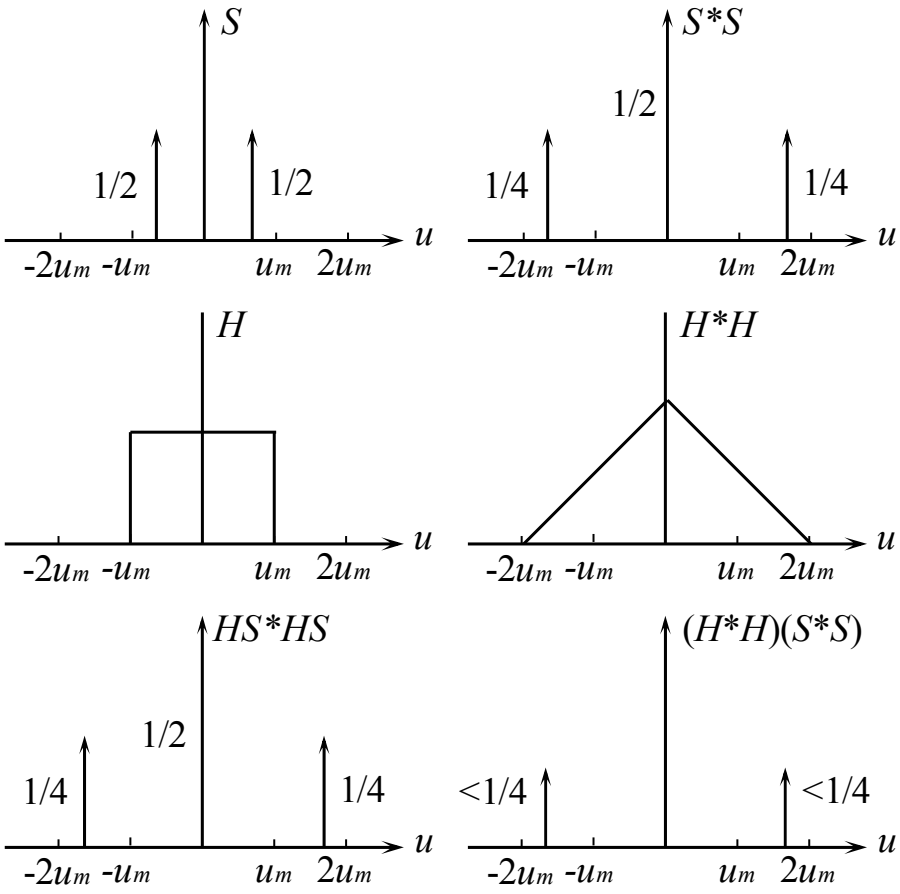


Рис. 1.13. Оценка спектра интенсивности косинусоидальной амплитудной решетки

Разрешение двух точек

Согласно известному критерию Рэлея, два некогерентных источника «разрешаются» дифракционно ограниченной системой, если центр диска Эйри, созданного одним источником, совпадает с нулем первого порядка дифракционной картины, созданной вторым источником. Минимальное разрешаемое расстояние между двумя точками для некоторой оптической системы может быть представлено в виде

$$\delta = 1{,}22\frac{\lambda d_i}{D}, \qquad (1.4.8)$$

где D – диаметр входного зрачка. На рис. 2.17 представлено распределение интенсивности для этого минимального расстояния. Провал в центре составляет величину порядка 20% от максимальной интенсивности.

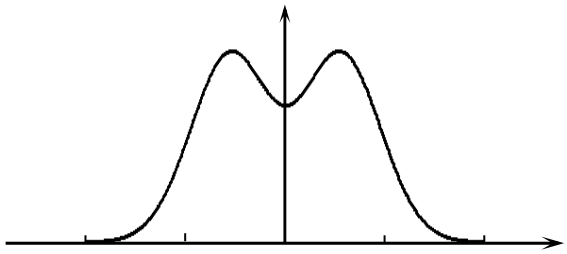


Рис. 1.14. Распределение интенсивности в изображении двух некогерентных точечных источников, находящихся на минимальном разрешаемом расстоянии

Какой тип излучения, когерентное или некогерентное, способствует лучшему различению двух точечных источников, разделенных расстоянием Рэлея? Ответ на этот вопрос зависит от распределения фазы, связанного с предметом. Распределение интенсивности в изображении можно записать в нормированных координатах

$$I(x) = \left| 2\frac{J_1[\pi(x-0{,}61)]}{\pi(x-0{,}61)} + e^{j\varphi} 2\frac{J_1[\pi(x+0{,}61)]}{\pi(x+0{,}61)} \right|^2, \qquad (1.4.9)$$

где φ – относительная разность фаз двух точечных источников, J_1 – функция Бесселя первого порядка.

На рис. 1.15 приведены распределения интенсивности изображения для точечных источников, излучающих в фазе ($\varphi = 0$), в противофазе ($\varphi = \pi$) и с разностью фаз в четверть периода ($\varphi = \pi/2$). Распределение интенсивности в изображении для источников, сдвинутых по фазе на $\pi/2$, тождественно распределению для некогерентных источников

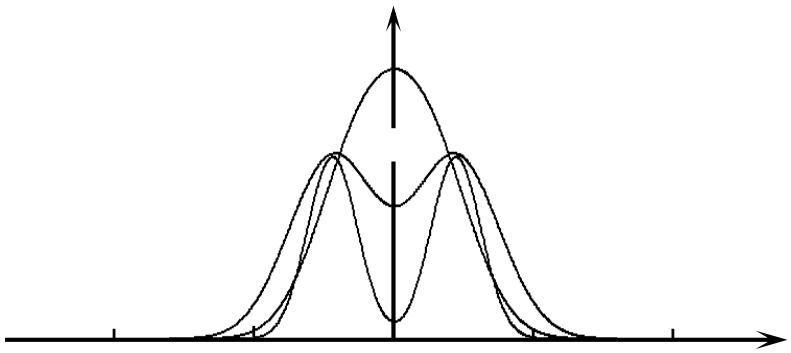
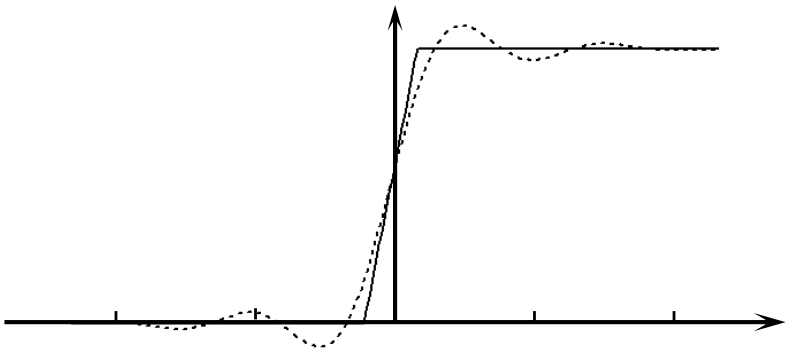


Рис. 1.15. Распределение интенсивности в изображении двух взаимно когерентных точечных источников, находящихся на минимальном расстоянии

. Когда разность фаз $\varphi = 0$, провал на кривой интенсивности изображения отсутствует и, следовательно, две точки труднее различить, чем в случае некогерентного освещения. Однако, если точечные источники излучают в противофазе, то провал составляет больше 20% и при когерентном освещении две точки различимы лучше, чем при некогерентном. Можно заключить, что некогерентное освещение следует использовать при разности фаз $0 \le \varphi < \pi/2$, а когерентное при $\pi/2 < \varphi \le \pi$.

Другие эффекты

Отклики когерентных и некогерентных систем на возмущение, распространяющееся от резко очерченного края, существенно отличаются друг от друга. На рис. 1.16 представлены теоретические отклики системы с квадратным входным зрачком на функцию Хэвисайда (резкий скачок в пространстве предмета).



-2

Рис. 1.16. Теоретические кривые интенсивности откликов системы на функцию Хэвисайда при когерентном (пунктир) и некогерентном (сплошная) освещении

Передаточная функция когерентной системы имеет резкие переходы, тогда как изменение этой функции в некогерентной системе происходят плавно. Из графиков на рис. 1.16 видно, что кажущиеся положения края (т.е. абсциссы точек, в которых интенсивности отклика равны половине максимальной) различны для двух видов освещения, причем когерентное освещение вносит незначительную погрешность.

При использовании освещения с высокой степенью когерентности значимым становится эффект спеклов, так как размер отдельных спеклов может быть равен размеру разрешаемого элемента изображения, что негативно сказывается при исследовании объектов малых размеров.

Освещение с высокой степенью когерентности особенно чувствительно к оптическим дефектам, которые могут встретиться на пути распространения света. Например, мельчайшие частички пыли на линзе могут привести к появлению ярко выраженных дифракционных колец, которые будут накладываться на изображения. Влияние подобных эффектов можно значительно уменьшить, если при освещении использовать подвижный рассеиватель, подобный матовому стеклу.

1.5 Принципы преобразования оптических полей

В настоящем разделе кратко рассмотрены основные принципы модуляции фазы и сдвига оптической частоты. Известно, что модуляция сигналов позволяет значительно повысить помехозащищенность системы, что особенно важно при создании высокочувствительных систем.

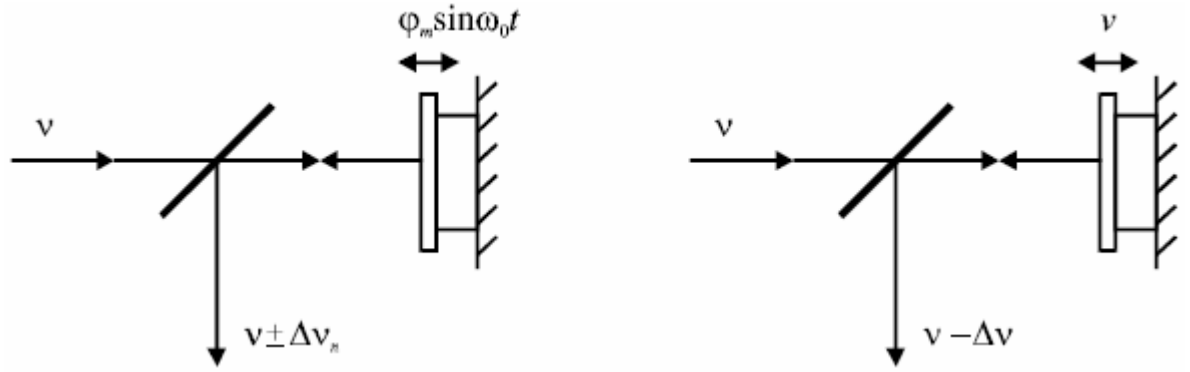
Методы, используемые для осуществления изменения фазы оптических волн, подразделяются по принципу действия на следующие основные группы: электромеханические, электрооптические и акустооптические. При этом чаще всего используются закон отражения света, поляризационные свойства света и эффекты дифракции на движущейся решетке.

На рис. 1.17 иллюстрируется принцип формирования световой волны с фазовой модуляцией (а) и сдвигом частоты (б) при использовании отражения света и перемещении отражателя.

При модуляции фазы напряженность поля E световой волны можно представить в форме

$$E(t) = \frac{1}{2}\{A\exp[-j(2\pi\upsilon t + \varphi_m \sin\omega_0 t)] + A^* \exp[j(2\pi\upsilon t + \varphi_m \sin\omega_0 t)]\}, \quad (1.5.1)$$

где ω_0 – частота модуляции, φ_m – индекс модуляции.



а *б*

Рис 1.17. Принцип формирования световой волны с фазовой модуляцией (а) и сдвигом частоты (б) при использовании отражения света и перемещении отражателя

Учитывая, что функцию вида $\exp(\varphi_m \sin\omega_0 t)$ можно разложить в ряд, а именно

$$\exp(\varphi_m \sin\omega_0 t) = I_0(\varphi_m) + 2\sum_{k=0}^{\infty}(-1)^k I_{2k+1}(\varphi_m)\sin[(2k+1)\omega_0 t] +$$

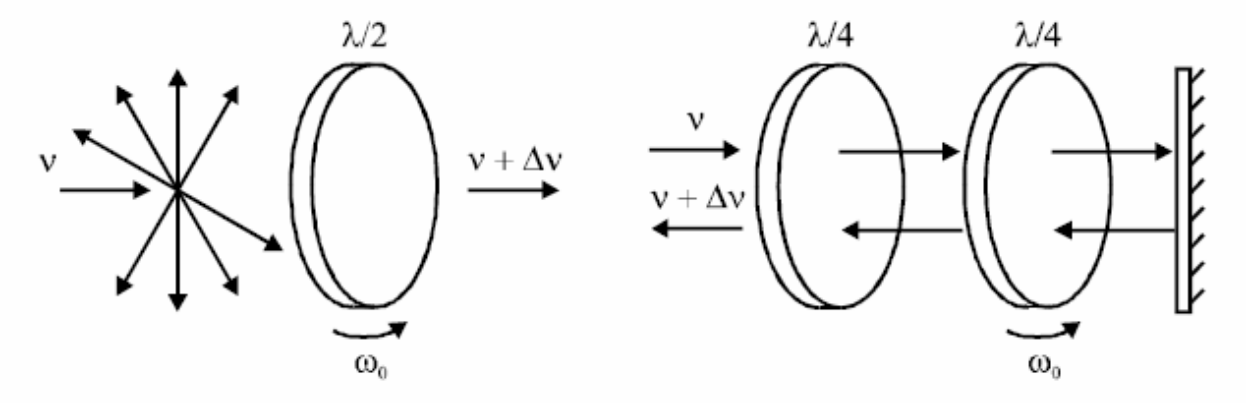
$$+ 2\sum_{k=1}^{\infty}(-1)^k I_{2k}(\varphi_m)\cos[2k\omega_0 t], \qquad (1.5.2)$$

где $I_i(\varphi_m)$ – модифицированные функции Бесселя порядка i, можно получить в результате волну с преобразованным спектром колебаний, составляющие которого сдвинуты относительно исходной частоты υ на частотные интервалы $\Delta\upsilon_n = n\upsilon_0$, где $n = 1, 2, \ldots$, $\upsilon_0 = \omega_0 / 2\pi$. Амплитуды отдельных составляющих спектра зависят от индекса модуляции φ_m. В модуляционных схемах интерферометров обычно используют первые две гармоники в (1.5.2), причем обеспечивают равенство амплитуд этих гармоник, задавая $\varphi_m = 0{,}84\pi$. Известно, что при индексах модуляции $\varphi_m < 0{,}5\pi$ модуляция является узкополосной. Поэтому при выбранном режиме фазовой модуляции энергия спектра сосредоточена главным образом именно в первой и второй гармонических составляющих.

Сдвиг частоты исходной волны можно получить при равномерном перемещении опорного отражателя со скоростью v (рис. 1.16, *б*). При этом вследствие эффекта Доплера частота отраженной волны сдвигается на величину $\Delta\upsilon = 2v/\lambda$.

Модуляционные схемы с использованием поляризационных свойств света чаще всего строятся на основе вращающихся полуволновых или четвертьволновых пластинок.

На рис. 1.18 показаны принципы формирования световых колебаний со сдвигом частоты.



а *б*

Рис. 1.18. Методы сдвига частоты световых колебаний

Свет, поляризованный по кругу (рис. 1.18, *а*), проходит через вращающуюся с частотой ω_0 полуволновую пластинку. При этом частоты проходящего света изменяется на величину сдвига $\Delta\upsilon = \omega_0 / \pi$. Математически такое преобразование удобно характеризовать при помощи поляризационных матриц. Излучение, поляризованное по кругу, определяется поляризационной матрицей

$$P_1 = \frac{A}{\sqrt{2}} \exp(j 2\pi\upsilon t) \begin{pmatrix} 1 \\ -j \end{pmatrix}, \qquad (1.5.3)$$

где A – амплитуда электрического поля исходной световой волны.

Поляризационная матрица излучения на выходе вращающейся с частотой ω_0 фазовой пластинки, вносящей сдвиг фазы волны, равный ψ, имеет вид

$$P_2 = \frac{A}{\sqrt{2}} \begin{pmatrix} \exp(j 2\pi\upsilon t \cos[\psi/2]) + j \exp(j 2[\pi\upsilon - \omega_0] t \sin[\psi/2]) \\ -j \exp(j 2\pi\upsilon t \cos[\psi/2]) - \exp(j 2[\pi\upsilon - \omega_0] t \sin[\psi/2]) \end{pmatrix}. \qquad (1.5.4)$$

Для полуволновой пластинки $\psi = \pi$, и матрица (1.33) преобразуется к форме

$$P_2 = \frac{A}{\sqrt{2}} \exp(j 2\pi\upsilon t) \begin{pmatrix} j \exp(-j 2\omega_0 t) \\ -\exp(-j 2\omega_0 t) \end{pmatrix}. \qquad (1.5.5)$$

Сравнение (1.5.3) и (1.5.5) показывает, что на выходе полуволновой пластинки излучение также поляризовано по кругу, а частота изменяется на величину $\Delta\upsilon = \omega_0 / \pi$

Другой способ получения сдвига частоты (рис. 1.18, *б*) состоит в том, что используют неподвижную и вращающуюся с частотой ω_0 последовательно установленные четвертьволновые пластинки. При двукратном прохождении вращающейся четвертьволновой пластинки возникает сдвиг частоты как и при использовании полуволновой пластинки.

Сдвиг частоты на основе эффектов дифракции достигается в движущейся решетке (рис. 1.19). Если пространственная частота решетки равна u_0, то сдвиг частоты $\Delta\upsilon$ в первом порядке дифракции по отношению к нулевому порядку (исходная частота υ) будет равен $\Delta\upsilon = u_0 v$, где v – скорость движения решетки. В самом деле, напряженность поля исходной световой волны представим с точностью до сопряженной компоненты в виде

$$E = A\exp j2\pi\upsilon t, \qquad (1.5.6)$$

тогда напряженность поля на выходе движущейся решетки в первом порядке дифракции составит

$$E = A\exp j2\pi(\upsilon - u_0 v)t, \qquad (1.5.7)$$

где $2\pi u_0 vt$ – фазовый сдвиг, соответствующий изменяющейся оптической разности хода Δ лучей, исходящих от соседних щелей решетки, в первом порядке дифракции.

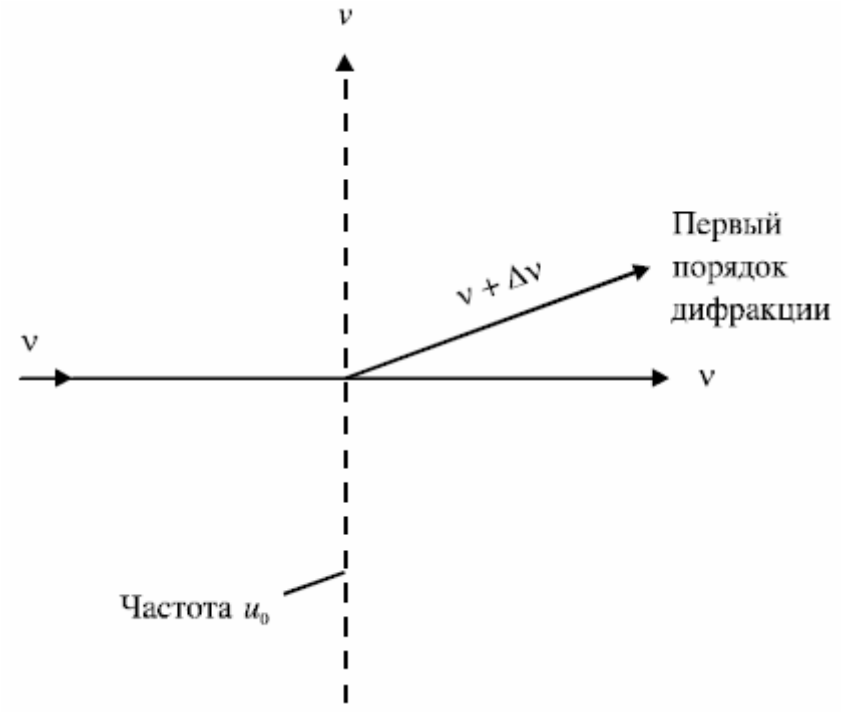


Рис. 1.19. Сдвиг частоты на основе эффектов дифракции

Наряду с достоинствами – простотой конструкции и высоким коэффициентом полезного действия (КПД) – электромеханические модуляторы имеют недостаточно высокую стабильность характеристик при длительной эксплуатации и изменяющихся условиях внешней среды. В ряде случаев наблюдается нелинейность преобразования. В частности, для пьезоэлектрических преобразователей характерны последствия и нелинейность гистерезиса. Другой существенный недостаток состоит в невысоком быстродействии, поскольку из-за инерционных механических элементов и узлов достигаемая скорость их перемещения обычно не

позволяет получить сдвиг частоты более 1-2 кГц для электродинамических и 20-30 кГц для малогабаритных пьезоэлектрических модуляторов.

Быстродействие процесса модуляции световых волн можно значительно увеличить при использовании принципов электрооптической модуляции света. В электрооптических модуляторах осуществляется преобразование фазы световой волны на основе использования свойств монокристаллов. Электрооптические кристаллы позволяют осуществлять модуляцию в полосе частот до сотен мегагерц. В интерферометрических системах обычно используют кристаллы с электрооптическим эффектом Поккельса.

При использовании анизотропного кристалла в общем случае справедливо уравнение для оптической индикатрисы:

$$\alpha x^2 + \beta y^2 + \gamma z^2 = 1, \qquad (1.5.8)$$

где $\alpha = 1/n_x$; $\beta = 1/n_y$; $\gamma = 1/n_z$; n_x, n_y, n_z – показатели преломления вдоль осей координат, совпадающих с главными осями монокристалла.

Вследствие двойного лучепреломления в электрическом поле уравнение (1.36) видоизменяется и принимает вид

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{31}zx = 1, \qquad (1.5.9)$$

где a_{ik} – поляризационные константы монокристалла, которые в случае эффекта Поккельса линейно зависят от напряженности внешнего электрического поля E_0. При этом их изменения равны

$$\Delta a = \rho_{ik} E_0, \qquad (1.5.10)$$

где ρ_{ik} – электрооптические коэффициенты.

Если внешнее электрическое поле приложено вдоль оси *z*, совпадающей с осью кристалла, то уравнение оптической индикатрисы принимает вид

$$\alpha(x^2 + y^2) + \gamma z^2 + 2\rho_{63}E = 1. \qquad (1.5.11)$$

Эта индикатриса представлена на рис. 1.20. В направлении главных осей эллипсоида (1.5.11) показатель преломления равен

$$n_{xy} \approx n_0(1 \pm n_0^2 \rho_{63} E/2). \qquad (1.5.12)$$

то есть линейно зависит от *E*. При прохождении света, поляризованного под углом $\pi/4$ к оси *x*, происходит фазовая модуляция, причем

$$\Delta\varphi(E) = \pi n_0^3 \rho_{63} El/\lambda. \qquad (1.5.13)$$

где *l* – длина монокристалла.

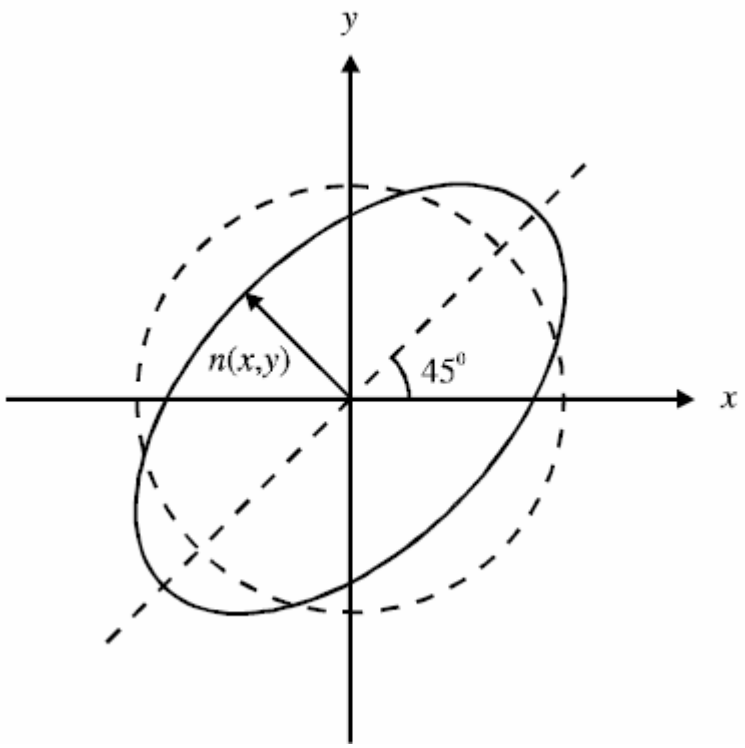


Рис. 1.20. Оптическая индикатриса монокристалла в случае приложения внешнего электрического поля вдоль оси *z*, совпадающей с осью монокристалла

Из (1.5.12) следует, что полуволновое напряжение U_π в случае продольного электрооптического эффекта равно

$$U_\pi = \lambda / 2n_0^3 \rho_{63}, \qquad (1.5.14)$$

при этом для кристалла KDP при длине волны $\lambda = 0{,}63$ мкм оно составит примерно 8 кВ. Столь высокое полуволновое напряжение осложняет получение требуемой глубины модуляции интерференционных сигналов.

Полуволновое напряжение значительно снижеается при использовании поперечного электрооптического эффекта, когда свет распространяется не вдоль оси *z*, а вдоль одной из наведенных осей – *x* или *y* (рис. 1.20). В этом случае фазовый сдвиг равен

$$\Delta\varphi(E) = 2\pi l(n_e - n_o)/\lambda + \pi n_0^3 \rho_{63} El/\lambda, \qquad (1.5.15)$$

где первое слагаемое обусловлено естественным двойным лучепреломлением кристалла, а второе характеризует воздействие электрического поля.

Для поперечного электрооптического эффекта

$$U_\pi = k\lambda / n_0^3 \rho_{63}, \qquad (1.5.16)$$

где $k = d/l$ – соотношение поперечного и продольного размеров кристалла.

Сравнивая (1.5.14) и (1.5.16), можно увидеть, что в последнем случае полуволновое напряжение снижается в $2k$ раз. Однако требуемое

модулирующее напряжение для реальных кристаллов и в том случае достигает сотен вольт.

Недостатками электрооптических модуляторов с гармоническим возбуждением являются невозможность обеспечить сдвиг частоты без появления комбинационных гармоник, а также нестабильность характеристик во времени вследствие значительного тепловыделения. Получение необходимого индекса фазовой модуляции требует использования высоких питающих напряжений.

Эти недостатки устранены в акустооптических модуляторах. Акустическая ультразвуковая волна, возбуждаемая в твердом теле или жидкости, создает локальные сжатия и разрежения среды. При этом имеют место изменения показателя преломления, то есть образуется периодическая фазовая решетка с шагом, равным длине ультразвуковой волны. При прохождении света через такую решетку возникает дифракция, и в первом порядке дифракции имеет место сдвиг частоты $\Delta\upsilon = \omega_0 / 2\pi$, где ω_0 – частота возбуждения акустооптического модулятора (АОМ) (рис. 1.21). Представленный на рисунке режим модуляции соответствует дифракции Брэгга на объемной периодической фазовой структуре, когда дифракционные максимумы наблюдаются в нулевом и, главным образом, в первом порядках, причем интенсивность света в других порядках дифракции близка к нулю.

Фазовая структура возникает при возбуждении в среде ультразвуковой волны, то есть волны механических напряжений. При наличии механического напряжения эллипсоид оптической индикатрисы материала деформируется, причем при одномерной деформации фотоупругая среда становится одноосной с оптической осью, направленной вдоль оси деформации. Для света, распространяющегося перпендикулярно этой оси, среда приобретает свойство двойного лучепреломления. Изменение показателя преломления можно выразить как

$$\Delta n = n_e - n_o = -n_0^3 p\zeta, \qquad (1.5.17)$$

где p – эффективный коэффициент фотоупругости, ζ – величина упругих деформаций в среде.

Дифракция на периодической объемной структуре с изменяющимся показателем преломления, согласно (1.5.17), характеризуется следующим соотношение интенсивности падающей I_0 и дифрагированной I_1 волн:

$$I_1 = I_0 \sin^2\left[\frac{\pi\sqrt{P_a l M / 2H}}{\lambda_0 \cos\alpha_Б}\right], \qquad (1.5.18)$$

где $M = n_0^6 p^2 / \rho v^3$ – константа материала, ρ – плотность материала, v – скорость звука, P_a – акустическая мощность, l – длина взаимодействия света с акустической волной, H – ширина звукопровода.

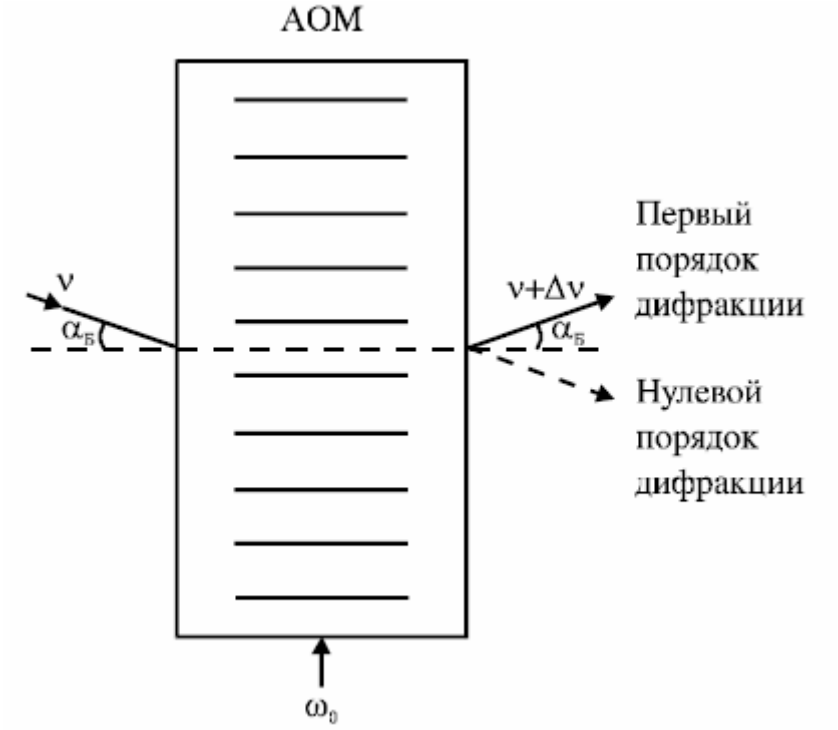


Рис. 1.21. Принцип акустооптической модуляции

Интенсивность света в первом спектральном порядке наибольшая, если свет падает под углом Брэгга к волновому фронту акустической волны:

$$\alpha_{Б} = \arcsin(\lambda / 2B), \qquad (1.5.19)$$

где λ – длина световой волны, B – длина акустической волны. Режим дифракции Брэгга имеет место при высоких ультразвуковых частотах и большой длине взаимодействия света с акустической волной.

Для типичных значений размеров H из (1.5.19) следует, что дифракционная эффективность оказывается близкой к единице при акустической мощности в несколько ватт. Быстродействие модулятора ограничено только значением времени пробега акустической волной сечения светового пучка. Акустические модуляторы позволяют осуществить сдвиг оптической частоты без появления побочных спектральных составляющих, которые возникают в электрооптических фазовых модуляторах. Питающее напряжение для возбуждения акустооптической модуляции, как правило, не превышает десятков ватт.

1.6 Принципы фотоэлектрической регистрации когерентных оптических полей

Основой процесса фотоэлектрической регистрации интерференционных полей является взаимодействие суммарного поля,

определяемого комплексными амплитудами измерительной $A_1(\mathbf{r},t)$ и опорной $A_2(\mathbf{r},t)$ волн, с материалом светочувствительной площадки фотодетектора (рис. 1.22). Фототок на выходе фотодетектора определяется операцией интегрирования по площади σ фотодетектора

$$\xi(t) = \frac{e}{h\upsilon}\iint_{\sigma}\eta(\mathbf{r})I(\mathbf{r},t)d^2\mathbf{r} + n(t), \qquad (1.6.1)$$

где e – заряд электрона, $h\upsilon$ – энергия фотона, $n(t)$ – флуктуационная компонента фототока, обусловленная стохастическим характером появления фотоэлектронов, $\eta(\mathbf{r})$ – квантовая эффективность материала чувствительной площадки фотодетектора.

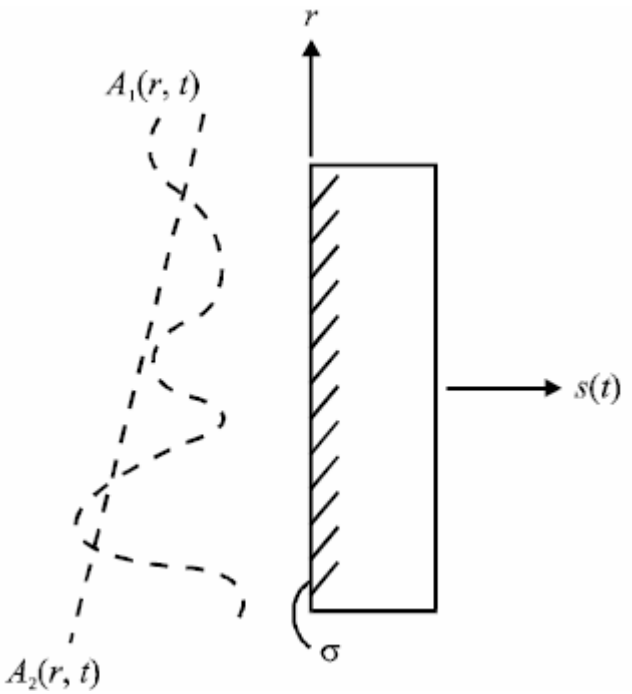


Рис. 1.22. Регистрация интерференционного поля на чувствительной площадке фотодектора

Основная составляющая фототока равна

$$s(t) = s_0 + \frac{e}{h\upsilon}(\mathbf{p}_1\cdot\mathbf{p}_2)\iint_{\sigma}\eta(\mathbf{r})[A_1(\mathbf{r},t)A_2^*(\mathbf{r},t) + A_1^*(\mathbf{r},t)A_2(\mathbf{r},t)]d^2\mathbf{r}, \qquad (1.6.2)$$

где

$$s_0 = \frac{e}{h\upsilon}\iint_{\sigma}\eta(\mathbf{r})[|A_1(\mathbf{r},t)|^2 + |A_2(\mathbf{r},t)|^2]d^2\mathbf{r}, \qquad (1.6.3)$$

$\mathbf{p}_1, \mathbf{p}_2$ – единичные векторы, определяющие направления поляризации интерферирующих волн.

Если квантовая эффективность и комплексные амплитуды интерферирующих волн не зависят от $\mathbf{r}$ и t, то вторую составляющую (1.6.2) можно записать в форме

$$s(t) = \frac{e}{h\upsilon}\sigma(\mathbf{p}_1 \cdot \mathbf{p}_2)\left|A_1\right\|A_2\left|\cos[\delta(t)]\right., \qquad (1.6.4)$$

где $\delta(t)$ – разность фаз между измерительной и опорной волнами, которая может изменяться во времени в случае контроля динамических объектов. При использовании подходящего метода модуляции оптической разности хода полезную составляющую (1.6.4) несложно выделить методами спектральной фильтрации. При этом можно найти амплитуду и фазу полезного интерференционного сигнала.

Влияние флуктуационной составляющей $n(t)$ в (1.6.1) определяется ее корреляционными свойствами. Вследствие конечной протяженности во времени импульсного отклика фотодетктора длина корреляции составляющей $n(t)$ возрастает, и шум имеет ограниченную полосу частот. Значения полезной составляющей сигнала (1.6.4) пропорциональны амплитуде опорной волны. При определенных условиях увеличение этого параметра позволяет повысить отношение сигнал-шум.

Пусть опорная волна $A_2(\mathbf{r},t)$ имеет регулярную модуляцию – амплитудную, частотную, фазовую или поляризационную. Тогда ее временная зависимость в точке $\mathbf{r}$ в плоскости фотодектора приводит к соответствующей модуляции полезной составляющей (1.6.2)

$$s(t) = \frac{2\eta e}{h\upsilon}\sigma[\mathbf{p}_1 \cdot \mathbf{p}_2(t)]A_1A_2(t), \qquad (1.6.5)$$

то есть поляризационная модуляция в результате подобна амплитудной. Спектр интерференционного сигнала определяется законами модуляции $\mathbf{p}_2(t)$ и $A_2(t)$. При использовании модуляции спектр полезного сигнала смещается в область высоких частот, что обеспечивает повышенную помехозащищенность и чувствительность системы при обработке сигнала на переменном токе.

При фотоэлектрической регистрации интерферирующих волн имеет важное значение согласование их волновых фронтов в пределах площадки фотодетектора (рис. 1.22).

Пусть волны поляризованы в одной плоскости, то есть $(\mathbf{p}_1 \cdot \mathbf{p}_2) = 1$. При этом, опуская временную зависимость, запишем выражение для информационной составляющей сигнала

$$A_1(\mathbf{r}) = \left|A_1\right|\exp(j\mathbf{k} \cdot \mathbf{r})\exp(j\varepsilon), \qquad (1.6.6)$$

в форме

$$s = \frac{2\eta e}{h\upsilon}\iint_{\sigma} A_1(\mathbf{r})A_2^*(\mathbf{r})d^2\mathbf{r}. \qquad (1.6.7)$$

При интерференции плоских волн можно получить

$$A_1(\mathbf{r}) = \left|A_1\right|\exp(j\mathbf{k}_1 \cdot \mathbf{r})\exp(j\varepsilon), \qquad (1.6.8)$$

$$A_2(\mathbf{r}) = \left|A_2\right|\exp(j\mathbf{k}_2 \cdot \mathbf{r}), \qquad (1.6.9)$$

где $\mathbf{k}_1$ и $\mathbf{k}_2$ – волновые векторы. Из (1.6.8)–(1.6.9) получим

$$s = \frac{2\eta e}{h\upsilon}|A_1\|A_2|\iint_{\sigma}\exp[j(\mathbf{k}_1 - \mathbf{k}_2)\mathbf{r} + \varepsilon]d^2\mathbf{r}\,. \qquad (1.6.10)$$

Если чувствительная площадка имеет квадратную форму с размерами $2b \times 2b$, из (1.6.10) можно получить

$$s = \frac{2\eta e}{h\upsilon}|A_1\|A_2|\int_{-b}^{b}\int\exp[j(\Delta k_x x + \Delta k_y y + \varepsilon)]dxdy =$$

$$= \frac{2\eta e}{h\upsilon}|A_1\|A_2|\sigma\,\mathrm{sinc}(\Delta k_x b)\mathrm{sinc}(\Delta k_y b)\exp(j\varepsilon)\,, \qquad (1.6.11)$$

где $\sigma = 4b^2$ – площадь чувствительной площадки, $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$, Δk_x и Δk_y – проекции вектора $\Delta\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2$ на оси координат (x, y) площадки фотодетектора.

Из (1.6.11) следует, что максимум полезного сигнала достигается при условии $(\Delta k_x b), (\Delta k_x b) \to 0$. Это означает, во-первых, что необходимо минимизировать значения Δk_x и Δk_y, то есть обеспечивать одинаковый угол падения волн на чувствительную площадку, и, во-вторых, уменьшать размеры этой площадки ($b \to 0$). Уменьшение размеров площадки снижает полезный сигнал в квадратичной зависимости вследствие влияния множителя σ в (1.6.11). Поэтому во избежание энергетических потерь необходимо использовать многоэлементные приемники излучения, содержащие большое число независимых фоточувствительных ячеек.

Если волновые фронты не являются идеально плоскими, то вместо (1.39) следует записать

$$A_1(\mathbf{r}) = |A_1(\mathbf{r})|\exp[j\varphi_1(\mathbf{r})]\,, \qquad (1.6.12)$$

$$A_2(\mathbf{r}) = |A_2(\mathbf{r})|\exp[j\varphi_2(\mathbf{r})]\,, \qquad (1.6.13)$$

где $\varphi_1(\mathbf{r})$ и $\varphi_2(\mathbf{r})$ определяют фазовые отклонения по волновому фронту. При этом выражение (1.52) преобразуется к виду

$$s = \frac{2\eta e}{h\upsilon}\iint_{\sigma}|A_1(\mathbf{r})\|A_2(\mathbf{r})|\cos[\Delta\varphi(\mathbf{r})]d^2\mathbf{r}\,, \qquad (1.6.14)$$

где $\Delta\varphi(\mathbf{r}) = \varphi_2(\mathbf{r}) - \varphi_1(\mathbf{r})$. Следовательно, максимум сигнала (1.57) достигается при следующем условии согласования волновых фронтов:

$$\varphi_2(\mathbf{r}) = \varphi_1(\mathbf{r}) + 2\pi n,\ n = 0, \pm1, \pm2, \ldots\,, \qquad (1.6.15)$$

Фактически это означает, что волновые фронты в пределах площадки σ должны быть одинаковыми. Условие локального согласования волновых фронтов может быть обеспечено упомянутым выше способом – использованием многоэлементных приемников излучения, когда локальные рассогласование волновых фронтов достаточно мало.

Допустимый локальный угол рассогласования нормалей α к волновым фронтам можно оценить из условия, которое следует из (1.6.11), в форме

$$\frac{2\pi}{\lambda}\alpha b \le \frac{\pi}{4}, \qquad (1.6.16)$$

то есть $\alpha_{\max} \le \lambda/8b$. Следовательно, при уменьшении размера b элементарной чувствительной площадки возрастают допустимые пределы рассогласования волновых фронтов.

Рассмотрим влияние флуктуационной составляющей интерференционного фотоэлектрического сигнала.

Известно, что флуктуации числа фотоэлектронов, учитываемые слагаемым $n(t)$ в (1.6.1), подчиняются распределению Пуассона. Число регистрируемых фотоэлектронов зависит от общей интенсивности света I, поэтому вероятность регистрации N фотоэлектронов в общем случае следует определить как

$$P(N) = \int_0^\infty P(N/I)\,p_I(I)dI = \int_0^\infty \frac{(\beta I)^N}{N!}\exp(-\beta I)\,p_I(I)dI, \qquad (1.6.17)$$

где $P(N/I)$ – условная вероятность, β – параметр распределения Пуассона. Из (1.6.17) следует, что плотность вероятности $P(N)$, вообще говоря, может не подчиняться закону Пуассона.

Рассмотрим статистические характеристики фотоотсчетов при использовании многоэлементного приемника излучения, следуя методике.

Запишем вначале выражения для интерференционного сигнала, полученного при фотоэлектрической регистрации колебаний в фиксированной точке **r**, когда направления волновых векторов измерительной и опорной плоских волн совпадают ($\Delta\mathbf{k} = 0$), то есть

$$s = \left\langle EE^* \right\rangle = \mu|A_1|^2 + \mu|A_2|^2 + 2\mu|A_1||A_2|\cos(2\pi\Delta/\lambda) =$$
$$= \mu(I_1 + I_2)[1 + V\cos(2\pi\Delta/\lambda)], \qquad (1.6.18)$$

где μ, как и ранее, постоянный коэффициент,

$$V = \frac{2\sqrt{I_1 I_2}}{I_1 + I_2}, \qquad (1.6.19)$$

представляет собой видность интерференционных полос.

При использовании монохроматического источника излучения искомым параметром является оптическая разность хода Δ. В случае источника с ограниченной когерентностью, учитывая (1.5.18), информационным параметром может быть видность полос $V(\Delta)$, причем $V \to V_{\max}$ при $\Delta \to 0$.

Рассмотрим особенности определения параметров Δ и V при обработке сигналов многоэлементного приемника излучения.

Прежде всего, оптическую разность хода требуется «развернуть» в плоскости регистрации (x, y), задавая некоторый малый угол между волновыми векторами и соблюдая условие (1.6.16). При этом возникает

зависимость интерференционного сигнала от пространственных координат, а именно

$$s(x,y) = s_0\{1 + V(x,y)\cos[2\pi(ux + vy) + \varepsilon]\}, \quad (1.6.20)$$

где $s_0 = \mu(I_1 + I_2)$, пространственные частоты (u, v) определяются из соотношений

$$u = \frac{\Delta k_x}{2\pi},\ v = \frac{\Delta k_y}{2\pi}, \quad (1.6.21)$$

ε – начальная фаза сигнала в точке $x = 0$, $y = 0$. Для одномерного случая можно переписать (1.6.20) в форме

$$s(x) = s_0[1 + V(x)\cos(2\pi ux + \varepsilon)]. \quad (1.6.22)$$

Пусть сигнал (1.6.22) сформирован многоэлементным приемником излучения, состоящим из K фоточувствительных ячеек, линейно расположенных без зазоров с достаточно малым размером отдельной ячейки b, таким, чтобы амплитуду и фазу интерференционного поля можно было считать постоянными. Тогда (1.6.22) можно переписать в виде

$$s(k) = s_0[1 + V\cos(2\pi mk / K + \varepsilon)], \quad (1.6.23)$$

где $m = ubK$ – известное число периодов интерференционного сигнала в пределах поля зрения фотодетектора.

Найдем косинусную и синусную составляющие $s(k)$ в (1.6.23) в форме, соответственно

$$C = \frac{1}{K}\sum_{k=0}^{K-1} s(k)\cos(2\pi mk / K) = \frac{s_0}{2} V\cos\varepsilon, \quad (1.6.24)$$

$$S = \frac{1}{K}\sum_{k=0}^{K-1} s(k)\sin(2\pi mk / K) = \frac{s_0}{2} V\sin\varepsilon. \quad (1.6.25)$$

Искомые параметры можно вычислить из (1.6.24)–(1.6.25), используя простые соотношения

$$\hat{\varepsilon} = \operatorname{arctg}(S / C), \quad (1.6.26)$$

$$\hat{V} = \sqrt{S^2 + C^2} / s_0, \quad (1.6.27)$$

где s_0 вычисляется усреднением значений $s(k)$ в (1.6.23).

Следует подчеркнуть, что выражения (1.6.26) и (1.6.27) записаны в предположении отсутствия флуктуаций фотоэлектронов в процессе фотодетектирования интерференционного поля. Для определения ограничений точности оценок параметров (1.6.26) и (1.6.27) необходимо учесть влияние таких флуктуаций.

На рис. 1.23 показана векторная диаграмма, иллюстрирующая влияние флуктуаций на видность V и фазу ε.

Амплитуда полезной составляющей $s_m = \sqrt{S^2 + C^2}$ оценивается с погрешностью, обусловленной отклонениями σ_C и σ_S, которые вызваны влиянием шума. При достаточно большой амплитуде $s_m >> \sigma_C, \sigma_S$ можно выразить как

$$C/\sigma_C = V\sqrt{s_0/2}\,, \qquad (1.6.28)$$

где s_0 – среднее суммарное число фотоэлектронов, генерируемых при регистрации измерительной и опорной волн.

Погрешность определения фазы, очевидно, выражается соотношением

$$\sigma_\varepsilon \approx \sigma_s / C = (1/V)\sqrt{2/s_0}\,. \qquad (1.6.29)$$

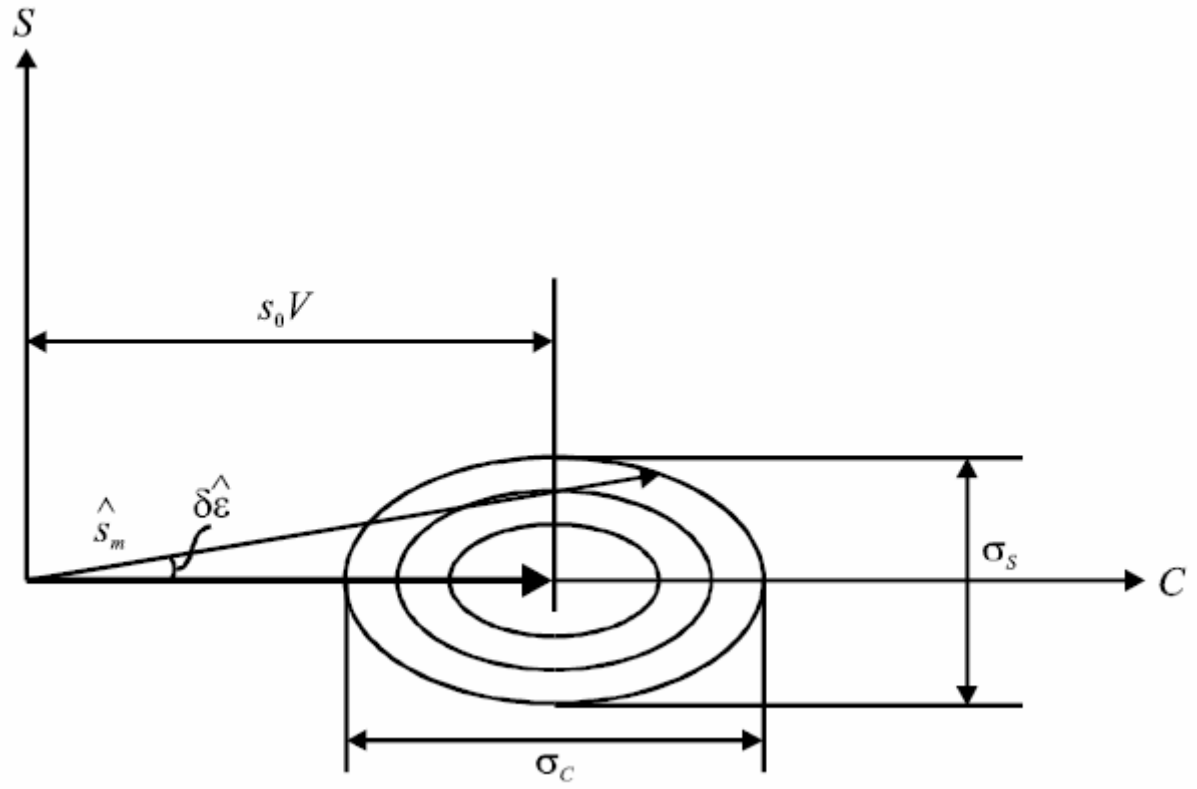


Рис. 1.23. Векторная диаграмма, показывающая влияние случайных флуктуаций на видность *V* и фазу ε

Таким образом, отношение сигнал-шум (1.6.28) пропорционально корню квадратному из полного числа фотособытий, что соответствует свойствам распределения Пуассона, и пропорционально видности интерференционных полос. Для сохранения заданного соотношения сигнал-шум при уменьшении видности V необходимо увеличивать интенсивность излучения источника.

Погрешность определения фазы обратно пропорциональна квадратному корню из полного числа фотособытий и обратно пропорциональна видности интерференционных полос.

Рассмотрим изменение интерференционного сигнала при отклонении фазы ε, обусловленных изменением оптической разности хода интерферирующих волн.

При использовании многоэлементного приемника излучения сигналы от отдельных светочувствительных ячеек можно рассматривать как

многомерный вектор наблюдений, компоненты которого изменяются в соответствии с выражением

$$s_i(k) = s_{0i}[1 + V_i \cos(2\pi mk / K + \varepsilon_i)], \qquad (1.6.30)$$

где i – номер светочувствительной ячейки. Используя рассмотренную выше методику, можно оценить погрешности применительно к параметрам многомерного сиганла (1.6.30), как это иллюстрируется обобщенной схемой процесса фотоэлектрической регистрации многомерного интерференционного сигнала, показанной на рис. 1.24.

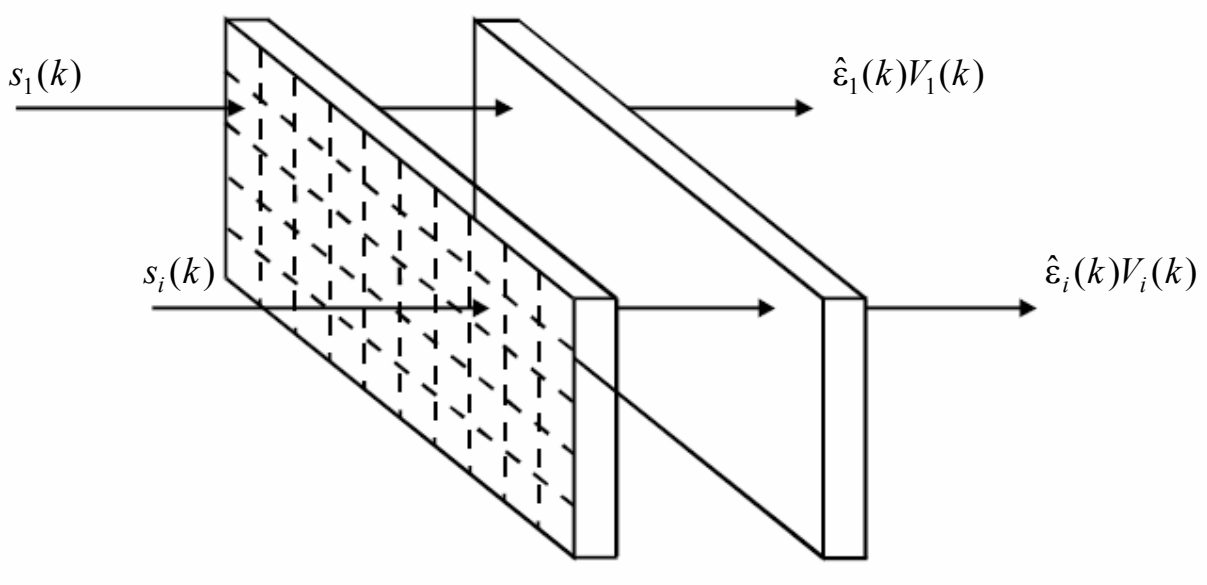


Рис. 1.24. Обобщенная схема регистрации и обработки многомерного интерференционного сигнала

Компоненты $s_i(k)$ многомерного сигнала изменяются в соответствии с заданными фазовыми сдвигами $\Delta\varphi_k$ опорного отражателя интерферометра. Зарегистрированные данные анализируются в системе обработки с получением оценок искомых параметров сигнала. В частности, параметр $\hat{\varepsilon}$ показывает отклонения формы волнового фронта измерительной волны и, следовательно, характеризует профиль поверхности контролируемого объекта.

1.7 Виды шумов в физических системах

При формировании и регистрации оптических сигналов имеет место влияние шумов различной физической природы, которое необходимо учитывать при создании систем и устройств, особенно для обеспечения высокой чувствительности и точности.

Для некоторых физических процессов характеристики шумов известны заранее. К основным видам шумов в физических системах, в том

числе электронных, относятся тепловой шум, дробовой шум и фликер-шум.

Тепловой шум характерен для электронных цепей, он возникает вследствие случайных флуктуаций скорости носителей заряда. Этот механизм шума часто относят к известной модели броуновского движения (применительно к носителям заряда), обусловленного тепловой энергией в материале и являющегося механизмом сохранения термодинамического равновесия.

Дробовой шум может быть связан с прохождением тока, то есть отдельных носителей заряда через потенциальный барьер (например, в области p-n перехода) и в этом смысле такой шум является разновидностью шума неравновесного характера. Наиболее наглядно этот вид шума иллюстрируется случайным процессом эмиссии электронов из катода (например, термоэлектронном диоде).

Интересно заметить, что несмотря на различную физическую природу теплового и дробового шума, структура шумовых сигналов обоих типов сходна: она представляет собой последовательность случайных импульсов, похожих по форме и случайно распределенных во времени.

Если форма одиночного импульса описывается некоторой известной функцией $h(t)$, причем $h(t) = 0$ при $t < 0$ (так называемое условие каузальности или причинности), то результирующая последовательность $s(t)$ есть линейная суперпозиция импульсов

$$s(t) = \sum_k a_k h(t - t_k), \tag{1.7.1}$$

где a_k – амплитуда k-ого импульса. Распределение t_k подчиняется закону Пуассона, согласно которому плотность вероятности определяется формулой

$$p(m, \tau) = \frac{(\nu\tau)^m}{m!}\exp(-\nu\tau), \tag{1.7.2}$$

где m – среднее число событий (импульсов) за интервал времени τ. Первый статистической момент распределения Пуассона обозначается как

$$\langle m \rangle = \sum_{m=0}^{\infty} m p(m, \tau) = \nu\tau, \tag{1.7.3}$$

откуда ясен смысл параметра ν: это средняя скорость появления импульсов. Дисперсия распределения равна

$$\sigma^2 = \left\langle (m - \langle m \rangle)^2 \right\rangle = \left\langle m^2 \right\rangle - \langle m \rangle^2 = \sum_{m=0}^{\infty} m^2 p(m, \tau) - \langle m \rangle^2. \tag{1.7.4}$$

После подстановки выражения (1.7.2) в (1.7.4), можно получить соотношение

$$\sigma^2 = \langle m \rangle, \tag{1.7.5}$$

т.е. дисперсия такого процесса равна его среднему значению.

Спектральные свойства шумовых процессов

Форма сигнала, состоящего из импульсов $h(t)$, характеризуется спектральной плотностью мощности

$$S(\omega) = 2\nu\left\langle a^2 \right\rangle \left|H(j\omega)\right|^2, \tag{1.7.6}$$

где ν – средняя частота событий, $\left\langle a^2 \right\rangle$ – среднее значение квадрата амплитуды, $H(j\omega)$ – результат преобразования Фурье функции $h(t)$. Для бесконечно коротких импульсов $h(t)$, т.е. δ-импульсов, $\left|H(j\omega)\right| \to 1$, поэтому спектральная плотность (1.7.6) равна

$$S(\omega) = 2\nu\left\langle a^2 \right\rangle. \tag{1.7.7}$$

Таким образом, импульсы, возникающие в результате дискретных событий, имеют равномерные спектральные плотности.

Значение спектральной плотности теплового шума для датчика сопротивлением R составляет

$$\left\langle S(\omega) \right\rangle = 4kTR. \tag{1.7.8}$$

где kT – средняя тепловая энергия на каждую степень свободы, k – постоянная Больцмана, T – абсолютная температура. Последнюю формулу часто называют формулой Найквиста.

Спектральная плотность дробового шума при среднем токе I равна

$$\left\langle S(\omega) \right\rangle = 2eI. \tag{1.7.9}$$

где e – заряд электрона. Эту формулу легко получить, полагая, что среднее число импульсов в единицу времени равно I/e, все амплитуды импульсов равны e, что позволяет записать $\left\langle a^2 \right\rangle = e^2$.

Особенностью фликер-шума является то, что его спектральная плотность изменяется пропорционально степени частоты $f^{-\alpha}$, где $\alpha = 0{,}8$–$1{,}2$. Этот вид шума наиболее заметен на инфранизких частотах $f^{-\alpha}$ (на более высоких частотах происходит маскирование другими видами шума).

Фликер-шум можно представить как последовательность случайных импульсов с определенной функцией формы, а именно

$$h(t) = u(t)/\sqrt{t}. \tag{1.7.10}$$

где $u(t)$ – единичная ступенчатая функция, спектр которой имеет вид

$$H(j\omega) = \sqrt{\frac{\pi}{\omega}} \exp(\pm j\pi/4). \tag{1.7.11}$$

где $\omega \neq 0$, знак плюс соответствует значениям $\omega < 0$, а знак минус – $\omega > 0$. Таким образом, выражение для спектральной плотности фликер-шума можно записать в виде

$$S(\omega) = 2\nu\left\langle a^2 \right\rangle \left|H(j\omega)\right|^2 \approx 1/\left|\omega\right|. \tag{1.7.12}$$

Шумы в оптических системах

Характеристики шумов в оптических системах имеют ванное отличие: в оптике часто требуется исследовать световые поля и, следовательно, составляющие шума не только во временной, но и в пространственной области. Типичные виды шумов представлены в табл. 1.1.

Таблица 1.1. Виды шумов в оптических системах

Физический процесс	Характеристики	Примечания
1. Шум спонтанного лазерного излучения	$\langle n_s(t)\rangle = 0$ $\langle n_s^2(t)\rangle = \sigma_s^2$ $R_s(\tau) = \sigma_s^2 \exp(-B_s\lvert\tau\rvert)$	Лоренцев шум с шириной полосы B_s
2. Шум источника белого света	$\langle n_s(t)\rangle = 0$ $\langle n_s^2(t)\rangle = \sigma_s^2$ $R_s(\tau) = \sigma_s^2 \exp(-B_s\tau^2)\cos(\omega_0\tau)$	Гауссов шум с шириной полосы B_s и центральной частотой ω_0
3. Фазовый шум спеклов	$\langle\gamma\rangle = 0$ $\langle\gamma^2\rangle = \sigma_\gamma^2$ $p(\gamma) = (1/\sqrt{2\pi\sigma_\gamma^2})\exp(-\gamma^2/2\sigma_\gamma^2)$ $R_\gamma(\tau) = \sigma_\gamma^2 \exp(-r^2/r_0^2)$	Гауссов шум с длиной корреляции r_0
4. Интерференционная картина со случайной фазой $I = I_0 + \tilde{I}$ $\tilde{I} = I_m \cos\Phi$ $\Phi = \varepsilon + u_0 x + \nu_0 y + \omega_0 t$	$\langle I\rangle = I_0$ $\langle I^2\rangle = I_0^2 + I_m^2/2$ $p(\varepsilon) = 1/\sqrt{2\pi}$, $-\pi < \varepsilon < \pi$ $\sigma^2 = I_m^2/2$ $R_\gamma(\alpha,\beta,\tau) = \sigma^2\cos\Phi$	Гармонический случайный процесс
5. Фазовые флуктуации в оптической системе	$\langle\psi(t)\rangle = 0$ $\langle\psi^2(t)\rangle = \sigma_\psi^2$ $R_\psi(\tau) = (\sigma_\psi^2/2B_\psi)\delta(\tau)$	Гауссов некоррелированный шум с полосой B_ψ
6. Дробовой шум приемника излучения	$\langle n(t)\rangle = 0$ $\langle n^2(t)\rangle = \sigma_n^2 = 2e\langle s\rangle B_n$ $R_n(\tau) = (\sigma_n^2/2B_n)\delta(\tau)$	Гауссов шум с полосой B_n. Дисперсия пропорциональная среднему значению сигнала

Рассмотрим влияние помех, вносимых источником излучения, оптической системой и приемником излучения на примере интерферометрической системы. Учитываться будут два основных вида шумов: шум лазерного источника и фазовый пространственный шум. Для лазерного источника наиболее характерен высокочастотный (выше 10 кГц) Лоренцев шум $n_s(t)$. При этом выражения для комплексных амплитуд двух интерферирующих волн записываются в виде

$$A_1 = [1 + n_s(t+\tau)]\left|A_1\right|\exp(j\varepsilon_1), \tag{1.7.13}$$

$$A_2 = [1 + n_s(t)]\left|A\right|\exp(j\varepsilon_2). \tag{1.7.14}$$

Влияние рассеивания на оптических поверхностях, то есть пространственный фазовый шум $\gamma(x)$, учитывается в форме (для одномерной модели)

$$A_1 = \mathrm{rect}(x/D)\left|A_1\right|\exp(j2\pi u_0 x)\exp(j\gamma(x)), \tag{1.7.15}$$

$$A_2 = \mathrm{rect}(x/D)\left|A_2\right|\exp(-j2\pi u_0 x), \tag{1.7.16}$$

где

$$\mathrm{rect}(x/D) = \begin{cases} 1, |x| \le D/2, \\ 0, |x| > D/2 \end{cases} \tag{1.7.17}$$

представляет собой прямоугольную функцию окна, которая задает границы световых пучков в пределах области *D*.

При интерференции регистрируется результирующая интенсивность света, которая определяется квадратом амплитуды, а именно

$$\begin{aligned} I(x,t,\varepsilon) &= (A_1 + A_2)(A_1 + A_2)^* = \\ &= \frac{1}{2} I_1 [1 + n_s(t+\tau)]^2 \cos[\varepsilon + 4\pi u_0 x + \gamma(x)] + \\ &+ \frac{1}{2} I_2 [1 + n_s(t)]^2 \cos[\varepsilon + 4\pi u_0 x + \gamma(x)] + \\ &+ \frac{1}{2} I_m [1 + n_s(t) + n_s(t+\tau)]^2 \cos[\varepsilon + 4\pi u_0 x + \gamma(x)] + \\ &+ \frac{1}{2} n_s(t) n_s(t+\tau) \cos[\varepsilon + 4\pi u_0 x + \gamma(x)] + \frac{1}{\mu} n(t), \end{aligned} \tag{1.7.18}$$

где $I_1 = \mathrm{rect}(x/D)\left|A_1\right|^2$, $I_2 = \mathrm{rect}(x/D)\left|A_2\right|^2$, $I_m = 2\sqrt{I_1 I_2}$, $\varepsilon = \varepsilon_1 - \varepsilon_2$, $n(t)$ – дробовой шум, приведенный ко входу.

Влияние шумов можно уменьшить за счет усреднения учитывая, что

$$\langle n_s(t) \rangle = \langle n_s(t+\tau) \rangle = 0, \tag{1.7.19}$$

$$\langle n_s^2(t) \rangle = \sigma_n^2, \tag{1.7.20}$$

$$\langle n_s(t) n_s(t+\tau) \rangle = R_n(\tau). \tag{1.7.21}$$

Из табл. 1.1 следует, что корреляционная функция лоренцева шума $R_s(\tau) = \sigma_s^2 \exp(-B_s|\tau|)$. При $B_s \le 10$ кГц и $\tau < 1$ мкс (что соответствует интервалу времени распространения светового луча на расстояние 300 м) $R_s(\tau) \approx \sigma_s^2$, следовательно, модель (1.7.18) можно представить в упрощенной форме

$$I(x,t,\varepsilon) = \frac{1}{2}\{I_1' + I_2' + I_m'\}\cos[\varepsilon + 4\pi u_0 x + \gamma(x)]\} + \frac{1}{\mu}n'(t), \qquad (1.7.22)$$

где $I_i' = I_i(1+\sigma_n^2)$, $n'(t)$ – остаточный шум с дисперсией $\sigma_n^2 = \sigma_0^2 / 2BT$, где B – ширина полосы, T – интервал усреднения.

Умножив левую и правую часть на коэффициент преобразования μ, можно получить модель интерферометрического сигнала

$$s(x,t,\varepsilon) = s_0 + s_m \cos[\varepsilon + 4\pi u_0 x + \gamma(x)]\} + n'(t). \qquad (1.7.23)$$

Это пример модели процесса почти периодического по переменной x и случайного по переменной t.

При анализе и обработке данных во многих случаях требуется определять параметры процесса, а именно начальную фазу ε, частоту u_0 или амплитуду s_m.

Контрольные вопросы

1. Приведите зависимость длины когерентности от ширины спектра источника излучения.
2. Какие виды проекций используются при рассмотрении особенностей формирования некогерентных изображений?
3. Перечислите основные виды искажений в реальных системах формирования изображений.
4. Объясните, каким образом учитывается степень влияния шероховатости поверхности на диаграмму направленности отраженного излучения.
5. Какова степень взаимной корреляции амплитуды и фазы спекл-поля?
6. Напишите выражение для дисперсии интенсивности в картине спеклов.
7. Какому закону распределения вероятностей подчиняются значения спекл-интерференционной картины?
8. Проведите сравнение формирования оптических полей в когерентной и некогерентной системах.
9. Перечислите основные физические явления, лежащие в основе методов преобразования световых полей.

Список литературы

1.1. Борн М., Вольф Э. Основы оптики. – М.: Наука, 1973.

1.2. Гудмен Дж. Статистическая оптика. – М.: Мир, 1988.

1.3. Гудмен Дж. Введение в Фурье-оптику. – М.: Мир, 1970.

1.4. Исимару А. Распространение и рассеяние волн в случайно-неоднородных средах. – М.: Мир, 1981.

1.5. Порфирьев Л.Ф. Основы теории преобразования сигналов в оптико-электронных системах. – Л.: Машиностроение, 1989.

1.6. Гуров И.П. Формирование и анализ стохастических интерференционных полей. В кн.: Проблемы когерентной и нелинейной оптики / Под ред. И.П. Гурова и С.А. Козлова. – СПб.: СПбГУИТМО, 2000. – С. 67–87.

1.7. Ван дер Зил. Шумы при измерениях. – М.: Мир, 1979.

1.8. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложениях к интерферометрическим системам. – СПб: БХФВ Санкт-Петербургу, 1998.

Раздел 2. Математические модели сигналов и систем

2.1 Системные преобразования сигналов в фотонике

При обработке сигналов в системах фотоники требуется определять значения параметров сигналов, которые известным образом связаны со свойствами исследуемых объектов. Обработка согласно выбранным критериям качества осуществляется на основе математических моделей оптических полей и систем.

Рассмотрим обобщенную структурную схему (рис. 2.1), в которой можно выделить две основные части: входное измерительное преобразование и определение искомых параметров объекта. Первую операцию можно представить в форме

$$y = \mathrm{T}\{x\}, \tag{2.1.1}$$

где x – независимая переменная («вход»), y – отклик («выход»), $\mathrm{T}\{.\}$ – обобщенный оператор системы.

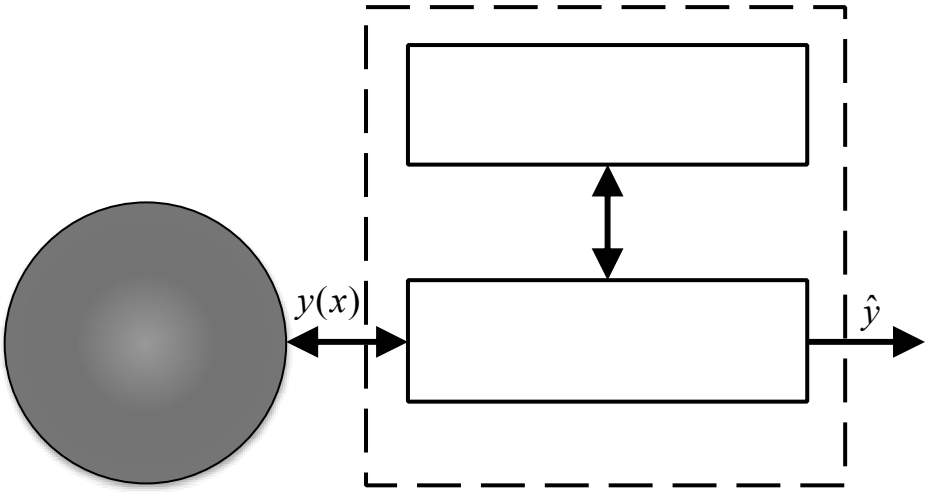


Рис. 2.1. Обобщенная структурная схема системы

При входном преобразовании, а именно, освещении объекта и регистрации суммарного поля (рассеянной и опорной волн), в результате получают на выходе распределение интенсивности $\hat{y} \to I$, зависящее от свойств источника излучения $x \to I_0$. Свойства объекта учитываются оператором системы, а именно

$$y = \mathrm{T}(\boldsymbol{\theta}) I_0, \tag{2.1.2}$$

где $\mathrm{T}(\boldsymbol{\theta})$ рассматривается как аппаратная функция, аргументом которой является искомый вектор параметров объекта $\boldsymbol{\theta}$. Поэтому вторая из упомянутых выше операций, а именно, восстановление искомых

параметров объекта, заключается в параметрической идентификации аппаратной функции системы.

В условиях рассеянной измерительной волны аппаратная функция $\mathrm{T}(\boldsymbol{\theta})$ является, вообще говоря, нелинейной случайной функцией, зависящей от пространственных координат и времени. Этим определяется сложность решения обратной задачи идентификации и, в результате, определения свойств объекта.

Если априорная информация об исследуемом процессе не вносит необходимых ограничений, то, вообще говоря, существует множество различных моделей, удовлетворяющих данным. В ряде случаев вид аппаратной функции является известным, поэтому модели отличаются значениями параметров в ограниченной области их изменения. Тогда проблема носит название идентификации в узком смысле и сводится к задачам оценивания параметров и фильтрации параметров сигналов и изображений.

В активной оптической системе интенсивность I_0 источника излучения заранее известна. При известном коэффициенте преобразования μ интенсивности света в электрический сигнал $s = \mu I = \mu' I / I_0$ выражение (2.1.2) можно записать в форме

$$s = \mathrm{T}(\mathbf{r}, \boldsymbol{\theta}), \qquad (2.1.3)$$

где $\mathbf{r}$ – точка регистрации сигнала, $\boldsymbol{\theta}$ – вектор параметров аппаратной функции размером $M \times 1$.

Уравнение (2.1.3) является нелинейным, поэтому для обеспечения возможности использования для решения обратной задачи математического аппарата теории линейных систем необходимо применить процедуру линеаризации задачи.

Пусть имеется некоторое априорное приближение $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_M)^{\mathrm{T}}$ для значений компонентов θ_i вектора параметров $\boldsymbol{\theta}$ $(i = 1, ..., M)$. Тогда, решая прямую задачу, можно вычислить соответствующее этому приближению теоретические значения отсчетов $\hat{s}_k = \hat{s}(x_k)$ сигнала

$$s_k = \mathrm{T}(x_k, \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_M), k = 1, ..., K. \qquad (2.1.4)$$

Разложение (2.4) в ряд Тейлора в окрестности $\hat{\boldsymbol{\theta}}$ пространства параметров позволяет получить

$$\begin{aligned} s_k &= \mathrm{T}(x_k, \hat{\theta}_1 + \delta\hat{\theta}_1, \hat{\theta}_2 + \delta\hat{\theta}_2, ..., \hat{\theta}_M + \delta\hat{\theta}_M) = \\ &= \hat{s}_k + \frac{\partial \mathrm{T}}{\partial \hat{\theta}_1}\Delta\hat{\theta}_1 + \frac{\partial \mathrm{T}}{\partial \hat{\theta}_2}\Delta\hat{\theta}_2 + ... + \frac{\partial \mathrm{T}}{\partial \hat{\theta}_M}\Delta\hat{\theta}_M + O(\Delta), \end{aligned} \qquad (2.1.5)$$

где $O(\Delta)$ – остаток ряда, содержащий слагаемые, содержащие производные более высоких порядков. Если $\Delta\hat{\theta}_i$ достаточно малы при всех k, то остатком ряда (2.5) можно пренебречь, при этом

$$s_k - \hat{s}_k = \frac{\partial \mathrm{T}}{\partial \hat{\theta}_1}\Delta\hat{\theta}_1 + \frac{\partial \mathrm{T}}{\partial \hat{\theta}_2}\Delta\hat{\theta}_2 + ... + \frac{\partial \mathrm{T}}{\partial \hat{\theta}_M}\Delta\hat{\theta}_M, \quad (2.1.6)$$

или

$$\mathbf{s} = \mathbf{D}\Delta\boldsymbol{\theta}, \quad (2.1.7)$$

где $\mathbf{s}$ – вектор разностей между измеренными и расчетными значениями сигнала размера $K \times 1$, D – матрица частных производных (матрица чувствительности) размера $K \times M$, $\Delta\boldsymbol{\theta}$ – вектор приращений параметров размера $M \times 1$. Модель (2.1.7) является линейной, и искомые значения компонентов вектора параметров можно найти в результате сложения $\hat{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}$. Вычисление (2.1.7) выполняется итерационно для одной реализации данных s_k, или с использованием рекуррентной процедуры, осуществляемой при поступлении новых отсчетов данных. Расчет матрицы чувствительности $\mathbf{D}$ должен выполняться на каждом шаге обработки, при этом каждый раз требуется использовать решение (2.1.4).

Следует подчеркнуть, что разложение (2.1.6) непосредственно применимо только для детерминированных сигналов. Если компоненты вектора параметров $\boldsymbol{\theta}$ являются случайными величинами, операции дифференцирования в (2.1.6) могут быть некорректными, поэтому необходимо рассматривать стохастические информационные сигналы, описываемые нелинейными стохастическими дифференциальными уравнениями Ито-Стратоновича, которые определяют процесс эволюции плотности вероятности $p(\boldsymbol{\theta})$. Зная плотность вероятности, несложно рассчитать значения вектора параметров на каждом шаге обработки.

Отдельной задачей является исследование влияния и редукция шума наблюдений, при котором измеренные значения s_k в (2.1.5) отличаются от истинных, что может нарушить стабильность процесса вычислений.

Реализация процедур итерационной и рекуррентной обработки требует обеспечения устойчивости решения при минимальном времени вычислительных операций. Для этого следует использовать корректные методы предварительной обработки и анализа данных с получением надежной априорной информации в рамках адекватной модели исследуемого процесса.

Системные преобразования сигналов

В зависимости от вида и объема априорной информации о характеристиках полезного сигнала и помех можно выделить три основных подхода к синтезу методов и алгоритмов обработки данных.

При первом подходе предполагается детерминированный характер значений сигнала $s(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\theta}')$, где $\mathbf{x}$ – вектор независимых переменных, $\boldsymbol{\theta}$ – вектор искомых детерминированных параметров, $\boldsymbol{\theta}'$ – вектор сопутствующих детерминированных параметров. Помеха n считается

аддитивной и имеющей нулевое среднее значение на интервале измерения. Операторное уравнение системы в этом случае имеет вид

$$\hat{\boldsymbol{\theta}} = \mathrm{T}\{\boldsymbol{\theta}_a, s(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\theta}') + n(\mathbf{x})\} = a\mathrm{T}\{\boldsymbol{\theta}_a, s(\mathbf{x},\boldsymbol{\theta})\}, \quad (2.1.8)$$

где $\boldsymbol{\theta}_a$ – вектор параметров линейного усреднения сигнала s, a – собственные значения оператора $\mathrm{T}\{.\}$, соответствующие гармоническим составляющим интерференционного сигнала. Следовательно, задача сводится к синтезу сравнительно простого оператора обработки T, инвариантного к вектору параметров $\boldsymbol{\theta}'$ (см. рис. 2.2).

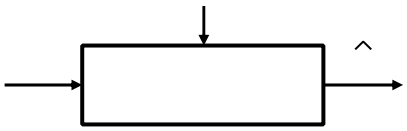


Рис. 2.2. Схема системного преобразования сигнала

Второй подход (см. рис. 2.3) основывается на предположении о сигнале как реализации случайного процесса $\{s(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\theta}')\}$ с известной априорной плотностью вероятности $p(\boldsymbol{\theta},\boldsymbol{\theta}')$. Необходимо с допустимой погрешностью получить оценку

$$\hat{\boldsymbol{\theta}} = \mathrm{T}\{s(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\theta}') + n(\mathbf{x})\}, \quad (2.1.9)$$

при обработке зарегистрированной реализации случайного процесса на фоне шума наблюдений $n(\mathbf{x})$ с известными статистическими характеристиками. Синтез требуемого оператора обработки в (2.1.9) может быть выполнен на основе достаточно общих критериев, например, критерия минимума среднего риска, правила максимума функционала правдоподобия или критериев, асимптотически приближающихся по характеристикам к критерию максимума апостериорной плотности вероятности.

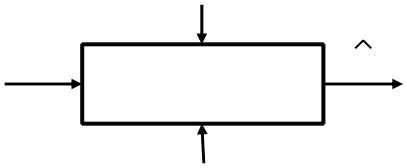


Рис. 2.3. Схема системного преобразования сигнала

Третий подход (см. рис. 2.4) к обработке сигналов применяется в тех случаях, когда априорно неизвестны даже статистические характеристики сигнала и помехи. В этом случае необходимо использовать принципы адаптации измерительных систем. Согласно этому подходу, в процессе обработки по зарегистрированным данным должны быть найдены параметры, характеризующие реализацию $s(\mathbf{x},\boldsymbol{\theta})$, причем параметры заменяются их оценочными значениями, то есть

$$\hat{\boldsymbol{\theta}} = \mathrm{T}_{pr}\{\mathrm{T}\{s(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\theta}') + n(\mathbf{x})\}\} = \mathrm{T}_0\{s(\mathbf{x},\boldsymbol{\theta},\hat{\boldsymbol{\theta}}') + \hat{n}(\mathbf{x})\}, \quad (2.1.10)$$

где $\hat{\boldsymbol{\theta}}'$ и $\hat{n}(\mathbf{x})$ – оценки вектора сопутствующих параметров сигнала и шума, соответственно, получаемые на стадии предварительной обработки, определяемой оператором $\mathrm{T}_{pr}\{.\}$.

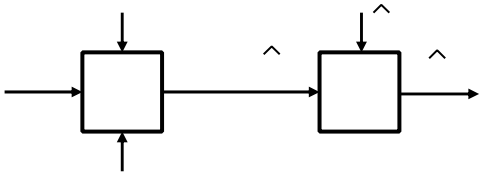


Рис. 2.4. Схема системного преобразования сигнала

Во всех случаях достигаемая точность и сложность обработки в значительной мере определяются выбором адекватных моделей сигналов и помех в (2.1.8) – (2.1.10), то есть решением задачи идентификации в широком смысле, которое возможно на основе рассмотрения физических принципов и особенностей построения систем компьютерной фотоники.

Синтез оператора обработки T может осуществляться в частотной области или в области независимых переменных.

Линейные системы в оптике

Первоначально теория и техника обработки изображений основывалась на хорошо известных принципах анализа линейных систем. Изображение при этом рассматривалось как двумерный сигнал. На этой основе можно реализовать соответствующий подход к обработке изображений.

В качестве примера рассмотрим расфокусированную оптическую систему, схема которой показана на рис 2.5.

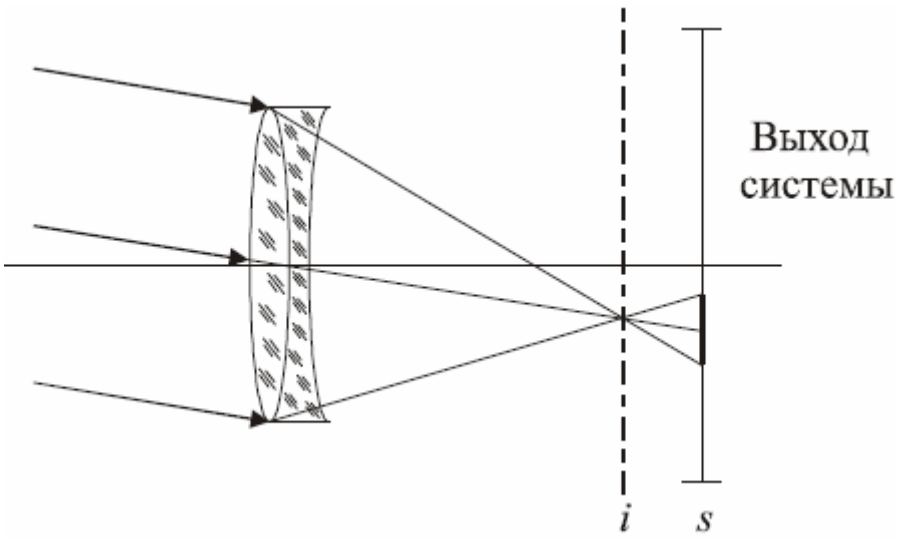


Рис. 2.5. Оптическая система формирования изображения

Изображение $g(x,y)$, полученное при помощи такой системы, можно считать преобразованным вариантом идеального изображения $f(x,y)$, получаемого в идеально сфокусированной системе. Если освещение изменится так, что яркость идеального изображения $f(x,y)$, то яркость расфокусированного изображения $g(x,y)$ также удвоится. Далее, если несколько сместить оптическую систему так, что идеальное изображение несколько сместится в плоскости изображения, то расфокусированное изображение также сместится подобным образом. Поэтому принято говорить, что переход от идеального изображения к расфокусированному является линейной пространственно-инвариантной операцией.

Система формирования изображений называется линейной, если выполняется принцип суперпозиции, который можно выразить в форме

$$\alpha f_1(x,y) + \beta f_2(x,y) \to [Система] \to \alpha g_1(x,y) + \beta g_2(x,y), \qquad (2.1.11)$$

для произвольных α и β.

Система называется пространственно-инвариантной, если для произвольных a и b реакция системы на смещенный входной сигнал $f(x-a,y-b)$ представляет собой смещенный сигнал $g(x-a,y-b)$:

$$f(x-a,y-b) \to [Система] \to g(x-a,y-b)). \qquad (2.1.12)$$

На практике изображения ограничены по размерам, поэтому пространственная инвариантность соблюдается только для ограниченных сдвигов a и b.

Преобразование исходного изображения линейной пространственно-инвариантной системой определяется операцией свертки

$$g(x,y) = \iint_R f(x-\xi, y-\eta) h(\xi,\eta) d\xi d\eta, \qquad (2.1.13)$$

где $h(\xi,\eta)$ – реакция системы на единичный импульс, представляемый в виде дельта-функции $\delta(x,y)$

$$h(x,y) = \iint_R \delta(x-\xi, y-\eta) h(\xi,\eta) d\xi d\eta. \qquad (2.1.14)$$

Поскольку интеграл от функции $\delta(x,y)$ представляет собой единичный импульс в точке с координатами $x=0$, $y=0$, то функция $h(x,y)$ показывает, как реальная система размывает точку. Поэтому $h(x,y)$ называют функцией рассеяния точки.

Известно, что при переходе к частотному представлению операция свертки сводится к перемножению спектров. Рассмотрим понятие частоты, применительно к двумерным сигналам. Пусть входной сигнал системы определяется так называемой пространственной гармонической функцией

$$f(x,y) = \cos 2\pi(u_0 x + v_0 y). \qquad (2.1.15)$$

где u_0, v_0 – пространственные частоты, имеющие размерность $[u_0] = [v_0] = \text{м}^{-1}$.

На рис. 2.6 показаны линии максимумов (сплошные) и минимумов (пунктирные) функции $f(x,y)$, которые соответствуют уравнению

$$2\pi(u_0 x + v_0 y) = \pi n\,,\ n = 0,\pm1,\pm2.... \qquad (2.1.16)$$

Период Λ пространственной гармоники в направлении φ, перпендикулярном линиям экстремумов, очевидно, равен.

$$\Lambda = 1/\sqrt{u_0^2 + v_0^2}\,. \qquad (2.1.17)$$

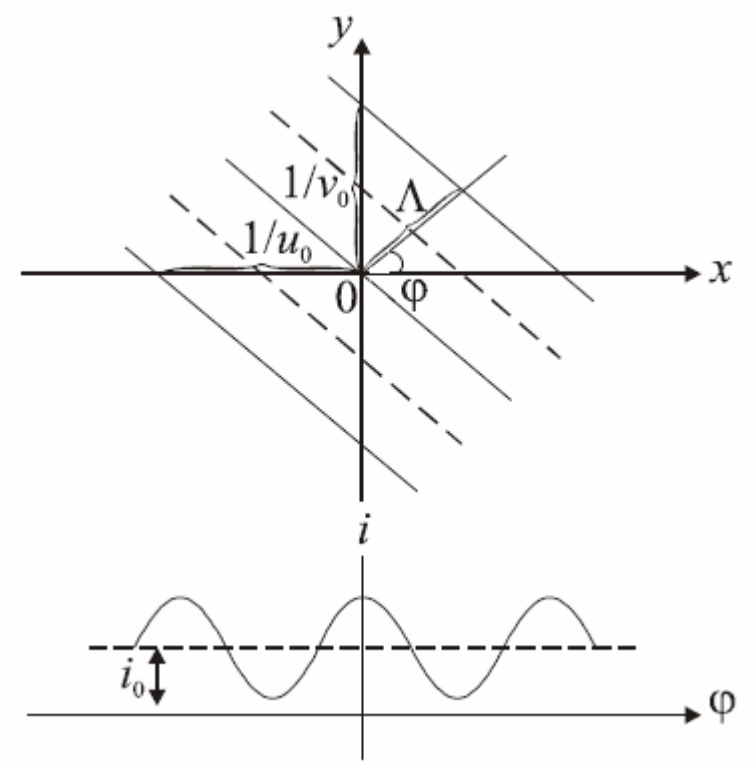


Рис. 2.6. Вид пространственной гармонической функции

Ориентация линий экстремумов определяется формулой

$$\varphi = \frac{\pi}{2} - \operatorname{arctg}\frac{u_0}{v_0}. \qquad (2.1.18)$$

Поскольку значения яркости не могут быть отрицательными, всегда имеется фоновая составляющая f_0.

Для того, чтобы учесть возможный фазовый сдвиг пространственного гармонического сигнала вводится комплексная функция

$$f_0(x,y) = \exp[j2\pi(ux+vy)] = \cos 2\pi(ux+vy) + j\sin 2\pi(ux+vy). \qquad (2.1.19)$$

На выходе линейной пространственно-инвариантной системы можно получить сигнал

$$f(x,y) = \iint_R \exp\{j2\pi[u(x-\xi)+v(y-\eta)]\}h(\xi,\eta)d\xi d\eta =$$

$$= \exp[j2\pi(ux+vy)]\iint_R \exp[-j2\pi(u\xi+v\eta)]h(\xi,\eta)d\xi d\eta =$$

$$= \exp[j2\pi(ux+vy)]H(u,v), \qquad (2.1.20)$$

где функция $H(u,v) = \mathrm{FT}\{h(x,y)\}$ носит название оптической передаточной функции системы. Последнее выражение показывает, что пространственная гармоническая функция является собственной функцией операции свертки в двумерной системе. Операция преобразования такой функции сводится к умножению на масштабный множитель $|H(u,v)|$:

$$\exp[j2\pi(ux+vy)] \to [Система] \to H(u,v)\exp[j2\pi(ux+vy)]. \qquad (2.1.21)$$

Это подтверждает целесообразность представления изображения в виде пространственных гармонических составляющих с различными амплитудами, частотами (ориентацией) и фазами.

Расфокусировка изображения может иметь различный характер, однако часто такой вид искажений нормированной моделируется гауссовой функцией рассеяния точки

$$h(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \qquad (2.1.22)$$

Эта функция обладает круговой симметрией, поскольку зависит от $r^2 = x^2 + y^2$. Оптическая передаточная функция в этом случае разделяется по пространственным переменным с точностью до множителя $1/4\pi^2$ равна

$$H(u,v) = \iint_R \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)\exp[-j2\pi(ux+vy)]dxdy =$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp(-j2\pi ux)dx \times$$

$$\times \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{y^2}{2\sigma^2}\right)\exp(-j2\pi vy)dy. \qquad (2.1.23)$$

Поскольку справедливо соотношение

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp(-j2\pi ux)dx = \sigma\exp\left(\frac{-u^2\sigma^2}{2}\right), \qquad (2.1.24)$$

можно показать, что передаточная функция $H(u,v)$ также обладает круговой симметрией

$$H(u,v) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{u^2+v^2}{2\sigma^2}\right). \qquad (2.1.25)$$

Следовательно, составляющие изображения с низкими пространственными частотами $u,v \to 0$ передаются при расфокусировке без изменения, тогда как высочкочастотные составляющие заметно ослабляются, начиная с частот $\approx 1/\sigma$. Однако σ – это характерный размер функции рассеяния точки. Поэтому, чем больше величина размывания, тем

более низкими становятся частоты подавления составляющих изображения.

Другим видом искажения изображений являются искажения, вызванные движением. При этом точка смазывается и превращается в черту. Пусть смещение изображения происходит вдоль оси x и составляет $2l$. В этом случае функция рассеяния точки имеет вид произведения

$$h_x(x,y) = [\beta(x+l) - \beta(x-l)]\delta(y)/2l, \qquad (2.1.26)$$

где $\beta(z)$ – единичная ступенчатая функция (рис. 2.7).

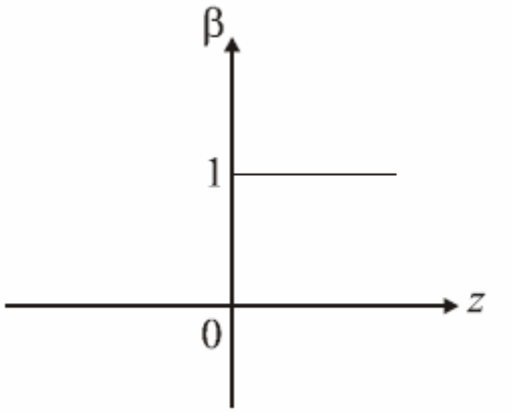


Рис. 2.7. Единичная ступенчатая функция

Передаточная функция искажающей системы в этом случае определяется выражением

$$H(u,v) = \int_{-\infty}^{\infty} \frac{1}{2l}[\beta(x+l) - \beta(x-l)]\exp(-j2\pi ux)dx \int_{-\infty}^{\infty} \delta(y)\exp(-j2\pi vy)dy =$$

$$= \frac{1}{2l}\int_{-l}^{l} \exp(-j2\pi ux)dx = \operatorname{sinc}(ul). \qquad (2.1.27)$$

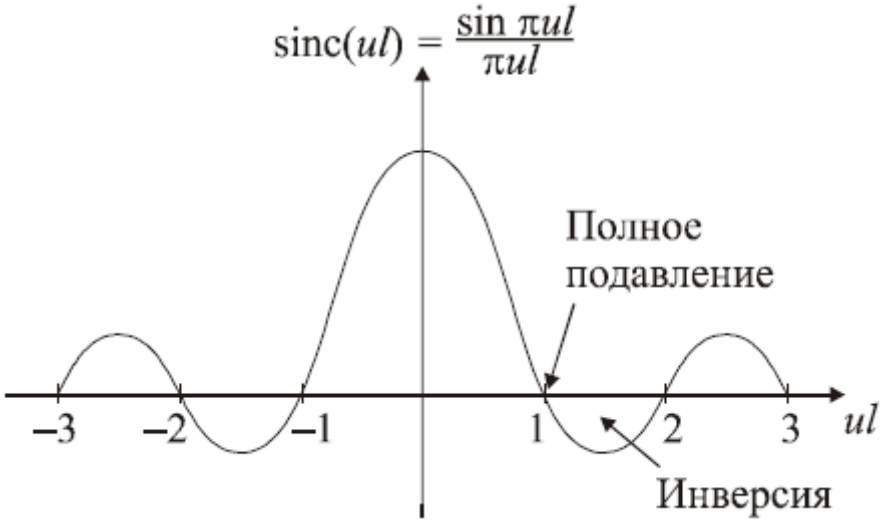


Рис. 2.8. Частотная характеристика системы в условиях искажений, вызванных движением

Следовательно, составляющие с низкими частотами вновь не изменяются, а с высокими – гасятся (см. рис. 2.8). В точках $ul = n$, $n = 0,\pm1,\pm2...$ происходит полное подавление. На частотных интервалах, соответствующих отрезкам между нечетными и четными $|n|$, составляющие спектра изображения инвертируются.

2.2 Модели сигналов и систем

Обработка сигналов в системах компьютерной фотоники основывается на теоретических исследованиях, результаты которых выражаются математическими моделями сигналов и шумов. Получение предварительных знаний в форме математических моделей носит название идентификации (отождествления) в широком смысле.

Во многих случаях необходимо получать оценки параметров модели применительно к конкретному исследуемому процессу, что обеспечивается решением задачи идентификации в узком смысле.

Принято различать детерминированные и стохастические модели в зависимости от соответствующих характеристик исследуемого процесса.

При динамическом оценивании параметров сигналов осуществляется оперативная идентификация в узком смысле.

В настоящем разделе рассматриваются методы анализа и идентификации математических моделей с точки зрения их сравнения с данными и с целью получения моделей практически обоснованного качества. Общим принципы моделирования применимы к широкому кругу задач, в том числе синтеза и анализа сигналов в информационных системах.

Моделью называется компактное и количественно верное представление опытных данных. Представление данных не всегда является однозначным, поэтому включение ряда моделей в состав программного комплекса приводит к необходимости оценки и качества и сравнения моделей между собой. Основную роль вне зависимости от целей моделирования играет точность опытных и расчетных данных как критерий качества модели. Отсюда следует необходимость разумного соотношения точности теоретических расчетов и точности экспериментальных данных.

Идентификация в широком смысле необходима в случаях, когда априорная информация отсутствует или является очень бедной, поэтому приходится решать задачи выбора структуры системы и задания класса моделей, определения степени линейности и стационарности процесса, определения вида влияния входных переменных на выходные, выбора информативных переменных и т.д. Методы идентификации в широком

смысле стали разрабатываться лишь в последние годы при использовании возрастающих вычислительных возможностей компьютерной техники.

Идентификация в узком смысле состоит в оценивании параметров модели по данным наблюдения. Эта задача является основной применительно к проблематике компьютерной фотоники. Правильный выбор модели имеет исключительное значение при параметрическом описании системы, поскольку в модели в большей степени, чем в непараметрических методах описания сосредоточено «концентрированное априорное знание» об исследуемом процессе. В результате системы, построенные на основе параметрических моделей, могут рассматриваться как более специализированные и, следовательно, имеющие преимущество в разрешающей способности, быстродействии, простоте структуры и т.д.

Качество математических моделей и критерии идентификации

Качество модели наиболее последовательно и строго оценивается методами математической статистики. Основная методика состоит в следующем.

Данные представляются суммой регулярной составляющей процесса и случайной помехи (шума) с некоторой функцией распределения вероятности, а именно

$$\boldsymbol{\xi} = \mathbf{s}(x, \boldsymbol{\theta}) + \mathbf{n}\,,\ \mathbf{n} \in \mathbf{F}(\boldsymbol{\theta}_n)\,, \qquad (2.2.1)$$

где $\boldsymbol{\xi}$ – вектор измеренных выходных переменных (значений реального интерференционного сигнала), $\mathbf{s}(x, \boldsymbol{\theta})$ – векторная функция, представляющая идеализированную модель, $\mathbf{n}$ – реализация случайного вектора шума, $\mathbf{F}(\boldsymbol{\theta}_n)$ – векторная функция многомерного распределения шума с параметрами $\boldsymbol{\theta}_n$, x – независимая переменная, $\boldsymbol{\theta}$ – вектор параметров модели. На основе предварительных исследований устанавливается вектор параметров шума $\boldsymbol{\theta}_n$ и вид функции $\mathbf{F}(\boldsymbol{\theta}_n)$.

Исходя из требований несмещенности, эффективности, состоятельности, формируется критерий Q идентификации модели, являющейся для конкретной выборки опытных данных функцией параметров модели $\boldsymbol{\theta}$. Минимизация (максимизация) критерия Q приводит к получению оценки $\hat{\boldsymbol{\theta}}$ вектора параметров, наиболее приближающей расчетные значения модели к данным.

Проверяется статистическая гипотеза о соответствии полученных отклонений модели от экспериментальных данных выбранной функции $\mathbf{F}(\boldsymbol{\theta}_n)$. Если эта гипотеза не отвергается, то говорят об адекватности модели. При этом нужно учитывать следующие особенности:

- при выборе функции $\mathbf{F}(\boldsymbol{\theta}_n)$ в общем случае неизвестна полезная составляющая, поскольку полезный сигнал и шум раздельно наблюдаемы;
- проводится замена вопроса об адекватности теоретической модели вопросом об адекватности модели эксперимента.

Можно сделать следующий вывод: для преодоления упомянутых сложностей целесообразно рассматривать более тонкую структуру разности расчетных и зарегистрированных значений сигнала, так как именно эта разность является мерой сходства оригинала и модели.

Точность определения параметра обычно принято оценивать значением среднего квадратичного отклонения. Пусть ряд наблюдений есть $\boldsymbol{\xi} = (\xi_1,...,\xi_K)^{\mathrm{T}}$. Предположим, что на основе указанного ряда наблюдений должен быть вычислен скалярный параметр *s*. Ясно, что оценка по методу наименьших квадратов в общем случае отличается, например, от оценки по критерию минимума модуля отклонений. Поэтому важно определить, какова в принципе достигаемая точность для ряда наблюдений $\boldsymbol{\xi}$. Можно показать, что ряд наблюдений $\boldsymbol{\xi}$ характеризуется так называемой информацией в определении Фишера, а именно, дисперсия оценки, соответствующая выбранному алгоритму оценивания $Z\{\xi\}$ параметра *s*, не может быть меньше, чем обратная величина информации Фишера. Как уже отмечалось выше, эта величина носит название границы Крамера-Рао. Важное свойство критерия состоит в том, что граница определяется только функцией плотности вероятности наблюдений и не зависит от алгоритма. Поэтому отношение квадратного корня из значения границы Крамера-Рао к среднему квадратическому отклонению, обеспечиваемому выбранным алгоритмом оценивания, является критерием его эффективности.

Проанализируем структуру разности расчетных и зарегистрированных интерферометрических данных, которая на практике является векторной величиной $\delta\mathbf{s}$, заданной в точках измерения.

Для отдельного опыта можно записать

$$\delta\mathbf{s} = \boldsymbol{\xi} - \mathbf{s}(x,\boldsymbol{\theta}), \qquad (2.2.2)$$

где $\boldsymbol{\xi}$ – вектор измеренных значений размера $K \times 1$, $\mathbf{s}$ – векторная функция расчетных значений модели размерности *K*, $\boldsymbol{\theta}$ – вектор уточняемых в эксперименте параметров модели размера $M \times 1$. Разность между значениями в идеализированной модели $\mathbf{s}(x)$ и значениями в реальной модели $\mathbf{s}_e(x,\boldsymbol{\theta})$ обозначим как

$$\delta\mathbf{s}_0 = \mathbf{s}(x) - \mathbf{s}_e(x,\boldsymbol{\theta}), \qquad (2.2.3)$$

В свою очередь

$$\delta\mathbf{s}_m = \boldsymbol{\xi} - \mathbf{s}(x), \qquad (2.2.4)$$

является погрешностью измерения. Отсюда следует

$$\delta\mathbf{s} = \delta\mathbf{s}_0 - \delta\mathbf{s}_m, \quad (2.2.5)$$

В качестве характеристики отклонения реальной модели с параметрами $\boldsymbol{\theta}_e$ от истинной можно использовать разность

$$\delta\mathbf{s}_M = \mathbf{s}(x,\boldsymbol{\theta}) - \mathbf{s}(x,\boldsymbol{\theta}_e), \quad (2.2.6)$$

где … – погрешность моделирования. Ее выделение из $\delta\mathbf{s}$ приводит к появлению еще одной составляющей, $\delta\mathbf{s}$:

$$\delta\mathbf{s}_e = \mathbf{s}(x,\boldsymbol{\theta}) - \mathbf{s}_e(x,\boldsymbol{\theta}_e), \quad (2.2.7)$$

Вектор $\mathbf{s}_e(x,\boldsymbol{\theta}_e)$ в (2.2.6) следует интерпретировать как истинные значения выходных переменных процесса, которые имели бы место, если в реальной модели на самом деле были бы точно идентифицированы все истинные компоненты вектора параметров. Таким образом, $\delta\mathbf{s}_M$ характеризует реальную модель процесса, а $\delta\mathbf{s}_e$ – качество проведения эксперимента. Повышение сложности модели ведет к уменьшению $\delta\mathbf{s}_M$ и увеличению $\delta\mathbf{s}_e$, поскольку в эксперименте требуется определять большее число величин. Увеличение состава контролируемых переменных повышает достигаемую точность, то есть повышение качества эксперимента снижает $\delta\mathbf{s}_e$. Следовательно, качество эксперимента можно оптимизировать, варьируя сложность модели.

Выделение $\delta\mathbf{s}_M$ и $\delta\mathbf{s}_e$ в общей погрешности $\delta\mathbf{s}$ приводит к ее разложению на три составляющих, связанных с качеством модели, качеством постановки конкретного эксперимента и точностью измерения

$$\delta\mathbf{s} = \delta\mathbf{s}_M + \delta\mathbf{s}_e + \delta\mathbf{s}_m. \quad (2.2.8)$$

Погрешность $\delta\mathbf{s}_M$ может носить систематический и случайных характер. Источниками $\delta\mathbf{s}_M$ могут быть, в частности, неадекватная формализация задачи, неучет отдельных влияющих факторов, погрешность компьютерной обработки данных. Источники $\delta\mathbf{s}_e$ характеризуются отклонениями учитываемых в реальной модели параметров $\boldsymbol{\theta}_e$ относительно идеализированного вектора $\boldsymbol{\theta}$. Источники погрешности и методы определения составляющей $\delta\mathbf{s}_m$ являются предметом теории измерений. Как правило, получении оценки $\delta\mathbf{s}_m$ возможно независимо от других составляющих $\delta\mathbf{s}$.

2.3 Структура двумерных дискретных сигналов, многомерные и векторные представления сигналов

На практике при решении задач обработки информации в компьютерной фотонике приходится иметь дело с большим количеством величин, в том числе случайных. Возможности удобного представления и

обработки этих данных дает математический аппарат векторного и матричного представления.

Применительно к фотонике, получение дискретных представление оптических полей связано с выполнением операций взятия дискретных отсчетов в двумерной (пространственной) или в трехмерной (пространственно-временной) области. Особенности операций дискретизации рассматриваются в других учебных дисциплинах (прежде всего, в теории информационных систем, систем обработки изображений [3.1]). Настоящая глава посвящена более общему рассмотрению структуры многомерных дискретных сигналов.

Двумерный дискретный сигнал можно представить [3.2] значениями функции, определенной на совокупности упорядоченных пар целых чисел

$$s = \{s(n_1, n_2), -\infty < n_1, n_2 < \infty\}. \quad (2.3.1)$$

В соответствии с таким определением двумерные последовательности имеют бесконечную протяженность, однако на практике эти последовательности являются финитными, поскольку значения отсчетов сигнала известны только в конечной области плоскости (n_1, n_2).

Свойства двумерных последовательностей в некоторых отношениях существенно отличаются от свойств одномерных.

Элементарные двумерные последовательности

Выделим некоторые виды последовательностей, которые имеют особенно важное значение для задач обработки изображений.

Единичный импульс (рис. 2.10, а) определяется следующим образом:

$$\delta(n_1, n_2) = \begin{cases} 1, n_1 = n_2 = 0, \\ 0, n_1, n_2 \neq 0. \end{cases} \quad (2.3.2)$$

Двумерный линейный импульс (рис. 2.10, в) представляет собой последовательности вида

$$s_x(n_1, n_2) = \delta(n_1), \quad (2.3.3)$$

$$s_y(n_1, n_2) = \delta(n_2). \quad (2.3.4)$$

Двумерная единичная ступенчатая функция (рис. 2.10, б) определяется как

$$\sigma(n_1, n_2) = \begin{cases} 1, n_1 \geq 0, n_2 \geq 0, \\ 0, n_1, n_2 < 0. \end{cases} \quad (2.3.5)$$

Эту функцию можно рассматривать также в форме произведения

$$\sigma(n_1, n_2) = \sigma(n_1)\sigma(n_2), \quad (2.3.6)$$

в котором

$$\sigma(n) = \begin{cases} 1, n \geq 0, \\ 0, n < 0. \end{cases} \quad (2.3.7)$$

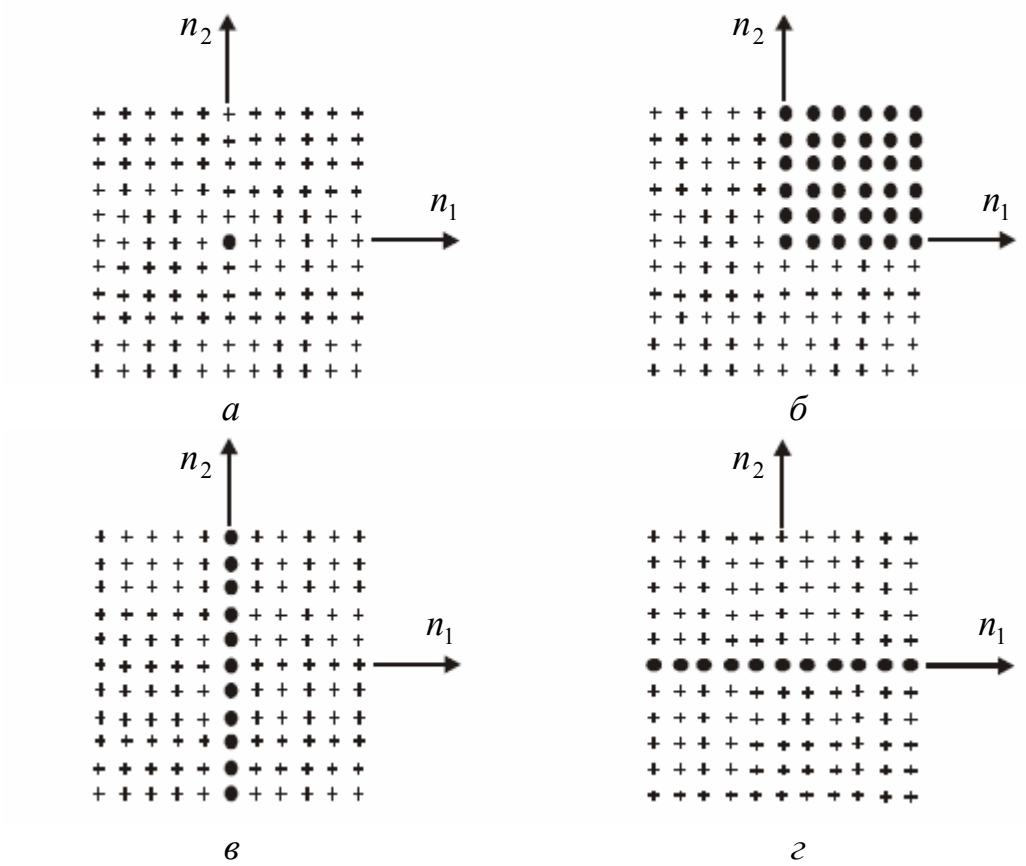


Рис. 2.10. Элементарные двумерные последовательности: двумерная единичная импульсная функция (а), двумерная единичная ступенчатая функция (б)
и примеры двумерных дискретных импульсов (в-г)

Экспоненциальные последовательности определяются следующим образом

$$s(n_1,n_2)=a^{n1}b^{n2},-\infty<n_1,n_2<\infty, \qquad (2.3.8)$$

где a и b являются комплексными числами. Если $|a|=|b|=1$, то можно записать

$$a=\exp(j2\pi u_1), \qquad (2.3.9)$$

$$b=\exp(j2\pi u_2), \qquad (2.3.10)$$

где u_1,u_2 – пространственные частоты. В этом случае экспоненциальная последовательность представляет собой комплексную синусоидальную последовательность

$$s(n_1,n_2)=\exp[j2\pi(u_1n_1+u_2n_2)]=$$
$$=\cos[2\pi(u_1n_1+u_2n_2)]+j\sin[2\pi(u_1n_1+u_2n_2)]. \qquad (2.3.11)$$

Экспоненциальные двумерные последовательности представляют особый интерес, так как они являются собственными функциями двумерных пространственно-инвариантных линейных систем.

Рассмотренные выше элементарные последовательности можно представить в виде

$$s(n_1, n_2) = s_1(n_1) s_2(n_2). \tag{2.3.12}$$

Любую последовательность, которую можно представить в форме произведения одномерных последовательностей, называют разделимой.

На практике лишь немногие двумерные последовательности являются разделимыми. Однако любое двумерное множество с конечным числом ненулевых отсчетов можно записать в виде суммы конечного числа разделимых последовательностей, а именно

$$s(n_1, n_2) = \sum_{p=1}^{P} s_{p1}(n_1) s_{p2}(n_2). \tag{2.3.13}$$

где P – число ненулевых строк или столбцов. Простейший пример имеет вид

$$\begin{cases} s_{p1}(n_1) = s(n_1, p), \\ s_{p2}(n_2) = \delta(n_2 - p). \end{cases} \tag{2.3.14}$$

Эти уравнения представляют собой суммы отдельных строк двумерной последовательности.

Реальные последовательности данных, представляющие изображения, соответствуют сигналам, значения которых являются равными нулю вне области конечной протяженности в (n_1, n_2)-плоскости. Эту область иногда называют опорной областью двумерного сигнала.

На рис 2.11 показан пример финитной последовательности вида

$$s(n_1, n_2) = \begin{cases} s(n_1, n_2), & n_1 \in [0, N_1], \quad n_2 \in [0, N_2], \\ 0, & n_1 \notin [0, N_1], \quad n_2 \notin [0, N_2]. \end{cases} \tag{2.3.15}$$

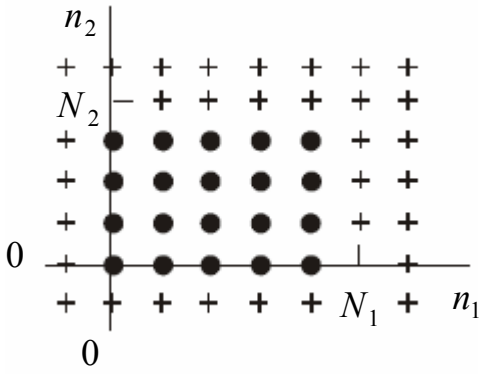


Рис. 2.11. Двумерная последовательность конечной протяженности с опорной областью прямоугольной формы

Периодические последовательности

Двумерная периодическая последовательность представляет собой сигнал, регулярно повторяющийся в пространстве. Формальное определение периодической двумерной последовательности является более сложным, чем одномерной последовательности.

Рассмотрим двумерную последовательность $\tilde{s}(n_1, n_2)$, удовлетворяющую условиям

$$\tilde{s}(n_1, n_2 + N_2) = \tilde{s}(n_1, n_2), \tag{2.3.15}$$

$$\tilde{s}(n_1 + N_1, n_2) = \tilde{s}(n_1, n_2). \tag{2.3.16}$$

Эта последовательность обладает двойной периодичностью; ее значения повторяются если переменная n_1 увеличивается на N_1 или если переменная n_2 увеличивается на N_2.

На рис. 3.3 приведено изображение такой последовательности. Значения N_1 и N_2, представляющие собой минимальные положительные целые числа, для которых справедливы выражения (2.3.15) и (2.3.16) называются горизонтальным и вертикальным интервалами периодичности последовательности $\tilde{s}(n_1, n_2)$.

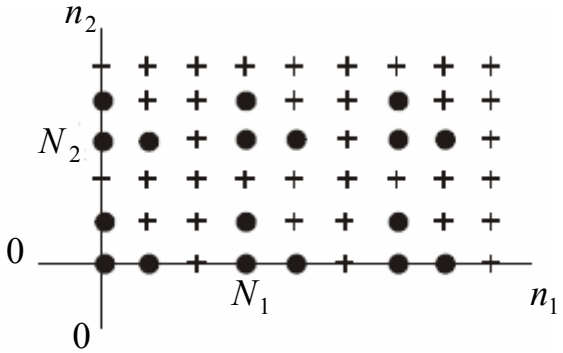


Рис. 2.12. Двумерная периодическая последовательность с $N_1 = N_2 = 3$

Из всех отсчетов последовательности только $N_1 N_2$ отсчетов оказываются независимыми. Остальные же отсчеты определяются в соответствии с условиями периодичности (2.3.17)–(2.3.18). Периодом последовательности $\tilde{s}(n_1, n_2)$ называется любая связная область плоскости (n_1, n_2), содержащая $N_1 N_2$ отсчетов, если значения этих отсчетов независимы. Часто наиболее удобной формой определения периода является прямоугольник со сторонами N_1 и N_2, однако, это не единственная возможность.

Рассмотрим двумерную последовательность $\tilde{s}(n_1, n_2)$, которая удовлетворяет более общим условиям периодичности

$$\tilde{s}(n_1 + N_{11}, n_2 + N_{21}) = \tilde{s}(n_1, n_2), \tag{2.3.17}$$

$$\tilde{s}(n_1 + N_{12}, n_2 + N_{22}) = \tilde{s}(n_1, n_2). \tag{2.3.18}$$

причем

$$D = N_{11} N_{22} - N_{12} N_{21} \neq 0. \tag{2.3.19}$$

Упорядоченные пары $(N_{11}, N_{21})^{\mathrm{T}}$ и $(N_{12}, N_{22})^{\mathrm{T}}$ можно рассматривать как векторы $\mathbf{N}_1$ и $\mathbf{N}_2$, представляющие собой смещения от любого отсчета к соответствующим отсчетам двух других периодов (см. рис. 2.13). Один

период такой последовательность заключен в области, имеющей форму параллелограмма, смежные стороны которого образованы векторами $\mathbf{N}_1$ и $\mathbf{N}_2$.

Понятие периодичности легко обобщается на случай *M*-мерных сигналов. Для простоты обозначим через **n** упорядоченную группу из *M* целочисленных переменных $(n_1, n_2, ..., n_M)^{\mathrm{T}}$. Тогда $\tilde{s}(\mathbf{n})$ представляет собой *M*-мерную периодическую последовательность при условии, что существует *M* таких линейно независимых *M*-мерных целочисленных векторов $\mathbf{N}_1, \mathbf{N}_2, ..., \mathbf{N}_M$, что

$$\tilde{s}(\mathbf{n} + \mathbf{N}_i) = \tilde{s}(\mathbf{n}),\ i = 1, ..., M.. \qquad (2.3.20)$$

Векторы $\mathbf{N}_i$ называются векторами периодичности; их можно использовать в качестве столбцов матрицы **N** размерностью $M \times M$, называемой матрицей периодичности

$$\mathbf{N} = [\mathbf{N}_1 | \mathbf{N}_2 | ... | \mathbf{N}_M]. \qquad (2.3.21)$$

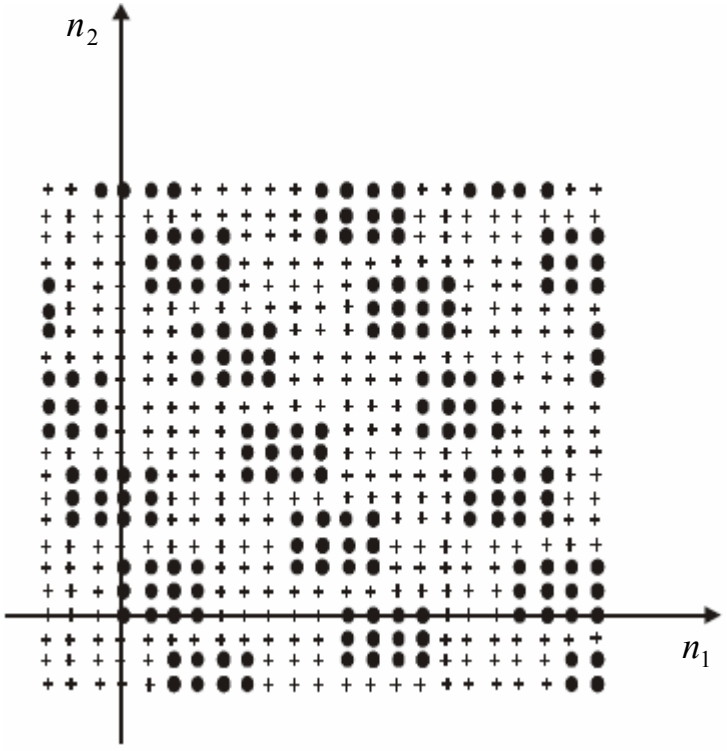


Рис. 2.13. Периодическая последовательность с векторами периодичности $(7,2)^{\mathrm{T}}$ и $(-2,4)^{\mathrm{T}}$

Требование линейной независимости векторов периодичности эквивалентно требованию наличия у матрицы **N** определителя. В частном случае, когда **N** – диагональная матрица, можно сказать, что последовательность $\tilde{s}(\mathbf{n})$ прямоугольно периодична. Выше рассмотрен именно этот случай.

Если $\tilde{s}(\mathbf{n})$ периодична с матрицей периодичности **N**, то для любого целочисленного вектора справедливо

$$\tilde{s}(\mathbf{N}+\mathbf{Nr})=\tilde{s}(\mathbf{n}). \qquad (2.3.22)$$

Отсюда следует, что если **P** – некоторая целочисленная матрица, то **NP** также будет матрицей периодичности для $\tilde{s}(\mathbf{n})$. Таким образом, любая периодическая последовательность имеет не единственную матрицу периодичности. Отметим, что абсолютное значение определителя матрицы периодичности дает число отсчетов последовательности $\tilde{s}(\mathbf{n})$, содержащееся в одном периоде. Это обстоятельство используется при построении *M*-мерного дискретного преобразования Фурье.

Основные операции над многомерными сигналами

Сигналы можно объединять или изменять с помощью множества операций. Пусть ω и *s* – двумерные дискретные сигналы. Эти сигналы можно сложить и получить третий сигнал ξ. Сложение выполняется поэлементно, так что значение каждого отсчета $\xi(n_1,n_2)$ получается путем сложения двух соответствующих отсчетов $\omega(n_1,n_2)$ и $s(n_1,n_2)$

$$\xi(n_1,n_2)=\omega(n_1,n_2)+s(n_1,n_2). \qquad (2.3.23)$$

Умножая двумерные последовательности на константу, можно также получить новую последовательность. Если *c* – константа, мы можем образовать двумерную последовательность ξ из скаляра *c* и двумерной последовательности *s*, умножив значение каждого отсчета на *c*

$$\xi(n_1,n_2)=cs(n_1,n_2). \qquad (2.3.24)$$

Двумерную последовательность s можно подвергнуть линейному сдвигу, что также приведет к образованию новой последовательности ξ. Операция сдвига переносит всю последовательность s на новый участок плоскости (n_1,n_2). Значения отсчетов ξ связаны в этом случае со значениями отсчетов s соотношением

$$\xi(n_1,n_2)=s(n_1-m_1,n_2-m_2), \qquad (2.3.25)$$

где (m_1,m_2) – величина сдвига.

Используя базовые операции сложения, скалярного умножения и сдвига (2.3.23)–(2.3.25), можно разложить любую двумерную последовательность на сумму взвешенных и сдвинутых двумерных единичных импульсов

$$s(n_1,n_2)=\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty}s(k_1,k_2)\delta(n_1-k_1,n_2-k_2), \qquad (2.3.26)$$

где $\delta(n_1-k_1,n_2-k_2)$ представляет собой единичный импульс, сдвинутый так, что его ненулевой отсчет находится в точке (k_1,k_2). Значения $s(k_1,k_2)$

можно рассматривать как скалярные множители для соответствующих единичных импульсов.

Стоит упомянуть еще о двух основных операциях над двумерными последовательностями. Одну из них, которую мы назовем пространственным маскированием, можно рассматривать, как обобщение скалярного умножения. Значение каждого отсчета двумерной последовательности s умножается на число $c(n_1,n_2)$, значение которого зависит от положения соответствующего отсчета

$$\xi(n_1,n_2)=c(n_1,n_2)s(n_1,n_2). \qquad (2.3.27)$$

Совокупность чисел $c(n_1,n_2)$ можно рассматривать как двумерную последовательность. Тогда правая часть равенства (2.3.27) представляет собой поэлементное произведение двух последовательностей.

Система обработки, в том числе двумерный фильтр, имеет реакцию на единичный импульс (2.3.2), определяемую в общем случае как

$$h_{k_1,k_2}(n_1,n_2)=\mathbf{T}\{\delta(n_1-k_1,n_2-k_2)\}, \qquad (2.3.28)$$

где T представляет собой оператор системы. Если система является линейной и инвариантной к сдвигу, то выходная последовательность связана с входной последовательностью следующим образом

$$\xi(n_1,n_2)=\sum_{k_1=-\infty}^{\infty}\sum_{k_1=-\infty}^{\infty}s(k_1,k_2)h(n_1-k_1,n_2-k_2). \qquad (2.3.29)$$

Это соотношение известно под названием двумерной дискретной свертки. В сущности здесь выполняется разложение входной последовательности $s(n_1,n_2)$ на взвешенную сумму сдвинутых импульсов в соответствии с равенством (2.3.26). Система преобразует каждый импульс в сдвинутую копию импульсного отклика $h(n_1,n_2)$. Суперпозиция этих взвешенных и сдвинутых импульсных откликов образует выходную последовательность, причем весовыми коэффициентами являются значения отсчетов входной последовательности $s(n_1,n_2)$. Равенство (3.15) записано в предположении, что фильтр полностью характеризуется своим импульсным откликом $h(n_1,n_2)$.

Выполнив замену переменных $n_1-k_1=l_1$ и $n_2-k_2=l_2$, равенство (2.3.29) можно переписать в форме

$$\xi(n_1,n_2)=\sum_{l_1=-\infty}^{\infty}\sum_{l_1=-\infty}^{\infty}h(l_1,l_2)s(n_1-l_1,n_2-l_2). \qquad (2.3.30)$$

Отсюда видно, что свертка – это коммутативная операция. Операция свертки обычно обозначается при помощи символа «*».

С помощью векторных обозначений выходную последовательность M-мерной системы можно представить как M-мерную свертку входной последовательности и импульсного отклика

$$\xi(\mathbf{n})=\sum_{k}s(\mathbf{k})h(\mathbf{n}-\mathbf{k}). \qquad (2.3.31)$$

Операция двумерной свертки видоизменяется в случаях, если импульсный отклик двумерного фильтра является разделимым. Разделимая входная последовательность представлена выше.

Пусть импульсный отклик системы можно представить следующим выражением

$$h(n_1,n_2)=h_1(n_1)h_2(n_2). \qquad (2.3.32)$$

Тогда выходной сигнал системы имеет вид

$$\xi(n_1,n_2)=\sum_{k_1=-\infty}^{\infty}\sum_{k2=-\infty}^{\infty}s(n_1-k_1,n_2-k_2)h_1(k_1)h_2(k_2)=$$

$$=\sum_{k_1=-\infty}^{\infty}h_1(k_1)\sum_{k_2=-\infty}^{\infty}s(n_1-k_1,n_2-k_2)h_2(k_2). \qquad (2.3.33)$$

Внутренняя сумма представляет собой двумерный массив. Если определить $g(n_1,n_2)$ в виде

$$g(n_1,n_2)=\sum_{k_2=-\infty}^{\infty}s(n_1,n_2-k_2)h_2(k_2), \qquad (2.3.34)$$

то можно переписать выражение (3.18) как

$$\xi(n_1,n_2)=\sum_{k_1=-\infty}^{\infty}h_1(k_1)g(n_1-k_1,n_2). \qquad (2.3.33)$$

Массив $g(n_1,n_2)$ можно вычислить, выполняя одномерную свертку каждого столбца $s(n_1=const)$ с одномерной последовательностью h_2. Тогда выходной массив ξ вычисляется путем свертки каждой строки $g(n_2=const)$ с одномерной последовательностью h_1. Можно поступить и наоборот, выполнив сначала свертку по строкам, а затем по столбцам. Выходной сигнал в любом случае не изменится.

M-мерный случай мало отличается от двумерного. Разделимая система и здесь может быть реализована с помощью одномерных сверток, однако число операций свертки быстро растет с увеличением размерности сигнала. Рассмотрим, например, *M*-мерную последовательность $s(n_1,n_2,...,n_M)$, определенную на гиперкубе размера $N_1\times N_2\times...\times N_M$. При свертке этого сигнала с разделимой последовательностью вида $h_1(n_1)h_2(n_2)...h_M(n_M)$ для получения выходной последовательность потребуется выполнить MN^{M-1} одномерных сверток.

2.4 Матричные представления в теории обработки данных

Векторное представление оказывается полезным при описании последовательностей данных, получаемых в процессе дискретизации непрерывных сигналов [3.3]. Пусть необходимо принимать во внимание конечное число точек дискретизации *K* в выборке данных. Тогда отдельное

значение в выборке может рассматривать как компонент вектора размера $K \times 1$.

Если точки дискретизации обозначить как $x_1, x_2, ..., x_K$, то вектор, представляющий значения функции **s**(x) можно представить в виду

$$\mathbf{s}(x) = \begin{bmatrix} s(x_1) \\ s(x_2) \\ \vdots \\ s(x_K) \end{bmatrix}. \tag{2.4.1}$$

Если выборочная функция $s(x)$ является случайной, то есть представляет собой реализацию случайного процесса, то каждый компонент вектора наблюдения **s** представляет собой случайную величину. Можно определить ковариационную матрицу размером $K \times K$, которая характеризует ковариацию между случайными величинами $s(x_i)$ и $s(x_j)$, $i, j = 1, ..., K$

$$\mathbf{R}_s = \mathrm{M}\{\mathbf{s}\mathbf{s}^{\mathrm{T}}\} = \mathrm{M}\begin{bmatrix} s(x_1)s(x_1) & s(x_1)s(x_2) & \cdots & s(x_1)s(x_K) \\ s(x_2)s(x_1) & s(x_2)s(x_2) & \cdots & s(x_2)s(x_K) \\ \vdots & \vdots & \ddots & \vdots \\ s(x_K)s(x_1) & s(x_K)s(x_2) & \cdots & s(x_K)s(x_K) \end{bmatrix}, \tag{2.4.2}$$

где M{.} представляет собой оператор взятия математического ожидания. После усреднения случайных элементов этой матрицы по ансамблю реализаций получаем значение автоковариационной матрицы $\mathbf{R}_s(x_i, x_j)$ случайного процесса $\{s(x)\}$

$$\mathbf{R}_s = \begin{bmatrix} R_s(x_1, x_1) & R_s(x_1, x_2) & \cdots & R_s(x_1, x_K) \\ R_s(x_2, x_1) & R_s(x_2, x_2) & \cdots & R_s(x_2, x_K) \\ \vdots & \vdots & \ddots & \vdots \\ R_s(x_K, x_1) & R_s(x_K, x_2) & \cdots & R_s(x_K, x_K) \end{bmatrix}, \tag{2.4.3}$$

где $R_s(x_i, x_j) = M\{s(x_i)s(x_j)\}$.

В случае стационарности процесса $\{s(x)\}$ все компоненты матрицы $\mathbf{R}_s$ становятся функциями интервала

$$\zeta_k = x_1(k-1)\Delta x, \tag{2.4.4}$$

где $\Delta x = x_{i+1} - x_i$ представляет собой шаг дискретизации, причем матрица $\mathbf{R}_s$ является симметричной, так как

$$\mathbf{R}_s(k\Delta x) = \mathbf{R}_s(-k\Delta x). \tag{2.4.5}$$

Во многих случаях используют понятие корреляционной матрицы

$$\mathbf{C}_s = \mathrm{M}\{(\mathbf{s} - \mathrm{M}\{\mathbf{s}\})(\mathbf{s} - \mathrm{M}\{\mathbf{s}\})^{\mathrm{T}}\}. \tag{2.4.6}$$

Корреляционная матрица связана с ковариационной матрицей соотношением

$$\mathbf{C}_s = \mathbf{R}_s - \mathrm{M}\{\mathbf{s}\}\mathrm{M}\{\mathbf{s}^{\mathrm{T}}\}. \tag{2.4.7}$$

Для центрированного случайного процесса $\mathbf{C}_s = \mathbf{R}_s$.

Для процесса, стационарного в широком смысле, дисперсии на любой диагонали одинаковы $\sigma_i^2 = \sigma_j^2 = \sigma_s^2$, $i, j = 1,...,K$. Если представить коэффициенты корреляции между каждой парой величин как $\rho_{ij} = \rho_{|i-j|}$, $i, j = 1,...,K$, то корреляционную матрицу можно записать в форме

$$\mathbf{C}_s = \sigma_s^2 \mathbf{T} = \sigma_s^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_K \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{K-1} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{K-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_K & \rho_{K-1} & \rho_{K-2} & \cdots & 1 \end{bmatrix}, \tag{2.4.8}$$

где **T** представляет собой матрицу Теплица, в которой главная диагональ заполнена единицами, а значения на диагоналях, параллельных ей, одинаковы. Такую матрицу можно умножить на вектор за $O(n \log n)$ операций. Матрица Теплица также используется для осуществления дискретного преобразования Фурье, так как такой матрицей можно представить оператор умножения на многочлен из синусов и косинусов, спроецированный на конечномерное пространство.

Векторная обработка данных

Использование векторных обозначений особенно удобно, когда случайные величины выбираются из различных случайных процессов

$$\mathbf{s}(x) = \begin{bmatrix} s_1(x) \\ s_2(x) \\ \vdots \\ s_K(x) \end{bmatrix}. \tag{2.4.9}$$

Ковариационная матрица в этом случае определяется в форме

$$\mathbf{R}_s(\zeta) = \mathrm{M}\left[\mathbf{s}(x)\mathbf{s}^{\mathrm{T}}(x+\zeta)\right] = \begin{bmatrix} R_{11}(\zeta) & R_{12}(\zeta) & \cdots & R_{1K}(\zeta) \\ R_{21}(\zeta) & R_{22}(\zeta) & \cdots & R_{2K}(\zeta) \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1}(\zeta) & R_{K2}(\zeta) & \cdots & R_{KK}(\zeta) \end{bmatrix}. \tag{2.4.10}$$

Здесь ковариационные функции зависят от непрерывного сдвига ζ, а не от числа шагов дискретизации, как в случае выборки из реализации одного

случайного процесса. При одном случайном процессе говорят о многомерной обработке, при нескольких – о многоканальной обработке.

Ковариационные матрицы играют важную роль при определении совместной плотности вероятности K случайных величин, принадлежащих Гауссовскому процессу

$$p(\mathbf{s}) = f[s(x_1), s(x_2), ..., s(x_K)] =$$

$$= \frac{1}{(2\pi)^{K/2}\left|\mathbf{C}_s\right|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{s}^{\mathrm{T}} - \langle \mathbf{s} \rangle^{\mathrm{T}})\mathbf{C}_s^{-1}(\mathbf{s} - \langle \mathbf{s} \rangle) \right). \qquad (2.4.11)$$

где $\left|\mathbf{C}_s\right|$ – определитель матрицы $\mathbf{C}_s$, а $\mathbf{C}_s^{-1}$ – обратная ей матрица.

Если выборочная функция $s(x)$ является случайной, то есть представляет

Математическое описание и линейные преобразования дискретных изображений.

Компьютерная обработка оптических полей осуществляется после преобразования значений интенсивности в матрицу отсчетов с положительными элементами. Дальнейшие преобразования матрицы отсчетов необходимы для выделения полезной информации. Использование методов матричной алгебры позволяет построить эффективные алгоритмы обработки. Как правило, компьютерная система разрабатывается для обработки не единственной матрицы отсчетов, а совокупности видеокадров. Поэтому наряду с подходом на основе модели детерминированного двумерного дискретного сигнала широко используется также статистический подход на основе многомерных случайных процессов.

Пусть $\mathbf{S}$ – матрица отсчетов исходного дискретного квантованного изображения размером $K \times K$. Обобщенное преобразование этой матрицы можно представить в виде

$$\mathbf{Y} = \mathbf{U}^{\mathrm{T}}\mathbf{S}\mathbf{V}, \qquad (2.4.12)$$

где Y – преобразованное изображение, $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K)$, $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K)$ – это так называемые унитарные матрицы, такие, что $\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$, $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$, где $\mathbf{I}$ – единичная матрица, $\mathbf{u}_k$, $\mathbf{v}_k$ – векторы, образованные из столбцов **U** и **V**, которые являются ортогональными.

Исходное изображение определяется обратным преобразованием

$$\mathbf{S} = (\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K)\mathbf{Y}(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_K)^{\mathbf{T}}. \qquad (2.4.13)$$

Представим матрицу **Y** в виде суммы матриц

$$\mathbf{Y} = \begin{pmatrix} s_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} + \begin{pmatrix} 0 & s_{12} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} + \ldots. \qquad (2.4.14)$$

При этом можно записать

$$\mathbf{Y} = \sum_{k=1}^{K} \sum_{m=1}^{K} s_{km} \mathbf{u}_k \mathbf{v}_m^{\mathrm{T}}. \qquad (2.4.15)$$

Произведение вида $\mathbf{u}\mathbf{v}^{\mathrm{T}}$ носит название внешнего произведения векторов **u** и **v** и представляет собой матрицу (в отличие от внутреннего – скалярного – произведения векторов). Следовательно, преобразованное изображение **Y** есть линейная комбинация внешних произведений базисных векторов, взятых с весами s_{km}.

Следует заметить, что базисные векторы матриц **U** и **V** для пространственно разделимых изображений могут выбираться из одних и тех же или различных базисов (например, **U** – из базиса гармонических функций, **V** – из базиса прямоугольных функций Уолша).

Разложение изображения (2.4.14)–(2.4.15), очевидно, осуществляется с использованием заранее выбранного базиса внешних произведений заданных векторов и в этом смысле не зависит от характеристик конкретного изображения.

Изображение можно преобразовать иным образом – путем разложения по сингулярным значениям (SVD), а именно:

$$\mathbf{Y} = \sum_{k=1}^{R} s_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}. \qquad (2.4.16)$$

если S – диагональная матрица коэффициентов s_k ранга R. В этом выражении $s_k \geq 0$ представляют собой квадратные корни из сингулярных значений $\mathbf{S}\mathbf{S}^{\mathrm{T}}$; $\mathbf{u}_k$ и $\mathbf{v}_k$ – это сингулярные векторы.

Как и в предыдущем случае, изображение раскладывает на ортогональные базисные изображения в виде матриц ранга 1, однако разложение по сингулярным значениям учитывает характеристики конкретного изображения – модели **S** и поэтому обеспечивает единственное разложение по оптимальному базису. Важным достоинством SVD-разложения является устойчивость этого преобразования по отношению к ошибкам вычислений.

Разложение изображения вида (2.4.14)–(2.4.15) или (2.4.16) иллюстрируется на рис. 2.14.

s_{11} s_{ij} s_K

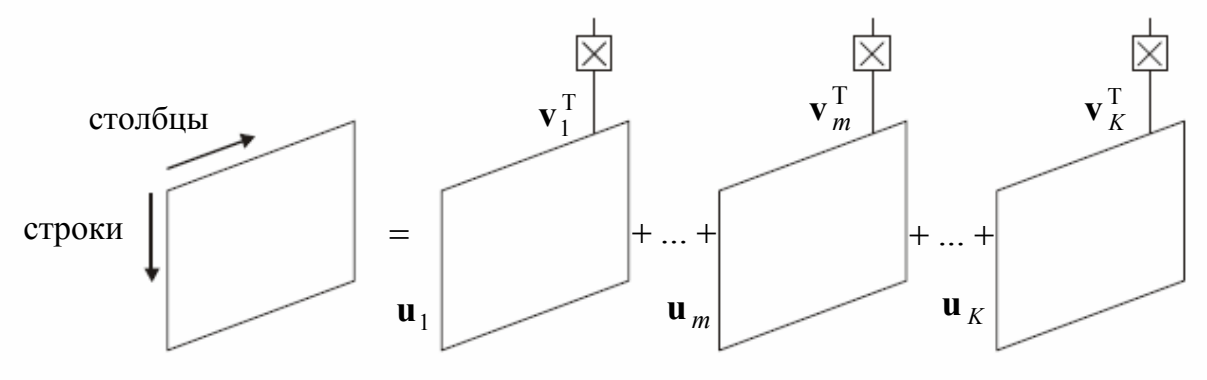


Рис. 2.14. Линейная декомпозиция изображения в ортогональном базисе «элементарных» изображений

Располагая веса в порядке убывания их значений, можно затем отбросить малые составляющие и закодировать изображение в виде усеченного набора значимых коэффициентов для известных ортогональных базисных изображений. Такой метод носит название кодирования посредством преобразований.

Заметим, что в ряде случаев используется разложение изображений по собственным значениям матрицы $\mathbf{S}\mathbf{S}^{\mathrm{T}}$. Если эта матрица симметрическая и квадратная, то ее собственные значения являются действительными числами.

Преобразование Карунена-Лоэва

Преобразование Карунена-Лоэва или разложение по собственным векторам используется для кодирования и анализа дискретных сигналов и изображений.

В общем виде преобразование Карунена-Лоэва записывается в форме

$$S(u,v) = \sum_{j=0}^{N-1}\sum_{k=0}^{N-1} F(j,k)A(j,k;u,v), \qquad (2.5.1)$$

где $A(j,k;u,v)$ – ядро преобразования, удовлетворяющее уравнению

$$\lambda(u,v)A(j,k;u,v) = \sum_{j'=0}^{N-1}\sum_{k'=0}^{N-1} K_F(j,k;j',k')A(j',k';u,v), \qquad (2.5.2)$$

где $K_F(j,k;j',k')$ – ковариационная функция дискретного изображения, а $\lambda(u,v)$ является постоянной при фиксированных u и v. Функции $A(j,k;u,v)$ являются собственными функциями ковариационной матрицы, а $\lambda(u,v)$ – ее собственными значениями. В большинстве случаев выразить собственные функции явно не представляется возможным.

Известно, что если ковариационную функцию можно разделить следующим образом:

$$K_F(j,k;j',k') = K_C(j,j')K_R(k,k'), \qquad (2.5.3)$$

то ядро преобразования Карунена-Лоэва разделяется аналогичным образом, то есть

$$A_F(j,k;j',k') = A_C(j,j')A_R(k,k'). \quad (2.5.4)$$

Можно записать уравнения для строк и столбцов матрицы, описывающей эти ядра:

$$\lambda_R(v)A_R(k,v) = \sum_{k'=0}^{N-1} K_R(k,k')A_R(k',v), \quad (2.5.5)$$

$$\lambda_C(u)A_C(j,u) = \sum_{j'=0}^{N-1} K_C(j,j')A_C(j',u). \quad (2.5.6)$$

В векторном виде прямое и обратное преобразования Карунена-Лоэва могут быть записаны как

$$\mathbf{s} = \mathbf{A}\mathbf{f}, \quad (2.5.7)$$

$$\mathbf{f} = \mathbf{A}^{\mathrm{T}}\mathbf{s}, \quad (2.5.8)$$

где s и f – представленные в векторном виде результат преобразования и исходное изображение, соответственно, а A представляет собой матрицу преобразования, удовлетворяющую уравнению

$$\mathbf{A}\mathbf{K}_\mathbf{f} = \mathbf{\Lambda}\mathbf{A}, \quad (2.5.9)$$

где $\mathbf{K}_\mathbf{f}$ – ковариационная матрица вектора f, A – матрица, каждая строка которой является собственным вектором матрицы $\mathbf{K}_\mathbf{f}$, $\mathbf{\Lambda}$ – диагональная матрица из собственных значений матрицы $\mathbf{K}_\mathbf{f}$, которая записывается как

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{N^2} \end{pmatrix}. \quad (2.5.10)$$

Если матрица $\mathbf{K}_\mathbf{f}$ разделима, то матрицу преобразования A можно записать в виде

$$\mathbf{A} = \mathbf{A}_C \otimes \mathbf{A}_R, \quad (2.5.11)$$

а матрицы $\mathbf{A}_C$ и $\mathbf{A}_R$ удовлетворяют условиям

$$\mathbf{A}_R\mathbf{K}_R = \mathbf{\Lambda}_R\mathbf{A}_R, \quad (2.5.12)$$

$$\mathbf{A}_C\mathbf{K}_C = \mathbf{\Lambda}_C\mathbf{A}_C, \quad (2.5.13)$$

причем собственные значения в матрицах $\mathbf{\Lambda}_R$ и $\mathbf{\Lambda}_C$ удовлетворяют условию

$$\lambda_k = \lambda_{R,i}\lambda_{C,j} \quad (2.5.14)$$

при $i, j = 1, 2, \ldots N$.

Важное свойство преобразования Карунена-Лоэва заключается в том, что в среднем по ансамблю изображений основная часть энергии сосредоточена в наименьшем количестве коэффициентов s_{ij}, что не выполняется для других видов линейных преобразований.

Контрольные вопросы

1. Сформулируйте сущность системных преобразований сигналов в фотонике для задач бесконтактного контроля объектов.
2. Приведите три основных класса систем в зависимости от объема доступной априорной информации.
3. Приведите примеры линейных систем в оптике.
4. Объясните сущность понятия идентификации системы.
5. Перечислите и объясните три основные составляющие погрешности при математическом моделировании физических систем.
6. Перечислите основные виды двумерных последовательностей, представляющих сигналы в пространственной области.
7. Приведите примеры периодических двумерных последовательностей.
8. Напишите формулу разложения изображения по внешним произведениям базисных векторов.

Список литературы

2.1. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложениях к интерферометрическим системам. – СПб: БХФВ Санкт-Петербургу, 1998.

2.2. И.Н. Матвеев, А.Н. Сафронов, И.Н. Троицкий, Н.Д. Устинов Адаптация в информационных оптических системах. – М.: Радио и связь, 1984.

2.3. Цыпкин Я.З. Информационная теория идентификации. – М.: Наука, 1995.

2.4. Гуров И.П. Формирование и анализ стохастических интерференционных полей. В кн.: Проблемы когерентной и нелинейной оптики / Под ред. И.П. Гурова и С.А. Козлова. – СПб.: СПбГУИТМО, 2000. – С. 67–87.

2.5. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. – М.: Мир, 1988.

2.6. K.R. Rao, P.C. Yip The Transform and Data Compressing Handbook. – Boca Raton: CRC Press, 2001.

Раздел 3. Интегральные преобразования сигналов

3.1 Преобразования Фурье и Хартли

Прежде чем переходить к рассмотрению преобразования Фурье, следует рассмотреть понятие тригонометрического ряда Фурье. Если некоторая функция удовлетворяет условию

$$s(t) = s(t + mT), \tag{3.1.1}$$

где t – независимая переменная, m – некоторая целочисленная константа, то такая функция называется периодической с периодом T.

Из всех периодических функций вида (3.1.1) образуем пространство функций и введем для него скалярное произведение

$$(s_1, s_2) = \int_0^T s_1(t), s_2^*(t) dt, \tag{3.1.2}$$

где звездочкой обозначено комплексное сопряжение. Система функций $\{s_n(t)\}_{n=0..\infty}$, удовлетворяющая условию

$$(s_n, s_k) = \int_0^T s_n(t), s_k^*(t) dt = \delta_{n,k}, \tag{3.1.3}$$

где $\delta_{n,k}$ обозначает дельта-символ Кронекера

$$\delta_{n,k} = \begin{cases} 0, & n \neq k, \\ 1, & n = k, \end{cases} \tag{3.1.4}$$

называется ортонормированным базисом.

Любая периодическая функция может быть разложена в ряд по базисным функциям, удовлетворяющим условию (3.1.3),

$$s(t) = \sum_{n=0}^{\infty} c_n s_n(t), \tag{3.1.5}$$

где c_n представляют собой коэффициенты ряда, которые могут быть найдены как скалярные произведения исходной функции и соответствующих базисных функций

$$c_n = (s, s_n) = \int_0^T s(t), s_n^*(t) dt. \tag{3.1.6}$$

Представление (3.1.5) принято называть обобщенным рядом Фурье. Использование в качестве ортонормированного базиса набора функций

$$\frac{1}{\sqrt{T}}, \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi nt}{T}\right), \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi nt}{T}\right), n = 1...\infty \tag{3.1.7}$$

позволяет получить так называемый тригонометрический ряд Фурье

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(2\pi nt / T) + b_n \sin(2\pi nt / T)], \tag{3.1.8}$$

где коэффициенты a_0, a_n и b_n могут быть найдены с учетом (3.1.6) как

$$a_0 = \int_0^T s(t)dt, \tag{3.1.9}$$

$$a_n = \int_0^T s(t)\cos(2\pi nt/T)dt, \tag{3.1.10}$$

$$b_n = \int_0^T s(t)\sin(2\pi nt/T)dt. \tag{3.1.11}$$

Набор (3.1.7) называется тригонометрическим базисом. При $n = 1...\infty$ функции из набора (3.1.7) являются гармоническими с периодом T/n, то есть тригонометрический ряд Фурье является разложением по гармоническим функциям.

Ряд Фурье можно записать в эквивалентной форме

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cos(n\omega t + \varphi_n), \tag{3.1.12}$$

где $\omega = 2\pi/T$ – основная частота ряда, A_n и φ_n – амплитуда и фаза соответствующих гармоник, которые могут быть рассчитаны из соответствующих коэффициентов Фурье как

$$A_n = \sqrt{a_n^2 + b_n^2}, \tag{3.1.13}$$

$$\varphi_n = \begin{cases} 2\pi - \arccos(a_n / A_n), & b_n \geq 0, \\ \arccos(a_n / A_n), & b_n < 0. \end{cases} \tag{3.1.14}$$

Используя формулу для записи косинуса как полусуммы двух комплексно сопряженных экспонент, выражение (3.1.12) можно переписать в следующей форме:

$$s(t) = \frac{a_0}{2} + \frac{1}{2}\sum_{n=1}^{\infty} A_n \exp(jn\omega t + j\varphi_n) + \frac{1}{2}\sum_{n=1}^{\infty} A_n \exp(-jn\omega t - j\varphi_n) =$$

$$= \frac{a_0}{2} + \frac{1}{2}\sum_{n=1}^{\infty} A_n \exp(j\omega_n t + j\varphi_n) + \frac{1}{2}\sum_{n=1}^{\infty} A_n \exp(j\omega_{-n} t + j\varphi_{-n}), \tag{3.1.15}$$

где j – комплексная единица.

Это позволяет записать ряд Фурье в комплексной форме

$$s(t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} A_n \exp(j\omega_n t + j\varphi_n) = \sum_{n=-\infty}^{\infty} C_n \exp(j\omega_n t), \tag{3.1.16}$$

где $C_0 = a_0/2$, $C_n = A_n \exp(j\varphi_n)/2$ для $n = \pm 1, \pm 2, ...$ – комплексные амплитуды гармоник. Для новой записи (3.1.16) переменные с отрицательными индексами определяются следующим образом: $\omega_{-n} = -\omega_n = -\omega n$

$$\omega_{-n} = -\omega_n = -\omega n, \tag{3.1.17}$$

$$A_{-n} = A_n, \tag{3.1.18}$$

$$\varphi_{-n} = -\varphi_n, \tag{3.1.19}$$

$$C_{-n} = C_n^* . \qquad (3.1.20)$$

Отрицательная частота в (3.1.17) не имеет физического смысла и является математическим понятием, определяемым способом представления комплексных чисел. Комплексную амплитуду C_n можно представить с помощью выражения

$$C_n = \frac{1}{T}\int_0^T s(t)\exp(-jn\omega t)dt . \qquad (3.1.21)$$

Вычисление коэффициентов Фурье для функции $s(t)$, заданной на отрезке $[0, T]$ называется гармоническим анализом. Совокупность гармонических составляющих, образующих ряд Фурье, называется спектром функции $s(t)$. На практике часто рассматривают спектр амплитуд или спектр фаз, которые представляют собой, соответственно, зависимости A_n и φ_n от частоты ω_n.

Непрерывное преобразование Фурье

Преобразование Фурье является обобщением понятия ряда Фурье на случай практически произвольных непериодических функций. Такие функции можно рассматривать как функции с бесконечным периодом T. Областью определения непериодических функций является вся ось вещественных чисел $(-\infty, \infty)$. С учетом периодичности функции (3.1.1) выражение (3.1.21) можно записать в эквивалентной форме

$$C_n = \frac{1}{T}\int_0^T s(t)\exp(-jn\omega t)dt = \frac{1}{T}\int_{-T/2}^{T/2} s(t)\exp(-jn\omega t)dt . \qquad (3.1.22)$$

После предельного перехода $T \to \infty$ основная частота $\omega = 2\pi/T$, которая определяет расстояние между соседними отсчетами спектра $\delta\omega = \omega_{n+1} - \omega_n = \omega$, будет стремиться к нулю, в результате чего, отсчеты спектра будут плотно заполнять всю вещественную ось частот, то есть спектр сигнала будет представлять собой непрерывную функцию от частоты. С учетом этого, выражение (3.1.16) можно переписать в виде

$$s(t) = \sum_{n=-\infty}^{\infty} C_n \exp(j\omega_n t) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} G(\omega_n)\exp(j\omega_n t)\delta\omega , \qquad (3.1.23)$$

где функция $G(\omega_n)$ определяется для бесконечного дискретного набора частот и записывается как

$$G(\omega_n) = C_n T = 2\pi\frac{C_n}{\delta\omega} = \int_{-T/2}^{T/2} s(t)\exp(-jn\omega t)dt . \qquad (3.1.24)$$

После предельного перехода $T \to \infty$, аналогичном $\delta\omega \to 0$, бесконечная сумма в (3.1.23) переходит в интегральную сумму

$$\lim_{\delta\omega\to 0}\frac{1}{2\pi}\sum_{n=-\infty}^{\infty}G(\omega_n)\exp(j\omega_n t)\delta\omega=\frac{1}{2\pi}\int_{-\infty}^{\infty}G(\omega)\exp(j\omega t)d\omega. \qquad (3.1.25)$$

Выражения (3.1.23) и (3.1.24) теперь можно переписать в форме

$$G(\omega)=\int_{-\infty}^{\infty}s(t)\exp(-j\omega t)dt=\mathrm{F}\{s(t)\}, \qquad (3.1.26)$$

$$s(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}G(\omega)\exp(j\omega t)d\omega=\mathrm{F}^{-1}\{G(\omega)\}. \qquad (3.1.27)$$

Видно, что функция $G(\omega)$, представляющая собой непрерывный спектр, теперь определена на всей вещественной оси частот.

Полученные интегральные уравнения (3.1.26) и (3.1.27) называются преобразованием Фурье, которое принято обозначать как оператор $\mathrm{F}\{.\}$. Функции, связанные этим преобразованием называются сопряженными по Фурье.

Следует отметить, что если независимая переменная t обозначает некоторую координату в пространстве, то переменная ω называется, соответственно, пространственной частотой.

Некоторые полезные свойства преобразования Фурье перечислены ниже.

1. Взаимная однозначность. Для каждой функции существует только один соответствующий ей Фурье-образ, по которому может быть восстановлена эта функция. Это свойство можно записать как

$$\mathrm{F}^{-1}\{\mathrm{F}\{s(t)\}\}=s(t), \qquad (3.1.28)$$

$$\mathrm{F}\{\mathrm{F}^{-1}\{G(\omega)\}\}=G(\omega). \qquad (3.1.29)$$

2. Линейность. Преобразование Фурье линейной комбинации двух функций равно линейной комбинации преобразований Фурье этих функций

$$\mathrm{F}\{\alpha s_1(t)+\beta s_2(t)\}=\alpha\mathrm{F}\{s_1(t)\}+\beta\mathrm{F}\{s_2(t)\}, \qquad (3.1.30)$$

где α и β – некоторые константы.

3. Теорема о свертке. Операция свертки двух функций обозначается как

$$s_1(t)\otimes s_2(t)=\int_{-\infty}^{\infty}s_1(\tau)s_2(t-\tau)d\tau. \qquad (3.1.31)$$

В соответствии с прямой теоремой о свертке преобразование Фурье свертки двух функций эквивалентно произведению их Фурье-образов.

$$\mathrm{F}\{s_1(t)\otimes s_2(t)\}=\mathrm{F}\{s_1(t)\}\mathrm{F}\{s_2(t)\}. \qquad (3.1.32)$$

Обратная теорема о свертке в свою очередь утверждает, что преобразование Фурье от произведения двух функций представляет собой свертку Фурье-образов этих функций

$$\mathrm{F}\{s_1(t)s_2(t)\} = \frac{1}{2\pi}\mathrm{F}\{s_1(t)\} \otimes \mathrm{F}\{s_2(t)\}. \quad (3.1.33)$$

4. Теорема смещения. Это свойство определяется то, как смещается спектр функции (или сама функция, если речь идет об обратном преобразовании), при умножении ее на экспоненту с соответствующей частотой

$$\mathrm{F}\{s(t)\exp(j\omega_0 t)\} = G(\omega - \omega_0), \quad (3.1.34)$$

$$\mathrm{F}^{-1}\{G(\omega)\exp(j\omega t_0)\} = s(t - t_0). \quad (3.1.35)$$

5. Теорема Парсеваля для интеграла Фурье. Это свойство равенства энергий исходной функции и ее спектра

$$\int_{-\infty}^{\infty} |s(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |G(\omega)|^2 d\omega, \quad (3.1.36)$$

где функция $S(\omega) = |G(\omega)|^2$ называется спектральной плотностью энергии сигнала.

Преобразование Фурье легко обобщается на многомерный случай. Так для функции $s(x, y)$, которая может представлять, например, двумерное распределение интенсивности оптического излучения, прямое и обратное преобразование Фурье можно записать как

$$G(\omega_x, \omega_y) = \int_{-\infty}^{\infty} s(x, y)\exp(-j\omega(x + y))dxdy, \quad (3.1.37)$$

$$s(x, y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} G(\omega_x, \omega_y)\exp(j(\omega_x + \omega_y)t)d\omega_x d\omega_y, \quad (3.1.38)$$

где ω_x и ω_y – соответствующие координатам x и y пространственные частоты.

Дискретное преобразование Фурье

Пусть некоторая функция $s(t)$, определенная на промежутке $0 \le t < T$, задана набором из N дискретных отсчетов $s_0, s_1, s_2, ..., s_{N-1}$, которые взяты в соответствующих точках $t_0, t_1, t_2, ..., t_{N-1}$, для которых выполняется условие эквидистантности

$$t_k = m\Delta t, \quad m = 0,1,..., N - 1. \quad (3.1.39)$$

По аналогии с (3.1.26)–(3.1.27) можно записать дискретное преобразование Фурье (ДПФ), связывающее дискретное представление функции $s(t)$ с ее дискретным спектром G_k:

$$G_k = \sum_{m=0}^{N-1} s_m \exp(-j2\pi mk), \quad (3.1.40)$$

$$s_m = \frac{1}{N}\sum_{k=0}^{N-1} G_k \exp(j2\pi mk). \qquad (3.1.41)$$

Так как полученный спектр представлен дискретной последовательностью отсчетов G_k, он должен соответствовать некоторой периодической последовательности. Так как исходная последовательность содержит только N отсчетов, при анализе ее спектра, делается предположение о том, что она периодична, с периодом N

$$s_{pN+m} = s_m, \quad m = 0,1,...,N-1, \qquad (3.1.42)$$

где p – любое целое число. Таким образом, если исследуется

ДПФ широко используется для задач обработки сигналов, в теории информации и других областях. Его популярность связана с разработкой большого количества алгоритмов быстрого преобразования Фурье (БПФ), первые из которых были опубликованы в 1965 году. Алгоритмы БПФ обеспечивают значительный выигрыш в скорости работы по сравнению с прямым методом вычисления ДПФ.

Преобразование Хартли

Одним из недостатков преобразования Фурье является то, что в результате применения этого преобразования к вещественным сигналам получают комплексное представление его спектра. Это не всегда удобно, так как при анализе спектров приходится осуществлять трудоемкие вычисления в области комплексных чисел, а также хранить информацию, как о вещественной, так и о мнимой части каждого спектрального отсчета, что ведет к увеличению объема требуемой памяти.

Существует вещественный аналог преобразования Фурье, известный как преобразование Хартли. Главным его преимуществом является тот факт, что оно позволяет получить из вещественного сигнала вещественное представление его спектра, что является удобным в ряде случаев.

Прямое и обратное преобразование Хартли определяются как

$$H(f) = \int_{-\infty}^{\infty} s(t)\mathrm{cas}(2\pi ft)dt, \qquad (3.1.43)$$

$$s(t) = \int_{-\infty}^{\infty} H(t)\mathrm{cas}(2\pi ft)df, \qquad (3.1.44)$$

где $H(f)$ представляет собой вещественную функцию от частоты f, функция $\mathrm{cas}(t)$ представляет собой специальную введенную Хартли функцию, которая представляет собой сумму функций синуса и косинуса от одного аргумента

$$\mathrm{cas}(t) = \sin(t) + \cos(t). \qquad (3.1.45)$$

Прямое и обратное преобразование Хартли в данном случае неразличимы.

Представим результат преобразование Хартли $H(f)$ в виде суммы четной $E(f)$ и нечетной $O(f)$ компонент.

$$H(f) = E(f) + O(f). \tag{3.1.46}$$

Четную компоненту $E(f)$ можно определить как

$$E(f) = \frac{H(f) + H(-f)}{2} = \int_{-\infty}^{\infty} s(t)\cos(2\pi ft)dt, \tag{3.1.47}$$

где функция $H(-f)$ представляет собой зеркальное отражение функции $H(f)$. Нечетная компонента $O(f)$ в свою очередь определяется как

$$O(f) = \frac{H(f) - H(-f)}{2} = \int_{-\infty}^{\infty} s(t)\sin(2\pi ft)dt. \tag{3.1.48}$$

При этом несложно показать, что $E(-f) = E(f)$, а $O(-f) = -O(f)$.

При известной функции $H(f)$ несложно получить результат преобразования Фурье от функции $s(t)$

$$\mathrm{F}\{s(t)\} = E(f) - jO(f) = \int_{-\infty}^{\infty} s(t)[\cos(2\pi ft) - j\sin(2\pi ft)]dt =$$

$$= \int_{-\infty}^{\infty} s(t)\exp(-2j\pi ft)dt, \tag{3.1.49}$$

где $\mathrm{F}\{.\}$ – оператор преобразования Фурье, j – мнимая единица. Вещественная и мнимая части преобразования Фурье равны соответственно

$$\mathrm{Re}\,\mathrm{F}\{s(t)\} = E(f), \tag{3.1.50}$$

$$\mathrm{Im}\mathrm{F}\{s(t)\} = -O(f). \tag{3.1.51}$$

Преобразование Хартли в свою очередь так же может быть получено при известном преобразовании Фурье как разность его вещественной и мнимой части

$$H(f) = \mathrm{Re}\mathrm{F}\{s(t)\} - \mathrm{Im}\mathrm{F}\{s(t)\}. \tag{3.1.52}$$

Следует отметить, что преобразование Хартли представляет собой вещественную функцию только в случае, когда исходный сигнал так же является вещественным.

Если некоторая функция $s(t)$ имеет преобразование Хартли $H(f)$ то для этого преобразования справедливы следующие соотношения.

1. Линейность. Преобразование Хартли линейной комбинации двух функций равно линейной комбинации преобразований этих функций

$$\mathrm{H}\{\alpha s_1(t) + \beta s_2(t)\} = \alpha\mathrm{H}\{s_1(t)\} + \beta\mathrm{H}\{s_2(t)\}, \tag{3.1.53}$$

где $\mathrm{H}\{.\}$ – оператор преобразования Хартли, α и β – некоторые константы.

2. Масштабирование. Преобразование Хартли растянутой по временной шкале функции имеет вид

$$\mathrm{H}\{s(t/\alpha)\} = |T|H(\alpha f), \tag{3.1.54}$$

где α – масштабирующий коэффициент.

3. Сдвиг во временной области. Для преобразования Хартли сдвинутой по временной шкале функции справедливо соотношение

$$\mathrm{H}\{s(t-\alpha)\} = \sin(2\pi\alpha f)H(-f) + \cos(2\pi\alpha f)H(f), \qquad (3.1.55)$$

где α – величина сдвига.

4. Модуляция. Если обеспечить модуляцию исходного сигнала $s(t)$ некоторой периодической функцией с частотой f_0, то преобразование Хартли этой функции можно записать как

$$\mathrm{H}\{s(t)\cos(2\pi f_0 t)\} = \frac{1}{2}[H(f-f_0) + H(f+f_0)]. \qquad (3.1.56)$$

5. Теорема о свертке. Преобразование Хартли свертки двух функций выражается следующим образом:

$$\mathrm{H}\{s_1(t) \otimes s_2(t)\} = \frac{1}{2}[H_1(f)H_2(f) - H_1(-f)H_2(-f)] +$$

$$+ \frac{1}{2}[H_1(f)H_2(-f) + H_1(-f)H_2(f)]. \qquad (3.1.57)$$

Дискретное преобразование Хартли

Заменяя интегралы в (3.1.43) и (3.1.44) на суммы, можно получить прямой и обратное дискретное преобразование Хартли:

$$H_k = \frac{1}{N}\sum_{m=0}^{N-1} s_m \mathrm{cas}(2\pi fm/N), \qquad (3.1.58)$$

$$s_m = \sum_{k=0}^{N-1} H_k \mathrm{cas}(2\pi fk/N), \qquad (3.1.59)$$

где m и k – номера дискретных отсчетов временного и частотного представлений исходного сигнала.

Получаемые в результате дискретного преобразования Хартли последовательности являются вещественными, что снижает требования памяти при численных расчетах.

3.2 Вейвлет-преобразование

Вейвлет-преобразование представляет собой разложение сигнала в ряд по базисным функциям особого вида – вейвлетам (англ. wavelet – маленькая волна). Вейвлеты являются масштабируемыми функциями и позволяют исследовать локальные свойства сигналов, разделяя их по различным частотным компонентам на малых интервалах в области независимой переменной и затем изучать особенности этих сигналов с разрешением, соответствующим выбранному масштабу.

Ненулевая часть вейвлета локализована на ограниченном интервале. Вейвлет-функция свертывается с сигналом, то есть представляет собой

скользящее вдоль сигнала локализованное «окно» заданной формы. При этом вейвлет-преобразование обеспечивает возможность выделять местоположение особенностей сигнала.

Благодаря перечисленным свойствам вейвлет-анализ нашел широкое применение во многих областях, таких как локализация и различение особых точек сигналов, вычисление их различных характеристик, подавление шумов, сжатие данных и другие.

Вейвлет-преобразование определяется следующим образом:

$$W(a,x)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(t)w^*\left(\frac{t-x}{a}\right)dt, \qquad (3.2.1)$$

где результатом преобразования является функция, зависящая от двух переменных: координаты x и масштаба a. Используемая здесь функция $w^*(.)$ является базисной функцией вейвлет-преобразования, а звездочкой обозначается комплексное сопряжение.

Базисная вейвлет-функция должна удовлетворять двум условиям. Во-первых, она должна быстро убывать в области независимой переменной (то есть ненулевая часть вейвлета должна быть сосредоточена на отрезке конечной протяженности). Во-вторых, интеграл базисной функции должен быть равен нулю.

На каждом отдельном масштабе базисная функция w растягивается по горизонтали и сжимается по вертикали. Затем она сдвигается в точку x исследуемой функции и свертывается с ней.

Вейвлет преобразование является, как и преобразование Фурье, линейным, то есть выполняется условие

$$\mathrm{W}\{f_1(x)+f_2(x)\}=\mathrm{W}\{f_1(x)\}+\mathrm{W}\{f_2(x)\}, \qquad (3.2.2)$$

где $\mathrm{W}\{.\}$ обозначает операцию вейвлет-преобразования.

Сдвиг и масштабирование исходной функции отражаются на результате ее вейвлет-преобразования следующим образом:

$$\mathrm{W}\{f(x-u)\}=W(a,x-u), \qquad (3.2.3)$$

$$\mathrm{W}\{f_1(cx)\}=\frac{1}{\sqrt{c}}W(ca,cx). \qquad (3.2.4)$$

Пусть волнистой линией обозначается результат преобразования Фурье некоторой функции. Тогда для результата вейвлет-преобразования справедливо следующее соотношение:

$$\widetilde{W}(a,x)=\sqrt{a}\,\widetilde{f}(\nu)\widetilde{w}^*(a\upsilon). \qquad (3.2.5)$$

С изменением масштаба a вейвлет-функция $w^*(a\nu)$ сжимается и расширяется по соответствующим осям, сохраняя при этом свою форму.

Первоначальная функция $f(t)$ может быть восстановлена с помощью формулы обратного преобразования:

$$f(t)=\frac{1}{C_\chi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{a}}W(a,x)w\left(\frac{t-x}{a}\right)\frac{dadx}{a^2}, \tag{3.2.6}$$

где C_χ – норма базисной функции w

$$C_\chi=\int_0^\infty\frac{\widetilde{w}^2(\nu)}{\nu}d\nu. \tag{3.2.7}$$

Восстановление сигнала возможно, только если норма вейвлет-функции корректно определена (условие допустимости). Очевидно, для этого необходимо, чтобы выполнялось условие $\widetilde{w}(0)=0$. Формулы вейвлет-преобразования могут быть обобщены и для комплексных функций.

Простейшим примером вейвлета может служить функции, на основе которой строится базис Хаара:

$$f(x)=\begin{cases}1\ :x\in[-1,0),\\ -1:\ x\in[0,1),\\ 0\ :x\notin[-1,1).\end{cases} \tag{3.2.8}$$

Более сложным примером вейвлета является вейвлет-функция «сомбреро», определенная как вторая производная функции Гаусса:

$$w(t)=\frac{1}{\sigma^3\sqrt{2\pi}}\left(\frac{t^2}{\sigma}-1\right)\exp\left(-\frac{t^2}{2\sigma^2}\right), \tag{3.2.9}$$

где σ – параметр, определяющий ширину функции Гаусса.

Базовым вейвлетом называют такую базисную функцию, из которой путем перемасштабирования и сдвига получают набор вейвлет-функций. На рис. 3.1 базисный вейвлет обозначен сплошной линией, а перемасштабированный – пунктирной.

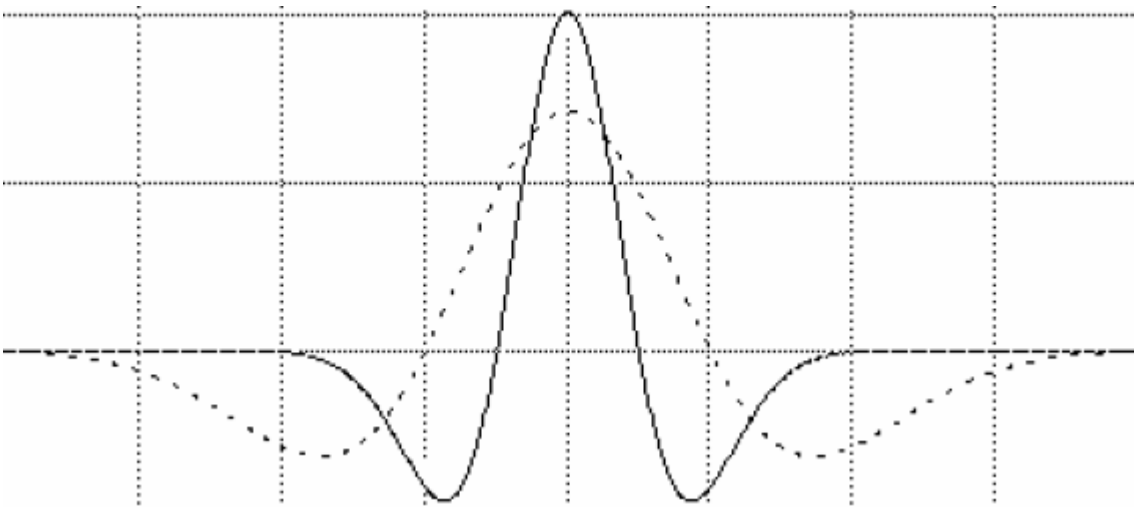


Рис. 3.1. Вейвлет «сомбреро» до (сплошная линия) и после (пунктирная линия) перемасштабирования

Из рис. 3.1 видно, что частота вейвлет-функции после перемасштабирования уменьшилась.

Еще одним примером вейвлет-функции является вейвлет Морле, который задается комплексной функцией

$$w(t) = \exp(jat)\exp\left(-\frac{t^2}{2\sigma^2}\right), \qquad (3.2.9)$$

где a – параметр модуляции. Пример этого вейвлета представлен на рис. 3.2.

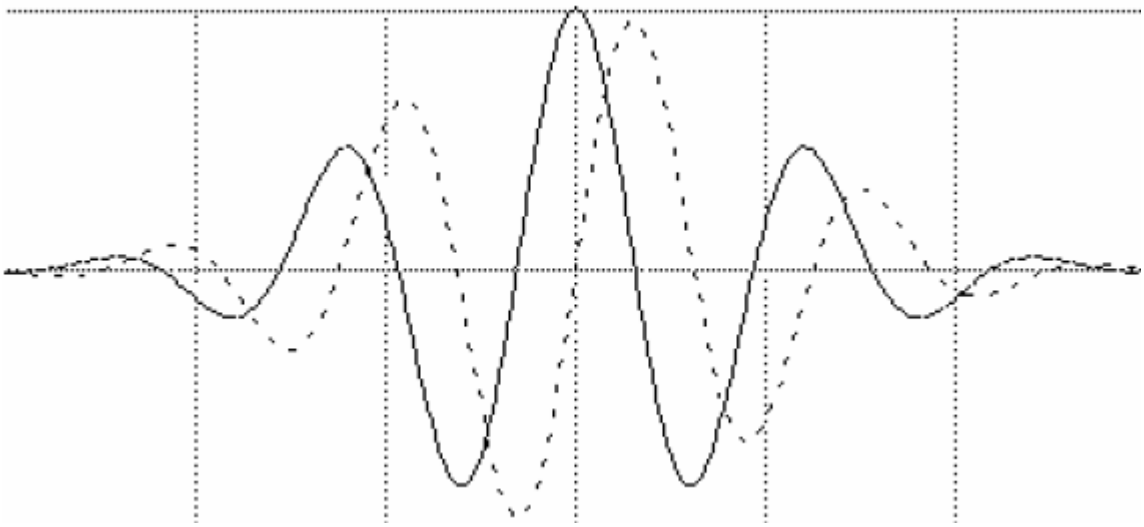


Рис. 3.2. Вейвлет Морле: действительная (сплошная линия) и мнимая (пунктирная линия) части

При увеличении параметра модуляции количество квазипериодов в вейвлете Морле увеличивается. Это показано на рис. 3.3.

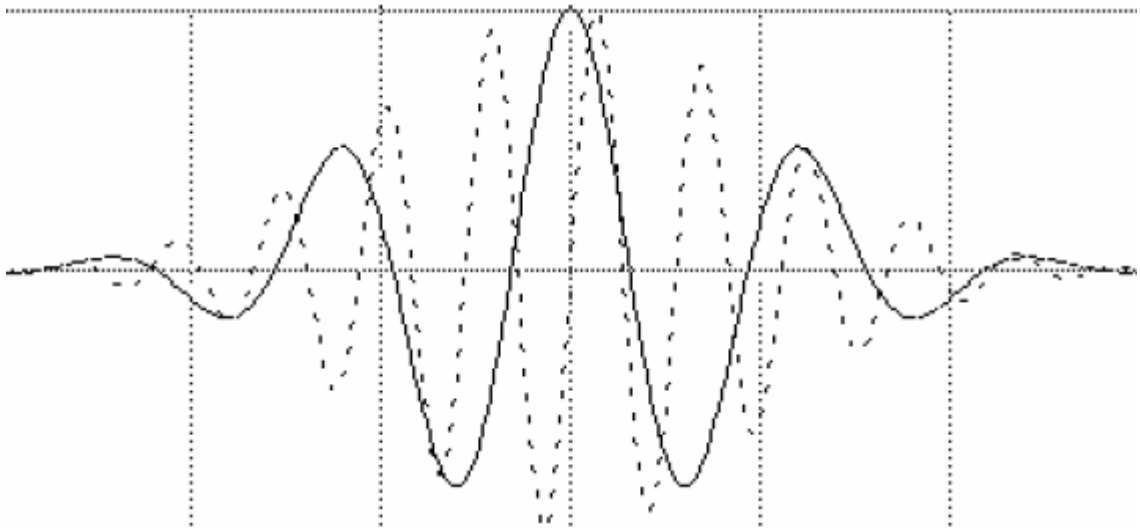


Рис. 3.3. Вейвлет Морле при увеличении параметра модуляции: действительная (сплошная линия) и мнимая (пунктирная линия) части

Фурье-спектр вейвлета Морле сконцентрирован в окрестности некоторой частоты f, а спектр сжатого в n раз вейвлета будет сконцентрирован вокруг частоты f/n, как показано на рис. 3.4. Это связано со свойством подобия преобразования Фурье.

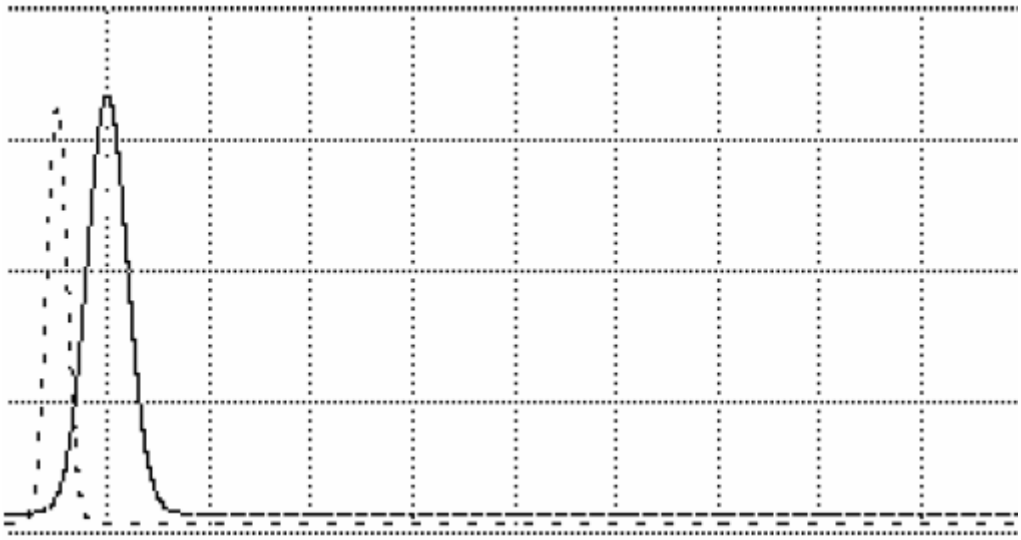


Рис. 3.4. Изменение частоты при перемасштабировании вейвлета

Каждый вид вейвлет-функции имеет определенные частоты, на которые он настроен (спектр вейвлета, вообще говоря, может иметь и более одного пика). Эти частоты изменяются при изменении масштаба в n раз. Таким образом, на разных масштабах вейвет-преобразование позволяет различить различные особенности анализируемых сигналов.

Рис. 3.5 иллюстрирует частотно-временные особенности вейвлет-преобразования. Каждый блок на рисунке соответствует значению вейвлет-преобразования на фазовой плоскости. Все точки на этой плоскости, которые попадают в один блок, представляются одним значением вейвлет-преобразования.

Видно, что, хотя форма ячеек изменяется, площадь остается постоянной. В результате блоки эквивалентны, но «смешивают» в разных пропорциях частоту и время. На низких частотах высота ячеек ниже, что приводит к меньшей неоднозначности, а, следовательно, лучшему разрешению по частоте и худшему по времени. На более высоких частотах ширина блоков меньше, то есть разрешающая способность по времени лучше, но высота блоков больше, что ведет к снижению разрешающей способности по частоте.

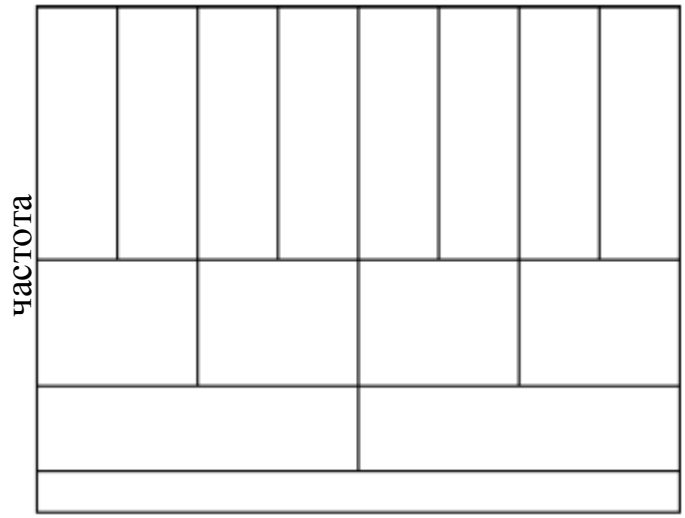


Рис. 3.5. Покрытие фазовой плоскости вейвлетами

Вейвлет-функции имеют различные наборы основных частот. От удачного выбора вейвлета для каждой конкретной задачи зависит информативность результата. Четкий критерий выбора существует далеко не для всех ситуаций. Часто поиск подходящей вейвлет-функции осуществляется подбором.

3.3 Преобразование Лапласа и Z-преобразование

Непрерывное преобразование Лапласа

Одним инструментов решения задач математики и физики является преобразование Лапласа. Оно связывает некоторую функцию от вещественной переменной $f(t)$, называемую функцией-оригиналом, и функцию от комплексной переменной $F(s)$, называемую функцией-изображением, при помощи интегрального оператора $\mathrm{L}\{.\}$

$$F(s) = \mathrm{L}\{f(t)\} = \int_0^\infty f(t)\exp(-st)dt, \qquad (3.3.1)$$

где s – комплексная переменная. Обратное преобразование Лапласа можно определить как

$$f(t) = \mathrm{L}^{-1}\{F(s)\} = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma-iT}^{\gamma+iT} F(s)\exp(st)ds, \qquad (3.3.2)$$

где γ – некоторое вещественное число. Эта пара преобразований важна в оптике и компьютерной фотонике, потому что в общем виде задает связь между временным и частотным представлением сигналов, в том числе многомерных. Распространенность преобразования Лапласа связана с тем,

что многие операции над функциями-оригиналами упрощаются в области изображений.

Необходимым условием существования преобразования Лапласа является дифференцируемость функции-оригинала на интервале $[0,\infty)$, однако оно может быть определено и для отрицательных значений аргумента t. Такое преобразование называется двусторонним преобразованием Лапласа и записывается как

$$\mathrm{B}\{f(t)\} = \int_{-\infty}^{\infty} f(t)\exp(-st)dt. \qquad (3.3.3)$$

Двустороннее преобразование Лапласа переходит в преобразование Фурье при замене

$$s = i\omega. \qquad (3.3.4)$$

где ω – частотная переменная. Таким образом, преобразование Фурье является частным случаем преобразования Лапласа.

Некоторые полезные свойства преобразования Лапласа приведены ниже.

1. Линейность. Преобразование Лапласа линейной комбинации двух функций равно линейной комбинации преобразований Лапласа этих функций

$$\mathrm{L}\{\alpha f_1(t) + \beta f_2(t)\} = \alpha\mathrm{L}\{f_1(t)\} + \beta\mathrm{L}\{f_2(t)\}, \qquad (3.3.5)$$

где α и β – некоторые константы.

2. Сдвиг в области вещественного аргумента. Для преобразования Лапласа справедливо соотношение

$$\mathrm{L}\{f(t-\alpha)u(t-\alpha)\} = \exp(-\alpha s)F(s), \qquad (3.3.6)$$

где $u(t)$ – функция Хэвисайда, определяемая как

$$u(t) = \begin{cases} 0, & \text{если } t < 0 \\ 1, & \text{если } t \geq 0 \end{cases}. \qquad (3.3.7)$$

Использование функции Хэвисайда в (3.3.6) приравнивает к нулю интеграл функции-оригинала для всех отрицательных значений аргумента, что обеспечивает существования ее изображения.

3. Сдвиг в области комплексной переменной. Сдвиг функции изображения может быть обеспечен умножением функции-оригинала на экспоненту от аргумента αt, где α – величина сдвига:

$$\mathrm{L}\{\exp(\alpha t)f(t)\} = F(s-\alpha). \qquad (3.3.8)$$

4. Теорема о свертке. Преобразование Лапласа от свертки двух функций-оригиналов может быть выражено как умножение преобразований Лапласа соответствующих функций:

$$\mathrm{L}\{f_1(t) \otimes f_2(t)\} = \mathrm{L}\{f_1(t)\}\mathrm{L}\{f_2(t)\}. \qquad (3.3.9)$$

5. Преобразование Лапласа периодических функций. Если функция-оригинал является периодической с периодом T, то есть удовлетворяет условию

$$f(t+T)=f(t), \tag{3.3.10}$$

то преобразование Лапласа этой функции может быть записано в форме

$$\mathrm{L}\{f(t)\}=\frac{1}{1-\exp(-sT)}\int_0^T f(t)\exp(-st)dt. \tag{3.3.11}$$

Преобразование Лапласа имеет важное значение в теории вероятностей. Если имеется случайная величина X с функцией плотности вероятности p, то преобразование Лапласа этой функции задается математическим ожиданием:

$$\mathrm{L}\{p(s)\}=E[\exp(-sX)]. \tag{3.3.12}$$

Обратное преобразование Лапласа в данном случае позволяет восстановить функцию распределения вероятностей

$$P_X(x)=\mathrm{L}^{-1}\left\{\frac{1}{s}E[\exp(-sX)]\right\}. \tag{3.3.13}$$

Z-преобразование

В компьютерной фотонике, как правило, рассматриваются дискретные сигналы. Это приводит к необходимости использования дискретных реализаций интегральных преобразований для анализа таких сигналов. Преобразование Лапласа (3.3.1) в дискретной форме представляет собой разложение в бесконечный ряд:

$$F(s)=\mathrm{L}\{f_k\}=\sum_{k=0}^{\infty} f_k\exp(-sk\Delta t). \tag{3.3.14}$$

где функция $f(k)$ представляет собой последовательность отсчетов, связанную с непрерывной функцией-оригиналом $f(t)$ соотношением

$$f_k=f(t)\delta(t-k\Delta t), \tag{3.3.15}$$

где $\delta(t-k\Delta t)$ – дельта-функция, k – номер дискретного отсчета, Δt – шаг дискретизации.

На практике удобнее пользоваться так называемым Z-преобразованием, которое может быть получено из дискретного преобразования Лапласа путем замены переменной

$$z=\exp(s\Delta t), \tag{3.3.16}$$

что позволяет получить из (3.3.14) степенной ряд вида

$$F(z)=\mathrm{Z}\{f_k\}=\sum_{k=0}^{\infty} f_k z^{-k}, \tag{3.3.17}$$

где $\mathrm{Z}\{.\}$ обозначает оператор Z-преобразования. Z-преобразование также называют преобразованием Лорана.

Существует Z-преобразование, определенное и для отрицательных значений k. Такое преобразование называется двусторонним и записывается как

$$F(z) = \mathrm{Z}\{f_k\} = \sum_{k=-\infty}^{\infty} f_k z^{-k} \ . \qquad (3.3.18)$$

Обратное *Z*-преобразование определяется как

$$f_k = \frac{1}{2\pi j} \oint_C F(z) z^{k-1} dz \, , \qquad (3.3.19)$$

где C – это замкнутый контур, который огибает область сходимости бесконечной суммы (3.3.17).

На практике часто неудобно использовать выражение (3.3.19), поэтому в зависимости от вида подынтегральной функции, оно может заменяться более простыми эквивалентами. Например, если подынтегральная функция имеет особенности только в виде полюсов, то, используя теорему о вычетах, (3.3.19) можно записать как

$$f_k = \mathrm{Z}^{-1}\{F(z)\} = \sum_n \mathrm{Re}[F(z_n) z_n^{k-1}] \, , \qquad (3.3.20)$$

где суммирование ведется по всем полюсам функции.

Контрольные вопросы

1. Перечислите основные свойства преобразования Фурье и приведите примеры применительно к задачам анализа изображений.
2. Приведите формулы сравнения преобразований Фурье и Хартли.
3. Каким условиям должны удовлетворять функции, называемые вейвлетами?
4. Напишите формулу, определяющую вейвлет Морле.

Список литературы

3.1. Прэтт У. Цифровая обработка изображений. – М.: Мир, 1982.

3.2. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. – М.: Мир, 1988.

3.3. Шполянский Ю.А. Численные методы для модлирования оптических материалов и процессов. Часть 1. Элементы теории. – СПб: СПбГУ ИТМО, 2008.

3.4. Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления сверток. – М.: Радио и связь, 1985.

3.5. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложениях к интерферометрическим системам. – СПб: БХФВ Санкт-Петербургу, 1998.

3.6. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005.

3.7. Марпл С.Л. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990.
3.8. Брейсуэлл Р. Преобразование Хартли. – М.: Мир, 1990.

Раздел 4. Решение обратных и некорректных задач

4.1 Обратные и некорректные задачи в фотонике

Прямая задача в оптических исследованиях заключается в определении особенностей распространения излучения от источника. Обратная задача состоит в нахождении характеристик объекта по данным регистрируемого излучения от исследуемого объекта.

Сущность обратной оптической задачи хорошо известна: о размерах, форме, поверхностной структуре объектов судят посредством анализа отраженного (поглощенного) ими излучения. Другими примерами служат реставрация смазанных оптических изображений, восстановление фазы по интенсивности интерференционных полос и т.п.

Пусть источники и рассеиватели описываются множеством пространственно-временных функций

$$G = (g_1, g_2, ..., g_n), \tag{4.1.1}$$

которые называются функциями источников (предполагается, что рассеиватели рассматриваются как вторичные источники). Результирующее распределение излучения описывается множеством пространственно-временных функций

$$F = (f_1, f_2, ..., f_n), \tag{4.1.2}$$

называемых данными, которые могут быть проверены с помощью измерений. Из функций источников g_i можно однозначно получить данные f_i с помощью прямых соотношений

$$f_i = \mathbf{E}_i(g_1, g_2, ..., g_n), \tag{4.1.3}$$

где множество E операторов $\mathbf{E}_i$ является отображением G в F

$$E : G \to F. \tag{4.1.4}$$

Например, в когерентной оптике $\mathbf{E}_i$ соответствует некоторым интегральным преобразованиям, а g_i и f_i – источникам и, например, амплитудам в дальней зоне и их корреляциям. Решение прямой задачи означает вычисление данных f_i по известным функциям источников g_i с использованием прямых соотношений (4.1.3).

Решение обратной задачи состоит в нахождении функции источников, которые соответствуют полученным данным f_i и согласуются с физическими законами или экспериментами, то есть с так называемой априорной информацией.

Априорная информация сужает класс возможных функций источников. Например, можно часто допускать, что источник имеет конечный объем, исследуемый объект – ограниченные размеры.

Существует два противоположных подхода к решению указанной выше задачи. Первый заключается в установлении формулы или алгоритма, которые позволяют найти функции источников обращением отображения (4.1.3), а именно:

$$E^{-1} : F \to G. \qquad (4.1.5)$$

Название обратной задачи обычно закреплено за этим подходом.

Второй подход заключается в поиске модели функций источников методом подбора и подгонке неизвестных параметров таким образом, чтобы они соответствовали экспериментальным данным. Этот подход позволяет решить обратную задачу при помощи прямых процедур.

На практике переход от обратных процедур к прямым становится менее заметным с увеличением априорной информации.

Известно, что обращенные отображения (4.1.5) включает в себя математические вопросы существования, единственности и устойчивости решений. Например, экстраполяция данных в оптических изображениях принадлежит к классу задач (обычно называемых «некорректно поставленными»), в которых решение зависит от данных однозначно, но не непрерывно. Небольшие погрешности в данных приводят к большим ошибкам, если дополнительно не поставлены подходящие стабилизирующие условия, т.е. если не привлечены не требующие доказательств априорные сведения о решении. Конечно, ошибки неизбежны в экспериментальных данных.

Искомые характеристики источников или рассеивателей обычно получаются из данных об интенсивности и фазе посредством соответствующего обратного соотношения с учетом имеющейся априорной информации. Обратное соотношение, или алгоритм обращения, очевидно, основывается на соответствующей теории распространения и рассеяния излучения. Детекторы обеспечивают получение данных об интенсивности. Таким образом, следует получить необходимую информацию о фазе из распределений интенсивности. Эта задача называется восстановлением фазы.

Данные об интенсивности могут также включать в себя некоторые величины из когерентной и квантовой оптики, такие как модуль степени когерентности, автокорреляцию интенсивности и статистические свойства фотонов. Корректное оценивание сигналов, получаемых при помощи детектора требует привлечения теории фотодетектирования и включает в себя другую обратную задачу – восстановление статистических свойств падающего излучения из статических свойств фотоэлектронов.

Под априорной информацией здесь подразумеваются любые сведения о функциях источников до проведения эксперимента, а не регистрируемые данные, полученные по наблюдаемой реализации. Эту информацию можно получить из общих принципов, гипотез, результатов других экспериментов и естественных ограничений, обусловленных планируемой процедурой

эксперимента. Априорная информация является определяющей для единственности и устойчивости решения обратной задачи. Более того, природа априорной информации в значительной степени определяет характер задачи. Если априорной информации достаточно для построения хороших моделей источников, то можно надеяться на успех в решении обратной задачи методом перебора параметров при решении прямой задачи.

Обратные оптические задачи можно разделить на два класса:

- задачи, имеющие целью получение информации о пространственных изменениях функций источников (пространственно-частотных спектров), таких как профиль интенсивности или степень пространственной когерентности и другие пространственные корреляции;
- задачи, имеющие целью получение информации о временных изменениях, то есть динамике функций источников, или временных частотных спектрах, таких как спектральная плотность или степень временной когерентности и другие временные корреляции.

Другая возможная классификация обратных задач может быть основана на статистическом аспекте излучения. Так, известны обратные задачи в классическом переносе излучения, когерентной и квантовой оптике.

Некорректные задачи

Большинство обратных задач в фотонике являются также и некорректными, что затрудняет поиск их решения. Чтобы строго определить понятие некорректной задачи следует рассмотреть уравнение вида

$$A\mathbf{y} = \mathbf{f}, \qquad (4.1.6)$$

где $\mathbf{y}$ – искомое решение, $\mathbf{f}$ – заданная правая часть, A – заданный непрерывный оператор.

Задача решения уравнения вида (4.1.6) называется корректно поставленной, если выполняются следующие условия (условия корректности по Адамару):

1) решение существует;
2) решение единственно;
3) решение непрерывно зависит от данных (начальных и граничных условий, коэффициентов).

Если хотя бы одно из этих условий не выполняется, то задача является некорректной.

Примером невыполнения первого условия может служить переопределенная система линейных уравнений (СЛАУ), в которой

количество уравнений превышает количество неизвестных, а «лишние» уравнения не являются линейной комбинацией других уравнений СЛАУ.

Недоопределенная СЛАУ (случай, когда количество уравнений меньше количества неизвестных) в свою очередь является примером невыполнения второго условия, так как имеется бесконечное множество решений.

Примером невыполнения третьего условия может служить СЛАУ, малые изменения коэффициентов в которой приводят к большим изменениям получаемого решения. Задачи, решение которых незначительно изменяется при незначительных возмущениях входных данных, называются хорошо обусловленными.

В качестве меры обусловленности задачи обычно используют число обусловленности

$$\mathrm{cond}(\mathrm{A}) = \frac{\lambda_{\max}}{\lambda_{\min}}, \tag{4.1.7}$$

где $\lambda_{\max}$ и $\lambda_{\min}$ – соответственно, максимальное и минимальное собственное значение оператора A.

В общем случае для того, чтобы задача была корректной, необходимо выполнение следующих условий:

$$\mathrm{cond}(\mathrm{A}) = \| \mathrm{A} \| \| \mathrm{A}^{-1} \|, \tag{4.1.8}$$

$$\frac{\|\delta \mathbf{y}\|}{\|\mathbf{y}\|} \le \frac{\mathrm{cond}(\mathrm{A})}{1 - \mathrm{cond}(\mathrm{A})\dfrac{\|\delta \mathrm{A}\|}{\|\mathrm{A}\|}} \left(\frac{\|\delta \mathrm{A}\|}{\|\mathrm{A}\|} + \frac{\|\delta \mathbf{f}\|}{\|\mathbf{f}\|} \right), \tag{4.1.9}$$

где $\mathrm{A}\mathbf{y} = \mathbf{f}$ – основное уравнение, а $(\mathrm{A} + \delta\mathrm{A})\mathbf{y} = \mathbf{f} + \delta\mathbf{f}$ – уравнение с возмущениями.

Выполнение условий обусловленности задачи в фотонике необходимо, так как физические данные, как правило, определяются из эксперимента приближенно.

В зависимости от того, какое условие корректности задачи нарушено, используются различные методы решения таких задач:

1) если решение не существует, то применяется метод наименьших квадратов Гаусса, в результате чего получают псевдорешение;
2) если решение не единственно, то используется метод псевдообратной матрицы Мура-Пенроуза, в результате чего получают нормальное решение;
3) если решение неустойчиво, то применяются методы регуляризации (например, регуляризация Тихонова) и фильтрации (например, фильтр Винера).

Все эти методы рассмотрены ниже.

4.2 Решение обратных задач на основе метода наименьших квадратов

Классический метод наименьших квадратов

Рассмотрим случай нарушения первого условия корректности по Адамару. Пусть имеется система из m линейных алгебраических уравнений относительно n неизвестных:

$$\mathbf{A}\mathbf{y} = \mathbf{f}, \tag{4.2.1}$$

где A – матрица размера $m \times n$, причем $m > n$, а ранг дополненной матрицы $(\mathbf{A}\,|\,\mathbf{f})$ выше ранга матрицы **A**, то есть система является переопределенной, **y** – искомый вектор-столбец размера $n \times 1$, f – вектор-столбец свободных членов размера $m \times 1$.

Такая СЛАУ не имеет решений, то есть не существует такого **y**, для которого справедливо условие равенства невязки нулю

$$\left\|\mathbf{A}\mathbf{y} - \mathbf{f}\right\| = 0. \tag{4.2.2}$$

Для решения таких СЛАУ используют метод наименьших квадратов Гаусса, в котором вместо (4.2.2) используют условие

$$\left\|\mathbf{A}\overline{\mathbf{y}} - \mathbf{f}\right\| = \min_{\mathbf{y}}, \tag{4.2.3}$$

где $\overline{\mathbf{y}}$ – псевдорешение переопределенной СЛАУ.

Пусть в качестве нормы используется Евклидова норма. Минимизировать (4.2.3) можно путем приравнивания нулю производной нормы по **y**:

$$2(\mathbf{A}\overline{\mathbf{y}} - \mathbf{f})\mathbf{A} = 0, \tag{4.2.4}$$

или с учетом правил умножения матриц

$$\mathbf{A}^*(\mathbf{A}\overline{\mathbf{y}} - \mathbf{f}) = 0, \tag{4.2.5}$$

где $\mathbf{A}^*$ – эрмитово-сопряженная матрица для матрицы **A**. В случае, если все компоненты **A** вещественные, эрмитово-сопряженная матрица совпадает с транспонированной, то есть $\mathbf{A}^* = \mathbf{A}^{\mathrm{T}}$.

Выражение (4.2.5) можно переписать в виде

$$\mathbf{B}\overline{\mathbf{y}} = \mathbf{u}, \tag{4.2.6}$$

где

$$\mathbf{B} = \mathbf{A}^*\mathbf{A}, \tag{4.2.7}$$

$$\mathbf{u} = \mathbf{A}^*\mathbf{f}. \tag{4.2.8}$$

Полученная новая система (4.2.8) называется нормальной СЛАУ. Псевдорешение $\overline{\mathbf{y}}$ может быть найдено стандартными методами решения СЛАУ (4.2.6), например, в форме

$$\overline{\mathbf{y}} = \mathbf{B}^{-1}\mathbf{u}. \tag{4.2.9}$$

Рекуррентный метод наименьших квадратов

Классический метод наименьших квадратов (МНК) Гаусса традиционно используется для оценивания параметров сигнала по наблюдаемым значениям в соответствии с априорно заданной моделью.

Обработка данных при помощи метода наименьших квадратов использует детерминированное представление данных. К примеру, интерферометрический сигнал определяется моделью

$$s_k = B + A\cos(\varepsilon + 2\pi f_0 x_k), \tag{4.2.10}$$

где $\Phi_k = \varepsilon + 2\pi f_0 x_k$, ε – начальная фаза, f_0 – частота, $x_k = k\Delta x$ – значения независимой переменной, Δx – шаг дискретизации (который без потери общности рассмотрения можно принять равным единице), характеризуется вектором $\mathbf{y} = (B, A, \varepsilon, f_0)$, компоненты которого считаются постоянными, но не известными величинами, которые требуется определить.

Работа алгоритма состоит в уточнении априорной (грубой) оценки вектора параметров $\overline{\mathbf{y}} = (\overline{B}, \overline{A}, \overline{\varepsilon}, \overline{f}_0)$ по критерию минимизации суммы квадратов разностей между измеренными значениями интерферометрического сигнала и значениями, предсказанными в соответствии с моделью сигнала (4.2.10). Поправка к вектору параметров вычисляется по формуле

$$\Delta\mathbf{y} = \mathbf{y} - \overline{\mathbf{y}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\Delta\mathbf{s}, \tag{4.2.11}$$

где

$$\mathbf{A} = \begin{bmatrix} 1 & \cos\Phi_1 & -A_{pr}\sin\Phi_1 & -A_{pr}x_1\sin\Phi_1 \\ & & \cdots & \\ 1 & \cos\Phi_K & -A_{pr}\sin\Phi_K & -A_{pr}x_K\sin\Phi_K \end{bmatrix}, \tag{4.2.12}$$

представляет собой матрицу чувствительности (матрица производных модели сигнала (4.2.10) по параметрам), $\Delta\mathbf{s}$ – $(K\times 1)$ вектор-столбец разностей измеренных значений сигнала и значений, рассчитанных согласно модели (4.2.10) для вектора параметров $\overline{\mathbf{y}}$ (невязка), K – количество зарегистрированных отсчетов сигнала.

В рассмотренном выше классическом МНК для расчетов необходимо наличие полной реализации сигнала, что не всегда выполняется в оптических системах. В некоторых случаях необходимо осуществлять оценивание в процессе эксперимента, так как данные поступают постепенно.

Ниже рассмотрена рекуррентная реализация МНК на примере обработки данных в интерферометрических системах, в которых данные поступают на вход алгоритма последовательно (шаг за шагом).

В отличие от классического МНК, его рекуррентная реализация позволяет получать оценки вектора параметров для каждого

последующего отсчета K+1, если получены K предыдущих отсчетов сигнала. Для этого выражение (4.2.11) должно быть переписано в виде

$$\Delta\mathbf{y}(K) = [\mathbf{A}^{\mathrm{T}}(K)\mathbf{A}(K)]^{-1}\mathbf{A}^{\mathbf{T}}(K)\Delta\mathbf{s}(K). \qquad (4.2.13)$$

При поступлении следующего отсчета (K+1) в матрице (4.2.12) добавляется еще одна строка, а к вектору $\Delta\mathbf{s}$ – дополнительный элемент. Это может быть проиллюстрировано при помощи следующих соотношений:

$$\mathbf{A}(K+1) = \begin{bmatrix} \mathbf{A}(K) \\ \mathbf{v}^{\mathrm{T}}(K+1) \end{bmatrix}, \qquad (4.2.14)$$

где

$$\mathbf{y} = (1, \quad \cos\Phi_{K+1}, \quad -A\sin\Phi_{K+1}, \quad -Ax_{K+1}\sin\Phi_{K+1})^{\mathrm{T}}, \qquad (4.2.15)$$

$$\Delta\mathbf{s}(K+1) = \left[\Delta\mathbf{s}(K), \Delta s(K+1)\right]^{\mathrm{T}}, \qquad (4.2.16)$$

где $s(K+1)$ – поступивший дополнительньый отсчет сигнала.

Рекуррентная оценка $\Delta\mathbf{y}(K+1)$ вычисляется согласно (4.2.11) в форме

$$\Delta\mathbf{y}(K+1) = [\mathbf{A}^{\mathrm{T}}(K+1)\mathbf{A}(K+1)]^{-1}\mathbf{A}^{\mathbf{T}}(K+1)\Delta\mathbf{s}(K+1) =$$

$$= [\mathbf{A}^{\mathrm{T}}(K)\mathbf{A}(K) + \mathbf{v}^{\mathrm{T}}(K+1)\mathbf{y}(K+1)]^{-1} \times$$

$$\times [\mathbf{A}(K)\Delta\mathbf{s}(K) + \mathbf{v}^{\mathrm{T}}(K+1)\Delta\mathbf{s}(K+1)]. \qquad (4.2.17)$$

Это выражение можно переписать в виде

$$\Delta\mathbf{y}(K+1) = \mathbf{y}(K) + \mathbf{P}(K)[\Delta\mathbf{s}(K+1) - \mathbf{v}^{\mathrm{T}}(K+1)\Delta\mathbf{y}(K+1)], \qquad (4.2.18)$$

где выражение в квадратных скобках есть разность между прогнозированным (модельным) и зарегистрированным значениями сигнала, а $\mathbf{P}(K)$ – вектор весовых коэффициентов (“коэффициентов усиления”), который представляется в форме

$$\mathbf{P}(K) = [\mathbf{A}^{\mathrm{T}}(K+1)\mathbf{A}(K+1)]^{-1}\mathbf{v}^{\mathrm{T}}(K+1). \qquad (4.2.19)$$

Поскольку справедливо соотношение

$$\mathbf{A}^{\mathrm{T}}(K)\mathbf{A}(K) = \sum_{k=1}^{K}\mathbf{v}^{\mathrm{T}}(k)\mathbf{v}(k), \qquad (4.2.20)$$

левая часть (4.2.17) оказывается вырожденной при малых K, что может вести к большим вычислительным погрешностям. Вследствие этого начало рекуррентной процедуры должно соответствовать устойчивым значениям $\mathbf{P}(K)$, начиная с K_1, при этом

$$\Delta\overline{\mathbf{y}}(K_1) = [\mathbf{A}^{\mathrm{T}}(K_1)\mathbf{A}(K_1)]^{-1}\mathbf{A}^{\mathrm{T}}(K_1)\Delta\mathbf{s}(K_1). \qquad (4.2.21)$$

Точность априорного приближения имеет прямое влияние на погрешность оценивания параметров при помощи рекуррентного метода наименьших квадратов, что является ограничением возможности применения данного метода для решения задач с неизвестными начальными условиями и неизвестным законом изменения параметров сигнала.

4.3 Метод псевдообратной матрицы Мура-Пенроуза

Метод псевдооратной матрицы Мура-Пенроуза применяется в случае нарушения второго условия корректности задача по Адамару.

Пусть имеется недоопределенная СЛАУ

$$\mathbf{A}\mathbf{y} = \mathbf{f}, \tag{4.3.1}$$

где $\mathbf{A}$ – матрица размера $m \times n$, $\mathbf{y}$ – искомый вектор размера n, $\mathbf{f}$ – вектор размера m, причем $m < n$. Для такой СЛАУ существует множество решений.

Такая СЛАУ может быть решена методом псевдообратной матрицы Мура-Пенроуза [4.7], который позволяет отыскать среди множества решений решение с минимальной нормой

$$\|\bar{\mathbf{y}}\| = \min_{\mathbf{y}}. \tag{4.3.2}$$

Такое решение принято называть нормальным решением СЛАУ. Оно является единственным и его можно найти при помощи выражения

$$\bar{\mathbf{y}} = \mathbf{A}^{+}\mathbf{f}, \tag{4.3.3}$$

где $\mathbf{A}^{+}$ – псевдообратная матрица Мура-Пенроуза, которая удовлетворяет условию

$$\mathbf{A}\mathbf{A}^{+}\mathbf{A} = \mathbf{A}. \tag{4.3.4}$$

Асимптотически она может быть найдена с помощью выражения

$$\mathbf{A}^{+} = \lim_{\alpha \to 0} (\alpha \mathbf{E} + \mathbf{A}^{*}\mathbf{A})^{-1}\mathbf{A}^{*}. \tag{4.3.5}$$

Поиск псевдообратной матрицы при помощи формулы (4.3.5) является в большинстве случаев неудобным и трудоемким. На практике пользуются итерационным алгоритмом поиска таких матриц. Он состоит в последовательном вычислении приближения обратной матрицы при помощи формулы

$$\mathbf{A}_{i+1} = 2\mathbf{A}_i - \mathbf{A}_i\mathbf{A}\mathbf{A}_i. \tag{4.3.6}$$

Начальная матрица $\mathbf{A}_0$ определяется выражением

$$\mathbf{A}_i\mathbf{A} = (\mathbf{A}_0\mathbf{A})^{*} \tag{4.3.7}$$

или

$$\mathbf{A}_0 = \alpha\mathbf{A}^{\mathrm{T}}. \tag{4.3.8}$$

Последовательность матриц $\mathbf{A}_i$ сходится к истинной псевдообратной матрице $\mathbf{A}^{+}$ только в случае, когда α лежит в интервале

$$0 < \alpha < 2/\sigma_1^2(\mathbf{A}), \tag{4.3.9}$$

где $\sigma_1(\mathbf{A})$ – это наибольшее сингулярное значение матрицы $\mathbf{A}$, которое может быть получено при помощи сингулярного (SVD) разложения.

4.4 Регуляризация Тихонова

Метод регуляризации Тихонова [4.7] является обобщением рассмотренных выше метода наименьших квадратов Гаусса и метода псевдообратной матрицы Мура-Пенроуза. Пусть имеется СЛАУ

$$\mathbf{A}\mathbf{y} = \mathbf{f}, \tag{4.4.1}$$

где $\mathbf{A}$ – матрица, описывающая линейный непрерывный оператор, $\mathbf{y}$ – иском решение, $\mathbf{f}$ – заданная правая часть. При этом вместо точных значений $\mathbf{A}$ и $\mathbf{f}$ известны только их приближения $\overline{\mathbf{A}}$ и $\overline{\mathbf{f}}$, удовлетворяющие условиям

$$\left\|\overline{\mathbf{A}} - \mathbf{A}\right\| \le \mu, \tag{4.4.2}$$

$$\left\|\overline{\mathbf{f}} - \mathbf{f}\right\| \le \delta, \tag{4.4.3}$$

где μ и δ – верхние границы погрешностей оператора и правой части. Таким образом, (4.4.1) можно переписать в виде

$$\overline{\mathbf{A}}\overline{\mathbf{y}} = \overline{\mathbf{f}}, \tag{4.4.4}$$

где $\overline{\mathbf{y}}$ – оценка искомого решения.

Метод регуляризации Тихонова реализует выполнение двух условий: минимизации невязки вида (4.2.3) и минимизации нормы решения вида (4.3.2).

Полученная задача условной минимизации решается методом неопределенных множителей Лагранжа, в результате чего можно получить уравнение

$$\left\|\overline{\mathbf{A}}\overline{\mathbf{y}} - \overline{\mathbf{f}}\right\|^2 + \lambda\left\|\overline{\mathbf{y}}\right\|^2 = \min_{\overline{\mathbf{y}}}, \tag{4.4.5}$$

где $\lambda > 0$ – параметр регуляризации, играющий роль неопределенного множителя Лагранжа.

При нарушении третьего условия корректности по Адамару условие (4.4.5) приводит к уравнению Тихонова

$$(\alpha\mathbf{E} + \overline{\mathbf{A}}^*\overline{\mathbf{A}})^{-1}\overline{\mathbf{y}}_\lambda = \overline{\mathbf{A}}^*, \tag{4.4.6}$$

где $\mathbf{E}$ – единичная матрица, $\mathbf{A}^*$ – эрмитово-сопряженная матрица для матрицы $\mathbf{A}$.

Из (4.4.5) видно, что при $\lambda = 0$ метод регуляризации Тихонова эквивалентен методу наименьших квадратов Гаусса, то есть находится решение, с минимальной невязкой. С увеличением параметра α решение становится более гладким и устойчивым. На практике обычно используется некоторое компромиссное решение между гладким решением и решением, с малой невязкой.

Методы выбора параметра регуляризации

Наиболее популярными методами выбора параметра регуляризации являются метод невязки и метод подбора. Первый из них заключается в выборе параметра регуляризации при условии точно известного оператора **A** или равенства нулю уравнения (4.4.2). В этом случае (4.4.3) можно переписать в форме

$$\left\|\mathbf{A}\bar{\mathbf{y}}_\lambda - \mathbf{f}\right\| = \delta. \qquad (4.4.7)$$

Если $\left\|\mathbf{f}\right\| \geq \delta$, то решение уравнения (4.4.5) относительно λ существует и является единственным.

На практике часто для поиска параметра регуляризации используется метод подбора, который заключается в переборе большого количества значений λ до тех пор, пока не будет выполнено условие

$$\min_{\bar{\mathbf{y}}} - \alpha \leq \varphi(\lambda) \leq \min_{\bar{\mathbf{y}}} + \alpha, \qquad (4.4.8)$$

где

$$\varphi(\lambda) = \left\|\overline{\mathbf{A}}\bar{\mathbf{y}} - \bar{\mathbf{f}}\right\|^2 + \lambda\left\|\bar{\mathbf{y}}\right\|^2, \qquad (4.4.9)$$

α – коэффициент, задающий требуемую точность выполнения условия (4.4.5), При $\alpha = 0$ полученное решение будет оптимальным.

Если (4.4.5) монотонна, то для подбора параметра регуляризации можно использовать метод градиентного спуска, который заключается в следующем:

- задание начального значения параметра регуляризации λ;
- вычисление значения функции (4.4.5);
- до тех пор, пока условие (4.4.5) не выполнено, увеличивать параметр регуляризации на некоторую величину, если $\varphi(\lambda) < \min_{\bar{\mathbf{y}}} - \alpha$, и уменьшать, его, если $\varphi(\lambda) > \min_{\bar{\mathbf{y}}} + \alpha$.

Для подбора параметра регуляризации могут быть использованы и другие методы оптимизации, например, метод касательных Ньютона.

Регуляризация для решения уравнений типа свертки

Пусть имеется интегральное уравнение типа свертки

$$\mathrm{A}\{y(x)\} \equiv \int_{-\infty}^{\infty} K(x-s)y(s)ds = f(x). \qquad (4.4.10)$$

где A представляет собой интегральный оператор свертки с ядром $K(x)$, $x \in (-\infty, \infty)$. Условие минимизации сглаживающего функционала для уравнения (4.4.10) может быть записано как

$$\int_{-\infty}^{\infty}[\mathrm{A}\{y(x)\}-f(x)]^2\,dx+\alpha\int_{-\infty}^{\infty}M(w)\left|Y(w)\right|^2dw=\min_{y}. \qquad (4.4.11)$$

где $Y(w)$ – результат преобразования Фурье функции $y(x)$

$$Y(w)=\mathrm{F}\{y(x)\}=\int_{-\infty}^{\infty}y(x)\exp(-jwx)dx, \qquad (4.4.12)$$

где F{.} – оператор преобразования Фурье. Член $M(w)$ в (4.4.11) – это регуляризатор q-го порядка, определяемый как

$$M(w)=\left|w\right|^{2q}. \qquad (4.4.13)$$

Условие (4.4.11) приводит к следующему выражению решения

$$\bar{y}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\lambda(-w)F(w)}{\left|\lambda(w)\right|^2+\alpha M(w)}\exp(jws)dw=\mathrm{F}^{-1}\left\{\frac{\lambda(-w)F(w)}{\left|\lambda(w)\right|^2+\alpha M(w)}\right\}, \qquad (4.4.14)$$

где $\lambda(w)$ и $F(w)$ – результаты преобразования Фурье ядра свертки $K(x)$ и функции $f(x)$ соответственно

$$\lambda(w)=\mathrm{F}\{K(x)\}=\int_{-\infty}^{\infty}K(x)\exp(-jwx)dx, \qquad (4.4.15)$$

$$F(w)=\mathrm{F}\{f(x)\}=\int_{-\infty}^{\infty}f(x)\exp(-jwx)dx. \qquad (4.4.16)$$

Для нахождения решения $\bar{y}(x)$ на практике пользуются численными методами, основанными на замене интегралов в (4.4.11)–(4.4.16) конечными суммами и переходе тем самым от непрерывного преобразования Фурье к дискретному и дальнейшему использованию алгоритмов быстрого преобразования Фурье.

Связь регуляризации Тихонова и оптимальной фильтрации Винера

Пусть имеется интегральное уравнение типа свертки, аналогичное уравнению (4.4.10)

$$\int_{-\infty}^{\infty}K(x-s)y(s)ds=f(x)+v(x), \qquad (4.5.1)$$

где $v(x)$ – случайная погрешность правой части с известными статистическими характеристиками.

Для применения фильтрации Винера необходимо сделать ряд предположений:

- искомая функция $y(s)$ и погрешность решений $v(x)$ являются реализациями некоррелированных стационарных случайных процессов;
- спектральные плотности мощности этих процессов известны и задаются как

$$S_y(w) = \lim_{T\to\infty} \frac{1}{2T} E\left[\left|\int_{-T}^{T} y(x)\exp(-jwx)dx\right|^2\right], \quad (4.5.2)$$

$$S_v(w) = \lim_{T\to\infty} \frac{1}{2T} E\left[\left|\int_{-T}^{T} v(x)\exp(-jwx)dx\right|^2\right], \quad (4.5.3)$$

где $S_y(w)$ и $S_v(w)$ – спектральные плотности мощности исходной функции и погрешности решений. В отличие от регуляризации Тихонова в оптимальном линейном фильтре Винера решение ищется исходя из условия минимизации среднего квадратического отклонения результата фильтрации от точного решения

$$E[\bar{y}(x) - y(x)]^2 = \min_y, \quad (4.5.4)$$

где $y(x)$ – точное решение, а $\bar{y}(x)$ – результат фильтрации, который можно записать как

$$\bar{y}(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\lambda(-w)F(w)}{|\lambda(w)|^2 + S_v(w)/S_y(w)}\exp(jws)dw = ,$$

$$= \mathrm{F}^{-1}\left\{\frac{\lambda(-w)F(w)}{|\lambda(w)|^2 + S_v(w)/S_y(w)}\right\}. \quad (4.5.5)$$

Среднее квадратическое отклонение (4.5.4) можно записать как

$$E[\bar{y}(x) - y(x)]^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S_v(w)}{|\lambda(w)|^2 + S_v(w)/S_y(w)}\exp(jwx)dw. \quad (4.5.6)$$

Фильтр Винера называется оптимальным, так как значений (4.5.6) является минимально возможным.

Несложно заметить, что фильтр Винера переходит в метод регуляризации Тихонова при выполнении условия

$$\alpha M(w) = S_v(w)/S_y(w). \quad (4.5.7)$$

Устойчивость фильтрации Винера растет с увеличением отношения помехи к сигналу $S_v(w)/S_y(w)$, то есть увеличение помехи ведет к стабилизации фильтра. Напротив, увеличение помехи ведет к увеличению (4.5.6.), что позволяет сделать вывод о том, что в фильтре Винера имеет место компромисс между устойчивостью решения и его точностью.

Функцию спектральной плотности мощности помехи $S_v(w)$ можно получить при помощи многократных измерений, что, однако, неверно для спектральной плотности мощности сигнала $S_y(w)$. Таким образом, фильтрация Винера является теоретическим методом, позволяющим найти оптимальное решение для уравнения вида (4.5.1).

Контрольные вопросы

1. В чем заключается сущность прямой и обратной задач применительно к проблематике бесконтактного контроля объектов?
2. Сформулируйте условия некорректности обратной задачи по Адамару.
3. В чем состоят достоинства и недостатки рекуррентного метода наименьших квадратов по сравнению с классическим методом?
4. Объясните условия использования регуляризации решения обратной задачи методом Тихонова.

Список литературы

4.1. Болтс Г.П. Обратные задачи в оптике. – М.: Машиностроение, 1984.

4.2. Фриден Б. Компьютеры в оптических исследованиях. – М.: Мир, 1983.

4.3. Гуров И.П. Компьютерная обработка видеоинформации. Методы обработки неподвижных изображений. Учебное пособие. – СПб.: БХВ, 1998.

4.4. Сизиков В.С. Устойчивые методы обработки результатов измерений. Учебное пособие. – СПб: «СпецЛит», 1999. – 240 с.

4.5. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005.

УНИВЕРСИТЕТ ИТМО

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

КАФЕДРА КОМПЬЮТЕРНОЙ ФОТОНИКИ И ВИДЕОИНФОРМАТИКИ

Достижения в оптической науке, технике и технологиях за последние годы способствовали появлению нового направления – фотоники. Этот термин охватывает область науки и техники, связанную с использованием светового излучения (или потока фотонов) в оптических элементах, устройствах и системах.

На рубеже XX – XXI веков электронные информационные технологии достигли фундаментальных и технических пределов производительности при продолжающемся росте потребительского спроса на скорость и объем обрабатываемой и передаваемой информации. Решение данной проблемы потребовало разработки нового поколения информационно – телекоммуникационных систем, основанных на технологиях фотоники. В фотонике появилось новое динамично развивающееся направление, определяющее прогресс мировой науки и техники, – «оптоинформатика». Под «оптоинформатикой» понимают область науки и техники, связанную с исследованием, разработкой, созданием и эксплуатацией новых материалов, технологий, приборов и устройств, направленных на передачу, прием, обработку, хранение и отображение информации.

Изучение фотоники основывается на знании принципов формирования, преобразования, анализа изображений, теории построения информационных систем. Интеграция фотоники и компьютерных технологий позволяет создавать методы, которые возможно реализовать исключительно средствами компьютерной фотоники, обеспечивая развитие технологий качественно нового уровня.

По многим направлениям фотоники и оптоинформатики Россия находится на уровне промышленно – развитых стран (интегральная оптика, системы приема, обработки и отображения информации и др.), а, по некоторым – даже опережает. Приоритетными направлениями являются: волоконная оптика (работы академика Дианова Е.М. – ИОФ РАН), голография (академик Денисюк Ю.Н. – ГОИ им. С.И. Вавилова),

полупроводниковые лазеры (академик Алферов Ж.И – ФТИ РАН им. А.Ф. Иоффе), полифункциональные оптические материалы (академик Петровский Г.Т. – ГОИ им. С.И. Вавилова) и др.

Ввиду большого научного и практического значения направления "Фотоника и оптоинформатика", а также спроса на него на потребительском рынке, в 2002 г. в СПбГУ ИТМО был организован факультет «Фотоники и оптоинформатики» под руководством доктора физ.-мат. наук, профессора С.А. Козлова. По инициативе профессорско-преподавательского состава, начиная с 2005 года, на факультете стала работать выпускающая кафедра «Компьютерной фотоники», которую возглавил доктор технических наук, профессор И.П. Гуров.

История кафедры началась в 1946 году. На всех этапах развития результаты научных исследований, проводимых сотрудниками кафедры, неизменно использовались в учебном процессе. Совершенствовались направления подготовки студентов, изменялось название кафедры, но всегда кафедра гордилась своими выпускниками.

Выпускники кафедры занимают видное место в оптической науке: академик РАН Ю.Н. Денисюк, изобретатель трехмерной голографии; член-корр. РАЕН, профессор Н.Г. Бахшиев, известный специалист в области спектроскопии межмолекулярных взаимодействий; Заслуженный деятель науки РФ, профессор Г.Н. Дульнев, крупный ученый в области теплофизики, долгие годы бывший ректором ЛИТМО; профессор И.М. Нагибина, исследования которой в области физической оптики получили широкое признание.

Одной из важнейших задач кафедры является организация учебного процесса и подготовка профессионалов в области компьютерной фотоники. Направление работы кафедры определяется развитием информационных технологий и компьютерных систем в области формирования, синтеза, обработки и анализа изображений на основе интеграции эффективных компьютерных систем с системами фотоники.

Проводимые исследования в области компьютерной обработки когерентных и некогерентных изображений обеспечивают решение научно-технических задач оптической томографии, цифровой голографии, синтеза, анализа, распознавания и классификации изображений.

Научным консультантом работ кафедры в области компьютерной обработки изображений – иконики – является член-корреспондент РАН М.М. Мирошников.

Кафедра проводит работы в рамках международных научных проектов в сотрудничестве с ведущими зарубежными университетами, институтами и исследовательскими лабораториями Италии, Финляндии, Франции, Германии, Великобритании, Японии, США и других стран в области оптической когерентной томографии для биомедицинских исследований, цифровой голографии для исследования микро- и

наноструктур, трехмерной фотографии микро- и макроскопических объектов, гиперспектральной обработки изображений.

В последнее время на кафедре активно развивается новое направление – видеоинформатика. Ввиду этого в 2010 году кафедра была переименована в кафедру Компьютерной фотоники и видеоинформатики. В 2011 году Университет получил статус Национального исследовательского университета, в этом есть и заслуга преподавательского коллектива кафедры Компьютерной фотоники и видеоинформатики. Эти обстоятельства позволяют обеспечивать и в дальнейшем подготовку высококлассных востребованных на рынке специалистов в области Компьютерной фотоники и видеоинформатики.

Гуров Игорь Петрович

ФОРМИРОВАНИЕ И АНАЛИЗ СИГНАЛОВ В СИСТЕМАХ КОМПЬЮТЕРНОЙ ФОТОНИКИ

Учебно-методическое пособие

В авторской редакции
Редакционно-издательский отдел Университета ИТМО
Зав. РИО Н.Ф. Гусарова
Подписано к печати
Заказ №
Тираж
Отпечатано на ризографе