

УДК 550.83+535.12

ИТЕРАТИВНЫЙ АЛГОРИТМ РАСЧЕТА СКОРОСТИ И ЗАТУХАНИЯ ТРУБНЫХ ВОЛН ПО ДАННЫМ АКУСТИЧЕСКОГО КАРОТАЖА

© 2001 Н.Л. Казанский, П.Г. Серафимович, С.И. Харитонов

Институт систем обработки изображений РАН, г. Самара

Предложен новый итеративный алгоритм расчета скорости и затухания трубной акустической волны, измеренной малым количеством приемников. Особенностью алгоритма является то, что оптимизация оцениваемых параметров производится и во временной, и в частотной плоскостях. Это, с одной стороны, обеспечивает высокую устойчивость алгоритма к шумам, а с другой стороны позволяет выполнять дисперсионный анализ волновых мод, например, волны Стоунли. Приведено сравнение нового алгоритма с гомоморфным алгоритмом.

Введение

Оценка скорости и затухания волны является базовой задачей при обработке данных акустического каротажа [1]. По этим параметрам акустической волны определяют физические и механические свойства околоскважинной породы (например, [2]).

Распространенным методом обработки данных акустического каротажа являются метод Прони [3] и его модификации [4, 5]. Недостатками метода Прони являются большое количество выбросов в значениях оцениваемых параметров и большая погрешность в оценке коэффициента затухаемости. Чтобы преодолеть эти недостатки был разработан гомоморфный метод [1], который будет подробно рассмотрен далее. Однако при малом количестве приемников и высокой зашумленности данных, как будет показано ниже, гомоморфный метод приводит к большой погрешности в оценке затухаемости волны. В качестве шумов могут присутствовать как измерительная погрешность аппаратуры, так и искажения вследствие наложения различных типов волн.

Для обработки данных с высокой зашумленностью был разработан метод предсказания [6, 7]. Этот метод устойчив к шумам данных, но позволяет оценивать только скорость волны. Кроме того, он не позволяет проводить дисперсионный анализ и, таким образом, может с достаточной надежностью применяться в анализе только продольных и поперечных волн, исключая дисперсные волны Стоунли.

Общей особенностью всех этих методов является то, что в них изначально выбирается или временная или частотная плоскость для последующих расчетов. Так в методе Прони и его модификациях расчеты выполняются в частотной плоскости. Это приводит к тому, что результаты оценок параметров базовым методом Прони по зашумленным данным имеют большие выбросы. Сглаживание получаемых оценок в гомоморфном методе [1] делает их более устойчивыми к шумам. Однако точность оценок в некоторых полосах частот снижается. Кроме того, гомоморфный метод применим только в том случае, если отсутствуют искажения вследствие наложения различных типов волн.

В методе предсказания [6] расчеты выполняются только во временной плоскости. Это исключает возможность проведения дисперсионного анализа.

Предлагаемый в данной работе метод использует в расчетах и временную и частотную плоскости. Поэтому он сочетает достоинства упомянутых выше методов. Он устойчив к шумам, позволяет работать с малым количеством приемников и выполнять дисперсионный анализ.

Идея использования и временной и частотной плоскостей в итеративном подходе к оценке требуемых параметров не нова. Она применялась в оптике, астрономии [8, 9], при расчете дифракционных оптических элементов [10, 11].

Результаты работы предложенного итеративного алгоритма сравниваются с гомо-

морфным методом. Приведем описание гомоморфного метода.

Гомоморфный метод определения скорости и затухания трубной волны

В общем случае для нахождения требуемых параметров волны необходимо решить волновое уравнение [12]:

$$s(z, t) = \frac{1}{(2\pi)^2} \times \int_{-\infty}^{\infty} d\omega \exp(-i\omega t) \int_{-\infty}^{\infty} dk X(k, \omega) \exp(-ikz) \quad (1)$$

где $s(z, t)$ - отклик, записанный на приемнике, z - расстояние между излучателем и приемником, t - время, ω - частота, k - волновое число. Выражение $X(k, \omega)$ включает в себя информацию о геометрии модели, излучателе, граничных условиях и т.д. Для реализации гомоморфного и итеративного алгоритмов не требуется знать точный вид выражения $X(k, \omega)$.

Записывая (1) в частотной плоскости и выполняя интегрирование по контуру в плоскости комплексного волнового числа, получим выражение, состоящее из суммы вычетов, охваченных контуром, и двух ветвящихся интегралов \int_{cut} [13]:

$$S(z, \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk X(k, \omega) \exp(-ikz) = i \sum_j R_j \exp(-ikz) + \int_{cut} \quad (2)$$

В некоторых частотных диапазонах основной вклад в (2) вносит лишь один вычет. Экспериментально установлено [1], что так происходит для трубной волны в диапазоне от 0 до 5 kHz.

Пусть спектр волны на первом приемнике имеет вид:

$$s_1(\omega) = A(\omega) \exp(i\varphi(\omega)). \quad (3)$$

Спектр волны на втором приемнике описывается в виде:

$$s_2(\omega) = s(\omega) \exp(ik(\omega)l), \quad (4)$$

где l - расстояние между приемниками, $k(\omega)$

- комплексное волновое число.

Запишем $k(\omega)$ в виде:

$$k(\omega) = k_r(\omega) + i\alpha(\omega), \quad (5)$$

где $k_r(\omega)$ - описывает скорость волны, $\alpha(\omega)$

- затухание волны. Тогда (4) можно переписать в виде:

$$s_2(\omega) = A(\omega) \exp(-\alpha(\omega)l) \times \exp(i(\varphi(\omega) + k_r(\omega)l)) \quad (6)$$

Необходимо оценить параметры $k_r(\omega)$ и $\alpha(\omega)$ по измеренным сигналам на приемниках во временной плоскости.

Вычислим натуральный логарифм $s_2(\omega)$ и выделим действительную и мнимую части:

$$\Re[\ln s_2(\omega)] = \ln A(\omega) - \alpha(\omega)l, \quad (7)$$

$$\Im[\ln s_2(\omega)] = \text{Развернутое значение } [\varphi(\omega) - k_r(\omega)l]. \quad (8)$$

Выражения (7) и (8) описывают прямые линии. Выражение "развернутое значение" в (8) подразумевает [1], что полученные по модулю 2π значения "разворачиваются" в прямую линию добавлением чисел $2\pi n$, где n - целое, т.е. выполняется операция, обратная взятию по модулю 2π . Коэффициенты этих прямых можно найти, используя метод наименьших квадратов (МНК). В этом и заключается определение скорости и затухания акустической волны гомоморфным методом.

Итеративный алгоритм определения скорости и затухания трубной волны

Применение гомоморфного метода дало хорошие результаты при использовании большого количества приемников (12) и малом затухании волны [1]. Чтобы более точно оценивать требуемые параметры, предлагаем интерпретировать вычисления, выполняемые в гомоморфном методе, как "половину итерации". То есть, имея отклик на нескольких приемниках во временной плоскости, переходим в частотную плоскость, выполняя преобразование Фурье. Далее, в соответствии с гомоморфным методом, находим коэффици-

енты прямых - $k_r(\omega)$ и $\alpha(\omega)$, но не заканчиваем на этом расчеты, а переходим опять во временную плоскость. Здесь, получившуюся фазу мы оставляем неизменной, а результатирующую амплитуду заменяем на амплитуду отклика на приемниках. На этом единичная итерация заканчивается.

Кроме того, в процессе итераций мы полагаем, что $k_r(\omega)$ и $\alpha(\omega)$ являются линейными функциями частоты. Такая "линеаризация" увеличивает быстродействие алгоритма и делает его более устойчивым к шумам. Для выполнения более точного дисперсионного анализа значения $k_r(\omega)$ и $\alpha(\omega)$ можно аппроксимировать полиномами степени большей, чем первая.

Формально предлагаемый алгоритм для двух приемников записывается в виде:

Пусть $_1s(t)$ и $_2s(t)$ - отклики на первом и втором приемниках, соответственно.

$_1S(\omega) = \Phi\{_1s(t)\} = _1A(\omega) \exp[i(_1\phi(\omega))]$ - Фурье-спектр отклика на первом приемнике.

1. Полагаем отклик на втором приемнике начальным приближением $_2s(t) = _2s_0(t)$.

2. Находим Фурье-спектр

$$_2S_n(\omega) = \Phi\{_2s_n(t)\} = _2A_n(\omega) \exp[i(_2\phi_n(\omega))],$$

где n - номер итерации.

3. Находим значения α_n и c_n из условия

$$\min_{\alpha_n, c_n} \left\{ \sum_{\omega} \left[\ln \frac{_2A_n(\omega)}{_1A(\omega)} + \alpha_n \omega - c_n \right]^2 \right\}; \quad (9)$$

находим значения k_n из условия

$$\min_{k_n} \left\{ \sum_{\omega} [{}_1\phi(\omega) + k_n \omega - {}_2\phi_n(\omega)]^2 \right\}. \quad (10)$$

4. Выполняем обратное Фурье-преобразование

$${}_2s_{n+1}(t) = \Phi^{-1} \left\{ {}_1A(\omega) \exp(-\alpha_n \omega + c_n) \times \exp({}_1\phi(\omega) + k_n \omega) \right\}$$

5. Выполняем замену

$${}_2s_{n+1}^P(t) = \frac{{}_2s(t)}{|{}_2s_{n+1}(t)|} [{}_2s_{n+1}(t)].$$

Переходим к шагу 2 или заканчиваем итерации, если изменение ${}_2s_n(t)$ меньше заданной погрешности.

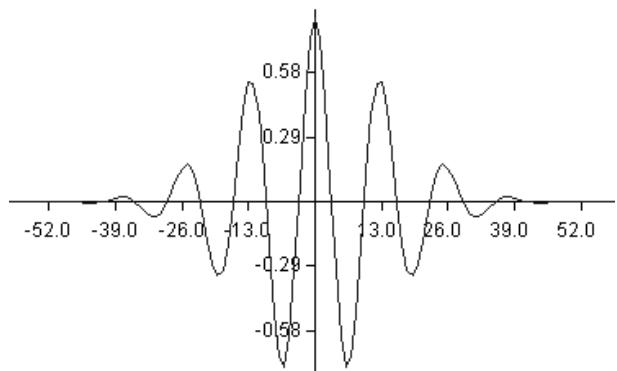


Рис. 1. Сигнал на первом приемнике

Численные результаты

В качестве тестового сигнала на первом приемнике использовался (рис.1)

$$_1s(t_i) = \cos\left(\frac{20\pi}{N}\left(i - \frac{N}{2}\right)\right) \exp\left(-\frac{(i - \frac{N}{2})^2}{\Delta^2}\right). \quad (11)$$

где $N = 128$ - количество отсчетов сигнала, $\Delta = 20$, $i = 1\dots 128$.

Умножив Фурье-спектр сигнала (11) на коэффициент, описывающий затухание, и выполнив обратное Фурье-преобразование, получим сигнал на втором приемнике (без запаздывания)

$${}_2s(t_i) = \Phi^{-1} \{ \exp(\alpha i) \Phi[{}_1s(t_i)] \}, \quad (12)$$

На рис.2 показаны амплитуды сигнала на первом приемнике и, получившегося в результате операции (12), сигнала на втором приемнике.

На рис.3 показана фаза сигнала на втором приемнике. Заметно, что операция (12) приводит к небинарной фазе. Подобное искажение фазы может происходить, например,

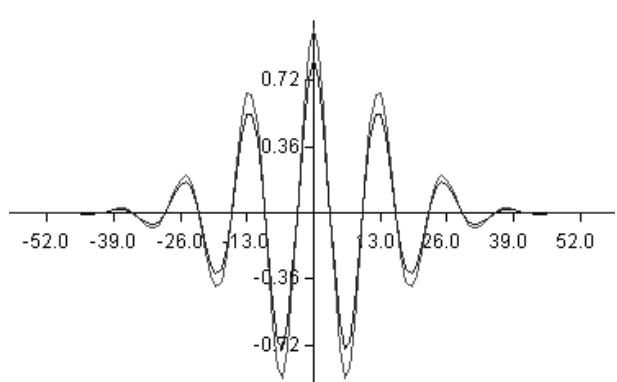


Рис. 2. Сигнал на первом приемнике и сигнал на втором приемнике с коэффициентом затухания $\alpha = 0.003$

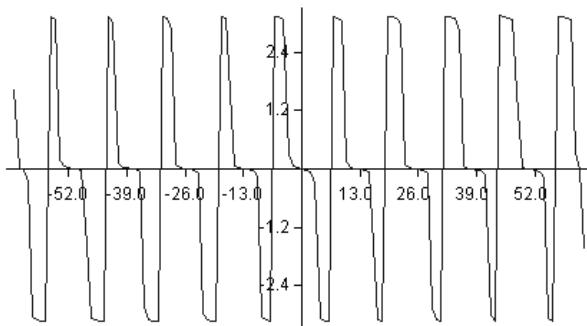


Рис. 3. Фаза сигнала на втором приемнике

при наложении друг на друга поперечной волны и волны Стоунли. Выражения (7) и (8) для модели (6) не учитывают возможности такого искажения фазы. Таким образом, цель предлагаемого итеративного процесса – точное восстановление сигнала (в том числе – фазы на втором приемнике, рис.3, включая оценку значений запаздывания и затухания сигнала, рис.4).

На рис.4 показаны амплитуды сигнала на первом приемнике и, получившегося в результате операции (12) и сдвига по времени для $k = 0.1$, сигнала на втором приемнике.

Эффективность работы предложенного итеративного алгоритма проверялась на зашумленных сигналах, показанных на рис.5. На этом рисунке представлены сигналы на первом и втором приемниках с коэффициентом затухания $\alpha = 0.003$ и запаздыванием $k=0.1$, искаженные гауссовым шумом с дисперсией $d = 0.01$.

Для генерации гауссова шума из равномерной случайной выборки использовался алгоритм Бокса-Мюллера. Генерация равномерной случайной выборки выполнялась с помощью алгоритма Ранеку [14].

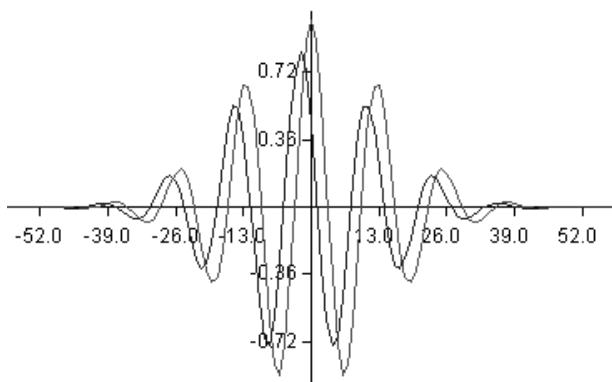


Рис. 4. Сигнал на первом приемнике и сигнал на втором приемнике с коэффициентом затухания $\alpha = 0.003$ и запаздыванием $k = 0.1$

В результате численных экспериментов было установлено, что предложенный алгоритм обладает различными скоростями сходимости для двух оцениваемых параметров. Алгоритм сходился к правильному значению запаздывания за две итерации, в то время как для точной оценки затухания требовалось 10–15 итераций. Такое различие можно объяснить различным характером воздействия шума на амплитуду и фазу сигнала.

Действительно, зашумленный сигнал на приемнике можно записать в виде:

$$s(\omega) = (A(\omega) \exp(-\alpha(\omega)l) + \varepsilon_a) \times \exp(i(\varphi(\omega) + k_r(\omega)l + \varepsilon_f)),$$

где ε_a и ε_f – шумовые компоненты амплитуды и фазы, соответственно. После логарифмирования вариация фазы записывается в виде:

$$\Delta(\Im[\ln s(\omega)]) = \Delta(\varepsilon_f).$$

Таким образом, преобразование не меняет вариацию фазы, и, следовательно, оценку скорости.

Используя разложение в ряд, вариация амплитуды записывается в виде:

$$\Delta(\Re[\ln s(\omega)]) \approx \Delta(\varepsilon_a) \frac{\exp(\alpha(\omega)l)}{A(\omega)}.$$

Т.е. вариация амплитуды растет экспоненциально с ростом расстояния l между приемниками, что затрудняет оценку величины затухания.

Для примера, представленного на рис.5, при оценке величины запаздывания алгоритм сошелся к правильному значению за две итерации, причем после первой итерации (что соответствует оценке гомоморфным методом)

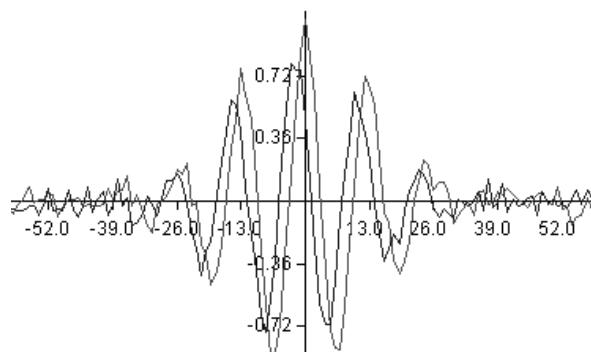


Рис. 5. Сигнал на первом приемнике и сигнал на втором приемнике с коэффициентом затухания $\alpha = 0.003$ и запаздыванием $k = 0.1$, искаженные гауссовым шумом с дисперсией $d = 0.01$

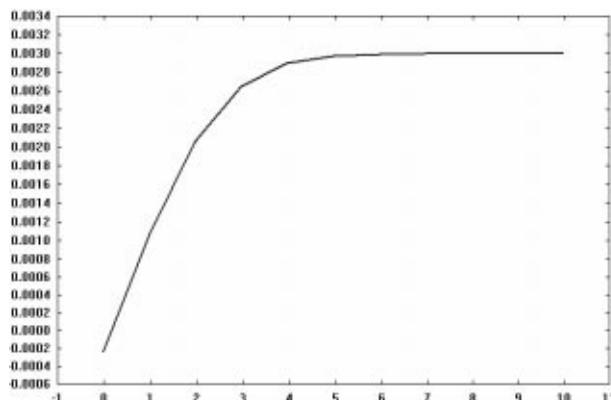


Рис.6. Зависимость оценки коэффициента затухаемости от числа итераций

ошибка составляла всего 2%. Для оценки коэффициента затухаемости потребовалось 10 итераций, чтобы получить ошибку менее 1% (рис.6).

Выводы

Численные эксперименты показали работоспособность предлагаемого метода. Особенno эффективен новый метод по сравнению с гомоморфным методом при оценке затухания зашумленной акустической волны. Предполагаются дальнейшие исследования, направленные на разделение различных типов волн, присутствующих в сигнале, и оценку их параметров.

СПИСОК ЛИТЕРАТУРЫ

1. Ellefsen K.J., Burns D.R., Cheng C.H. Homomorphic processing of the tube wave generated during acoustic logging// Geophysics. V.58. 1993.
2. Cheng C.H., Zhang J., Burns D.R. Effects of in-situ permeability on the propagation of Stoneley (tube) waves in a borehole // Geophysics. V.52. 1987.
3. Prony R. Essai experimental et analytique // L'ecole Polytech. V.1. 1975.
4. Lang S.W., Kurkjian A.L., McClellan J.H., Morris C.F., Parks T.W. Estimating slowness dispersion from arrays of sonic logging waveforms // Geophysics. V.52. 1987.
5. Ellefsen K.J., Cheng C.H., Tubman K.M. Estimating phase velocity and attenuation of guided waves in acoustic logging data // Geophysics. V.54. 1989.
6. Tang X.M. Predictive processing of array acoustic waveform data// Geophysics, Soc. of Expl. Geophys. V.62. 1997.
7. Chunduru R.K., Tang X.M. Automated velocity analysis of array acoustic waveform data// SEG Expanded Abstracts. 1998.
8. Gerchberg R.W., Saxton W.O. A practical algorithm for the determination of the phase from image and diffraction plane pictures // Optik. V.35. 1972. №237.
9. Fienup J.R. Phase retrieval algorithms: a comparison// Applied Optics. V.21. 1982. №15.
10. Методы компьютерной оптики// Подред. В.А. Сойфера. М.: Физматлит, 2000.
11. Kotlyar V.V., Seraphimovich P.G., Soifer V.A. An iterative weight-based method for calculating kinoforms // Proc. SPIE "Image Processing and Computer Optics". V.2363. 1994.
12. Cheng C.H., Toksoz M.N. Elastic wave propagation in a fluid-filled borehole and synthetic acoustic logs// Geophysics. V.46. 1981.
13. Kurkjian A.L. Numerical computation of individual far-field arrivals excited by an acoustic source in a borehole // Geophysics. V.50. 1985.
14. L'Ecuyer P. // Commun. ACM. 1988.

ITERATIVE ALGORITHM FOR CALCULATING VELOCITY AND ATTENUATION OF THE TUBE WAVE MEASURED DURING ACOUSTIC LOGGING

© 2001 N.L. Kazanskiy, P.G. Seraphimovich, S.I. Kharitonov

Image Processing Systems Institute of the Russian Academy of Sciences, Samara

We develop a new iterative algorithm to obtain estimates for the velocity and attenuation of the tube wave, which is measured during acoustic logging by a few receivers. A distinguishing feature of the algorithm is that optimization of the estimated parameters is performed in both the space plane and the frequency plane. On the one hand, it provides a high degree of tolerance of our algorithm of noise and on the other permits one to conduct dispersion analysis of wave modes, for example, Stoneley waves. A comparison of the new algorithm and the homomorphic algorithm is made.

ГРАДИЕНТНЫЙ МЕТОД ОПТИМИЗАЦИИ В ЗАДАЧЕ СИНТЕЗА БИНАРНОЙ МИКРООПТИКИ

© 2001 В.В. Котляр¹, Д.В. Нестеренко²

¹Институт систем обработки изображений РАН, г. Самара

²Самарский государственный аэрокосмический университет

Рассмотрено применение гибридного метода граничных – конечных элементов для синтеза двумерных дифракционных оптических элементов (ДОЭ). Гибридный метод позволяет эффективно рассчитывать векторную дифракцию от элементов микрооптики. В статье обсуждается применение градиентного метода оптимизации к матричной записи гибридного метода. Рассмотрен синтез бинарных линз, имеющих размеры порядка длины волны падающего света.

Введение

Когда минимальный характерный размер дифракционных структур приближается к длине волны освещения, необходимо применять векторную модель для анализа дифракции от них. Также векторная модель требуется, когда представляющая интерес область находится вблизи дифракционного элемента и, тем более, внутри него. Аналитические векторные модели дифракции могут быть определены для некоторых периодических структур, для апериодических структур граничные условия на электромагнитные поля делают аналитическую модель невозможной. Следовательно, для моделирования дифракции от апериодических дифракционных оптических элементов (ДОЭ) должны использоваться численные технологии, основанные на решении уравнений Maxwella. В работах [1], [2] и [3] описаны подходы к решению задач ближнего поля в свободном пространстве и в волноводе соответственно. Показано, что при использовании рефракционных микролинз наблюдается смещение максимума интенсивности от геометрического фокуса. При построении дифракционных микролинз из рефракционных наблюдается сдвиг положения фокуса и снижение эффективности дифракционной микролинзы [4]. Невозможно использование рефракционных микролинз в задачах с фокусировкой вне оптической оси вследствие их симметричности. В связи с этим возникает необходимость применения оптимизационных методов для синтеза дифракционных оптических элементов,

способных определить оптимальный профиль для заданного распределения интенсивности, принимая во внимание ограничения на форму рельефа. В данной статье мы представляем градиентный алгоритм для синтеза бинарной микрооптики.

Метод расчета

Используя гибридный метод конечных элементов – граничных элементов [1], рассмотрим дифракцию света в свободном пространстве.

Для неоднородной диэлектрической двумерной среды для двух случаев поляризации можно записать уравнение Гельмгольца для монохроматической электромагнитной волны:

$$\frac{\partial}{\partial x} \left[\frac{1}{\alpha} \frac{\partial \psi(x, y)}{\partial x} \right] + \frac{\partial}{\partial y} \left[\frac{1}{\alpha} \frac{\partial \psi(x, y)}{\partial y} \right] + k_0^2 \beta \psi(x, y) = 0, \quad (1)$$

где для TE-поляризации $\psi = E_z$, $\alpha = \frac{\mu}{\mu_0}$,

$\beta = \frac{\epsilon(x, y)}{\epsilon_0}$, $k_0 = \omega \sqrt{\mu_0 \epsilon_0} = \frac{\omega}{c} = \frac{2\pi}{\lambda_0}$ – волновое число в вакууме, λ_0 – длина волны падающего света, ω – циклическая частота излучения; ϵ , ϵ_0 , μ , μ_0 – диэлектрические и магнитные проницаемости вакуума и среды;

\vec{E}, \vec{H} – вектора электрической и магнитной напряженности электрического и магнитного полей; для TM-поляризации: $\psi = H_z$,

$$\alpha = \frac{\varepsilon(x, y)}{\varepsilon_0}, \beta = \frac{\mu}{\mu_0}.$$

Решение уравнения (1) равносильно минимизации функционала, рассчитываемого в ограниченной области Ω , включающей свою границу S [1]:

$$F(\psi) = \iint_{\Omega} \left[\frac{1}{\mu(x, y)} \nabla \psi(x, y) \nabla \psi(x, y) - k_0^2 \varepsilon(x, y) \psi^2(x, y) \right] dx dy - 2 \int_S \frac{\partial \psi}{\partial \mathbf{n}} \psi dl. \quad (2)$$

Приведем уравнение (2) к системе линейных алгебраических уравнений, используя следующую расчетную схему. Покроем область Ω треугольной сеткой. После этого в каждом из треугольников построим линейный полином. Для однозначного определения полиномов достаточно задать их значения в вершинах треугольников. Тогда базисом будет система функций $\omega_k(x, y)$. Геометрически функция $\omega_k(x, y)$ представляется как треугольная пирамида с центром в точке (x_k, y_k) .

Представим поле ψ через данный базис как

$$\psi(x, y) = \sum_{k=1}^N \psi_k \omega_k^\Omega(x, y), \quad (3)$$

где $k = 1, \dots, N + M$; N – число внутренних узлов сетки, M – число граничных узлов, приходим к системе линейных алгебраических уравнений:

Подставляя разложение (3) в функционал (2) и применяя разделение общего поля на сумму падающего и рассеянного полей, приравниваем нулю производные

$$\frac{\partial F(\psi^{sc})}{\partial \psi_k^{sc}} = 0, \quad (x, y) \in \Omega \quad (4)$$

и приходим к системе линейных алгебраических уравнений [5]:

$$\mathbf{A}(\psi_\Omega^{sc} + \psi_\Omega^{in}) + \mathbf{B}(\mathbf{v}^{sc} + \mathbf{v}^{in}) = 0, \quad (5)$$

где $\psi_\Omega = (\psi_1, \dots, \psi_N)$; $\psi_k = \psi_k^{sc} + \psi_k^{in}$, ψ_k^{sc} , ψ_k^{in} – полное, неизвестное рассеянное и известное падающее световые поля; $\mathbf{v} = (v_1, \dots, v_M)$; $v_k = \frac{\partial \psi_k}{\partial \mathbf{n}}$ – производная поля по внеш-

ней нормали на границе S .

Элементы матрицы \mathbf{A} вычисляются по формулам:

$$a_{k,j} = \iint_{\Omega_{k,j}} \left(\frac{1}{\mu_k} \left[\frac{\partial \omega_k^{\Omega_{k,j}}(x, y)}{\partial x} \frac{\partial \omega_j^{\Omega_{k,j}}(x, y)}{\partial x} + \frac{\partial \omega_k^{\Omega_{k,j}}(x, y)}{\partial y} \frac{\partial \omega_j^{\Omega_{k,j}}(x, y)}{\partial y} \right] - k_0^2 \varepsilon_k \omega_k^{\Omega_{k,j}}(x, y) \omega_j^{\Omega_{k,j}}(x, y) \right) dxdy, \quad k, j = 1, \dots, N + M, \quad (6)$$

где $\Omega_{k,j}$ – треугольная область, содержащая k и j узлы сетки.

Элементы матрицы \mathbf{B} вычисляются по формулам:

$$b_{m,p} = - \oint_{S_{m,p}} \omega_m^{S_{m,p}} \omega_p^{S_{m,p}} dl, \quad m, p = 1, \dots, M. \quad (7)$$

Получаем систему линейных уравнений, которую необходимо дополнить соотношением для поля и его производных на границе [6]:

$$\oint_S \left\{ \frac{\partial \psi^{sc}}{\partial \mathbf{n}} G_2 - \psi^{sc} \frac{\partial G_2}{\partial \mathbf{n}} \right\} dl = -\frac{1}{2} \psi^{sc}, \quad (x, y) \in S, \quad (8)$$

где функция Грина для двумерных световых полей (цилиндрическая волна) равна функции Ханкеля второго рода нулевого порядка:

$$G(x, x', y, y') = -\frac{i}{4} H_0^{(2)} \left(k_0 \sqrt{(x-x')^2 + (y-y')^2} \right). \quad (9)$$

Равенство (8) может быть представлено в матричном виде:

$$[\mathbf{C}] \psi_S^{sc} + [\mathbf{D}] \mathbf{v}_S^{sc} = 0, \quad (10)$$

где элементы матриц $[\mathbf{C}]$ и $[\mathbf{D}]$:

$$c_{m,p} = -h \int_0^1 \omega_p^S (\rho_p + [\rho_{p+1} - \rho_p] \xi) \times \frac{\partial G(\rho_m, \rho_p + [\rho_{p+1} - \rho_p] \xi)}{\partial \mathbf{n}'} d\xi - h \int_0^1 \omega_p^S (\rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) \times \frac{\partial G(\rho_m, \rho_{p-1} + [\rho_p - \rho_{p-1}] \xi)}{\partial \mathbf{n}'} d\xi + \gamma_m \delta_{mp}, \quad (11)$$

$$\begin{aligned} d_{m,p} = & h \int_0^1 \omega_p^S (\rho_p + [\rho_{p+1} - \rho_p] \xi) \times \\ & \times G(\rho_m, \rho_p + [\rho_{p+1} - \rho_p] \xi) d\xi + \\ & + h \int_0^1 \omega_p^S (\rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) \times \\ & \times G(\rho_m, \rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) d\xi, \end{aligned} \quad (12)$$

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

где \mathbf{n}' – вектор внутренней нормали; $\rho_m = (x_m, y_m) \in S$ – координаты m -ой точки; h – шаг сетки; параметр $\gamma_m = \phi_m / (2\pi)$, где ϕ_m – внутренний угол в точке (x_m, y_m) границы S .

После линеаризации уравнения (8) получаем полную систему линейных алгебраических уравнений вида:

$$\begin{aligned} & \begin{bmatrix} [\mathbf{A}_{G,G}] & [\mathbf{A}_{S,G}] & 0 \\ [\mathbf{A}_{G,S}] & [\mathbf{A}_{S,S}] & [\mathbf{B}] \\ 0 & [\mathbf{C}] & [\mathbf{D}] \end{bmatrix} \begin{bmatrix} \psi_G^{sc} \\ \psi_S^{sc} \\ \mathbf{v}^{sc} \end{bmatrix} = \\ & = - \begin{bmatrix} [\mathbf{A}_{G,G}] & [\mathbf{A}_{S,G}] & 0 \\ [\mathbf{A}_{G,S}] & [\mathbf{A}_{S,S}] & [\mathbf{B}] \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \psi_G^{in} \\ \psi_S^{in} \\ \mathbf{v}^{in} \end{bmatrix}, \end{aligned} \quad (13)$$

где G – область Ω без границы S , подматрица $\mathbf{A}_{G,G}$ размерностью $N \times N$ включает коэффициенты вклада внутренних узлов, подматрицы $\mathbf{A}_{G,S}$ и $\mathbf{A}_{S,G}$ размерностью $N \times M$ и $M \times N$ включают коэффициенты вклада внутренних узлов с внешними, подматрица $\mathbf{A}_{S,S}$ размерностью $M \times M$ включает коэффициенты вклада внешних узлов, ψ_G и ψ_S – вектора значений поля во внутренних и граничных узлах сетки.

Представим систему уравнений (13) в матричном виде:

$$\mathbf{T} \psi^{sc} = \mathbf{U} \psi^{in}, \quad (14)$$

$$\psi^{sc} = \mathbf{T}^{-1} \mathbf{U} \psi^{in}, \quad (15)$$

где

$$\mathbf{T} = \begin{bmatrix} [\mathbf{A}_{G,G}] & [\mathbf{A}_{S,G}] & 0 \\ [\mathbf{A}_{G,S}] & [\mathbf{A}_{S,S}] & [\mathbf{B}] \\ 0 & [\mathbf{C}] & [\mathbf{D}] \end{bmatrix}, \quad \psi^{sc} = \begin{bmatrix} \psi_G^{sc} \\ \psi_S^{sc} \\ \mathbf{v}^{sc} \end{bmatrix}$$

$$\mathbf{U} = - \begin{bmatrix} [\mathbf{A}_{G,G}] & [\mathbf{A}_{S,G}] & 0 \\ [\mathbf{A}_{G,S}] & [\mathbf{A}_{S,S}] & [\mathbf{B}] \\ 0 & 0 & 0 \end{bmatrix}, \quad \psi^{in} = \begin{bmatrix} \psi_G^{in} \\ \psi_S^{in} \\ \mathbf{v}^{in} \end{bmatrix}.$$

После нахождения поля ψ^{sc} на границе S поле в любой точке окружающего пространства находится по формуле:

$$\psi(x, y) = \psi^{in}(x, y) - \oint_S \left\{ \frac{\partial \psi^{sc}}{\partial \mathbf{n}} G_2 - \psi^{sc} \frac{\partial G_2}{\partial \mathbf{n}} \right\} dl, \quad (x, y) \notin \Omega \quad (16)$$

В матричном виде поле в окружающем пространстве представляется в виде:

$$\psi = \psi^{in} + \mathbf{W} \psi^{sc} = \psi^{in} + \mathbf{WT}^{-1} \mathbf{U} \psi^{in}. \quad (17)$$

где матрица \mathbf{W} размером $1 \times (N + 2M)$ равна:

$$\mathbf{W} = [0 \quad [\mathbf{E}] \quad [\mathbf{F}]],$$

где элементы подматриц \mathbf{E} и \mathbf{F} :

$$\begin{aligned} e_p(x, y) = & h \int_0^1 \omega_p^S (\rho_p + [\rho_{p+1} - \rho_p] \xi) \times \\ & \times \frac{\partial G((x, y), \rho_p + [\rho_{p+1} - \rho_p] \xi)}{\partial \mathbf{n}'} d\xi + \\ & + h \int_0^1 \omega_p^S (\rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) \times \\ & \times \frac{\partial G((x, y), \rho_{p-1} + [\rho_p - \rho_{p-1}] \xi)}{\partial \mathbf{n}'} d\xi, \end{aligned} \quad (18)$$

$$\begin{aligned} f_p(x, y) = & -h \int_0^1 \omega_p^S (\rho_p + [\rho_{p+1} - \rho_p] \xi) \times \\ & \times G((x, y), \rho_p + [\rho_{p+1} - \rho_p] \xi) d\xi - \\ & - h \int_0^1 \omega_p^S (\rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) \times \\ & \times G((x, y), \rho_{p-1} + [\rho_p - \rho_{p-1}] \xi) d\xi, \end{aligned} \quad (19)$$

$p = 1, \dots, M$.

Метод оптимизации

Определение оптимального профиля состоит в расчете параметров (y_1, \dots, y_k) профиля бинарной микрооптики (рис.1). Это достигается разработкой алгоритма систематического поиска оптимального решения. В качестве метода оптимизации мы выбрали градиентный метод.

Для построения процедуры расчета введем функцию ошибки $\varepsilon(\mathbf{p})$, характеризующую отличие рассчитанных значений интенсивности \mathbf{I}_n от требуемых значений \mathbf{I}' [7]:

$$\varepsilon(\mathbf{p}) = \varepsilon(\mathbf{I}(\mathbf{p}), \mathbf{I}'),$$

где $\mathbf{p} = (y_1, \dots, y_k)$ – вектор параметров бинарного профиля; k – число параметров профи-

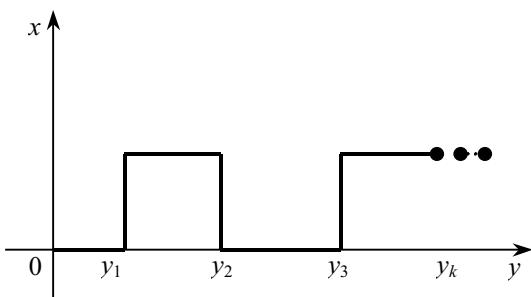


Рис. 1. Бинарный профиль микрооптики

ля; $(\mathbf{I}_n)_{-M}^M, (\mathbf{I}')_{-M}^M$ – вектора рассчитанных и требуемых значений интенсивности.

Градиентная процедура минимизации функции ошибки $\varepsilon(\mathbf{p})$ состоит в итерационной коррекции параметров профиля микрооптики по правилу:

$$\mathbf{p}_{n+1} = \mathbf{p}_n - t \cdot \nabla \varepsilon(\mathbf{p}_n),$$

где n – номер итерации, t – шаг градиентного алгоритма,

$$\nabla \varepsilon(\mathbf{p}) = \left(\frac{\partial \varepsilon(\mathbf{p})}{\partial y_1}, \dots, \frac{\partial \varepsilon(\mathbf{p})}{\partial y_k} \right)^T$$

градиент функции ошибки.

Точно определить шаг t не представляется возможным, т.к. неизвестно значение минимума функции ошибки $\varepsilon(\mathbf{p})$. Поэтому после вычисления градиента функции ошибки $\nabla \varepsilon(\mathbf{p})$ из исходной точки \mathbf{p}_n делается шаг заданной длины, находятся координаты новой точки \mathbf{p}_{n+1} . Если значение $\varepsilon(\mathbf{p}_{n+1})$ лучше, чем $\varepsilon(\mathbf{p}_n)$, то шаг считается удачным, и новая точка становится исходной, иначе – шаг неудачный, и его величина уменьшается. Если величина шага становится меньше определенного предела, то алгоритм останавливается, и последняя найденная точка – решение, иначе процедура уменьшения шага повторяется.

Вычисление градиента функции ошибки $\nabla \varepsilon(\mathbf{p})$ может вестись двумя способами. Первый заключается в непосредственном

нахождении частных производных $\frac{\partial \varepsilon(\mathbf{p})}{\partial x_i}$ в разностном виде.

Второй способ состоит в представлении $\varepsilon(\mathbf{p})$ через вектор электрического поля $\psi(\mathbf{p})$, учитывая зависимость $\mathbf{I} = \psi \psi^*$.

Электрическое поле ψ определяется из системы (17), где $\mathbf{T}(\mathbf{p}), \mathbf{U}(\mathbf{p}), \mathbf{W}$ – комплекс-

ные матрицы; $\psi(\mathbf{p})$ – комплексный вектор.

Из (15) следует:

$$\frac{\partial \psi^{sc}(\mathbf{p})}{\partial p_i} = \mathbf{T}^{-1} \frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \psi^{in} + \frac{\partial \mathbf{T}^{-1}(\mathbf{p})}{\partial p_i} \mathbf{U} \psi^{in},$$

отсюда:

$$\frac{\partial \mathbf{T}^{-1}(\mathbf{p})}{\partial p_i} = \left[\frac{\partial \psi^{sc}(\mathbf{p})}{\partial p_i} - \mathbf{T}^{-1} \frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \psi^{in} \right] (\mathbf{U} \psi^{in})^{-1}. \quad (20)$$

Из (14) следует:

$$\begin{aligned} \frac{\partial (\mathbf{T}(\mathbf{p}) \psi^{sc}(\mathbf{p}))}{\partial p_i} &= \mathbf{T} \frac{\partial \psi^{sc}(\mathbf{p})}{\partial p_i} + \frac{\partial \mathbf{T}(\mathbf{p})}{\partial p_i} \psi^{sc} = \\ &= \frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \psi^{in}, \end{aligned}$$

отсюда

$$\frac{\partial \psi^{sc}(\mathbf{p})}{\partial p_i} = \mathbf{T}^{-1} \left[\frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \psi^{in} - \frac{\partial \mathbf{T}(\mathbf{p})}{\partial p_i} \psi^{sc} \right]. \quad (21)$$

Подставляя (21) в (20), получим:

$$\frac{\partial \mathbf{T}^{-1}(\mathbf{p})}{\partial p_i} = -\mathbf{T}^{-1} \frac{\partial \mathbf{T}(\mathbf{p})}{\partial p_i} \psi^{sc} (\mathbf{U} \psi^{in})^{-1}. \quad (22)$$

Найдем выражение для производной поля в любой точке окружающего пространства по параметру \mathbf{p} из системы (17), используя (22):

$$\begin{aligned} \frac{\partial \psi(\mathbf{p})}{\partial p_i} &= \frac{\partial (\mathbf{W} \mathbf{T}^{-1}(\mathbf{p}) \mathbf{U}(\mathbf{p}))}{\partial p_i} \psi^{in} = \\ &= \mathbf{W} \left[\frac{\partial \mathbf{T}^{-1}(\mathbf{p})}{\partial p_i} \mathbf{U} + \mathbf{T}^{-1} \frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \right] \psi^{in} = \\ &= \mathbf{W} \mathbf{T}^{-1} \left[-\frac{\partial \mathbf{T}(\mathbf{p})}{\partial p_i} \psi^{sc} + \frac{\partial \mathbf{U}(\mathbf{p})}{\partial p_i} \psi^{in} \right]. \end{aligned} \quad (23)$$

Учитывая (13) можно записать:

$$\begin{aligned} \frac{\partial \psi(\mathbf{p})}{\partial p_i} &= \\ &= \mathbf{W} \mathbf{T}^{-1} \left[\frac{\partial \mathbf{A}_{I,I}(\mathbf{p})}{\partial p_i} \psi^{in} - \frac{\partial \mathbf{A}_{I,I}(\mathbf{p})}{\partial p_i} \psi^{sc} \right] = \\ &= \mathbf{W} \mathbf{T}^{-1} \frac{\partial \mathbf{A}_{I,I}(\mathbf{p})}{\partial p_i} (\psi^{in} - \psi^{sc}) \end{aligned} \quad (24)$$

Элементы вектора градиента $\frac{\partial \varepsilon(\mathbf{p})}{\partial p_i}$, где p_i – компонента вектора \mathbf{p} , имеют вид:

$$\frac{\partial \varepsilon(\mathbf{p})}{\partial p_i} = \sum_j \frac{\partial \varepsilon(\mathbf{I}, \mathbf{I}')}{\partial I_j} \frac{\partial I_j(\mathbf{p})}{\partial p_i},$$

$$\frac{\partial I_j(\mathbf{p})}{\partial p_i} = \psi_j \frac{\partial \psi_j^*(\mathbf{p})}{\partial p_i} + \psi_j^* \frac{\partial \psi_j(\mathbf{p})}{\partial p_i}. \quad (25)$$

Для определения $\nabla \varepsilon(\mathbf{p})$ уравнение (24) подставляется в уравнение (25).

Отметим, что данный подход позволяет определять любые другие параметры, подлежащие оптимизации.

Проведен ряд экспериментов, показывающий эффективность использования градиентного метода оптимизации бинарных микролинз.

Конструкция дифракционной линзы

Рассмотрим цилиндрическую рефракционную микролинзу (рис.2).

Положим, что ее длина L_x настолько больше ширины L_y , что длину линзы по сравнению с шириной можно считать бесконечной. Рассмотрим численный пример. Определим апертуру линзы как $a = 8$ мкм, радиус кривизны $R = 5$ мкм, показатель преломления $n = 2$ для длины волны $\lambda_0 = 1$ мкм, что примерно соответствует хлориду серебра ($n_{AgCl} = 2,02239$ для $\lambda_0 = 1$ мкм).

Для данной цилиндрической линзы построим линзу с фазовой функцией, приведенной по модулю 2π , и дифракционную линзу с двумя степенями градации фазовой функции (рис.3).

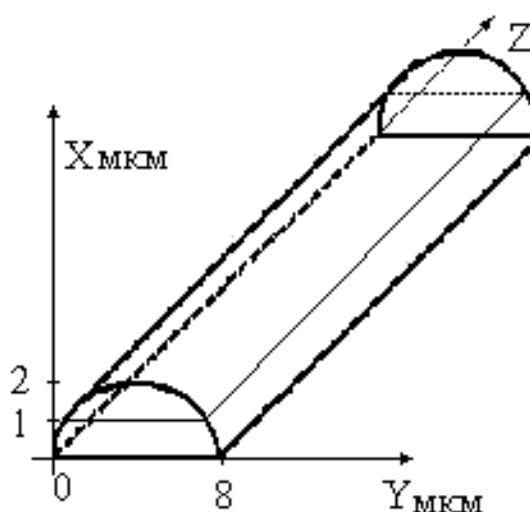


Рис. 2. Цилиндрическая рефракционная микролинза

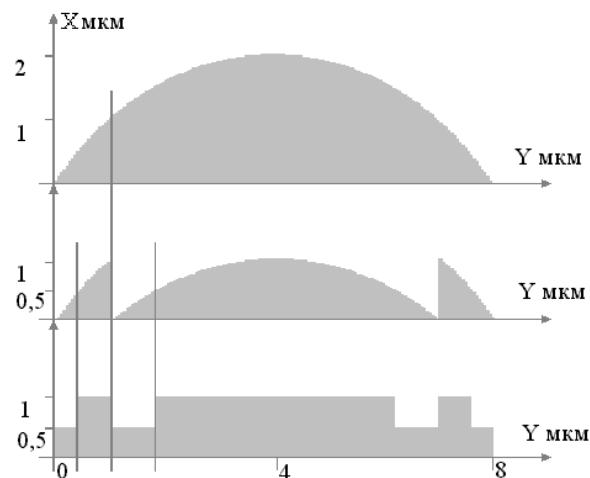


Рис. 3. Построение линзы с фазовой функцией, приведенной по модулю 2π , и бинарной дифракционной линзы

Численное моделирование и оптимизация

Поставим вычислительный эксперимент, в котором плоская волна ТЕ-поляризации падает на бинарные линзы в свободном пространстве. В качестве оптимизируемых параметров профиля брались координаты выступов.

На рис.4 показано распределение интенсивности дифракции плоской волны на построенной однофокусной бинарной линзе (рис.3) с дифракционной эффективностью 54,3 % и энергетической эффективностью 43,5 %. Ее профиль был использован в качестве начального для процедуры оптимизации и был оптимизирован до дифракционной эффективности 72,3 % и энергетической эффективности 64,0 %.

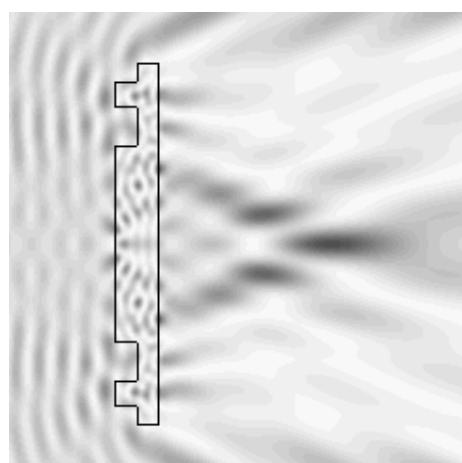


Рис. 4. Распределение интенсивности света при дифракции на бинарной линзе начального построения

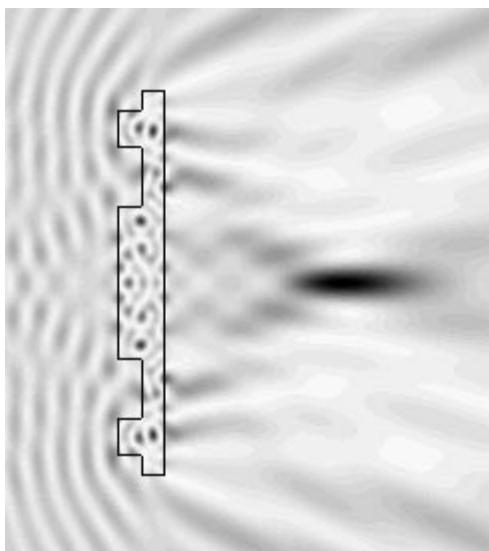


Рис. 5. Распределение интенсивности света при дифракции на оптимизированной бинарной линзе

Под энергетической эффективностью мы понимаем отношение энергии, попавшей в область фокуса, к энергии, падающей на микролинзу. Дифракционная эффективность – это отношение энергии, попавшей в область фокуса, к энергии, прошедшей микролинзу. Интенсивность – это квадрат модуля проекции электрического вектора на ось z .

Рассмотрим синтез бинарной линзы, дающей в плоскости, располагающейся на расстоянии 3,5 мкм, три максимума равной интенсивности шириной 1 мкм: один в центре и два на расстоянии 2,5 мкм от центра. Были использованы несколько начальных профилей линзы с тремя выступами, расположившимися симметрично относительно

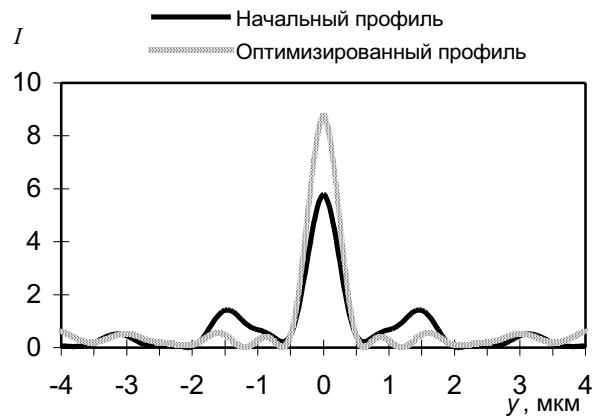


Рис. 6. Распределение интенсивности света в фокальной плоскости бинарных линз: с начальным профилем и оптимизированным

оси симметрии, для нахождения по возможности всех локальных экстремумов функции ошибки. В результате работы процедуры оптимизации было найдено два вида профиля, дающих максимальные значения эффективности. На рис.7 показано распределение интенсивности дифракции плоской волны на оптимизированных трехфокусных бинарных линзах. Линза, показанная на рис.7 a , дает дифракционную эффективность 86,2 % и энергетическую эффективность 79,3 %. Линза, показанная на рис.7 b , дает дифракционную эффективность 85,8 % и энергетическую эффективность 84,1 %. Дифракционная эффективность в этом случае равна отношению энергии в трех заданных порядках к энергии, прошедшей микролинзу. Под энергией понимается сумма интенсивностей на выбранной линии вдоль оси y .

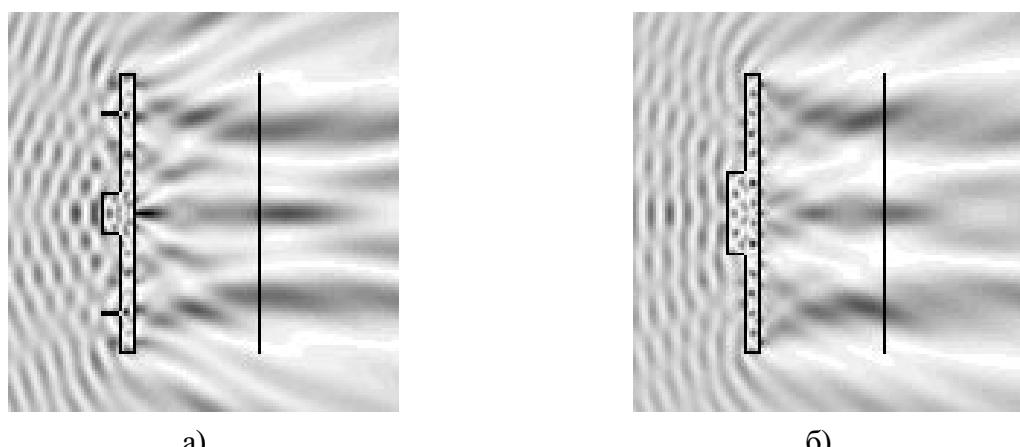


Рис. 7. Распределение интенсивности света при дифракции на оптимизированных трехфокусных бинарных линзах

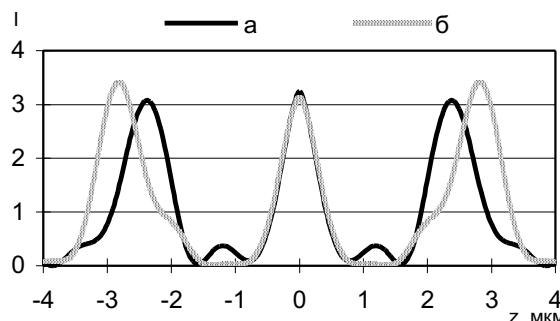


Рис. 8. Распределение интенсивности света при дифракции на оптимизированных трехфокусных бинарных линзах в указанной плоскости

Таким образом, представленный алгоритм позволил найти профиль трехфокусной бинарной линзы (рис.7б) с высокой дифракционной эффективностью и глобальным максимумом энергетической эффективности, более простой в физической реализации, чем линзы с начальным профилем. Линза, показанная на рис.7а, с глобальным максимумом дифракционной эффективности дает форму поля дифракции наиболее соответствующую заданной.

Выводы

В работе рассмотрен градиентный метод расчета профиля диэлектрического микрообъекта с использованием совместного метода конечных – граничных элементов. Работоспособность метода продемонстрирована на примере расчета бинарных однофокусной и трехфокусной микролинз.

СПИСОК ЛИТЕРАТУРЫ

1. Котляр В.В., Нестеренко Д.В. Анализ задачи дифракции света на микрооптике гибридным методом конечных элементов – граничных элементов // Компьютерная оптика. 2000. Вып.20.
2. Головашкин Д.Л., Сойфер В.А. Анализ прохождения электромагнитного излучения через дифракционную линзу // Автометрия. 1999. №6.
3. Kotlyar V. V., Nesterenko D. V. Analyzing the light diffraction by binary micro-optics using a combination of boundary element method and finite element method / Laser physics and spectroscopy, Saratov fall meeting, Proceedings of SPIE. Vol. 4003. 2000.
4. Головашкин Д.Л., Котляр В.В., Нестеренко Д.В. Анализ дифракции света на микролинзах в свободном пространстве и волноводе // Компьютерная оптика. 2001. Вып.21.
5. Kotlyar V.V., Nesterenko D.V. Modeling the light diffraction by microoptics elements using the finite element method / Laser physics and spectroscopy, Saratov fall meeting, Proceedings of SPIE. Vol. 4002. 1999.
6. Бреббия К., Теллес Ж., Вроубел Л. Методы граничных элементов. М.: Мир, 1987.
7. Методы компьютерной оптики. Под ред. В.А. Сойфера. М.: Физматлит, 2000.

SYNTHESIS OF BYNARY MICROOPTICS BY GRADIENT METHOD OF OPTIMIZATION

© 2001 V.V. Kotlyar¹, D.V. Nesterenko²

¹Image Processing Systems Institute of Russian Academy of Science, Samara

²Samara State Aerospace University

We apply a gradient method as synthesis algorithm to the two-dimensional hybrid finite element-boundary element method as vector model of diffraction for the design of DOEs. The hybrid method is capable of modeling inhomogeneous DOEs in unbounded freespace in computationally efficient manner. In this paper we discuss the application of gradient method of optimization to matrix notation of hybrid method. In such application it is possible to analyze profiles with large number of features. This allows to overcome the limitations of time of calculation dependent on amount of modifications of DOE. We use the gradient method to design of binary-phase lenses that has subwavelength features. Although we have considered only binary-phase lenses, presented gradient method is capable of design inhomogeneous DOEs or coatings.

УДК 535.31

ИССЛЕДОВАНИЕ МАТРИЧНОГО МЕТОДА ВЫЧИСЛЕНИЯ ВЫтянутых СФЕРОИДАЛЬНЫХ ФУНКЦИЙ НУЛЕВОГО ПОРЯДКА

© 2001 С.Н. Хонина

Институт систем обработки изображений РАН, г. Самара

Проводится исследование матричного метода вычисления вытянутых сфериодальных функций нулевого порядка, реализованного с использованием программного обеспечения Matlab. Точность вычисления определяется по выполнению основных свойств вытянутых сфероидальных функций, таких как инвариантность к интегральному преобразованию с sinc-ядром и преобразованию Фурье. Предложен алгоритм правильного упорядочивания функций по собственным значениям. Определены границы работоспособности данной реализации метода.

Введение

Вытянутые сфероидальные функции (ВСФ) нулевого порядка $\psi_n(c, x)$ являются собственными функциями интегрального уравнения с sinc-ядром [1-4] и преобразования Фурье [5], что позволяет эффективно использовать их во многих областях науки и техники: теория синтеза антенн, восстановление объектов по изображению, сверхразрешение, теория резонаторов [1-10].

Подходы к вычислению ВСФ произвольного порядка отличаются сложностью и трудоемкостью: представление через ряды по другим, более простым базисам [11-13], решение дифференциального уравнения методом ортогональной дифференциальной прогонки с переносом граничных условий из особых точек [14]. Для расчета ВСФ нулевого порядка можно дополнительно воспользоваться присущими только им свойствами и решать интегральное уравнения на собственные значения, например, итерационно [15].

Учитывая трудоемкость упомянутых подходов, в одном случае [16,17] и чувствительность к начальному приближению, в другом [15], желательно иметь простой и устойчивый метод расчета ВСФ нулевого порядка. В работах [8,10] предлагается вместо интегрального уравнения на собственные значения решать матричное уравнение, полученное путем дискретизации.

В данной работе проводится исследование матричного метода вычисления ВСФ нулевого порядка, реализованного с исполь-

зованием программного обеспечения Matlab. Работоспособность оценивается по выполнению основных свойств рассчитанных функций: ортогональность, инвариантность к интегральному преобразованию с sinc-ядром и преобразованию Фурье. Правильность упорядочивания функций определяется по убыванию собственных значений. Целью работы является определение границ работоспособности метода.

Основные свойства ВСФ нулевого порядка

ВСФ произвольного порядка являются решением скалярного волнового уравнения в вытянутых сфероидальных координатах. После разделения переменных трехмерного волнового уравнения в этих координатах возникает дифференциальное уравнение [18]:

$$\frac{d}{dx} (1-x^2) \frac{d\psi(x)}{dx} + \left(\chi - c^2 x^2 - \frac{m^2}{1-x^2} \right) \psi(x) = 0, \quad (1)$$

которое имеет непрерывное вещественное решение $\psi(x)$, ограниченное при любых x . С точностью до константы решение единственное. Здесь и далее, c – произвольное положительное число, m – неотрицательное целое.

Обозначим специальные значения χ , при которых существуют непрерывные решения через $\chi_{m,n}$, $n = m, m+1, \dots$. Эти собственные значения могут быть так обозначены, поскольку $0 < \chi_{m,m} < \chi_{m,m+1} < \dots$. Соответствующие им решения обозначим $\psi_{m,n}(x)$. Собственные функции $\psi_{m,n}(x)$ четные, если $n-m$ – четное число и нечетные, если $n-m$ – нечетное

число. Также они имеют ровно n -т нулей на открытом интервале $|x|<1$.

Функции $\psi_{m,n}(x)$ известны [11–13] как, вытянутые сфероидальные волновые функции порядка m . Однако особый интерес представляют ВСФ нулевого порядка. Благодаря ряду замечательных свойств [1-5] они играют важную роль во многих прикладных задачах [1-10].

Функции $\psi_{0,n}(x)$, которые далее будем обозначать просто $\psi_n(x)$ удовлетворяют интегральному уравнению [18]:

$$\lambda_n \psi_n(y) = \int_{-1}^1 \frac{\text{sinc}(y-x)}{\pi(y-x)} \psi_n(x) dx, \quad (2)$$

где λ_n – собственные числа:

$$1 > \lambda_0 > \lambda_1 > \dots > 0. \quad (3)$$

При этом собственные числа интегрального преобразования определяют количество энергии соответствующей собственной функции, концентрирующейся на данном ограниченном интервале. Собственные числа близкие к единице показывают, что данная ВСВФ имеет за пределами данного интервала малую долю энергии. Если рассматривать λ_n как функцию от n , то собственные числа имеют характерную особенность поведения: до некоторого $n_0=2c/\pi$ они имеют значения близкие к единице, а после – резко спадают до нуля.

ВСФ нулевого порядка являются также собственными функциями преобразования Фурье на ограниченном интервале [5,7,18]:

$$\alpha_n \psi_n(y) = \int_{-1}^1 \exp[i c x y] \psi_n(x) dx, \\ \lambda_n = \frac{C}{2\pi} |\alpha_n|^2, \quad (4)$$

В [1] предложено следующее правило нормировки:

$$\int_{-T/2}^{T/2} \psi_n(x) \psi_m(x) dx = \delta_{mn} \lambda_n, \\ \int_{-\infty}^{+\infty} \psi_n(x) \psi_m(x) dx = \delta_{mn}. \quad (5)$$

Также ВСФ нулевого порядка являются собственными функциями для sinc-преобразования (2) и преобразования Фурье (4) на

бесконечном интервале [5, 7, 18].

Исследование матричного метода

Рассмотрим модифицированное интегральное уравнение (2):

$$\lambda_n \psi_n(y) = \int_{-T/2}^{T/2} \psi_n(x) \frac{\sin\left[\frac{2c}{T}(y-x)\right]}{\pi(y-x)} dx, \quad (6)$$

где параметр $c = \pi WT$, T – ширина интервала в объектной области, W – ширина интервала в частотной области.

Пусть N – число точек дискретизации, тогда получим дискретный вариант уравнения (6)

$$\lambda_n \psi_n(y_k) = \sum_{l=0}^{N-1} \psi_n(x_l) \frac{\sin\left[\frac{2c}{T}(y_k - x_l)\right]}{\pi(y_k - x_l)}, \\ k = \overline{0, N}, \quad (7)$$

который можно записать в матричном виде:

$$\lambda_n \psi_n = S \cdot \psi_n. \quad (8)$$

Решать уравнение (8) можно с помощью прикладных пакетов программ, например, Matlab. Заметим, что матрица S квадратная, и число точек дискретизации N определяет также общее количество вычисляемых функций.

Проверка свойства ортогональности.

Для проверки свойства ортогональности функций, рассчитанных матричным методом (7)- (8) вычислялись следующие величины:

$$\Delta_{nm} = \frac{\sum_{k=0}^{N-1} \psi_n(x_k) \psi_m(x_k)}{\sum_{k=0}^{N-1} \psi_n^2(x_k)}, \quad n, m = \overline{0, M}. \quad (9)$$

Критерием выполнения свойства ортогональности (5) может служить величина:

$$D = \sum_{n=0}^M \sum_{m \neq n} |\Delta_{nm}| + \sum_{n=0}^M |1 - \Delta_{nn}|, \quad (10)$$

которая в идеале д.б. равна нулю. В таблице 1 приведены значения величины D для десяти первых функций (в ур. (10) $M=9$) с различным параметром c .

Таблица 1. Выполнение свойства ортогональности для $M = 9, N = 101$

	$c=2$	$c=5$	$c=10$	$c=20$
D	$1,0463 \cdot 10^{-14}$	$1,4003 \cdot 10^{-14}$	$9,6979 \cdot 10^{-15}$	$1,9848 \cdot 10^{-14}$

Из таблицы 1 можно сделать вывод о выполнении свойства ортогональности для полученных собственных функций.

Проверка выполнения свойства инвариантности к преобразованию с sinc-ядром.

Выполнение этого свойства оценивалось по среднеквадратичному отклонению "выходной" $\hat{\psi}_n(Y)$ функции после преобразования (7) от "входной" $\psi_n(x)$:

$$\delta = \sqrt{\frac{\sum_{k=0}^{N-1} [\hat{\psi}_n(x_k) - \psi_n(x_k)]^2}{\sum_{k=0}^{N-1} \psi_n^2(x_k)}}, \quad (11)$$

где функции $\hat{\psi}_n(Y)$, $\psi_n(x)$ предварительно выровнены по максимуму.

Результаты для некоторых значений параметра $c=10, 20$ приведены в таблице 2 и 3,

соответственно. Значения λ_n в таблицах 2 и 3 вычислялись исходя из ур. (2) как отношение максимумов функций $\hat{\psi}_n(Y)$ и $\psi_n(x)$ на одинаковых интервалах, то есть

$$\lambda_n = \frac{\max_{[-T/2T/2]} \{\hat{\psi}_n(Y)\}}{\max_{[-T/2T/2]} \{\psi_n(x)\}}. \quad (12)$$

Из таблиц 2, 3 видно, что не выполняется свойство о не возрастании собственных чисел λ_n . Ранжировка полученных функций была проведена в соответствии с (3), что скорректировало порядок, установленный встроенной функцией Matlab.

Из таблиц 2 и 3 также видно, что свойство (2) нарушается, начиная с некоторого номера n_c , зависящего от параметра c , при чем $n_c > n_0$. Так, например,

- для $c=2$: $n_0=1, n_c=8$,
- для $c=5$: $n_0=3, n_c=11$,

Таблица 2. Значения λ_n и ошибки δ (11) для $c=10, N=101, M=18$

Номер функции (Matlab)	Скорректированный номер функции	Значение λ_n	Ошибка, δ
6	0	0,99999	$1,1781 \cdot 10^{-15}$
5	1	0,99998	$7,7543 \cdot 10^{-16}$
4	2	0,99993	$7,2443 \cdot 10^{-16}$
3	3	0,9982	$6,5013 \cdot 10^{-16}$
2	4	0,9778	$7,3094 \cdot 10^{-16}$
0	5	0,8419	$7,0206 \cdot 10^{-16}$
1	6	0,4678	$1,0162 \cdot 10^{-15}$
7	7	0,1254	$2,7448 \cdot 10^{-15}$
8	8	0,0173	$1,0685 \cdot 10^{-14}$
9	9	0,0016	$7,1481 \cdot 10^{-14}$
10	10	0,0001	$1,2566 \cdot 10^{-12}$
11	11	$5,7821 \cdot 10^{-6}$	$1,8201 \cdot 10^{-11}$
12	12	$2,6109 \cdot 10^{-7}$	$4,2788 \cdot 10^{-10}$
13	13	$9,9312 \cdot 10^{-9}$	$1,5854 \cdot 10^{-8}$
14	14	$3,2293 \cdot 10^{-10}$	$3,7692 \cdot 10^{-7}$
15	15	$9,0837 \cdot 10^{-12}$	$1,6731 \cdot 10^{-5}$
16	16	$2,2325 \cdot 10^{-13}$	$4,3169 \cdot 10^{-4}$
17	17	$4,8268 \cdot 10^{-15}$	0,0204
18	18	$1,5834 \cdot 10^{-16}$	2,6302

Таблица 3. Значения λ_n и ошибки δ (11) для $c = 20$, $N = 101$, $M = 27$

Номер функции (Matlab)	Скорректированный номер функции	Значение λ_n	Ошибка, δ
16	0	0,999999	$9,7149 \cdot 10^{-15}$
15	1	0,999998	$8,6180 \cdot 10^{-16}$
14	2	0,999996	$1,4720 \cdot 10^{-15}$
13	3	0,999992	$1,0531 \cdot 10^{-15}$
12	4	0,99999	$1,2603 \cdot 10^{-15}$
11	5	0,99997	$8,5340 \cdot 10^{-16}$
9	6	0,99993	$1,2798 \cdot 10^{-15}$
8	7	0,9999	$8,0873 \cdot 10^{-16}$
6	8	0,9998	$5,4342 \cdot 10^{-16}$
5	9	0,9981	$1,0142 \cdot 10^{-15}$
3	10	0,9840	$2,2465 \cdot 10^{-15}$
1	11	0,9019	$1,3170 \cdot 10^{-15}$
0	12	0,6365	$2,2044 \cdot 10^{-15}$
4	13	0,0621	$1,1603 \cdot 10^{-14}$
2	14	0,2667	$2,6452 \cdot 10^{-15}$
7	15	0,0095	$2,7672 \cdot 10^{-14}$
10	16	0,0011	$2,3090 \cdot 10^{-13}$
17	17	$1,0868 \cdot 10^{-4}$	$1,7534 \cdot 10^{-12}$
18	18	$9,1571 \cdot 10^{-6}$	$2,9481 \cdot 10^{-11}$
19	19	$6,7550 \cdot 10^{-7}$	$4,1404 \cdot 10^{-10}$
20	20	$4,4129 \cdot 10^{-8}$	$5,2743 \cdot 10^{-9}$
21	21	$2,5752 \cdot 10^{-9}$	$1,0669 \cdot 10^{-7}$
22	22	$1,3519 \cdot 10^{-10}$	$1,7211 \cdot 10^{-6}$
23	23	$6,4209 \cdot 10^{-12}$	$3,0271 \cdot 10^{-5}$
24	24	$2,7739 \cdot 10^{-13}$	$8,1059 \cdot 10^{-4}$
25	25	$1,0941 \cdot 10^{-14}$	0,0195
26	26	$4,9774 \cdot 10^{-16}$	0,4351
27	27	$3,9142 \cdot 10^{-16}$	0,9099

- для $c=10$: $n_0=6$, $n_c=16$,

- для $c=20$: $n_0=12$, $n_c=24$.

Индексы n_c , n_0 определяют границы применимости метода:

- для $n < n_0$ значения $\lambda_n \approx 1$, и величина ошибки $\delta << 1\%$ и имеет значение отличное от нуля в четырнадцатом порядке;

- для $n_0 \leq n \leq n_c$ значение ошибки δ приемлемо и порядок ее не более 1%, причем числа λ_n имеют значения отличные от нуля в порядке меньше или равном пятнадцати;

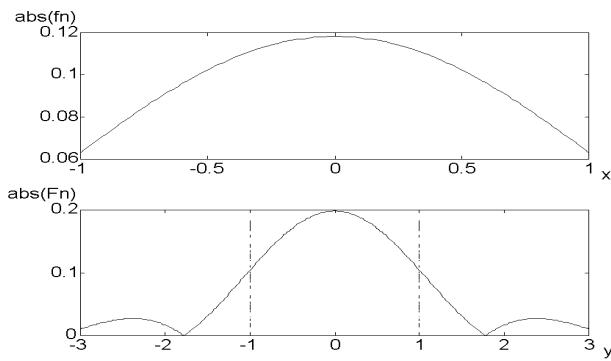
- для $n > n_c$ величина ошибки δ становится не приемлемой, а числа λ_n имеют значения отличные от нуля в порядке больше или равном шестнадцати;

Следует отметить, что в прикладных задачах наибольший интерес представляют случаи, когда $\lambda_n \approx 1$, когда рассматриваемый метод демонстрирует хорошие результаты.

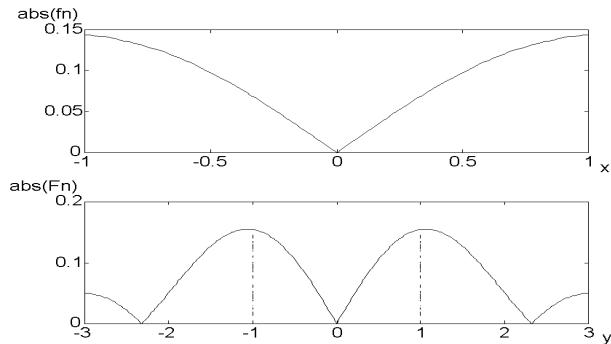
Проверка выполнения свойства инвариантности к преобразованию Фурье.

Выполнение этого свойства оценивалось по среднеквадратичному отклонению "выходной" $\hat{\psi}_n(y)$ функции после преобразования (4) от "входной" $\psi_n(x)$.

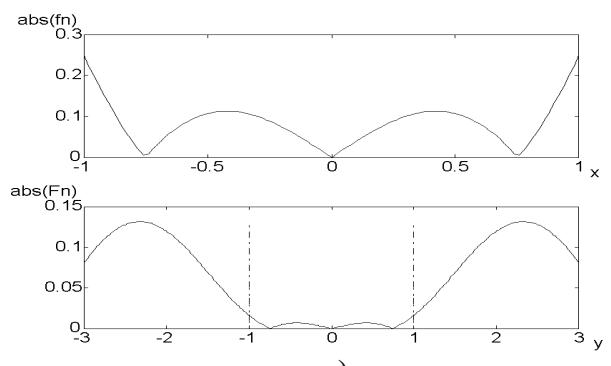
На рис.1, 2 приведены графики абсолютных значений собственных функций $\psi_n(x)$ и их Фурье-образов $\hat{\psi}_n(y)$ для различных параметров. Из рисунков 1 и 2 видно, что при



а)



б)



в)

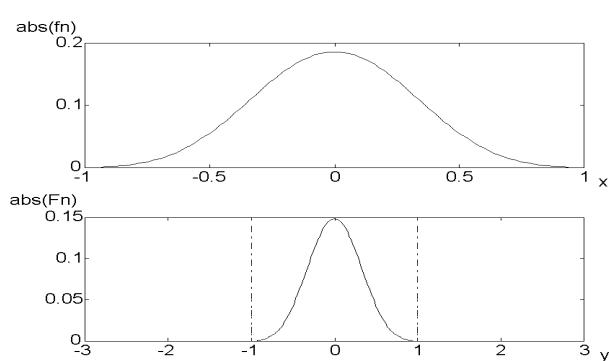
Рис. 1. Графики абсолютных значений собственных функций $\psi_n(x)$ ($abs(fn)$) и их Фурье-образов $\hat{\psi}_n(y)$ ($abs(Fn)$) для $N=101$, $c=2$, $n=0$ (а), $n=1$ (б), $n=3$ (в)

значениях $\lambda_n \approx 1$ энергия ВСФ сконцентрирована на интервале $[-T/2, T/2]$, и, наоборот, при значениях $\lambda_n \approx 0$ вне этого интервала. Однако, вид функции внутри интервала после преобразования Фурье не меняется с точностью до множителя.

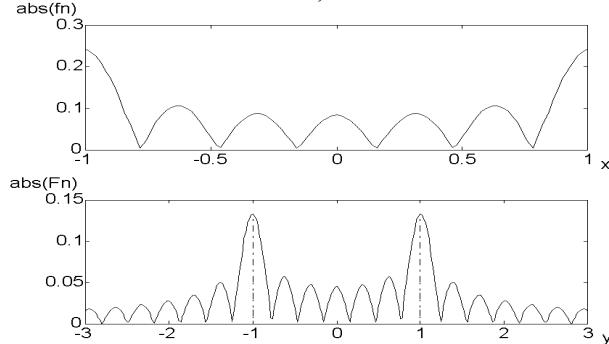
Из таблицы 4 видно, что ошибка δ становится не приемлемой только, если собственные значения λ_n намного меньше единицы, при значениях λ_n близких к единице значение ошибки стремится к нулю.

Недостатки метода

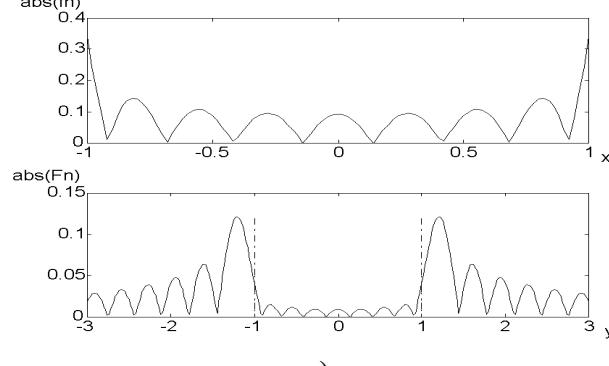
Выше было показано, что полученные собственные функции удовлетворяют основ-



а)



б)



в)

Рис. 2. Графики абсолютных значений собственных функций $\psi_n(x)$ ($abs(fn)$) и их Фурье-образов $\hat{\psi}_n(y)$ ($abs(Fn)$) для $N=101$, $c=10$, $n=6$ (а), $n=8$ (б), $n=8$ (в)

ным свойствам ВСФ при $\lambda_n \approx 1$.

Однако, из таблиц 2-4 видно несоответствие номеров собственных функций порядку собственных чисел. Правильно отранжировать ВСФ можно, используя (12) и (3).

При уменьшении λ_n наблюдается не только нарушение свойств ВСФ, но и искашение вида функций. Например, собственные функции на рис.3. Нужно при этом отметить фильтрующее свойство преобразования Фурье - после его применения функции приняли стандартный вид. Преобразование с sinc-ядром такого эффекта не дало. Однако "выглаживание" рассчитанных функций с помощью (4) снижает ошибку в выполнении

Таблица 4. Значения отклонения абсолютного значения ВСФ и их Фурье-образа для $M = 9, N = 101$

Номер функции (Matlab)	$c=2$		$c=10$	
	Собственные числа, λ_n	Ошибка, δ	Собственные числа, λ_n	Ошибка, δ
0	0,8929	0,0046	0,8394	0,0419
1	0,3678	0,0047	0,4719	0,0537
2	0,0395	0,0169	0,9796	0,0374
3	0,0014	0,0439	0,9984	0,0307
4	0,0000	0,0804	0,9988	0,0228
5	0,0000	0,1244	0,9996	0,0154
6	0,0000	0,1743	0,9999	0,0099
7	0,0000	0,2286	0,1422	0,0728
8	0,0000	0,2840	0,0225	0,1253
9	0,0000	0,2805	0,0024	0,1866

(2) только при $\lambda_n \approx 1$, в остальных случаях ошибка даже увеличивается.

Заключение

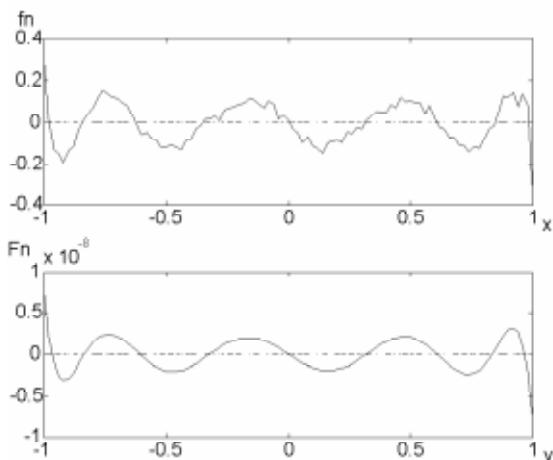
В данной работе реализован и исследован матричный метод вычисления ВСФ нулевого порядка с использованием программного обеспечения Matlab. Исследование метода заключается в проверке получаемых функций на выполнение основных свойств. Результатом данной проверки являются следующие выводы:

1. Данный метод позволяет получать удовлетворительную точность только в некотором диапазоне изменения индекса n , обеспечивающие $\lambda_n \approx 1$. Верхней границей диапазона можно считать $n_0 = 2c/\pi$.

2. Выполнение свойства инвариантности к интегральному преобразованию с sinc-ядром обеспечивается с более высокой точностью, чем свойство инвариантности к преобразованию Фурье. В первом случае отклонение составляет менее 1% даже для функций с $\lambda_n \approx 10^{-13}$. Во втором случае отклонение было меньше 5% только для $n < n_0$.

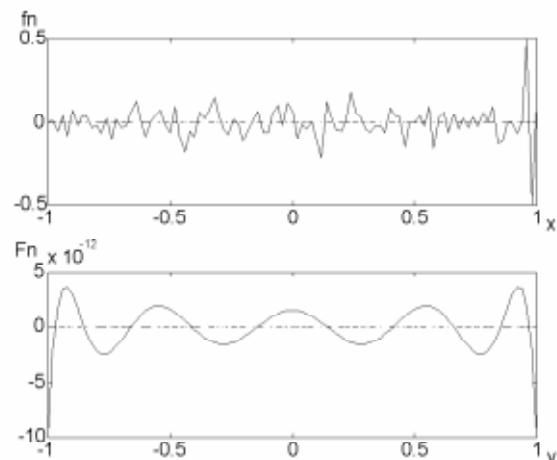
3. Недостатком метода также является неправильное упорядочивание вычисляемых функций, которую нужно корректировать в соответствии с убыванием λ_n .

4. При $\lambda_n \approx 0$ наблюдается не только нарушение свойств ВСФ, но и искажение вида функций. "Выгладить" такие функции можно с помощью преобразования Фурье, однако ошибка при этом снижается только при $\lambda_n \approx 1$.



a)

Рис. 3. Графики абсолютных значений собственных функций $\psi_n(x)$ (fn) и их Фурье-образов $\hat{\psi}_n(y)$ (Fn) для $N = 101$, $c = 2$, $n = 9$ (a), $n = 10$ (б)



б)

Тем не менее, учитывая простоту реализации, рассмотренный метод может успешно применяться в практических задачах, использующих ВСФ с собственными числами близкими к единице.

Работа поддержанна Российским фондом фундаментальных исследований (гранты 00-15-96114, 00-01-00031).

СПИСОК ЛИТЕРАТУРЫ

1. *Slepian D., Pollak H.O.* Prolate spheroidal wave functions. Fourier Analysis and Uncertainty – I // Bell Syst. Tech. J. V.40. 1961.
2. *Landau H.J., Pollak H.O.* Prolate spheroidal wave functions. Fourier Analysis and Uncertainty – II // Bell Syst. Tech. J. V.40. 1961.
3. *Slepian D., Pollak H.O.* Prolate spheroidal wave functions, Fourier analysis and uncertainty – III: The dimension of essentially time-and band-limited signals // Bell Syst. Tech. J. V.41. 1962.
4. *Хургин Я.И., Яковлев В.П.* Методы теории целых функций в радиофизике, теории связи и оптике. М.: Физматгиз, 1962.
5. *De Santis P., Palma C.* Degrees of freedom of aberrated images // Opt. Acta. V.23. 1976. №9.
6. *Bertero M., Viano G.A., De Mol C.* Resolution beyond the diffraction limit for regularized object restoration // Opt. Acta. V.27. 1980. №3.
7. *Bertero M., Pike E.R.* Resolution in diffraction-limited imaging, a singular value analysis. I. The case of coherent illumination // Opt. Acta. V.29. 1982. №6.
8. *Barbosa L.C.* A maximum-energy concentration spectral window // IBM J. Res. Develop. V.30. 1986. №3.
9. *Rogers M.E., Erkkila J.H.* Resonator mode analysis using linear prolate functions // Appl. Opt. V.22. 1983. №13.
10. *Latham W.P., Tilton M.L.* Calculation of prolate functions for optical analysis // Appl. Opt. V.26. 1987. №13.
11. *Фламмер К.* Таблицы волновых сфероидальных функций. М.: ВЦ АН СССР, 1962.
12. *Комаров И. В., Пономарев Л. И., Славинов С. Ю.* Сфероидальные и кулоновские сфероидальные функции. М.: Наука, 1976.
13. *Лоуэн А.* Сфероидальные волновые функции / Справочник по специальным функциям под ред. М. Абрамовица и И. Стиган. М.: Наука, 1979.
14. *Абрамов А.А., Дышко А.Л., Конюхова Н.Б., Пак Т.В., Парицкий Б.С.* Вычисление вытянутых сфероидальных функций решением соответствующих дифференциальных уравнений // Ж. вычисл. матем. и матем. физ. Т.24. 1984. № 1.
15. *Хонина С.Н., Комляр В.В.* Вытянутые сфероидальные функции в дифракционной оптике // Материалы международной молодежной школы по оптике, лазерной физике и биофизике. Саратов: Сарат. ун-т, 2000.
16. *Хонина С.Н.* Приближение сфероидальных волновых функций конечными рядами // Компьютерная оптика. Самара, 1999. Вып.19.
17. *Волотовский С.Г., Казанский Н.Л., Хонина С.Н.* Анализ и разработка методов вычисления собственных значений вытянутых сфероидальных функций нулевого порядка // Труды 5-ой международной конференции РОАИ-5-2000. Т.4. Самара, 2000.
18. *D. Slepian*, Some asymptotic expansions for prolate spheroidal wave functions // J. Math. & Phys. V.44. 1965.

STUDIES INTO A MATRIX TECHNIQUE FOR CALCULATING PROLATE SPHEROIDAL FUNCTIONS OF ZERO ORDER

© 2001 S.N. Khonina

Image Processing Systems Institute of Russian Academy of Science, Samara

A matrix technique for calculating prolate spheroidal functions of zero order that is implemented using the Matlab tools is studied. The calculational accuracy is assessed on the basis of fulfillment of the major properties of the prolate spheroidal functions: invariance to the integral transform with sinc-kernel and to the Fourier transform. An algorithm for properly ordering the functions using their eigenvalues is proposed. Applicability limits of the proposed technique implementation are determined.

ИТЕРАТИВНОЕ ОЦЕНИВАНИЕ С ИСПОЛЬЗОВАНИЕМ ИНФОРМАЦИИ ОБ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЯХ НЕВЯЗОК

© 2001 Э.И. Коломиец¹, А.В. Устинов², В.А. Фурсов²

¹ Самарский государственный аэрокосмический университет

² Институт систем обработки изображений РАН, г. Самара

Рассматриваются итерационные алгоритмы, в которых для уточнения оценок используется дополнительная информация, содержащаяся в невязках. Получены сравнительные характеристики эффективности некоторых критериев для принятия решений о классах распределений по реализациям невязок и рассмотрены основанные на них процедуры корректировки данных.

Введение

При решении задач оценивания по малому числу наблюдений распределения ошибок на отдельных реализациях не обладают статистической устойчивостью, поэтому использование априорных вероятностных моделей помех не вполне правомерно. Если о распределении ошибок ничего не известно, то, опираясь на восходящее к Гауссу мнение, применяют метод наименьших квадратов (МНК). Однако известно, что если распределение ошибок на конкретных реализациях существенно отличается от гауссовского, МНК-оценки, оказываются неработоспособными [1].

Для преодоления указанного недостатка строят так называемые робастные оценки в расчете на "наихудшее" (например, в смысле фишеровской информации) распределение из некоторого класса [2]. Такой подход обеспечивает повышение надежности оценок, однако возможна существенная потеря точности по сравнению с классическими МНК-оценками в ситуациях, когда распределение ошибок на конкретной реализации в действительности оказывается близким к гауссовскому.

Более адекватным существу задачи в данном случае представляется адаптивный подход, заключающийся в подборе наиболее подходящей модели ошибок для каждой реализации в ходе оценивания. Центральной проблемой при этом является получение необходимой для реализации процедуры перестройки дополнительной информации. Такая информация может быть получена, напри-

мер, путем дополнительных измерений [3]. В настоящей работе исследуется возможность получения дополнительной информации путем изучения распределений реализаций вектора невязок, получаемых на промежуточных шагах так называемого итерационного МНК [4].

Постановка задачи

Рассматривается задача оценивания $M \times 1$ -вектора \mathbf{c} параметров модели типа линейной регрессии:

$$\mathbf{y} = \mathbf{X}\mathbf{c} + \mathbf{o}, \quad (1)$$

где $N \times M$ -матрица $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ и $N \times 1$ -вектор \mathbf{y} составлены из доступных для наблюдения измерений, а \mathbf{o} - неизвестный $N \times 1$ -вектор ошибок. Предполагается, что N мало (превышает M не более чем в 2-3 раза).

Для получения оценок реализуется схема итерационного МНК в виде следующей последовательности шагов. Вычисляются начальная МНК-оценка

$$\hat{\mathbf{c}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} \quad (2)$$

вектора параметров \mathbf{c} модели (1) и вектор невязок

$$\hat{\mathbf{o}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}}. \quad (3)$$

Далее по вектору невязок строится так называемая весовая матрица

$G = G(\hat{\mathbf{c}}) = \text{diag}(g_1, g_2, \dots, g_N)$,
осуществляется преобразование исходных данных:

$$\tilde{\mathbf{X}} = \mathbf{G}\mathbf{X}, \quad \tilde{\mathbf{y}} = \mathbf{G}\mathbf{y}, \quad (4)$$

и вновь вычисляется МНК-оценка (2) по \tilde{x} и \tilde{y} .

Если в рамках этой схемы весовые коэффициенты g_i , $i=1, N$ на каждом шаге процедуры (2)-(4) вычисляются по правилу

$g_i = |\xi_i|^{v-2}$, $0 < v < 2$, она совпадает с итерационным МНК, предложенным в работе [5] для вычисления L_v - оценок. К этой же схеме можно свести вычисление оценок максимального правдоподобия параметров сдвига c и масштаба s плотности $p(\xi(c), s)$.

Действительно, пусть функция потерь

$$F(\xi(c), s) = -\ln p(\xi(c), s) \quad (5)$$

удовлетворяет условиям:

$$\left. \begin{array}{l} F_\xi'(\xi(c), s)\xi^{-1} > 0 \\ \lim_{\xi \rightarrow \infty} g(\xi, s) = 0, \\ \lim_{\xi \rightarrow 0} g(\xi, s) < \infty \end{array} \right\}, \quad (6)$$

где

$$g(\xi, s) = (F_\xi'(\xi(c), s)\xi^{-1})^{\frac{1}{2}}. \quad (7)$$

Тогда искомые оценки \hat{c} , \hat{s} :

$$Q(\xi(\hat{c}), \hat{s}) = \min_{c, s} \sum_{i=1}^N F(\xi_i(c), s)$$

являются решением системы ($M+1$) уравнений вида

$$\begin{aligned} & \left[\sum_{i=1}^N g^2(\xi_i(\hat{c}), \hat{s}) x_i x_i^T \right] \hat{c} - \\ & - \left[\sum_{i=1}^N g^2(\xi_i(\hat{c}), \hat{s}) x_i y_i \right] = 0, \end{aligned} \quad (8)$$

$$\sum_{i=1}^N F_s'(\xi_i(\hat{c}), \hat{s}) = 0. \quad (9)$$

Попеременное решение уравнений (8) и (9), по существу, является схемой последовательных МНК-оценок (2)-(4). В частности, решение уравнения (8) доставляет взвешенные МНК-оценки, а элементы весовой матрицы могут быть вычислены по соотношению (7) при $s = \hat{s}$, которое является решением нелинейного уравнения (9). Формальное совпадение вычислительных схем послу-

жило поводом для вывода о том, что оценивание методом взвешивания (2)-(4) на одном малом наборе данных согласуется со статистической теорией оценивания [3, стр.244]. Необходимо подчеркнуть, что такое сходство носит исключительно внешний характер.

Таким образом, весовая матрица преобразования G может быть построена, если задана априорная гипотеза о виде плотности распределения вероятности вектора ошибок. Поскольку априорное распределение вектора ошибок никогда не известно, поставим следующую задачу.

Зададим совокупность гипотез в виде n функций распределения

$$\{F_1(\xi, s), F_2(\xi, s), \dots, F_n(\xi, s)\}$$

или плотностей

$$\{p_1(\xi, s), p_2(\xi, s), \dots, p_n(\xi, s)\}, \quad (10)$$

где ξ - вектор невязок (3), а s - неизвестный параметр (масштаба). Будем на каждой очередной итерации алгоритма решать задачу выбора наиболее правдоподобной гипотезы относительно эмпирического распределения вектора невязок. Далее с использованием выбранной функции $p(\hat{\xi}(c), s)$ по указанным выше соотношениям вычисляются весовые коэффициенты $g(\xi, s)$ и строится оценка вектора параметров на следующей итерации.

В рамках этой постановки задачи возникают, по крайней мере, две проблемы. Во-первых, в какой мере в рамках указанной выше вычислительной схемы правомерно использовать гипотезу $p(\hat{\xi}(c), s)$ вместо неизвестной плотности $p(\xi(c), s)$ для вычисления весовых коэффициентов. В качестве обоснования такой "подмены" можно привести следующие соображения. Можно показать, что $\xi = T_0 T_0^T \xi$, где T_0 - $N \times (N-M)$ - матрица, составленная из нормированных собственных векторов, соответствующих нулевым собственным значениям матрицы XX^T . Из этого равенства видно, что если вектор ошибок принадлежит нуль-пространству матрицы X^T , он совпадает с вектором невязок. Есть осно-

вания полагать, что в некоторой окрестности нуль-пространства всегда существует множество векторов ошибок достаточно близких к соответствующим векторам невязок. Для этого должны выполняться обычные требования слабой сопряженности векторов ошибок и полезных сигналов (столбцов матрицы \mathbf{X}). Признаком принадлежности вектора ошибок окрестности нуль-пространства может служить сравнительное большое (при соответствующей нормировке) значение нормы вектора невязок.

Вторая проблема связана с применением в данном случае классических критериев проверки гипотез при малом числе наблюдений. Деликатность ситуации состоит в том, что даже в случае, когда ошибки независимы и одинаково распределены, невязки являются зависимыми и неодинаково распределенными. Поэтому на вектор наблюдений (невязок) следует смотреть как на реализацию некоторого случайного процесса, а не как на выборку в классическом смысле.

В настоящей работе предлагаются некоторые эмпирические критерии принятия решений о наиболее правдоподобном семействе функций распределения (плотностей) для вычисления весовых коэффициентов с использованием компонент вектора невязок. При этом не принимается во внимание в полной мере существование отмеченной выше второй проблемы и предлагаемые критерии во многом аналогичны классическим. Учитывая этот факт, проводится сравнительный анализ критериев и экспериментально показывается, в какой степени информация об эмпирическом распределении невязок является полезной и может быть использована для повышения точности оценок.

Критерии выбора наиболее правдоподобной гипотезы

В качестве набора гипотез (10) рассмотрим симметричные относительно оси ординат параметрические семейства функций, удовлетворяющих условиям (5)-(7) и зависящих (при $\xi_0=0$) лишь от одного параметра масштаба s .

Среднеквадратичный критерий.

В качестве меры для различия гипо-

тез о принадлежности эмпирической функции распределения $F_{\text{эмп}}(\xi)$ к одному из заданных однопараметрических семейств $F_j(\xi, s)$ будем рассматривать критерий вида:

$$J_j(s) = \int_{-\infty}^{+\infty} [F_j(\xi, s) - F_{\text{эмп}}(\xi)]^2 d\xi. \quad (11)$$

Наиболее правдоподобной считается та гипотеза, для которой величина критерия (11) является наибольшей.

Критерий нормированного квантиля

Для набора априорных вероятностных моделей (гипотез) (10) наряду с (11) будем рассматривать критерии различия, основанные на непосредственном вычислении характеристик эмпирического распределения.

Зададимся некоторой вероятностью p_0 . Квантиль, соответствующий данной вероятности, зависит от вида функции распределения и пропорционален параметру масштаба s или с.к.о. σ (если оно существует). Поэтому произвести выбор можно по величине $K_{q,j} = F_j^{-1}(p_0)/\sigma$. В данном случае по выборке вычисляется эмпирическое значение этой величины и сравнивается с теоретическими значениями для заданных распределений. Выбирается то распределение, у которого $K_{q,j}$ наиболее близко к эмпирическому.

Критерий нормированного среднего модулей невязок.

Смысл критерия поясним на рисунке 1, где приведена непрерывная функция распределения $F(\xi)$. В данном случае в качестве признака принадлежности эмпирического распределения к одному из заданных априорных семейств используются величины площадей, в областях слева и справа относительно центра. На рисунке, эти области обозначены S и заштрихованы.

Можно показать, что суммарная (слева и справа от ординаты) площадь равна математическому ожиданию модуля величины ξ . Этот факт очень важен при практическом использовании, так как позволяет избежать построения вариационного ряда и эмпирической функции распределения. Данная площадь имеет ту же размерность, что и сама

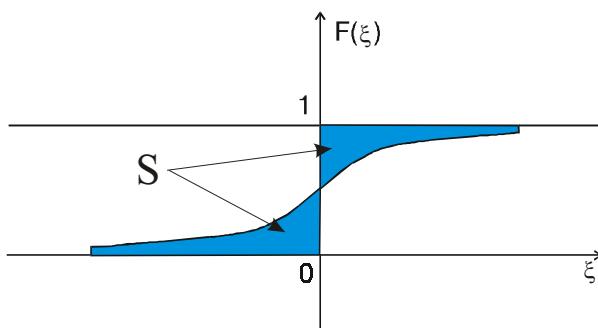


Рис. Непрерывная функция распределения $F(\xi)$

случайная величина. Сравнение распределений производится по безразмерным величинам, также как в случае критерия нормированного квантиля.

Будем рассматривать два способа нормировки среднего модулей невязок:

a) *Нормировка на среднеквадратичное отклонение* В этом способе величиной для сравнения распределений является S/σ .

b) *Нормировка на площадь равномерного распределения.*

В этом способе уже при теоретическом рассмотрении непрерывного распределения используются не бесконечные, а конечные пределы интегрирования - минимальное и максимальное значения невязок $\hat{\xi}_{\min}$ и $\hat{\xi}_{\max}$. Нормировка производится на площадь, заключенную под функцией распределения равномерной плотности в тех же пределах. Сравнение распределений производится так же, как и в предыдущем способе. Далее приводятся результаты сравнительного исследования эффективности описанных критериев.

Проверка гипотез и процедура оценивания с использованием среднеквадратического критерия

В таблице 1 приведены некоторые из известных симметричных относительно центра параметрических семейств распределений, удовлетворяющих условиям (5)-(7) [6]. Для упрощения рассматриваются функции, зависящие лишь от одного параметра масштаба - s . Во втором столбце приведены выражения для функций функции $p(\hat{\xi}(\mathbf{c}), s)$, в третьем - соответствующие им функции потерь (5), а в четвертом - выражения для квадратов весовых коэффициентов (7).

Прежде всего, проиллюстрируем принципиальную работоспособность среднеквадратического критерия (11), который в дискретном случае можно переписать в виде

$$\chi^2_j(s) = \left\{ \sum_{i=1}^{N-1} [F_j(s, \hat{\xi}_i) - p_i]^2 * (\hat{\xi}_{i+1} - \hat{\xi}_i) \right\}^{1/2} \quad (12)$$

Здесь $\hat{\xi}_i$ - отсчеты вариационного ряда, построенного из компонент вектора невязок $\xi, \hat{\xi}_i \in [\xi_i, \xi_{i+1}]$, p_i - отсчеты эмпирической функции распределения $F_j(\hat{\xi}, s) \mid_{j=1, N}$.

Заметим, что "весовые" множители $\hat{\xi}_{i+1} - \hat{\xi}_i$ по смыслу близки к множителям n / p_i при вычислении статистики χ^2 (если N - число разрядов, а n - объем выборки). При малом числе наблюдений использование свойств χ^2 -распределения, конечно, необоснованно. Более того, теряет смысл и раз-

Таблица 1. Параметрические семейства распределений

	Функция $p(\hat{\xi}(\mathbf{c}), s)$	Функция потерь	Функция $g(\hat{\xi}, s)$
1	$\frac{1}{s\sqrt{2\pi}} e^{-\frac{\hat{\xi}^2}{2s^2}}$	$\ln(\sqrt{2\pi}s) + (-\hat{\xi}^2/2s^2)$	$\frac{1}{s^2}$
2	$\frac{1}{4s} \left(1 + \frac{ \hat{\xi} }{s}\right) e^{-\frac{ \hat{\xi} }{s}}$	$\ln 4s + \hat{\xi} /s - \ln(1 + \hat{\xi} /s)$	$\frac{1}{s(s + \hat{\xi})}$
3	$\frac{s}{\pi} * \frac{1}{s^2 + \hat{\xi}^2}$	$\ln(s^2 + \hat{\xi}^2) + \ln(\pi/s)$	$\frac{2}{s^2 + \hat{\xi}^2}$

бение выборки на разряды, поэтому мы полагаем $N=p$. Однако указанная аналогия может быть полезна при необходимости верификации результатов при больших N .

Для проверки эффективности критерия (12) проводился эксперимент, заключавшийся в генерации малых выборок с заданными в таблице 1 законами распределения и последующей их аппроксимацией набором тех же параметрических семейств функций. При построении набора данных в качестве случайного значения $\hat{\xi}_i$ принимался квантиль, соответствующий вероятности

$$P_i = \frac{i - 0.5}{N}, i = \overline{1, N}.$$

При этом способе построения выборки длина отрезка интегрирования $[a; b]$ растёт с ростом N как $\sim \sqrt{\ln N}$ при распределении 1, $\sim \ln N$ при распределении 2 и $\sim N$ при распределении 3. Поэтому оценка сверху погрешности традиционных квадратурных формул может не выполняться. Тем не менее, отношение $(b-a)/N$ для распределений 1 и 2 стремится к нулю, поэтому выражение под корнем в правой части (12) является интегральной суммой для интеграла (11). При этом вместо кажущегося естественным выбора $\hat{\eta}_i$ по формуле центральных прямоугольников: $\hat{\eta}_i = (\hat{\xi}_i + \hat{\xi}_{i+1})/2$, величины $\hat{\eta}_i \in [\xi_i, \xi_{i+1}]$ определялись равными квантилям соответствующих вероятностей:

$$P_{1i} = \frac{i}{N}, i = \overline{1, N-1}.$$

Это позволяет исключить методическую ошибку аппроксимации и, в частности, обеспечивает равенство нулю критерия (12) в случае, когда генерируемое и аппроксимирующее семейства совпадают.

Аппроксимация построенных для каждого набора данных эмпирических функций распределения осуществлялась по принципу "каждая" "каждой". Для оценки параметра масштаба s методом диахотомии решалось уравнение $dJ(\hat{s})/ds = 0$. Результаты исследования для $N=25$ приведены в таблице 2. Целые числа 1-3 в обозначениях строк и столбцов соответствуют номерам исходных распределений в таблице 1. При дальнейшем увеличении N они практически не изменяются.

Результаты моделирования подтверждают предположение о возможности непосредственного использования критерия (12) для оценки априорных гипотез при небольших N ($N \leq 25$). Значение критерия при аппроксимации множества компонент вектора невязок "своим" параметрическим семейством действительно минимально и отличается от нуля, по-видимому, лишь из-за погрешностей округления.

Исследовалась также частота совпадений параметрических семейств распределений компонент векторов ошибок и невязок. Эксперимент проводился по следующей схеме. Определялось оценивание вектора \mathbf{c} параметров уравнения (1) по матрице \mathbf{X} и вектору \mathbf{y} в присутствии вектора ошибок $\boldsymbol{\xi}$. Размерности векторов \mathbf{y} и $\boldsymbol{\xi}$ 25×1 . Генерировалось по 100 реализаций вектора выхода модели и шума указанной длины для каждого заданного параметрического семейства распределений. После получения оценки $\hat{\mathbf{c}}$ для каждой реализации вычислялся вектор невязок (10) и строилась эмпирическая функция распределения его компонент. Затем она аппроксимировалась каждым из указанных выше трех (1, 2 и 3) параметрических семейств распределений и подсчитывалось значение критерия (12). При моделировании

Таблица 2. Результаты исследования эффективности критерия (12) для $N=25$

Вид	1		2		3	
распределения	s	$J_{\hat{s}}(s)$	s	$J_{\hat{s}}(s)$	s	$J_{\hat{s}}(s)$
1 (s=1,000)	1,000	0,00011	0,554	0,02213	0,543	0,07834
2 (s=0,546)	1,071	0,03682	0,546	0,00020	0,510	0,06869
3 (s=0,480)	1,665	0,21448	0,578	0,14051	0,480	0,00005

Таблица 3. Относительное число совпадений распределений ошибок и невязок

Распределение	Априорная гипотеза	Относительное число решений K_{cp} (в %)	Средняя норма вектора невязок
1	1	54,3	4,439
	2	43,4	4,405
	3	2,3	3,578
2	1	28,2	4,614
	2	66,4	4,753
	3	5,4	4,555
3	1	0,4	2,236
	2	26,1	9,843
	3	73,5	10,747

векторов ошибок искусственно вводилась рандомизация, а результаты усреднялись по 100 экспериментам. В таблице 3 приведены данные об относительном числе случаев K_{cp} , в которых каждое распределение оказывалось наилучшей, в смысле минимума критерия (12), аппроксимацией.

Приведенные данные хорошо согласуются с высказанными ранее соображениями качественного характера о том, что значение нормы вектора невязок свидетельствует о близости его распределения к распределению вектора ошибок. Приведем теперь результаты оценивания вектора параметров.

Эксперимент проводился для той же

модели и при тех же размерностях векторов, как и выше. Особенность эксперимента заключалась в следующем. Параметрические классы распределений, задаваемые на этапе моделирования, считались известными, а оценивание осуществлялось по трехшаговой схеме с использованием весовых коэффициентов, соответствующих классам 1-3 из таблицы 1. Для анализа вектора невязок и принятия решения о типе преобразования взвешивания использовался среднеквадратический критерий. Вычислялись средние (по 100 реализациям) погрешности оценок вектора параметров \hat{c} , на первом - ε_1 и третьем - ε_3 шагах процедуры оценивания в процентах. Результаты приведены в таблице 4.

Указанный в таблице выбор значений параметра s связан с тем, что при аппроксимации функции распределения (см. таблицу 2) оптимальное значение s для распределений 2 и 3 примерно вдвое меньше, чем для распределения 1 (при аппроксимации по методу моментов для распределения 2 значение s будет точно в два раза меньше). Приведенные результаты подтверждают принципиальную возможность существенного повышения точности при использовании стратегии МНК-оценивания при малом числе наблюдений.

Описанный метод проверки гипотез по среднеквадратическому критерию обладает существенным недостатком - очень большим

Таблица 4. Погрешности на первой и третьей итерации трёхшаговой схемы оценивания

Исходное распределение 1											
s	0,01	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
ε_1	0,2374	2,458	4,646	7,736	9,131	12,88	14,66	19,75	22,52	21,54	21,57
ε_3	0,2374	2,458	4,646	7,736	9,131	12,88	14,66	19,75	22,52	21,54	21,57
Исходное распределение 2											
s	0,01	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
ε_1	0,4982	2,509	4,443	8,065	9,717	14,39	13,64	15,48	18,58	27,58	26,22
ε_3	0,5111	2,568	4,194	7,707	9,645	14,58	13,30	15,05	18,34	25,78	26,05
Исходное распределение 3											
s	0,01	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
ε_1	1,395	5,480	10,94	21,62	22,62	29,21	38,57	45,82	50,85	56,12	68,07
ε_3	0,723	2,987	6,73	11,04	12,49	15,48	21,74	23,83	27,68	27,44	28,85

объёмом вычислений. Эксперименты показывают, что на определение вида распределения затрачивается примерно в 4,4 раза больше времени, чем на все остальные вычисления. Далее приводятся результаты исследования других, предложенных в разделе 2 критериев, основанных на непосредственном вычислении характеристик эмпирического распределения.

Проверка гипотез по критерию нормированного квантиля

Прежде чем вести речь о других критериях проверки гипотез и излагать методику сравнения критериев, следует заметить, что ошибки при проверке гипотез неравнозначны. Ошибки принятия гипотезы 2, когда в действительности имело место распределение типа 1 (и наоборот) менее существенны, чем принятие гипотезы 3, когда было распределение типа 1 и уж тем более принятие гипотезы 1, в ситуации, когда имело место распределение 3. Это, вообще говоря, ясно из интуитивных соображений и подтверждается данными таблицы 3.

С учетом сказанного, целесообразно установить ограничение на число ошибок, заключающихся в принятии гипотезы 3, в то время как имело место распределение 1 и наоборот - гипотезы 1 при распределении 3. В частности, потребуем, чтобы относительное число таких ошибок составляло не более 1-2%. При выполнении этого условия будем сравнивать критерии по суммарному числу правильно принятых решений на векторах невязок, полученных моделированием трёх указанных выше распределений (1, 2 и 3) одинаковое число раз.

Еще одно замечание. Для всех исследуемых далее эмпирических критериев, фигурирующий в них параметр масштаба s находится без решения уравнений. Для этого используются соотношения, полученные при проведении экспериментов по схеме п.3 для всех распределений:

$$\text{№1: } s = \sigma; \text{ №2: } s = 0,546 \cdot \sigma \text{ и } \text{№3: } s = 0,48 \cdot \sigma. \quad (13)$$

Теперь конкретизируем процедуру выбора вида распределения по эмпирическому значению величины нормированного кван-

тиля $K_{q,j} = F_j^{-1}(p_0)/\sigma$. Исходя из вида различаемых нами трёх распределений, лучше всего было бы взять p_0 малым, чтобы квантиль был за пределами $2 \div 2,5\sigma$, где различие трёх квантилей достаточно велико, однако при малой выборке это нереализуемо. Подсчёты по непрерывным плотностям показали, что можно взять $p_0 = 1/3$. Алгоритм работает следующим образом:

1) Вычисляется выборочное с.к.о. $\hat{\sigma}$ и среднее \bar{x} .

2) Значения выборки упорядочиваются по возрастанию и центрируются.

3) Вычисляется значение квантиля $F^{-1}(\frac{1}{3})$ (оно отрицательно). В силу симметрии проверяемых распределений для повышения точности в качестве результирующего квантиля берётся

$$Q_{v,j} = \left(F_j^{-1}\left(\frac{1}{3}\right) + F_j^{-1}\left(\frac{2}{3}\right) \right) / 2.$$

Кроме того, так как при малом N округление от деления на 3 существенно, при вычислениях $F^{-1}(\frac{1}{3})$ и $F^{-1}(\frac{2}{3})$ делается коррекция в зависимости от остатка от деления N на 3.

4) Величина $Q_v/\hat{\sigma}$ сравнивается с теоретическими значениями для каждого из распределений. Они равны соответственно 0,43; 0,388; 0,277.

5) Значение параметра s вычисляется по формуле (13) с учётом уже найденного вида распределения, вместо σ подставляется $\hat{\sigma}$.

Проверка гипотез по критерию нормированного среднего модулей невязок

Рассмотрим особенности применения этого критерия для двух указанных выше способов нормировки. Вначале рассмотрим случай нормировки на среднеквадратичное отклонение. В этом способе величиной сравнения является S_α/σ . Приведём выражения для S_α и σ при трёх используемых распределений.

$$\text{Распределение 1. - } S_\alpha = \sqrt{(2/\pi)}s; \sigma = s$$

Распределение 2. - $S_\alpha = 1,5 \cdot s; \sigma = 2s$

Распределение 3.

В этом случае оба интеграла расходятся, поэтому ограничим интегрирование отрезком $[-A; A]$.

$$\tilde{S}_\alpha = \frac{s}{\pi} \ln \left(1 + \frac{A^2}{s^2} \right); \tilde{\sigma}^2 = \frac{2s}{\pi} \left(2A - s \cdot \arctg \frac{A}{s} \right)$$

В асимптотическом варианте, когда A велико:

$$\tilde{S}_\alpha \cong \frac{2s}{\pi} \ln \frac{A}{s}; \tilde{\sigma} \cong \frac{2}{\sqrt{\pi}} \sqrt{As}$$

Нормированные величины будут следующими: $\sqrt{\frac{2}{\pi}}, \frac{3}{4}, \sqrt{\frac{1}{\pi} \ln \frac{A}{s}} \cdot \sqrt{\frac{s}{A}}$.

Опишем алгоритм:

- 1) Вычисляется выборочное с.к.о. $\hat{\sigma}$ и среднее $\bar{\xi}$.
- 2) Значения выборки центрируются.
- 3) Находится среднее значение

$$|\hat{\xi}| = |\bar{\xi} - \bar{\xi}|$$

4) Величина $|\hat{\xi}|/\hat{\sigma}$ сравнивается с теоретическими значениями. При сравнении с распределением №3 берётся

$$A = (\max \hat{\xi} - \min \hat{\xi}) / 2; s = 0,48 \cdot \hat{\sigma}.$$

5) Значение параметра s вычисляется по соотношениям (13).

Экспериментальные исследования показывают, что лучшие результаты, обычно, получаются при отличающихся от теоретических значениях нормированных величин:

$\sqrt{2/\pi} + 0,036; 3/4 + 0,02; 0,675$. Для гипотезы 3 (распределение Коши) это значение существенно отличается от исходного теоретического, по-видимому, из-за того, что при малом объёме выборки N асимптотическое приближение неправомерно.

Теперь рассмотрим случай нормировки на площадь равномерного распределения. Вначале приведём выражения для площадей при различных распределениях. Отрезок интегрирования $[-A; B]$.

Распределение 1.

$$S_1 = \frac{s}{\sqrt{2\pi}} \left(2 - \exp \left(-\frac{A^2}{2s^2} \right) - \exp \left(-\frac{B^2}{2s^2} \right) \right). \quad (14)$$

Распределение 2.

$$S_2 = \frac{3s}{2} - \exp \left(-\frac{A}{s} \right) * \left[\frac{3}{8} (A + s) + \frac{A^2}{4s} \right] - \exp \left(-\frac{B}{s} \right) * \left[\frac{3}{8} (B + s) + \frac{B^2}{4s} \right]. \quad (15)$$

Распределение 3.

$$S_3 = \frac{s}{2\pi} \left(\ln \left(1 + \frac{A^2}{s^2} \right) + \ln \left(1 + \frac{B^2}{s^2} \right) \right). \quad (16)$$

Равномерное распределение.

$$S_{\text{равн}} = \frac{A^2 + B^2}{2(A + B)}. \quad (17)$$

Алгоритм реализуется в виде следующих шагов:

- 1) Вычисляется выборочное с.к.о. $\hat{\sigma}$ и среднее $\bar{\xi}$.

2) Значения выборки центрируются.

3) Находится среднее значение $|\hat{\xi}|$, $A = -\min \hat{\xi}$, $B = \max \hat{\xi}$. Площадь, соответствующая равномерному распределению, считается по формуле (17).

4) Величина $|\hat{\xi}| / S_{\text{равн}}$ сравнивается с теоретическими значениями.

5) Значение s вычисляется по формуле (13).

В отличие от нормировки на с.к.о. здесь нельзя заранее вычислить теоретические значения нормированных величин. Из выражений (14)-(17) видно, что пороговые значения зависят не только от выбранного параметрического класса распределений, но и от интервала $[-A; B]$ ошибок. В свою очередь, этот интервал, для каждого заданного параметрического семейства, зависит от объема данных.

В качестве начального приближения для порогового значения брались значения, получаемые по формулам (14)-(17). A для вычисления границ интервала A и B использо-

Таблица 5. Проверка качества различия видов распределений

Тактическое распределение	Количество решений о гипотезе №:			Сумма правильных решений
	1	2	3	
Квадратичный критерий				
1	54,3	43,4	2,3	155,33
2	28,3	66,3	5,4	
3	0,4	26,1	73,5	
Критерий нормированного квантиля				
1	56,0	32,4	11,6	143,59
2	38,0	39,0	23,0	
3	2,3	13,1	84,6	
Критерий площади, отнесённой к с,к,о,				
1	54,7	44,4	0,9	171,05
2	28,8	67,1	4,1	
3	0,4	7,6	92,0	
Критерий площади, отнесённой к площади при равномерном распределении				
1	61,45	38,52	0,03	170,53
2	29,34	70,58	0,08	
3	0,23	18,63	81,14	

валась следующая эмпирическая формула

$$A=B=k(N)*s. \quad (18)$$

Здесь $k(N)$ - зависит от вида распределения и объёма выборки N так, что $-A$ является квантилем $1/2N$, а B -квантилем $1-(1/2N)$. Наилучшие результаты при $N = 25$ получились при нормированных величинах 0,896; 0,645; 0,355 соответственно для распределений 1, 2, 3.

Сравнение критериев

Проводилась сравнительная экспериментальная оценка критериев. В ходе эксперимента на каждой из 100 реализаций \mathbf{X} , у осуществлялись следующие этапы:

1. Моделирование реализации вектора ошибок для одного из заданных типов распределений (№1, 2 или 3). Параметры масштаба для них задавались на основе результатов исследований, приведенных в таблице 2. В частности, для распределения №1, $s = \hat{\sigma}$, для №2, $s = 0,546 \cdot \hat{\sigma}$ и для №3, $s = 0,48 \cdot \hat{\sigma}$, где $\hat{\sigma}$ - выборочное СКО.

2. Решение задачи оценивания и вычис-

ление соответствующей реализации вектора невязок.

3. Определение (по невязкам) текущего значения критерия (одного из трёх) и выбор типа распределения в соответствии с заданным пороговым значением при моделировании трёх указанных выше распределений (1, 2 и 3) одинаковое число раз. Для сравнения критериев осуществлялась проверка качества различия видов распределений на тех же наборах данных. Для этого на исходные данные накладывался шум с распределением заданного класса (1, 2 или 3) и подсчитывалось среднее по 100 экспериментам относительное число правильных и ошибочных решений для каждого распределения. Результаты приведены в таблице 5.

Заключение

Приведенные результаты экспериментов показывают, что параметрические семейства распределений компонент векторов ошибок и невязок в ситуациях, когда принимается гипотеза 3, соответствующая плотности распределения Коши, в большом числе случаев совпадают. По-видимому, это является следствием того, что для этого распределения характерным является больший вес числа выделяющихся ошибок, для которых такая связь проявляется более устойчиво.

С точки зрения надежности принимаемых решений наиболее эффективным является критерий нормированного среднего модулей невязок. Эффективность различия гипотез будет существенно больше, если вместо двух плохо различимых гипотез 1 и 2 использовать лишь одну (любую из них).

Использование описанных процедур проверки гипотез о параметрическом распределении позволяет повысить надежность и точность оценивания на каждой реализации. Их использование полезно на этапе адаптации алгоритмов, когда априорная информация о распределении ошибок в исходных данных отсутствует.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (РФФИ, грант № N 99-01-00079).

СПИСОК ЛИТЕРАТУРЫ

1. Калман Р.Е. Идентификация систем с шумами // Успехи математических наук. Т.40. 1985. Вып.4(244).
2. Цыпкин Я.З. Основы информационной теории идентификации. М.: Наука, 1983.
3. Теряев Е.Д., Шамриков Б.М. Цифровые системы и поэтапное адаптивное управление. М.: Наука, 1999.
4. Демиденко Е.З. Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981.
5. Fletcher R., Grant J.A., Heblen H.D. The calculation of linear least Lp-approximations // Computer Journal. V.14. 1971. № 3.
6. Фурсов В.А. Идентификация моделей систем формирования изображений по малому числу наблюдений. Самара: СГАУ, 1998.

**ITERATED ESTIMATION WITH USAGE OF THE INFORMATION
ABOUT EMPIRICAL DISTRIBUTIONS OF DISCREPANCIES**

© 2001 E.I. Kolomiec¹, A.V. Ustinov², V.A. Fursov²

¹Samara State Aerospace University

²Image Processing Systems Institute of Russian Academy of Sciences

The iterative algorithms are considered, in which for refinement of estimations the additional information contained in discrepancies uses. The comparative characteristics of efficiency of some criteria for decision making about classes of distributions on implementations of discrepancies are obtained and the procedures, grounded on them, of updating(adjusting) of the data surveyed.

РЕШЕНИЕ ПЛОХО ОБУСЛОВЛЕННЫХ ЗАДАЧ ИДЕНТИФИКАЦИИ ЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ МЕТОДОМ ИНСТРУМЕНТАЛЬНЫХ ПЕРЕМЕННЫХ

© 2001 А.И. Жданов, С.Ю. Гоголева

Самарский государственный аэрокосмический университет

Решается задача оценивания параметров динамического объекта, описываемого стохастическими линейными разностными уравнениями, методом инструментальных переменных, имеющего важное прикладное значение в многочисленных задачах идентификации систем. Предлагается эффективный (по точности и числу арифметических операций) численный метод нахождения этих оценок.

Введение

Метод инструментальных переменных находит применение в математической статистике, а именно, в регрессионном анализе для оценки неизвестных параметров регрессии с ошибками в независимых переменных. При применении метода инструментальных переменных в регрессионном анализе мы сталкиваемся с трудностью выбора самих инструментальных переменных. Поэтому метод инструментальных переменных нашел большое применение в области идентификации динамических систем, описываемых стохастическими линейными разностными уравнениями, где задачи идентификации динамических систем относятся к классу регрессионных задач и инструментальные переменные выбираются естественным образом.

Определение параметров по методу инструментальных переменных приводит нас к решению плохо обусловленных систем линейных алгебраических уравнений, которые обычно решаются методом Гаусса. Метод Гаусса не является наилучшим для решения плохо обусловленных систем. В данной работе рассматривается прямой проекционный метод (ППМ) для решения задач методом инструментальных переменных. Показаны преимущества этого метода по сравнению с другими методами. Применение ППМ основано на использовании расширенной системы линейных алгебраических уравнений [4], эквивалентной системе нормальных уравнений, которая никогда ранее не рассматривалась для решения задач методом инструментальных переменных.

Постановка задачи

Рассматривается скалярный по выходу и входу стационарный линейный динамический объект с дискретным временем $t = 1, 2, \dots$, описываемый уравнением

$$y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + e(t), \quad (1)$$

где $y(t)$ – выходной сигнал, $u(t)$ – входной сигнал, а $e(t)$ – неконтролируемое возмущение,

$$A(q^{-1}) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a},$$
$$B(q^{-1}) = b_1 q^{-1} + \dots + b_{n_b} q^{-n_b},$$

где q^{-1} – оператор сдвига назад:

$$q^{-1}u(t) = u(t-1).$$

Для системы, описываемой уравнением (1) выполняются предположения:

1) объект, описываемый уравнением (1), предполагается устойчивым, т.е.

$$A(z) = 0 \Rightarrow \|z\| > 1;$$

2) $\{e(t)\}_{t \geq 1}$ – последовательность независимых случайных величин с

$$M\{e(t)\} = 0, M\{e^2(t)\} = \sigma^2 > 0;$$

3) $\{u(t)\}_{t \geq 1}$ удовлетворяет условиям постоянного возбуждения.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n u^T(t) u(t) = P_n$$

является невырожденной конечной матрицей.

При сделанных предположениях 1) – 3) требуется оценить неизвестный вектор параметров уравнения (1):

$$\theta = (a_1, a_2, \dots, a_{n_a}, b_1, b_2, \dots, b_{n_b})^T \in R^{n_a+n_b}.$$

Описание алгоритма

Введем обозначение

$$\varphi(t) = (-y(t-1), \dots, -y(t-n_a), -u(t-1), \dots, -u(t-n_b))^T.$$

Общий класс методов инструментальных переменных может быть представлен с помощью выражения

$$\hat{\theta} = \operatorname{Argmin}_{\theta} \left\| \sum_{t=1}^n \xi(t) \varphi^T(t) \right\| \theta - \sum_{t=1}^n \xi(t) y(t) \right\|_2^2,$$

где $\|\cdot\|_2$ - евклидова норма, n - число пар данных, $\xi(t)$ - вектор инструментальных переменных, удовлетворяющий условиям :

- 1) $M\xi(t)\varphi^T(t)$ - невырожденная матрица;
- 2) $M\xi(t)e(t) = 0$.

Существует несколько вариантов выбора инструментальной переменной [5,6]. Для системы (1) инструментальные переменные определяются как

$$\xi(t) = (u(t-1), \dots, u(t-n_a - n_b))^T.$$

Такой выбор инструментальных переменных для системы (1) является наиболее подходящим [5]. Оценка $\hat{\theta}$, соответствующая методу инструментальных переменных, может быть найдена из решения системы уравнений [3]

$$\left[\sum_{t=1}^n \xi(t) \varphi^T(t) \right] \theta = \sum_{t=1}^n \xi(t) y(t). \quad (2)$$

Известно, что систему уравнений (2) обычно решают с помощью метода Гаусса.

Матрица системы уравнений (2) $\xi^T \varphi^T$ является несимметричной в отличие от задачи наименьших квадратов ($\xi^T \varphi^T = \varphi^T \varphi$), где матрица системы (2) - симметричная. Это осложняет решение системы (2), которая в большинстве случаев является плохо обусловленной [5]. Поэтому предлагаются рассмотреть систему линейных алгебраических уравнений

$$\begin{cases} e(t) = y(t) - \varphi^T(t)\theta \\ \xi(t)e(t) = 0, \end{cases}$$

которые можно записать в матричном виде

$$\begin{pmatrix} E & \varphi^T(t) \\ \xi^T(t) & 0 \end{pmatrix} \begin{pmatrix} e(t) \\ \theta \end{pmatrix} = \begin{pmatrix} y(t) \\ 0 \end{pmatrix} \quad (3)$$

Показано, что исходная задача (2) эквивалентна системе линейных алгебраических уравнений (3).

$$\text{Пусть } A = \begin{pmatrix} E & \varphi^T(t) \\ \xi^T(t) & 0 \end{pmatrix}$$

Собственные числа матрицы A будут

$$\lambda_i = 1/2 + \sqrt{1/4 + \mu_i}, \text{ где } \mu_i \text{ - собственные числа матрицы } \xi^T \varphi^T.$$

Пусть σ_i - сингулярные числа матрицы A. Так как модули собственных чисел матрицы заключены между наибольшим и наименьшим сингулярными числами [2], то

$$\operatorname{cond}_2 A = \frac{\sigma_{\max}}{\sigma_{\min}} \geq \frac{|\lambda_{\max}|}{|\lambda_{\min}|} = \frac{1/2 + \sqrt{1/4 + \mu_{\max}}}{1/2 - \sqrt{1/4 + \mu_{\min}}}. \quad (4)$$

Рассмотрим систему уравнений

$$\begin{pmatrix} E & \alpha \varphi^T(t) \\ \alpha \xi^T(t) & 0 \end{pmatrix} \begin{pmatrix} \alpha e(t) \\ \theta \end{pmatrix} = \begin{pmatrix} \alpha y(t) \\ 0 \end{pmatrix} \quad (5)$$

где $\alpha > 0$ - произвольный варьируемый множитель.

Тогда собственные числа матрицы

$$A(\alpha) \text{ будут } \lambda_i = \alpha/2 + \sqrt{\alpha^2/4 + \mu_i} \text{ и}$$

$$\operatorname{cond}_2 A(\alpha) \geq \frac{\alpha/2 + \sqrt{\alpha^2/4 + \mu_{\max}}}{\alpha/2 - \sqrt{\alpha^2/4 + \mu_{\min}}}. \quad (6)$$

Легко показать, что нижняя граница (6)

достигается при $\alpha = \tilde{\alpha} = \sqrt{\frac{2}{\mu_{\min}}}$:

$$\operatorname{cond}_2 A(\alpha) \geq \frac{+ \sqrt{+ \operatorname{cond}_2 \xi^T \varphi^T}}{+ \sqrt{+ \operatorname{cond}_2 \xi^T \varphi^T}}. \quad (7)$$

Рассмотрим пример. Пусть

$$\xi(t) = \begin{pmatrix} 1000 & -2000 \\ 3 & 4 \end{pmatrix},$$

$$\varphi^T(t) = \begin{pmatrix} 3000 & 4000 \\ 1 & 2 \end{pmatrix},$$

тогда

$$cond_2(\xi(t)\varphi^T(t)) = 0.224410^7,$$

$$\tilde{\alpha} = 0.8185, cond_2(A) = 0.126710^5.$$

Из (7) и рассмотренного примера видно преимущество использования системы (5) по сравнению с (2) по численной устойчивости.

Для решения (5) предлагается прямой проекционный метод [1]. С помощью данного метода первые $n - (n_a + n_b)$ уравнений системы (5) решаются аналитически. Это показывает превосходство рассмотрения системы (5) по числу операций над системой (2), которая обычно решается методом Гаусса [5].

Заключение

Вычисляя оценки параметров динамического объекта, описываемого уравнением (1), по методу инструментальных перемен-

ных, система (2) обычно решалась методом Гаусса [6]. Использование подхода на основе расширенной системы дает преимущество как по численной устойчивости, так и по числу арифметических операций (при применении прямого проекционного метода).

СПИСОК ЛИТЕРАТУРЫ

1. Жданов А.И. Прямой рекуррентный метод решения систем линейных алгебраических уравнений // Докл. РАН. 1997.
2. Беклемищев Д.В. Дополнительные главы линейной алгебры. М.: Наука, 1983.
3. Льюис Л. Идентификация систем. Теория для пользователя. М.: Наука, 1991.
4. Bjork A. Numerical stability of methods for solving augmented systems // Contemporary Mathematics 204. 1997.
5. Stoica P., Friedlander B. and Soderstrom T. Instrumental variable methods for ARMA models. In Control and Dynamic Systems – Advances in Theory and Applications. Vol.XXIV. New York: Academic Press, 1986.
6. Soderstrom T., Stoica P. Comparison of some instrumental variable methods. Consistency and accuracy aspects // Automatica. Vol.17. 1981.

SOLVING THE ILL - CONDITIONED IDENTIFICATION PROBLEMS OF THE LINEAR DYNAMIC SYSTEMS BY THE INSTRUMENTAL VARIABLE METHOD

© 2001 A.I. Zhdanov, S.Yu. Gogoleva.

Samara State Aerospace University

The problem of estimating the parameters of the dynamic object described by stochastic linear difference equations are solved by the instrumental variable method being important in many identification problems. The effective (on accuracy and operation number) numerical method of the finding this estimates are suggested.

УДК 621.372, 681.142

РАЗРАБОТКА ИНСТРУМЕНТАЛЬНОЙ СИСТЕМЫ ДЛЯ СОЗДАНИЯ МУЛЬТИАГЕНТНЫХ ПРИЛОЖЕНИЙ В СЕТИ ИНТЕРНЕТ

© 2001 С.В. Батищев¹, О.И. Лахин¹, И.А. Минаков²
Г.А. Ржевский², П.О. Скобелев¹

¹ НПК "Генезис Знаний", г. Самара и Самарский филиал Физического
института им. П.Н. Лебедева РАН

² Маджента Корпорейшн, г. Лондон, Англия

Рассматриваются основные функции и возможности разработанного программного комплекса для построения мультиагентных систем в сравнении с аналогичными системами, анализируется его структура, архитектура и особенности реализация, предлагаются возможные применения и перспективы развития.

В настоящее время многофункциональные Internet-порталы становятся важным инструментом научной и инженерной деятельности. Однако при разработке такого рода систем часто возникает целый ряд сложных практических задач:

– Организация развитых online сервисов для пользователей (например, электронные магазины, форумы, системы дистанционного обучения, библиотеки, биржа труда, служба новостей).

– Прозрачная интеграция сервисов. Например, "получить все статьи в библиотеке по тематике форума", "заказать описанный в статье оптический элемент", "найти эксперта по установке заказанного оборудования").

– Индивидуальный подход к пользователям сайта. Возможность описывать свои предпочтения и интересы в предметной области для более эффективного доступа к информации. Возможность настраивать работу системы под свои нужды и под свой график работ.

– Обеспечение различных механизмов доступа и организации информации (стандартный браузер, e-mail, мобильные устройства).

– Максимально короткий цикл разработки.

– Управление системой "на лету", включая возможность легко уточнять описания предметной области, внедрять новые компоненты и сервисы.

Для описания систем такого класса весь-

ма эффективен мультиагентный подход. В рамках этого подхода Internet-ресурс описывается как набор небольших программных компонент - агентов, которые обладают следующими базовыми свойствами:

– Способность динамически планировать поведение на основе формализованного описания предметной области и стратегий поведения. Такое представление принято называть онтологией. В рамках теории искусственного интеллекта выработан целый ряд подходов, в той или иной степени удобных для описания онтологии (фреймы, правила вывода, семантические сети, сценарии).

– Способность взаимодействовать друг с другом для обмена информацией, поиска взаимно-оптимальных решений и стратегий, коллективного достижения целей. Взаимодействие обычно осуществляется на основе протоколов переговоров, которые являются частью онтологии.

– Способность взаимодействовать с "внешним миром". В случае агентов портала эта способность может выражаться в доступе к внешней базе данных, обработке запроса удаленного пользователя, формировании и отправке HTML (Hyper Text Markup Language) страницы (сообщения по e-mail, SMS сообщения или WAP-страницы для мобильных устройств) и т.д.

В рамках наших исследований разработано и реализовано мультиагентное ядро и набор визуальных инструментов для быстрой и эффективной разработки сложных инфор-

мационных ресурсов в Internet¹. В этой работе мы рассмотрим основные функции и возможности ядра в сравнении с аналогичными системами, проанализируем его структуру, архитектуру и особенности реализации.

Обзор архитектуры мультиагентного ядра

С точки зрения архитектуры, можно выделить 2 основных типа мультиагентных систем, поддерживающих работу в Internet:

– *Системы с мобильными агентами*. В этом случае исполняемая часть агента загружается с сервера на машину клиента, что дает возможность серьезного повышения производительности за счет реальной многозадачности. Однако из-за серьезных проблем поддержки безопасности, надежности и переносимости такие системы пока используются в ограниченном числе приложений. В основном такие системы используют язык Java для переносимого описания агентов.

– *Системы с серверными агентами*. В этом случае агенты исполняются на сервере, осуществляя взаимодействие с удаленными пользователями по стандартным протоколам Internet. При этом значительно легче решаются проблемы переносимости (клиенту достаточно иметь стандартный browser) и безопасности (исполнение осуществляется только на сервере и легко контролируется). Помимо этого, такие системы легче интегрировать с существующими решениями, такими как, системы управления базами данных (СУБД), серверы приложений, web серверы. Кроме того серверные системы могут использовать как стандартные языки (Java, C++) для описания поведения агентов, так и предлагать собственные решения (языки сценариев, правила вывода и т.п.).

Из описанных соображений в нашей системе предпочтение отдано второму подходу (более подробно проблемы выбора архитектуры изложены в [1]). При этом, ядро технически реализовано как Windows NT Service (Windows 2000 Service), имеет стандартный интерфейс для связи с web сервером Microsoft Internet Information Server (IIS), связи с СУБД по протоколу ODBC (Open Database Connectivity) и систему интерфейсов COM

(Component Object Model) для связи с другими приложениями.

Средства описания онтологий

Как было отмечено ранее, агенты используют формализованное описание предметной области – "онтологию" – для планирования поведения и реализации своих планов в рамках определенной стратегии поведения. Например, простейшим описанием онтологии может быть набор правил поведения ("предусловие"- "действие").

В мультиагентном ядре онтология агентов описывается как семантическая сеть. На базовом уровне определено несколько фундаментальных концептов, связанные с описанием семантических сетей: "объект", "сценарий", "предусловие", "отношение", "событие", "действие" и некоторые другие. Расширения ядра (внешние модули) имеют возможность посредством стандарта СОМ и любого языка программирования (поддерживающего эту технологию) дополнить (специализировать) функции этих концептов. Можно выделить два основных направления расширения функций ядра:

– *Создание универсальных объектов для описания мультиагентных систем*. Самым важным примером такого модуля является подсистема обмена сообщениями, которая вводит объекты "сообщение", "почтовый ящик", "протокол переговоров", позволяя тем самым различным агентам взаимодействовать друг с другом посредством переговоров. Кроме того, фундаментальные концепты ядра спроектированы таким образом, что дополнительные модули ядра могут относительно легко расширять их функциональность. Например, одно из стандартных расширений мультиагентного ядра вводит набор понятий для описания агентов на высоком уровне, например, "класс", "экземпляр класса", "атрибут", "наследование классов", "действие над экземпляром класса" и т.д. Таким образом можно вводить различные базовые системы понятий, сравнивая их эффективность, полноту, избыточность, удобство конструирования предметной области и т.д.

– *Создание инструментальных объектов, специфичных для определенного типа*

приложений или операционной системы. Типичными примерами являются объекты для доступа к СУБД (запрос, таблица), для формирования страниц HTML (шаблон, URL) и т.д.

Основные функции и компоненты мультиагентного ядра

Принятые архитектурные решения и средства описания онтологий, определяют следующий набор базовых, неотъемлемых сервисов мультиагентного ядра:

- подсистема загрузки и компиляции описания онтологий;
- подсистема диспетчера параллельного исполнения сценариев;
- подсистема взаимодействия с внешними модулями;
- вспомогательные сервисы (средства конфигурации, средства протоколирования).

Подсистема загрузки и компиляции описаний онтологий содержит два подмодуля:

– *Анализ и компиляция описаний семантических сетей*. Этот модуль выполняет восстановление всех концептов и связей между ними. Для описания семантических сетей используется декларативный язык, который описывает сеть как вложенные списки объектов с перекрестными ссылками. После этого выполняется перевод описания в форму, удобную для эффективной работы на этапе исполнения.

– *Компиляция сценариев*. Этот модуль выполняет перевод языка сценариев в форму байт-кода, оптимальную для быстрой работы на этапе исполнения. Сам процедурный язык сценариев основан на синтаксисе C++.

Главная задача системы параллельного исполнения сценариев – выполнять диспетчеризацию активных сценариев, т.е. хранение и переключение контекстов сценариев для псевдопараллельного исполнения. Кроме этого система отслеживает условия запуска, приостановки ("заморозки"), продолжения и завершения исполнения сценариев. Авторы внешних модулей ядра могут легко управлять процессами псевдопараллельного исполнения (например, приостановить вызывающий сценарий на время исполнения запроса к СУБД).

Подсистема взаимодействия с внешними модулями осуществляет двунаправленное взаимодействие ядра с внешними модулями (вызов функций/методов, реакция на события, предоставление стандартных функций для внешних модулей). Разработана также специальная библиотека для компиляторов C++, которая определяет набор базовых классов, шаблонов и макроопределений для эффективного описания внешних компонентов.

Как уже было отмечено ранее, реализован также ряд стандартных внешних модулей ядра:

- подсистема переговоров;
- подсистема доступа к СУБД;
- подсистема формирования HTML страниц (для сервера web и включения в e-mail);
- подсистема расширенного описания онтологий.

Средства визуальной разработки и отладки

Итак, онтологии для мультиагентного ядра описываются на специализированном языке. Для упрощения построения онтологий нами разработана система поддержки визуальной разработки ("визуальный конструктор"). Отметим основные функции этой системы:

- визуальное построение семантической сети (как направленного графа), добавление, редактирование и удаление вершин и отношений, привязка сценариев;
- прозрачная интеграция построенных семантических сетей в виде файлов проекта ("проектный подход");
- проверка построенных онтологий на синтаксическую корректность (в будущем планируется ввести возможность предварительного семантического анализа, например, на непротиворечивость);
- поддержка отладки сценариев (установка точек прерывания, пошаговой отладки);
- поддержка библиотеки стандартных объектов на основе списка активных модулей расширения ядра;
- поддержка библиотеки стандартных агентов.

На рисунке представлен пользовательский интерфейс конструктора.

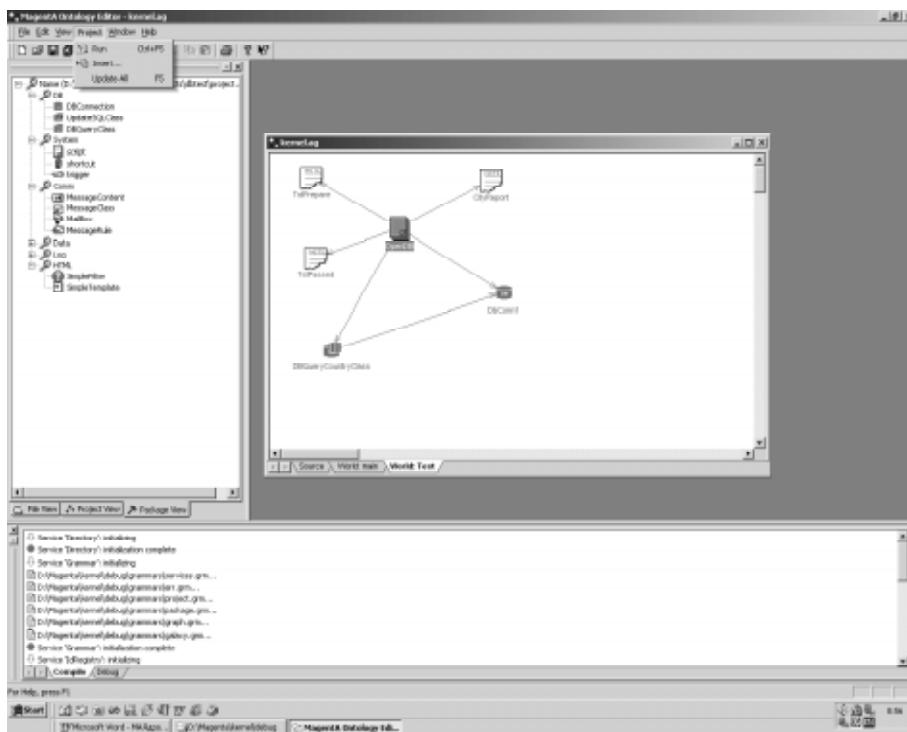


Рис. Пользовательский интерфейс конструктора

Для поддержки отладки создана локальная версия ядра, которая способна запускаться отдельно от web-сервера (или взаимодействовать с ним, если он доступен).

Интересно отметить, что сам визуальный инструмент построен на специализированной версии мультиагентного ядра, т.е. отдельные компоненты инструмента (вершины графа, объекты библиотеки) являются агентами. Это позволяет, например, легко описать особенности визуального редактирования для сложных объектов (таких как "шаблон HTML страницы", который обычно требует интерактивного импорта структуры шаблона из внешнего файла-описания). А также делает инструмент более гибким, гарантируя возможность поддержки будущих библиотек и подсистем.

Приложения на базе мультиагентного ядра и стандартные библиотеки агентов

В настоящее время на основе универсального мультиагентного ядра разрабатывается большое число разнообразных приложений. Основное внимание при этом уделяется созданию стандартных библиотек агентов для выполнения наиболее типичных задач, возникающих при разработке информацион-

ной системы в Internet. Примером такой библиотеки является набор агентов для осуществления диалога с пользователями Internet портала.

В рамках этой системы каждому ресурсу информационного пространства портала становится в соответствие агент, например, "агент статьи", "агент формы-анкеты", "агент новостей", "агент форума". Такие агенты ресурсов могут вступать в переговоры с агентами пользователя и агентами дизайна.

Задача агентов пользователей – выявлять предпочтения конкретного пользователя и представлять их в переговорах с агентами ресурсов. В результате таких переговоров происходит:

- корректировка формы представления информации ("Если вы заинтересовались этой темой, то мы можем выводить такие новости в первую очередь / присыпать извещения по e-mail");
- предложение дополнительной информации по теме ("Сейчас в форуме идет обсуждение изложенных в статье идей");
- формирование вопросов к пользователю для непрерывного уточнения знаний о нем ("Вы попросили присыпать новости, но не сообщили e-mail при регистрации").

Ожидается, что информационные сис-

темы на базе такой технологии будут существенно более эффективны и просты в управлении, чем их аналоги на базе традиционных систем.

Совместно с системой управления диалогом разработаны и реализованы ряд других мультиагентных библиотек:

- система для дистанционного обучения;
- система для поддержки приложений электронной коммерции.

В качестве примера практического использования таких систем можно привести портал "OPTIC CITY" для интеграции знаний по оптике и лазерной физике (см., например, [2]). Также большой эффект может дать применение этих библиотек в системах для электронной коммерции для корпоративных сайтов.

Заключение

Применение мультиагентного подхода может значительно повысить эффективность разработки и эксплуатации информационных систем в сети Internet. В рамках описанных в этой работе исследований был спроектирован и реализован пакет программного обеспечения для построения мультиагентных си-

стем ("мультиагентное ядро"). На основе мультиагентного ядра разработан ряд практически важных приложений и стандартных библиотек агентов.

¹ Все права на мультиагентное ядро и систему визуального проектирования принадлежат компании MagentA, plc. Разработка ряда приложений (портал OPTIC CITY) осуществляется совместно СФ ФИАН и НПК "Генезис Знаний".

СПИСОК ЛИТЕРАТУРЫ

1. Батищев С.В., Скобелев П.О. Проблемы реализации мультиагентных систем дистанционного обучения в сети Интернет // Материалы Международной научно-практической конференции "Развитие новых технологий в системе образования РФ". Самара. 2000.
2. Batishchev S.V., Boerner P., Cristoff J., Doue J., Grosse V., Skobelev P.O. The development of multiagent internet portal for E-commerce "Optic city" // Труды 2-ой международной конференции "Проблемы управления и моделирования в сложных системах". Самара. 2000.

DESIGN AND IMPLEMENTATION OF THE SOFTWARE PACKAGE FOR DEVELOPMENT OF MULTIAGENT INTERNET APPLICATIONS

© 2001 S.V. Batishchev¹, O.I. Lakhin¹, I.A. Minakov²
G.A. Rzevski², P.O. Skobelev¹

¹MEC "Knowledge Genesis", Samara and Samara Branch
Physics Institute named for P.N. Lebedev of Russian Academy of Sciences,
²MagentA Corporation, plc, England, London

This article dwells upon design and implementation of the software package for multiagent applications development. The main features of the system are described in comparison with similar systems. The structure, architecture and implementation are analyzed. Possible applications and prospects of development are proposed.

КЛАССИФИКАЦИЯ И ПРИНЦИПЫ ПОСТРОЕНИЯ МНОГОКОМПОНЕНТНЫХ ОПТРОННЫХ СТРУКТУР

© 2001 С.А. Матюнин

Самарский государственный аэрокосмический университет

В статье рассматриваются принципы спектрального взаимодействия многокомпонентных оптоэлектронных структур. Приводятся примеры реализации принципа спектрального взаимодействия в ряде оптоэлектронных преобразователей и элементов – в датчиках больших и малых перемещений, стабильных источниках и приемниках оптического излучения, устройствах гальванической развязки, коммутаторах оптических сигналов.

Общепризнанно, что применение оптоэлектроники в контрольно-измерительной технике и системах управления позволяет добиться высоких метрологических характеристик, устойчивости к электромагнитным воздействиям, позволяет создавать помехоустойчивые каналы связи, элементы практически идеальной гальванической развязки измерительных, управляющих и силовых цепей и т.д. Так как обобщенная структура оптоэлектронного устройства (ОЭУ) в каждом конкретном случае реализуется лишь частью блоков (рис.1), то можно выделить следующие группы ОЭУ: 1- оптоэлектронные элементы (ОЭЭ - источники излучения (ИИ), оптическая система (ОС), передающая и преобразующая излучение, приемники излучения (ПИ)); 2 – элементарные оптроны (используют блоки ИИ-ОС-ПИ); 3 – оптоэлект-

ронные преобразователи или устройства (содержат совокупность элементарных оптров, устройства формирования и управления).

Воздействие информационных параметров в ОЭУ происходит, в основном, двумя способами: 1 - путем пространственно-временной модуляции светового потока некогерентного излучения (к этому же, в конечном счете, сводится и изменение степени и направления поляризации излучения); 2 - путем интерференционного взаимодействия когерентных компонент излучения и их модуляции. Например, в устройствах гальванической развязки аналоговых сигналов – оптронах, воздействие информационного параметра осуществляется посредством управления интенсивностью излучения ИИ, а в оптоэлектронных аналого-цифровых датчиках

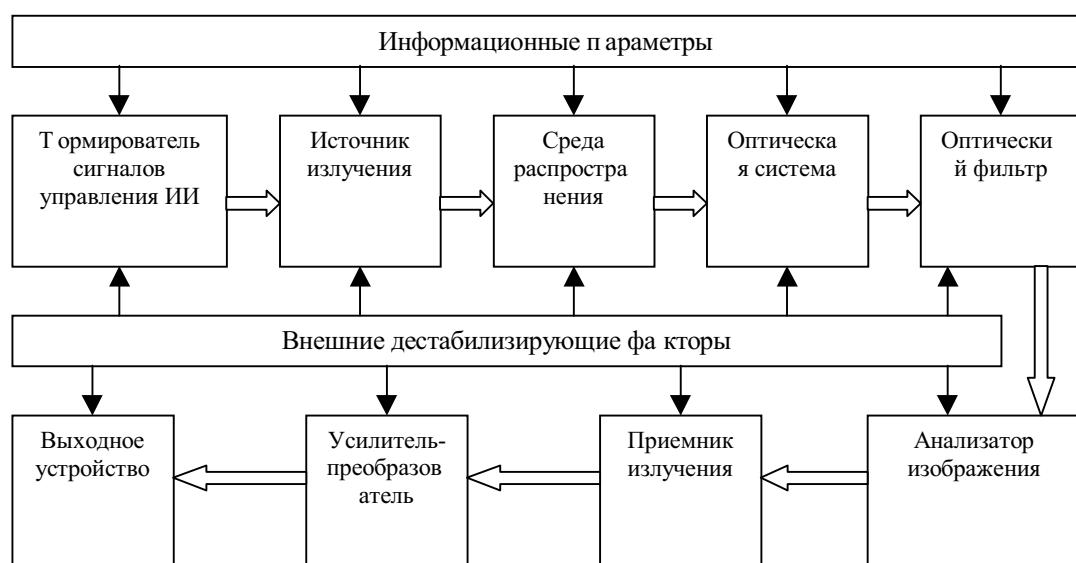


Рис. 1. Обобщенная структура ОЭУ

перемещения – посредством пространственной модуляции излучения. В тоже время, использование принципа спектрального взаимодействия позволяет получить простые по конструктивному исполнению устройства с уникальными характеристиками.

В многокомпонентных оптронных структурах спектрального взаимодействия (МОС) на поверхность оптоэлектронных элементов (рис.2) ИИ (1), ОС (4) и ПИ (7), наносится дополнительный ряд компонент (2,3,5,6), выполняемых, обычно, в виде многослойных тонкопленочных покрытий (МТП). Эти компоненты, выполняя спектроформирующую функцию, функцию избирательной, временной и пространственной фильтрации, модуляции и коммутации, существенно изменяют свойства ОЭЭ. Здесь воздействие информационных параметров происходит вследствие спектрального взаимодействия многокомпонентных оптоэлектронных элементов (МОЭ). В результате такого взаимодействия становится возможным добиться инвариантности параметров ОЭУ к внешним дестабилизирующим факторам (ВДФ), реализовать функцию временной и пространственной модуляции, коммутации, спектрального уплотнения и создать прецизионные ОЭЭ и ОЭУ [1-3].

Многокомпонентные оптронные структуры (многокомпонентные оптоэлектронные элементы) спектрального взаимодействия – это оптоэлектронные устройства (элементы), в оптическую цепь которых введены (на поверхность которых нанесены) многокомпонентные оптические структуры, выполняющие спектроформирующую, спектровзаимодействующую и термокомпенсирующую функции, функции пространственной и временной модуляции и др., воздействие информационных параметров в которых происходит в результате взаимодействия их спектральных характеристик.

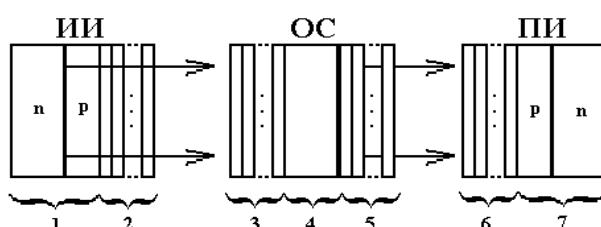


Рис. 2. Структурная схема МОС

Если обозначить

$$\Phi^*(\lambda, t, \tau, U_\phi), T^*(\lambda, t, \tau, U_T), S^*(\lambda, t, \tau, U_S)$$

- абсолютные спектральные характеристики излучения ИИ, пропускания ОС и чувствительности ПИ, U_ϕ, U_T, U_S - управляющие сигналы ИИ, ОС и ПИ, а

$$\Phi_n^{*i}(\lambda, t, \tau), T_n^{*j}(\lambda, t, \tau), S_n^{*k}(\lambda, t, \tau)$$

- абсолютные спектральные характеристики пропускания i, j, k-го МТП, нанесенных на поверхности ИИ, ОС и ПИ соответственно, то для последовательно соединенных МОЭ (рис.2) спектральный состав излучения $I^*(\lambda, t, \tau)$, падающего на ПИ, определится из выражения:

$$\begin{aligned} I^*(\lambda, t, \tau) = & \{\Phi^*(\lambda, t, \tau, U_\phi) \times \\ & \times \prod_{i \in I} \Phi_n^{*i}(\lambda, t, \tau) \} \cdot \{T^*(\lambda, t, \tau, U_T) \times \\ & \times \prod_{j \in J} T_n^{*j}(\lambda, t, \tau) \} \cdot \{S^*(\lambda, t, \tau, U_S) \times \\ & \times \prod_{k \in K} S_n^{*k}(\lambda, t, \tau)\}, \end{aligned}$$

где I, J, K – множество индексов компонент МТП, нанесенных на ИИ, элементы ОС и ПИ соответственно. Проведенные автором теоретические и экспериментальные исследования МОЭ позволяют с достаточной степенью точности и не зависимо от физической природы процессов, происходящих в них, аппроксимировать их абсолютные спектральные характеристики разностными функциями вида:

$$\Phi^*(\lambda, t, \tau, U_\phi) = \Phi_m(t, \tau, U_\phi) \cdot \Phi[\lambda - \lambda_\phi(t, \tau)]$$

$$T^*(\lambda, t, \tau, U_T) = T_m(t, \tau, U_T) \cdot T[\lambda - \lambda_T(t, \tau)]$$

$$S^*(\lambda, t, \tau, U_S) = S_m(t, \tau, U_S) \cdot S[\lambda - \lambda_S(t, \tau)]$$

где:

$\Phi[\lambda - \lambda_\phi(t, \tau)], T[\lambda - \lambda_T(t, \tau)], S[\lambda - \lambda_S(t, \tau)]$ - относительные спектральные характеристики мощности излучения ИИ, пропускания ОС и чувствительности ПИ соответственно;

$\Phi_m(t, \tau, U_\phi), T_m(t, \tau, U_T), S_m(t, \tau, U_S)$ - температурно-временные зависимости изменения максимумов их абсолютных спектраль-

ных характеристиках; $\lambda_\phi(t, \tau)$, $\lambda_T(t, \tau)$, $\lambda_s(t, \tau)$ - температурно-временные зависимости длин волн максимумов соответствующих спектральных характеристик. После введения в цепь оптической связи МТП спектральные характеристики МОЭ определяются из выражения:

$$\begin{aligned} \Phi^* &= \Phi_m(t, \tau) U_\phi \cdot \Phi[\lambda - \lambda_\phi(t, \tau)] * \\ &\times \prod_{i \in I} \{\Phi_{mn}^i(t, \tau) U_{n\phi}^i\} \cdot \Phi_n^i[\lambda - \lambda_{\phi-n}^i(t, \tau) U_{n\lambda\phi}^i] \\ T^* &= T_m(t, \tau) U_T \cdot T[\lambda - \lambda_T(t, \tau)] * \\ &\times \prod_{j \in J} \{T_{mn}^j(t, \tau) U_{nT}^j\} \cdot T_n^j[\lambda - \lambda_{T-n}^j(t, \tau) U_{n\lambda T}^j] \\ S^* &= S_m(t, \tau) U_S \cdot S[\lambda - \lambda_s(t, \tau)] * \\ &\times \prod_{k \in K} \{S_{mn}^k(t, \tau) U_{ns}^k\} \cdot S_n^k[\lambda - \lambda_{s-n}^k(t, \tau) U_{n\lambda s}^k] \end{aligned}$$

где $U_{n\phi}^i U_{nT}^j U_{ns}^k$ - управляющие воздействия на максимумы пропускания соответствующих МТП, $U_{n\lambda\phi}^i U_{n\lambda T}^j U_{n\lambda s}^k$ - управляющие воздействия на сдвиг спектральных характеристик МТП.

Так как в общем случае для МОЭ наблюдается так же координатная зависимость их спектральных характеристик (по поверхности), то их абсолютные спектральные характеристики (обозначены знаком "**)") запишутся в виде:

$$\begin{aligned} \Phi^{**}(x, y, z, \lambda, t, \tau) &= \\ &= \Phi^*(\lambda, t, \tau) U_\phi U_{n\phi}^i U_{n\lambda\phi}^i \cdot G_\phi(x, y, z) \\ T^{**}(x, y, z, \lambda, t, \tau) &= \\ &= T^*(\lambda, t, \tau) U_T U_{nT}^j U_{n\lambda T}^j \cdot G_T(x, y, z), \quad (1) \\ S^{**}(x, y, z, \lambda, t, \tau) &= \\ &= S^*(\lambda, t) U_S U_{ns}^k U_{n\lambda s}^k \cdot G_S(x, y, z) \end{aligned}$$

где: $G_\phi(x, y, z)$, $G_T(x, y, z)$, $G_S(x, y, z)$ - весовые функции, учитывающие координатную зависимость соответствующих спектральных характеристик ИИ, ОС, ПИ, а $U_\phi^{**} U_T^{**} U_S^{**}$ - множество управляющих воздействий МОЭ. Тогда с учетом [4] выражения для освещенности чувствительной пло-

щадки P_P ПИ от смешенного относительно него на величину x_0, y_0 ИИ с излучающей площадкой P_H , получим обобщенное выражение для величины выходного сигнала ПИ в виде:

$$\begin{aligned} J &= \int_{\lambda_H P_H}^{\lambda_B} \int \frac{1}{z_0^2} \int \frac{\Phi^{**}(x_0, y_0, z_0, \lambda, t, \tau) U_\phi^{**}}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]} \times \\ &\times \frac{T^{**}(x, y, z, \lambda, t, \tau) U_T^{**}}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]} \times \\ &\times \frac{S^{**}(x, y, z, \lambda, t, \tau) U_S^{**} \cdot dx_0 dy_0}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]} dx dy d\lambda \end{aligned} \quad (2)$$

Выражение (2) позволяет учесть как влияние информационных и управляющих воздействий на МОЭ, так и влияние ВДФ.

В таблице 1 с позиций теории чувствительности приведены выражения для определения чувствительность МОС $\alpha_I = \frac{\Delta J}{J}$ к изменению управляющих сигналов и ВДФ. В таблице 1 приняты следующие условные обозначения:

$$\alpha_{\phi M}^\tau = \frac{1}{\Phi_m(t, \tau) U_\phi} \frac{\partial \Phi_m(t, \tau) U_\phi}{\partial \tau},$$

$$\alpha_{SM}^\tau = \frac{1}{S_m(t, \tau) U_S} \frac{\partial S_m(t, \tau) U_S}{\partial \tau},$$

$$\alpha_{TM}^\tau = \frac{1}{T_m(t, \tau) U_T} \frac{\partial T_m(t, \tau) U_T}{\partial \tau}$$

- инерционность ИИ, ОС, ПИ;

$$\alpha_{\phi Mn}^{ix} = \frac{1}{\Phi_{mn}^i(t, \tau) U_\phi} \frac{\partial \Phi_{mn}^i(t, \tau) U_\phi}{\partial \tau},$$

$$\alpha_{TMn}^{ix} = \frac{1}{T_{mn}^j(t, \tau) U_T} \frac{\partial T_{mn}^j(t, \tau) U_T}{\partial \tau},$$

$$\alpha_{SMn}^{kt} = \frac{1}{S_{mn}^k(t, \tau) U_T} \frac{\partial S_{mn}^k(t, \tau) U_T}{\partial \tau}$$

- инерционность ИИ, ОС, ПИ по цепи управления МТП;

$$\alpha_{\phi M}^t = \frac{1}{\Phi_m(t, \tau) U_\phi} \frac{\partial \Phi_m(t, \tau) U_\phi}{\partial t},$$

Таблица 1. Чувствительность МОС к управляющим воздействиям и ВДФ

Воздействие	Наименование воздействия	Коэффициент чувствительности
1	2	3
$\Delta\tau$	время	$\alpha_{\phi M}^{\tau} + \alpha_{SM}^{\tau} + \alpha_{TM}^{\tau} + \sum_i \alpha_{\phi Mn}^{i\tau} + \sum_j \alpha_{TMn}^{j\tau} + \sum_k \alpha_{SMn}^{k\tau} + \frac{1}{A} \int_{\lambda_H}^{\lambda_B} A_1 [\lambda_{\phi} \beta_{\lambda\phi} \alpha_{\lambda\phi}^{\tau} + \lambda_S \beta_{\lambda S} \alpha_{\lambda S}^{\tau} + \lambda_T \beta_{\lambda T} \alpha_{\lambda T}^{\tau} + \sum_{i \in I} \lambda_{\phi n}^i \beta_{\lambda\phi n}^i \alpha_{\lambda\phi n}^{i\tau} + \sum_{j \in J} \lambda_{Tn}^j \beta_{\lambda Tn}^j \alpha_{\lambda Tn}^{j\tau} + \sum_{k \in K} \lambda_{Sn}^k \beta_{\lambda Sn}^k \alpha_{\lambda Sn}^{k\tau}] d\lambda$
Δt	температура	$\alpha_{\phi M}^t + \alpha_{SM}^t + \alpha_{TM}^t + \sum_i \alpha_{\phi Mn}^{it} + \sum_j \alpha_{TMn}^{jt} + \sum_k \alpha_{SMn}^{kt} + \frac{1}{A} \int_{\lambda_H}^{\lambda_B} A_1 [\lambda_{\phi} \beta_{\lambda\phi} \alpha_{\lambda\phi}^t + \lambda_S \beta_{\lambda S} \alpha_{\lambda S}^t + \lambda_T \beta_{\lambda T} \alpha_{\lambda T}^t + \sum_{i \in I} \lambda_{\phi n}^i \beta_{\lambda\phi n}^i \alpha_{\lambda\phi n}^{it} + \sum_{j \in J} \lambda_{Tn}^j \beta_{\lambda Tn}^j \alpha_{\lambda Tn}^{jt} + \sum_{k \in K} \lambda_{Sn}^k \beta_{\lambda Sn}^k \alpha_{\lambda Sn}^{kt}] d\lambda$
$\Delta U_{\phi}, \Delta U_{n\phi}^i$	интенсивность излучения ИИ	$\alpha_{\phi M}^{U\phi} \Delta U_{\phi} + \sum_i \alpha_{\phi Mn}^{iU\phi} \Delta U_{n\phi}^i$
$\Delta U_T, \Delta U_{nT}^j$	пропускание ОС	$\alpha_{TM}^{UT} \Delta U_T + \sum_j \alpha_{TMn}^{jUT} \Delta U_{nT}^j$
$\Delta U_S, \Delta U_S^k$	чувствительность ПИ	$\alpha_{SM}^{US} \Delta U_S + \sum_k \alpha_{Smn}^{kUS} \Delta U_S^k$
$\Delta U_{n\lambda\phi}^i$	сдвиг спектральной характеристики МТП ИИ	$\frac{1}{A} \int_{\lambda_H}^{\lambda_B} A_1 [\sum_{i \in I} \lambda_{\phi n}^i \beta_{\lambda\phi n}^i \alpha_{\lambda\phi n}^{iU\lambda} \Delta U_{n\lambda\phi}^i] d\lambda$
$\Delta U_{n\lambda T}^j$	сдвиг спектральной характеристики МТП ОС	$\frac{1}{A} \int_{\lambda_H}^{\lambda_B} A_1 [\sum_{j \in J} \lambda_{Tn}^j \beta_{\lambda Tn}^j \alpha_{\lambda Tn}^{jU\lambda} \Delta U_{n\lambda T}^j] d\lambda$
$\Delta U_{n\lambda S}^k$	сдвиг спектральной характеристики МТП ПИ	—

$$\alpha_{TM}^{\tau} = \frac{1}{T_M(\tau, \tau U_T)} \frac{\partial T_M(\tau, \tau U_T)}{\partial \tau},$$

$$\frac{1}{S_M(\tau, \tau U_S)} \frac{\partial S_M(\tau, \tau U_S)}{\partial \tau}$$

- температурная чувствительность ИИ, ОС, ПИ;

$$\alpha_{\phi Mn}^{it} = \frac{1}{\Phi_{mn}^i(\tau, \tau U_{\phi})} \frac{\partial \Phi_{mn}^i(\tau, \tau U_{\phi})}{\partial \tau},$$

$$\alpha_{TM}^{jt} = \frac{1}{T_{mn}^j(\tau, \tau U_T)} \frac{\partial T_{mn}^j(\tau, \tau U_T)}{\partial \tau},$$

$$\alpha_{SM}^{kt} = \frac{1}{S_{mn}^k(\tau, \tau U_T)} \frac{\partial S_{mn}^k(\tau, \tau U_T)}{\partial \tau}$$

- температурная чувствительность ИИ, ОС, ПИ по цепи МТП;

$$A_1 = \Phi[\lambda - \lambda_{\phi}(t, \tau)] \times \\ \times \prod_{i \in I} \Phi_n^i[\lambda - \lambda_{\phi n}^i(t, \tau, U_{n\lambda\phi}^i)] \cdot T[\lambda - \lambda_T(t, \tau)] \cdot \\ \prod_{j \in J} T_n^j[\lambda - \lambda_{Tn}^j(t, \tau, U_{n\lambda T}^j)] \cdot S[\lambda - \lambda_S(t)] \times \\ \times \prod_{k \in K} S_n^k[\lambda - \lambda_{Sn}^k(t, \tau, U_{n\lambda S}^k)] d\lambda$$

$$A = \int_{\lambda_H}^{\lambda_B} A_1 d\lambda;$$

$$\alpha_{\lambda\phi}^{\tau} = \frac{1}{\lambda_{\phi}(\tau, \tau)} \frac{\partial \lambda_{\phi}(\tau, \tau)}{\partial \tau},$$

$$\alpha_{\lambda T}^{\tau} = \frac{1}{\lambda_T(\tau, \tau)} \frac{\partial \lambda_T(\tau, \tau)}{\partial \tau},$$

$$\alpha_{\lambda s}^{\tau} = \frac{1}{\lambda_s(t, \tau)} \frac{\partial \lambda_s(t, \tau)}{\partial \tau}$$

- инерционность изменения длин волн максимумов спектральных характеристик ИИ, ОС, ПИ;

$$\alpha_{\lambda \phi n}^{ir} = \frac{1}{\lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i} \frac{\partial \lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i}{\partial \tau},$$

$$\alpha_{\lambda Tn}^{jr} = \frac{1}{\lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j} \frac{\partial \lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j}{\partial \tau},$$

$$\alpha_{\lambda Sn}^{kr} = \frac{1}{\lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k} \frac{\partial \lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k}{\partial \tau}$$

- инерционность изменения длин волн максимумов спектральных характеристик МТП соответственно ИИ, ОС, ПИ;

$$\alpha_{\lambda \phi}^t = \frac{1}{\lambda_{\phi}(t, \tau)} \frac{\partial \lambda_{\phi}(t, \tau)}{\partial t},$$

$$\alpha_{\lambda T}^t = \frac{1}{\lambda_T(t, \tau)} \frac{\partial \lambda_T(t, \tau)}{\partial t},$$

$$\alpha_{\lambda s}^t = \frac{1}{\lambda_s(t, \tau)} \frac{\partial \lambda_s(t, \tau)}{\partial t}$$

- температурный коэффициент изменения длин волн максимумов спектральных характеристик ИИ, ОС, ПИ;

$$\alpha_{\lambda \phi n}^{it} = \frac{1}{\lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i} \frac{\partial \lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i}{\partial t},$$

$$\alpha_{\lambda Tn}^{jt} = \frac{1}{\lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j} \frac{\partial \lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j}{\partial t},$$

$$\alpha_{\lambda Sn}^{kt} = \frac{1}{\lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k} \frac{\partial \lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k}{\partial t}$$

- температурные коэффициенты изменения длин волн максимумов спектральных характеристик МТП соответственно ИИ, ОС, ПИ;

$$\beta_{\lambda \phi} = \frac{\partial \Phi[\lambda - \lambda_{\phi}]}{\partial \lambda_{\phi}},$$

$$\beta_{\lambda T} = \frac{\partial T[\lambda - \lambda_T]}{\partial \lambda_T},$$

$$\beta_{\lambda s} = \frac{\partial S[\lambda - \lambda_s]}{\partial \lambda_s}$$

- крутизна спектральной характеристики ИИ, ОС, ПИ;

$$\beta_{\lambda \phi n}^i = \frac{\partial \Phi_n^i[\lambda - \lambda_{\phi n}^i]}{\partial \lambda_{\phi n}^i},$$

$$\beta_{\lambda Tn}^j = \frac{\partial T[\lambda - \lambda_{Tn}^j]}{\partial \lambda_{Tn}^j},$$

$$\beta_{\lambda s}^k = \frac{\partial S[\lambda - \lambda_{Sn}^k]}{\partial \lambda_{Sn}^k}$$

- крутизна спектральной характеристики МТП соответственно ИИ, ОС, ПИ;

$$\alpha_{\phi M}^{U\phi} = \frac{1}{\Phi_M(t, \tau) U_{\phi}} \frac{\partial \Phi_M(t, \tau) U_{\phi}}{\partial U_{\phi}},$$

$$\alpha_{TM}^{UT} = \frac{1}{T_M(t, \tau) U_T} \frac{\partial T_M(t, \tau) U_T}{\partial U_T},$$

$$\alpha_{SM}^{US} = \frac{1}{S_M(t, \tau) U_S} \frac{\partial S_M(t, \tau) U_S}{\partial U_S}$$

- чувствительность максимумов спектральных характеристик ИИ, ОС, ПИ к соответствующим управляющим воздействиям;

$$\alpha_{\phi Mn}^{iU\phi} = \frac{1}{\Phi_{Mn}(t, \tau) U_{\phi n}^i} \frac{\partial \Phi_{Mn}(t, \tau) U_{\phi n}^i}{\partial U_{\phi n}^i},$$

$$\alpha_{TMM}^{jUT} = \frac{1}{T_{Mn}(t, \tau) U_{Tn}^j} \frac{\partial T_{Mn}(t, \tau) U_{Tn}^j}{\partial U_{Tn}^j},$$

$$\alpha_{SMM}^{kUS} = \frac{1}{S_{Mn}(t, \tau) U_{Sn}^k} \frac{\partial S_{Mn}(t, \tau) U_{Sn}^k}{\partial U_{Sn}^k}$$

- чувствительность максимумов спектральных характеристик МТП соответственно ИИ, ОС, ПИ к соответствующим управляющим воздействиям;

$$\alpha_{\lambda \phi n}^{iU\lambda} = \frac{1}{\lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i} \frac{\partial \lambda_{\phi n}^i(t, \tau) U_{n\lambda\phi}^i}{\partial U_{n\lambda\phi}^i},$$

$$\alpha_{\lambda Tn}^{jU\lambda} = \frac{1}{\lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j} \frac{\partial \lambda_{Tn}^j(t, \tau) U_{n\lambda T}^j}{\partial U_{n\lambda T}^j},$$

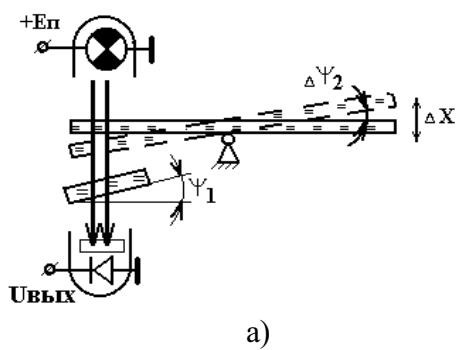
$$\alpha_{\lambda Sn}^{kU\lambda} = \frac{1}{\lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k} \frac{\partial \lambda_{Sn}^k(t, \tau) U_{n\lambda S}^k}{\partial U_{n\lambda S}^k}$$

- чувствительность изменения длин волн максимумов спектральных характеристик МТП соответственно ИИ, ОС, ПИ к соответствую-

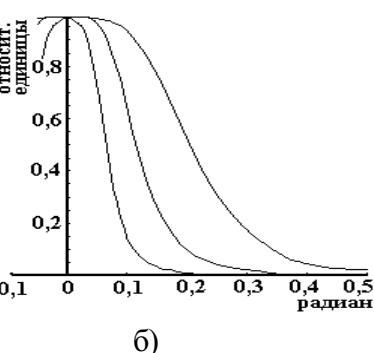
щим управляющим воздействиям. Как видно из (2) и таблицы 1 выходной сигнал ПИ определяется величиной излучаемого ИИ светового потока, спектральными характеристиками МОЭ и взаимодействием этих спектральных характеристик. Причем, этим взаимодействием можно управлять как путем изменения соответствующих управляющих воздействий, так и изменением крутизны и взаимного расположения спектральных характеристик МОЭ. Так, например, примем, что МОС состоит из обычных ИИ, ПИ и двухкомпонентного МОЭ оптической системы, что ИИ и ПИ питаются от стабильных источников питания и характеризуется равномерной плотностью потока излучения по поверхности, а управляющие сигналы воздействуют только на МОЭ оптической системы. Тогда выходной сигнал ПИ определится из выражения:

$$\begin{aligned} J(U_{n\lambda T}^{1/2} U_{n\lambda T}^{1/2} U_{Tn}^{1/2} U_{Tn}^{1/2}) = \\ = \{\Phi_m \cdot T_m \cdot S_m\} \cdot T_{mn}^{1/2} U_{nT}^{1/2} \cdot T_{mn}^{1/2} U_{nT}^{1/2} \cdot \\ \int_{\lambda_B}^{\lambda_E} \Phi[\lambda - \lambda_\phi] T[\lambda - \lambda_T] B[\lambda - \lambda_s] T_n^{1/2} \times \\ \times [\lambda - \lambda_{Tn}^{1/2} U_{n\lambda T}^{1/2}] T_n^{1/2} [\lambda - \lambda_{Tn}^{1/2} U_{n\lambda T}^{1/2}] d\lambda \quad (3) \end{aligned}$$

Пусть, для определенности, МОЭ сформированы на базе интерференционных фильтров, для которых характерно изменение длины волны максимума пропускания при изменении оптической толщины резонансных слоев (например, путем изменения углов наклонов ψ_1, ψ_2 фильтров к потоку излучения), тогда:



a)



б)

Рис. 3. Схема и позиционные характеристики преобразователя угловых перемещений

$$T_n^1 [\lambda - \lambda_{Tn}^1 U_{n\lambda T}^{1/2}] \models \frac{1}{1 + B_1 \cdot \text{Sin} \frac{d_1}{\lambda} \text{Cos} \psi_1};$$

$$T_n^2 [\lambda - \lambda_{Tn}^2 U_{n\lambda T}^{1/2}] \models \frac{1}{1 + B_2 \cdot \text{Sin} \frac{d_2}{\lambda} \text{Cos} \psi_2}$$

где B_1, B_2 - определяют полуширины пропускания фильтров (определяются конструктивными параметрами); d_1, d_2 - оптические толщины резонансных слоев фильтров. Тогда, если зафиксировать один фильтр неподвижно, а второй смещать относительно первого на некоторый угол, то в результате спектрального взаимодействия их характеристик выходной сигнал ПИ будет определяться изменением угла наклона второго фильтра. На рис. 3 изображено семейство позиционных характеристик такого преобразователя угловых перемещений при различной крутизне спектральной характеристики фильтров. Как видно, даже при небольших наклонах фильтра (наклон в 0,1 радиана при размере фильтра 10 мм соответствует линейному перемещению в 1 мм) наблюдается существенное изменение выходного сигнала ПИ.

Если принять в (2), что управляющие воздействия на МОЭ неизменны во времени, отсутствует координатная зависимость спектральных характеристик элементов, а ВДФ проявляются только в температурных изменениях характеристик, то температурную зависимость выходного сигнала ПИ можно определить из (4). Откуда следует, что температурная зависимость выходного сигнала

ПИ определяется температурными зависимостями спектральных характеристик МОЭ и их взаимодействием.

$$J(t) = \Phi_m(t) \mathbb{F}_m^{-1}(t) \cdot \int_{\lambda_H}^{\lambda_B} \Phi[\lambda - \lambda_{\phi_m}(t)] [\lambda - \lambda_{m_n}(t)] \times \lambda \times [\lambda - \lambda_{s_m}(t)]^{-1} [\lambda - \lambda_{t_n}(t)] d\lambda, \quad (4)$$

Если учесть, что при изменении температуры МОЭ наблюдается сдвиг их спектральных характеристик, то из условия:

$$\int_{t_H}^{t_B} [J(t) - J_0]^2 dt \rightarrow \min$$

можно определить оптимальную характеристику МТП $\mathbb{F}_m^{-1}(t) \lambda_n^{-1} [\lambda - \lambda_{t_n}(t)]$, обеспечивающую необходимую температурную стабильность параметров МОС.

Проведенные автором экспериментальные и теоретические исследования, позволили создать ряд термокомпенсированных аналоговых МОЭ спектрального взаимодействия и прецизионных измерительных устройств на их основе. Так на рис.4 приведены экспериментальные температурные зависимости мощности излучения арсенидо-галлиевого светодиода, термокомпенсированного по цепи оптического канала путем нанесения на наружную поверхность кристалла светодиода МТП и температурные зависимости текущей относительной погрешности термо компенсации $P(t)$. Как видно, при толщине МТП $d=0,916$ мкм в диапазоне температур от 0 до 110 градусов температурная погрешность ИИ не превышает 0,3%.

Классификация МОС спектрального вза-

имодействия (рис. 5) отражает наиболее важные классы.

По виду конструктивного исполнения МТП можно выделить на три класса: МОС дискретного исполнения, МОС интегрального исполнения и комбинированные МОС. В первом случае МТП выполняются в виде отдельных конструктивно законченных оптических спектро-формирующих элементов (СФЭ), во втором случае МТП наносятся непосредственно на рабочую поверхность МОЭ, в третьем случае часть МОЭ выполняется по первому, а часть – по второму вариантам. Очевидно, что во втором варианте исполнения МОС, благодаря малой толщине МТП (единицы-доли мкм) и непосредственному контакту с рабочей поверхностью МОЭ обеспечиваются не только минимальные габаритно-массовые показатели, но и идеальный тепловой контакт МТП с поверхностью МОЭ (последнее обстоятельство позволило создать ряд МОС и МОЭ, термокомпенсированных по оптическому каналу). Выполнение же МТП в едином технологическом процессе с изготовлением МОЭ гарантирует высокую повторяемость их характеристик и высокую надежность. По виду исполнения СФЭ можно выделить: МОС с объемными СФЭ – здесь формирование спектральных характеристик происходит за счет поглощения, отражения и рассеивания излучения в веществе СФЭ; МОС с толстопленочными СФЭ – здесь формирование спектральных характеристик происходит так же за счет объемных эффектов в веществе, но СФЭ выполняется по толстопленочной технологии

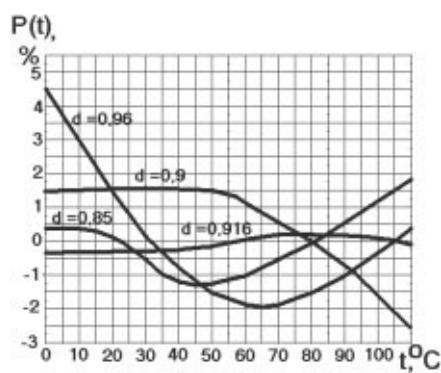
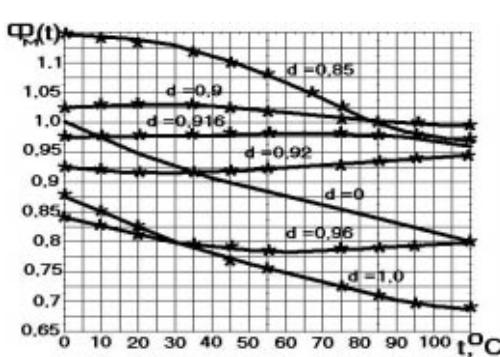


Рис. 4. Температурные характеристики и погрешность термо компенсации многокомпонентного арсенидо-галлиевого светодиода при различных толщинах МТП (мкм)

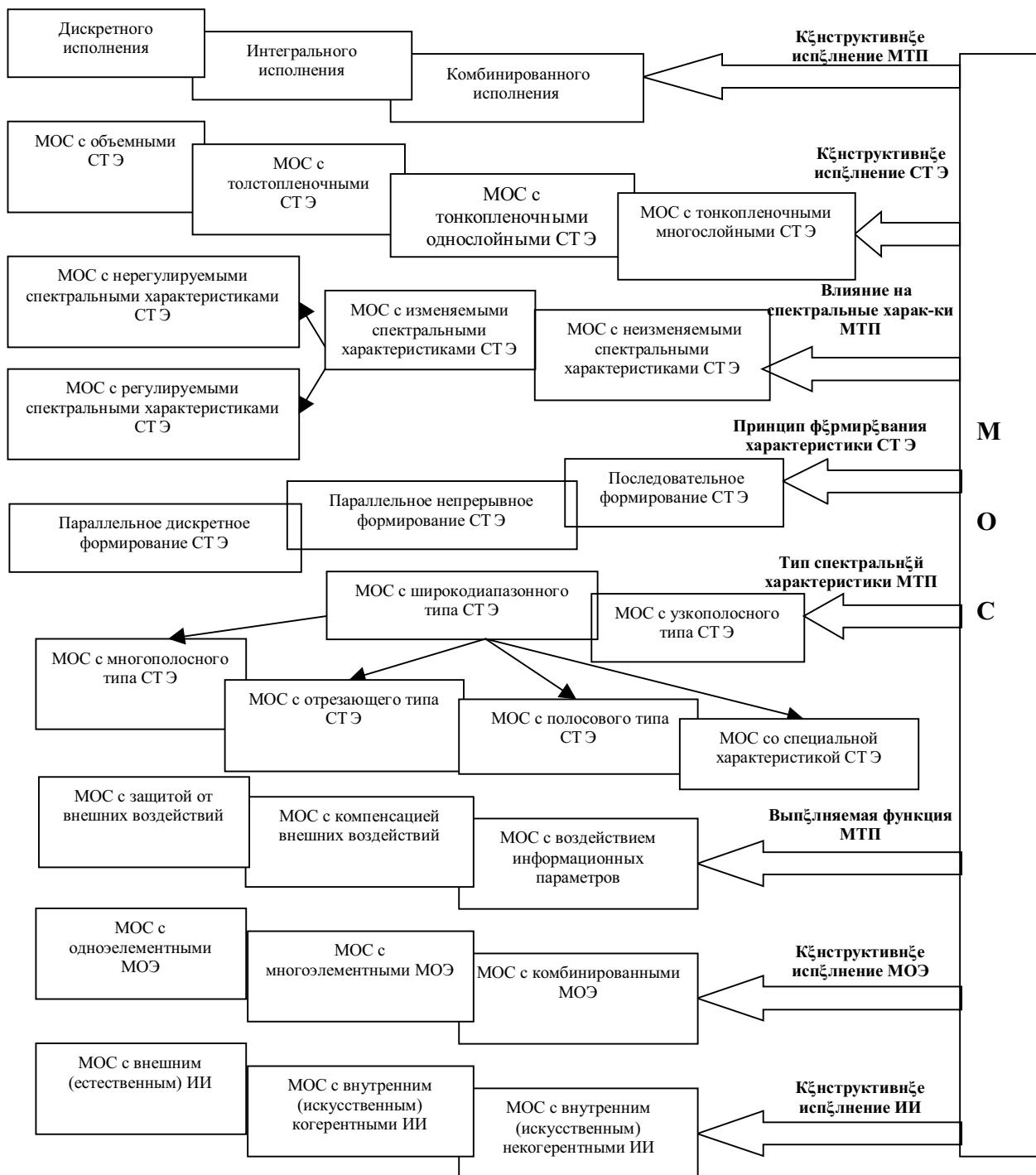


Рис. 5. Классификационная схема МОС

путем нанесения на поверхность МОЭ; МОС с тонкопленочными одно- и многослойными СФЭ – здесь формирование спектральных характеристик происходит, в основном, за счет интерференционных явлений в пленочных структурах. По виду влияния внутренних управляющих или ВДФ на спектральные характеристики МОЭ можно выделить: МОС с неизменными спектральными характеристиками и МОС с изменяемыми спектральными характеристиками. Последние делятся на

МОС с регулируемыми (управляемыми) спектральными характеристиками СФЭ под действием специально сформированных тепловых, электромагнитных, звуковых и т.д. полей, и на МОС с изменением спектральных характеристик под действием ВДФ (температуры, вибрации и т.д.). По принципу формирования спектральной характеристики СФЭ можно выделить: МОС с параллельно-дискретным, с параллельно-непрерывным и последовательным формированием спектраль-

ной характеристики СФЭ. В МОС с *параллельно-дискретным* формированием спектральной характеристики СФЭ представляет из себя (рис.6) множество одновременно освещаемых параллельным световым потоком элементарных дискретных спектроформирующих элементов СФЭ₁ ... СФ_m, каждый из

которых реализует свою часть итоговой спектральной характеристики. Здесь суммирование сигналов, прошедших через СФЭ, осуществляется в ПИ. В этом случае, при пространственно равномерном световом потоке ИИ выходной сигнал ПИ определяется из выражения:

Таблица 2. Основные характеристики элементов и устройств МОС

Тип МОС	Взаимодействие	Характеристика	Значение
Модуль гальванической развязки сигналов датчиков	Спектральная термокомпенсация ИИ-ПИ	Количество гальванически развязанных каналов	от 1 до 16
		Напряжение развязки, кВ	от 1 до 6
		К-т нелинейности характеристики, %	0,02
		Приведенный температурный коэффициент передаточной характеристики, не более %/°C	0,005
		Напряжение питания, В	+12,6; -12,6
		Габаритные размеры модуля, мм	100x150
Кодирующий фотоприемник	Спектральная термокомпенсация ИИ-ПИ	Число двоичных разрядов фотоприемника	10
		Число двоичных разрядов выходного кода датчика	14
		Диаметр фотоприемника, мм	40
		Максимальная частота считывания, кГц	100
		Диапазон измеряемых перемещений, град.	0...360
		Диапазон рабочих температур, °C	-20...+60
Аналоговый ПИ угловых перемещений	Спектральная термокомпенсация ИИ-ПИ	Диапазон измеряемых перемещений, град.	0...30
		Основная погрешность, %	0,25
		Дополнительная температурная погрешность, %	0,3
		Диапазон рабочих температур, °C	-20...+60
Аналоговый датчик угловых перемещений	Спектральное взаимодействие при угловом перемещении	Диапазон измеряемых перемещений, град.	0...3
		Основная погрешность, угл.сек.	25
		Дополнительная температурная погрешность, угл.сек.	30
		Диапазон рабочих температур, °C	-20...+60
Аналоговый датчик линейных перемещений	Спектральное взаимодействие при угловом перемещении	Диапазон измеряемых перемещений, мм	0...1
		Основная погрешность, %	0,3
		Дополнительная температурная погрешность, %	0,3
		Диапазон рабочих температур, °C	-20...+60
Коммутатор оптических сигналов	Электрическое управление спектральной характеристикой	Длина волны излучения, мкм	0,98
		Полоса пропускания, мкм	0,1
		Количество каналов	100 (10*10)
АЦП линейных перемещений со спектральным кодированием	Пространственная модуляция спектральной характеристики МТП	Диапазон измеряемых перемещений, мм	0...50
		Количество двоичных разрядов	10
		Спектральный диапазон излучения, мкм	0,4-0,98
		Диапазон рабочих температур, °C	-20...+60
Оптический коммутатор-демультиплексор	Электрическое управление спектральной характеристикой	Длина волны излучения, мкм	0,98
		Количество каналов	10
		Полоса пропускания, мкм	0,1
		Диапазон рабочих температур, °C	-20...+60

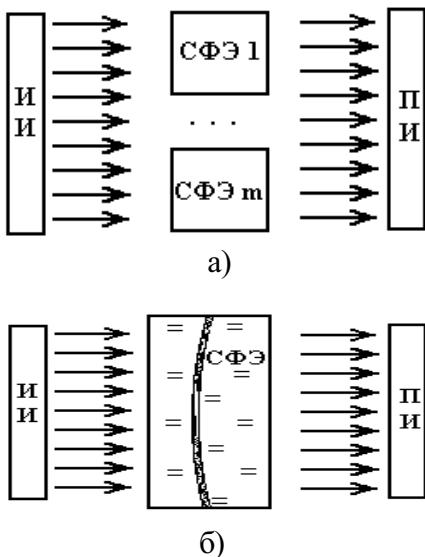


Рис. 6. Параллельно-дискретное и параллельно-непрерывное формирование спектральной характеристики

$$J(t) = F \left\{ \int_{\lambda_H}^{\lambda_B} J^*(\lambda, t) d\lambda \right\},$$

$$J^*(\lambda, t) = \Phi^*(\lambda, t) \cdot \sum_{j \in J} T_n^{*j}(\lambda, t) \cdot S^*(\lambda, t).$$

МОС с параллельно-непрерывным формированием спектральной характеристики (рис.6) отличается тем, что СФЭ выполнен в виде единого конструктива, реализующего необходимую зависимость. Здесь используется зависимость спектральной характеристики МТП от угла падения излучения на их поверхность [5]. Для данного варианта МОС при пространственно равномерном световом потоке ИИ выходной сигнал ПИ определяется из выражения:

$$\begin{aligned} J^*(\lambda, t) &= \Phi^*(\lambda, t) \cdot S^*(\lambda, t) \times \\ &\times \int_x \int_y T_n^{*j}(\lambda, x, y, t) dx dy, \end{aligned}$$

а профиль подложки СФЭ $\psi(x)$ определяется из интегрального уравнения Фредгольма:

$$T(\lambda) = \iint \frac{C \cdot dx \cdot dy}{B A + D \cdot \cos \left\{ \frac{B}{\lambda} \operatorname{Cos}[\arctg \frac{dy}{dx} + \arctg \frac{dy}{dx}] \right\}}$$

где: A, B, C, D – константы, определяемые параметрами МТП.

МОС с последовательным формированием спектральной характеристики (рис.7) представляет из себя множество последовательно расположенных конструктивно закон-

ченных элементарных дискретных или выполненных в виде единого конструктива спектоформирующих элементов. Здесь результирующая спектральная характеристика получается в результате произведения спектральных характеристик элементарных спектоформирующих элементов:

$$J^*(\lambda, t) = \Phi^*(\lambda, t) \cdot \prod_{j \in J} T_n^{*j}(\lambda, t) \cdot S^*(\lambda, t)$$

По виду спектральной характеристики спектоформирующего элемента можно выделить: МОС с узкополосной и с широкополосной спектральной характеристикой спектоформирующего элемента. МОС с широкополосной характеристикой спектоформирующего элемента делятся на многополосные, полосовые и отрезающие. Особо необходимо выделить класс МОС со специально сформированной спектральной характеристикой спектоформирующего элемента - как класс МОЭ с уникальными эксплуатационными характеристиками.

По назначению МТП можно выделить: МОС, в которых МТП выполняет функцию защиты элемента от внешних воздействий - спектоформирующая функция является не основной, а побочной; МОС, в которых МТП выполняет функцию компенсации внешних воздействий на характеристики элементов; МОС, в которых действие информационных параметров на спектоформирующий элемент приводит к спектральной, пространственной и временной модуляции светового потока.

По виду конструктивного исполнения МОЭ можно выделить: МОС с одно- и многоэлементными МОЭ и комбинированные МОС, в которых часть МОЭ представляет из себя элементарные источники и приемники излучения, а часть – многоэлементные МОЭ.

По виду используемого ИИ можно вы-

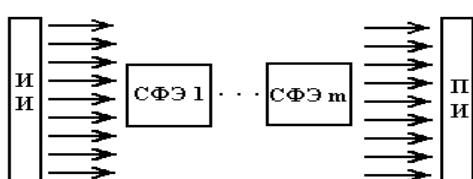


Рис.7. Последовательное формирование спектральной характеристики

делить: МОС с внешним естественным источником излучения; МОС с внутренним (искусственным) источником когерентного или некогерентного излучения.

На основе описанного принципа спектрального взаимодействия автором разработан и исследован ряд МОЭ и МОС, характеристики некоторых из них приведены в таблице 2.

СПИСОК ЛИТЕРАТУРЫ

1. *Матюнин С.А.* Многокомпонентные оптоэлектронные аналоговые и аналого-цифровые преобразователи // Труды научно-технической конференции "Датчики и преобразователи информации систем измерения, контроля и управления". Судак, 2001.
2. *Плют А.А., Матюнин С.А.* Способ температурной стабилизации световых характеристик оптоэлектронных устройств // Радиотехника. 1981. Т.36. №2.
3. *Конюхов Н.Е., Леонович Г.И., Матюнин С.А.* Оптоэлектронные цифровые преобразователи перемещений с интерференционными элементами коррекции и обработка оптических сигналов // Приборы и системы управления. 1990. №9.
4. *Леонович Г.И.* Оптоэлектронные цифровые датчики перемещений для жестких условий эксплуатации. Самара: СГАУ, 1998.
5. *Матюнин С.А.* Интерференционный светофильтр с профилированной подложкой // Известия вузов. Приборостроение, 1982. №11.

PRINCIPLES OF CONSTRUCTION OF MULTICOMPONENT OPTRON SYSTEMS OF SPECTRAL INTERPLAY

© 2001 S.A. Matjunin

Samara State Aerospace University

In the article the principles of spectral interplay of multicomponent optoelectronic patterns are esteemed. The examples of implementation of a principle of spectral interplay in a series of optoelectronic converters and members - in sensors of large and small movements, stable sources and receivers of optical radiation, devices of a galvanic isolation, switches of optical signals are resulted.

НЕПАРАМЕТРИЧЕСКИЙ АНАЛИЗ В ЗАДАЧАХ ОХРАНЫ ЗДОРОВЬЯ

© 2001 В.А. Акулов

Самарский государственный технический университет

Технология охраны здоровья, ориентированная на проблемы ранней диагностики, профилактики и лечения заболеваний, получившая название "Двойной скрининг-тест", основана на выполнении и компьютерном анализе двух скрининг-тестов в условиях, максимально приближенных к реальным, связанным с профессиональной деятельностью. Первое обследование выполняется в начале работы (перед прогнозируемым стрессом), а второе – в конце работы (после стресса). По реакции организма на реальные нагрузки, обусловленные производственной деятельностью (смещению жизненно важных показателей), осуществляется заключение о состоянии. Для оценки коллективного здоровья (здоровья в однородной группе) применяется непараметрический анализ данных.

Введение

Как показывает мировой и отечественный опыт, накопленный за последние 15 – 20 лет, здравоохранение многими из своих успехов обязано медицинским технологиям. Важнейшими составляющими таких технологий являются технические средства врача, в частности, специализированные информационные системы. Одна из таких технологий, не имеющая ближайших аналогов в СНГ и получившая название "Двойной скрининг – тест" (ДСТ), разработана группой специалистов под руководством автора [1, 5]. ДСТ является технологией охраны здоровья трудоспособного населения. Ее основу составляет анализ реактивности организма, т. е. смещение жизненно важных параметров под действием реальных нагрузок, обусловленных производственной деятельностью пациентов. Необходимая диагностическая информация накапливается в результате двух обследований, разделенных по времени (схема ("До" - "После"). Например, начало работы ("До"), ее окончание ("После"). Это определило название технологии ("двойной").

Применение упомянутой схемы обследования способствует значительному (более, чем в два раза) увеличению объема диагностической информации по сравнению с традиционными схемами. С подобным увеличением связаны два важных обстоятельства: повышенная эффективность технологии и острая потребность в компьютерном анализе данных. Проблеме создания и апробации специализированного интерфейса и посвящена настоящая статья.

ДСТ и оценка коллективного здоровья

ДСТ является технологией, обеспечивающей количественную и качественную оценку состояния организма и степень его адаптации к конкретному виду профессиональной деятельности. С точки зрения принятия административных решений представляют интерес ряд задач, относящихся к проблеме так называемого коллективного здоровья. "Здорова ли группа в целом?", т. е. кто определяет "лицо" группы: здоровые или больные. "Кто из работников по объективным данным обладает наиболее слабым здоровьем и нуждается в адекватной помощи?" и т. п.

Так как тестирование выполняется дважды (схема "До" - "После"), что весьма принципиально, графически результаты представляются в виде 2-х кривых. В качестве типового примера на рис.1 показаны результаты тестирования по группы пациентов по параметру SYS. В роли пациентов выступали операторы многоканального телефона одной из служб экстренной помощи населению.

Как видно, картина отличается значительным разнообразием. Если пациенты №№ 1, 4, 5, 11 практически не отреагировали на рабочие нагрузки, то пациенты №№ 3, 6, 14, 16 показали значительную реактивность. Отмечены переходы из одной нозологической группы в другую. Так из группы "норма" в группу "риск" переместились пациенты №№ 3, 7, 15, а из группы "риск" в группу "больные" - №№ 6, 9, 14.

Совершенно очевидно, что подобные зависимости строятся и для других парамет-

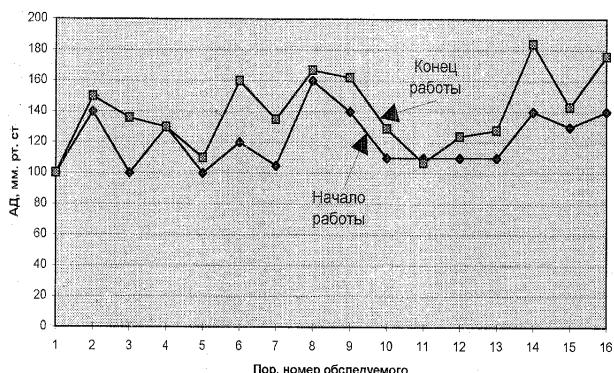


Рис. 1. Пример смещения (реактивности) АД в течение рабочего дня. По горизонтали отложены порядковые номера пациентов, по вертикали - систолическое артериальное давление (SYS)

ров, измеренных в количественной шкале. В итоге формируется задача анализа множества парных выборок. Специфика ДСТ такова, что число пациентов в группе, как правило, невелико и составляет 10 – 40 человек (малые и средние предприятия, распределения по возрасту, полу и другим признакам). Кроме того, неизвестным является закон распределения. Все это требует привлечения непараметрических методов анализа данных [2-4]. В результате проведенных обоснований [1, 5], были выбраны два из них: критерий знаков и критерий Вилкоксона.

В силу ряда причин разработчики интерфейса были ограничены рамками электронного процессора Excel, в котором отсутствует блок непараметрического анализа. Кроме того, как показал опыт эксплуатации и литературные данные [3], формулы описательной статистики, встроенные в базовый пакет, не отвечают требованиям со стороны ДСТ (низкая надежность и скорость вычислений). В итоге была сформулирована и решена серия задач, включающая в себя адаптацию непараметрических критериев к специфике ДСТ, контроль массивов исходных данных на предмет их соответствия требованиям со стороны непараметрического анализа, "прозрачность" математических моделей и простота работы на уровне начинающего пользователя системы Windows.

Адаптация критерия знаков к условиям ДСТ

Пусть X – исходная выборка, полученная при первом (начальном) тестировании, а

Y – повторная выборка, полученная при повторном тестировании пациентов. Тогда моделью ДСТ – исследований становится множество парных выборок вида:

$$X = \{x_1, x_2, \dots, x_n\}, \quad Y = \{y_1, y_2, \dots, y_n\}. \quad (1)$$

Здесь x_i, y_i – результат измерения соответствующего параметра, например SYS (см. рис.1) i –го пациента, где $i = 1, 2, \dots, n$, n – количество пациентов в группе (объем выборки). В итоге, задача оценки смещений медицинских показателей (реактивность) сводится к типовым или близких к типовым задачам описательной и индуктивной статистики [2]. В целях сокращения изложения ограничимся интерпретацией критерия знаков применительно к ДСТ.

Вычислим разности $z_i = x_i - y_i$ и представим их в виде аддитивной модели

$$z_i = \Theta_i + e_i, \quad (2)$$

где Θ_i неслучайная величина, характеризующая эффект смещения центров распределений (1), e_i – стандартная, нормально распределенная случайная величина.

Сформулируем нулевую гипотезу об отсутствии эффекта смещения H_0 и ее альтернативную H_1 – статистически значимое смещение центров распределения X и Y :

$$H_0: \Theta_i = 0, \quad H_1: \Theta_i \neq 0. \quad (3)$$

Если справедлива гипотеза H_0 (см. (3)), то с учетом (2) имеем:

$$P(x_i < y_i) = P(x_i > y_i) = 0,5, \quad (4)$$

где $P(*)$ – вероятность соответствующего события.

Из (4) следует, что число положительных и отрицательных разностей z_i не должно значительно отличаться от $n/2$. Обозначим их соответственно S_+ , S_- . В итоге, моделью ДСТ становится широко известная схема Бернулли для случая $P = 0,5$. Таким образом, интерпретация парных данных (статистический вывод) принимает вид:

$$\text{Max } \{S_+, S_-\} < S_{kp} \Rightarrow H_0, \quad (5)$$

$$\text{Max } \{S_+, S_-\} > S_{kp} \Rightarrow H_1. \quad (6)$$

Критическое значение S_{kp} определяется по таблицам, причем

$$S_{kp} = S_{kp}(\alpha, (n - r)), \quad (7)$$

где α - уровень значимости, r - число нулевых разностей z_i [2].

Суммы S_+ , S_- , r вычисляются по формулам:

$$S_{+(-)} = \sum \begin{cases} 1, & \text{если } z_i > 0, (z_i < 0) \\ 0, & \text{если } z_i < 0, (z_i > 0). \end{cases} \quad (8)$$

$$r = n - S_+ - S_-. \quad (9)$$

Программный интерфейс

Эффективность ДСТ в значительной степени определяется эффективностью и доступностью программного интерфейса врача, что обусловлено массовым характером обработки и большим числом решаемых задач. Структурно интерфейс оформлен в виде блоков и построен в среде Excel, как получившей широкое распространение в медицинских учреждениях. Каждый из блоков включает в себя одну или несколько страниц и выполняет функции хранения исходных данных, предварительной обработки, рабочего поля и критериальных оценок.

Основным экраном является рабочее поле, на котором и осуществляется обработка

и интерпретация данных (рис.2). Рабочее поле состоит из нескольких областей. Область (B9:C60) отведена под исходные данные (X, Y) (на рис.2 показан фрагмент). Несколько выше (A2:C7) выводятся показатели описательной статистики, позволяющие в оперативном порядке оценить крайние значения обрабатываемого параметра (Мин, Макс), центр распределения (М.ожид) и разброс. Если выборки не отвечают требованиям со стороны анализа, в частности, непарные выборки, в блоке (A8:C8) выводятся соответствующие аварийные сообщения.

В поле F25 формулируется статистический вывод. В приведенном примере выборки не имеют существенных отличий (гипотеза H_0 , см. формулы (3)). Значительную часть экрана занимает парный график анализируемых выборок (X, Y), существенно облегчающий оценку состояний как отдельных пациентов, так и группы в целом.

Интерфейс ДСТ отличается исключительной простотой применения. Так, например, работа с экраном "рабочее поле" заключается в выполнении 3 – х процедур, доступных начинающему пользователю:

- Очистка полей исходных данных (нажатие кнопки в ячейке A10).

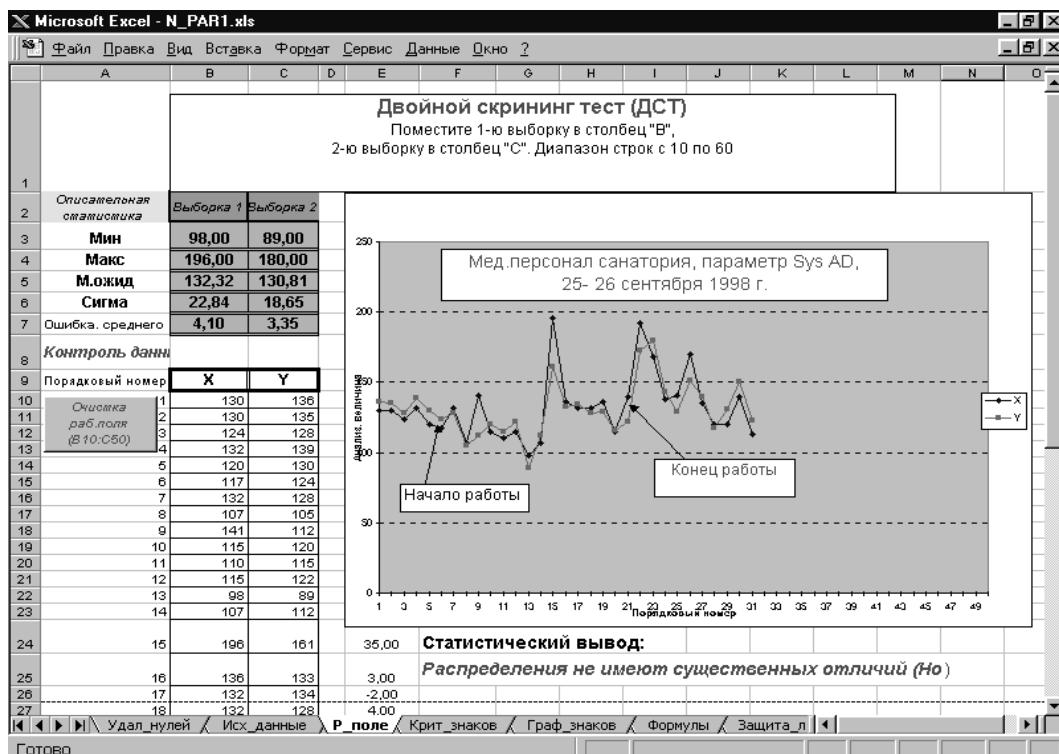


Рис. 2. Типовой экран "Рабочее поле" интерфейса ДСТ

- Ввод исходных данных X, Y с помощью обычных для Excel приемов, известных начинающему пользователю. По окончании ввода выдаются соответствующие результаты обработки, включая, таблицу с показателями описательной статистики, графическую иллюстрацию и статистический вывод.

- Чтение и распечатка результатов анализа.

Второй особенностью является компоновочная схема экрана. В отличие от широко распространенных "бухгалтерских" схем, результаты расчетов и, соответственно, расчетные формулы предшествуют исходным данным. Подобный простейший прием повысил наглядность, существенно сократил число рутинных операций и свел к минимуму вероятность ошибочных действий, возможных при ручном переносе блока формул, что важно при массовом характере обработки массивов переменной длины.

Третья особенность заключается в создании собственного блока непараметрической статистики (формулы 1 -9), поскольку в базовой системе Excel она не предусмотрена. Как показал опыт эксплуатации, а также литературные данные [3], процедуры описательной статистики организованы в Excel нерационально, а, главное, ненадежно. В связи с этим разработан собственный блок описательной статистики.

Четвертая особенность состоит в графической иллюстрации статистического вывода в виде отдельного экрана. Экран содержит графическую интерпретацию таблицы критических значений статистики S_{kp} , а также точку, соответствующую фактическому значению S_+ . Подобное простое решение позволяет пользователю, не являющемуся специалистом в области непараметрической статистики, оценить "надежность" статистических выводов. В частности, если фактическое значение приближается к критическому, следует принять меры к накоплению статистических данных.

Апробация ДСТ- технологии в реальных условиях

ДСТ – технология в варианте, ориентированном на диагностику заболеваний системы кровообращения, и ее интерфейс были апробированы в реальных условиях. В общей

сложности обследовано около тысячи человек, в числе которых, водители служебного и общественного транспорта, врачи санатория Волжский Утес (лечебное учреждения Управления делами Президента РФ), служащие банка, операторы многоканального телефона одной из экстренных служб крупного города, студенты, артисты джаза и т. д. Апробация подтвердила высокую эффективность ДСТ как технологии целом, так и ее программного интерфейса.

В результате компьютерного анализа получены уникальные данные. Так, например, при кажущейся близости условий труда, уровень возмущений и, соответственно, реактивность организма могут существенно отличаться. Примером является система кровообращения артистов одного из профессиональных джаз-оркестров (рис.3). Прежде всего, отметим уникальность самого эксперимента, когда получены количественные данные о состоянии системы кровообращения артистов на сцене. Кроме того, отметим, что артисты разделились на две группы. Если для первой группы характерна незначительная реактивность (пациенты № 1 - 8), то для другой наоборот- она оказалась ярко выраженной (№ 9 - 12). Как было установлено, в первую группу входили артисты оркестра, которые в ходе выступления мало волновались ("обычная работа"), а во вторую - солисты, которые испытывали значительный стресс.

Таким образом, при наблюдении за здоровьем необходимо тщательное обследование специфики рабочего места пациента. Даже в пределах одного помещения (сцена в рассматриваем случае) ситуация может отличаться самым коренным образом

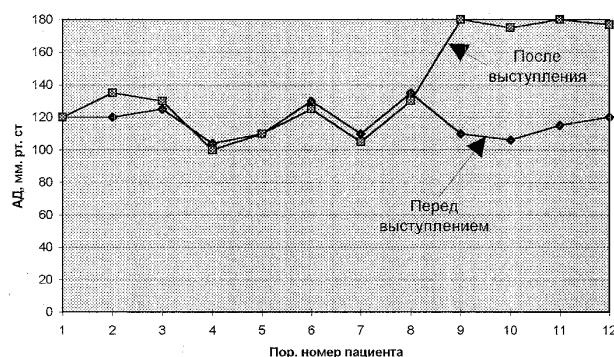


Рис. 3. Реактивность системы кровообращения артистов джаза по параметру SYS

Другим важным результатом является то, что реактивность является ранним предвестником заболевания, когда оно находится в начальной стадии и не ощущается пациентом. Третьим результатом является то, что состояние некоторых групп пациентов может улучшаться в течение рабочего дня в случаях умеренной физической нагрузки и благоприятного морального климата. Примером служат метеозависимые люди.

Выводы

1. Разработан программный интерфейс, обеспечивающий информационное сопровождение технологии охраны здоровья, известной как "Двойной скрининг – тест". Технология предусматривающей количественную и качественную оценку состояний и степень адаптации организма к реальным нагрузкам, обусловленным производственной деятельностью людей (нагрузочный тест по схеме "До – После"). Оценка состояний осуществляется как по величине ряда жизненно важных параметров, так и их смещениям, вызванным реальными нагрузками (реактивность организма).

2. Интерфейс реализован в среде Excel, что потребовало создания серии специализированных блоков, ориентированных на массовую обработку парных данных. В их числе, блок подготовки исходных данных (проверка их корректности с точки зрения математических моделей), блок визуализации исходных данных и результатов расчетов, блок описательной статистики, блок непараметрического анализа. При разработке бло-

ков найдены оригинальные и простые решения, направленные на повышение достоверности расчетов, простоты и удобства пользователей, не являющихся специалистами в области информатики.

3. Технология и ее интерфейс апробированы в реальных условиях (водители, операторы многоканальных телефонов одной из экстренной службы, врачи санатория Волжский Утес, артисты, студенты и т. д). В общей сложности обследовано около тысячи пациентов. Подтверждена высокая эффективность как самой технологии охраны здоровья, так и ее программного интерфейса.

СПИСОК ЛИТЕРАТУРЫ

1. Акулов В.А. Двойной скрининг тест: концепция, модель, предварительные результаты // Труды IV Международной конференции "Новые информационные технологии в медицине и экологии IT + Se' 98". Ялта - Гурзуф, 26 мая - 4 июня 1998.
2. Runyon R. Nonparametric statistics a Contemporary Approach. Massachusetts: Addison-Wesley Publishing Company, 1977.
3. Кулаичев А.П. Методы и средства анализа данных в среде Windows Stadia 6.0. М: Информатика и компьютеры, 1996.
4. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. М: Финансы и статистика, 1995.
5. Akulov V. DST - Technology and its usage while evaluating the collective health // Proceedings of the Foton / Bion International Conference. Samara 25 - 30 June. 2000.

NONPARAMETRIC ANALYSIS IN THE FIELD OF HEALTH PROTECTION

© 2001 V.A. Akulov

Samara State Technical University

A technology of health protection directed to early diagnosing, prophylaxis and treating illnesses caused by patients' professional activity was developed in Samara. The technology which was called "Double Screening Test" (DST) suggests the performance of two screening tests in the conditions very close to real ones. The first test is performed at the beginning of the working day (or before a predicted stress), the other one is done at the end of the day (after the stress). On comparing the data one finds patients with higher reactivity to real loads. They are given medical aid if necessary.