ОПТИМИЗАЦИЯ РАЗРЕШАЮЩЕЙ СПОСОБНОСТИ ОПТИЧЕСКИХ ПРИБОРОВ

А.Г. Рамм и С.А. Родионов

Среди всех линейных изопланатических приборов найден прибор, «оптимально» в определенном смысле различающий два произвольных сигнала на фоне коррелированных помех. Дан критерий качества прибора, решающего задачу различения.

1. Постановка задачи

Пусть имеется линейный изопланатический оптический прибор, функция рассеяния которого есть $h(x)$ [1] Это значит, что сигнал $f(y)$ на поверхности объекта преобразуется прибором в сигнал на поверхности изображения

$$s(x) = \int h(x-y) f(y) dy \tag{1.1}$$

Здесь $x = (x_1, x_2)$ и $y = (y_1, y_2)$ – обобщенные координаты предмета и изображения, приведенные на одну поверхность для того, чтобы исключить из преобразования (1.1) различия масштабов предмета и изображения. Интеграл без указания пределов берется по всей плоскости. Через $\widetilde{h}(\lambda)$ будем обозначать преобразование Фурье функции $h(x)$

$$\widetilde{h}(\lambda) = \int \exp(i\lambda x) h(x) dx, \;\; \lambda x = \lambda_1 x_1 + \lambda_2 x_2 \tag{1.2}$$

Из теории образования оптического изображения вытекают следующие ограничения:

для когерентных сигналов:	для некогерентных сигналов:	
$\int \lvert h(x)\rvert^2 dx = E < \infty$,	$\int h(x) dx = E < \infty$	(1.3)
$\widetilde{h}(\lambda) = 0$ вне D,	$\widetilde{h}(\lambda) = 0$ вне $D \otimes D$	(1.4)

где D – область зрачка прибора, $D \otimes D$ – область, в которой отлична от нуля свертка характеристических функций области D, E – светосила прибора. Первое условие ограничивает светосилу прибора, а второе является обычно используемой идеализацией фильтрующих свойств прибора [1]. Кроме того, при некогерентном освещении накладывается дополнительное ограничение: $f(y)$, $s(x)$ и $h(x)$ – неотрицательны.

Пусть $n(x)$ – гауссов шум (случайное поле) с нулевым средним значением и корреляционной функцией $R(x,y)$, принадлежащей классу $\Re$ введенному в работе [2]

[1] предполагается, что функция рассеяния h(x) описывает прибор в целом, т. е. собственно изображающий прибор и приемник (анализатор) изображения. Такое объединение, возможно при условии линейности всего процесса.

$$R = (x, y) = \int F(\lambda)\Phi(x, y, \lambda)d\rho(\lambda) \qquad (5)$$

В [2] определены величины $\Phi(x, y, \lambda)$, $d\rho(\lambda)$. Для наших целей достаточно иметь в виду, что если шум однороден, то $R(x, y) = R(x - y)$, $d\rho(\lambda) = d\lambda$, $\Phi(x, y, \lambda) = (2\pi^2 \times \exp\{i\lambda(x - y)\}$. Для простоты будем считать в дальнейшем шум однородным. Пусть на вход прибора поступает один из двух сигналов $f_0(y)$ или $f_1(y)$. На поверхности изображения образуется сигнал $u(x) = s_1 + n(x)$, $j = 0,1$, где

$$s_j(x) = \int h(x - y) f_j(y) dy \qquad (6)$$

Шум $n(x)$ можно, например, считать внутренним шумом прибора или шумом регистрирующего сигнал устройства. Требуется по принятой реализации сигнала $u(x)$ дать ответ на вопрос о том, какой из двух сигналов поступил на вход прибора. Более подробно следует указать метод обработки сигнала $u(x)$ и найти функцию рассеяния $h(x)$ (среди функций, удовлетворяющих ограничениям (1.3), (1.4)) так, чтобы минимизировать вероятность ошибки при различении двух гипотез H_0, H_1. Здесь H_j – гипотеза, состоящая в том, что $u(x) = s_j(x) + n(x)$. Вероятность ошибки или функционал, ее определяющий, характеризуют качество прибора при различении двух сигналов или «разрешающую способность» прибора в данной задаче.

Классические понятия разрешающей способности оптических приборов [8] соответствуют в указанном смысле частным задачам: разрешающая способность по Релею – задаче различения сигнала в виде одной точки и сигнала в виде двух близких точек; разрешающая способность по Фуко – задаче различения периодического сигнала определенной частоты и постоянного сигнала.

Поставленная задача отличается от изученной в [3,6] тем, что в [3,6] не учитывались внутренние шумы прибора или шум приемника изображения. В работе [4] рассматривалась задача о разрешающей способности при наличии шума приемника. В этой работе минимизировалась правая часть неравенства Рао-Крамера для дисперсии уклонения оценки некоторого параметра от этого параметра. Поскольку, вообще говоря, правая часть неравенства Рао-Крамера есть лишь одна из нижних границ упомянутой дисперсии и существуют примеры, в которых точная нижняя грань дисперсии значительно превышает границу, даваемую неравенством Рао-Крамера, минимизация нижней границы в неравенстве Рао-Крамера не обязательно ведет к улучшению оценки параметра.

Можно указать соображение другого рода, которое (в применении к большим выборкам) до некоторой степени оправдывает подход к проблеме различения сигналов, состоящий в максимизации информационной матрицы (информационного количества) Фишера, входящей в неравенство Рао-Крамера. Дело в том, что при некоторых ограничениях. [5,7] справедливо утверждение, согласно которому отношение правдоподобия при $n \to \infty$ имеет асимптотически гауссово распределение с матрицей корреляций, совпадающей с информационной матрицей. Таким образом, если информационная матрица приведена к диагональному виду и $\lambda_j > 0$ – ее собственные числа, то $\lambda_j^{-1} = \sigma_j^2$, где σ_j^2 – дисперсия j-й компоненты случайного вектора наблюдений. Вместе с тем с принципиальной точки зрения важно иметь в виду следующие свойства оценки максимального правдоподобия, которые свидетельствуют с точки зрения практики в пользу этой оценки [5]: 1) если существует достаточная оценка параметра, то метод максимального правдоподобия (ММП) дает именно эту оценку; 2) если существует эффективная оценка, т. е. оценка с минимальной дисперсией, то ММП дает именно эту оценку; 3) при $n \to \infty$ (большая выборка) оценка $\hat{\theta}$ по ММП асимптотически нормальна, причем если оцениваемый параметр единственный, то $\overline{\hat{\theta}} = \theta$, $D[\hat{\theta}] = -(\overline{\partial^2 l / \partial\theta^2})$, где l – логарифм отношения правдоподобия; 4) оценка ММП инварианта (при $n >> 1$) относительно преобразования параметра: если $q = q(\theta)$, то $\hat{q} = q(\hat{\theta})$. Другие пороговые оценки, в частности, оценка Неймана-Пирсона, не обладают многими из указанных свойств.

2. Решение задачи в общем случае

В нижеследующих рассуждениях существенно используются результаты §2, 3 работы [3]. В этой работе показано, что оптимальное решающее правило для различения гипотез H_0, H_1 состоит в следующем: если выполнено неравенство

$$\int_{D_1} u(x)\upsilon(x)dx \geq 0.5\int_{D_1} [s_0(x) + s_1(x)]\upsilon(x)dx \tag{2.1}$$

то принимается гипотеза H_1. В противном случае принимается гипотеза H_0. Здесь $\upsilon(x)$ есть решение уравнения

$$\int_{D_1} R(y,x)\upsilon(x)dx = s(y), y \in D_1, s = s_1 - s_0 \tag{2.2}$$

Область D_1 есть область обработки сигнала $u(x)$, т.е. поле изображения прибора.

Сформулированное решающее правило оптимально в том смысле, что оно получено по методу максимального правдоподобия, т. е. максимизирует апостериорную плотность вероятности распределения случайного поля $u(x)$.

Ошибка первого рода этого правила, т. е. вероятность принять гипотезу H_1 когда на самом деле осуществилась гипотеза H_0, вычисляется так:

$$\begin{aligned}\alpha = p(\gamma_1 \mid H_0) = p\left(\int_{D_1}[s_0 + n(x)]\upsilon(x)dx \geq 0.5\int_{D_1}(s_0 + s_1)\upsilon(x)dx\right) = \\ = p\left(\int_{D_1} n(x)\upsilon(x)dx \geq 0.5\int_{D_1} s(x)\upsilon(x)dx\right)\end{aligned} \tag{2.3}$$

Заметим, что (черта означает статистическое усреднение)

$$\overline{\int\limits_{D_1} n(x)\upsilon(x)dx = 0}\,,\; d^2 = \overline{\left(\int\limits_{D_1} n(x)\upsilon(x)dx\right)^2} = \int\limits_{D_1}\int\limits_{D_1} R(y,x)\upsilon(y)\upsilon(x)dydx \quad (2.4)$$

Отсюда, учитывая, что случайная величина $\int\limits_{D_1} n(x)\upsilon(x)dx$ распределена нормально, находим

$$\alpha = \frac{1}{d\sqrt{2\pi}}\int\limits_{0.5d^2}^{\infty}\exp\left(-\frac{t^2}{2d^2}\right)dt = \frac{1}{\sqrt{2\pi}}\int\limits_{0.5d}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx \qquad (2.5)$$

Здесь было учтено равенство

$$d^2 = \int\limits_{D_1} s(y)\upsilon(y)dy \qquad (2.6)$$

которое вытекает из уравнения (2.2) и формулы (2.4). Из формулы (2.5) следует, что α достигает минимума при тех функциях $h(x)$ при которых d достигает максимума.

Для ошибки второго рода имеем формулу

$$\beta = p(\gamma_0 \mid H_1) = p\left(\int\limits_{D_1}(s_1+n)\upsilon dx \prec 0.5\int\limits_{D_1}(s_0+s_1)\upsilon dx\right) =$$

$$= p\left(\int\limits_{D_1} n\upsilon dx \prec -0.5\int\limits_{D_1} s(x)\upsilon(x)dx\right), s = s_1 - s_0$$

Отсюда и из формулы (2.4) выводим

$$\beta = \frac{1}{d\sqrt{2\pi}}\int\limits_{-\infty}^{-d^2/2}\exp\left(-\frac{t^2}{2d^2}\right)dt = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{-d/2}\exp\left(-\frac{x^2}{2}\right)dx = \frac{1}{\sqrt{2\pi}}\int\limits_{d/2}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx = \alpha$$

Таким образом, для нахождения функции рассеяния прибора, оптимально разрешающего два сигнала, получаем вариационную задачу

$$d^2 = J = \int\limits_{D_1}\int\limits_{D_1} R(x,y)\upsilon(y)\upsilon(x)dydx = \max \qquad (2.7)$$

при условиях

$$\int R(x,y)\upsilon(y)dy = \int h(x-y)f(y)dy; f(y) = f_1(y) - f_0(y) \qquad (2.8)$$

где $h(x)$ удовлетворяет ограничениям (1.3), (1.4). Все рассуждения значительно упростятся, если с самого начала считать $D_1 = E_1$, т. е. считать поле изображения прибора неограниченным. При этом предположении условие (2.8) позволяет найти $\upsilon(y)$, если шум однороден, т. е. $R(x,y) = R(x-y)$, $\widetilde{R}(\lambda) = F(\lambda)$. Преобразуя равенство (2.8) по Фурье, получаем

$$\widetilde{\upsilon}(\lambda) = \widetilde{h}(\lambda)\widetilde{f}(\lambda)F^{-1}(\lambda)$$

Отсюда из равенства Парсеваля и условия (2.7), (2.8) выводим

$$J=\frac{1}{(2\pi)^2}\int F^{-1}(\lambda)\left|\tilde{f}\right|^2\left|\tilde{h}(\lambda)\right|^2 d\lambda=\frac{1}{(2\pi)^2}\int_D F^{-1}(\lambda)\left|\tilde{f}(\lambda)\right|^2\left|\tilde{h}(\lambda)\right|^2 d\lambda=\max \quad (2.9)$$

Полученное выражение для функционала J позволяет найти оптимальный для задачи различения двух объектов прибор, а также сравнить между собой различные приборы. Именно для различения двух данных объектов тот прибор будет лучше, для которого функционал (2.9) принимает большее значение. Таким образом, величина J является критерием качества прибора («разрешающей способностью») в задаче различения двух сигналов f_0, f_1. В следующих разделах полученные результаты применяются для исследования когерентных и некогерентных приборов.

3. Различение когерентных сигналов

В когерентном случае функции предмета $f(y)$, изображения $s(x)$ и рассеяния $h(x)$ рассматриваются как комплексные функции, описывающие распределение комплексной амплитуды электромагнитного поля.

Для когерентных оптических сигналов функция рассеяния $h(x)$ удовлетворяет ограничениям (3), (4)

$$\int\left|h(x)\right|^2 dx=E<\infty;\tilde{h}(\lambda)=0$$

Преобразование Фурье $\tilde{h}(\lambda)$ функции рассеяния есть зрачковая функция оптического прибора. Обращаясь к выражению (2.9), находим, что функционал J достигает максимума, если

$$\left|\tilde{h}\right|^2=\begin{cases}\gamma\left|\tilde{f}(\lambda)\right|^2 F^{-1}(\lambda) & \text{при } \lambda\in D\\ 0 & \text{при } \lambda\notin D\end{cases} \quad (3.1)$$

Константа γ определяется из условия (3)

$$\gamma=E\left(\frac{1}{(2\pi)^2}\int_D\left|\tilde{f}(\lambda)\right|^2 F^{-1}(\lambda)d\lambda\right)^{-1}$$

Окончательно получаем выражение для зрачковой функции оптимального оптического прибора при когерентном освещении

$$\tilde{h}(\lambda)\equiv\begin{cases}2\pi E\dfrac{\tilde{f}(\lambda)\exp[i\varphi(\lambda)]}{\sqrt{F(\lambda)}(\int_D\left|\tilde{f}(\lambda)\right|^2 F^{-1}(\lambda)d\lambda)^{1/2}} & \text{при } \lambda\in \mathrm{D}\\ 0 & \text{при } \lambda\notin \mathrm{D}\end{cases} \quad (3.2)$$

где $\varphi(\lambda)$ – произвольная функция.

Ошибка различения вычисляется по формуле (2.5), причем величина d^2 для оптимального прибора равна

$$d^2 = \frac{1}{(2\pi)^2}\int \widetilde{s}\,\widetilde{\upsilon}\,d\lambda = \frac{1}{(2\pi)^2}\int_D \left|\widetilde{h}\widetilde{f}\right|^2 F^{-1}(\lambda)d\lambda =$$

$$= E^2 \int_D \frac{\left|\widetilde{f}\right|^4}{F^2(\lambda)}d\lambda \left\{\int_D \left|\widetilde{f}(\lambda)\right|^2 F^{-1}(\lambda)d\lambda\right\}^{-1} \qquad (3.3)$$

Выражения (2.9), (3.2) показывают, что качество оптического прибора определяется только модулем зрачковой функции и не зависит от ее фазы. Но модуль зрачковой функции определяется размерами и формой зрачка и функцией пропускания энергии по зрачку, а фаза – аберрациями прибора. Таким образом, в когерентном случае качество прибора не зависит от его аберраций. Однако, как известно, от аберрации прибора существенно зависит протяженность функции рассеяния, т. е. характерный размер области, в которой она существенно отлична от нуля. При произвольных аберрациях этот размер может быть как угодно большим. С другой стороны, формула (1) показывает, что размер изображения «равен сумме размеров предмета и функции рассеяния». Поэтому при произвольных аберрациях вывод о том, что они не влияют на качество прибора, справедлив только при бесконечной области интегрирования в (2.7), т. е. при бесконечно большом поле изображения прибора. При малых аберрациях, когда протяженность функции рассеяния мала, для справедливости вывода достаточно, чтобы область интегрирования была лишь несколько больше (на величину протяженности функции рассеяния) размеров предмета.

Пример (разрешение по Релею). Требуется по наблюдаемому изображению ответить на вопрос о том, является ли объект двумя точками или одной точкой, т. е. $f_1(x) = (1/2)[\delta(x-a) + \delta(x+a)]; f_0(x) = \delta(x)$, где $\delta(x)$ – дельта-функция Дирака. Пусть $F(\lambda) = \sigma^2/S$, где S – площадь спектральной полосы шума, σ^2 – его дисперсия, $f(x) = f_1(x) - f_0(x)$. Тогда $\widetilde{f} = (1/2)[\exp(i\lambda a) + \exp(-i\lambda a)] - 1 = \cos\lambda a - 1 = -2\sin^2(\lambda a/2)$. Если $a \to 0$, то $\widetilde{f} \sim -(\lambda a)^2/2$. В этом случае зрачковая функция оптимального прибора имеет вид $\widetilde{h}_{opt}(\lambda) \sim \lambda^4$ т.е. в когерентном приборе, оптимально различающем две близкие точки, пропускание по зрачку должно возрастать к краям от нуля в центре пропорционально четвертой степени расстояния точки зрачка от центра.

«Разрешающая способность» для оптимального прибора пропорциональна

$$d^2 = J \sim E\frac{S}{\sigma^2}(ar)^4$$

где r – радиус круга, вне которого $\widetilde{h}(\lambda)$ равна нулю.

4. Различение некогерентных сигналов

У оптических приборов при некогерентном освещении функция рассеяния есть действительная неотрицательная функция, удовлетворяющая ограничениям

$$h(x) \geq 0, \int h(x)dx = E < \infty,$$

$$\widetilde{h}(\lambda) = 0 \text{ вне } D \otimes D.$$

Преобразование Фурье $\widetilde{h}(\lambda)$ функции рассеяния в этом случае есть оптическая передаточная функция (ОПФ) прибора, модуль этого преобразования есть частотно-контрастная характеристика (ЧКХ), а аргумент – частотно-фазовая характеристика (ЧФХ). Ограничение $h(x) \geq 0$ не позволяет получить аналитическое решение вариационной задачи (2.9). Однако можно указать численный метод решения этой задачи. Будем искать $h(x)$ в виде

$$h(x) = \left| \sum_{j=1}^{n} c_j \int e^{i\lambda x} \Phi_j(\lambda) d\lambda \right|^2 = \left| \sum_{j=1}^{n} c_j \psi_j(x) \right|^2 \geq 0 \qquad (4.1)$$

где множество функций $\{\Phi_j(\lambda)\}$ образуют ортонормированный базис в $L_2(D \times D)$, а постоянные c_j подлежат определению из условия (2.9) при ограничении $\int h(x)dx = E < \infty$. Таким образом, задача сводится к максимизации функции $J(c_1, \ldots, c_n)$ при ограничении вида $F(c_1, \ldots, c_n) \leq E$, причем известная функция $F(c_1, \ldots, c_n)$ есть квадратичная форма переменных $c_1, \ldots, c_n$.

Из выражения (2.9) следует, что и для некогерентных предметов качество прибора определяется только модулем $\widetilde{h}(\lambda)$. В некогерентном случае $\left|\widetilde{h}(\lambda)\right|$ есть частотно-контрастная характеристика, следовательно, качество прибора не зависит от частотно-фазовой характеристики. Требования к размерам поля изображения здесь менее жесткие, чем в когерентном случае. Это объясняется тем, что при условии $h(x) \geq 0$ модуль $\widetilde{h}(\lambda)$ (ЧКХ) и фаза $\widetilde{h}(\lambda)$ (ЧФК) не являются независимыми и размер пятна рассеяния определяется в основном ЧКХ, причем для оптимальных приборов, удовлетворяющих условию (2.7), этот размер почти всегда мал. Поэтому в практических приложениях достаточно рассматривать поле изображения, лишь незначительно большее размеров предметов.

5. Выводы

Найдена функция рассеяния линейного изопланатического прибора, оптимально различающего два сигнала. В когерентном случае дано аналитическое выражение для функции рассеяния оптимального прибора, в некогерентном — численный метод расчета этой функции.

Дан метод оценки качества прибора, решающего задачу различения. Развитая теория учитывает конечность поля зрения прибора, наличие шумов приемника изображения. Комбинируя полученные результаты с результатами работ [2,3], в которых учитывалось, что входные сигналы поступают в смеси с шумом, можно провести анализ, учитывающий оба сорта шумов.

ЛИТЕРАТУРА

1. М. Борн, Э. Вольф. *Основы оптики*. М. «Мир». 1970.
2. А. Г. Рамм. *Дифференциальные уравнения*. *7*. 1971. 2234; 9. 1973. 931.
3. А. Г. Рамм. *Проблемы передачи информации.* 9. 1973. 22.
4. Л. А. Халфин. *Опт. и спектр. 26*. 1969. 1065
5. Г. Крамер. *Математические методы статистики. ИЛ*. М., 1948.
6. А. Г. Рамм. *Опт. и спектр.* 27. 1969. 508, 881; 29. 1970. 594; 29. 1970. 794.
7. Б.Р. Левин. *Теоретические основы статистической радиотехники*. М. «Советское радио». 1968.
8. А. И. Тудоровский. *Теория оптических приборов*. М.–Л. ч. II. АН СССР. 1948.