

МЕТОДИЧЕСКИЕ ЗАМЕТКИ

535+519.876.5+621.373.826

О ФАЗОВОЙ ПРОБЛЕМЕ В ОПТИКЕ***Т. И. Кузнецова*****СОДЕРЖАНИЕ**

1. Введение	677
2. Алгоритм Гершберга — Сэкстона и другие близкие к нему методы решения фазовой задачи	677
3. Алгоритм для восстановления фазовых характеристик, моделирующий действие оптической схемы	679
4. Математические вопросы, связанные с фазовой проблемой	681
5. Некоторые приложения фазовой задачи. Киноформ	685
6. Восстановление временных характеристик излучения	686
7. Заключение	689
Список литературы	689

1. ВВЕДЕНИЕ

Проблема получения фазовых характеристик световых полей возникает в разнообразных оптических исследованиях. Трудности непосредственного измерения фазы в оптическом диапазоне заставляют оптиков искать обходные пути: пытаться извлекать фазовую информацию из данных об интенсивности. Разумеется, попытки получить тут простые рецепты решения задачи были обречены на неудачу. Однако за последние 15 лет наметилось серьезное продвижение в проблеме восстановления фазовых характеристик световых полей. Появились работы, в которых с помощью вычислительной техники реально выполняется восстановление фазовых характеристик по характеристикам интенсивности. Важно подчеркнуть, что при этом привлекаются дополнительные данные о поле, например, во многих случаях используются два (а не одно) распределения интенсивности, относящиеся к двум сечениям поля.

В той же мере, как об успехах в оптике, можно было бы говорить о прогрессе в обработке данных рентгеноструктурного анализа, электронной микроскопии, однако мы здесь ограничимся оптическими задачами.

Начало этому новому направлению исследований фазовой проблемы положила работа ¹. С состоянием фазовой проблемы в предшествующий период можно ознакомиться, например, по работам ²⁻⁴. Ниже мы переходим к изложению работы ¹.

**2. АЛГОРИТМ ГЕРШБЕРГА — СЭКСТОНА
И ДРУГИЕ БЛИЗКИЕ К НЕМУ МЕТОДЫ РЕШЕНИЯ ФАЗОВОЙ ЗАДАЧИ**

Для восстановления фазовой информации Гершберг и Сэкстон первыми стали привлекать данные об интенсивности, относящиеся не к одной, а к двум плоскостям. Особенно удачным у них оказался итерационный способ нахождения фазы ¹, который заключается в следующем.

Допустим, что исходное монохроматическое поле, несущее информацию об исследуемом объекте, было зарегистрировано в двух плоскостях: в плоскости изображения и в фурье-плоскости. Будем считать (в соответствии с исходными предположениями работы¹⁾, что поле зависит только от одной пространственной координаты. Обозначим поле в плоскости изображения через $\mathcal{E}(x) = a(x) \exp(i\varphi(x))$, в фурье-плоскости — через $\mathcal{F}(x) = A(x) \exp(i\Phi(x))$. Для операции преобразования Фурье введем символ

$$L\mathcal{E}(x) = \frac{1}{\sqrt{2\pi}} \int \mathcal{E}(x') \exp(-ixx') dx',$$

так что

$$F(x) = L\mathcal{E}(x).$$

Пусть при регистрации поля получена информация о функциях $a(x)$ и $A(x)$, а информация о фазовых множителях отсутствует. Задача состоит в построении комплексной функции по заданному ее модулю и модулю ее фурье-образа. Для решения этой задачи Гершберг и Сэкстон предложили итерационную процедуру, которая заключалась в следующем. В качестве пробной функции бралась функция, модуль которой совпадал с заданным в плоскости изображения модулем $a(x)$, а фазовый множитель $\exp(i\psi(x))$ брался произвольным; в работе он строился с помощью генератора случайных чисел. Пробную функцию удобно обозначить через $y_0(x)$, в дальнейшем индекс будет совпадать с числом выполненных итераций, перед началом итераций имеем

$$y_0(x) = a(x) \exp(i\psi(x)).$$

Для этой функции строилось ее преобразование Фурье

$$Ly_0 = A_0(x) \exp(i\Phi_0(x)).$$

Затем у полученной функции модуль заменялся на правильный, т. е. на $A(x)$, а фаза сохранялась. Так получали функцию в фурье-плоскости

$$Y_0(x) = A(x) \exp(i\Phi_0(x)),$$

что можно представить как

$$Y_0 = \frac{ALy_0}{|Ly_0|}.$$

Затем выполнялось обратное преобразование Фурье и снова исправлялся модуль, при этом получалась функция в плоскости изображения в первом приближении

$$y_1 = \frac{aL^{-1}Y_0}{|L^{-1}Y_0|}.$$

На следующем шаге выполнялись такие же преобразования. Вся итерационную процедуру можно представить формулами

$$Y_n = \frac{ALy_n}{|Ly_n|}, \quad y_{n+1} = \frac{aL^{-1}Y_n}{|L^{-1}Y_n|}. \quad (1)$$

Если процесс окажется сходящимся, то функцию, получающуюся в качестве предела итераций, будем обозначать через y

$$y = \lim_{n \rightarrow \infty} y_n.$$

Во многих случаях процесс, действительно, сходил, и функция y совпадала с исходным комплексным полем $a \exp(i\varphi)$. Иллюстративный материал мы приведем в п. 3, где будет изложен наш алгоритм, имеющий много общего с алгоритмом Гершберга — Сэкстона. Отметим, что в методе Гершберга — Сэкстона не гарантировалось ни существования решения, ни его единственности. Тем не менее многочисленные удачи в численном эксперименте при-

вели к тому, что алгоритм стали широко использовать другие авторы. Более того, алгоритм Гершберга — Сэкстона стали применять к несколько иным задачам, возникли его модификации, обусловленные как измененной постановкой задач, так и стремлением ускорить численную процедуру. Остановимся на некоторых из этих модификаций.

Прежде всего здесь следует назвать алгоритм Финапа, предназначенный для восстановления вещественной неотрицательной функции по модулю ее фурье-спектра⁵. Подчеркнем, что в отличие от алгоритма Гершберга — Сэкстона, здесь не требуется знания модуля поля в объектной плоскости, т.е. функции $a(x)$. Особенности итерационной процедуры состоят в том, что значение функции, полученное на очередном шаге, при пересчете от фурье-плоскости к плоскости изображения умножается на нуль, если оно не удовлетворяет условию неотрицательности, и оставляется неизменным в противном случае. Оказывается, что условие неотрицательности является столь сильным, что с его использованием восстановление в ряде случаев проходит не хуже, чем восстановление комплексных функций по модулям, заданным в двух плоскостях. Следует отметить, что алгоритм Финапа успешно применялся не только к одномерным, но и к двумерным функциям^{6,7}. Хотя, на первый взгляд, двумерные задачи представляются более сложными, чем одномерные, и при вычислениях для них на каждом шаге итераций необходим больший объем вычислений, но в двумерном случае алгоритмы работают более эффективно. К особенностям двумерных задач мы вернемся в п. 4. Сравнительно недавно в литературе появилось сообщение о восстановлении по модулю фурье-спектра двумерных функций, на которые не налагается условия неотрицательности, т.е. функции могут быть комплексными и в плоскости изображения, и в фурье-плоскости⁸.

3. АЛГОРИТМ ДЛЯ ВОССТАНОВЛЕНИЯ ФАЗОВЫХ ХАРАКТЕРИСТИК, МОДЕЛИРУЮЩИЙ ДЕЙСТВИЕ ОПТИЧЕСКОЙ СХЕМЫ

Поставим вопрос, нельзя ли процедуру, описанную в п. 2, выполнить с помощью оптических преобразований. Операция преобразования Фурье, как известно, легко осуществляется в оптике с помощью линзы. Умножение на функции (на $a(x)$ и $A(x)$) легко выполнить, если эти функции представлены в виде пропускания транспарантов. Что же касается операции выравнивания модуля комплексной амплитуды, то ее можно произвести лишь приближенно. Для этого нужно использовать нелинейный элемент типа затемняющегося фильтра (например, элемент с двухфотонным поглощением или с генерацией ВКР либо усилитель с насыщающимся усилением). Схема оптического устройства для восстановления фазы, объединяющего указанные элементы и транспаранты в кольцевом резонаторе, была описана нами в^{9,10}. Алгоритм, соответствующий действию этого оптического устройства, задается формулой

$$y_{n+1} = \beta L^{-1} A N (\beta L a N (y_n)); \quad (2)$$

здесь y_n и y_{n+1} — поля на последовательных итерациях, β — константа, означающая коэффициент усиления, функции a и A и оператор L определены в п. 2. Через N обозначен нелинейный оператор, который сглаживает распределение интенсивности, оставляя фазовое распределение неизменным, имеет вид

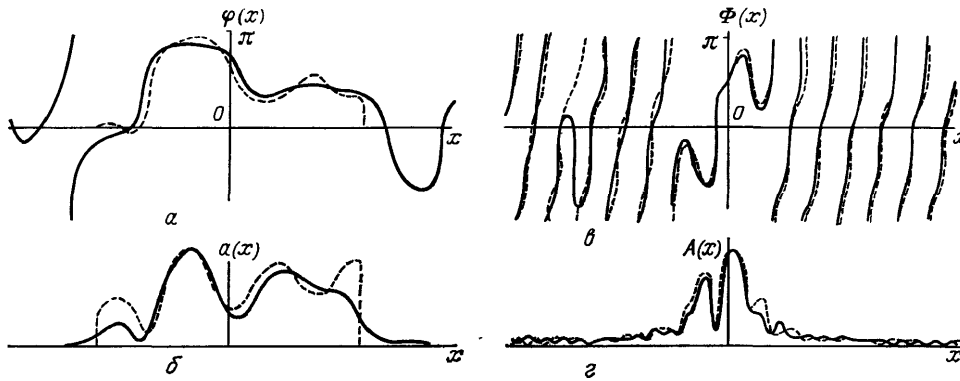
$$N(y) = \frac{y}{(1 + yy^*)^{1/2}}. \quad (3)$$

Такое нелинейное преобразование соответствует нелинейному элементу, потери в котором обусловлены двухфотонным поглощением или генерацией второй оптической гармоники.

Следует обратить внимание, что при очень сильных полях ($|y|^2 \gg 1$) имеем $N(y) = y/|y|$ и алгоритм (2) переходит в алгоритм (1). В наших расчетах мы начинали с малых интенсивностей, при которых нелинейность не изменяла поле, т. е. в начале расчетов мы имели $N(y) \approx y$. В зависимости от коэффициента β либо интенсивность поля при итерациях нарастала, что приводило к «включению» нелинейности, либо интенсивность уменьшалась от шага к шагу, так что нелинейность не действовала в ходе всей итерационной процедуры. В этом последнем случае, который мы назвали допороговым, за 8–10 итераций достигалось устойчивое распределение поля по координате x , а среднее значение затухало от итерации к итерации с постоянной скоростью. В допороговом режиме установившееся распределение $y(x)$ и исходное поле $a(x) \exp(i\varphi(x))$ имели интеграл перекрытия $\eta = 0,75–0,90$. Интеграл перекрытия определяется формулой

$$\eta = \frac{|\langle y a \exp(-i\varphi) \rangle|}{(\langle y y^* \rangle \langle a^2 \rangle)^{1/2}},$$

в которой угловые скобки означают усреднение по координате x . Установившееся распределение не зависело от пробной функции, с которой начинались



Пример восстановления комплексного поля по двум модулям.

Зависимость фазы и модуля поля от координаты: $a, б$ — в плоскости изображения, $в, г$ — в фурье-плоскости. Штриховые линии относятся к исходной функции, сплошные — к восстановленной.

расчеты. Недостатком схемы являлось невысокое значение интеграла перекрытия. Заметно лучшее качество восстановления достигалось в нелинейном (или надпороговом) режиме. В этом случае, который осуществлялся при больших значениях коэффициента β , происходило следующее. В самом начале расчетов вместе с изменением структуры поля росло среднее значение интенсивности. Затем, после нескольких итераций, включалась нелинейность, рост интенсивности замедлялся и, наконец, устанавливалось стационарное значение интенсивности и устойчивое распределение поля по x . При этом получалось значительно более высокое значение интеграла перекрытия функций $y(x)$ и $a(x) \exp(i\varphi(x))$, чем в допороговом режиме. Обычно — без сколько-нибудь тщательного подбора значений β и $\langle y_0 y_0^* \rangle$ — легко достигались значения интеграла перекрытия $\eta = 0,997$, они устанавливались за 15–20 итераций. В работе¹ необходимое число итераций было порядка ста. Так же, как в¹, мы восстанавливали одномерные комплексные функции. Исходные поля строились с помощью генератора случайных чисел и операций сглаживания по координате. Пример объектного поля и результата восстановления (в надпороговом режиме) приведен на рисунке. Высокая скорость сходимости нашего алгоритма достигается за счет линейной стадии, на которой быстро формируется приближенное решение, которое затем уточняется в ходе нелинейной стадии.

4. МАТЕМАТИЧЕСКИЕ ВОПРОСЫ, СВЯЗАННЫЕ С ФАЗОВОЙ ПРОБЛЕМОЙ

Параллельно с работами по вычислительным методам решения фазовой задачи в литературе публиковалось большое количество работ, посвященных аналитическим методам исследования проблемы.

Наиболее полное изложение полученных здесь результатов можно найти в ^{11,12}. Ниже мы приведем некоторые из них. Мы изложим современные представления о неоднозначности решения фазовой задачи, т.е. о том, с какой степенью неопределенности восстанавливается распределение фазы по одному распределению интенсивности. Начало исследований этого вопроса восходит к работам ^{13,14}. Современное состояние исследований имеется в ^{11,12}. По существу, мы будем следовать работам ^{11,12}, однако дадим упрощенное изложение вопроса, не требующее обращения к аппарату теории аналитических функций. Упрощения возникают за счет того, что берутся комплексные поля менее общего вида, чем в ^{11,12}. В известной мере наш подход близок к подходу, использованному в работе ¹⁵.

Будем рассматривать монохроматические поля, передающие информацию о каких-либо стационарных объектах; информация заключена в поперечной структуре полей. Прежде всего следует подчеркнуть, что поля, которые изучаются в оптике, принадлежат определенному классу функций — функций с ограниченным спектром. Каков бы ни был объект, изображение которого мы исследуем, каков бы ни был истинный спектр его пространственных частот, — из-за конечных апертур оптических элементов в формировании изображения всегда участвует только конечная спектральная полоса. И даже в отсутствие апертурных ограничений возникают ограничения, связанные с фильтрующими свойствами свободного пространства (высшие пространственные гармоники являются затухающими волнами и переносимая ими информация исчезает при удалении от объекта на расстоянии, больше длины волны излучения). Пусть поле монохроматического источника освещает некоторый объект конечной протяженности. Обозначим поле в плоскости, лежащей в непосредственной близости от объекта через $\mathcal{E}_0(x_0)$, а фурье-образ функции $\mathcal{E}_0(x_0)$ — через $\mathcal{F}_0(\xi)$,

$$\mathcal{F}_0(\xi) = \frac{1}{\sqrt{2\pi}} \int \mathcal{E}_0(x_0) \exp(-i\xi x_0) dx_0. \quad (4)$$

Как уже было сказано выше, в формировании изображения участвует лишь часть истинного спектра, имеющая конечный носитель. Эту часть спектра мы обозначим через $\mathcal{F}(\xi)$. Будем считать, что

$$\begin{aligned} \mathcal{F}(\xi) &= \mathcal{F}_0(\xi) \quad \text{при } |\xi| < b, \\ &= 0 \quad \text{при } |\xi| > b. \end{aligned} \quad (5)$$

Изображение, сформированное этим спектром, обозначим $\mathcal{E}(x)$.

$$\mathcal{E}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{F}(\xi) \exp(ix\xi) d\xi = \frac{1}{\sqrt{2\pi}} \int_{-b}^b \mathcal{F}(\xi) \exp(ix\xi) d\xi. \quad (6)$$

Теперь мы сделаем основное упрощающее предположение. Допустим, что спектральная функция $\mathcal{F}_0(\xi)$ может быть представлена на отрезке $[-b, b]$ конечным числом фурье-гармоник:

$$\mathcal{F}_0(\xi) = \sum_{m=-n/2}^{n/2} G_m e^{im(\pi/b)\xi}, \quad (7)$$

где G_m — произвольные комплексные константы. При этом спектральная функция изображения, $\mathcal{F}(\xi)$, связанная с $\mathcal{F}_0(\xi)$ соотношением (5), описы-

вадается произведением

$$\mathcal{F}(\xi) = \text{rect} \frac{\xi}{2b} \cdot \sum_{m=-n/2}^{n/2} G_m e^{im(\pi/b)\xi}. \quad (8)$$

Здесь мы используем ступенчатую функцию, которая часто используется в Фурье-оптике (см., например, ¹⁶):

$$\begin{aligned} \text{rect } \xi' &= 1 & | \xi' | < 1, \\ &= 0 & | \xi' | > 1. \end{aligned}$$

Переходя по формуле (6) от спектра $\mathcal{F}(\xi)$ к изображению $\mathcal{E}(x)$, из (8) получаем

$$\mathcal{E}(x) = \sum_{m=-n/2}^{n/2} G_m \frac{\sin(bx + \pi m)}{bx + \pi m}. \quad (9)$$

Выражение (9) можно преобразовать к виду

$$\mathcal{E}(x) = \sin bx \cdot \sum_{m=-n/2}^{n/2} \frac{G_m (-1)^m}{bx + \pi m} = \sin bx \cdot \frac{Q_n(x)}{\prod_{m=-n/2}^{n/2} (bx + \pi m)}, \quad (10)$$

где Q_n — многочлен степени n с комплексными коэффициентами. Обозначая корни многочлена через z_j ($j = 1, 2, \dots, n$), а через Cb^{n+1} — коэффициент при старшей степени многочлена, мы можем вместо (10) записать

$$\mathcal{E}(x) = C \sin bx \cdot \frac{\prod_{j=1}^n (x - z_j)}{x \prod_{m=1}^{n/2} \left(x - \frac{\pi m}{b} \right) \left(x + \frac{\pi m}{b} \right)}. \quad (11)$$

Из формулы (11) видно, что — с точностью до константы — изображение $\mathcal{E}(x)$ полностью определяется корнями многочлена z_j и шириной спектра b . Число корней многочлена, n , как следует из (8) и (10), определяется числом гармоник, участвующих в формировании изображения. Из (6) видно, что число n определяет также число существенно различных элементов изображения, которое можно называть числом его степеней свободы. Если теперь перейти к модулю поля, точнее, к квадрату модуля, или интенсивности, то на основе (11) можно получить для интенсивности следующее представление

$$I(x) = \mathcal{E}(x) \mathcal{E}^*(x) = CC^* \left[\frac{\sin bx}{x \prod_{m=1}^{n/2} \left(x^2 - \frac{\pi^2 m^2}{b^2} \right)} \right]^2 \prod_{j=1}^n (x - z_j) (x - z_j^*). \quad (12)$$

Из (12) видно, что если функция $I(x)$ известна, то можно определить корни z_j, z_j^* . Если же пытаться по этим корням найти поле $\mathcal{E}(x)$, то возникает неоднозначность: для каждой пары комплексно сопряженных корней z_j и z_j^* существует две возможности, можно считать корнем функции \mathcal{E} либо z_j , либо z_j^* . Перебирая все n пар комплексно сопряженных корней, мы можем построить 2^n различных функций $\mathcal{E}(x)$. Таким образом, число допустимых решений составляет 2^n . Подчеркнем, что добавление к заданному модулю произвольного фазового множителя могло бы вывести решение за пределы нужной спектральной полосы. Число 2^n определяет степень произвола в выборе решения. Имеется 2^n различных функций, совместимых и с заданным модулем и с заданной шириной спектра.

Отметим, что можно было бы, пользуясь разложением

$$\sin bx = bx \prod_{m=1}^{\infty} \left(1 - \frac{b^2 x^2}{m^2 \pi^2}\right)$$

представить поле $\mathcal{E}(x)$ в виде

$$\mathcal{E}(x) = \text{const} \cdot \prod_{j=1}^n (x - z_j) \prod_{m=\frac{n}{2}+1}^{\infty} \left(x - \frac{\pi m}{b}\right) \left(x + \frac{\pi m}{b}\right). \quad (13)$$

Если рассматривать функцию $\mathcal{E}(x)$ на комплексной плоскости переменной x , то можно было бы говорить, что она обращается в нуль при $x = z_j$, $j = 1, 2, \dots, n$ и при $x = \pm \frac{\pi m}{b}$, $m = \frac{n}{2} + 1, \frac{n}{2} + 2, \dots$. Нули при $x = \pm \frac{\pi m}{b}$ связаны только с шириной спектральной полосы и не зависят от коэффициентов G_m . Нули при $x = z_j$, $j = 1, 2, \dots, n$ несут существенную информацию о поле.

Приведенные здесь соображения справедливы не только для функций, фурье-образ которых дается формулой (8), они носят более общий характер. Требуется лишь, чтобы функция имела ограниченный спектр. Из теории аналитических функций известно, что функцию с ограниченным спектром можно представить в виде

$$\mathcal{E}(x) = \text{const} \cdot \exp\left(i \frac{a+b}{2} x\right) \prod_{j=-\infty}^{\infty} (x - z_j), \quad (14)$$

причем носителем спектра является отрезок $[a, b]$. Необходимые пояснения и литературные ссылки можно найти в ^{11,12}, краткое изложение вопроса имеется в ряде монографий, например, в ¹⁷. Известно (см. ^{11,12}), что вид функции $\mathcal{E}(x)$ существенно зависит лишь от положения конечного числа нулей. Это число, n , связано с числом степеней свободы функции, однако в общем случае эта связь не выглядит столь прозрачной, как для нашего частного случая — для функций вида (13). Остальные нули (за исключением указанных n нулей), непременно находятся в одной полуплоскости комплексного переменного, лежат на конечном расстоянии от вещественной оси, причем в асимптотике последовательные нули сдвинуты один относительно другого на отрезок длины $2\pi/(b-a)$, параллельный вещественной оси. Для функции вида (14) известно, что восстановление функции по ее модулю дает неоднозначность 2^n . Утверждается, впрочем, (см. ¹¹), что если рассматривать продолжение спектра такой функции за пределы апертуры $[a, b]$, то только у одной из 2^n функций спектр будет убывать с ростом модуля частоты на вещественной оси. Однако условие убывания спектральной функции на высоких частотах не удается ввести в итерационные алгоритмы. Поэтому при численных расчетах по восстановлению комплексных функций оказывается, что данных, относящихся к одному сечению поля, недостаточно, приходится привлекать данные, относящиеся ко второму сечению.

Строгая теория, к сожалению, пока не может указать, какой объем дополнительной информации нужен для того, чтобы снять неоднозначность. В литературе существует мнение, основанное лишь на опыте работы с алгоритмами, что в задаче «двух модулей», а также во многих случаях восстановления вещественной неотрицательной функции однозначность обеспечивается ¹⁸.

Связь задачи восстановления фазы с отысканием положения нулей комплексной функции была выяснена уже в старых работах, которые не использовали формулу (14), а опирались на преобразование Гильберта (см., напри-

мер, ², с. 625). Приведем формулу из обзора ², которая выражает связь фазы и модуля комплексной функции

$$\varphi(x) = \frac{x}{\pi} P \int_{-\infty}^{\infty} \frac{|\mathcal{E}(x')| dx'}{x'(x-x')} + \frac{1}{i} \ln \prod_{j=1}^n \frac{x-z_j}{x-z_j^*}. \quad (15)$$

Здесь предполагается, что функция $\mathcal{E}(x)$ регулярна в верхней полуплоскости комплексного переменного x , z_j — нули функции $\mathcal{E}(x)$, лежащие в верхней полуплоскости, P — символ главного значения интеграла. Как видно из (15), в этом подходе тоже требуется решать вопрос об отнесении нулей. Следует при этом иметь в виду, что если положение нулей известно, то функция может быть получена с помощью произведения Адамара (14), так что обращение к преобразованию Гильберта становится излишним.

Очень интересен в фазовой проблеме переход от одного измерения к двум и большему числу измерений. Особенности двумерных фазовых задач исследовались многими авторами. Из первых работ на эту тему назовем ¹⁵, из более поздних — ^{19,20} (см. также ссылки в ^{18,20}). Поведение функции двух переменных, имеющей ограниченный спектр, также определяется ее нулями. Однако в отличие от одномерного случая нули больше не являются изолированными точками и не могут переставляться порознь, поэтому источник неоднозначности ликвидируется. Выведем для двумерного случая формулы, аналогичные (9), (10) и (12); получаем

$$\mathcal{E}(x, y) = \sum_{m=-n/2}^{n/2} \sum_{l=-n/2}^{n/2} G_{ml} \frac{\sin(bx + \pi m)}{bx + \pi m} \frac{\sin(by + \pi l)}{by + \pi l}, \quad (9a)$$

$$\mathcal{E}(x, y) = \frac{C \sin bx \cdot \sin by}{\prod_{m=-n/2}^{n/2} \left(x + m \frac{\pi}{b}\right) \prod_{l=-n/2}^{n/2} \left(y + l \frac{\pi}{b}\right)} Q_{n,n}(x, y), \quad (10a)$$

$$I(x, y) = CC^* \left[\frac{\sin bx \cdot \sin by}{\prod_{m=-n/2}^{n/2} \left(x + m \frac{\pi}{b}\right) \prod_{l=-n/2}^{n/2} \left(y + l \frac{\pi}{b}\right)} \right]^2 Q_{n,n}(x, y) Q_{n,n}^*(x, y). \quad (12a)$$

В формуле (10a) (в отличие от (10)) стоит полином от x, y . Вероятность того, что при произвольных коэффициентах G_{ml} полином $Q_{n,n}$ окажется приводимым, мала. Поэтому чаще всего заданному распределению интенсивности будут соответствовать только две функции для поля, которые переходят одна в другую с помощью комплексного сопряжения. Эта очевидная неоднозначность обычно не вызывает затруднений и, как правило, в литературе даже не упоминается. Отметим, что в работах по восстановлению изображений принято говорить о «форме изображения» ⁸ и считать в этом плане эквивалентными функции, которые получаются одна из другой с помощью комплексного сопряжения, сдвига или инверсии. Таким образом, в двумерном случае фаза связана с интенсивностью «почти всегда» однозначно; исключения составляют функции, для которых зависимость от двух координат разделяется, и другие факторизуемые функции ²⁰. Такие исключения обсуждаются, например, в ²¹, где идет речь о неоднозначности, которая не снимается даже условием неотрицательности.

Отметим, что в изложенных особенностях двумерной фазовой задачи прослеживается аналогия с задачей о дислокациях волнового фронта ^{22–25}. В задаче о дислокациях переход от одной поперечной координаты к двум также приводит к качественным особенностям за счет изменения свойств нулей поля (в реальном — а не комплексном — пространстве). Имеющиеся к настоящему времени исследования позволяют пока говорить лишь о внешней аналогии.

Заканчивая изложение этого раздела, упомянем еще один математический вопрос — о зависимости решений фазовой задачи от малых отклонений в исходных данных. Поскольку в эксперименте всегда присутствуют шумы, то зарегистрированное распределение интенсивности поля отличается от истинного, и в какой-то мере эти отличия будут влиять на восстанавливаемое распределение фазы. Вопрос о роли шумов поднимался в ряде работ по восстановлению фазы^{18,26–29}. В других оптических задачах исследования устойчивости, методы регуляризации находят все более широкое применение (см.³⁰). В фазовой задаче влияние шумов еще подлежит количественному изучению.

5. НЕКОТОРЫЕ ПРИЛОЖЕНИЯ ФАЗОВОЙ ЗАДАЧИ. КИНОФОРМ

Проблема восстановления зависимости фазы поля от координаты связана с разнообразными прикладными вопросами. Из них наиболее полно освещены в литературе вопросы, возникающие при построении изображений в астрономии. Здесь в ряде случаев регистрация сигналов выполняется таким образом, что обработка данных сводится к восстановлению неотрицательной функции по модулю ее фурье-спектра. При этом оказывается целесообразным использование алгоритма Финапа^{5,6}. Работы этого направления отражены во многих сборниках и монографиях (см., например,^{31–33}), поэтому мы не будем на них останавливаться.

Обратим внимание на задачу, которая в литературе описана не столь обстоятельно. Это задача воздействия на пространственные характеристики излучения. Она состоит в расчете и построении таких фазовых элементов, которые могли бы сфокусировать излучение лазерного источника в заданную линию, фигуру, в какую-то заранее предписанную пространственную картину. Такие фазовые элементы принято называть киноформами³⁴. Если ограничиться математической стороной вопроса, то можно сказать, что при синтезе киноформа решается задача подбора фазовой функции по двум распределениям интенсивности. В ближней зоне берется равномерное распределение (обычно кусочно-постоянное, иногда — гауссов профиль, $\exp(-r^2/r_0^2)$) и заданная сложная пространственная картина в дальней зоне. Формально мы здесь имеем те же исходные данные, что и в задаче восстановления фазы поля, интенсивность которого зарегистрирована в двух плоскостях. Поэтому при построении киноформов применяются алгоритмы типа Гершберга — Сэкстона. Ссылки по использованию алгоритмов можно найти в³⁵. Есть, впрочем, и другой подход к расчету киноформов (см., например,³⁶).

Имеется, однако, важное отличие фазовой задачи от задачи синтеза киноформа. В случае восстановления фазы всегда известно, что решение существует. Действительно, в фазовой задаче мы имеем два распределения интенсивности, полученные для объектного поля, которое реально существовало: эти два распределения можно назвать согласованной парой. В случае киноформа заранее не ясно, существует ли такое комплексное поле, которое содержит оба заданных распределения интенсивности.

Численные расчеты по киноформам показали, что если требуется заданное для дальней зоны пространственное распределение воспроизвести с низкой точностью, то решение обычно удается построить. Однако задача повышения точности, как правило, оказывается невыполнимой (заданные входные и выходные распределения не являются согласованной парой). Эта ситуация описана в³⁵, предложен такой путь улучшения характеристик киноформов: высокая точность в центральной области выходного распределения интенсивности достигается за счет того, что допускается некоторый произвол в периферийной области. При этом увеличение площади периферийной области приводит к более точному результату и к повышению скорости сходимости алгоритма.

6. ВОССТАНОВЛЕНИЕ ВРЕМЕННЫХ ХАРАКТЕРИСТИК ИЗЛУЧЕНИЯ

Вопросы, которые мы обсуждали в предыдущих пунктах, относились к пространственной структуре световых полей. Близкие вопросы возникают и в отношении **временной** структуры излучения.

Развитие лазеров и нелинейной оптики выдвинуло задачу исследования излучения с необычными **временными** характеристиками, например, световых импульсов очень малой длительности, импульсов с сильной фазовой модуляцией. Возникла необходимость измерения **временных** характеристик с высокой точностью. Напомним, что в 1967 г. были предложены^{37,38} и стали применяться косвенные методы регистрации излучения, основанные на измерении автокорреляционных функций интенсивности. В те годы электронная техника не позволяла проводить прямые измерения с необходимым временным разрешением. Поэтому косвенные методы быстро завоевали многочисленных сторонников. В последующие годы прямые методы шагнули далеко вперед. Важнейшие достижения в этой области изложены, например, в³⁹. Однако и сейчас электронная техника отстает от потребностей лазерной физики, которые к настоящему времени переместились из пикосекундного в фемтосекундный диапазон. В связи с этим косвенные методы измерения **временных** характеристик по-прежнему широко используются в лабораториях. При использовании этих методов, т.е. при обработке автокорреляционных функций с целью получения **временных** характеристик излучения, возникают такие же проблемы, как и при обработке пространственных характеристик полей, направленной на восстановление пространственного распределения фазы. Однако эти аналогии и методы, применяемые для обработки пространственных характеристик, при **временных** исследованиях не используются. Более того, при изучении ряда вопросов, касающихся временных измерений, исследователи опираются на неверные представления, хотя аналогичные вопросы на языке пространственных функций уже выяснены и изложены в литературе. В качестве примера укажем ссылку на принцип «максимальной энтропии» в работе⁴⁰, изложение возможностей корреляционных методик в монографиях, например в⁴¹. В этом плане очень полезна было бы рассматривать временные измерения как один из аспектов фазовой проблемы.

Напомним, что при корреляционных методах исследования излучения проводятся следующие измерения. Исследуемый световой пучок расщепляется на два пучка с контролируемой разностью хода, оба пучка направляются в нелинейную среду, в которой возбуждается квадратичный по интенсивности света эффект, и на основе этого регистрируется функция

$$g(\tau) = \frac{1}{T} \int_0^T I(t) I(t + \tau) dt; \quad (16)$$

здесь $I(t)$ — изменяющаяся во времени интенсивность излучения, T — полное время регистрации, τ — разность хода двух пучков; корреляционная функция $g(\tau)$ измеряется в относительных (а не в абсолютных) единицах. Особенностью задачи является наличие двух существенно различных временных масштабов: величины T — полной длительности (которую можно связать с периодом, определяемым резонатором лазера, а в схеме с затвором, выделяющим один импульс из цуга, просто приравнять периоду), составляющей $\approx 10^{-8}$ с, и величины $\tau_{\text{корр}}$, характеризующей ширину пика корреляционной функции и имеющей порядок 10^{-12} с. Чтобы не усложнять рассмотрение, мы здесь не касаемся того обстоятельства, что чаще всего в экспериментах регистрируется сумма функции $g(\tau)$ и некоторого постоянного «фона». Подробное рассмотрение этого обстоятельства и вычисление коэффициента пропорциональности, связывающего «фон» с величиной $g(0)$ для случая двухфотонной люминесценции, имеется в⁴²⁻⁴⁴. Таким образом, в простей-

шем варианте задача состоит в определении функции $I(t)$ по измеренной функции вида (16).

Если заранее известно, что излучение представляет собой одиночный импульс, то по ширине пика функции $g(\tau)$ можно найти длительность этого импульса. Если же дополнительная информация о свойствах излучения отсутствует, то доказать, что импульс был одиночный, по таким измерениям совсем не легко. Трудности связаны с тем, что существенно различные сигналы (например, одноимпульсный и многоимпульсный, или одноимпульсный с нулевым фоном и одноимпульсный на широком пьедестале) имеют мало различающиеся корреляционные функции, причем различия могут теряться из-за шумов при регистрации. Здесь требуется высокая точность измерения, функции $g(\tau)$ и надлежащая обработка измерений.

Рассмотрим, например, следуя⁴²⁻⁴⁴, вопрос о том, какая точность нужна, чтобы по автокорреляционной функции оценить мгновенную мощность, которая достигается в исследуемом излучении. В общем случае ширина пика автокорреляционной функции (которая обычно только и измеряется) дает лишь характерный масштаб изменения интенсивности: если ширина пика составляет $\tau_{\text{корр}}$, то, пренебрегая множителями $\sim \sqrt{2}$, можно считать, что во **временной** картине излучения имеются выбросы длительности $\tau_{\text{корр}}$. Максимальная мгновенная мощность достигалась бы, если бы имелся только один такой выброс на нулевом фоне; при этом мгновенная мощность превышала бы среднюю в $T/\tau_{\text{корр}}$ раз;

$$I_{\text{мгн}} \approx \frac{I_{\text{ср}} T}{\tau_{\text{корр}}}.$$

В лазерах с самосинхронизацией мод это превышение может составлять 10^3-10^4 ($T/\tau_{\text{корр}} \approx 10^3-10^4$). Максимальное значение соответствует случаю полной синхронизации мод. При неполной синхронизации величина $I_{\text{мгн}}/I_{\text{ср}}$ может лежать в диапазоне от 1-2 до 10^3-10^4 .

Введем величину, с помощью которой можно охарактеризовать усредненное значение мгновенной мощности:

$$t_{\text{эфф}} = \left(\int_0^T I(t) dt \right)^2 \left(\int_0^T I^2(t) dt \right)^{-1} \equiv T \langle I \rangle^2 \langle I^2 \rangle^{-1}. \quad (17)$$

Ясно, что $I_{\text{мгн}}/I_{\text{ср}} \sim T/t_{\text{эфф}}$. Величину $t_{\text{эфф}}$ мы называем эффективной длительностью, она дает характерную величину **временного** промежутка, на котором сосредоточена **большая** часть энергии излучения. Выразим величину $t_{\text{эфф}}$ через корреляционную функцию $g(\tau)$. Из формулы (16) получаем

$$\int_{-T}^T g(\tau) d\tau = \frac{1}{T} \left(\int_0^T I(t) dt \right)^2,$$

$$g(0) = \frac{1}{T} \int_0^T I^2(t) dt;$$

отсюда и из определения (17) следует

$$t_{\text{эфф}} = \frac{1}{g(0)} \int_{-T}^T g(\tau) d\tau.$$

Нетрудно теперь связать ошибку в измерении эффективной длительности с точностью измерения корреляционной функции

$$\delta(t_{\text{эфф}}) = \frac{T \delta g}{g(0)}. \quad (18)$$

Если требуется доказать, что $t_{\text{эфф}} \approx t_{\text{корр}}$, то эта ошибка, разумеется, не должна превышать $t_{\text{корр}}$, $\delta(t_{\text{эфф}}) \leq t_{\text{корр}}$ или $T\delta g/g(0) \leq t_{\text{корр}}$. Это означает, что допустимая погрешность в измерении корреляционной функции составляет очень малую величину $\delta g/g(0) \leq t_{\text{корр}}/T \approx 10^{-3}$. Впрочем, требования к точности могли бы быть несколько снижены, если бы одновременно с измерением корреляционной функции проводились прямые измерения **временного** хода интенсивности излучения. Пусть прямые методы имеют разрешение $t_{\text{разр}}$ и указывают на наличие единственного импульса длительности $t_{\text{разр}}$ на периоде T ; тогда в оценке (18) величину T можно заменить на $t_{\text{разр}}$. В соответствии с этим требования к точности измерения автокорреляционной функции принимают вид

$$\frac{\delta g}{g(0)} \leq \frac{t_{\text{корр}}}{t_{\text{разр}}}.$$

Обратим внимание, что задача восстановления **временного** хода интенсивности по автокорреляционной функции может рассматриваться как пример фазовой задачи. В самом деле, если ввести спектр интенсивности, т. е. функцию $\tilde{I}(\Omega)$, с которой искомая функция $I(t)$ связана преобразованием

$$I(t) = \frac{1}{\sqrt{2\pi}} \int \tilde{I}(\Omega) e^{i\Omega t} d\Omega,$$

то пользуясь теоремой Винера — Хинчина или же непосредственно выполняя фурье-преобразование от (16), можно получить

$$|\tilde{I}(\Omega)|^2 = \frac{1}{\sqrt{2\pi}} \int g(\tau) e^{-i\Omega\tau} d\tau.$$

Таким образом, в задаче имеется информация о величине $|\tilde{I}(\Omega)|$, т. е. о модуле фурье-спектра искомой функции, а фаза фурье-спектра неизвестна; кроме того, интенсивность $I(t)$ является неотрицательной функцией, имеющей конечный носитель $[0, T]$. Это значит, что мы здесь имеем задачу восстановления неотрицательной функции по модулю ее фурье-спектра. Так что опыт, накопленный при решении подобных задач, может быть использован при нахождении **временных** характеристик излучения. Здесь нужно иметь в виду и достижения, связанные с применением алгоритмов^{5,6}, и трудности, относящиеся к тем случаям, когда решение не единственно²¹. Применение алгоритмов не снимает, конечно, требования к точности измерения корреляционной функции.

Методы измерения **временных** характеристик, основанные на получении корреляционных функций, во всех своих вариантах имеют дело не с полем $\mathcal{E}(t)$, а с интенсивностью поля $I(t) = \mathcal{E}(t)\mathcal{E}^*(t)$. В традиционной постановке эти методы направлены на восстановление **временного** хода модуля поля $|\mathcal{E}(t)|$. Информация о модуле поля очень важна, часто оказывается, что только она и нужна в экспериментальных исследованиях. Однако в ряде случаев при прохождении световых импульсов через среду с нелинейной рефракцией возникает сильная частотная самомодуляция⁴⁵⁻⁴⁸. Детальные сведения о параметрах модуляции могут оказаться существенными для изучения последующих процессов, в которых будут участвовать световые импульсы.

Допустим, стоит вопрос об исследовании частотной самомодуляции, т. е. о восстановлении **временного** хода фазы поля. Восстановление **временного** хода фазы может быть сведено к обсуждавшейся выше задаче на восстановление фазы по двум модулям. Действительно, пусть получены данные о **временном** ходе интенсивности излучения и об интенсивности спектра поля. Будем считать при этом, что спектр зарегистрирован прибором с высоким спектральным разрешением, как, например, в работе⁴⁹. Такие данные можно было бы обрабатывать с помощью алгоритма Гершберга — Сэкстона

или с помощью алгоритма, изложенного в п. 3. Необходимым условием применения алгоритма является, конечно, высокая точность измерений и хорошее разрешение как во **временнóй**, так и в спектральной картине. Вопрос о допустимой погрешности измерений в настоящее время не ясен. Ясно лишь, что эта задача примыкает к ряду других некорректно поставленных задач, которые возникают в оптике³⁰.

7. ЗАКЛЮЧЕНИЕ

Основная цель, которая ставилась при написании обзора, состояла в привлечении внимания к успехам исследований последних лет в области фазовой проблемы. Работы по фазовой проблеме публикуются, в основном, в специализированных сборниках, посвященных обработке оптических изображений, и недостаточно известны физикам других специальностей. Значительная часть работ, выполненных в этой области, носит эмпирический характер и направлена на разработку и усовершенствование численных алгоритмов для решения того или иного варианта фазовой задачи. Имеются и математические работы по фазовой проблеме. В этих работах очень интересны результаты, касающиеся неоднозначности решений в одномерном случае и особенностей, возникающих при переходе к двумерному случаю. Математические работы по фазовой проблеме пока еще не дают окончательных заключений о существовании и единственности решений для ряда практически работающих алгоритмов. Недостает в исследованиях и подробного анализа устойчивости решений относительно малых изменений в исходных данных.

Тем не менее, несмотря на отсутствие законченной теоретической базы, результаты работ по фазовой проблеме находят широкое применение для решения многих прикладных задач (восстановление изображений, прошедших через турбулентную атмосферу, управление пространственными характеристиками лазерного излучения). Важно, чтобы имеющиеся здесь достижения были восприняты и в других областях физических исследований. Особенно важно было бы перенести идеи, сформировавшиеся при решении фазовых задач, в область исследования временных характеристик излучения.

Приношу искреннюю благодарность Б. Я. Зельдовичу и Н. Г. Преображенскому за обсуждение изложенных здесь вопросов.

Физический институт им. П. Н. Лебедева
АН СССР

СПИСОК ЛИТЕРАТУРЫ

1. Gerchberg R. W., Saxton W. O. // *Optik*. 1972. Bd 35. S. 237.
2. Вольф Э., Манделъ Л. // *УФН*. 1966. Т. 88. С. 619.
3. Нуссенцвайг Х. М. *Причинность и дисперсионные соотношения*. — М.: Мир, 1976.
4. Клаудер Дж., Сударшан Э. *Основы квантовой оптики*. — М.: Мир, 1970.
5. Fienup J. R. // *Opt. Lett.* 1978. V. 3. P. 27.
6. Fienup J. R. // *Appl. Opt.* 1982. V. 21. P. 2758.
7. Bates R. H. T., Fright W. R. // *J. Opt. Soc. Am.* 1983. V. 73. P. 358.
8. Bates R. H. T., Tan D. G. H. // *Ibidem Ser. A*. 1985. V. 2. P. 2013.
9. Кузнецова Т. И., Кузнецов Д. Ю. // *КЭ*. 1985. Т. 12. С. 2507.
10. Кузнецова Т. И., Кузнецов Д. Ю. // *Кр. сообщ. физ., ФИАН СССР*. 1986. № 2. С. 12; *Opt. Commun.* 1987. V. 61. P. 374.
11. Ferwerda H. A. // *Inverse Source Problems in Optics* / Ed. H. P. Baltes. — Berlin; Heidelberg; New York: Springer-Verlag, 1978. — (Topics in Current Physics, V 9); перевод: Ферверда Х. А. *Проблема восстановления фазы волнового фронта по амплитудному распределению и функциям когерентности. Обратные задачи в оптике* / Под ред. Г. П. Болтса. — М.: Машиностроение, 1984.
12. Ross G., Fiddy M. A., Nieto-Vesperinas M. // *Inverse Scattering Problems in Optics* / Ed. H. P. Baltes. — Berlin; Heidelberg; New York: Springer-Verlag, 1980. — (Topics in Current Physics, V 20).
13. O'Neill E. L., Wahlter A. // *Opt. Acta*. 1963. V. 10. P. 33.
14. Wahlter A. // *Ibidem*. P. 41.

15. Вгуск Yu. M., Sodin L. G.//Opt. Commun. 1979. V. 30. P. 304.
16. Гудмен Дж. Введение в фурье-оптику.— М.: Мир, 1970.
17. Василенко Г. И., Тараторин А. М. Восстановление изображений.— М.: Радио и связь, 1986.
18. Вагакат R., Newsam G.//J. Opt. Soc. Am. Ser. A. 1985. V. 2. P. 2027.
19. Hayes M. H., McClellan J. H.//Proc. IEEE. 1982. V. 70. P. 197.
20. Sanz J. L. C., Huang T. S.//J. Opt. Soc. Am. 1983. V. 73. P. 1446.
21. Фиенур J. R.//Ibidem. Ser. A. 1986. V. 3. P. 284.
22. Баранова Н. Б., Зельдович Б. Я.//ЖЭТФ. 1981. Т. 80. С. 1789.
23. Баранова Н. Б., Зельдович Б. Я., Мамаев А. В., Пилипецкий Н. Ф., Шкунов В. В.//Письма ЖЭТФ. 1981. Т. 33. С. 206.
24. Баранова Н. Б., Зельдович Б. Я., Мамаев А. В., Пилипецкий Н. Ф., Шкунов В. В.//ЖЭТФ. 1982. Т. 83. С. 1702.
25. Ваганова Н. В., Мамаев А. В., Пилипетский Н. Ф., Шкунов В. В., Зельдович В. Я.//J. Opt. Soc. Am. 1983. V. 73. P. 525.
26. Фиенур J. R.//Ibidem. P. 1421.
27. Sanz J. L. C., Huang T. S., Sukierman F.//Ibidem. P. 1442.
28. Аристов В. В., Ерко А. И., Копецкий Ч. В., Кузнецов С. М., Ушаков Н. Г.//Опт. и спектр. 1987. Т. 62. С. 1105.
29. Бакут П. А., Пахомов А. А., Ряхин А. Д., Свиридов К. Н., Устинов Н. Д.//ДАН СССР. 1986. Т. 290. С. 89.
30. Bertero M., De Mol C., Viano G. A.//¹².
31. Построение изображений в астрономии по функциям когерентности/Под ред. К. ван Схонвелда.— М.: Мир, 1982.
32. Компьютеры в оптических исследованиях/Под ред. Б. Фридена — М.: Мир, 1983.
33. Устинов Н. Д., Матвеев И. Н., Протопопов В. В. Методы обработки оптических полей в лазерной локации.— М.: Наука, 1983.
34. Кольер Р., Беркхарт К., Лин Л. Оптическая голография.— М.: Мир, 1973.
35. Hiroshi Akaŋogi//Appl. Opt. 1986. V. 25. P. 802.
36. Гончарский А. В., Данилов В. А., Попов В. В., Прохоров А. М., Сисакян И. Н., Соيفер В. А., Степанов В. В.//ДАН СССР. 1983. Т. 273. С. 605.
37. Armstrong J. A.//Appl. Phys. Lett. 1967. V. 10. P. 16.
38. Giordmaine J. A., Rentzepis P. M., Shapiro S. L., Wecht K. W.//Ibidem. V. 11. P. 216.
39. Щелев М. Я.//Тр. ФИАН СССР. 1985. Т. 155. С. 3.
40. Anderson T., Eng S. T.//Opt. Commun. 1983. V. 47. P. 288.
41. Сверхкороткие световые импульсы/Под ред. С. Шапиро.— М.: Мир, 1981.
42. Кузнецова Т. И. Об особенностях некоторых методов регистрации длительности сверхкоротких световых импульсов: Препринт ФИАН СССР № 47.— Москва, 1968.
43. Кузнецова Т. И.//ЖЭТФ. 1968. Т. 54. С. 2453.
44. Кузнецова Т. И.//Тр. ФИАН СССР. 1975. Т. 84. С. 62.
45. Островский Л. А.//Письма ЖЭТФ. 1967. Т. 6. С. 807.
46. Коробкин В. В., Малютин А. А., Прохоров А. М.//Ibidem. 1970. Т. 12. С. 216.
47. Stolen R. H., Lin C.//Phys. Rev. Ser. A. 1978. V. 17. P. 1448.
48. Дианов Е. М., Карасик А. Я., Прохоров А. М., Серкин В. Н.//Изв. АН СССР. Сер. физ. 1986. Т. 50. С. 1042.
49. Малышев В. И., Масалов А. В., Сычев А. А.//ЖЭТФ. 1970. Т. 59. С. 48.