

РАСЧЕТ, ПРОЕКТИРОВАНИЕ И ПРОИЗВОДСТВО ОПТИЧЕСКИХ СИСТЕМ

УДК 535.317.2

УНИВЕРСАЛЬНЫЕ МЕТОДЫ РЕШЕНИЯ МОДЕЛЬНЫХ УРАВНЕНИЙ ПРИ ГАБАРИТНОМ РАСЧЕТЕ ОПТИЧЕСКИХ СИСТЕМ

© 2007 г. А. В. Иванов, канд. техн. наук

Санкт-Петербургский государственный университет информационных технологий,
механики и оптики, Санкт-Петербург

E-mail: ivanov@mail.ifmo.ru

Предложены высокоэффективные универсальные методы параметрического синтеза центрированных оптических систем в гауссовой области, использующие алгебраический, численный и смешанный подходы к решению систем полилинейных модельных уравнений. Основу алгебраического метода составляет полиномиальное упрощение путем применения результатов многочленов или построения базиса Гребнера идеала, сгенерированного многочленами. Численный алгоритм базируется на сформулированных автором необходимых условиях существования решений. Приводятся примеры работы предложенных методов параметрического синтеза.

Коды OCIS: 080.0080.

Поступила в редакцию 26.12.2006.

Как впервые было показано в работе [1], габаритный расчет центрированных оптических систем в подавляющем большинстве случаев можно осуществлять с помощью обобщенной модели, построенной на основе полилинейных функций. Это свойство позволяет использовать для решения модельных уравнений универсальные методы, отличающиеся высокой эффективностью.

С целью рассмотрения указанных методов представим каждое из уравнений, входящих в систему условий параметрического синтеза, в виде равенства нулю полилинейной функции:

$$f(\mathbf{X}) = f(x_1, x_2, \dots, x_n) = c_0 + \sum_{i=1}^n c_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij} x_i x_j + \\ + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n c_{ijk} x_i x_j x_k + \dots + c_{123\dots n} x_1 x_2 \dots x_n = 0, \quad (1)$$

где $f(\mathbf{X})$ – полилинейная функция n переменных, $\mathbf{X}$ – вектор варьируемых параметров $x_1, x_2, \dots, x_n$; $c_0, c_1, \dots, c_n, c_{12}, \dots, c_{12\dots n}$ – постоянные коэффициенты. Функция $f(\mathbf{X})$ в простейшем случае может представлять собой отклонение координаты нулевого луча от заданного значения, в более сложных ситуациях – линейную комбинацию различных координат. В качестве переменных $\mathbf{X}$ могут выступать оптические силы поверхностей (компонентов), осевые расстояния между ними, а также отрезки, определяющие положение предмета и входного зрачка.

Используя функции вида (1), можно сформулировать задачу габаритного расчета оптической системы в виде проблемы отыскания вектора n переменных, одновременно удовлетворяющего n условиям

$$f_p(x_1, x_2, \dots, x_n) = 0, \quad p = 1, 2, \dots, n. \quad (2)$$

Учитывая выводы работы [1], посредством (2) можно описать большинство известных требований к параксиальным и габаритным характеристикам оптической системы или ее части, в том числе требования к оптической силе, увеличению, сумме Петцваля, положению изображения и выходного зрачка, продольным и поперечным габаритам, а также, при определенных условиях, к расстоянию от предмета до изображения. Уравнения (2) пригодны для создания систем как с постоянными, так и с переменными характеристиками.

В основу процесса поиска решений может быть положен алгебраический или численный подход.

При использовании алгебраического подхода в системе полилинейных уравнений (2) проводится последовательное исключение переменных, в результате чего проблема поиска корней сводится к решению одного уравнения с одной неизвестной и нахождению оставшихся переменных по способу обратной подстановки.

На первый взгляд, изложенная концепция, знакомая еще из курса элементарной алгебры, кажется весьма понятной и простой. Однако, если степень

многочленов в системе (2) превышает 2, каждый этап ее реализации становится крайне трудно осуществимым “вручную” и требует применения современных математических методов. С этой целью автором предлагается использовать ряд прогрессивных достижений и результатов компьютерной алгебры, ранее не применявшихся в практике оптических расчетов.

В частности, для исключения переменных в (2) предлагается воспользоваться построением результатов полилинейных функций $f_p(\mathbf{X})$ или базиса Гребнера идеала, сгенерированного этими многочленами [2, 3].

1. Метод результатов

При работе по данному методу многочлены, входящие в систему уравнений, на каждом этапе преобразований полагают зависящими только от одной переменной. Для каждой пары таких многочленов находится результат (например, определитель матрицы Сильвестра, матрицы Безу или на основе более современных алгоритмов [2]). Приравняв его нулю, можно получить из двух уравнений одно, в котором выбранная переменная будет отсутствовать.

Описанный метод достаточно прост, позволяет осуществить исключение переменных за известное конечное число этапов. Однако метод обладает рядом особенностей вычислений (в частности, может давать “паразитные” решения) и приводит к весьма громоздким выражениям.

2. Метод построения и использования базиса Гребнера

Этот современный метод относится к недавним достижениям компьютерной алгебры и получил широкое распространение в профессиональных пакетах программ символьных вычислений “Maple”, “Mathematica”, “Reduce” и др. Принцип его работы несколько напоминает решение системы линейных уравнений способом исключения по Гауссу. В результате процесса преобразования исходных полиномов (генераторов идеала) на основе алгоритма Бухбергера [2] получается подмножество многочленов, имеющих такие же корни, что и исходные, но обладающих необходимыми свойствами для простейшего решения системы уравнений (переменные последовательно исключены). Анализ показывает, что благодаря полилинейности многочленов построение базиса Гребнера проходит весьма эффективно, не приводя к промежуточным громоздким выражениям. Однако время работы алгоритма заранее определить довольно сложно.

Для решения проблемы нахождения действительных корней многочлена одной переменной на всей числовой оси или на заданном промежутке автором предлагается применять сочетание алгебраического и численного методов. Дело в том, что использование для этой цели чисто численных алгоритмов, не требующих отделения корней (типа Берстоу, Лобачевского, Мюллера [4, 5]), не всегда оправдано из-за плохой обусловленности задачи [6], а применение исключительно алгебраических подходов возможно лишь для многочлена не выше четвертой степени.

Суть предлагаемого процесса решения заключается в том, что отделение корней проводится на основе алгебраических свойств многочлена, а уточнение корней – численным путем. В частности, для изолирования корней можно использовать теорему Штурма, теорему Бюдана–Фурье, правило Декарта [3, 7] и другие способы. Уточнение решений целесообразно осуществлять прямыми итерационными методами: либо оптимальными по скорости сходимости регуляризованными методами Ньютона или секущих, либо медленнее сходящимися, но более простыми методами дихотомии (бисекции), хорд и др. [7, 8].

Последний этап реализации алгебраического подхода к решению (2), заключающийся в поиске всех неизвестных “в обратном ходе”, оказывается в значительной степени методически связанным с двумя предыдущими. При использовании базиса Гребнера уравнения для всех переменных, кроме одной, оказываются линейными. При применении результатов приходится вычислять корни многочленов подобно тому, как это делалось на втором этапе поиска решений.

Рассмотрим на примере работу алгебраических методов. Пусть требуется определить оптические силы трех компонентов оптической системы с переменными характеристиками типа “фотообъектив” (предмет находится в бесконечности, изображение – на конечном расстоянии). Средний компонент является подвижным и обеспечивает трехкратный перепад заднего фокусного расстояния. Известные параметры и характеристики объектива приведены в таблице.

Исходные данные для габаритного синтеза трехкомпонентного фотообъектива

№ состояния	Расстояния между главными плоскостями компонентов		f' , мм	s'_F , мм
	d_1 , мм	d_2 , мм		
0	25	25		65,0
1	45	5	33,33	
2	5	45	100,0	

В таблице f' – заднее фокусное расстояние, s'_F – расстояние от задней главной плоскости третьего компонента до заднего фокуса объектива, d_1 , d_2 – расстояния между главными плоскостями компонентов.

С целью нахождения неизвестных составим систему модельных уравнений (заметим, что эту систему можно построить в полуавтоматическом режиме на основе метода неопределенных коэффициентов [1])

$$\begin{cases} f_1(x_1, x_2, x_3) = -0,01 + x_1 + x_2 + x_3 - 5x_1x_2 - 50x_1x_3 - 45x_2x_3 + 225x_1x_2x_3 = 0 \\ f_2(x_1, x_2, x_3) = -0,03 + x_1 + x_2 + x_3 - 45x_1x_2 - 50x_1x_3 - 5x_2x_3 + 225x_1x_2x_3 = 0 \\ f_3(x_1, x_2, x_3) = 1 - 115x_1 - 90x_2 - 65x_3 + 2250x_1x_2 + 3250x_1x_3 + 1625x_2x_3 - 40625x_1x_2x_3 = 0, \end{cases} \quad (3)$$

где через x_1, x_2, x_3 обозначены оптические силы компонентов.

Вначале для решения системы (3) применим метод результатов. Исключим переменную x_1 , приравняв нулю результаты (Res) многочленов первого и второго, а также первого и третьего уравнений. С этой целью обозначим

$$\begin{aligned} a_0 &= 225x_2x_3 - 5x_2 - 50x_3 + 1, \\ a_1 &= -45x_2x_3 + x_2 + x_3 - 0,01, \\ b_0 &= 225x_2x_3 - 45x_2 - 50x_3 + 1, \end{aligned}$$

$$\begin{aligned} b_1 &= -5x_2x_3 + x_2 + x_3 - 0,03, \\ c_0 &= -40625x_2x_3 + 2250x_2 + 3250x_3 - 115, \\ c_1 &= 1625x_2x_3 - 90x_2 - 65x_3 + 1. \end{aligned}$$

Тогда

$$\begin{cases} f_4(x_2, x_3) = \text{Res}_{x_1}(f_1, f_2) = \begin{vmatrix} a_0 & a_1 \\ b_0 & b_1 \end{vmatrix} = a_0b_1 - b_0a_1 = 0, \\ f_5(x_2, x_3) = \text{Res}_{x_1}(f_1, f_3) = \begin{vmatrix} a_0 & a_1 \\ c_0 & c_1 \end{vmatrix} = a_0c_1 - c_0a_1 = 0 \end{cases}$$

или

$$\begin{cases} f_4(x_2, x_3) = -0,02 - 0,3x_2 + 40x_2^2 + x_3 + 75,5x_2x_3 - 2000x_2^2x_3 - 2000x_2x_3^2 + 9000x_2^2x_3^2 = 0, \\ f_5(x_2, x_3) = -0,15 + 42,5x_2 - 1800x_2^2 + 32,5x_3 - 4406,25x_2x_3 + 113500x_2^2x_3 + 91000x_2x_3^2 - 1,4625 \times 10^6 x_2^2x_3^2 = 0. \end{cases} \quad (4)$$

Далее исключим переменную x_2 . Вводя обозначения

$$\begin{aligned} r_0 &= 9000x_3^2 - 2000x_3 + 40, \\ r_1 &= -2000x_3^2 + 75,5x_3 - 0,3, \quad r_2 = x_3 - 0,02, \\ s_0 &= -1,4625 \times 10^6 x_3^2 + 113500x_3 - 1800, \\ s_1 &= 91000x_3^2 - 4406,25x_3 + 42,50, \\ s_2 &= 32,5x_3 - 0,15, \end{aligned}$$

получим уравнение относительно x_3 :

$$f_6(x_3) = \text{Res}_{x_2}(f_4, f_5) = \begin{vmatrix} r_0 & r_1 & r_2 & 0 \\ 0 & r_0 & r_1 & r_2 \\ s_0 & s_1 & s_2 & 0 \\ 0 & s_0 & s_1 & s_2 \end{vmatrix} = 0$$

или

$$\begin{aligned} 725,8 - 218595x_3 + 2,52431 \times 10^7 x_3^2 - 1,58695 \times 10^9 x_3^3 + \\ + 6,67893 \times 10^{10} x_3^4 - 2,05756 \times 10^{12} x_3^5 + \\ + 4,00509 \times 10^{13} x_3^6 - 3,28536 \times 10^{14} x_3^7 = 0. \end{aligned} \quad (5)$$

Данное уравнение имеет следующие действительные корни:

$$\begin{aligned} 1) x_3 &= 0,00778565, \\ 2) x_3 &= 0,018623, \\ 3) x_3 &= 0,0211111, \\ 4) x_3 &= 0,0222222, \\ 5) x_3 &= 0,0481782. \end{aligned}$$

Подстановка этих значений в (3) и (4) приводит к трем действительным решениям исходной системы уравнений:

$$\begin{aligned} 1) \begin{cases} x_1 = -0,0184886 \\ x_2 = 0,0190301 \\ x_3 = 0,00778565, \end{cases} \quad 2) \begin{cases} x_1 = 0,0254363 \\ x_2 = -0,0733858 \\ x_3 = 0,018623, \end{cases} \quad 3) \begin{cases} x_1 = 0,0577836 \\ x_2 = -0,0520539 \\ x_3 = 0,0481782. \end{cases} \end{aligned} \quad (6)$$

Применим теперь другой алгебраический метод к системе (3). Построим базис Гребнера идеала, сгенерированного конечным числом полиномов f_1, f_2, f_3 относительно лексикографического порядка. Используя алгоритм Бухбергера, получим

$$\begin{cases} g_1(x_3) = -382 + 79765x_3 - 5,10383 \times 10^6 x_3^2 + \\ + 1,9378 \times 10^8 x_3^3 - 6,3739 \times 10^9 x_3^4 + 8,112 \times 10^{10} x_3^5, \\ g_2(x_2, x_3) = 5,03142 \times 10^9 - 5,37998 \times 10^{10} x_2 - \\ - 6,08977 \times 10^{11} x_3 + 1,79597 \times 10^{13} x_3^2 - \\ - 8,62357 \times 10^{14} x_3^3 + 1,41538 \times 10^{16} x_3^4, \\ g_3(x_1, x_3) = 20243 + 368775x_1 - 2,05946 \times 10^6 x_3 + \\ + 6,17689 \times 10^7 x_3^2 - 2,72051 \times 10^9 x_3^3 + 4,056 \times 10^{10} x_3^4, \end{cases} \quad (7)$$

где подмножество многочленов g_1, g_2, g_3 образует базис Гребнера. Заметим, что полином g_1 зависит только от одной переменной x_3 , а многочлены g_1, g_2 содержат изолированные переменные x_1, x_2 (обе они могут быть выражены через x_3 , если приравнять многочлены нулю). Это говорит о принципиальной разрешимости системы уравнений (3). Изолированные переменные входят в базисные многочлены с относительно низкими степенями по сравнению с методом результатов за счет возможностей алгоритма.

Решая уравнение $g_1(x_3) = 0$, выделим по методу Декарта три сегмента с изолированными действительными корнями $[0; 0,015625]$, $[0,015625; 0,03125]$, $[0,03125; 0,0625]$, а затем проведем их итерационное уточнение:

- 1) $x_3 = 0,00778565$,
- 2) $x_3 = 0,018623$,
- 3) $x_3 = 0,0481782$.

Полученные корни с точностью до 6 знаков мантиссы совпадают с тремя найденными из (5). Но в отличие от (5) в методе базиса Гребнера нет “паразитных” промежуточных решений $x_3 = 0,0211111$, $x_3 = 0,0222222$, которые не удовлетворяют исходной системе уравнений (что устанавливается при вычислении x_1, x_2 посредством (4) и (3)). Поэтому и с этой позиции метод базиса Гребнера оказывается более эффективным, чем метод результатов.

Рассмотренный алгебраический подход к решению систем уравнений вида (1) прельщает возможностью использования универсального математического и программного обеспечения, позволяет найти все действительные корни. Тем не менее он является достаточно сложным и громоздким. При этом часто в результате получаются решения, не пригодные с точки зрения их реализации в оптической системе, – ведь процесс нахождения корней в интересую-

щих пользователя промежутках контролируется только по завершении всех преобразований.

В этой связи автором был разработан численный метод решения систем полилинейных уравнений, который, обладая глобальными свойствами нахождения полного множества действительных решений, практически мгновенно может дать точную оценку возможности существования или отсутствия корней в области изменения переменных, заданной в виде гиперпараллелепипеда.

Переходя к описанию метода, общие принципы которого были первоначально изложены в работе [1], необходимо вновь вернуться к свойствам полилинейной функции $f(\mathbf{X})$ в (1). Заметим, что матрица Гессе (матрица вторых частных производных) $f(\mathbf{X})$ имеет главную диагональ, состоящую из нулей. Следовательно, след (шпур) матрицы равен нулю и она обладает в общем случае собственными числами разных знаков [7]. Отсюда можно сделать вывод, что $f(\mathbf{X})$ не имеет локальных экстремумов, а достигает своего наибольшего и наименьшего значений на границе области изменения переменных.

Пусть теперь некоторый параметр x_k варьируется в пределах сегмента $[x_k^{\min}, x_k^{\max}]$, а остальные переменные будут “заморожены”. В этой ситуации функция $f(\mathbf{X})$ достигает своего максимального и минимального значений тогда, когда величина x_k находится на одной из границ области изменения ($x_k = x_k^{\min}$ или $x_k = x_k^{\max}$). Если далее взять гиперпараллелепипед $\mathbf{H}$, который описывается системой неравенств

$$\begin{cases} x_1^{\min} \leq x_1 \leq x_1^{\max}, \\ x_2^{\min} \leq x_2 \leq x_2^{\max}, \\ \dots \\ x_n^{\min} \leq x_n \leq x_n^{\max}, \end{cases} \quad (8)$$

то понятно, что $f(\mathbf{X})$ может принять свое наибольшее $\max_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X})$ и наименьшее $\min_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X})$ значения только в его “углу”.

С другой стороны, ввиду свойств непрерывных функций [4, 7] внутри гиперпараллелепипеда может находиться по крайней мере один корень многочлена $f(\mathbf{X})$ тогда и только тогда, когда условия $\max_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X}) \geq 0$ и $\min_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X}) \leq 0$ выполняются одновременно, т. е.

$$\max_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X}) \min_{\mathbf{X} \in \mathbf{H}} f(\mathbf{X}) \leq 0. \quad (9)$$

Поэтому необходимым условием существования хотя бы одного из корней полилинейной функции $f(\mathbf{X})$ внутри или на гранях гиперпараллелепипеда $\mathbf{H}$ является смена знака или нулевое значение этой функции в его угловых точках.

Отсюда вытекает правило: система уравнений типа (2) может иметь решение только тогда, когда все функции $f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_n(\mathbf{X})$ поменяют свой знак или примут нулевое значение в углах гиперпараллелепипеда (8):

$$\max_{\mathbf{X} \in \mathbf{H}} f_p(\mathbf{X}) \min_{\mathbf{X} \in \mathbf{H}} f_p(\mathbf{X}) \leq 0, \quad p = 1, 2, \dots, n. \quad (10)$$

Установленное правило имеет двоякое применение. Во-первых, с его помощью можно, предварительно вычислив значения многочленов $f_p(\mathbf{X})$ при различных комбинациях наименьших и наибольших значений переменных, сразу ответить на вопрос о наличии решений системы полилинейных уравнений. Во-вторых, правило позволяет постепенно сужать область локализации решений (вплоть до требуемой точности), если проводить дробление гиперпараллелепипеда на части (скажем, делением пополам) и отбрасывать “пустые” блоки. Однако оказывается, что такая локализация без применения специальных приемов является весьма трудоемкой и требует большого объема памяти в компьютере для хранения промежуточных результатов.

Дело в том, что свойство (9) по сути только свидетельствует о прохождении через гиперпараллелепипед поверхности нулевого уровня функции $f(\mathbf{X})$. Но даже если через область $\mathbf{H}$ пройдут поверхности нулевого уровня всех функций $f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_n(\mathbf{X})$, то это не означает, что они будут иметь общее пересечение. Иными словами, правила вида (10) формулируют необходимые, но далеко не достаточные условия существования решения полилинейных уравнений. Это приводит к тому, что многие “пустые” блоки приходится запоминать и считать перспективными для поиска корней, так как через них проходят поверхности нулевого уровня всех многочленов.

Чтобы ускорить отбрасывание “пустых” блоков в процессе разделения гиперпараллелепипеда, можно дополнить (10) целым рядом необходимых условий, “сгустив” их, если так можно выразиться, почти до уровня достаточности. Для вывода этих условий отметим, что если в пределах области $\mathbf{H}$ найдется точка, в которой все функции $f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_n(\mathbf{X})$ одновременно обращаются в нуль, то через эту область обязательно пройдет поверхность нулевого уровня любой линейной комбинации названных функций. Кроме того, эта линейная комбинация будет обладать свойством полилинейности. Поэтому необходимые условия наличия решений в системе уравнений (2) можно записать в таком виде:

$$\max_{\mathbf{X} \in \mathbf{H}} \sum_{k=1}^n z(i_k) f_k(\mathbf{X}) \min_{\mathbf{X} \in \mathbf{H}} \sum_{k=1}^n z(i_k) f_k(\mathbf{X}) \leq 0, \quad (11)$$

$$\forall i_1, \forall i_2, \dots, \forall i_n \in \{1, 2, \dots, m\},$$

где z – некоторая функция, а m – задаваемое количество ее значений, используемых в качестве коэффициентов при многочленах. Можно положить, например, $z(i) = 1 + (-1)^{i-1}$, $i = 1, 2, \dots, 15$. При начальной нормировке (выравнивании значений) полиномов $f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_n(\mathbf{X})$ в одном из углов гиперпараллелепипеда условия (11) позволяют в значительной степени ужесточить отбор перспективных блоков по сравнению с (10).

Иллюстрации успешной работы предложенного способа отыскания корней полилинейных уравнений приведены в работе [1].

Завершая обсуждение методов решения (2), следует указать на то обстоятельство, что положительные стороны алгебраического и численного подходов взаимно дополняют друг друга. Поэтому целесообразно применять их в комплексе. Численный метод при этом будет давать моментальную информацию об отсутствии решений или областях возможного существования решений. Этой информации вполне достаточно для предварительного анализа приемлемости синтезируемой конструкции. Алгебраический подход позволит найти решения с нужной степенью точности, которая необходима для завершения процесса синтеза.

ЛИТЕРАТУРА

1. *Ivanov A. V.* Generalized method for first-order lens layout // Proc. SPIE. 1999. V. 3780. P. 199–206.
2. *Дэвенпорт Дж., Сирэ И., Турнье Э.* Компьютерная алгебра. М.: Мир, 1991. 352 с.
3. *Курош А. Г.* Курс высшей алгебры. СПб.: Лань, 2006. 432 с.
4. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. М.: Наука, 1984. 832 с.
5. Математическое обеспечение ЕС ЭВМ. Вып. 3 / Ин-т мат. АН БССР. Минск, 1973. 240 с.
6. *Мак-Кракен Д., Дорн У.* Численные методы и программирование на ФОРТРАНЕ. М.: Мир, 1977. 584 с.
7. *Бронштейн И. Н., Семендяев К. А.* Справочник по математике для инженеров и учащихся втузов. М.: Наука, 1986. 544 с.
8. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. М.: Мир, 1980. 280 с.