

Запорізький національний технічний університет



Радіоелектроніка Інформатика Управління

1(19)'2008

Науковий журнал

Видається двічі на рік

Зареєстрований **29 січня 2003 року**
Державним комітетом інформаційної політики,
телебачення та радіомовлення України.

Свідоцтво – серія **КВ № 6904**

Засновник і видавник – *Запорізький національний технічний університет*

Запоріжжя, ЗНТУ
2008

Передплатний індекс – 22914

Журнал «Радіоелектроніка. Інформатика. Управління» видається з 1999 року.

Журнал входить до переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата технічних наук та фізико-математичних наук за спеціальністю «Радіофізика». Статті, що публікуються в журналі, реферуються в базах даних та РЖ ВІНТІ (Росія) і «Джерело» (Україна)

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор – д.т.н. Потапенко Є. М.

Заст. головного редактора – к.т.н. Дубровін В. І.

Члени редколегії :

д.ф.-м.н. Ахметшин А. М.
д.т.н. Волков О. В.
д.ф.-м.н. Горбань О. М.
д.ф.-м.н. Горр Г. В.
д.т.н. Гостєв В. І.
д.ф.-м.н. Дробахин О. О.
д.т.н. Карпуков Л. М.
д.ф.-м.н. Ковальов О. М.

д.ф.-м.н. Корніч Г. В.
д.ф.-м.н. Онуфрієнко В. М.
д.т.н. Павлов О. А.
д.т.н. Піза Д. М.
д.ф.-м.н. Погосов В. В.
д.т.н. Толок В. О.
д.т.н. Труфанов І. Д.
д.ф.-м.н. Чумаченко В. П.

Редакційно-видавнича рада:

Беліков С. Б., Внуков Ю. М., Піза Д. М., Яримбаш С. Т., Потапенко Є. М., Дубровін В. І., Петруша Ю. П., Флора В. Д., Соболев Ю. О., Дедков М. В., Рибіна Ю. А., Табунщик Г. В., М'ясников М. Л., Богданова Л. Ф., Кобзар Н. М., Федоренко Е. С.

Рекомендовано до видання вченою радою Запорізького національного технічного університету, протокол № 10 від 26.05.2008 р.

Рукописи проходять незалежне рецензування з залученням провідних фахівців, за результатами якого редакційна колегія приймає рішення про опублікування

Журнал зверстаний редакційно-видавничим відділом
Запорізького національного технічного університету

Адреса редакції: 69063, м. Запоріжжя, вул. Жуковського, 64, ЗНТУ,
редакція журналу «РІУ»

Тел: (061) 769-82-96 – редакційно-видавничий відділ

Факс: (061) 764-21-41

E-mail: rvv@zntu.edu.ua

TQM@zntu.edu.ua

Web-сайт журналу розташований в Internet за адресою: <http://www.zntu.edu.ua/base/news2005/ric.htm>

© 2007 Запорізький національний технічний університет

ЗМІСТ

РАДІОФІЗИКА	5
<i>А. В. Крапивной, В. П. Чумаченко, Я. В. Чумаченко</i> ОБ УТОЧНЕНИИ ОДНОЙ ОЦЕНКИ СКАЛЯРНОГО ВОЛНОВОГО ПОЛЯ РАССЕЯННОГО ИМПЕДАНСНОЙ СФЕРОЙ..... 5	
РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ	9
<i>О. С. Антропов, С. М. Вовк, В. Ф. Борулько</i> МЕТОД МИНИМУМА ПРОСТРАНСТВЕННОЙ ПРОТЯЖЕННОСТИ ДЛЯ ОПРЕДЕЛЕНИЯ ДИАГРАММЫ НАПРАВЛЕННОСТИ АНТЕННЫ ПО ИЗМЕРЕНИЯМ БЛИЗНЕГО ПОЛЯ	9
<i>Л. М. Карпуков, Р. Ю. Корольков</i> МЕТОД РАСЧЕТА МИКРОВОЛНОВЫХ СТУПЕНЧАТЫХ ФИЛЬТРОВ НА СОРАЗМЕРНЫХ ОТРЕЗКАХ ЛИНИЙ	13
<i>Ю. А. Крисан, А. А. Крисан</i> ИССЛЕДОВАНИЕ ЭЛЕКТРОННОГО КЛЮЧА С ГАЛЬВАНИЧЕСКОЙ РАЗВЯЗКОЙ	16
<i>В. М. Кришук, Г. М. Шило, Б. А. Артюшенко</i> РОЗПАРАЛЕЛЕННЯ ГЕНЕТИЧНОГО АЛГОРИТМУ ПАРАМЕТРИЧНОГО СИНТЕЗУ АНТЕННОЇ ГРАТКИ НА КОМП'ЮТЕРНІЙ МЕРЕЖІ.....	20
<i>Д. М. Пиза, А. А. Степаненко</i> СЕГМЕНТАЦИЯ УЛЬТРАЗВУКОВЫХ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ МЕТОДА НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ.....	24
<i>И. М. Прохорец</i> МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДЛЯ АНАЛИЗА СТАТИЧЕСКИХ ХАРАКТЕРИСТИК МНОГОЭЛЕМЕНТНЫХ ПОЛУПРОВОДНИКОВЫХ ДЕТЕКТОРОВ	27
<i>С. С. Самойлик, В. П. Бондарев</i> ДОБРОТНОСТЬ ПРЯМОУГОЛЬНОГО РЕЗОНАТОРА С ПЕРЕМЕННЫМ ЧИСЛОМ ДИЭЛЕКТРИЧЕСКИХ НЕОДНОРОДНОСТЕЙ...	32
<i>Г. М. Шило, Д. А. Коваленко, М. П. Гапоненко</i> ПРИЗНАЧЕННЯ НОРМАЛЬНИХ ДОПУСКІВ МЕТОДОМ ВІДОБРАЖЕНЬ	37
МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ	42
<i>Н. І. Біла, Н. О. Нечипоренко, Л. О. Бондаренко</i> ПРО РЕГУЛЯРИЗАЦИЮ ЧИСЕЛЬНОГО ДИФЕРЕНЦИУВАННЯ	42
<i>А. Н. Довбня, С. Г. Удовенко, А. А. Шамраев</i> ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ МОДЕЛИ МАГНИТНОГО СПЕКТРОМЕТРА ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ	45
<i>В. И. Дубровин, О. И. Юськив</i> МОДЕЛИ И МЕТОДЫ ОПТИМИЗАЦИИ ВЫБОРА ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ	49
<i>А. Б. Кунгурцев, С. Л. Зиноватная</i> ИМИТАЦИОННАЯ МОДЕЛЬ ДЛЯ ИССЛЕДОВАНИЯ ЭФФЕКТИВНОСТИ ДЕНОРМАЛИЗАЦИИ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ В ИНФОРМАЦИОННОЙ СИСТЕМЕ ..60	
<i>В. Й. Лобов, С. А. Рубан</i> РОЗРОБКА ПРИНЦИПІВ КЕРУВАННЯ ТЕМПЕРАТУРНИМ РЕЖИМОМ ПРОЦЕСУ ВИПАЛЮВАННЯ КОТУНІВ З ВИКОРИСТАННЯМ ПРОГНОЗУЮЧИХ ANFIS-МОДЕЛЕЙ.....	69
<i>В. І. Мороз</i> АНАЛІЗ РАЦІОНАЛЬНОГО ПОРЯДКУ АПРОКСИМАЦІЙНОГО ПОЛІНОМА ДЛЯ ВІДНОВЛЕННЯ ІНФОРМАЦІЇ ЗА ЇЇ ДИСКРЕТНИМИ ВІДЛІКАМИ	74
<i>А. Г. Овский, В. А. Толок</i> ИСПОЛЬЗОВАНИЕ СИСТЕМЫ КОМПЬЮТЕРНОЙ МАТЕМАТИКИ MAPLE ДЛЯ ДОКАЗАТЕЛЬСТВА ЗАКОНА ОРТОГОНАЛЬНОСТИ МАТРИЦ ПРЯМОГО И ОБРАТНОГО ПРЕОБРАЗОВАНИЙ, СОСТАВЛЕННЫХ ВЛАСОВЫМ В. З.	78

Е. С. Решевская, С. Н. Гребенюк
МОДЕЛИРОВАНИЕ НАПРЯЖЕННО-
ДЕФОРМИРОВАННОГО СОСТОЯНИЯ
ТРЕХМЕРНЫХ КОНСТРУКЦИЙ НА ОСНОВЕ
МЕТОДА КОНЕЧНЫХ ЭЛЕМЕНТОВ
С ИНТЕРПОЛИРУЮЩИМ ПОЛИНОМОМ
ЭРМИТА.....85

НЕЙРОИНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ 92

Е. В. Бодянский, Е. А. Викторов
ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ
ПОСЛЕДОВАТЕЛЬНОСТЕЙ
НА ОСНОВЕ КАСКАДНОЙ ОРТОГОНАЛЬНОЙ
НЕЙРОННОЙ СЕТИ.....92

Т. В. Киприч, В. И. Дубровин
ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МЕТОДИКИ
САМООГРАНАДУЮЩИХСЯ КАРТ КОХОНЕНА
ДЛЯ ДИАГНОСТИРОВАНИЯ СОСТОЯНИЙ
ПРОЦЕССОВ И СИСТЕМ98

Б. Б. Нестеренко, М. А. Новотарський
МОДЕЛЮВАННЯ ДИСКРЕТНИХ КЛІТКОВИХ
НЕЙРОННИХ МЕРЕЖ 103

И. В. Новосельцев, Н. Г. Аксак
ПРИМЕНЕНИЕ МНОГОПРОЦЕССОРНЫХ
СИСТЕМ ДЛЯ РЕАЛИЗАЦИИ МНОГОСЛОЙНОГО
ПЕРСЕПТРОНА 108

С. О. Субботін, А. О. Олійник
СТРУКТУРНИЙ СИНТЕЗ НЕЙРОМОДЕЛЕЙ НА
ОСНОВІ ПОЛІМОДАЛЬНОГО ЕВОЛЮЦІЙНОГО
ПОШУКУ 111

М. Ю. Юрич, А. Н. Щербаков, А. Ф. Камышанов
АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОЙ
СИСТЕМЫ НЕЙРОУПРАВЛЕНИЯ
ЛЕТАТЕЛЬНЫМ ОБЪЕКТОМ
С ЭВОЛЮЦИОННЫМ ПРОГНОЗОМ
ПРЕДПОЧТИТЕЛЬНОГО ПОВЕДЕНИЯ 118

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ..... 122

С. А. Зайцев, С. К. Корниенко
ОПЕРАЦИОННАЯ СИСТЕМА ДЛЯ
ВСТРАИВАЕМЫХ УСТРОЙСТВ 122

А. В. Неласая, Г. Л. Козина, Н. А. Молдовян
ПРОТОКОЛЫ КОЛЛЕКТИВНОЙ ЦИФРОВОЙ
ПОДПИСИ НА ЭЛЛИПТИЧЕСКИХ
И ГИПЕРЭЛЛИПТИЧЕСКИХ КРИВЫХ 127

В. И. Слепцов, Л. М. Карпуков
ПРАВОВАЯ ПОДГОТОВКА КАДРОВ
В ОБЛАСТИ ИНФОРМАЦИОННОЙ
БЕЗОПАСНОСТИ 133

КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ 138

А. Е. Казурова, Н. В. Ивочка
ОБЕСПЕЧЕНИЕ ЗАДАННОГО СООТНОШЕНИЯ
ПОКАЗАТЕЛЕЙ КАЧЕСТВА
ЭЛЕКТРОТЕХНИЧЕСКОЙ СИСТЕМЫ
С НЕПОЛНЫМ ИЗМЕРЕНИЕМ ВЕКТОРА
СОСТОЯНИЯ 138

Е. В. Душинова, В. И. Левыкина, Е. В. Васильева
ПРОСТОЙ МЕТОД ОЦЕНКИ
ПОТОКОСЦЕПЛЕНИЙ АСИНХРОННОГО
ДВИГАТЕЛЯ 143

А. В. Соломаха
ФОРМИРОВАНИЕ СИСТЕМЫ ВЕКТОРНОГО
УПРАВЛЕНИЯ АСИНХРОННЫМ
ПРИВОДОМ С КОМПЕНСАЦИЕЙ
НЕОПРЕДЕЛЕННОСТЕЙ 148

АВТОРЫ НОМЕРА 154

РАДІОФІЗИКА

РАДИОФИЗИКА

RADIOPHYSICS

УДК 537.874.6

А. В. Крапивной, В. П. Чумаченко, Я. В. Чумаченко

ОБ УТОЧНЕНИИ ОДНОЙ ОЦЕНКИ СКАЛЯРНОГО ВОЛНОВОГО ПОЛЯ РАССЕЯННОГО ИМПЕДАНСНОЙ СФЕРОЙ

На примере рассеяния плоской скалярной волны импедансной сферой исследуются возможности совершенствования нового метода, близкого к методу физической оптики и учитывающего (в исходном варианте) кривизну граничной поверхности в плоскости падения волны. Показано, что возможный в рамках этого подхода учет кривизны в плоскости перпендикулярной плоскости падения улучшает оценку рассеянного поля, однако, это улучшение не столь существенно, как в случае перехода от физической оптики к исходному варианту метода с учетом кривизны только в одной плоскости.

ВВЕДЕНИЕ

Недавно в работах [1, 2] был предложен новый подход к оцениванию коротковолнового поля, рассеянного выпуклым двумерным объектом, который, сохраняя простоту метода физической оптики, позволяет заметно улучшить оценки для не зеркальных направлений. В работах [3, 4] область его приложений была расширена на решение трехмерных задач рассеяния плоской волны сферическими препятствиями, а в [5] метод был обобщен на случай произвольного выпуклого импедансного тела. Подобно приближению физической оптики метод дает аналитическое выражение для поля у рассеивающей поверхности, однако, при этом учитывает ее кривизну в плоскости падения волны и не предполагает обращения в нуль поля на теневой части тела. В работе [3], в частном случае мягкой сферы, было показано, что в рамках развиваемого подхода может быть рассмотрена кривизна поверхности также и в плоскости перпендикулярной плоскости падения. В настоящей

работе возможности улучшения оценок путем учета кривизны поверхности в двух упомянутых плоскостях изучаются на примере решения задачи рассеяния для сферы с произвольным поверхностным импедансом.

РАСЧЕТНЫЕ ФОРМУЛЫ

Пусть (x, y, z) и (ρ, θ, φ) – прямоугольная и сферическая системы координат с полярной осью вдоль оси z , а S – сфера радиуса R , обладающая в общем случае поверхностным импедансом Z (рис. 1). Среда, в которой расположено препятствие, характеризуется волновым сопротивлением Z_0 и скоростью распространения волн c . Поле будем характеризовать переменной p и, опущенным в последующих соотношениях, временным множителем $e^{i\omega t}$, где ω – круговая частота колебаний. Если ось z направлена навстречу падающей плоской волне, то возбуждающее поле можно записать в виде

$$p_i = e^{ikz}, \quad (1)$$

где $k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$, а λ – длина волны в среде. Полное поле

$$p = p_i + p_s \quad (2)$$

удовлетворяет уравнению Гельмгольца вне сферы и импедансному граничному условию

$$\frac{\partial p}{\partial n} - ik\chi p = 0 \quad (3)$$

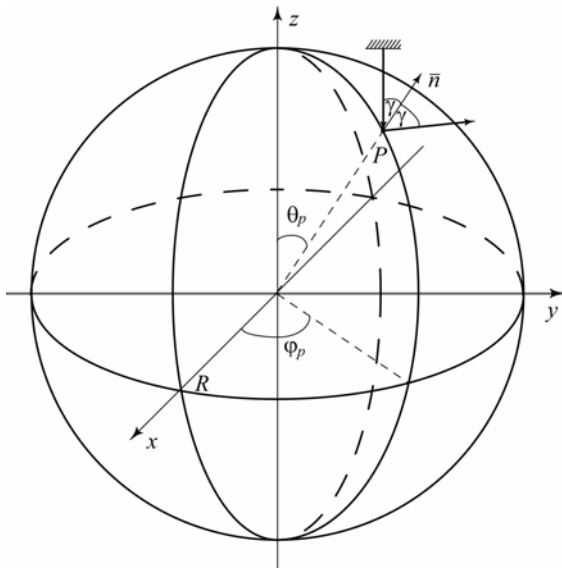


Рисунок 1 – Геометрия задачи

на ее поверхности. Здесь $\chi = \frac{Z_0}{Z}$, а направление нормали \vec{n} указано на рис. 1. Значение $\chi = 0$ отвечает случаю жесткой сферы.

Для мягкой сферы $\chi = \infty$ и

$$p = 0 \tag{4}$$

на границе S .

Рассеянное поле p_s вне сферы может быть выражено через значения p и $\partial p / \partial n$ на ее поверхности с помощью интеграла Киргофа. В дальней зоне это представление преобразуется к виду

$$p_s(M) \approx f(\varphi_M, \theta_M) \frac{e^{-ik|\vec{r}|}}{|\vec{r}|}, \tag{5}$$

где \vec{r} – радиус-вектор точки M , а амплитудная функция рассеяния $f(\varphi_M, \theta_M)$ определяется [3, 4] формулами

$$f(\varphi_M, \theta_M) = -\frac{ik}{4\pi} \iint_S [\chi(P) - \cos\alpha] p(P) e^{ikR \cos\alpha} dS_P \tag{6}$$

(для импедансной поверхности) и

$$f(\varphi_M, \theta_M) = \frac{ikZ_0}{4\pi} \iint_S v_n(P) e^{ikR \cos\alpha} dS_P \tag{7}$$

(для случая $\chi = \infty$). Здесь α – угол между \vec{r} и \overline{OP} (O – начало координат), а

$$v_n(P) = \frac{i}{kZ_0} \left. \frac{\partial p}{\partial \rho} \right|_P. \tag{8}$$

Чтобы улучшить оценки, получаемые для $p(P)$ и $v_n(P)$ методом физической оптики, приблизим сначала рассеивающую поверхность в окрестности точки P поверхностью кругового цилиндра, который перпендикулярен плоскости падения волны и ось которого проходит через центр сферы. Кривизна поверхности в направлении, перпендикулярном плоскости падения, при этом очевидно не учитывается. Следуя работам [3, 4], мы получим в случае импедансной поверхности следующее выражение для $p(P)$:

$$p = e^{ikR \cos\gamma} \left[1 + ik \frac{\chi - \cos\gamma}{X' - i\chi k} \right], \tag{9}$$

где γ – угол падения в точке P , а

$$X' = k \left[\frac{\mu}{kR} - \frac{H_{\mu+1}^{(2)}(kR)}{H_{\mu}^{(2)}(kR)} \right]. \tag{10}$$

Здесь $H_{\mu}^{(2)}(kR)$ – функция Ганкеля и

$$\mu = kR \sin\gamma. \tag{11}$$

Оценка для $v_n(P)$ в случае мягкой сферы имеет вид

$$v_n = \frac{ie^{ikR \cos\gamma}}{kZ_0} [ik \cos\gamma - X']. \tag{12}$$

Чтобы учесть кривизну поверхности в направлении, перпендикулярном плоскости падения, введем в окрестности точки P локальные сферические координаты (ρ, ϑ, ϕ) с началом отсчета в центре сферы и полярной осью вдоль оси упомянутого выше цилиндра. При этом точке P отвечают значения $\vartheta = \frac{\pi}{2}$ и $\phi = 0$, причем при $\vartheta = \frac{\pi}{2}$ угол ϕ определяется разностью $\phi = \theta - \theta_P$. Будем считать далее, что в малой окрестности точки P , как и в случае плоской поверхности, функциональные зависимости падающей и отраженной волн совпадают вдоль поверхности с точностью до постоянного множителя. Разделяя переменные в уравнении Гельмгольца, получим [3] при этом условии

$$p_s = C \frac{h_v^{(2)}(k\rho)}{h_v^{(2)}(kR)} e^{-i\mu\phi}, \tag{13}$$

где μ определяется формулой (11), а $h_v^{(2)}(k\rho)$ является сферической функцией Бесселя индекса

$$v = \sqrt{\mu^2 + 0,25} - 0,5. \tag{14}$$

Определив постоянную C с помощью граничного условия (3) или (4), сможем получить оценки для $p(P)$

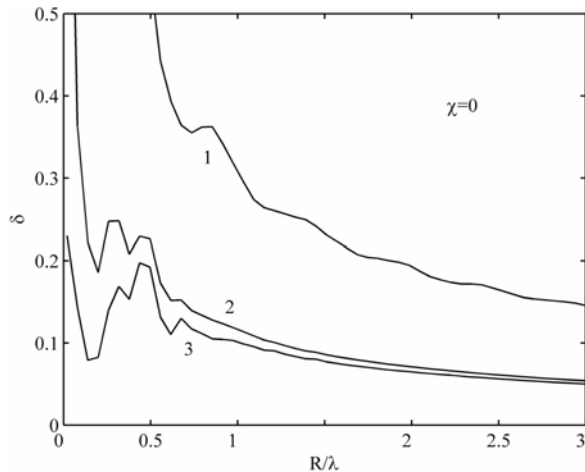


Рисунок 2 – Относительная среднеквадратическая ошибка для случая жесткой сферы:

1 – метод физической оптики, 2 – формула (10), 3 – формула (15)

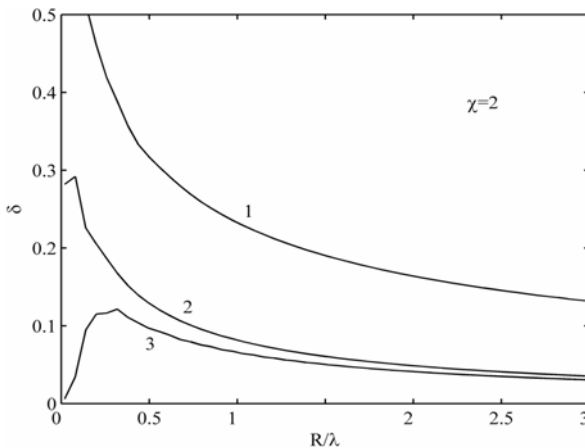


Рисунок 3 – Относительная среднеквадратическая ошибка для случая импедансной сферы:

1 – метод физической оптики, 2 – формула (10), 3 – формула (15)

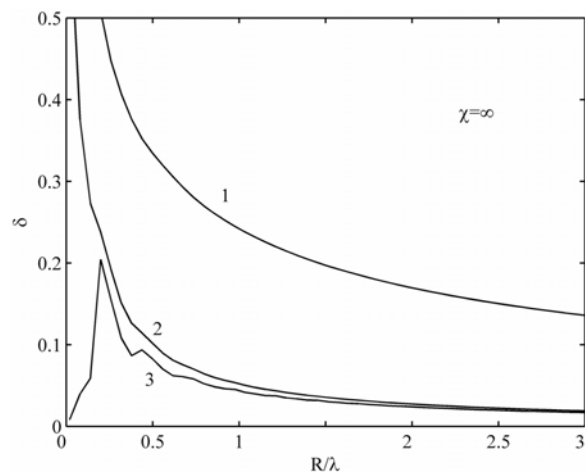


Рисунок 4 – Относительная среднеквадратическая ошибка для случая мягкой сферы:

1 – метод физической оптики, 2 – формула (10), 3 – формула (15)

и $v_n(P)$. Они снова описываются формулами (9) и (12), если только понимать под X' величину

$$X' = k \left[\frac{v}{kR} - \frac{h_{\nu+1}^{(2)}(kR)}{h_{\nu}^{(2)}(kR)} \right]. \quad (15)$$

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

На рис. 2, 3 и 4 представлены частотные зависимости относительной среднеквадратической ошибки

$$\delta = \frac{\sqrt{\int_0^{\pi} |f - f_{ex}|^2 d\theta_M}}{\sqrt{\int_0^{\pi} |f_{ex}|^2 d\theta_M}}, \quad (16)$$

которая возникает при оценивании функции рассеяния f с помощью выписанных выше приближений в случаях жесткой, импедансной и мягкой поверхностей, соответственно. Под точным решением f_{ex} мы понимаем решение, полученное в виде ряда методом разделения переменных [6]. Видно, что переход к приближению, основанному на учете кривизны рассеивающей поверхности в плоскости падения волны, ведет к значительному уменьшению ошибки по сравнению с методом физической оптики. Учет кривизны в двух плоскостях приводит к дальнейшему ее сокращению, однако, это сокращение в высокочастотной области уже не столь существенно.

Заметим, что из представленных данных также следует, что последнее приближение заметно лучше других приближений при $\frac{R}{\lambda} \rightarrow 0$ и обеспечивает приемлемую оценку функции f для шаров малых радиусов. На рис. 5 и 6 приведены угловые зависимости рассеянного

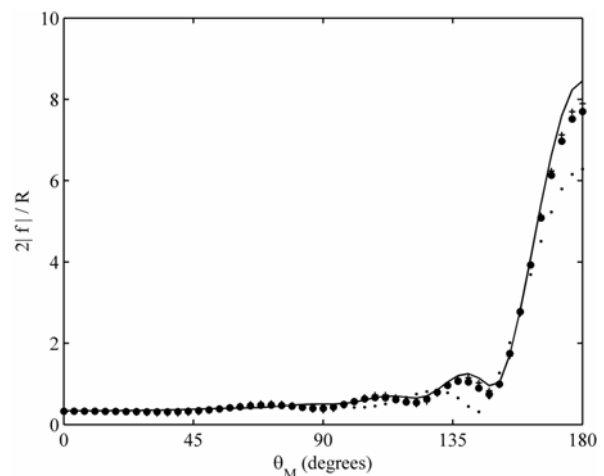


Рисунок 5 – Угловая зависимость рассеянного поля в дальней зоне при $R/\lambda = 1$ и $\chi = 2$:

сплошная линия – точно, пунктир – метод физической оптики, кружки – формула (10), крестики – формула (15)

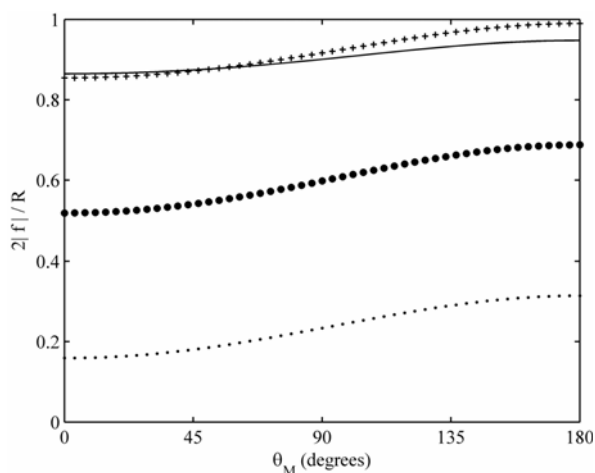


Рисунок 6 – Угловая зависимость рассеянного поля в дальней зоне при $R/\lambda = 0,05$ и $\chi = 2$:

сплошная линия – точно, пунктир – метод физической оптики, кружки – формула (10), крестики – формула (15)

поля в дальней зоне, которые характерны для больших и малых значений $\frac{R}{\lambda}$.

ЗАКЛЮЧЕНИЕ

На примере решения задачи рассеяния скалярной волны импедансной сферой рассмотрены возможности дальнейшего совершенствования предложенного недавно приближенного подхода, который близок к методу физической оптики и учитывает кривизну рассеивающей поверхности в плоскости падения волны. Показано, что возможное в рамках метода рассмотрение кривизны поверхности в плоскости, перпендикулярной плоскости падения, улучшает оценку рассеянного поля в высокочастотной части диапазона волн, однако, это улучшение не так существенно, как в случае перехода от физической оптики к варианту метода с учетом кривизны только в одной плоскости. Установлено также, что учет кривизны в двух плоскостях делает возможным использование развитого подхода и для оценок

характеристик рассеяния шаров малого, в сравнении с длиной волны, радиуса.

ПЕРЕЧЕНЬ ССЫЛОК

1. Chumachenko V. P. On the estimation of scattering from convex conducting cylinders // Microwave and Optical Technology Letters. – Vol. 45. – May 2005. – P. 191–194.
2. Крапивну А. В., Чумаченко В. П. On the estimation of scattering from convex impedance cylinders // Proceedings of the 11th Int. Conf. on Mathematical Methods in Electromagnetic Theory. – Kharkiv, Ukraine. – June 26–29, 2006. – P. 248–250.
3. Крапивной А. В., Чумаченко В. П. Об оценивании волнового поля рассеянного мягкой сферой // Радиоэлектроника. Информатика. Управление. – 2006. – № 2. – С. 5–8.
4. Крапивной А. В., Чумаченко В. П., Чумаченко Я. В. О рассеянии плоской скалярной волны импедансной сферой // Радиоэлектроника. Информатика. Управление. – 2007. – № 1. – С. 9–14.
5. Chumachenko V. P., Крапивну А. В., Чумаченко Я. В. An estimation method for acoustic scattering from convex bodies // Journal of Sound and Vibration. – 2008. – Vol. 310, Mar. – P. 845–854.
6. Морс Ф. М., Фешбах Г. Методы теоретической физики. – Том 1. – М.: Изд-во иностр. лит., 1958. – 930 с.

Надійшла 24.03.2008

На прикладі розсіювання плоскої скалярної хвилі імпедансною сферою досліджуються можливості вдосконалення нового методу, який близький до методу фізичної оптики і враховує (в початковому варіанті) кривизну граничної поверхні в площині падіння хвилі. Показано, що можливе в рамках цього підходу урахування кривизни в площині, перпендикулярній площині падіння, покращує оцінку розсіяного поля, але це покращення не настільки суттєве, як у випадку переходу від фізичної оптики до початкового варіанту методу з урахуванням кривини тільки в одній площині.

The problem of scalar plane-wave scattering off a reactively loaded sphere is examined. The features of a new approach, which is close to the physical optics approximation and considers (in its initial variant) the curvature of the target at the plane of incidence, are studied. It is shown, that possible consideration of the curvature at the plane perpendicular to the plane of incidence improves the estimate of the scattered field, but this improvement is not so essential as in the case of change from physical optics to the initial variant of the method considering the curvature at only one plane.

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

РАДИОЕЛЕКТРОНИКА И ТЕЛЕКОММУНИКАЦИИ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 621.391.8

О. С. Антропов, С. М. Вовк, В. Ф. Борулько

МЕТОД МИНИМУМА ПРОСТРАНСТВЕННОЙ ПРОТЯЖЕННОСТИ ДЛЯ ОПРЕДЕЛЕНИЯ ДИАГРАММЫ НАПРАВЛЕННОСТИ АНТЕННЫ ПО ИЗМЕРЕНИЯМ БЛИЖНЕГО ПОЛЯ

Рассмотрена задача определения диаграммы направленности антенны путем восстановления распределения токов в плоскости раскрытия антенны по данным ближнего поля. Для решения обратной задачи восстановления распределения токов применен метод минимума пространственной протяженности. Диаграмма направленности антенны получена путем преобразования Фурье поля в раскрытие. Представлены результаты численного моделирования для случая одномерной апертуры.

ВВЕДЕНИЕ

Одной из важных и актуальных задач теории антенн является задача определения диаграммы направленности антенны (ДНА) по результатам измерения ее ближнего поля. На практике этот вид измерений наиболее распространен, поскольку позволяет минимизировать искажающее влияние окружающей среды и дает возможность оценить ДНА посредством преобразования Фурье. Однако при практическом использовании результатов измерений ближнего поля возникают трудности, обусловленные конечным размером области измерений. Так, при использовании стандарт-

ной процедуры пересчета ближнего поля в дальнее поле, часть ближнего поля, которая не охвачена областью измерений, обычно полагается равной нулю [1]. Это приводит к неточному восстановлению распределения поля в дальней зоне вне некоторого диапазона углов, определяемого размерами антенны и области измерений, из-за появления в расчетах так называемой ошибки усечения. Для уменьшения ошибки усечения необходимо либо увеличить размер области измерений, либо ввести ограничения на физический размер исследуемых антенн. Это обуславливает необходимость поиска методов, позволяющих обойти указанные выше трудности.

Для решения указанной задачи было предложено несколько методов. Один из методов [2] использует априорную информацию о геометрии антенны и данные измерений ближнего поля для оценивания эквивалентного распределения магнитных токов на плоскости раскрытия антенны с последующим получением дальнего поля. Другой метод [3] использует дополнительную математическую обработку результатов измерений

ближнего поля на основе геометрических характеристик антенны. Идея этого метода основана на экстраполяции измеренных данных ближнего поля за пределы области измерений. Это позволяет получить эквивалентную область измерений, превышающую физически доступную. Так как данный метод основан на построении квадратичного функционала для модуля невязки измеренного поля и вычисляемого поля, то полученное решение будет неустойчивым и для получения устойчивого решения используется техника регуляризации Тихонова [4]. После восстановления поля в ближней зоне, оба метода используют прямой пересчет распределения поля в ближней зоне в распределение поля в дальней зоне с помощью стандартных алгоритмов преобразований, основанных на алгоритмах быстрого преобразования Фурье. Существенным ограничением указанных и других аналогичных методов является необходимость располагать точными априорными сведениями о геометрических характеристиках исследуемой антенны. Если же такие априорные сведения являются неточными или отсутствуют, то применение указанных методов становится проблематичным.

В [5] мы предложили новый подход к обработке результатов измерений ближнего поля. Идея этого подхода заключалась в замене стандартной техники регуляризации на новую технику, которая основана на критерии минимума пространственной протяженности решения. Применение этого подхода позволяет достичь существенного улучшения результатов восстановления поля точечных источников, расположенных в раскрыве антенны, при отсутствии априорных сведений об их местоположении. В данной работе предлагается использовать этот подход для решения задачи определения ДНА по результатам измерений ближнего поля антенны. Ключевым моментом предлагаемого подхода является предположение о финитности пространственного распределения токов в раскрыве антенны, представленного набором точечных источников при отсутствии априорных сведений об их местоположении.

1 ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Представление распределения поля в дальней зоне посредством преобразования Фурье от пространственного распределения токов в раскрыве антенны [6] является вычислительно эффективным и не содержит в себе погрешности, обусловленные ошибкой усечения. Следовательно, успешность решения задачи определения ДНА определяется результатом решения обратной задачи восстановления токов в раскрыве антенны. Для демонстрации общего подхода к решению данной обратной задачи используем простую модель, которая описывает измеренные данные $b(\xi)$ ближнего поля ан-

тенны и распределение токов $a(r)$ в области S апертуры антенны в рамках скалярной задачи уравнением:

$$b(\xi) = \int_{S_0} a(r) \frac{e^{-ik|\xi-r|}}{|\xi-r|} dr + N(\xi); \quad \xi \in M, \quad (1)$$

где M – область измерений, $N(\xi)$ – аддитивный пространственный шум, S_0 – неизвестная область распределения токов в области S , т. е. $S_0 \subset S$. Последнее означает, что если $r \in S_0$, то $a(r) \neq 0$, но если $r \notin S_0$, то $a(r) = 0$. Феноменологическая постановка обратной задачи восстановления токов в области апертуры антенны формулируется следующим образом: найти распределение токов $a(r)$ по известным значениям поля $b(\xi)$ в условиях, когда эти известные данные искажены шумом и область S_0 априорно неизвестна.

Отметим, что из (1) следует принципиальная невозможность точного определения $a(r)$ в условиях ограниченной области измерений M и наличия шума. Действительно, источники тока всегда ограничены в пространстве и создают поле во всем пространстве. Но если только часть этого пространства доступна для непосредственных измерений, то, во-первых, происходит потеря части полезной информации и, во-вторых, доступная полезная информация является искаженной из-за наличия измерительного шума.

Ясно, что потерянная информация может быть восстановлена только искусственно на основе априорной информации, объем которой должен быть адекватен объему потерянной информации. Априорная информация, которая явно присутствует в модели (1), заключается в том, что искомое решение $a(r)$ должно описываться финитной функцией, так как область S_0 априорно ограничена. Теоретически такая априорная информация является достаточной для восстановления функции $a(r)$ по любому известному сегменту данных $b(\xi)$. Однако наличие шума и отсутствие знания или неточное знание области S_0 не позволяют эффективно применять соответствующие известные схемы восстановления. В этих условиях мы предлагаем решать уравнение (1) таким образом, чтобы пространственная протяженность области S_0 , где сосредоточены источники, была минимальной.

Рассмотрим случай, когда источники тока расположены вдоль оси OX , а измерения ближнего поля выполняются вдоль линии, отстоящей от источников на расстоянии z , причем линия измерений параллельна оси OX и лежит в плоскости $y = 0$. Тогда можно переписать (1) в виде:

$$b(\xi) = \int_{S_0} a(x) \frac{e^{-ik\sqrt{(\xi-x)^2+z^2}}}{\sqrt{(\xi-x)^2+z^2}} dx + N(\xi). \quad (2)$$

Поскольку задача решения уравнения (2) относительно $a(x)$ является некорректной обратной задачей, то для получения устойчивого решения нужна процедура регуляризации. Вместо стандартной квадратичной регуляризации, основанной на добавлении члена в виде энергии решения, используем нелинейную регуляризацию, добавляя неэнергетический член в форме, описывающей пространственную протяженность решения [5]. Это приводит к формулировке следующей задачи минимизации:

$$\int_M \left| b(\xi) - \int_S a(x) \frac{e^{-ik\sqrt{(\xi-x)^2+z^2}}}{\sqrt{(\xi-x)^2+z^2}} dx \right|^2 d\xi + \gamma^2 \int_S \psi[a(x); \alpha, \beta, \dots] dx \rightarrow \min_{a(x)} \quad (3)$$

где γ^2 – параметр регуляризации. Первый член в (3) является мерой ошибки полученного решения по области измерений; второй член определяет пространственную протяженность решения в качестве регуляризирующего члена. Отметим, что функция ψ , представленная в (3), может иметь различные варианты [7]; здесь используем следующую форму представления:

$$\psi[s(x); \alpha, \beta] = [s(x)^2/A^2 + \alpha^2]^\beta - \alpha^{2\beta}; \quad 0 < \beta \leq 0,5, \quad (4)$$

где параметры α, β позволяют управлять поведением функции ψ . Результаты численного моделирования показали [5, 7, 8], что наиболее удачным является такой выбор: $\alpha \approx \sigma$; $\beta = 1/16$, где σ – стандартное отклонение измерительного шума.

Решение задачи (3) будем называть решением по методу минимума пространственной протяженности (РММПП). Из (3) следует, что если для всех x выполняется: $\alpha^2 \gg |a(x)|^2$, то РММПП стремится к решению, получаемому при обычной регуляризации Тихонова с использованием энергетического члена. Кроме того, увеличение γ^2 до очень больших значений приближает РММПП к нулевому решению, а использование малых γ^2 приводит к решению без регуляризации. Таким образом, можно утверждать, что для каждого значения дисперсии шума существует некоторое наилучшее значение параметра регуляризации γ^2 , при использовании которого абсолютная ошибка решения будет минимальной.

После решения задачи (3) с целью восстановления функции $a(x)$, описывающей распределение токов антенны, соответствующая ДНА определяется через преобразование Фурье:

$$F(u) = \frac{1}{2\pi} \int_S a(x) e^{iux} dx, \quad (5)$$

где $u = k \sin \theta = 2(\pi/\lambda) \sin \theta$; θ – угол между направлением излучения и внешней нормалью к апертуре антенны. После этого результат можно перевести в пространство углов с учетом только тех пространственных частот, которые соответствуют диапазону $[-\pi/2, \pi/2]$.

2 ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ

Цель проведения численного моделирования заключалась в сопоставлении результатов определения ДНА по измерениям ближнего поля при применении двух методов регуляризации решения обратной задачи: квадратичной регуляризации Тихонова [3, 4] и регуляризации по методу минимума пространственной протяженности. Для моделирования исходное распределение токов было задано в виде набора точечных источников, расположенных на оси OX в заданных узлах дискретной сетки, которая имела равномерный шаг Δx . Предполагалось, что измерение комплексной амплитуды ближнего поля было проведено вдоль отрезка прямой линии с шагом $\Delta \xi$, отстоящей от линии источников на расстояние z . Учитывая (2), для численного моделирования была использована следующая система уравнений:

$$b_p = \sum_{q=1}^Q a_q \frac{e^{-i2\pi \sqrt{[q(\Delta x/\lambda) - p(\Delta \xi/\lambda)]^2 + (z/\lambda)^2}}}{\lambda \sqrt{[q(\Delta x/\lambda) - p(\Delta \xi/\lambda)]^2 + (z/\lambda)^2}} + N_p; \quad p = \overline{1, P}, \quad (6)$$

где b_p – известные данные, представленные комплексными значениями, $P = 100$; $\Delta x/\lambda = 0,1$; $\Delta \xi/\lambda = 0,1$; $z/\lambda = 2$; $Q = 100$, причем амплитуды точечных источников $a_q = 1$ для $q = 40, 50, 60, 70, 80$, и $a_q = 0$ для других q , N_p – комплексный аддитивный гауссовский шум.

Полученные результаты численного решения обратной задачи (6) с использованием квадратичной регуляризации Тихонова и регуляризации по методу минимума пространственной протяженности приведены на рис. 1. На рис. 2 приведены графики соответствующих ДНА, которые были рассчитаны посредством дискретного преобразования Фурье и приведены к диапазону углов $[-\pi/2, \pi/2]$. При численном моделировании, выполненном для 100 различных реализаций шума при отношении сигнал/шум +20 дБ, средний квадрат ошибки расчета ДНА по методу стандартной регуляризации Тихонова относительно истинной ДНА составил величину 6,87 %, тогда как для регуляризации по методу

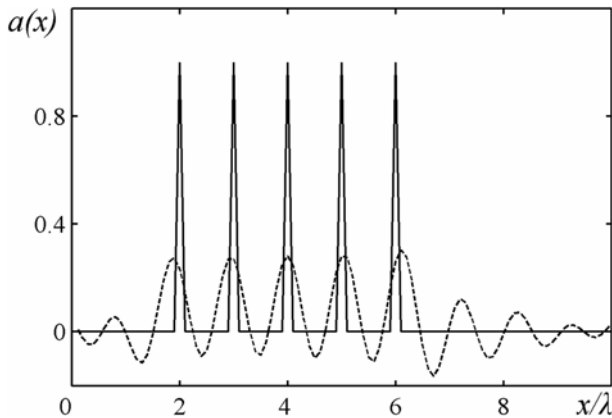


Рисунок 1 – Пространственное распределение токов:

пунктирная линия – стандартная регуляризация;
непрерывная линия – предлагаемая регуляризация

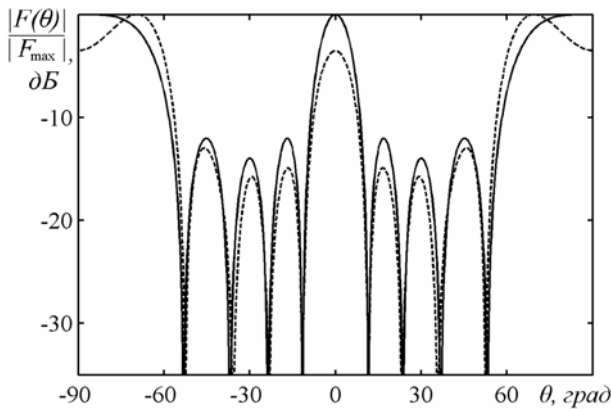


Рисунок 2 – Восстановленные ДНА:

пунктирная линия – стандартная регуляризация;
непрерывная линия – предлагаемая регуляризация

минимума пространственной протяженности эта величина составила менее 0,27 %, причем при применении предлагаемого метода наблюдалось визуальное совпадение полученной ДНА с истинной ДНА.

Таким образом, можно заключить, что предложенный метод позволяет уменьшить ошибку расчета ДНА по сравнению со стандартным подходом.

ВЫВОДЫ

Решение задачи определения ДНА целесообразно выполнять путем восстановления распределения токов в плоскости раскрыва с последующим преобразованием полученного распределения токов в значения ДНА. Данный подход обеспечивает возможность получения оценок ДНА в произвольном диапазоне углов, так как

при этом не возникает ошибка усечения данных из-за ограниченности области измерений ближнего поля. Применение метода минимума пространственной протяженности обеспечивает возможность решения обратной задачи при отсутствии априорных сведений о геометрии излучающей системы, причем применение функционала минимума пространственной протяженности в качестве регуляризирующего члена имеет преимущество по сравнению с традиционным подходом на основе квадратичной регуляризации. Результаты численного моделирования для случая одномерной апертуры подтвердили работоспособность и эффективность предложенного подхода.

ПЕРЕЧЕНЬ ССЫЛОК

1. Yaghjian A. D. An overview of near-field antenna measurements // IEEE Transactions Antennas Propagation. – 1986. – Vol. AP-34. – Pp. 30–45.
2. Petre P., Sarkar T. K. Near-field to far-field transformation using an equivalent magnetic current approach // IEEE Transactions Antennas Propagation. – 1992. – Vol. 40, No. 11. – Pp. 1348–1356.
3. Bolomey J.-C., Bucci O. M., Casavola L., D'Elia G., Migliore M. D., Ziyat A. Reduction of truncation error in near-field measurements of antennas of base-station mobile communication systems // IEEE Transactions Antennas Propagation. – 2004. – Vol. AP-52, No. 2. – Pp. 593–602.
4. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. – М.: Наука, 1986. – 286 с.
5. Вовк С. М., Борулько В. Ф. Регуляризація даних антенних вимірювань на основі обмеження просторової протяжності струмів. // Вісник Дніпропетровського національного університету. – 2006. – № 2/3. – С. 9–11.
6. Минкович Б. М., Яковлев В. П. Теория синтеза антенн. – М: Сов. Радио, 1969. – 269 с.
7. Вовк С. М., Борулько В. Ф. Метод минимума длительности для восстановления финитных сигналов // Известия вузов. Радиоэлектроника. – 1991. – Т. 34, № 9. – С. 66–69.
8. Вовк С. М., Борулько В. Ф. Відновлення рефлектограм за допомогою екстраполяції широкосмугових НВЧ-вимірювань методом мінімуму тривалості // Радіоелектроніка. Інформатика. Управління. – 2005. – № 1. – С. 5–8.

Надійшла 5.03.2008
Після доробки 9.04.2008

Розглянута задача визначення діаграми спрямованості антени шляхом відновлення розподілу струмів у площині розкриття антени за даними ближнього поля. Для розв'язку оберненої задачі відновлення розподілу струмів використано метод мінімуму просторової протяжності. Діаграма спрямованості антени отримана за допомогою перетворення Фур'є поля у площині розкриття. Представлені результати чисельного моделювання для випадку одновимірної апертури.

The problem of far-field pattern determination from near-field data measurements through restoration of currents distribution on the aperture plane is considered. Extension minimum method is applied for currents estimation inverse problem. Fourier transform applied afterwards to determine far-field pattern. Simulation results are presented for one-dimensional aperture case.

Л. М. Карпуков, Р. Ю. Корольков

МЕТОД РАСЧЕТА МИКРОВОЛНОВЫХ СТУПЕНЧАТЫХ ФИЛЬТРОВ НА СОРАЗМЕРНЫХ ОТРЕЗКАХ ЛИНИЙ

Предложен матричный метод расчета ступенчатых микроволновых фильтров. Метод основан на использовании матриц специального вида, составляемых из полиномов комплексной частоты, соответствующих знаменателю и числителю параметров рассеяния фильтра. Представлены результаты расчета фильтров с максимально-плоской и равноволновой характеристиками затухания.

ВВЕДЕНИЕ

Конструкции на основе линии передачи со ступенчатым изменением волнового сопротивления находят широкое применение в микроволновых устройствах, обеспечивающих функции фильтров нижних частот, полосно-пропускающих и полосно-заграждающих фильтров, фильтров гармоник, трансформаторов сопротивлений, согласующих цепей, корректоров частотных характеристик [1–3].

При расчете микроволновых устройств на ступенчатых линиях используются процедуры синтеза, в которых в качестве схемы-прототипа берется ступенчатый фильтр нижних частот с нормированной к единице частотой среза [1–3]. Процедуры синтеза включают в себя аппроксимацию частотной зависимости затухания, вносимого фильтром, функциями, удовлетворяющими условию физической реализуемости, с последующим составлением схемы фильтра, реализующей эти функции. Составление схемы микроволнового фильтра и определение значений параметров ее элементов производится по матрице рассеяния фильтра, полученной на этапе аппроксимации. Для определения значений параметров элементов фильтров со структурой в виде каскадного соединения четырехполюсников в настоящее время используется аппарат классических матриц передач [1–5]. Применение в расчетах фильтров классических матриц передачи (А-матриц) или волновых матриц передачи (Т-матриц) усложняет вычисления, требуя перевода матриц одного типа в другой. В данной статье задача расчета параметров четырехполюсников, составляющих структуру фильтра, решается с помощью специального вида матриц. Элементами этих матриц являются полиномы числителя и знаменателя передаточных функций параметров рассеяния. Простота и эффективность применения этого вида матриц проиллюстрирована на расчете ступенчатых фильтров низких частот с максимально-плоской и равноволновой характеристиками затухания.

пенчатых фильтров низких частот с максимально-плоской и равноволновой характеристиками затухания.

ПОСТАНОВКА ЗАДАЧИ

В инженерной практике при расчете микроволновых фильтров широко применяются максимально-плоская и равноволновая аппроксимации частотных характеристик затухания.

Для ступенчатого низкочастотного фильтра-прототипа с максимально-плоской характеристикой, реализуемого на соизмеримых отрезках линий, частотная зависимость коэффициента передачи по мощности определяется выражением [3]

$$|S_{12}|^2 = \frac{1}{1 + [\sin(\omega)/\alpha]^{2n}}, \quad (1)$$

где $\alpha = \sin(\omega_c)$, ω_c – частота среза, n – порядок фильтра.

Передаточные функции от комплексной частоты $p = j\omega$ для параметров рассеяния фильтра, вытекающие из (1), имеют вид

$$S_{11}(p) = \frac{A_{11}(p)}{B(p)} = \frac{(1-z)^n}{\prod_{i=1}^n (a_i - b_i z)}, \quad (2)$$

$$S_{12}(p) = \frac{A_{12}(p)}{B(p)} = \frac{(2\alpha)^n z^{n/2}}{\prod_{i=1}^n (a_i - b_i z)}, \quad (3)$$

где $a_i = \sqrt{1 - \alpha^2} e^{2j\theta_i} - j\alpha e^{j\theta_i}$, $b_i = \sqrt{1 - \alpha^2} e^{2j\theta_i} + j\alpha e^{j\theta_i}$, $\theta_i = (2i - 1)\pi/(2n)$, $z = e^{-2p}$.

Коэффициент передачи по мощности ступенчатого фильтра-прототипа с равноволновой характеристикой затухания определяется функцией [3]

$$|S_{12}|^2 = \frac{1}{1 + T_n[\sin(\omega)/\alpha]^2}, \quad (4)$$

где n – порядок фильтра, $T(x) = \cos(n \arccos(x))$ – полином Чебышева n -го порядка.

Из этого выражения вытекают соотношения для параметров рассеяния

$$S_{11}(p) = \frac{A_{11}(p)}{B(p)} = \prod_{i=1}^n \frac{d_i - q_i z}{g_i - t_i z}; \quad (5)$$

$$S_{12}(p) = \frac{A_{12}(p)}{B(p)} = (2\alpha)^n z^{n/2} \prod_{i=1}^n \frac{v_i}{g_i - t_i z}, \quad (6)$$

где $d_i = \sqrt{1 - \alpha^2 \cos^2(\theta_i)} - j\alpha \cos(\theta_i)$, $q_i = \sqrt{1 - \alpha^2 \cos^2(\theta_i)} + j\alpha \cos(\theta_i)$, $\theta_i = (2i - 1)\pi / (2n)$, $g_i = \sqrt{1 - \alpha^2 \cos^2(\phi_i)} - j\alpha \cos(\phi_i)$, $t_i = \sqrt{1 - \alpha^2 \cos^2(\phi_i)} + j\alpha \cos(\phi_i)$, $v_i = \sqrt{\eta^2 + \sin^2(j\pi/n)}$, $\phi_i = a \sin(j\eta) + \theta_i$.

По соотношениям (2), (3) или (5), (6) определяются число звеньев и значения параметров элементов фильтра. Традиционно эта задача решается с использованием А- или Т-матриц передачи [1–5]. Более рациональным является оперирование непосредственно с матрицей рассеяния фильтра и его звеньев. В связи с этим представляет интерес использование для расчета фильтров метода анализа устройств СВЧ, предложенного в [6]. Этот метод позволяет представить матрицу рассеяния каскадного соединения четырехполюсников произведением матриц, специальным образом составленных из числителей и знаменателей элементов матриц рассеяния четырехполюсников, входящих в соединение.

МЕТОД РАСЧЕТА ПАРАМЕТРОВ ЭЛЕМЕНТОВ ФИЛЬТРА С КАСКАДНОЙ СТРУКТУРОЙ

Составим из матриц рассеяния каскадного соединения n четырехполюсников матрицу со следующей структурой [6]:

$$S = \begin{bmatrix} A_{11_1} & A_{12_1} & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ A_{21_1} & A_{22_1} & -B_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -B_2 & A_{11_2} & A_{12_2} & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & A_{21_2} & A_{22_2} & -B_2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & -B_n & A_{11_n} & A_{12_n} \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & A_{21_n} & A_{22_n} \end{bmatrix}, \quad (7)$$

где A_{kl_r} – числитель, B_r – знаменатель элемента S_{kl_r} матрицы рассеяния r -го четырехполюсника.

Вычеркивая строки и столбцы, соответствующие свободным плечам соединения, определим миноры,

образующие числители \widehat{A}_{kl} и знаменатель \widehat{B} элемента \widehat{S}_{kl} матрицы рассеяния соединения. Раскрывая миноры, непосредственно из (7) получим соотношения для числителей коэффициентов передачи:

$$\widehat{A}_{12} = (-1)^{n-1} \prod_{r=1}^n A_{12_r}; \quad (8)$$

$$\widehat{A}_{21} = (-1)^{n-1} \prod_{r=1}^n A_{21_r}. \quad (9)$$

Коэффициенты отражения будут определяться следующим соотношением:

$$\begin{bmatrix} \widehat{A} & -\widehat{A}_{11} \\ \widehat{A}_{22} & \widehat{B} \end{bmatrix} = \prod_{r=1}^n \begin{bmatrix} A_r & -A_{11_r} \\ A_{22_r} & B_r \end{bmatrix}. \quad (10)$$

Здесь

$$\widehat{A} = \frac{\widehat{A}_{11} \widehat{A}_{22} - \widehat{A}_{12} \widehat{A}_{21}}{\widehat{B}}; \quad (11)$$

$$A_r = \frac{A_{11_r} A_{22_r} - A_{12_r} A_{21_r}}{B_r}. \quad (12)$$

Следует отметить, что по теореме Якоби [7] выражение (11) соответствует определителю матрицы (7), т. е. числитель в правой части выражения (11) должен содержать общий множитель, равный \widehat{B} . Аналогичное условие выполняется для выражения (12). В частности, при составлении матрицы рассеяния четырехполюсника по его эквивалентной схеме, будет иметь место [7, 8]

$$\left. \begin{aligned} S_{kl_r} &= \frac{A_{kl_r}}{B_r} = \frac{\Delta_{kl} - \delta_{ki} \Delta}{\Delta}, \\ A_r &= \frac{1}{B_r} \begin{bmatrix} A_{11_r} & A_{12_r} \\ A_{21_r} & A_{22_r} \end{bmatrix} = 4 - 2(\Delta_{11} + \Delta_{22}) + \Delta, \end{aligned} \right\} \quad (13)$$

где $\delta_{ki} = 0$ при $k \neq 1$ или 1 при $k = 1$; Δ , Δ_{ki} – определитель и алгебраические дополнения матрицы проводимостей или сопротивлений схемы.

Структура ступенчатого фильтра составляется из отрезков линий передачи, характеризующихся матрицей рассеяния

$$S_r = \frac{1}{1 - \Gamma_r^2 e^{-2p\tau_0}} \begin{bmatrix} \Gamma_r (1 - e^{-2p\tau_0}) & (1 - \Gamma_r^2) e^{-p\tau_0} \\ (1 - \Gamma_r^2) e^{-p\tau_0} & \Gamma_r (1 - e^{-2p\tau_0}) \end{bmatrix}, \quad (14)$$

где $\Gamma_r = \frac{\rho_r - \rho_0}{\rho_r + \rho_0}$ – коэффициент отражения сочленения линий, ρ_r – волновое сопротивление r -го отрезка линии, ρ_0 – волновое сопротивление тракта, в который включен отрезок линии, τ_0 – задержка сигнала отрезком линии.

Из (14) следует выражение для матрицы четырех-полюсника в форме, соответствующей (10):

$$M_r = \begin{bmatrix} A_r & -A_{11r} \\ A_{22r} & B_r \end{bmatrix} = \begin{bmatrix} \Gamma_r^2 - z & -\Gamma_r(1-z) \\ \Gamma_r(1-z) & -(1-\Gamma_r^2 z) \end{bmatrix}, \quad (15)$$

где $z = e^{-2p\tau_0}$.

В аналогичной форме составляется матрица рассчитываемого ступенчатого фильтра. Для фильтра с максимально плоской характеристикой, принимая во внимание (2), (3), (11), запишем

$$\widehat{M}_{\text{МП}} = \begin{bmatrix} \widehat{A} & -\widehat{A}_{11} \\ \widehat{A}_{22} & \widehat{B} \end{bmatrix} = \begin{bmatrix} \prod_{i=1}^n (b_i - a_i z) & -(1-z)^n \\ (1-z)^n & -\prod_{i=1}^n (a_i - b_i z) \end{bmatrix}. \quad (16)$$

Для фильтра с равноволновой характеристикой из (5), (6), (11) получим

$$\widehat{M}_{\text{РВ}} = \begin{bmatrix} \widehat{A} & -\widehat{A}_{11} \\ \widehat{A}_{22} & \widehat{B} \end{bmatrix} = \begin{bmatrix} \prod_{i=1}^n (t_i - g_i z) - \prod_{i=1}^n (d_i - q_i z) \\ \prod_{i=1}^n (d_i - q_i z) - \prod_{i=1}^n (g_i - t_i z) \end{bmatrix}. \quad (17)$$

При составлении определителя \widehat{A} в (16) и (17) учтено, что, в соответствии с (11), корни полинома $\widehat{A}_{11}^2(z) - \widehat{A}_{12}^2(z)$ включают в себя n корней знаменателя $\widehat{B}(z)$ и n сопряженных с ними корней.

Соотношения (10), (15)–(17) обеспечивают нахождение коэффициентов отражения Γ_r звеньев фильтра. Для этого достаточно n раз умножить (16) или (17) на матрицу, обратную матрице (15), и на каждом шаге

умножения определить значение коэффициента отражения, при котором степень полиномов в матрице \widehat{M} уменьшается на единицу.

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

В качестве примера рассмотрим расчет коэффициентов отражения Γ_r звеньев ступенчатого фильтра нижних частот 3-го порядка с максимально плоской и равноволновой характеристиками с частотой среза $F_0 = 3$ ГГц при запаздывании сигнала $\tau_0 = 20,83$ нс в отрезках линий с длиной $(\lambda_0/16)$.

Матрицы (16), (17) представим в виде

$$\widehat{M}_{\text{МП}} = \begin{bmatrix} -1 & 2,172 \\ -2,172 & 4,719 \end{bmatrix} + \begin{bmatrix} 4,471 & -6,517 \\ 6,517 & -7,216 \end{bmatrix} z + \begin{bmatrix} -7,216 & 6,517 \\ -6,517 & 4,719 \end{bmatrix} z^2 + \begin{bmatrix} 4,719 & -2,172 \\ 2,172 & -1 \end{bmatrix} z^3,$$

$$\widehat{M}_{\text{РВ}} = \begin{bmatrix} -6,228 & 1,58 \\ -1,58 & 2,496 \end{bmatrix} + \begin{bmatrix} 11,001 & -4,045 \\ 4,045 & -4,408 \end{bmatrix} z + \begin{bmatrix} -8,373 & 4,045 \\ -4,045 & 3,355 \end{bmatrix} z^2 + \begin{bmatrix} 2,496 & -1,58 \\ 1,58 & -1 \end{bmatrix} z^3.$$

Понижая порядок этих матричных многочленов путем умножения на матрицу, обратную матрице (15), найдем коэффициенты отражений Γ_r звеньев фильтров. В результате получим: для фильтра с максимально плоской характеристикой $\Gamma_1 = \Gamma_3 = 0,460$, $\Gamma_2 = -0,659$; для фильтра с равноволновой характеристикой $\Gamma_1 = \Gamma_3 = 0,633$, $\Gamma_2 = -0,439$ при неравномерности характеристики $\varepsilon = 0,4$. На рис. 1 представлены амплитудно-частотные характеристики параметров рассеяния рассчитанных фильтров: кривая 1 соответствует коэффициенту отражения $S_{11_{\text{МП}}}$, кривая 2 – коэффициенту передачи $S_{12_{\text{МП}}}$ при максимально-плоской ап-

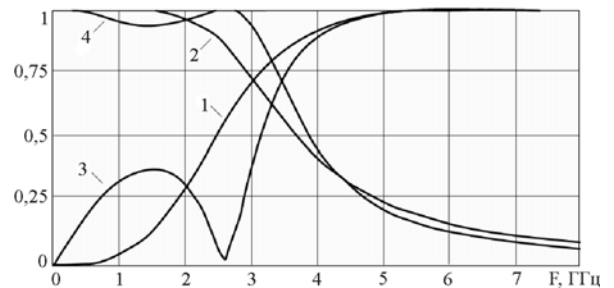


Рисунок 1 – Частотные зависимости параметров рассеяния трехзвенного ступенчатого фильтра нижних частот с максимально плоской и равноволновой характеристиками

проксимации; кривая 3 соответствует коэффициенту отражения $S_{11_{рв}}$, кривая 4 – коэффициенту передачи $S_{12_{рв}}$ при равноволновой аппроксимации затухания.

ЗАКЛЮЧЕНИЕ

Разработан метод расчета ступенчатых фильтров нижних частот по известным передаточным функциям параметров рассеяния, который является более экономичным по сравнению с известными методами. Метод оперирует с полиномами знаменателя и числителей параметров рассеяния фильтра, представленных дробно-рациональными функциями комплексной частоты, чем обеспечивается наглядность и простота процесса формирования структуры фильтра и синтеза параметров его элементов.

Разработанный метод может быть также использован для решения задач анализа и синтеза микроволновых устройств различного назначения, представленных структурой в виде каскадного соединения четырех полюсников.

ПЕРЕЧЕНЬ ССЫЛОК

1. Микроволновые устройства телекоммуникационных систем: В 2 т. / Згуровский М. З., Ильченко М. Е., Кравчук С. А. и др. – К.: Політехніка, 2003. – Т. 1: Распространение радиоволн. Антенные и частотно-избирательные устройства. – 456 с.

2. *Hong Jia-Sheng, Lancaster M. J.* Microstrip filters for RF / Microwave applicatios. – New York: John Wiley, 2001. – 476 p.
3. *Сысоев И. В.* Расчет полосковых фильтров. – М.: Специальная техника и связь, 2004. – 124 с.
4. *Роудз Дж. Д.* Теория электрических фильтров: Пер. с англ. / Под ред. А. М. Трахмана. – М.: Сов. радио, 1980. – 240 с.
5. *Сазонов Д. М., Гридин А. Н., Мишустин Б. А.* Устройства СВЧ. – М.: Высшая школа, 1981. – 295 с.
6. *Карпуков Л. М.* Символьный анализ устройств СВЧ методом подсхем // Электронное моделирование. – 1984. – № 3. – С. 81–84.
7. *Гантмахер Ф. Р.* Теория матриц. – М.: Физматлит, 2004. – 624 с.
8. *Карпуков Л. М.* Символьный анализ устройств СВЧ // Известия вузов. Радиоэлектроника. – 1982. – Т. 25. – № 6. – С. 85–87.

Надійшла 22.11.2007
Після доробки 17.12.2007

Запропонований матричний метод розрахунку східчастих мікрохвильових фільтрів. Метод заснований на використанні матриць спеціального вигляду, що складаються з поліномів комплексної частоти, відповідних знаменнику і чисельникам параметрів розсіяння фільтру. Наведені результати розрахунку фільтрів з максимально-плоскою і рівнохвильовою характеристиками загасання.

The matrix method of calculation step microwave filters is offered. The method is based on use of matrixes of the special kind made of polynoms of complex frequency, corresponding a denominator and numerators of parameters of dispersion of the filter. Results of calculation of filters with as much as possible-flat and equal wave characteristics of attenuation are presented.

УДК 621.382

Ю. А. Крисан, А. А. Крисан

ИССЛЕДОВАНИЕ ЭЛЕКТРОННОГО КЛЮЧА С ГАЛЬВАНИЧЕСКОЙ РАЗВЯЗКОЙ

Предложен электронный ключ, используемый как быстродействующий пороговый элемент с гальванической развязкой – электронное реле для контроля состояния приборов в полупроводниковых преобразователях, который удовлетворяет основным требованиям, предъявляемым к устройствам контроля.

ВВЕДЕНИЕ

Одной из проблем в преобразовательной технике является получение достоверной информации о состоянии приборов силовой схемы. Известные устройства не обеспечивают достаточной надежности и быстродействия, необходимых при конструировании преобразователей.

© Крисан Ю. А., Крисан А. А., 2008

Применяемые для таких целей устройства коммутации на основе реле и развязочных трансформаторов с магнитопроводом имеют известные недостатки [1]. К недостаткам гальванических развязок с использованием оптронов можно отнести низкую надежность работы из-за деградации светоизлучающих и светоприемных свойств оптронных пар. Происходит как деградация яркости, так и увеличение темнового тока оптронов. Оптроны критичны к режиму эксплуатации. Изменение режимов эксплуатации существенно влияет на надежность работы. При повышенных температурах эксплуатации деградация световых свойств оптронной пары происходит быстрее, что не обеспечивает работу устройства в течение гарантийного срока работы изделия [2, 3].

ИССЛЕДОВАНИЕ СХЕМЫ КЛЮЧА

Альтернативным решением вышеупомянутым устройством развязки является электронный ключ с гальванической развязкой (рис. 1), реализуемый на базе высокочастотного генератора G с выходным трансформатором $TV1$, вторичная обмотка которого нагружена через ключ K на резистор $R3$ и выпрямитель на диоде $VD1$.

Устройство работает следующим образом. При отсутствии напряжения на входе BX выходной ключ K разомкнут и контур, образованный вторичной обмоткой трансформатора $TV1$, входным ключом K и резистором $R3$, также разомкнут. На трансформаторе $TV1$ присутствует высокочастотное напряжение, наличие которого свидетельствует о разомкнутом состоянии ключевого элемента и отсутствии напряжения на входе. Высокочастотное напряжение с трансформатора $TV1$ через выпрямитель поступает на вход выходного ключа. При подаче напряжения на вход ключа K вторичная обмотка трансформатора $TV1$ нагружается на резистор $R3$. При этом изменяется эквивалентное резонансное сопротивление контура генератора. Условие самовозбуждения генератора нарушается и колебания срываются. Выпрямленное напряжение, поступающее в базовую цепь выходного ключа, реализованного на транзисторе $VT3$, исчезает и ключ закрывается, замыкая выходную цепь. При снятии напряжения со входа, генератор возобновляет свою работу, выходной ключ открывается, замыкая выходную цепь. Схема замещения трансформатора представлена рис. 2 [3].

Здесь r_1 – активное сопротивление первичной обмотки; L_1 – индуктивность первичной обмотки; M – взаимная индуктивность первичной и вторичной обмо-

ток; L_2 – индуктивность вторичной обмотки; r_2 – активное сопротивление вторичной обмотки; r_n, x_n – активное и реактивное сопротивление нагрузки.

Учитывая необходимость в большом быстродействии ключа, частота генератора должна быть высокой, при этом активные сопротивления первичной и вторичной обмоток трансформатора по сравнению с реактивным сопротивлением обмоток составляют малую величину и, с достаточной точностью для качественной и количественной оценки процессов, ими можно пренебречь. Тогда схема замещения упрощается и, с учетом элементов схемы ключа, будет иметь следующий вид (рис. 3, а).

Здесь $C_{ген}$ – емкость генератора; R – сопротивление нагрузки; S – ключ; $L_э$ – величина, характеризующая эквивалентную индуктивность трансформатора с учетом величин L_1, L_2, M, R ; $R_{вн}$ – это вещественная часть комплексного числа, характеризующая потери, вносимые сопротивлением.

При разомкнутом ключе S индуктивностью контура генератора является индуктивность первичной обмотки трансформатора L_1 , при этом потери в контуре отсутствуют, добротность контура высокая, что и обеспечивает устойчивую работу генератора. При замкнутом ключе S сопротивление нагрузки R вносит затухание в колебательный контур, величина которого зависит от соотношений величин L_1, L_2, M .

Определим степень вносимого затухания сопротивлением нагрузки R . Комплексное входное сопротивление трансформатора при замкнутом ключе S (с клемм 1, 2 рис. 3, а)

$$Z_{вх} = j\omega(L_1 - M) + \frac{[j\omega(L_2 - M) + R] \cdot j\omega M}{R + j\omega[(L_2 - M) + M]} \quad (1)$$

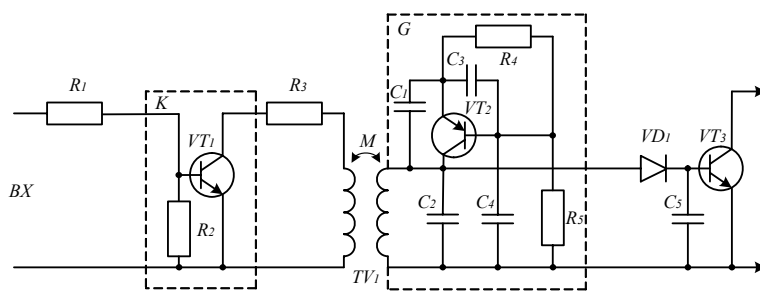


Рисунок 1

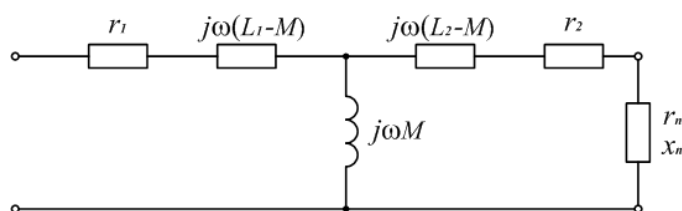


Рисунок 2

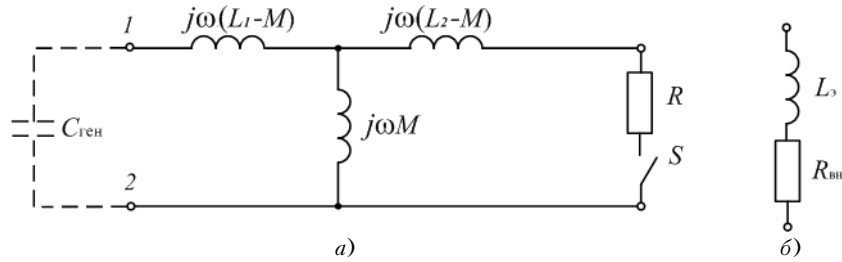


Рисунок 3

После преобразований получим выражение

$$Z_{\text{вх}} = j\omega(L_1 - M) + \frac{\omega^2 R M L_2 - \omega^2 (L_2 - M) M R}{R^2 + \omega^2 L_2^2} + \frac{j\omega R^2 M + j\omega^3 (L_2 - M) M L_2}{R^2 + \omega^2 L_2^2}. \quad (2)$$

Из выражения (2) видно, что $Z_{\text{вх}}$ представляет собой комплексное сопротивление

$$Z_{\text{вх}} = R_{\text{вн}} + j\omega L_3, \quad (3)$$

которое является эквивалентом схемы (рис. 3, а), и приведено на рис. 3, б.

Эквивалентное вносимое сопротивление потерь в контуре генератора равно

$$R_{\text{вн}} = \frac{\omega^2 R M L_2 - \omega^2 (L_2 - M) M R}{R^2 + \omega^2 L_2^2}, \quad (4)$$

после преобразования получаем выражение

$$R_{\text{вн}} = \frac{R}{R^2 / \omega^2 M^2 + L_2^2 / M^2}. \quad (5)$$

Из выражения (5) следует, что при $R = 0$ и при $R = \infty$ эквивалент вносимого сопротивления также равен нулю. При других значениях вносимое сопротивление имеет действительное значение и достигает при определенных значениях сопротивления нагрузки максимума. Для нахождения оптимального значения сопротивления, обеспечивающего максимальное затухание в контуре, определим производную из выражения и приравняем ее к нулю ($\frac{dR_{\text{вн}}}{dR} = 0$).

$$\frac{dR_{\text{вн}}}{dR} = \frac{L_2^2 / M^2 + R^2 / \omega^2 M^2 - 2R^2 / \omega^2 M^2}{(L_2^2 / M^2 + R^2 / \omega^2 M^2)^2} = 0. \quad (6)$$

Приравняв числитель к нулю

$$\frac{L_2^2}{M^2} + \frac{R^2}{\omega^2 M^2} - \frac{2R^2}{\omega^2 M^2} = 0, \quad (7)$$

находим $R_{\text{опт}}$:

$$R_{\text{опт}} = \omega L_2. \quad (8)$$

При этом значении сопротивления нагрузки трансформатор будет работать в согласованном режиме, внося максимальное затухание в контур.

Значение эквивалентной индуктивности L_3 определяется из выражения (2)

$$L_3 = L_1 - M + \frac{\omega^2 (L_2 - M) L_2 M + R^2 M}{R^2 + \omega^2 L_2^2}. \quad (9)$$

После преобразований получаем выражение

$$L_3 = L_1 - \frac{\omega^2 M^2 L_2}{R^2 + \omega^2 L_2^2} = L_1 - \frac{M}{R^2 / \omega^2 M L_2 + L_2 / M}. \quad (10)$$

Резонансное сопротивление контура $Za = \frac{L}{RC}$, при различных значениях R_2 , равно:

$$R_2 = 0, Z_1 = \frac{L_1(1 - K_{\text{св}}^2)}{C R_1} L_1; \quad (11)$$

$$R_2 = \infty, Z_2 = \frac{L_1}{C R_1}; \quad (12)$$

$$R_2 = \omega L_2, Z_3 = \frac{L(2 - K_{\text{св}}^2)}{C(2R_1 + L\omega K_{\text{св}}^2)}, \quad (13)$$

где $K_{\text{св}} = \frac{M}{\sqrt{L_1 L_2}}$.

Количественную оценку достигнутого эффекта определим через коэффициент эффективности

$$K_{\text{эф}} = \frac{K_2}{K_1}, \quad (14)$$

где K_1 и K_2 – степень изменения резонансного сопротивления контура при коммутации обмотки на коротко и на резистор $R = \omega L_2$.

$$K_1 = \frac{1}{(1 + K_{\text{св}}^2)}, \quad (15)$$

$$K_2 = \frac{2 + QK_{\text{св}}^2}{2 - K_{\text{св}}^2}. \quad (16)$$

Коэффициент эффективности после преобразования

$$K_{\text{эф}} = 1 + \frac{K_{\text{св}}^2(Q - 1) - QK_{\text{св}}^4}{2 - K_{\text{св}}^2}. \quad (17)$$

Из выражения видно, что, если числитель дроби приравнять к нулю, то коэффициент эффективности будет постоянным и равен 1

$$K_{\text{св}}^2(Q - 1) - QK_{\text{св}}^4 = 0. \quad (18)$$

Откуда после преобразований получаем

$$Q = \frac{1}{1 - K_{\text{св}}^2}. \quad (19)$$

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

При изменении значения $K_{\text{св}}$ от 0 до 1 определены зависимости добротности Q при $K_{\text{эф}} = 1$. По рассчитанным данным построен график $Q = f(K_{\text{св}})$ при $K_{\text{эф}} = 1$ (рис. 4). А также определены зависимости $K_{\text{эф}} = f(K_{\text{св}})$ при $Q = \text{const}$ (рис. 5). Анализируя эти зависимости, можно отметить, что при значениях $Q = 50$ и $K_{\text{св}} = 0,6 \div 0,8$ коэффициент эффективности почти на порядок выше, чем в первом случае и что обеспечивает надежность работы устройства в целом.

Надежность работы ключа оценивается по следующему критерию.

Амплитуда стационарных колебаний генератора определяется условием баланса амплитуд

$$|-S_{\text{ср}}| \cdot |Z| \cdot |\beta| = 1, \quad (20)$$

где $S_{\text{ср}}$ – средняя крутизна активного прибора, например, транзистора; Z – резонансное сопротивление контура; β – коэффициент передачи цепи обратной связи.

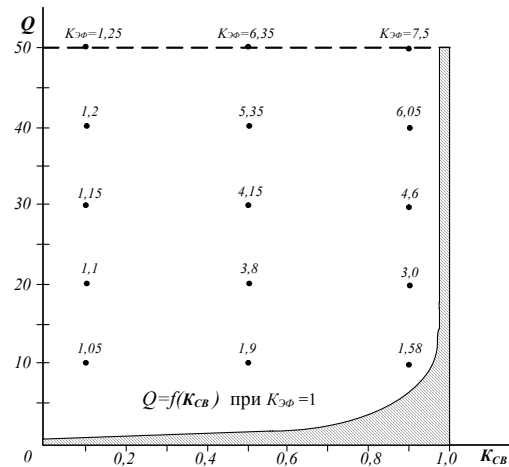


Рисунок 4

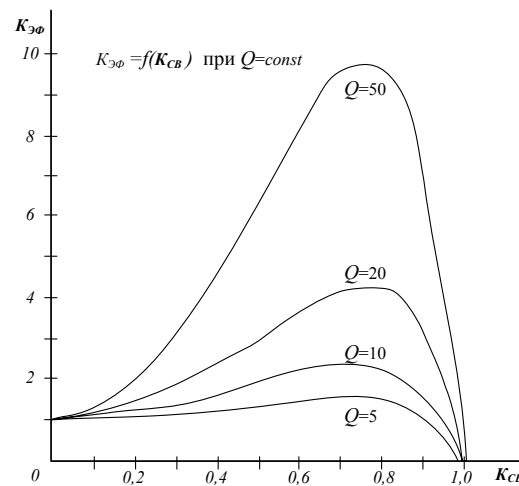


Рисунок 5

Практически нестабильными величинами являются все три составляющие $S_{\text{ср}}$, Z и β . Используя возможность воздействия на состояние генератора изменением резонансного сопротивления контура в достаточно большом диапазоне, можно исключить влияние дестабилизирующих факторов на $S_{\text{ср}}$ и β , что и обеспечит высокую надежность работы ключа.

Быстродействие ключа определяется по декременту затухания

$$d = \frac{R}{2L}T, \quad (21)$$

R – активное сопротивление контура; L – индуктивное сопротивление контура; T – период колебаний.

Декремент затухания при $R_{\text{опт}} = \omega L$

$$d = \frac{R}{2L}T = \frac{1}{2}. \quad (22)$$

Величина d достатньо велика, завдяки чому перехідний процес в схемі буде по характеру ближче к аперіодическому, що і забезпечує високе швидкодействие ключа.

ВИВОДИ

Предлагаемое технічне рішення може бути використано в автоматичній, обчислювальній і вимірній техніці, коли потрібно ключ з великим швидкодействием і великими напругами розв'язки «вхід – вихід».

Конструктивно електронний ключ розроблено на печатному трансформаторі з робочим напругою розв'язки 500 В (випробувальне напруга 2,5 кВ) з швидкодействием $t_{\text{вкл}} + t_{\text{выкл}} = 0,2$ мкс.

Характеристика ключа аналогічна релейному елементу з гістерезисом. Крім прямого використання ключ використовується як пороговий елемент для контролю стану тиристорів в перетворювачах енергії постійного і змінного струму.

УДК 621.396.6

В. М. Кришук, Г. М. Шило, Б. А. Артюшенко

РОЗПАРАЛЕЛЕННЯ ГЕНЕТИЧНОГО АЛГОРИТМУ ПАРАМЕТРИЧНОГО СИНТЕЗУ АНТЕННОЇ ГРАТКИ НА КОМП'ЮТЕРНІЙ МЕРЕЖІ

Для підвищення ефективності генетичного алгоритму (ГА) параметричного синтезу антенних решіток при заданих допусках запропоновано метод розпаралелення. Запропонований метод розроблено для роботи на комп'ютерній мережі та базується на паралельній моделі островів. Було досліджено вплив асинхронного, синхронного та адаптивного режимів роботи і обрано асинхронний режим.

ВСТУП

Використання генетичних методів при синтезі антенних систем останнім часом отримало значного поширення [1–4]. В основу роботи ГА покладено ідею запозичену з еволюційного процесу біології: найпристосованіші індивідууми популяції виживають, схрещуються між собою та дають нове, краще, покоління. Мутація, схрещення та відбір дозволяють видам пристосуватись до різноманітних умов існування. Ці ж проце-

ПЕРЕЧЕНЬ ССЫЛОК

1. Дугин Г. К. Схемы на феррит-полупроводниковых ячейках в аппаратуре связи. – М.: «Связь», 1968. – 312 с.
2. Применение оптоэлектронных приборов: Пер. с англ. С. Гейг, Д. Эванс, М. Хоуп, Х. Соренсен. – М.: «Радио и связь», 1981. – 344 с.
3. Барыбин А. А. Электроника и микроэлектроника. Физико-технологические основы. – М.: «Физматлит», 2007. – 424 с.
4. Попов В. П. Основы теории цепей. – М.: «Высшая школа», 2005. – 575 с.

Надійшла 29.03.2007
Після доробки 17.03.2008

Запропоновано електронний ключ, який використовується як швидкодіючий пороговий елемент з гальванічним розв'язанням – електронне реле для контролю стану приладів в напівпровідникових перетворювачах, що задовольняє основним вимогам до приладів контролю.

The electronic key used as a high-speed threshold element with a galvanic outcome – the electronic relay for the control of a devices condition over semi-conductor converters which meets the basic requirements shown to devices of the control is offered.

си дають можливість знаходити оптимальні рішення і для різноманітних прикладних задач.

Але ці методи потребують значних обчислювальних ресурсів. В зв'язку з чим постає задача розпаралелення обчислень [5]. Для цього можуть застосовуватись різноманітні апаратні системи, але, враховуючи наявні можливості конструкторських організацій, доцільно адаптувати паралельні обчислення саме для мереж персональних комп'ютерів (ПК).

Метою роботи є розробка засобів підвищення ефективності генетичного алгоритму параметричного синтезу антенних систем при заданих допусках шляхом розпаралелення ГА на комп'ютерній мережі.

Для вирішення поставленої задачі необхідно:

- дослідити особливості комп'ютерної мережі;
- дослідити можливі засоби розпаралелення ГА;
- адаптувати запропоновані засоби для задачі ГА синтезу антенних систем.

© Кришук В. М., Шило Г. М., Артюшенко Б. А., 2008

ОСОБЛИВОСТІ КОМП'ЮТЕРНОЇ МЕРЕЖІ

При паралельних обчисленнях на комп'ютерній мережі виникають суттєві затримки передачі даних. Так для мережі Fast Ethernet латентність складає 150 мкс і апаратна пропускна здатність не перевищує 100 Мб/с. Отже для пересилки опису амплітудно-фазового розподілу антенної ґратки (АГ) з 441 елементів час перевищуватиме 300 мкс (досліди показали, що насправді потрібно від 1 мс). При роботі на мережі ПК затримки та пропускні здатності між усіма ПК однакові, як наслідок, одночасна передача даних до одного ПК призведе до значних затримок при синхронізації. В роботі застосовувались наступні засоби: мережа Fast Ethernet з 10 ПК (процесор – AMD Athlon 1.66 GHz, оперативна пам'ять – 256 МБ, операційна система – QNX Neutrino 2.0). Обчислення діаграми напрямленості зазначеної антенної ґратки на цих ПК складає від 2 мс. Отже пересилка даних займає час того ж порядку, що і обчислення пристосованості одного геному.

З іншої сторони існує проблема програмної реалізації синхронізації. Вирішення цієї проблеми можливе застосуванням: багатопоточності, переважання переривання від мережевого пристрою чи циклічної перевірки приходу даних. Ці підходи досить складні при програмуванні. Наприклад, застосування багатопоточності пов'язано з проблемами розмежування доступу до даних. Операційна система QNX також дозволяє вирішити цю проблему при використанні мережевого протоколу Qnet, який дозволяє передавати сигнали POSIX. Отже для синхронізації достатньо переважити лише обробку сигналів.

РОЗПАРАЛЕЛЕННЯ ГЕНЕТИЧНИХ АЛГОРИТМІВ

Роботу ГА пояснює наступний запис алгоритму:

- 10** Створення початкової популяції P (0); // P (0) – початкова популяція
- 20** epoch=0; // номер покоління
- 30** Визначення пристосованості індивідуумів популяції P (0);
- 40 while** (не виконано критерій зупинки) // кількість поколінь, час обчислень
- 50** epoch = epoch + 1; // збільшення лічильника поколінь
- 60** P*(epoch) = селекція (P*(epoch – 1));
- 70** P (epoch) = схрещення та мутація (P*(epoch));
- 80** Визначення пристосованості індивідуумів популяції P (epoch);
- 90 end while;**

На 30-му та 80-му кроках визначається ступінь пристосованості індивідуумів – тобто оцінюються вихідні функції, які слід мінімізувати (максимізувати). На 60-

му кроці селекція обирає найпристосованіші індивідууми (тобто ті, для яких вихідна функція має мінімальне / максимальне значення). На 70-му кроці отримані індивідууми схрещуються між собою (тобто обмінюються частинами бітового коду їх машинного запису) та мутують (зазнають випадкових змін). Таким чином наступна популяція залежатиме від найпристосованіших індивідуумів попередньої. Але її індивідууми будуть відрізнятися від попередніх випадково.

Однією із часто згадуваних особливостей ГА є «природний» паралелізм, під яким розуміється можливість незалежного обчислення пристосованості індивідуумів популяції (80-й крок алгоритму).

Але генетичний алгоритм, з глобальною популяцією, над якою проводиться глобальна селекція та кросовер, важкий для розпаралелення: селекція (крок 60) проводиться послідовно на одному процесорі. Для вирішення цього недоліку застосовують пропорційну, турнамент та інші типи селекції [3]. Ці методи потребують значної міжпроцесорної взаємодії на пересилку даних (для аналізу пристосованості індивідууму клієнтом) та отримання результатів і, як наслідок, значних комунікаційних витрат.

З метою підвищення ефективності паралельної реалізації ГА було розроблено спеціальні версії генетичних методів. Наприклад, при дрібно-зернистій моделі [5] кожному з індивідуумів відповідає окремий процесор, що проводить мутацію, кросовер та обчислення пристосованості. Але селекція обмежується виключно сусідніми процесорами. Ця модель вимагає спеціалізованих масово-паралельних ЕОМ і, відповідно, не може бути застосована для мережі ПК.

Іншим методом є паралельна модель островів (pIGA – parallel island GA). При ній кожному з процесорів надається окрема популяція. Процесор виконує еволюцію за допомогою звичайного, послідовного ГА. Періодично процесор обмінюється частиною власних індивідуумів з іншими процесорами. Таким чином обсяг комунікацій значно скорочується та стає більш керованим. Як показано в [6], для задачі допускового синтезу антенних ґраток за наявності відмов або відхилень, найкращі результати показує СНС (Cross generational elitist selection Heterogeneous recombination (incest preventing) and Cataclysmic mutation [7]). Цей метод також найлегше реалізується для острівної моделі (Ip-SНС), цьому сприяють малий обсяг популяції, вбудовані засоби відсіву однакових індивідуумів та запобігання схрещенню близьких індивідуумів (при міграції кращих індивідуумів між популяціями досить частими можуть бути повтори).

Робота паралельного ГА за моделлю островів відбувається наступним чином [8]:

- 10** Головний процес (сервер) ініціює клієнтів та відсилає їм початкові популяції (генеруються випадково);

- 20** Клієнти проводять еволюцію отриманих популяцій (СНС) протягом заданої сервером кількості поколінь (у відповідності до попереднього алгоритму, за критерій зупинки обирається кількість поколінь);
- 30** Сервер очікує на сигнал від всіх клієнтів про завершення обчислень;
- 40** Сервер випадково обирає клієнта;
- 50** Обраний клієнт відсилає свої найпристосованіші індивідууми серверу;
- 60** Сервер пересилає найпристосованіші індивідууми іншим клієнтам;
- 70** Клієнти вносять отримані індивідууми до власних популяцій;
- 80** Повторення з кроку 20.

Через нерівномірність часу обчислення пристосованості індивідуумів для допускових задач, час обчислень клієнтами також варіює. З метою підвищення ефективності роботи цього методу слід вирішити ряд проблем: зменшити час взаємних очікувань процесорами та збільшити вірогідність розповсюдження найкращих відомих індивідуумів. Для цього можна внести наступні зміни:

1. Замість міграції кращих індивідуумів з одного клієнта, проводити глобальний пошук та міграцію найпристосованіших індивідуумів всіх клієнтів (на кроках 40–60 алгоритму 2), враховуючи незначний програв часу на проведення такого пошуку ця зміна використовується в подальшій частині;
2. Проводити на 30-му кроці синхронізацію не за принципом однакової кількості пройдених поколінь всіма клієнтами, а у відповідності до часу. Для цього за критерій зупинки (крок 40 першого алгоритму) слід обрати час, або сервер сигналом POSIX із заданою періодичністю зупиняє роботу клієнтів по еволюції, та починає обмін індивідуумів. В останньому випадку клієнтам достатньо перевантажити обробку сигналів.
3. Проводити на 30-му кроці синхронізацію по прибуттю сигналу лише з частини (наприклад з одного) клієнтів (асинхронний режим). Після приходу сигналу з цих клієнтів, сервер сигналом зупиняє роботу клієнтів над еволюцією та починає обмін, при цьому клієнти проводять сортування індивідуумів. Для асинхронного режиму відмічають уповільнення роботи ГА [9].
4. Проводити адаптацію параметрів генетичного алгоритму під час його роботи [7]. Для цього клієнти обмінюються не тільки кращими індивідуумами, а й

параметрами ГА (щільність мутації, величина еліти та популяції), після обирання найпристосованіших індивідуумів, клієнти змінюють параметри ГА, наближаючи їх до параметрів того з процесорів, який синтезував найпристосованіші індивідууми. Враховуючи нерівномірність кількості індивідуумів популяції, проводиться виключно синхронізація за часом.

ПРИКЛАДИ ЗАСТОСУВАННЯ ПАРАЛЕЛЬНИХ ОБЧИСЛЕНЬ

Було проведено дослідження задачі синтезу амплітудно-фазового розподілу (АФР) антенної ґратки з 21×21 ненапрямлених рівномірно розташованих елементів з початковим АФР [6] та 5 % допусками на амплітуду та фази. У якості критерію оптимальності обрано мінімізацію різниці між отриманою та заданою ДН при рівні бічних пелюстків у 20 дБ.

У табл. 1 приводяться результати роботи ІрСНС (величина популяції – 45, щільність мутації – 0,065, відношення попереднього та проміжного поколінь – 1:2, клієнти обмінюються 5 % найпристосованіших індивідуумів) за різної кількості процесорів та величин популяції при синхронізації після кожного покоління та глобальній міграції найбільш пристосованих індивідуумів. Як і в [7], маємо нелінійне прискорення роботи ГА при використанні паралельного ГА. При проведенні обчислень за допомогою синхронного паралельного СНС при кількості процесорів 2, час синхронізації дорівнював половині часу обчислень на одному з ПК (приблизно 2 с). При нарощуванні кількості ПК, час синхронізації поступово збільшувався. Таким чином, використання асинхронного режиму дозволяє значно скоротити час роботи і, незважаючи на дещо гірші результати при однаковій кількості поколінь (хоча це і не настільки суттєво як в роботі [9]) за малої кількості ПК, самі обчислення проводилися чи не вдвічі швидше, отже доцільно застосовувати асинхронний режим. Як видно з табл. 2, так зване «нелінійне пришвидчення» ГА при розпаралеленні за моделлю островів [9–11] частково спричинено збільшенням популяції. Таким чином, це пришвидчення матиме місце для задач, що потребують великої популяції (багатокритеріальні та багатопараметричні задачі).

Таблиця 1 – Результати роботи ІрСНС по завершенні 75 поколінь або по генерації $2 \times 45 \times 50$ індивідуумів (через похилу ризку)

Режим	Кількість процесорів				
	2	4	6	8	10
Синхронний режим	1,33	1,18 / 1,24	1,1 / 1,25	1,06 / 1,13	1,14 / 1,42
Асинхронний режим	1,57	1,17 / 1,43	1,26 / 1,59	1,24 / 1,42	1,07 / 1,37
Синхронізація за часом	1,43	1,32 / 1,48	1,37 / 1,45	1,22 / 1,42	1,08 / 1,37

Таблиця 2 – Результати роботи СНС при різній величині популяції по завершенні 75 поколінь або по генерації 2*45*50 індивідумів (через похилу риску)

Величина популяції	45	90	180	360
Результат роботи СНС	1,48	1,26 / 1,5	1,23 / 1,64	1,17 / 1,49

Таблиця 3 – СНС з адаптацією параметрів синхронного режиму по завершенні 75 поколінь

Кількість процесорів	2	4	5	6	7	8
Отриманий результат	1,20	1,66	0,99	1,4	1.11	1,29
Величина популяції	125	103	87	21	22	42
Величина еліти	52 %	47 %	22 %	23 %	50 %	40 %
Щільність мутації	0,05	0,064	0,06	0,05	0,11	0,1

Застосування ІрСНС з адаптацією параметрів хоч і дозволило дещо збільшити ефективність обчислень (табл. 3), але не дало можливості виявити оптимальні значення параметрів ГА, які з плином часу та між системами досить суттєво змінювались. Але слід зазначити, що при застосуванні синхронного адаптивного ІрСНС, виявлено тенденцію до поступового зменшення величини популяції з ростом кількості процесорів.

ВИСНОВКИ

В роботі запропоновано підхід до розпаралелення генетичного алгоритму параметричного синтезу антенних систем за наявності відхилень від номінальних значень для середовища комп'ютерної мережі. У якості моделі розпаралелення застосовано паралельний варіант моделі островів. Для цього підходу було досліджено вплив режиму синхронізації. Як показали отримані результати, з точки зору ефективності отриманих рішень, краще застосовувати асинхронний режим, що пов'язано з варіативністю часу обчислення пристосованості індивідумів. Так зване нелінійне пришвидження роботи паралельної моделі островів при зростанні кількості процесорів, характерне також і для задачі параметричного синтезу антен. Це пов'язано із великою кількістю параметрів та зростанням величини популяції із збільшенням кількості процесорів. Впровадження ГА з адаптацією параметрів дає незначний вигравш у якості знайдених рішень та не показало оптимальних значень параметрів генетичного алгоритму.

ПЕРЕЛІК ПОСИЛАНЬ

1. Артюшенко Б. А. Синтез лінійної антенної решітки з заданими допусками на її параметри за допомогою генетичного алгоритму // Радіоелектроніка. Інформатика. Управління. – № 1. – 2007. – С. 15–18.
2. Synthesis of Sparse Planar Arrays Using Modified Real Genetic Algorithm / Chen K., Yun X., He Z., Han C. // IEEE Transactions on Antennas and Propagation. – 2007. – Vol. 55, No. 4. – P. 1067–1073.
3. John A, Ammann M. J. Design of wide-band printed antenna using genetic algorithm on an array of overlapping sub-patches // IEEE Workshop on Antenna Techno-

- logy Small Antennas and Novel Metamaterials. – 2006. – P. 92–95
4. Chatteraj N., Roy J. S. Application of Genetic Algorithm to the Optimization of Microstrip Antennas with and without Superstrate // Application of Genetic Algorithm to the Optimization of Microstrip Antennas with and without Superstrate. – 2006. – Vol. 12, No. 2. – P. 32–35.
5. A Parallel Electromagnetic Genetic-Algorithm Optimization (EGO) Application for Patch Antenna Design / Vilegas F. J., Cwik T., Rahmat-Samii Y., Manteghi M. // IEEE Transactions on antennas and propagation – 2004. – Vol. 52, No. 9. – P. 2424–2435.
6. Artyushenko B. Genetic Algorithm for Antenna Array with Failed and Deviated Elements Optimization // Proc. of IEEE Int. Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application. – Dortmund (Germany). – 2007. – P. 228–231.
7. Eshelman L. The CHC Adaptive Search Algorithm // Foundation of Genetic Algorithms, G. Rawlings, ed. Morgan-Kaufmann. – 1991. – Pp. 256–283.
8. Whitley D. Cellular genetic algorithms. // In Proceedings of the Fifth International Conference on Genetic Algorithms – San Francisco, California (USA). – 1993 – P. 658.
9. Genetic Adventures in Parallel: Towards a Good Island Model under PVM [Електронний ресурс] / Keith Vertanen. – Електр. дан. – Режим доступу: http://www.keithv.com/papers/island_model_pvm.pdf, вільний. – Заголовок з екрану.
10. Tongchim S., Chongstitvatana P. Comparison between synchronous and asynchronous implementation of parallel genetic programming // In Proceedings of the 5th International Conference for Artificial Life and Robotics.– Japan. – 2000. – P. 251–254
11. Tongchim S., Chongstitvatana P. Parallel genetic algorithm with parameter adaptation // Information Processing Letters. – 2002. – Vol. 82, No. 1. – P. 47–54.

Надійшла 3.03.2008
Після доробки 18.03.2008

С целью повышения эффективности параметрического синтеза антенных решеток при заданных допусках предложен метод распараллеливания генетического алгоритма. Предложенный метод разработан для компьютерной сети и базируется на параллельной модели островов. Было рассмотрено влияние асинхронного, синхронного и адаптивного режимов работы и выбрано асинхронный режим.

To reduce time consumption of genetic algorithm for antenna arrays optimization with given tolerances on its parameters parallelization scheme is proposed. The proposed parallel genetic algorithm is developed for case of computer-network environment. After looking at the various available parallel models, the island model was chosen as the most appropriate. Synchronous, asynchronous and adaptive realizations were examined, and among them asynchronous variant was chosen.

Д. М. Пиза, А. А. Степаненко

СЕГМЕНТАЦИЯ УЛЬТРАЗВУКОВЫХ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ МЕТОДА НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ

Рассмотрен новый подход к повышению эффективности процедуры визуального анализа ультразвуковых медицинских эхо-импульсных изображений. Представлены результаты экспериментальной проверки информационных возможностей метода.

ВВЕДЕНИЕ

Ультразвуковая интроскопия относится к одному из основных и распространенных методов медицинской диагностики [1], в рамках которой основное внимание уделяется повышению чувствительности и достоверности процедуры визуализации ультразвуковых эхо-импульсных изображений. В работе [2] рассматривались возможности повышения чувствительности выделения границ разделов низкоконтрастных структур и нейтрализации влияния многократных переотражений ультразвуковых импульсов, затрудняющих визуальный анализ эхо-импульсных изображений. Однако, с точки зрения определения однородности, шероховатости, регулярности, наибольший интерес представляет процедура сегментации непосредственно самого ультразвукового изображения. Вопросы сегментации различных типов изображений относятся к области задач компьютерного зрения [3], однако характерной особенностью ультразвуковой диагностики является наличие модуляционных спекл-шумов, поэтому применение любого метода сегментации ультразвуковых изображений должно учитывать наличие этого фактора и сопровождаться использованием процедуры фильтрации модуляционных шумов.

Еще одной характерной особенностью ультразвуковой диагностики является то обстоятельство, что на ультразвуковых изображениях границы анатомических областей являются «размытыми», поскольку акустические импедансы соседних биоструктур могут быть близкими либо плавно изменяющимися. Это обстоятельство приводит к необходимости использования нечетких методов кластеризации. Нечеткая кластеризация вводит понятие нечетких кластеров и функцию принадлежности объектов к ним, изменяющуюся в интервале $[0,1]$, что позволяет оценить степень принадлежности объекта к тому либо иному классу. Основой нечеткой кластеризации служит метод FCM (Fuzzy C-Means), имеющий множество модификаций [4]. Метод

FCM базируется на использовании идей и математического аппарата нечеткой логики и широко используется применительно к задачам анализа медицинских радиологических изображений [5], однако информационные возможности метода применительно к анализу ультразвуковых изображений ранее не исследовались.

1 МЕТОД ИССЛЕДОВАНИЯ

В результате работы алгоритма FCM каждому пикселю изображения ставится в соответствие вектор из функций принадлежности к каждому классу, на основе которого можно делать выводы о природе данного объекта, но лишь при нейтрализации влияния шумовых факторов. В настоящее время отсутствует единый подход к решению этой задачи. В [6] предлагалось использовать винеровскую фильтрацию, а в [7] адаптивную анизотропную фильтрацию. Основной проблемой здесь является то обстоятельство, что для осуществления оптимальной фильтрации необходимо априорное знание спектральных характеристик как шума, так и самого незашумленного (т. е. идеального) изображения. В этой связи интерес представляет метод селективного сингулярного разложения автоморфного отображения (МССРАО) [8] зашумленных изображений, не требующий априорного знания спектральных характеристик, а лишь приблизительного знания интервала корреляции шума.

МССРАО относится к рамочным методам анализа изображений, причем размер квадратной рамки должен соответствовать интервалу корреляции. На основе МССРАО формируется многомерный массив, к которому применяется процедура декоррелирующего преобразования с формированием нового массива «собственных» изображений. В качестве отфильтрованного изображения принимается первое «собственное» изображение, соответствующее наибольшему сингулярному числу.

Приведенные в [8] экспериментальные результаты относились к анализу изображений геофизических полей и информационные возможности метода применительно к анализу ультразвуковых эхо-импульсных изображений остались невыясненными.

Поэтому для дальнейшего повышения эффективности медицинской диагностики необходимо исследование информационных возможностей МССРАО и нечеткой FCM кластеризации применительно к задачам сегментации медицинских ультразвуковых эхо-импульсных изображений.

2 РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Нечеткая кластеризация является обобщенным вариантом традиционного алгоритма К-средних, преодолевающим ограничения принадлежности выборок только к одной группе. В нечеткой кластеризации используется нечеткое множество для задания степени принадлежности выборок только к одной группе. Численное значение степени принадлежности лежит в диапазоне $[0-1]$, так что сумма всех значений принадлежности одного элемента (пикселя) ко всем группам равна единице.

Алгоритм FCM кластеризации базируется на использовании процедуры итеративной минимизации целевой функции вида

$$J(U, V) = \sum_{i=1}^c \sum_{n=1}^N u_{in}^m |x_n - v_i|^2, \quad (1)$$

где $V = \{v_1, \dots, v_c\}$ – центры кластеров; $U = [u_{in}]$ – матрица размером $c \times N$, где u_{in} есть i -я функция принадлежности n -го входа x_n , $m \in [1, \infty)$ – параметр фаззификации, c – задаваемое число кластеров. Значения функций принадлежности должны удовлетворять следующим ограничениям

$$0 \leq u_{in} \leq 1, \quad i = 1, 2, \dots, c, \quad n = 1, 2, \dots, N; \quad (2)$$

$$\sum_{i=1}^c u_{in} = 1, \quad n = 1, 2, \dots, N; \quad (3)$$

$$0 < \sum_{n=1}^N u_{in} < N, \quad i = 1, 2, \dots, c. \quad (4)$$

Алгоритм является итеративным и основан на последовательном вычислении следующих уравнений:

$$v = \frac{1}{\sum_{n=1}^N u_{in}^m} \sum_{n=1}^N u_{in}^m x_{in}, \quad i = 1, 2, \dots, c; \quad (5)$$

$$u = \frac{\left[\frac{1}{|x_n - v_i|^2} \right]^{1/(m-1)}}{\sum_{j=1}^c \left[\frac{1}{|x_n - v_j|^2} \right]^{1/(m-1)}}, \quad (6)$$

$i = 1, 2, \dots, c, \quad n = 1, 2, \dots, N.$

Для каждого пикселя ансамбля изображений, значение его функции принадлежности для каждого кластера зависит от расстояния до соответствующего центра кластера. Параметр фаззификации m уменьшает влияние небольших значений функции принадлежности. К достоинствам алгоритма нечетких С-средних можно отнести его гибкость, позволяющую работать со значениями функций принадлежности. К недостаткам можно отнести необходимость задания априорного числа кластеров и теоретическая неопределенность выбора начальных значений и параметра фаззификации, которая приводит к получению различных результатов кластеризации, а также чувствительность к шуму.

На рис. 1, $a-z$ в координатах «время – пространство» представлено ультразвуковое эхо-импульсное изображение головного мозга ребенка (a) и первые три «собственных» изображения МССРАО ($b-z$). Такие эхо-импульсные изображения рассматриваются, как правило, в виде набора сигнальных трасс (столбцов изображения). Из рассмотрения исходного изображения на рис. 1, a видно, что оно является зашумленным, тогда как на первом «собственном» изображении МССРАО (рис. 1, b) влияние модуляционной помехи фактически нейтрализовано. Рассматривая амплитудно-временные срезы (рис. 2, a и b) 50-й импульсной трассы (50-го столбца изображений рис. 1, a и b), заметно очищение сигнала (рис. 2, b) от излишних шумов (быстроизменяющейся амплитуды в пиках и спадах (рис. 2, a)).

На рис. 2, v, z показаны амплитудно-временные срезы второго и третьего «собственных» изображений МССРАО (рис. 1, v, z), соответствующих шумовым компонентам, и их отбрасывание фактически эквивалентно эффекту адаптивной фильтрации, поскольку никаких предположений о спектральных и статистических характеристиках модуляционных шумов не использовалось.

На рис. 3 представлены результаты применения метода нечеткой кластеризации для сегментации рис. 1, a и рис. 1, b на семь классов.

Из сопоставления рис. 1, a с рис. 3, a следует, что применение метода нечеткой кластеризации для сегментации зашумленного изображения фактически ничего не дает, тогда как его применение к отфильтрованному изображению (рис. 3, b) позволяет облегчить процедуру визуального анализа и интерпретации

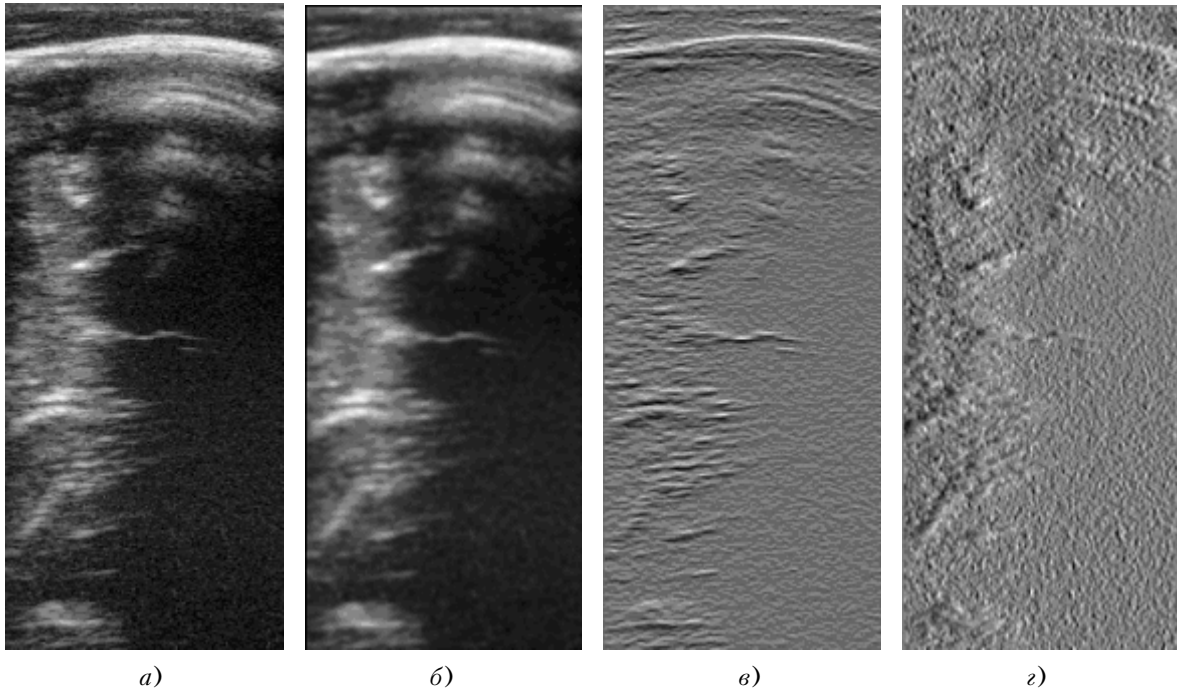


Рисунок 1 – Ультразвукове зображення головного мозку:
 а – оригінал; б-г – перші три «собственных» зображення, синтезовані на основі використання МССРАО
 (горизонтальна вісь – просторовна; вертикальна вісь – часова)

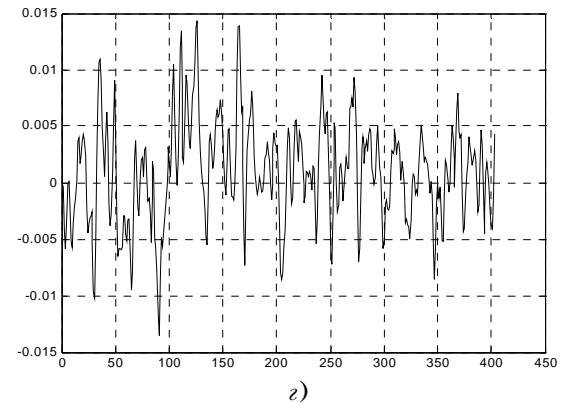
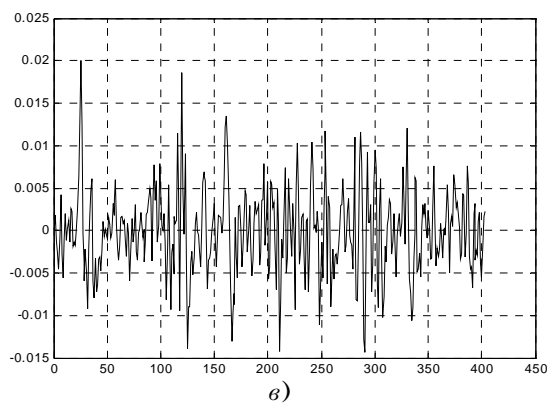
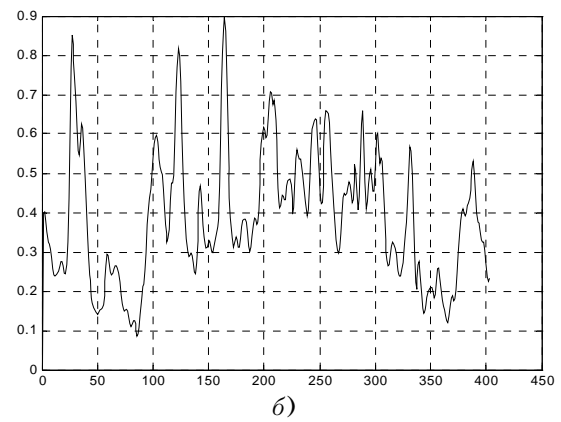
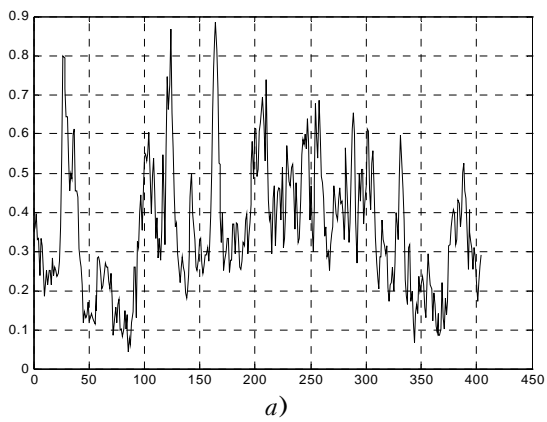
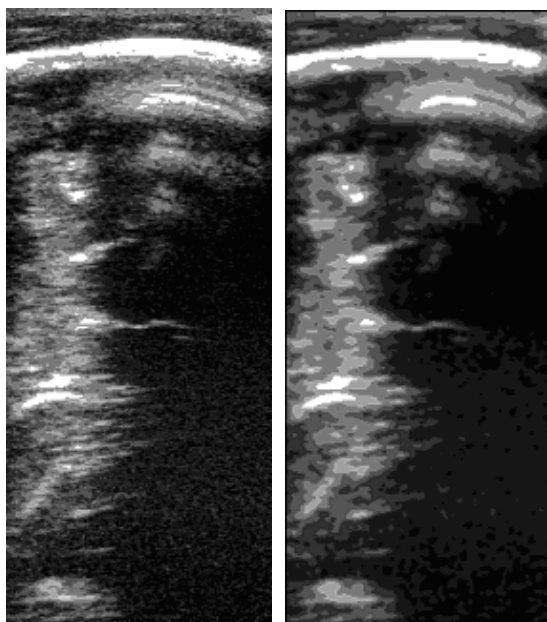


Рисунок 2 – Амплитудно-часові срези 50-ї сигнальної траси зображень на рис. 1, а-г відповідно
 (горизонтальна вісь – часова, вертикальна – відносна амплітуда)



а) б)
Рисунок 3 – Результат применения метода нечеткой кластеризации:
а – рис. 1, а; б – рис. 1, б

результатов диагностики посредством, например, варьирования числа классов.

ВЫВОДЫ

Нечеткая кластеризация является достаточно простым и удобным методом повышения эффективности процедуры визуального анализа ультразвуковых медицинских эхо-импульсных изображений. Необходимым условием успешного применения метода нечеткой кластеризации является нейтрализация влияния модуляци-

онных шумов. Метод селективного сингулярного разложения автоморфного отображения зашумленных изображений является эффективным способом фильтрации модуляционных шумов, поскольку не требует априорного знания их спектральных и статистических характеристик.

ПЕРЕЧЕНЬ ССЫЛОК

1. Физика визуализации изображений в медицине. Под ред. С. Уэбба. – М.: Мир. – 1991. – Т. 2. – 406 с.
2. Ахметшин А. М., Степаненко А. А. Повышение чувствительности ультразвуковой диагностики на основе метода параметрического спектрального анализа второго порядка. // Харьков: Клиническая информатика и телемедицина. – 2005. – Т. 2, № 1. – С. 98–100.
3. Форсайт Д., Понс Ж. Компьютерное зрение: современный подход. – М.: Вильямс. – 2004. – 926 с.
4. Леоненков А. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – Санкт-Петербург: БХВ-Петербург. – 2003. – 719 с.
5. Hall L. O., Bensaid A., Clarke L. P. A comparison of neural network and fuzzy clustering techniques in segmentation MRI of the brain // IEEE Trans. On Neural Networks. – 1992. – Vol. 3. – P. 672–682.
6. Применение ультразвука в медицине. Под ред. К. Хилла. – М.: Мир. – 1989. – 567 с.
7. Karatan M., Kutay A., Bozdagi G. An adaptive speckle suppression filter for medical ultrasonic imaging // IEEE Transaction on Medical Imaging. – 1995. – Vol. 14, No. 2. – P. 283–292.
8. Ахметшина Л. Г. Адаптивная фильтрация шумов в сигналах и изображениях: метод селективного сингулярного разложения автоморфного отображения // Искусственный интеллект. – 2005. – № 3. – С. 328–335.

Надійшла 24.03.2008

Розглянуто новий підхід до підвищення ефективності процедури візуального аналізу ультразвукових медичних ехо-імпульсних зображень. Представлено результати експериментальної перевірки інформаційних можливостей методу.

A new approach to increasing efficiency of visual analysis ultrasonic medical pulse-echo image is considered. Results experimental investigation information possibilities of the new method are presented.

УДК 621.382.002

И. М. Прохорец

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ДЛЯ АНАЛИЗА СТАТИЧЕСКИХ ХАРАКТЕРИСТИК МНОГОЭЛЕМЕНТНЫХ ПОЛУПРОВОДНИКОВЫХ ДЕТЕКТОРОВ

На основании теории направленных графов рассмотрены математические модели стрип-детектора для измерения межстриповых сопротивлений на постоянном токе. Приведены примеры расчета совместных линейных уравнений для электрических цепей, используемых для измерения межстриповых сопротивлений.

© Прохорец И. М., 2008

ВВЕДЕНИЕ

Многоэлементные полупроводниковые детекторы, к которым относятся полосковые или стрип-детекторы, пиксельные детекторы, относятся к приборам, получив-

ших наибольшее распространение в спектрометрии и визуализации изображений из-за их высокого пространственного разрешения, малого уровня шумов и наличия хорошо отработанной технологии. Для кремниевых стрип-детекторов существует целый ряд критических параметров, определяющих их характеристики – амплитудное, временное и пространственное разрешение. К ним относятся межстриповая емкость и емкость стрипа относительно противоположной плоскости [1], напряжение полного обеднения обратносмещенных $p-n$ переходов и их обратный ток, межстриповое сопротивление и т. д. Непосредственное измерение характеристик стрип-детектора затруднено из-за многих факторов: шаг внешних электродов находится в диапазоне 25–100 мкм; емкости малы по величине и не всегда допускают прямое измерение; сопротивления большие и их невозможно измерить без отключения от исследуемого объекта. Усложняет ситуацию и тот факт, что модуль детектора обычно работает в полях излучения, содержащих нейтроны, гамма-кванты и заряженные частицы. Действие излучения приводит к изменению внутренней структуры полупроводника. Следствием этих изменений является увеличение тока утечки $p-n$ переходов [2]

$$\Delta I_{\text{lead}}^{\text{rad}} = \alpha_{\text{rad}} \Phi_{\text{rad}},$$

где Φ_{rad} – поток излучения, а коэффициент α_{rad} равен $(2-3) \cdot 10^{-17}$ А·см⁻² для заряженных частиц и $(5-8) \times 10^{-17}$ А·см⁻² для нейтронов [2].

Конструктивно кремневые стрип-детекторы состоят из отдельных обратносмещенных $p-n$ переходов, емкостей и интегрированных сопротивлений смещения последних. Имитационное моделирование методики измерения емкостей детектора проведено в работах отечественных [1] и зарубежных авторов [3, 4].

Целью настоящей работы является создание математических моделей кремниевого детектора для измерения его характеристик на постоянном токе. Необходимость в этом возникает и при моделировании влияния поверхностных токов утечки в обычном планарном или стриповом детекторе на основе широкозонных полупроводниковых материалов CdTe, CdZnTe и др.

Практическая ценность работы определяется необходимостью создания полупроводниковых детекторов для дозиметрии и измерения полей ионизирующего излучения.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МОДЕЛЕЙ ДЛЯ ИЗМЕРЕНИЯ СТАТИЧЕСКИХ ПАРАМЕТРОВ КРЕМНИЕВЫХ СТРИП- ДЕТЕКТОРОВ

Как было показано в работе [1], для построения математической модели детектора, состоящего из отдель-

ных обратно смещенных переходов, интегрированных сопротивлений смещения последних и емкостей, можно использовать направленные графы [5–9]. В этой работе для анализа схем для измерения статических параметров многоэлементных детекторов была также использована теория графов.

В направленном графе каждая ветвь V_{ij} соответствует двухполюсному базовому элементу модели – межстриповому сопротивлению, сопротивлению смещения или источнику напряжения. Направление V_{ij} совпадает с направлением действия на этом элементе фазовой переменной – тока или напряжения. Каждая вершина графа S_{ij} является точкой соединения двух или более базовых элементов и соответствует узлу модели. Ветвь называют инцидентной вершине, если она начинается или заканчивается в этой вершине. Используя идентичность, информация о графе в вычислительной машине представляется в виде матрицы инцидентий [1] с $S = (m + 1)$ вершинами и $V = n$ ветвями. Основной особенностью матрицы инцидентий является то, что она позволяет записать законы Кирхгофа для токов и напряжений в электрической цепи, выраженные через матрицу главных сечений D и матрицу главных контуров B [7]. Если электрическая цепь состоит из n ветвей и $m + 1$ узлов, то ее граф содержит m ребер и $n - m$ хорд. Для связного графа с $m + 1$ узлами и n ветвями максимальное количество линейно-независимых строк M матрицы B равно $M = n - m$ [8].

Любая матрица главных контуров может быть разделена на две подматрицы $B = [B_t 1]$, где B_t соответствует ветвям дерева, а единичная матрица 1 порядка $n - m$ соответствует хордам. Если в цепи есть независимые источники напряжения, то их также включают в матрицу главных контуров B , которую в данном случае называют дополненной и обозначают B_d . Матрица B_d может быть разделена на следующие подматрицы $B_d = [B_E B]$, где B_E соответствует независимым источникам напряжения.

АНАЛИЗ СХЕМ ДЛЯ ИЗМЕРЕНИЯ СТАТИЧЕСКИХ ПАРАМЕТРОВ ДЕТЕКТОРОВ

Межстриповое сопротивление и сопротивление смещения стрип-детектора обычно измеряют на постоянном токе. Для измерения межстрипового сопротивления собирается схема, состоящая из основного источника смещения детектора e_1 и дополнительного источника e_2 , подключенного к стрипу, оторванного от основного источника смещения (рис. 1).

В качестве ребер графа этой схемы необходимо взять ветви, соответствующие источникам смещения. Граф для схемы измерения межстриповых сопротивлений содержит 3 узла и 7 ветвей. Следовательно, макси-

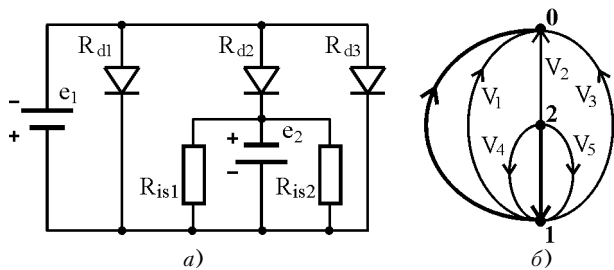


Рисунок 1 – Эквивалентная схема (а) и граф (б) для измерения межстрипового сопротивления с «оторванным» стрипом:

R_d – сопротивление обратно смещенного p-n перехода

Матрица	Контур	Ветви									
		e_1	e_2	V_1	V_8	V_2	V_3	V_3	V_5	V_6	V_7
B_D	1	1	0	1	0	1	0	0	0	0	0
	2	0	-1	-1	0	0	1	0	0	0	
	3	1	-1	0	0	0	0	1	0	0	
	4	0	1	0	1	0	0	0	1	0	
	5	0	1	0	1	0	0	0	0	1	
	6	1	0	0	1	0	0	0	0	0	

РАСЧЕТ ПАРАМЕТРОВ СТРИП-ДЕТЕКТОРОВ

Для определения численных параметров элементов детектора необходимо решить систему совместных линейных уравнений, выражающую в матричном виде второй закон Кирхгофа для электрической цепи:

$$B_D V = 0, \quad (1)$$

где V – вектор напряжений на ветвях графа. Для составления системы уравнений исследуемой цепи по методу контурных токов можно предположить, что напряжение на каждом пассивном элементе V_B выражается через его сопротивление Z_B и протекающий через него ток I_B в виде

$$V_B = Z_B I_B. \quad (2)$$

Тогда для всей цепи это соотношение можно записать в матричной форме

$$V_B = Z_B I_B, \quad (3)$$

где Z_B – диагональная матрица сопротивлений, I_B и V_B – векторы токов и напряжений, соответственно. Как показано в [6–9], токи I_X , протекающие через хорды, можно найти из системы уравнений

$$B Z_B B^t I_X = -B_E B_B, \quad (4)$$

где E_B – вектор известных источников напряжений в контурах. В свою очередь, вектор токов I_E , протекающих через источники напряжения, и вектор токов I_B можно найти из соотношений [6–9]

$$I_E = B_E^t I_X, \quad I_B = B^t I_X. \quad (5)$$

В качестве примера рассмотрим матрицу B_D , описывающую схему для измерения межстрипового сопротивления (рис. 3). Матрица B является единичной, следовательно, $B Z_B = Z_B$, $Z_B B^t = Z_B$. Вектор извест-

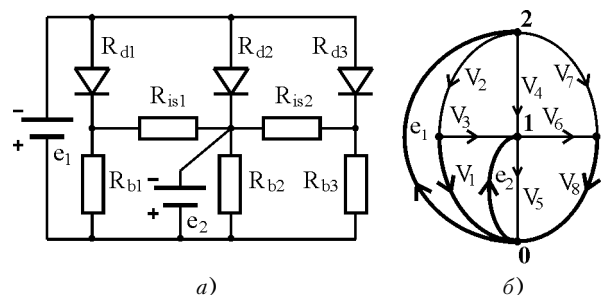


Рисунок 2 – Эквивалентная схема (а) и граф (б) для измерения межстрипового сопротивления без отсоединения сопротивления смещения:

R_b – сопротивление смещения; $R_d \gg R_b$

мальное количество линейно независимых уравнений для определения напряжения на элементах схемы равно 5, а матрица главных контуров B_D содержит 5 строк.

Матрица	Контур	Ветви						
		e_1	e_2	V_1	V_2	V_3	V_4	V_5
B_D	1	-1	0	1	0	0	0	0
	2	-1	-1	0	1	0	0	0
	3	-1	0	0	0	1	0	0
	4	0	-1	0	0	0	1	0
	5	0	-1	0	0	0	0	1

Для оценки межстриповых сопротивлений смещения без нарушения схемы собранного модуля стрип-детектора была предложена схема с использованием дополнительного источника напряжения e_2 . Этот источник совместно с сопротивлением смещения R_{b2} образует дополнительный главный контур (рис. 2).

Направленный граф схемы дает возможность записать для нее матрицу главных контуров, состоящую из 6 строк, и, следовательно, получить систему из 6 линейно независимых уравнений для определения неизвестных переменных.

ных источников равен $E_B = (e_1, e_2)^t$. Обозначим правую часть уравнения U . Тогда система уравнений будет иметь вид

$$Z_B I_X = U. \quad (6)$$

Так как матрица Z_B – диагональная, то система (6) для измерения межстрипового сопротивления с «оторванным стрипом» примет вид

$$E_B = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \text{ матрицы } B_E = \begin{pmatrix} -1 & 0 \\ -1 & -1 \\ -1 & 0 \\ 0 & -1 \\ 0 & -1 \end{pmatrix},$$

$$Z = \begin{pmatrix} R_{d1} & 0 & 0 & 0 & 0 \\ 0 & R_{d2} & 0 & 0 & 0 \\ 0 & 0 & R_{d3} & 0 & 0 \\ 0 & 0 & 0 & R_{is1} & 0 \\ 0 & 0 & 0 & 0 & R_{is2} \end{pmatrix}.$$

Рассмотрим правую часть системы уравнений (4)

$$B_E E_B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} e_1 \\ e_1 + e_2 \\ e_1 \\ e_2 \\ e_2 \end{pmatrix}.$$

Введем обозначение $-B_E E_B = U$. Тогда система уравнений (6) будет иметь вид $Z_B(k, k) I_X(k) = U(k)$, $k = 1 \dots 5$, т. е. она распадется на 5 независимых линейных уравнений, решение которых относительно неизвестных параметров не представляет сложности. Действительно, так как матрица Z_B – диагональная, то имеем

$$\begin{cases} R_{d1} I_{x1} = e_1, \\ R_{d2} I_{x2} = e_1 + e_2, \\ R_{d3} I_{x3} = e_1, \\ R_{is1} I_{x4} = e_2, \\ R_{is2} I_{x5} = e_2, \end{cases}$$

т. е. система (6) распадается на 5 независимых линейных уравнений и ее решение относительно неизвестных параметров имеет вид

$$I_x = \begin{pmatrix} I_{x1} \\ I_{x2} \\ I_{x3} \\ I_{x4} \\ I_{x5} \end{pmatrix} = \begin{pmatrix} e_1/R_{d1} \\ (e_1 + e_2)/R_{d2} \\ e_1/R_{d3} \\ e_2/R_{is1} \\ e_2/R_{is2} \end{pmatrix}.$$

Тогда

$$I_E = B_E^t I_x = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} I_{x1} \\ I_{x2} \\ I_{x3} \\ I_{x4} \\ I_{x5} \end{pmatrix} =$$

$$= \begin{pmatrix} -(I_{x1} + I_{x2} + I_{x3}) \\ -(I_{x2} + I_{x4} + I_{x5}) \end{pmatrix} = \begin{pmatrix} I_E(1) \\ I_E(2) \end{pmatrix},$$

где

$$I_E(1) = -\left(\frac{e_1}{R_{d1}} + \frac{e_1 + e_2}{R_{d2}} + \frac{e_1}{R_{d3}}\right);$$

$$I_E(2) = -\left(\frac{e_1 + e_2}{R_{d2}} + \frac{e_2}{R_{is1}} + \frac{e_2}{R_{is2}}\right).$$

При

$$R_{d1} = R_{d2} = R_{d3} = R_d, \quad I_E(1) = -\frac{2(e_1 + e_2)}{R_d}. \quad (7)$$

При

$$R_{is1} = R_{is2} = R_{is}, \quad R_{d2} \gg R_{is}, \quad I_E(2) = -\frac{2e_2}{R_{is}}. \quad (8)$$

Следовательно, в наиболее простом случае, когда можно предположить, что межстриповые сопротивления равны между собой, т. е. $R_{is1} = R_{is2}$, ток через источник e_2 связан с величиной сопротивлений простым соотношением (7), которое применялось для оценки межстриповых сопротивлений на ранней стадии исследования детекторов.

Решение системы уравнений (6) аналогичным способом для эквивалентной схемы на рис. 3 позволяет установить особенности измерения межстриповых сопротивлений в собранном модуле детектора.

Так, из рис. 3 видно, что, получив зависимость тока в цепи дополнительного источника смещения от его напряжения, можно определить значение межстрипового сопротивления. Из двух теоретических прямых, показанных на этом рисунке, следует, что при межстриповом сопротивлении $R_{is} = 50$ МОм и напряжении дополнительного источника 1 В изменение R_{is} на 1 МОм приводит к измерению тока в цепи источника на 0,5 нА. Это говорит о том, что чувствительность метода при указанных параметрах равна 0,5 нА/МОм.

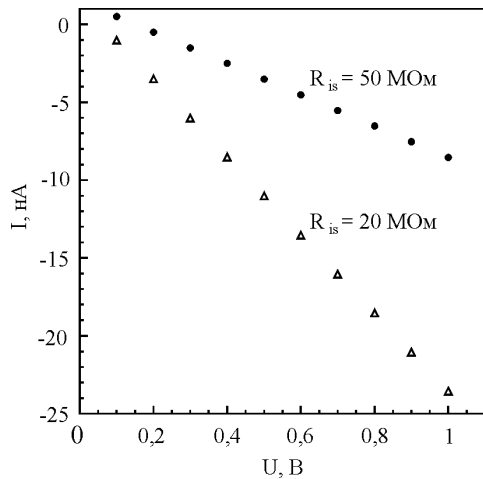


Рисунок 3 – Зависимость ток – напряжение для дополнительного источника смещения: R_{is} – межстриповое сопротивление

Хотя в некоторых частных случаях, аналогичных рассмотренному выше, решение системы (6) не представляет сложности, в общем случае необходимо применять специальные методы. Для решения системы линейных уравнений существует несколько общих методов, из которых наиболее часто используются метод исключения Гаусса, метод Зейделя. Конкретный алгоритм решения системы (6) зависит от набора параметров, характеризующих исследуемую схему, и размерности матрицы, описывающей граф схемы.

ТЕСТИРОВАНИЕ ПРОГРАММЫ ДЛЯ РАСЧЕТА СТАТИЧЕСКИХ ПАРАМЕТРОВ ДЕТЕКТОРА

На рис. 4 приведена тестовая схема и ее граф для расчета по методу контурных токов [10].

Параметры схемы: $R_1 = 6 \text{ Ом}$, $R_2 = 2 \text{ Ом}$, $R_3 = 1 \text{ Ом}$, $R_4 = 5 \text{ Ом}$, $R_5 = 4 \text{ Ом}$, $R_6 = 3 \text{ Ом}$, $e = 10 \text{ В}$. Число узлов $m + 1 = 5$, число ветвей $n = 7$, число неза-

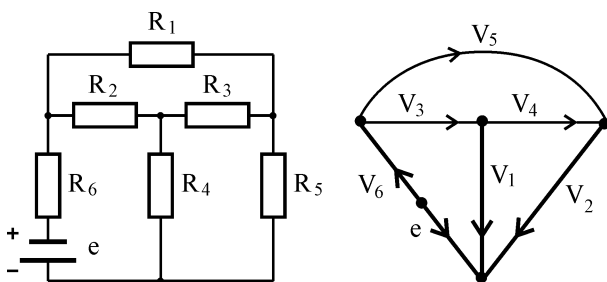


Рисунок 4 – Эквивалентная схема (а) и граф (б) для тестирования программы для расчета статических параметров детектора

висимых контуров $n - m = 3$. Матрица главных контуров имеет вид:

Матрица	Контур	Ветви						
		e	V_1	V_2	V_6	V_3	V_4	V_5
B_D	1	-1	1	0	1	1	0	0
	2	0	-1	1	0	0	1	0
	3	-1	0	1	1	0	0	1

Решая систему уравнений (6), получаем вектор токов $I_B =$

$$I_B = \begin{pmatrix} 0,694215 \\ 0,776859 \\ 1,471074 \\ 1,057851 \\ 0,363636 \\ 0,413223 \end{pmatrix}, \text{ совпадающий с приведенным в [10].}$$

ЗАКЛЮЧЕНИЕ

На основании теории направленных графов рассмотрены математические модели стрип-детектора для измерения межстриповых сопротивлений на постоянном токе. Приведены примеры расчета совместных линейных уравнений, выражающие закон Кирхгофа для электрической цепи, используемой для измерения межстриповых сопротивлений.

Приведенные в данной работе результаты исследований направлены на создание математических моделей и методов, сокращающих трудоемкость диагностического обеспечения современных полосковых детекторов на этапе тестирования их модулей. В работе показано, что для этих целей с успехом может применяться теория направленных графов, позволяющая создать топологические уравнения сложных электрических цепей.

ПЕРЕЧЕНЬ ССЫЛОК

1. Прохорец С. И., Хажмурадов М. А., Лукьянова В. П. Анализ емкостных параметров стрип-детекторов // Радиоэлектроника. Информатика. Управление. – 2004. – № 2. – С. 41–43.
2. Schwarz A. S. Heavy flavour physics at colliders with silicon strip vertex detectors // Physics reports Review (Section of Physics Letters). – 1994. – Vol. 238, No. 182. – P. 1–133.
3. Angarano M. M., Bader A., Greanza D. et al. Characterization and simulation of a single-sided, n+ on a n silicon microstrip detector before and after neutron irradiation // Nucl. Instr. and Meth. – 1999. – A 428. – P. 336–347.
4. Frautchi M. A., Hoferkamp H. R., Seidel S. C. Capacitance Measurements of Double-Sided Silicon Microstrip Detectors. CDF/DOC VTX/ CDF/2546. – The New Mexico Center for Particle Physics, Albuquerque, 1995. – 62 p.
5. Prokesch M., Szeles C. Accurate measurement of electrical bulk resistivity and surface leakage of CdZnTe radiation detector crystals // Journal of Applied Physics. – 2006. – No. 100. – P. 14503.

6. Атабеков Г. И. Основы теории цепей. – М.: Энергия, 1969. – 424 с.
7. Влах И., Сингхал К. Машинные методы анализа и проектирования электронных схем. – М.: Радио и связь, 1988. – 560 с.
8. Чуа Л. О., Лин Пен-Мин. Машинный анализ электронных схем. – М.: Энергия, 1980. – 640 с.
9. Алексеев О. В., Головкин А. А., Пивоваров И. Ю., Чавка Г. Г. Автоматизация проектирования радиоэлектронных средств. – М.: Высш. шк., 2000. – 480 с.
10. Геворкян Г. Х., Семенов В. Н. Электротехнические расчеты на языке Бейсик. – М.: Энергоатомиздат, 1989. – 184 с.

Надійшла 09.01.2008

На підставі теорії направлених графів розглянуто математичні моделі стрип-детектора для вимірювання міжстрипових опорів на постійному струмі. Наведено приклади розрахунку сумісних лінійних рівнянь для електричних ланцюгів, які використовуються для вимірювання міжстрипових опорів.

On the basis of graph theory mathematical models of strip detector for measurements of interstrip resistor were submitted. The examples of linear equation calculations for interstrip resistor measurements were reduced.

УДК 621.372.81

С. С. Самойлик, В. П. Бондарев

ДОБРОТНОСТЬ ПРЯМОУГОЛЬНОГО РЕЗОНАТОРА С ПЕРЕМЕННЫМ ЧИСЛОМ ДИЭЛЕКТРИЧЕСКИХ НЕОДНОРОДНОСТЕЙ

Расчет параметров прямоугольного резонатора со сложной формой диэлектрического заполнения решается методом интегральных уравнений макроскопической электродинамики. Проведенный численный анализ позволил вычислить добротности прямоугольного резонатора, для различных $H_{\text{мод}}$ типов колебаний в зависимости от параметров сложного диэлектрического заполнения.

ВВЕДЕНИЕ

Неотъемлемым элементом почти любой СВЧ экспериментальной установки является резонатор. Именно прогресс в их совершенствовании часто приводил к достижению качественно новых результатов. Так, появление мазеров было бы невозможно без применения высокодобротных СВЧ резонаторов. Высокодобротные резонаторы активно используются для сужения и стабилизации линии генерации, в качестве фильтров и дискриминаторов, в прецизионных физических экспериментах.

Потери электромагнитной энергии в резонаторе складываются из потерь в среде, заполняющей резонатор, и потерь в металлической оболочке резонатора. Кроме того, часть энергии из резонатора передается через элементы связи в устройства, связанные с резонатором.

Общие потери энергии в резонаторе:

$$W_{\text{п}} = W_{\text{мет}} + W_{\text{д}} + W_{\Sigma} + W_{\text{изл}}, \quad (1)$$

где $W_{\text{мет}}$ – энергия потерь за период колебания в оболочке резонатора; $W_{\text{д}}$ – энергия потерь в среде, запол-

няющей резонатор; W_{Σ} – энергия, отдаваемая резонатором во внешние устройства; $W_{\text{изл}}$ – энергия, теряемая на излучение.

Энергию потерь можно выразить как произведение мощности потерь, под которой подразумевается суммарная мощность потерь в металле, диэлектрике, а также мощность, отдаваемую резонатором во внешние устройства, и энергию, теряемую на излучение, на время, т. е. период колебания, соответствующий резонансу. Период колебаний связан с частотой колебаний $f_{\text{р}}$ обратным соотношением, в силу чего $W_{\text{п}} = P_{\text{п}\Sigma}/f_{\text{р}}$. При этом:

$$Q = 2\pi f_{\text{р}} \frac{W}{P_{\text{п}\Sigma}}, \quad (2)$$

где W – энергия, запасенная в колебательной системе.

С учетом (1) и (2), полную добротность (нагруженную) Q можно рассматривать как суперпозицию «частичных» добротностей:

$$\frac{1}{Q} = \frac{1}{Q_{\text{мет}}} + \frac{1}{Q_{\text{д}}} + \frac{1}{Q_{\Sigma}} + \frac{1}{Q_{\text{рад}}}, \quad (3)$$

где $Q_{\text{мет}} = 2\pi f_{\text{р}} \frac{W}{P_{\text{мет}}}$ – добротность, обусловленная по-

терями в стенках резонатора; $Q_{\text{д}} = 2\pi f_{\text{р}} \frac{W}{P_{\text{д}}}$ – добротность, обусловленная диэлектрическим заполнением

резонатора; $Q_{\Sigma} = 2\pi f_p \frac{W}{P_{\Sigma}}$ – внешняя добротность;

$Q_{\text{рад}} = 2\pi f_p \frac{W}{P_{\text{изл}}}$ – радиационная добротность.

Строгий расчет величины каждого из видов потерь в объемном резонаторе встречает большие трудности, ибо, как правило, не удастся найти решение трехмерного векторного волнового уравнения, если не пренебречь потерями в оболочке, в диэлектрическом заполнении, через элементы связи и т. д. Поэтому при анализе резонаторов обычно исходят из предположения, что небольшие общие потери, которые имеют место в резонаторе, не сказываются существенно на структуре полей в нем, т. е. предполагают, что в первом приближении структура поля в резонаторе с потерями и без них одинакова. В указанном приближении энергия, запасенная в резонаторе с малыми потерями и без потерь, практически одинакова. При этом потери в металле, среде, на излучение и потери, вызываемые передачей части энергии через элементы связи, можно рассчитывать независимо друг от друга.

1 ПОСТАНОВКА ЗАДАЧИ И МЕТОД РЕШЕНИЯ

В представленной статье расчет добротности прямоугольного резонатора со сложным диэлектрическим заполнением внутри закрытого резонатора проводится методом интегральных уравнений макроскопической электродинамики [1, 2]. Этот метод позволяет учитывать как комплексные параметры диэлектрического заполнения, так и его геометрическую форму.

В качестве математической модели выбран резонатор, содержащий диэлектрические пластины в форме параллелепипедов, с заданными комплексными диэлектрическими проницаемостями (рис. 1).

Электромагнитное поле в резонаторе удовлетворяет интегральным уравнениям для каждой из неоднородностей и может быть записано в следующем виде [1]:

– для первой диэлектрической неоднородности:

$$\begin{aligned} \vec{E}^{(1)}(\vec{r}, t) = & \vec{E}_0(\vec{r}, t) - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \int_0^t dt' \int_{V_1} G(\vec{r}, t, \vec{r}', t') \times \\ & \times [\varepsilon_1 - \varepsilon] \cdot \vec{E}^{(1)}(\vec{r}', t') d\vec{r}' - \\ & - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \int_0^t dt' \int_{V_2} G(\vec{r}, t, \vec{r}', t') \times \\ & \times [\varepsilon_2(t') - \varepsilon] \cdot \vec{E}^{(2)}(\vec{r}', t') d\vec{r}', \end{aligned} \quad (4)$$

– для второй диэлектрической неоднородности:

$$\begin{aligned} \vec{E}^{(2)}(\vec{r}, t) = & \vec{E}_0(\vec{r}, t) - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \int_0^t dt' \int_{V_1} G(\vec{r}, t, \vec{r}', t') \times \\ & \times [\varepsilon_1 - \varepsilon] \cdot \vec{E}^{(1)}(\vec{r}', t') d\vec{r}' - \\ & - \varepsilon_0 \mu_0 \frac{\partial^2}{\partial t^2} \int_0^t dt' \int_{V_2} G(\vec{r}, t, \vec{r}', t') \times \\ & \times [\varepsilon_2(t') - \varepsilon] \cdot \vec{E}^{(2)}(\vec{r}', t') d\vec{r}', \end{aligned} \quad (5)$$

где $\vec{E}_0(\vec{r}, t)$ – начальное поле в резонаторе, $G(\vec{r}, t, \vec{r}', t')$ – скалярная функция Грина прямоугольного резонатора; ε – диэлектрическая проницаемость свободного пространства; $\varepsilon_1, \varepsilon_2$ – диэлектрические проницаемости неоднородностей; ε_0, μ_0 – электрическая и магнитная постоянные.

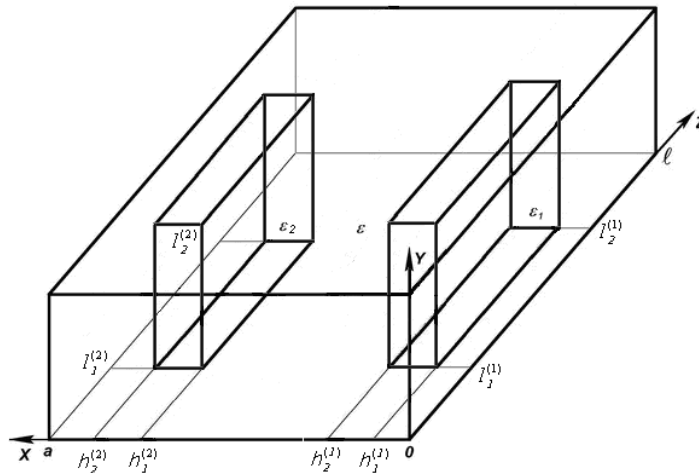


Рисунок 1 – Резонатор с диэлектрическим заполнением в виде двух диэлектрических пластин

Решение системы интегральных уравнений для напряженности электрического поля будем искать в виде собственных функций прямоугольной области:

$$E(\vec{r}, t) = \sum_{\alpha, \beta=1}^{\infty} E_{\alpha, \beta}(t) \cdot \sin\left(\frac{\alpha\pi}{a}x\right) \cdot \sin\left(\frac{\beta\pi}{l}z\right), \quad (6)$$

где a, l – размеры прямоугольного резонатора, $\alpha, \beta = 1, 2, 3, \dots$

Для рассматриваемой структуры функция Грина, в общем случае тензорная, имеет только диагональные элементы, которые должны удовлетворять граничным условиям на стенках резонатора, накладываемым на соответствующую компоненту поля. Поэтому функцию Грина $G(\vec{r}, t, \vec{r}', t')$ с условием ортогональности и нормировки в области можно представить в виде:

$$G(\vec{r}, t, \vec{r}', t') = \frac{4}{\varepsilon_1 \cdot a \cdot l \cdot \varepsilon_0 \mu_0} \times \\ \times \sum_{m, n=1}^{\infty} \frac{\sqrt{\varepsilon_1}}{\omega_{mn}} \sin\left(\frac{\omega_{mn}}{\sqrt{\varepsilon_1}}(t-t')\right) \cdot \Phi_{mn}(x, z) \cdot \Phi_{mn}(x', z'), \quad (7)$$

где $\omega_{mn} = c \cdot \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{l}\right)^2}$ – резонансная частота, Φ_{mn} – собственные функции области $0 \leq x \leq a, 0 \leq z \leq b$.

Подстановка такого представления поля (6) и функции Грина (7) для рассматриваемой структуры в (4), (5) приводит к системам алгебраических уравнений. Первая система является однородной. Вторая система является неоднородной и позволяет определить абсолютные значения амплитуд колебаний, возбуждаемых в резонаторе при заданном $E_0(r, t)$.

Рассмотрим однородную систему уравнений [3]:

$$\left\{ \begin{aligned} \sum_{q=1}^{\infty} \left[\left(\frac{(\varepsilon_1 - \varepsilon) \omega_r^2}{\varepsilon \omega_r^2 - \tilde{\omega}_p^2} A_{pq}^{(1)} + \delta_{pq} \right) U_q^{(+1)} + \right. \\ \left. + \left(\frac{(\varepsilon_2 - \varepsilon) \omega_r^2}{\varepsilon \omega_r^2 - \tilde{\omega}_p^2} A_{pq}^{(2)} \right) U_q^{(+2)} \right] = 0; \\ \sum_{q=1}^{\infty} \left[\left(\frac{(\varepsilon_1 - \varepsilon) \omega_r^2}{\varepsilon \omega_r^2 - \tilde{\omega}_p^2} A_{pq}^{(1)} \right) U_q^{(+1)} + \right. \\ \left. + \left(\frac{(\varepsilon_2 - \varepsilon) \omega_r^2}{\varepsilon \omega_r^2 - \tilde{\omega}_p^2} A_{pq}^{(2)} + \delta_{pq} \right) U_q^{(+2)} \right] = 0, \end{aligned} \right. \quad (8)$$

где $p = 1, 2, 3, \dots$

Решая задачу на собственные значения, определяем спектр и соответствующие отношения амплитуд возбуждаемых колебаний [3].

2 РАСЧЕТ ДОБРОТНОСТИ ЗАКРЫТОГО РЕЗОНАТОРА СО СЛОЖНЫМ ЗАПОЛНЕНИЕМ

Предположим, что связь резонатора с внешними устройствами полностью отсутствует, то есть $W_{\Sigma} = 0, Q_{\Sigma} = \infty$, и собственная добротность резонатора, как следует из (3), зависит от $Q_{мет}, Q_d, Q_{рад}$. В закрытых резонаторах радиационные потери отсутствуют, поэтому собственная (ненагруженная) добротность будет определяться выражением:

$$\frac{1}{Q_0} = \frac{1}{Q_{мет}} + \frac{1}{Q_d}. \quad (9)$$

Полная энергия, запасенная резонатором, определяется выражением [4]:

$$W = \frac{\mu_0}{2} \int_V (\vec{H})^2 dV = \frac{\mu_0}{2} \int_0^a \int_0^b \int_0^L (|H_x|^2 + |H_z|^2) dx dy dz. \quad (10)$$

Для колебаний H_{m0n} составляющие магнитного поля H_x и H_z , с учетом (6), примут вид:

$$\dot{H}_x = \frac{1}{i\omega\mu} \frac{\partial \dot{E}_y}{\partial z} = \frac{1}{i\omega\mu} \sum_{\alpha\beta=1}^{\infty} \frac{\beta\pi}{1} \times \\ \times E_{\alpha\beta} \sin\left(\frac{\alpha\pi}{a}x\right) \cdot \cos\left(\frac{\beta\pi}{1}z\right), \quad (11)$$

$$\dot{H}_z = \frac{1}{i\omega\mu} \frac{\partial \dot{E}_y}{\partial x} = -\frac{1}{i\omega\mu} \sum_{\alpha\beta=1}^{\infty} \frac{\alpha\pi}{a} \times \\ \times E_{\alpha\beta} \cos\left(\frac{\alpha\pi}{a}x\right) \cdot \sin\left(\frac{\beta\pi}{1}z\right). \quad (12)$$

Тогда для энергии W , запасенной прямоугольным резонатором, из (11) получим:

$$W = \frac{1}{\mu_0 \omega_p^2} \frac{alb}{8} \sum_{\alpha\beta=1}^{\infty} \left(\left(\frac{\alpha\pi}{a}\right)^2 + \left(\frac{\beta\pi}{l}\right)^2 \right) E_{\alpha\beta}^2. \quad (13)$$

Потери, вносимые диэлектрическим заполнением, можно представить в виде суммы

$$P_d = P_{d1} + P_{d2}, \quad (14)$$

где P_{d1} – мощность потерь в первой диэлектрической пластине, P_{d2} – мощность потерь во второй диэлектрической пластине.

Соответственно добротность, обусловленная диэлектрическим заполнением, может быть представлена в виде:

$$\frac{1}{Q_d} = \frac{1}{Q_{d1}} + \frac{1}{Q_{d2}}. \quad (15)$$

Мощность потерь в диэлектрике определяется выражением [4]:

$$W_d = \frac{\omega_p}{2} \varepsilon'' \varepsilon_0 \int_{V_1} |\vec{E}|^2 dV, \quad (16)$$

где ε'' – мнимая часть диэлектрической проницаемости, ω_p – резонансная частота.

Мощности потерь в N -й неоднородности, с учетом (6) и (16), будет определяться выражениями:

$$\begin{aligned} P_{dN} &= \frac{\omega}{2} \varepsilon_1'' \int_{V_1} |\vec{E}|^2 dV = \frac{1}{2} \omega \varepsilon_N'' \int_{h_1^{(N)} 0 l_1^{(N)}}^{h_2^{(N)} b l_2^{(N)}} |E_y|^2 dx dy dz = \\ &= \frac{1}{2} \omega \varepsilon_N'' \int_{h_1^{(N)} 0 l_1^{(N)}}^{h_2^{(N)} b l_2^{(N)}} \left| \sum_{\alpha\beta=1}^{\infty} E_{\alpha\beta} \sin\left(\frac{\alpha\pi}{a}x\right) \cdot \sin\left(\frac{\beta\pi}{l}z\right) \right|^2 dx dy dz. \end{aligned} \quad (17)$$

После преобразований и операций интегрирования получим окончательные выражения для определения мощности потерь в каждой из диэлектрических неоднородностей:

$$P_{dN} = \frac{b}{2} \omega \varepsilon_N'' \varepsilon_0 \sum_{\alpha, \beta=1}^{\infty} \sum_{m, n=1}^{\infty} E_{\alpha\beta} E_{mn} I_{m\alpha}^{(N)} \cdot J_{n\beta}^{(N)}, \quad (18)$$

где N – номер неоднородности,

$$\begin{aligned} I_{m\alpha}^{(N)} &= \frac{a}{2\pi} \left\{ \frac{\sin\left(\frac{(m-\alpha)\pi}{a}h_2^{(N)}\right)}{(m-\alpha)} - \frac{\sin\left(\frac{(m+\alpha)\pi}{a}h_2^{(N)}\right)}{(m+\alpha)} - \right. \\ &\quad \left. - \frac{\sin\left(\frac{(m-\alpha)\pi}{a}h_1^{(N)}\right)}{(m-\alpha)} + \frac{\sin\left(\frac{(m+\alpha)\pi}{a}h_1^{(N)}\right)}{(m+\alpha)} \right\}, \end{aligned}$$

$$\begin{aligned} J_{n\beta}^{(N)} &= \frac{1}{2\pi} \left\{ \frac{\sin\left(\frac{(n-\beta)\pi}{l}l_2^{(N)}\right)}{(n-\beta)} - \frac{\sin\left(\frac{(n+\beta)\pi}{l}l_2^{(N)}\right)}{(n+\beta)} - \right. \\ &\quad \left. - \frac{\sin\left(\frac{(n-\beta)\pi}{l}l_1^{(N)}\right)}{(n-\beta)} + \frac{\sin\left(\frac{(n+\beta)\pi}{l}l_1^{(N)}\right)}{(n+\beta)} \right\}, \end{aligned}$$

при $m = \alpha$ и $n = \beta$ получим:

$$I_{m\alpha}^{(N)} = \frac{1}{2} \left(h_2^{(N)} - h_1^{(N)} - \frac{a}{2m\pi} \left(\sin\left(\frac{2m\pi}{a}h_2^{(N)}\right) + \sin\left(\frac{2m\pi}{a}h_1^{(N)}\right) \right) \right),$$

$$J_{n\beta}^{(N)} = \frac{1}{2} \left(l_2^{(N)} - l_1^{(N)} - \frac{l}{2n\pi} \left(\sin\left(\frac{2n\pi}{l}l_2^{(N)}\right) + \sin\left(\frac{2n\pi}{l}l_1^{(N)}\right) \right) \right).$$

Тогда добротность, обусловленную диэлектрическим заполнением резонатора, можно вычислить из соотношения:

$$Q_{dN} = \frac{\frac{1}{\mu_0 \omega_p} \frac{alb}{8} \sum_{\alpha\beta=1}^{\infty} \left(\left(\frac{\alpha\pi}{a} \right)^2 + \left(\frac{\beta\pi}{l} \right)^2 \right) E_{\alpha\beta}^2}{\frac{b}{2} \omega \varepsilon_N'' \varepsilon_0 \sum_{\alpha, \beta=1}^{\infty} \sum_{m, n=1}^{\infty} E_{\alpha\beta} E_{mn} I_{m\alpha}^{(N)} \cdot J_{n\beta}^{(N)}}. \quad (19)$$

Мощность потерь, обусловленная конечной проводимостью стенок резонатора, определяется выражением [4]:

$$\begin{aligned} P_m &= \frac{1}{2} R_s \oint |\vec{H}|^2 dS = \\ &= R_s \left\{ \iint_{0,0}^{a,b} |H|_{z=0}^2 dx dy + \iint_{0,0}^{a,L} |H|_{y=0}^2 dx dz + \iint_{0,0}^{b,L} |H|_{x=0}^2 dy dz \right\}. \end{aligned} \quad (20)$$

После подстановки (11), (12) в выражение (20) и интегрирования по поверхности стенок резонатора получим:

$$P_m = \frac{1}{2} R_s \left\{ \frac{ab}{(\omega\mu_0)^2} \sum_{\alpha=1}^{\infty} \left(\sum_{\beta=1}^{\infty} \frac{\beta\pi}{l} E_{\alpha\beta} \right)^2 + \frac{1}{(\omega\mu_0)^2} \frac{al}{2} \sum_{\alpha\beta=1}^{\infty} \left(\left(\frac{\alpha\pi}{a} \right)^2 + \left(\frac{\beta\pi}{l} \right)^2 \right) E_{\alpha\beta}^2 + \frac{bl}{(\omega\mu_0)^2} \sum_{\beta=1}^{\infty} \left(\sum_{\alpha=1}^{\infty} \frac{\alpha\pi}{a} E_{\alpha\beta} \right)^2 \right\}, \quad (21)$$

где $R_s = \sqrt{\pi f_0 \mu_0 / \sigma_m}$ – поверхностное сопротивление металла на резонансной частоте.

Добротность, обусловленная потерями в металлических стенках резонатора, определяется выражением:

$$\begin{aligned} Q_m &= \frac{\frac{1}{\omega\mu_0} \frac{alb}{8} \sum_{\alpha\beta=1}^{\infty} \left(\left(\frac{\alpha\pi}{a} \right)^2 + \left(\frac{\beta\pi}{l} \right)^2 \right) E_{\alpha\beta}^2}{\frac{1}{2} R_s \left\{ \frac{ab}{(\omega\mu_0)^2} \sum_{\alpha=1}^{\infty} \left(\sum_{\beta=1}^{\infty} \frac{\beta\pi}{l} E_{\alpha\beta} \right)^2 + \frac{1}{(\omega\mu_0)^2} \frac{al}{2} \sum_{\alpha\beta=1}^{\infty} \left(\left(\frac{\alpha\pi}{a} \right)^2 + \left(\frac{\beta\pi}{l} \right)^2 \right) E_{\alpha\beta}^2 + \frac{bl}{(\omega\mu_0)^2} \sum_{\beta=1}^{\infty} \left(\sum_{\alpha=1}^{\infty} \frac{\alpha\pi}{a} E_{\alpha\beta} \right)^2 \right\}}. \end{aligned} \quad (22)$$

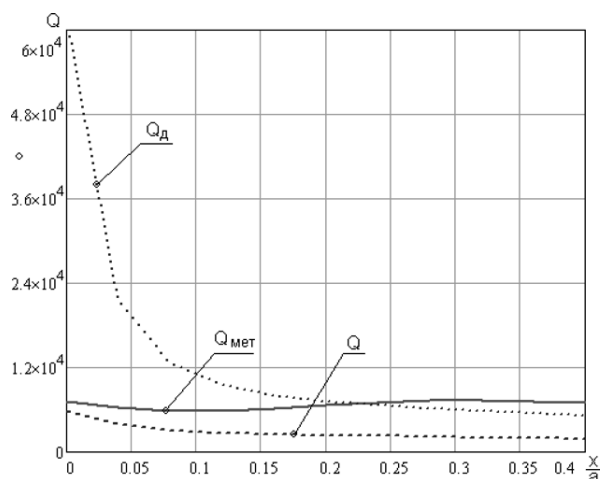


Рисунок 2 – Зависимость добротности от величины перемещения неоднородностей вдоль широкой стенки резонатора

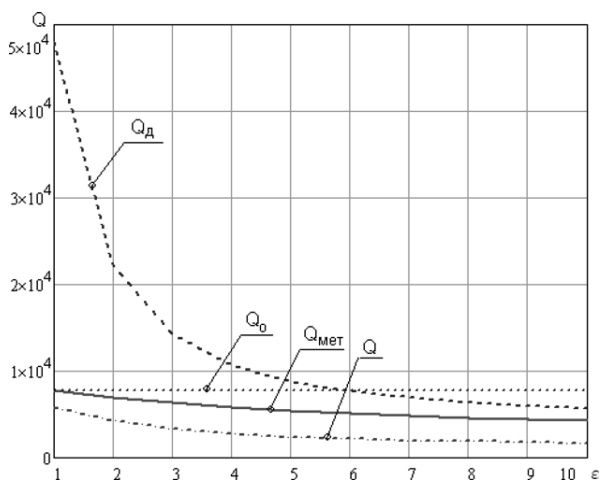


Рисунок 3 – Зависимость добротности от диэлектрической проницаемости пластин для первого типа волн

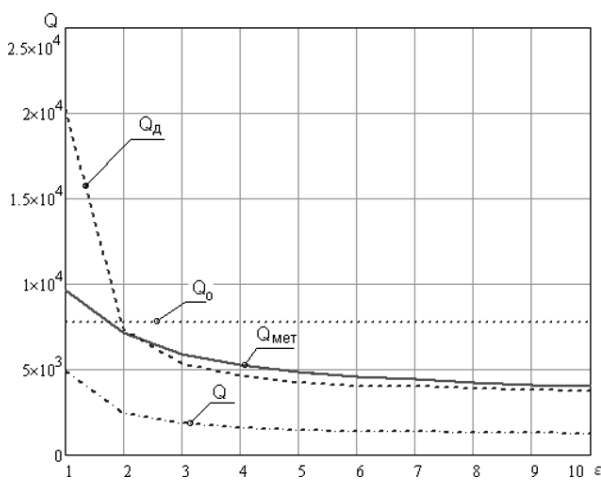


Рисунок 4 – Зависимость добротности от диэлектрической проницаемости пластин для второго типа волн

3 ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Так как размеры диэлектрических пластин сравнимы с размерами резонатора и резонансной длиной волны, то решение интегрального уравнения проводилось численно. Порядок алгебраических систем зависит от количества учитываемых собственных функций. При решении задачи количество собственных функций могло варьироваться в зависимости от параметров диэлектриков и их геометрических размеров.

Проведенный численный анализ позволил определить добротности для H_{m0n} типов колебаний в зависимости от расположения диэлектрических пластин в резонаторе и их диэлектрической проницаемости.

На рис. 2 показана зависимость добротности от местоположения пластин. При этом обе пластины с идентичными параметрами ($\epsilon = 4$, $\text{tg}\delta = 5 \cdot 10^{-4}$) и толщиной $0,1 \cdot a$, симметрично перемещались вдоль широкой стенки волновода из крайнего положения возле стенок, к центру резонатора, при этом диэлектрические пластины занимали 0,8 длины резонатора, геометрические размеры которого $23 \times 10 \times 25$ мм.

Добротность, обусловленная диэлектрическими неоднородностями у стенок резонатора, принимает наибольшее значение и резко убывает при перемещении пластин к центру резонатора, где напряженность H_{101} типа колебаний максимальна и, следовательно, максимальны диэлектрические потери. Таким же образом ведет себя и общая добротность. Добротность, обусловленная потерями в стенках резонатора (Q_M), убывает на отрезке $0 \dots 0,1$ после чего начинает возрастать т. к. диэлектрическая неоднородность оттягивает на себя поле от стенок резонатора и, соответственно, приводит к возрастанию добротности.

На рис. 3, 4 показаны зависимости добротностей от диэлектрической проницаемости пластин, при этом местоположение пластин было зафиксировано на расстоянии $0,1 \cdot a$ от боковых стенок резонатора и $0,1 \cdot L$ от продольных. Тангенсы угла диэлектрических потерь пластин выбраны равными $\text{tg}\delta = 5 \cdot 10^{-4}$. На рис. 3 представлены результаты расчетов добротности для первого типа колебаний, на рис. 2 – для второго типа колебаний. Закон изменения добротностей имеет одинаковый вид, но значения отличаются по величине (для первого типа колебаний добротности выше).

ВЫВОДЫ

Приведенный метод интегральных уравнений позволяет вычислять добротности различных H_{m0n} типов колебаний для произвольного числа диэлектрических неоднородностей в зависимости от их местоположения в резонаторе, диэлектрической проницаемости, геомет-

рической формы, а также размеров самого прямоугольного резонатора.

4. Капилевич Б. Ю., Требехин Е. Р. Волноводно-диэлектрические фильтрующие структуры: Справочник. – М.: Радио и связь, 1990. – 272 с.

Надійшла 25.02.2008
Після доробки 21.03.2008

ПЕРЕЧЕНЬ ССЫЛОК

1. Хижняк Н. А. Инженерные уравнения макроскопической электродинамики. – К.: Наукова думка, 1986. – 279 с.
2. Бухаров С. В. Расчет параметров частично заполненного резонатора при измерении диэлектрической проницаемости // Радиоэлектроника. – 2007. – № 6. – С.77–80 (Изв. вузов).
3. Samoilyk S. S., Bondaryev V. P. Complex frequencies of the rectangular resonator with the arbitrary number of the dielectric heterogeneities // 6th International Conference on Antenna Theory and Techniques. Conference Proceedings. – Sevastopol, 2007. – Pp. 1252–2054.

Розрахунок параметрів прямокутного резонатора зі складною формою діелектричного заповнення вирішується методом інтегральних рівнянь макроскопічної електродинаміки. Проведений чисельний аналіз дозволив обчислити добротності прямокутного резонатора для різних H_{m0n} типів коливань залежно від параметрів складного діелектричного заповнення.

The calculation of a rectangular resonator with a complex dielectric loading is solved by the method of integral equations of microscopic electrodynamics. The carried out numerical analysis enabled to calculate Q-qualities of the rectangular resonator, for various H_{m0n} types of fluctuations depending on parameters of complex dielectric filling.

УДК 621.396.6.004: 004.942

Г. М. Шило, Д. А. Коваленко, М. П. Гапоненко

ПРИЗНАЧЕННЯ НОРМАЛЬНИХ ДОПУСКІВ МЕТОДОМ ВІДОБРАЖЕНЬ

Розглядаються особливості формування допускових областей при нормальному законі розподілу параметрів та дії зовнішніх чинників. Розроблено алгоритм призначення номінальних допусків з урахуванням змін параметрів елементів при зовнішніх впливах. Проведені обчислення допусків на параметри елементів смугової фільтру.

ВСТУП

Задача призначення допусків на параметри елементів пов'язана з забезпеченням точності і стабільності при виробництві та експлуатації апаратури. Відхилення параметрів складаються з виробничих відхилень, спричинених недосконалістю технологічних процесів і неоднорідністю матеріалів, та відхилень, що викликані старінням матеріалів та впливом зовнішніх чинників в процесі експлуатації [1].

Розробка методів обчислення відхилень параметрів проводилась переважно без урахування дії зовнішніх впливів. Використовувались геометричні методи, в яких форма допускових областей визначалась законами розподілу параметрів елементів, а розміри областей обчислювались за координатами точок дотику цих областей до меж області працездатності [2–4]. Розроблено методи призначення інтервальних допусків, в яких закон розподілу параметрів задається на інтервалі, і допусків з нормальним законом розподілу параметрів.

Врахування зовнішніх впливів проводилось для лінійних або інтервальних моделей вихідних функцій [1, 5]. Одночасна дія кількох зовнішніх впливів на елементи електронних апаратів не враховувалась. Зокрема, не вибиралися найбільш несприятливі сполучення зовнішніх чинників.

Метою роботи є розробка методу призначення нормальних допусків з урахуванням дії зовнішніх впливів. Для досягнення поставленої задачі необхідно:

- розглянути особливості формування допускових областей при нормальному законі розподілу параметрів та дії зовнішніх чинників;
- розробити метод утворення інтервальних моделей зовнішніх впливів при найбільш несприятливому сполученні зовнішніх факторів;
- розробити алгоритм призначення нормальних допусків при заданих експлуатаційних обмеженнях області працездатності та визначених коефіцієнтах зовнішніх впливів.

1 ОСОБЛИВОСТІ ФОРМУВАННЯ ДОПУСКОВИХ ОБЛАСТЕЙ ПРИ ЗОВНІШНІХ ВПЛИВАХ

При нормальному законі розподілу параметрів допускова область приймає вигляд еліпсоїду [3]. Вплив

на параметри еліпсоїдної області одного зовнішнього фактора виражається співвідношенням [6]:

$$m = m_r d; \quad l = l_r d, \quad (1)$$

де m та l – математичні очікування параметрів елементів та напіввісі еліпсоїду при дії зовнішнього фактору; m_r та l_r – математичні очікування параметрів елементів та напіввісі еліпсоїду при нормальних умовах навколишнього середовища та відсутності старіння; $d = 1 + cz$ – відносне відхилення параметра елемента під дією зовнішнього впливу; c – коефіцієнт зовнішнього впливу; z – ширина інтервалу зміни зовнішнього впливу.

Для формування відносних відхилень параметрів при дії кількох зовнішніх чинників необхідно розглянути перетворення допускових областей під впливом кількох зовнішніх факторів. Ці перетворення показані на рис. 1, де x_1 і x_2 – параметри елементів; Ω_w і Ω_r – область працездатності і область номінальних допусків, яка утворюється при нормальних умовах навколишнього середовища. Области Ω_m та Ω_p формуються при нижньому та верхньому діапазонах зміни температур. Старінню параметрів елементів відповідає область Ω_s . Области Ω_{sm} і Ω_{sp} утворюються при одночасній зміні температури і старінні елементів. Гіперповерхні \underline{y}_r , \bar{y}_r , \underline{y}_e та \bar{y}_e формують межі області працездатності, які відповідають нижньому і верхньому номінальному та експлуатаційному відхиленню вихідної функції.

З рис. 1 видно, що зовнішні впливи призводять до дрейфу допускових областей, внаслідок чого значно розширюється область працездатності. Особливо значне розширення відбувається при сумісній дії зовнішніх факторів. Найбільш несприятливою є дія нижнього

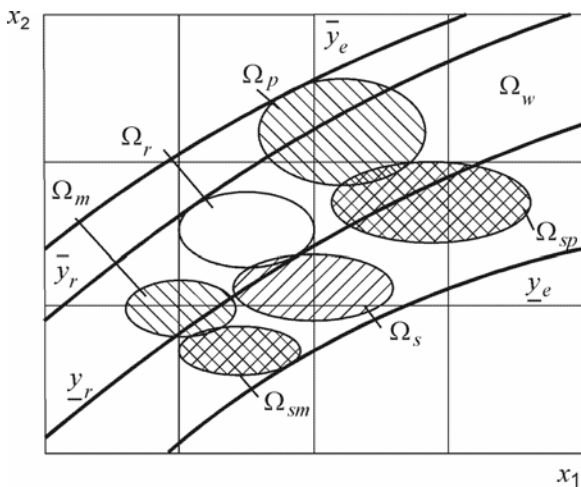


Рисунок 1 – Перетворення допускових областей при зовнішніх впливах

температурного діапазону і старіння, при якому утворюється допускova область Ω_{sm} . Верхня експлуатаційна межа області працездатності утворюється при дії верхнього температурного діапазону та відсутності старіння.

Формування інтервальних моделей відносних відхилень параметрів елементів при зовнішніх впливах доцільно проводити в точках дотику допускової області і гіперповерхонь області працездатності. Відносні відхилення параметрів оцінюються при одночасній дії зовнішнього чинника на всі параметри елементів в точках дотику. Для верхнього значення вихідної функції верхнє відносне відхилення визначається співвідношенням:

$$\bar{d}_{ij} = \begin{cases} 1 + c_{ij}z_{mj} & \text{якщо } y(\bar{X}_{mj}) > \bar{y}_r; \\ 1 + c_{ij}z_{pj} & \text{інакше,} \end{cases} \quad (2)$$

де z_m і z_p – ширина інтервалу зміни зовнішнього впливу в нижньому та верхньому діапазонах; $\bar{X}_{mj} = \{\bar{x}_{m1j}, \dots, \bar{x}_{mn}\}$ – множина координат точок дотику допускової області і верхньої межі області працездатності при дії j -го зовнішнього чинника; $\bar{x}_{mij} = \bar{x}_{ri} \times (1 + c_{ij}z_{mj})$; \bar{x}_{ri} – значення параметрів елементів в точці дотику допускової області і верхньої межі області працездатності при відсутності зовнішніх чинників.

Аналогічно визначаються нижні значення відносних відхилень при зовнішніх впливах:

$$\underline{d}_{ij} = \begin{cases} 1 + c_{ij}z_{mj} & \text{якщо } y(\underline{X}_{mj}) < \underline{y}_r; \\ 1 + c_{ij}z_{pj} & \text{інакше,} \end{cases} \quad (3)$$

де $\underline{X}_{mj} = \{\underline{x}_{m1j}, \dots, \underline{x}_{mn}\}$ – множина координат точок дотику допускової області і нижньої межі області працездатності при дії j -го зовнішнього впливу; $\underline{x}_{mij} = \underline{x}_{ri} \cdot (1 + c_{ij}z_{mj})$; \underline{x}_{ri} – значення параметрів в точці дотику допускової області і нижньої межі області працездатності при відсутності зовнішніх чинників.

Загальна відносна зміна параметрів елементів для кожної із меж області працездатності при сумісній дії кількох зовнішніх впливів має вигляд:

$$d_i = \prod_{j=1}^q d_{ij}, \quad (4)$$

де q – кількість зовнішніх впливів.

Співвідношення (2)–(4) дозволяють визначити відносні відхилення параметрів елементів при найбільш несприятливому сполученні зовнішніх чинників і мо-

жуть використовуватись в алгоритмах призначення номінальних допусків з урахуванням зовнішніх впливів.

2 СИНТЕЗ ДОПУСКІВ

Призначення номінальних допусків з урахуванням зовнішніх впливів відбувається при заданих експлуатаційних обмеженнях області працездатності, номінальних значеннях параметрів елементів, коефіцієнтах і діапазонах зміни зовнішніх впливів. Обмеження області працездатності може задаватись або межовими значеннями вихідної функції, або шириною інтервалу зміни вихідної функції впродовж життєвого циклу електронного апарату.

Коли призначаються допуски для заданого межового експлуатаційного значення вихідної функції, експлуатаційна межа гіперповерхня області працездатності є дотичною до еліпсоїдної допускової області, параметри якої сформовано з урахуванням дії зовнішніх впливів. При відсутності зовнішніх чинників точка дотику відображається на межу, що відповідає номінальному відхиленню вихідної функції. На початку процедури призначення допусків ця точка є невідомою. Тому для її визначення використовується алгоритм, геометрична інтерпретація якого при заданій верхній експлуатаційній межі області працездатності наведена на рис. 2.

Обчислення розпочинаються із визначення відхилень при межовому значенні вихідної функції \bar{y}_e . Для цього використовується алгоритм призначення нормальних допусків із заданим критерієм оптимізації [3]. Одночасно визначаються координати точки дотику \bar{B}_d

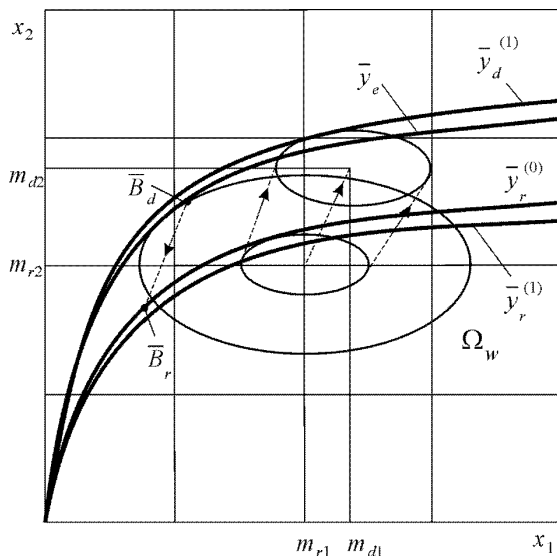


Рисунок 2 – Геометрична інтерпретація призначення допусків при зовнішніх впливах

граничної допускової області та межової гіперповерхні області працездатності:

$$\bar{X}_b = \{\bar{x}_{b1}, \dots, \bar{x}_{bn}\}.$$

Далі проводиться обчислення координат точки \bar{B}_r при відсутності дії зовнішніх чинників:

$$\bar{x}_{bri} = \bar{x}_{bi}/\bar{d}_i, \quad (5)$$

де значення \bar{d} визначається з використанням співвідношень (2), (4).

Після цього виконується перевірка умови реалізації алгоритму:

$$\bar{y}_r^{(0)} = y(\bar{X}_r^{(0)}) > y(m_r), \quad (6)$$

де $\bar{X}_r^{(0)} = \{\bar{x}_{br1}, \dots, \bar{x}_{brn}\}$.

Виконання умови (6) дозволяє призначити в першому наближенні номінальні допуски для меж області працездатності $\bar{y}_r^{(0)}$. Далі уточнюються результати з використанням ітераційного алгоритму, який містить кроки:

1. Отримані значення параметрів допускової області відображаються з допомогою співвідношень (1) в область зовнішніх впливів:

$$\bar{m}_{di} = m_r \bar{d}_i; \quad \bar{l}_{di}^{(k)} = l_{ri}^{(k)} \bar{d}_i.$$

2. Визначається граничне значення вихідної функції $\bar{y}_d^{(k)}$

$$\bar{y}_d^{(k)} = y(\bar{X}_d^{(k)}),$$

де $\bar{X}_d^{(k)}$ – точка дотику допускової області при зовнішніх впливах та межі області працездатності.

При цьому використовується алгоритм аналізу нормальних допусків [3].

3. Перевіряється умова завершення алгоритму:

$$\left| \frac{\bar{y}_d^{(k)} - \bar{y}_e}{\bar{y}_e} \right| \leq \varepsilon, \quad (7)$$

де \bar{y}_e – межове експлуатаційне значення вихідної функції; ε – точність обчислень.

При виконанні умови (7) – кінець роботи алгоритму.

4. Уточнюються значення номінальних меж області працездатності з допомогою співвідношень:

$$\Delta \bar{y}_d^{(k)} = \bar{y}_d^{(k)} - \bar{y}_e; \quad \bar{y}_r^{(k)} = \bar{y}_r^{(k-1)} - \frac{\Delta \bar{y}_d^{(k)}}{\bar{d}_y^{(k)}}, \quad (8)$$

де $\bar{d}_y^{(k)} = \bar{y}_d^{(k)} / \bar{y}_r^{(k-1)}$.

5. Призначаються номінальні параметри елементів. Використовується алгоритм призначення нормальних допусків.

6. Перехід до кроку 1.

Алгоритм призначення допусків при заданій нижній експлуатаційній межі області працездатності має аналогічний вигляд. Змінюється тільки умова реалізації алгоритму:

$$\underline{y}_r^{(0)} = y(\underline{X}_r^{(0)}) < y(m_r), \quad (9)$$

де $\underline{X}_r^{(0)} = \{x_{br1}, \dots, x_{brn}\}$; $x_{bri} = x_{bi}/d_i$; $\{x_{b1}, \dots, x_{bn}\}$ – координати точки дотику нижньої межі області працездатності та еліпсоїдної допускової області, сформованої при відсутності зовнішніх впливів, які забезпечуються межовими значеннями вихідної функції \underline{y}_e .

У випадку, коли обмеження області працездатності задається шириною інтервалу зміни вихідної функції, умова реалізації контролюється для кожної із меж області працездатності. Тоді алгоритм призначення номінальних допусків при нормальному законі розподілу параметрів набуває вигляду:

1. Призначаються відхилення параметрів елементів при заданій експлуатаційній ширині інтервалу зміни вихідної функції $w_{ye} = \bar{y}_e - \underline{y}_e$. Використовується алгоритм синтезу допусків із заданим критерієм оптимізації [2]. Одночасно визначаються координати точок дотику \underline{B}_d та \bar{B}_d граничної допускової області та межових гіперповерхонь області працездатності.

2. Проводиться відображення координат точок дотику в область номінальних значень з використанням співвідношень:

$$x_{bri} = \frac{x_{bi}}{d_i}; \quad \bar{x}_{bri} = \frac{\bar{x}_{bi}}{\bar{d}_i}.$$

Потім виконується перевірка умов реалізації (6) і (9).

3. Призначаються номінальні допуски в першому наближенні при заданих обмеженнях ширини інтервалу зміни вихідної функції:

$$w_{yr} = \bar{y}_r - \underline{y}_r, \quad (10)$$

де $\underline{y}_r = y(\underline{X}_r)$, $\bar{y}_r = y(\bar{X}_r)$.

4. Значення параметрів допускової області відображаються в області зовнішніх впливів з використанням співвідношень:

$$\underline{m}_{di} = m_{ri} d_i; \quad \underline{l}_{di}^{(k)} = l_{ri}^{(k)} d_i; \\ \bar{m}_{di} = m_r \bar{d}_i; \quad \bar{l}_{di}^{(k)} = l_{ri}^{(k)} \bar{d}_i.$$

5. Визначаються граничні значення $\underline{y}_d^{(k)}$, $\bar{y}_d^{(k)}$ і ширина відхилення вихідної функції при зовнішніх впливах:

$$w_{yd}^{(k)} = \bar{y}_d^{(k)} - \underline{y}_d^{(k)}.$$

Використовується алгоритм аналізу нормальних допусків [2].

6. Перевіряється умова завершення алгоритму:

$$\left| \frac{w_{yd}^{(k)} - w_{ye}}{w_{ye}} \right| \leq \varepsilon. \quad (11)$$

7. Уточнюється номінальне значення ширини вихідної функції:

$$\Delta w_{yd}^{(k)} = w_{yd}^{(k)} - w_{ye}; \quad w_{yr}^{(k)} = w_{yr}^{(k-1)} - \frac{\Delta w_{yd}^{(k)}}{d_{wy}^{(k)}}, \quad (12)$$

де $d_{wy}^{(k)} = w_{yd}^{(k)} / w_{yr}^{(k-1)}$.

8. Призначаються номінальні відхилення параметрів для зміненої номінальної ширини інтервалу вихідної функції. Використовується алгоритм призначення нормальних допусків [2].

9. Перехід до кроку 4.

Розроблені алгоритми дозволяють проводити обчислення допусків при заданих обмеженнях вихідної функції і найбільш несприятливому сполученні зовнішніх впливів. Можливе одностороннє обмеження експлуатаційного значення вихідної функції або обмеження експлуатаційної ширини інтервалу вихідної функції.

3 ЗАСТОСУВАННЯ МЕТОДУ

З використанням розробленого методу призначались допуски на параметри елементів активного смугового фільтра, схема якого надається на рис. 3.

Коефіцієнт загасання фільтра обчислюється з допомогою співвідношень [7]:

$$a(\omega) = \frac{1 + C_1/C_2}{\left(\frac{R_1}{R_2} \cdot \frac{C_1}{C_2}\right)} \sqrt{\left(\frac{1}{b_1 \omega} - \frac{b_2 \cdot \omega}{b_1}\right)^2 + 1}; \\ b_1 = (C_1 + C_2) \cdot \frac{R_0 R_2}{R_0 + R_2}; \quad b_2 = R_1 R_0 C_1 C_2,$$

де $\omega = 2\pi f$ – кутова частота; R_i , C_i – параметри елементів.

Номінальні значення параметрів елементів задавались у нормованому вигляді: $R_0 = 2$, $R_1 = 2$, $R_2 = 0,3$, $C_1 = 0,25$, $C_2 = 1$. Коефіцієнт загасання обчислювався на кутовій частоті $\omega = 1,6$. Забезпечувалась ширина експлуатаційного відхилення вихідної функції $w_{ye} = 10\%$. Враховувалась дія двох зовнішніх чинників: температури та старіння. Верхнє допустиме значення температури задавалось 60°C , а нижнє – -40°C . Тем-

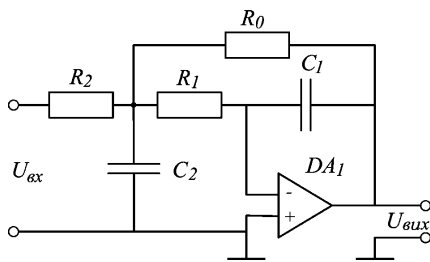


Рисунок 3 – Схема смугового фільтра

пературні коефіцієнти опору резисторів мали значення $100 \cdot 10^{-6} \text{ 1/K}$, а температурні коефіцієнти ємності конденсаторів – $-30 \cdot 10^{-6} \text{ 1/K}$. Мінімальний час напрацювання на відмову становив 15000 годин. Відхилення від номінальних значень параметрів резисторів і конденсаторів протягом часу напрацювання не перевищували 1 %.

Розрахунки проводились для максимального об'єму допускової області. Точність обчислень складала $\epsilon = 0,0001$. Оцінювання значень відхилень проводили при дії кожного зовнішнього чинника окремо та їх найбільш несприятливого сполучення. Через те, що номінальні відхилення резисторів та конденсаторів мають відповідати стандартизованим значенням за рядами перевагових чисел, далі призначали стандартизовані допуски. Для їх обчислення використовували алгоритм із ранжуванням відхилень і перерахунком їх неперевних значень [2].

Результати обчислень допусків наведені в табл. 1.

Таблиця 1 – Призначення допусків при зовнішніх впливах

Зовнішні впливи	Відхилення параметрів елементів, %				
	R_0	R_1	R_2	C_1	C_2
Відсутні	1,648	4,687	10,251	4,500	1,438
Температура	1,411	4,016	8,780	3,853	1,231
Старіння	0,966	2,755	6,020	2,644	0,843
Температура та старіння	0,729	2,080	4,544	1,996	0,636
Температура та старіння (стандартизовані допуски)	0,5	2,0	5,0	2,0	0,5

Як видно з табл. 1, дія зовнішніх впливів збільшує вимоги до точності елементів. Урахування зміни температури призводить до зменшення допусків на параметри в середньому на 14 % порівняно з випадком, коли впливи не враховуються. Якщо враховувати тільки старіння, то звуження допусків становить 41 %. Зменшення на 56 % відбувалось при комбінованій дії старіння та зміни температури.

ВИСНОВКИ

Розроблена інтервальні модель відносних відхилень параметрів, яка враховує одночасну дію зовнішнього впливу на всі елементи радіоелектронного пристрою. Ці моделі відповідають найбільш несприятливо-му сполученню дії зовнішніх чинників.

Призначення номінальних допусків відбувається з допомогою ітераційного алгоритму, в якому використовується відображення номінальних допускових областей в простір зовнішніх впливів. Корегування номінальної допускової області проводиться з урахуванням відносного відхилення вихідних функцій при зовнішніх впливах. Обчислюються нормальні допуски при нормальному законі розподілу параметрів, характерному для серійного виробництва електрорадіоелементів.

Результати обчислення допусків на параметри активного фільтра нижніх частот показали, що дія зовнішніх впливів призводить до зменшення допусків на параметри елементів порівняно з допусками, призначеними без урахування зовнішніх чинників. Найбільш впливає на зменшення допусків старіння елементів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Львович Я. Е., Фролов В. Н. Теоретические основы конструирования, технологии и надежности РЭА: Учеб. пособие для вузов. – М.: Радио и связь, 1986. – 192 с.
2. Воропай А. Ю., Гапоненко Н. П., Шило Г. Н. Назначение интервальных допусков с учетом особенностей элементной базы // Электроника и связь. – 2006. – № 2. – С. 11–15.
3. Шило Г. Н., Воропай А. Ю., Гапоненко Н. П. Расчет и назначение допусков методом касательных // Известия вузов «Радиоэлектроника». – 2006. – № 2. – С. 43–52.
4. Шило Г. Н. Геометрические методы назначения допусков // Проблемы управления и информатики. – 2007. – № 2. – С. 118–126.
5. Кришук В. М., Шило Г. М., Намлинський А. О., Гапоненко М. П. Вибір елементів при компенсації зовнішніх впливів // Радіоелектроніка. Інформатика. Управління. – 2004. – № 2. – С. 36–41.
6. Шило Г. М., Коваленко Д. А. Особливості компенсації зовнішніх впливів при призначенні нормальних допусків // Радіоелектроніка. Інформатика. Управління. – 2007. – № 1. – С. 44–47.
7. Знаменский А. Е., Теплюк И. Н. Активные RC-фильтры. – М.: Связь, 1970. – 280 с.

Надійшла 4.03.2008
Після доробки 18.03.2008

Рассматриваются особенности формирования допусковых областей при нормальном законе распределения параметров и воздействию внешних факторов. Разработан алгоритм назначения номинальных допусков с учетом изменений параметров элементов при внешних воздействиях. Проведены расчеты допусков на параметры элементов полосового фильтра.

Tolerance regions forming peculiarities under the normal distribution law and the external factors influence are being considered. An algorithm of a nominal tolerance assignment has been designed taking into account variations of elements parameters under external influences. The parameters tolerances calculations of active band-pass filter elements have been made.

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MATHEMATICAL AND COMPUTER MODELLING

УДК 519.65

Н. І. Біла, Н. О. Нечипоренко, Л. О. Бондаренко

ПРО РЕГУЛЯРИЗАЦИЮ ЧИСЕЛЬНОГО ДИФЕРЕНЦИУВАННЯ

Наводяться алгоритми, що дають найкраще за порядком наближення похідної таблично з аданої функції. Використання класичних методів квазірішення та нев'язки зводить задачу чисельного диференціювання до розв'язання задач лінійного програмування зі спеціальною структурою матриці обмежень, що дозволяє значно зменшити об'єм обчислень.

ПОСТАНОВКА ЗАДАЧІ

Розглянемо задачу обчислення похідної функції $f(t)$, що задана своїми значеннями \tilde{f}_k на рівномірній сітці $\{t_k\}_{k=1}^N$. Значення \tilde{f}_k відомі з похибкою ε , тобто мають місце нерівності

$$|\tilde{f}_k - f(t_k)| \leq \varepsilon, \quad k = \overline{1, N}. \quad (1)$$

Необхідно обчислити значення похідної $f'(t)$ у вузлах сітки $\{t_k\}_{k=1}^N$.

Відомо [1], що задача чисельного диференціювання функцій, значення яких можливо мати лише з якоюнебудь похибкою, є некоректною, тому потребує додаткової інформації та спеціальних алгоритмів для її ре-

гуляризації. Якщо значення ε відомо, маємо можливість регуляризувати задачу за методом нев'язки. Якщо ж значення ε не відомо, але є можливість отримати оцінки зверху другої похідної типу

$$\left| \frac{f(t_{k+1}) - 2f(t_k) + f(t_{k-1}))}{(t_{k+1} - t_k)^2} \right| \leq L_2, \quad k = 2, 3, \dots, N-1, \quad (2)$$

використовуємо метод квазірішень. Використання цих методів до задачі чисельного диференціювання у середньоквадратичній нормі наведено в [1] та [2]. В даній роботі наводяться алгоритми, що використовують норму простору l_∞ , що буває важливим для деяких практичних задач. Особливості алгоритмів дозволили створити комп'ютерні програми, що розв'язують задачу с заданими характеристиками якості [3].

МЕТОД НЕВ'ЯЗКИ

До значень функції \tilde{f}_k , $k = \overline{1, N}$, для яких виконуються нерівність (1) і та відомо значення похибки ε , за-

© Біла Н. І., Нечипоренко Н. О., Бондаренко Л. О., 2008

стосуємо алгоритм, що згладжує ці значення та базується на методі нев'язки. Значення $f'(t_k)$, $k = \overline{1, N}$ обчислюються за різницевою формулою з використанням згладжених значень. Вектор $F = \{f_1, f_2, \dots, f_N\}$ згладжених значень функції є рішенням такої екстремальної задачі:

$$\min_{f_k} \max_{2 \leq k \leq N-1} |f_{k-1} - 2f_k + f_{k+1}| \quad (3)$$

при обмеженнях

$$|f_k - \tilde{f}_k| \leq \varepsilon, \quad k = \overline{1, N}. \quad (4)$$

Задача (3)–(4) може бути зведена до вирішення такої задачі лінійного програмування:

$$\max x_{N+1} \quad (5)$$

при обмеженнях

$$\begin{aligned} x_k - 2x_{k+1} + x_{k+2} + x_{N+1} &\geq -B_k, \\ -x_k + 2x_{k+1} - x_{k+2} + x_{N+1} &\geq B_k, \\ k &= \overline{1, N-2}, \end{aligned} \quad (6)$$

$$\begin{aligned} 0 \leq x_k &\leq 2\varepsilon, \quad k = \overline{1, N}, \\ x_{N+1} &\geq 0, \end{aligned} \quad (7)$$

де $B_k = \tilde{f}_k - 2\tilde{f}_{k+1} + \tilde{f}_{k+2}$, $k = \overline{1, N-2}$; $x_k = f_k - \tilde{f}_k + \varepsilon$, $k = \overline{1, N}$.

Задача (5)–(7) вирішується методом послідовного покращення [4]. За початкове наближення беремо вектор $x_k = 0$, $k = \overline{1, N}$; $x_{N+1} = \max_k |B_k|$, тобто на першому кроці буде лише одне активне обмеження. На кожному наступному кроці число активних обмежень (а, значить, і активних змінних) збільшується на одиницю. На s -му кроці метода послідовних покращень необхідно розв'язувати дві системи алгебраїчних рівнянь порядку s . Обмеження (5) враховуються алгоритмічно, до активних входять лише обмеження з системи нерівностей (4). Тому матриця системи алгебраїчних рівнянь має таку структуру (при відповідному впорядкуванні рядків і стовпців системи): перші $s-1$ рядок мають кожен не більше 3 ненульових елементів, а останній рядок складається з s одиниць. Враховуючи вказану особливість матриці, можливо розв'язувати систему лінійних алгебраїчних рівнянь зі значною економією пам'яті та числа операцій. Розрахунки показали, що при розв'язанні системи число ненульових елементів збільшується менш, ніж у два рази.

Для розв'язання задачі лінійного програмування (5)–(7) можливо також використовувати метод послідовного покращання з перерахунком оберненої матриці. В цьому разі кількість операцій на кожному кроці, безперечно,

зменшується, однак, більші похибки перерахунку оберненої матриці при заміні рядків та стовпців призводять до значних відхилень у рішенні. В цьому сенсі метод послідовних покращань набагато точніше.

Матрицю обмежень (7) немає необхідності зберігати в пам'яті комп'ютера, необхідні значення елементів матриці обчислюються за допомогою виклику необхідної функції.

Після того, як знайдено вектор $X = \{x_1, x_2, \dots, x_{N+1}\}$ – розв'язок задачі (5)–(7), згладжені значення функції f_k обчислюються за формулою

$$f_k = \tilde{f}_k + x_k - \varepsilon, \quad k = \overline{1, N},$$

при цьому

$$\max_{1 \leq k \leq N-2} |f_{k+2} - 2f_{k+1} + f_k| = x_{N+1}.$$

Значення похідної $f'(t)$ у вузлах $\{t_k\}_{k=2}^{N-1}$ обчислюється за однією з різницевих формул

$$\begin{aligned} f'(t_k) &= \frac{f_{k+1} - f_k}{t_{k+1} - t_k}, \\ f'(t_k) &= \frac{f_{k+1} - f_{k-1}}{t_{k+1} - t_{k-1}}. \end{aligned}$$

Якщо відома оцінка другої похідної L_2 , то маємо таку оцінку похибки обчислення похідної:

$$\begin{aligned} \max_k \left| f'(t_k) - \frac{f_{k+1} - f_k}{t_{k+1} - t_k} \right| &\leq \\ &\leq \max_k \left\{ \left| f'(t_k) - \frac{f(t_{k+1}) - f(t_k)}{t_{k+1} - t_k} \right| + \right. \\ &+ \left| \frac{f(t_{k+1}) - f(t_k)}{t_{k+1} - t_k} - \frac{\tilde{f}_{k+1} - \tilde{f}_k}{t_{k+1} - t_k} \right| + \\ &+ \left. \left| \frac{\tilde{f}_{k+1} - \tilde{f}_k}{t_{k+1} - t_k} \right| \right\} \leq \frac{L_2 \Delta t}{2} + \frac{4\varepsilon}{\Delta t}, \end{aligned}$$

де $\Delta t = \max_k (t_{k+1} - t_k)$.

Ясно, що перед розв'язанням задачі, якщо N досить велике, належить збільшити крок сітки Δt , зменшуючи тим самим число N .

Проведені розрахунки з використанням наведеного алгоритму показали його ефективність.

Якщо про функцію відома деяка якісна інформація, а саме, інтервали, на яких функція зростає чи спадає, та (або) інтервали опуклості функції, то використання цієї інформації дозволить значно покращити якість апроксимації похідної. У цьому випадку до нерівностей (4) додамо такі:

$$(-1)^i (f_{k+1} - f_k) \geq 0, \quad k = \overline{1, N-1}, \quad i = \overline{1, m}, \quad (8)$$

$$(-1)^{n_j}(f_{k+1} - 2f_k + f_{k-1}) \geq 0, \quad k = \overline{2, N-1},$$

$$j = \overline{1, m_1}, \quad (9)$$

тут l_i та n_j можуть приймати значення 1 або 2 в залежності від зростання або спадання функції при $t = t_k$ і в залежності від виду опуклості функції відповідно. m – кількість інтервалів монотонності функції, m_1 – кількість інтервалів монотонності похідної функції.

Заміна змінних $x_k = f_k - \tilde{f}_k + \varepsilon$ зводить задачу (3), (4), (8), (9) до задачі лінійного програмування (5) при обмеженнях (6), (7) та

$$-x_k + x_{k+1} \geq (-1)^{l_i} d_k, \quad k = \overline{1, N-1},$$

$$i = \overline{1, m}, \quad (10)$$

$$x_k - 2x_{k+1} + x_{k+2} \geq (-1)^{n_j} B_k, \quad k = \overline{1, N-2},$$

$$j = \overline{1, m_1}, \quad (11)$$

де $d_k = \tilde{f}_k - \tilde{f}_{k+1}$.

Розв'язання останньої задачі лінійного програмування може бути здійснено так же, як і задачі (5)–(7). Матриці алгебраїчних систем зберігають свою структуру. В цьому випадку трохи збільшується кількість кроків методу послідовного покращання та порядок систем алгебраїчних рівнянь на останніх кроках.

МЕТОД КВАЗІРІШЕНЬ

Нехай для значень функції виконується нерівність (2) та константа L_2 відома. До заданих значень функції $\tilde{f}_k, k = \overline{1, N}$, застосуємо алгоритм, який оснований на методі квазірішень [2], що згладжує значення функції у вузлах. Значення похідної після цього обчислюються за різницевиими формулами. Вектор $F = \{f_1, f_2, \dots, f_N\}$ згладжених значень функції є рішенням такої екстремальної задачі

$$\min_{f_k} \max_{1 \leq k \leq N} |f_k - \tilde{f}_k| \quad (12)$$

при обмеженнях

$$|f_k - 2f_{k+1} + f_{k+2}| \leq L_2 h^2,$$

$$k = 1, 2, \dots, N-2. \quad (13)$$

Введемо такі позначення:

$$f_k - \tilde{f}_k = x_k, \quad \max_{1 \leq k \leq N} |x_k| = x_{N+1},$$

$$B_k = \tilde{f}_k - 2\tilde{f}_{k+1} + \tilde{f}_{k+2},$$

$$B_k^+ = -L_2 h^2 + B_k,$$

$$B_k^- = -L_2 h^2 - B_k.$$

Одержуємо задачу лінійного програмування:

$$\min x_{N+1} \quad (14)$$

при обмеженнях

$$-x_k + 2x_{k+1} - x_{k+2} = B_k^+, \quad k = \overline{1, N-2},$$

$$x_k - 2x_{k+1} + x_{k+2} \geq B_k^-, \quad k = \overline{1, N-2},$$

$$-x_k + x_{N+1} \geq 0, \quad x_k + x_{N+1} \geq 0,$$

$$x_{N+1} \geq 0, \quad k = \overline{1, N}. \quad (15)$$

Для того, щоб знайти початкове наближення, що задовольняє обмеженням (15), необхідно вирішити допоміжну задачу лінійного програмування. Задаємо $x_k = 0, k = \overline{1, N}$, та виділяємо індекси j , для яких $B_j^+ > 0$, та індекси i , для яких $B_i^- > 0$. Розв'язуємо задачу

$$\min \left\{ \sum_j (B_j^+ + (x_j - 2x_{j+1} + x_{j+2})) + \sum_i (B_i^- - (x_i - 2x_{i+1} + x_{i+2})) \right\}$$

при виконанні решти обмежень.

Розв'язання наведених задач, здійснюється методом послідовного покращання з перерахунком оберненої матриці. Зважаючи на розрідженість матриці обмежень, одержуємо алгоритм, що має на порядок менше кількість операцій, ніж повна задача лінійного програмування.

Аналогічно ставляться та розв'язуються задачі, що враховують обмеження на монотонність та опуклість функції.

ВИСНОВКИ

Як показали розрахунки, наведені алгоритми є досить ефективними. Методи нев'язки та квазірішень являються оптимальними за порядком точності, та для кожного з них має місце рівномірна збіжність обчисленої похідної до точної при $\varepsilon \rightarrow 0$.

ПЕРЕЛІК ПОСИЛАНЬ

1. Морозов В. А. Регулярные методы решения некорректно поставленных задач. – М.: Наука. гл. ред. физ.-мат. лит., 1987. – 240 с.
2. Гребенников А. И. Метод сплайнов и решение некорректных задач теории приближений. – М.: Изд-во МГУ, 1983. – 208 с.
3. Сергиенко И. В., Задирака В. К., Бабич М. Д. и др. Компьютерные технологии решения задач прикладной и вычислительной математики с заданными значениями характеристик качества // Кибернетика и системный анализ. – 2006. – № 5. – С. 33–41.

4. Булавский В. А., Звягина Р. А., Яковлева М. А. Численные методы линейного программирования. – М.: Наука, 1977. – 368 с.

Надійшла 20.11.2007
Після доробки 25.03.2008

Приводятся оптимальные по порядку алгоритмы приближения производной таблично заданной функции. Использование классических методов квази-решения и невязки сводит задачу численного дифференцирования к решению задач линейного программирования со специальной

структурой матрицы ограничений, которая позволяет значительно уменьшить объем вычислений.

The algorithms of the approximation the derivative of the function given by table are given. Classic methods quasi-solution and discrepancy are used to numerical differentiation and lead the task to linear programming problems with special structure of distraction matrix, what allow us to decrease greatly computational complexity.

УДК 621.365.036

А. Н. Довбня, С. Г. Удовенко, А. А. Шамраев

ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ МОДЕЛИ МАГНИТНОГО СПЕКТРОМЕТРА ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ

Разработан специализированный контроллер для цифрового управления магнитным спектрометром. По результатам активного эксперимента рассчитаны параметры модели анализирующего магнита, которая стала основой для синтеза адаптивного регулятора.

ВВЕДЕНИЕ

Одной из важных задач в технике линейных ускорителей является измерение энергетического спектра ускоренных частиц. Наиболее точным является метод измерений с использованием анализирующего магнита, устанавливаемого на выходе линейного ускорителя. Точность измерений определяется качеством системы управления магнитом, эффективность которой в значительной мере зависит от качества используемых при ее построении математических моделей, которые, с одной стороны, должны наиболее полно отражать свойства исследуемых объектов, а с другой быть удобными для реализации алгоритмов управления. Отсутствие полной информации об условиях функционирования объектов, а также об их динамических характеристиках и характере действующих помех обуславливают необходимость применения при управлении такими объектами адаптивного и робастного подходов, допускающих возможность использования при синтезе регуляторов упрощенных (в частности, линейных) моделей.

К основным методам построения математических моделей технических объектов можно отнести: эмпирический, который основывается на статистической обработке реальных данных, полученных в процессе функционирования объекта; аналитический, основанный на применении законов физики и химии; комбинированный, который объединяет рациональное планирование эксперимента, статистическую обработку экспери-

ментальных данных и основные физико-химические закономерности; автоматическое построение математической модели с помощью цифрового вычислителя, подключенного к объекту через датчики и преобразователи.

В данной работе рассматривается решение задачи динамической идентификации анализирующего магнита с применением специализированного контроллера в комплексе с персональной ЭВМ для автоматизированного измерения энергетического спектра пучка ускоренных электронов на выходе линейного ускорителя ЛУ-40.

ПОСТАНОВКА ЗАДАЧИ

В общем случае задача построения математической модели объекта состоит в выборе ее структуры и оценке ее параметров таким образом, чтобы при использовании критерия минимума некоторой функции разности расчетных и экспериментальных данных соблюдалось условие близости модели исследуемому процессу.

В соответствии с априорными данными об анализирующем магните как объекте управления, он был классифицирован как одномерный объект с самовыравниванием, наиболее эффективным методом определения параметров модели которого является активный эксперимент. Для получения передаточной функции по результатам эксперимента целесообразно использовать модифицированный метод площадей Симою [1].

СХЕМА ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА

На рис. 1 приведена часть схемы цифрового управления магнитным спектрометром, использованная для



Рисунок 1 – Система управления магнитным спектрометром



Рисунок 2 – Контроллер магнита

проведения активного эксперимента по определению динамических характеристик анализирующего магнита.

Текущее значение тока пучка, соответствующее плотности распределения электронов (выходной сигнал y), снимается с токоприемника, расположенного за анализирующим магнитом и оцифровывается восьмиразрядным аналого-цифровым преобразователем (АЦП) с частотой преобразования 30 МГц, который подключен к управляющей ЭВМ через шину ISA. Полученные данные принимаются ЭВМ для дальнейшей обработки и хранения.

Для обеспечения возможности дистанционного управления током магнита был разработан специализированный контроллер (рис. 2), формирующий сигнал опорного напряжения на входе схемы стабилизации тока анализирующего магнита (входной сигнал x). На-

пряжение формируется двенадцатиразрядным цифро-аналоговым преобразователем (ЦАП), обеспечивая дискретность шага по току магнита менее 0,1 А (0,05 % от максимального тока анализирующего магнита). Ядро системы реализовано на базе микроконтроллера MSP430F1611, основными задачами которого являются формирование опорного напряжения на выходе ЦАП, стабилизация тока анализирующего магнита по сигналу обратной связи с выхода бесконтактного датчика Холла и обеспечение связи с ЭВМ. Обмен данными с управляющей ЭВМ осуществляется по оптоволоконному кабелю в последовательном формате по протоколу RS232C. Применение оптического канала связи обеспечивает гальваническую развязку контроллера от линии связи, что требуется для поддержания соответ-

ствующего уровня безопасности и помехозащищенности системы управления спектрометром.

Электропитание магнита осуществляется от генератора постоянного тока ПН145 с независимым возбуждением. Для обеспечения стабилизации и регулировки тока магнита обмотка возбуждения генератора подключена к управляемому источнику тока, при этом величина тока в обмотке возбуждения генератора задается контроллером магнита. Управляемый источник тока построен на базе импульсного источника напряжения и линейного источника тока.

МЕТОДИКА ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА

Для проведения эксперимента по определению переходных характеристик объекта использовался цифровой запоминающий осциллограф Tektronix TDS2014B, первый канал которого подключался к входу АЦП (вход измерения тока анализирующего магнита), а второй канал к выходу ЦАП (управляющее воздействие, которое задает ток в обмотке возбуждения генератора). Весь диапазон изменения тока в обмотке возбуждения 0...600 мА был разбит на 12 интервалов с шагом 50 мА, внутри которых объект управления можно считать линейным. На рис. 3 показаны осциллограммы напряжений в указанных точках при задании уставки от 200 мА до 250 мА.

Результаты измерений использовались в качестве исходных данных для расчета параметров динамической модели объекта управления по модифицированному методу площадей Симоу.

ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА

Динамические свойства идентифицируемого объекта будем аппроксимировать моделью следующего вида:

$$W_M(s) = K \overline{W}_M(s) e^{-s\tau} = K \frac{1 + b_1 s + b_2 s^2 + \dots + b_m s^m}{1 + a_1 s + a_2 s^2 + \dots + a_n s^n} e^{-s\tau}, \quad (1)$$

где K – коэффициент усиления, τ – время запаздывания, a_i, b_i – коэффициенты передаточной функции.

Преобразуем модель (1) к нормированной передаточной функции с единичным коэффициентом усиления:

$$\overline{W}_M(s) = \frac{1 + b_1 s + b_2 s^2 + \dots + b_m s^m}{1 + a_1 s + a_2 s^2 + \dots + a_n s^n}, \quad (2)$$

$$K = \overline{W}_M(0) = 1. \quad (3)$$

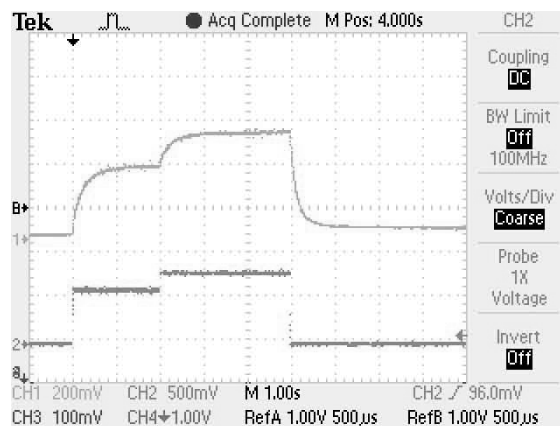


Рисунок 3 – Осциллограммы напряжений:
1 – отклик; 2 – уставка

Передаточная функция (2), полученная по кривой разгона, может быть однозначно преобразована к исходной модели (1).

Для получения кривой разгона в системе цифрового управления магнитным спектрометром, приведенной на рис. 1, устанавливается номинальный статический режим $x = x_n, y = y_n$. Затем система переводится в ручной режим (регулятор отключается, обратная связь разрывается) и на объект регулирования в начальный момент времени подается скачкообразное воздействие.

Величина запаздывания τ определяется непосредственно по кривой разгона, как время, за которое отклонение выходной величины $\Delta y(t)$ после подачи входного воздействия не превышает 0,5...1 % от $\Delta y_{уст}$.

Коэффициент усиления соответствует отношению:

$$K = \frac{\Delta y_{уст}}{\Delta x_{уст}}. \quad (4)$$

Для определения коэффициентов передаточной функции используем подход, предложенный в работе [2].

Рассмотрим инверсную нормированную передаточную функцию модели

$$\overline{W}_M^{-1}(s) = \frac{1}{\overline{W}_M(s)} = \frac{1 + a_1 s + a_2 s^2 + \dots + a_n s^n}{1 + b_1 s + b_2 s^2 + \dots + b_m s^m}. \quad (5)$$

Разложим $\overline{W}_M^{-1}(s)$ в ряд Тейлора в точке $s = 0$:

$$\begin{aligned} \overline{W}_M^{-1}(s) &= \frac{1 + a_1 s + a_2 s^2 + \dots + a_n s^n}{1 + b_1 s + b_2 s^2 + \dots + b_m s^m} = \\ &= 1 + S_1 s + S_2 s^2 + \dots + S_k s^k + \dots, \end{aligned} \quad (6)$$

где $\overline{W}_M^{-1}(0) = S_0 = 1$.

При известных площадях S_k можно определить коэффициенты передаточной функции a_i, b_i . Для этого умножим обе части равенства (6) на знаменатель $\overline{W}_M^{-1}(s)$, в результате чего получим

$$\begin{aligned} (1 + a_1s + a_2s^2 + \dots + a_ns^n) &= \\ = (1 + b_1s + b_2s^2 + \dots + b_ms^m) \times & \\ \times (1 + S_1s + S_2s^2 + \dots + S_ks^k + \dots). & \end{aligned} \quad (7)$$

После очевидных преобразований уравнения (7) получим степенной ряд

$$\begin{aligned} 1 + a_1s + a_2s^2 + \dots + a_ns^n &= \\ = 1 + (b_1 + S_1)s + (b_2 + b_1S_1 + S_2)s^2 + \dots + & \\ + \left(b_k + \sum_{i=1}^{k-1} b_iS_{k-i} + S_k \right) s^k \dots & \end{aligned}$$

Приравнявая в последнем равенстве коэффициенты при одинаковых степенях s слева и справа, получим линейную систему уравнений для определения коэффициентов модели a_i, b_i :

$$\begin{aligned} a_1 &= b_1 + S_1, \\ a_2 &= b_2 + b_1S_1 + S_2, \\ a_3 &= b_3 + b_2S_1 + b_1S_2 + S_3, \\ &\dots \\ a_k &= b_k + S_k + \sum_{i=1}^{k-1} b_iS_{k-i}. \end{aligned} \quad (8)$$

При определении порядка модели примем $m = n - 1$. Тогда

$$\begin{aligned} a_{k+1} &= a_{k+2} = \dots = 0, \\ b_k &= b_{k+1} = \dots = 0, \\ k &= n, n + 1, \dots \end{aligned} \quad (9)$$

Коэффициенты $a_i, b_i, i = 1, 2, \dots, n$ должны удовлетворять любому из уравнений (8). Сформируем систему из n уравнений (8) с учетом (9) для $k = n + 1, n + 2, \dots, 2n$:

$$\begin{aligned} 0 &= S_{n+1} + b_1S_n + b_2S_{n-1} + \dots + b_{n-1}S_2, \\ 0 &= S_{n+2} + b_1S_{n+1} + b_2S_n + \dots + b_{n-1}S_3, \\ &\dots \\ 0 &= S_{2n} + b_1S_{2n-1} + b_2S_{2n-2} + \dots + b_{n-1}S_{n+1}. \end{aligned} \quad (10)$$

Определитель системы (10) имеет вид:

$$\Delta_n = \begin{vmatrix} S_2 & S_3 & \dots & S_n & S_{n+1} \\ S_3 & S_4 & \dots & S_{n+1} & S_{n+2} \\ \dots & \dots & \dots & \dots & \dots \\ S_k & S_{k+1} & \dots & S_{n+k} & S_{n+k+1} \\ S_{n+1} & S_{n+2} & \dots & S_{2n-1} & S_{2n} \end{vmatrix}. \quad (11)$$

Из условия равенства нулю определителя (11), при котором система (10) имеет единственное решение, может быть определен порядок передаточной функции объекта.

При практических расчетах значения площадей S_i определяются с погрешностью, вызванной как неточностью определения переходной кривой, так и ошибками численного интегрирования. Поэтому условие $\Delta_n = 0$ может выполняться лишь приближенно.

Для оценки величины n целесообразно использовать критерий:

$$\frac{\Delta_n}{\Delta_{n-1}} \leq \delta, \quad (12)$$

где δ – некоторая малая величина.

Описанная методика была использована для обработки данных активного эксперимента на различных интервалах диапазона изменения тока в обмотке возбуждения анализирующего магнита. В результате получена нормированная передаточная функция вида (2), наилучшим образом аппроксимирующая динамические свойства идентифицируемого объекта, со следующими значениями коэффициентов, отличных от нуля: $a_0 = 1, a_1 = 2,764, a_2 = 2,136; b_0 = 1, b_1 = -0,152$. На рис. 4 приведены нормированные кривые разгона, построенные по экспериментальным и расчетным данным.

Предварительное исследование режимов работы магнитного спектрометра свидетельствует о квазистадио-

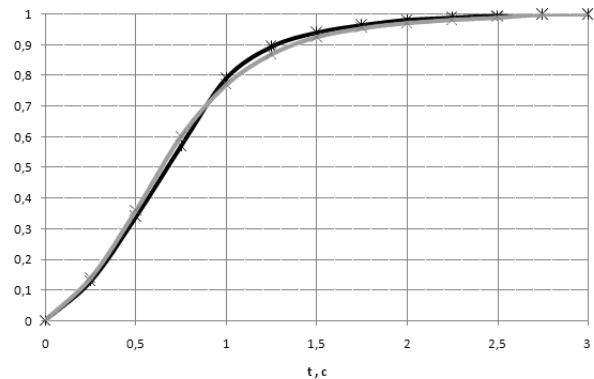


Рисунок 4 – Нормированная кривая разгона:

—*— экспериментальная кривая; —x— расчетная кривая

нарности его динамических характеристик. Возможность изменения характеристик идентифицируемого объекта во времени объясняется наличием неконтролируемых внешних и внутренних возмущений, действующих на систему автоматизированного измерения энергетического спектра пучка ускоренных электронов.

В связи с этим представляется целесообразным использование принципов адаптации в схеме цифрового управления магнитным спектрометром.

Квазистационарность свойств спектрометра делает возможным использование адаптивного байесовского идентификатора при синтезе цифрового регулятора [3]. Для оценки необходимости текущей коррекции коэффициентов передаточной функции анализирующего магнита предусмотрена вспомогательная процедура анализа данных, регистрируемых в соответствии со схемой, приведенной на рис. 1. В качестве начальной модели магнита может быть использована полученная в настоящей статье передаточная функция.

ВЫВОДЫ

Разработан специализированный контроллер для управления магнитным спектрометром, по экспериментальным данным рассчитаны параметры модели анализирующего магнита с использованием модифицированного метода площадей Симою. Полученная модель

использована для построения адаптивной системы цифрового управления магнитным спектрометром, испытания которой были проведены на базе линейного ускорителя электронов ЛУ-40 в НИК «Ускоритель» ННЦ «ХФТИ». Разработанная система управления показала удовлетворительные результаты при практическом применении в реальных измерениях. Дальнейшее развитие системы предполагает последовательное развитие ее программного обеспечения.

ПЕРЕЧЕНЬ ССЫЛОК

1. Симою М. П. Определение коэффициентов передаточных функций линеаризованных звеньев систем регулирования // Автоматика и телемеханика. – 1957. – № 6. – С. 514–527.
2. Волгин В. В. Методы расчета систем автоматического регулирования / Учебное пособие. – М.: Изд-во МЭИ, 1972. – 192 с.
3. Бодянский Е. В., Удовенко С. Г. Субоптимальное управление стохастическими процессами. – Харьков: Основа, 1997. – 140 с.

Надійшла 14.01.2008

Розроблено спеціалізований контролер для керування магнітним спектрометром. За результатами активного експерименту розраховано параметри моделі магніту, що стала основою для синтезу адаптивного регулятора.

The specialized controller for control of magnet spectrometer is developed. The parameters of analyzing magnet model will be used for adaptive regulator design are calculated by active experiment data.

УДК 681.3.06:330.322.54

В. И. Дубровин, О. И. Юськив

МОДЕЛИ И МЕТОДЫ ОПТИМИЗАЦИИ ВЫБОРА ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ

Рассмотрены основные составляющие оптимального портфеля (ожидаемая доходность портфеля и стандартное отклонение как мера риска), позволяющие агенту финансового рынка непрерывно реструктурировать портфель (максимизируя полезность промежуточного потребления и (или) конечного капитала) в соответствии со стохастически меняющимися инвестиционными возможностями. Проанализированы методы оптимизации (метод Марковица и Шарпа), а также рассмотрена роль функции полезности в формировании инвестиционного портфеля.

ВВЕДЕНИЕ

Финансовое инвестирование непосредственно связано с формированием инвестиционного портфеля. Финансовые рынки в современных условиях характери-

зуются нестационарными, стохастическими и кризисными явлениями различной природы [1, 2, 3]. В таких условиях традиционная портфельная теория [4] и классические методы финансовой математики [5] оказываются неадекватными и неспособными объяснить как поведение финансовых временных рядов, так и несоответствие практических рекомендаций по размещению капитала в рисковые активы теоретическим предсказаниям [6]. Кроме того, инвестирование неотделимо от потребления, а инвестиционная стратегия требует динамической реструктуризации портфеля. Поэтому возникает необходимость развития методов моделирования оптимального размещения капитала в рисковые активы в условиях изменения их доходности.

При осуществлении портфельного инвестирования перед субъектами инвестирования возникают проблемы эффективного вложения финансовых ресурсов. Важной проблемой при этом является выделение отдельных этапов процедуры сравнения экономической эффективности портфельного инвестирования и определения комплекса мер, которые непосредственно влияют на безопасность инвестиционных вложений.

Осуществляя инвестиции на рынке ценных бумаг, коммерческие банки формируют собственные или клиентские портфели ценных бумаг. Инвестиционный портфель – это набор ценных бумаг, приобретенных для получения доходов и обеспечения ликвидности. Управление портфелем заключается в поддержке равновесия между ликвидностью и прибыльностью.

Преимуществом портфельного инвестирования является возможность выбора портфеля для решения специфических инвестиционных проблем. Для этого используются различные виды портфелей ценных бумаг, в каждом из которых будет собственный баланс между риском, принятым для собственника портфеля, и ожидаемой отдачей (доходом) в определенный период времени [7]. Соотношение этих действий и дает возможность определить тип портфеля ценных бумаг. Тип портфеля – это его инвестиционная характеристика, которая базируется на соотношении дохода и риска.

1 РОЛЬ ФУНКЦИИ ПОЛЕЗНОСТИ В ПРИНЯТИИ РЕШЕНИЯ

За последние десятилетия в теории принятия решений интенсивно развивается новое направление, основанное на использовании функции полезности. Это направление основано на соединении некоторых положений теории принятия решений, методов системного анализа, исследования операций и методов квалиметрии. В основу его развития были положены исследования по теории полезности Джона фон Неймана и Morgenштерна [6]. При принятии решений обязательно учитывается неопределенность, связанная со статистическим характером анализируемых явлений, с неполнотой информации, отражаемой значениями тех или иных факторов [6].

При принятии решений человек оказывается перед выбором различных благ. Функция полезности вводится для того, чтобы одним числом описать степень удовлетворения, полученного лицом, принимающим решение.

Функцию полезности $U(x_1, x_2, \dots, x_N)$ можно представить как потребность, выраженную в количестве x_1, x_2, \dots, x_N , которую человек хочет удовлетворить.

Представление полезности в виде числа является удобным количественным выражением качественного исходного отношения предпочтения.

Рассмотрим общую схему определения функции полезности. Пусть X – множество точек $x = (x_1, x_2, \dots, x_N)$ пространства E_N .

Принятие решений осуществляется по такой схеме.

1) Ставится цель, например: получение прибыли, снижение затрат и т. д.

2) Определяются способы достижения цели, т. е. альтернативы, например: вложить наличную сумму x_0 в банк под гарантированный процент или инвестировать в проект, по которому с вероятностью 0,5 сумма удваивается до $2x_0$ или уменьшается до px_0 ; строить одно большое предприятие или два малых, или инвестировать в другой вид деятельности и т. п.

3) Проводится анализ решений, то есть возможных исходов (последствий) выбранной альтернативы.

Обозначим через a допустимую альтернативу, а через A – множество всех допустимых альтернатив. Каждому действию a из A ставится в соответствие n числовых показателей $x_1(a), \dots, x_N(a)$, то есть показатели x_1, x_2, \dots, x_N отображают каждое a из A в точку $x = (x_1, x_2, \dots, x_N)$ исходов (последствий) действий N -мерного пространства E_N . Выбирается та альтернатива, последствия которой обладают более предпочтительным, с точки зрения лица, принимающего решение, набором выделенных свойств [6].

Функция U , которая каждой точке $x = (x_1, x_2, \dots, x_N)$ пространства последствий ставит в соответствие действительное число $U(x)$, представляет структуру предпочтений лица, принимающего решение. Если U – функция полезности, отражающая предпочтение лица, принимающего решение, то задача о принятии решений может быть сформулирована в форме задачи оптимизации: найти $a \in A$, которое максимизирует $U(x(a))$. Функция полезности используется для того, чтобы помочь лицу, принимающему решение, четко выразить свои предпочтения.

Рассмотрим функции полезности, которые используются при исследовании только одной цели в управлении производством – показателя, отражающего доход от произведенной сделки.

На рис. 1, *a* изображена линейная функция. С ее помощью оценивается прирост полезности, происходящий пропорционально величине прироста капитала. Такой прирост происходит с одним и тем же коэффициентом пропорциональности a вне зависимости от того, сколько капитала используется в производстве. Это означает, что происходит насыщение полезности капитала.

Функция, представленная на рис. 1, *б*, описывает поведение лица, принимающего решение (ЛПР), относящегося осторожно к результату принятого решения. Чем больше средств имеется в распоряжении ЛПР, тем меньше оно заинтересовано в увеличении их объема путем дополнительного инвестирования.

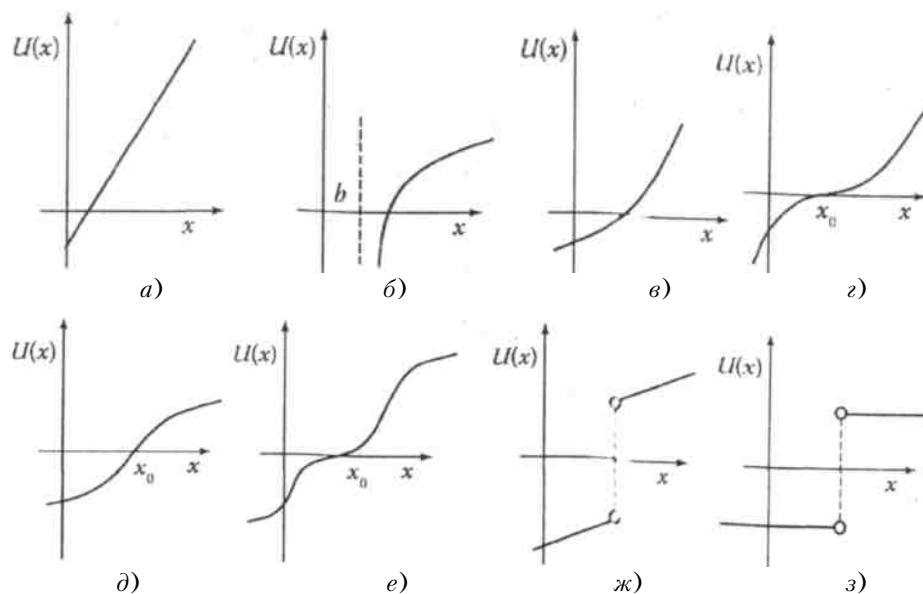


Рисунок 1 – Графики функции полезности

На рис. 1, в изображена функция полезности, которая описывает психологию мышления азартного ЛПР. Полезность дополнительной единицы прибыли для него увеличивается с ростом прибыли. Полученному доходу приписывается значительно большая ценность, чем она объективно существует. Лицо с такой функцией полезности неадекватно реагирует на потери, оно склонно преуменьшать размеры потерь.

ЛПР с функцией полезности, приведенной рис. 1, г, склонно преувеличивать предельную полезность дохода при его больших значениях и уменьшать при больших значениях проигрыша. При $x < x_0$ дополнительная единица дохода имеет меньшую ценность, чем предыдущая, а при $x > x_0$ – большую.

ЛПР с функцией полезности, изображенной на рис. 1, д, преуменьшает проигрыши и преувеличивает выигрыши. Эта функция описывает психологию ЛПР, отношение которого к большим выигрышам и к большим потерям носит осторожный характер.

Функция, изображенная на рис. 1, е описывает поведение «нормального» ЛПР. При нескольких больших по абсолютной величине значениях аргумента проявляется его умеренная азартность и осторожность. При больших значениях аргумента проявляется безразличие ЛПР к потерям.

Функция (рис. 1, ж) описывает психологию ЛПР, который, кроме объективного учета выигрыша, постоянно добавляет положительную за выигрыш и отрицательную за проигрыш «премию». Эта ситуация характерна для ЛПР-игрока.

Лицо с функцией полезности, изображенной на рис. 1, з, учитывает только выигрыш не менее определенной величины. Дальнейшее его повышение для такого ЛПР не играет роли.

На рис. 2 приведены наиболее часто встречающиеся на практике виды функции полезности логарифмического типа [6]. Графики этих кривых существенно зависят от параметра b (рис. 2, а, б; где b соответственно равен 1,5; 2; 3), их аналитическое выражение можно записать таким образом:

$$U(x) = \log_b(x + b) - 1.$$

Основанием логарифма в этих функциях можно выбирать любое число $b > 1$, единицу отнимают для удобства анализа, чтобы полезность нулевого дохода была равна нулю.

Логарифмические кривые обладают свойствами, существенными для их применения в оценке инвестиционной и производственной деятельности.

С помощью таких функций легко учитывать порог разорения, так как соответствующие им кривые определяют большие отрицательные значения полезности при приближении к соответствующим порогам чувствительности (порогам разорения). Параметр b можно рассматривать как относительный показатель размеров финансовых ресурсов компании. Большие значения b указывают на большой капитал, которым владеет компания. Соответственно, при возрастании капитала порог разорения перемещается в сторону больших потерь, а полезность выигрыша заметно падает.

Обычно функции полезности формируются на основе рационального общепринятого поведения менеджеров с учетом риска и порога разорения. Но функцию полезности можно также определять из равносильного соотношения

$$B^{U+1} - b = x.$$

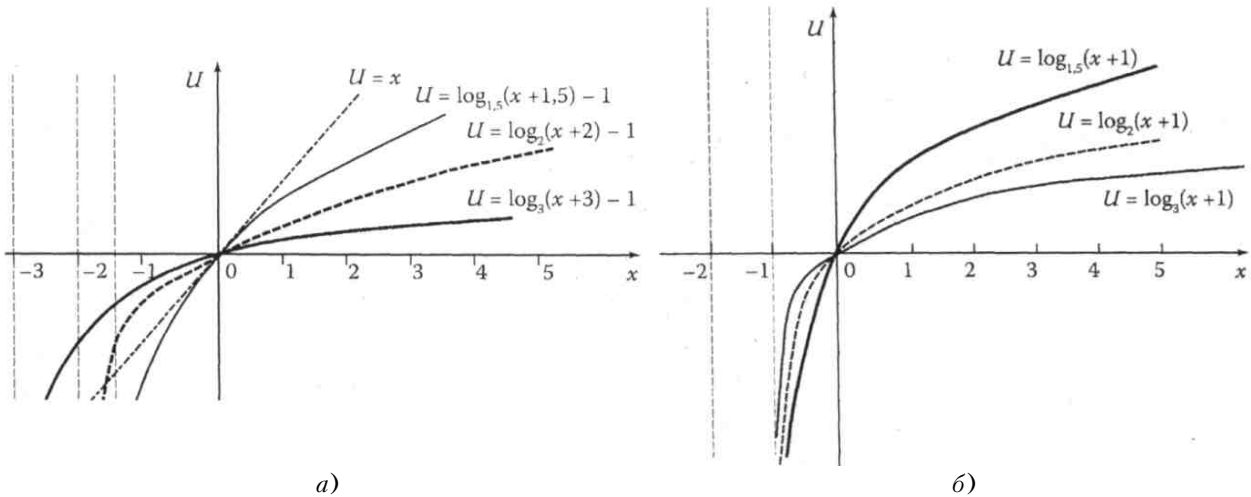


Рисунок 2 – Вид функции полезности логарифмического типа

Здесь параметр b характеризует финансовые ресурсы компании, весомость которых оценивается с точки зрения различных ЛПР.

На рис. 2, б изображены ситуации, когда менеджеры полностью согласны между собой относительно потерь (порог разорения $b = -1$), но их мнения резко расходятся в отношении полезности уровней выигрыша.

2 ФОРМИРОВАНИЕ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ С УЧЕТОМ ФУНКЦИЙ ПОЛЕЗНОСТИ

Для выбора наиболее оптимального портфеля в портфельной теории ценных бумаг используют кривые безразличия [7], которые являются монотонно возрастающими функциями. Для инвестора существует функция полезности, зависящая от двух аргументов: от ожидаемой доходности портфеля r_p и от стандартного (среднеквадратического) отклонения δ_p как меры риска:

$$U = U(r_p, \delta_p).$$

Все портфели, лежащие на одной линии безразличия, или линии уровня функции

$$U = U(r_p, \delta_p) = C,$$

являются равноценными для инвестора. Линии безразличия отражают отношение инвестора к риску и доходности портфеля и представляют собой кривые в координатах δ_p-r_p . Инвестор считает любой портфель, лежащий на линии безразличия выше и левее, более оптимальным, чем портфели, лежащие на линии безразличия, которая находится ниже и правее. Ожидаемая

доходность портфеля, состоящего из n ценных бумаг, равна [7]:

$$r_p = \sum_{i=1}^n x_i \cdot r_i,$$

где x_i – доля начальной стоимости портфеля, инвестированная в i -й вид ценных бумаг; r_i – ожидаемая доходность i -го вида ценных бумаг; n – количество видов ценных бумаг в портфеле.

Дисперсия доходности портфеля равна ковариации

$$\begin{aligned} D(r_p) &= \text{cov} \left(\sum_{i=1}^n x_i \cdot r_i, \sum_{j=1}^n r_j \cdot r_j \right) = \\ &= \sum_{i=1}^n \sum_{j=1}^n x_i \cdot x_j \cdot \text{cov}(r_i, r_j) = \sum_{i=1}^n \sum_{j=1}^n x_i \cdot x_j \cdot V_{ij}. \end{aligned}$$

Здесь $\text{cov}(r_p, r_j) = V_{ij}$ – ковариация ожидаемой доходности ценных бумаг i и j , вычисляется по формуле

$$\text{cov}(r_i, r_j) = \rho_{ij} \sqrt{D(r_i)D(r_j)} = \rho_{ij} \delta_j \delta_i, \quad i, j = 1, 2, \dots, n,$$

где ρ_{ij} – коэффициент корреляции между доходностями i -й и j -й ценных бумаг; D и δ – соответственно, дисперсия и стандартное (среднеквадратическое) отклонение доходностей портфеля ценных бумаг.

$$-1 \leq \rho_{ij} \leq 1.$$

Формула для стандартного отклонения портфеля имеет вид

$$\delta_p = \sqrt{D(r_p)}.$$

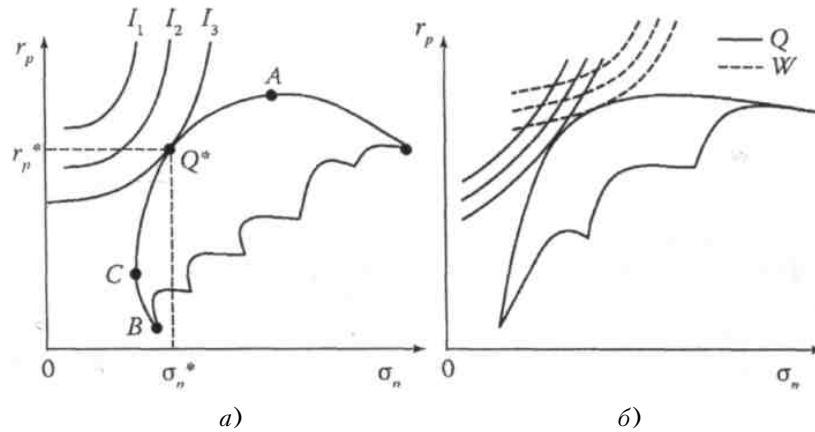


Рисунок 3 – Множество всех сформированных портфелей

На рис. 3, а показано достижимое множество, представляющее собой все портфели, которые можно сформировать из n видов ценных бумаг. Множество портфелей, обеспечивающих минимальный риск при меняющемся уровне ожидаемой доходности, находится на левой части границы достижимого множества, расположенного между точками A и B [7]. Инвестор выбирает свой оптимальный портфель, исходя из максимума ожидаемого дохода и минимума риска (среднеквадратического отклонения дохода). Инвестора удовлетворяют только портфели, находящиеся на верхней и левой границе достижимого множества.

Множество портфелей, которые может выбрать инвестор, представляет собой участок границы AC (рис. 3, а). Такое множество принято называть эффективным множеством портфелей, или множеством Парето. На этом множестве инвестор будет выбирать оптимальный с его точки зрения портфель.

Для выбора оптимального портфеля инвестор должен совместить свои линии безразличия с эффективным множеством. Оптимальный портфель будет соответствовать точке, в которой кривая безразличия касается эффективного множества: (портфель Q^* на кривой безразличия I_3).

Оптимальный портфель Q^* существенно зависит от формы линий безразличия, которые, в свою очередь, зависят от функции полезности, являющейся характеристикой стратегии инвестора. Если инвестор осторожен и стремится к уменьшению риска за счет снижения ожидаемой полезности доходности, то линии безразличия являются менее выпуклыми вниз (семейство линий Q на рис. 3, б).

Повышение риска инвестора для достижения более высокого уровня ожидаемой доходности выражается в том, что выпуклость вниз линий безразличия увеличивается (семейство линий W). Эта трактовка соответствует полезности Неймана – Моргенштерна, используемой для оценки действий инвестора.

3 ОПТИМИЗАЦИЯ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ ПО МОДЕЛИ Г. МАРКОВИЦА

В 1952 г. американский экономист Г. Марковиц опубликовал работу [8], которая легла в основу теории инвестиционного портфеля. Г. Марковиц исходил из предположения о том, что инвестирование рассматривается как однопериодовый процесс [7], т. е. полученный в результате инвестирования доход не реинвестируется. Другим важным исходным положением в теории Г. Марковица является идея об эффективности рынка ценных бумаг.

Марковиц считал, что инвестор, формируя свой портфель, оценивает лишь два показателя: $E(r)$ – ожидаемую доходность и σ – стандартное отклонение как меру риска. Следовательно, инвестор должен оценить доходность и стандартное отклонение каждого портфеля и выбрать наилучший портфель, который наиболее всего соответствует его требованиям – обеспечивает максимальную доходность r при допустимом значении риска σ . Какой при этом конкретный портфель предпочтет инвестор, зависит от его оценки соотношения «доходность – риск».

Цель любого инвестора – составить такой портфель ценных бумаг, который бы давал максимально возможную отдачу с минимально допустимым риском. Сравнение значений стандартных отклонений различных портфелей позволяет сделать два важных вывода: во-первых, при одних и тех же значениях коэффициентов корреляции между доходностями бумаг разных портфелей соответствуют разные величины σ , то есть при изменении соотношения ценных бумаг в портфеле меняется и риск портфеля. Во-вторых, для любого портфеля с понижением коэффициента корреляции бумаг уменьшается и риск портфеля (если, конечно портфель не состоит из одной ценной бумаги) [8].

Если брать различные количества ценных бумаг (3, 4, 5, ..., n), имеющих любые коэффициенты доходностей в пределах от (-1) до $(+1)$, и создавать из них

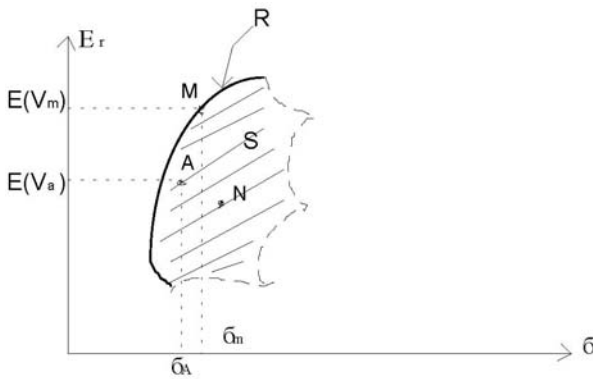


Рисунок 4 – Зона возможных существований портфелей

портфели, варьируя «вес» каждой ценной бумаги, то конкретному портфелю *A* будет соответствовать вполне определенное соотношение ожидаемой доходности $E(r_A)$ и риска (стандартное отклонение σ_A). Изображая эти соотношения на координатной плоскости с осями $E(r)$ и σ , получим точку *A* с координатами $[E(r_A); \sigma_A]$ на рис. 4 [8].

Заштрихованная площадь *S* представляет зону возможного существования портфелей, создаваемых из *n* выбранных ценных бумаг.

Для другого набора этих же ценных бумаг с определенным «весом» каждой бумаги получим другое соотношение ожидаемой доходности и риска (например, точка *N* на рис. 4). Можно показать, что из любого ограниченного набора ценных бумаг, выбранных инвестором, путем варьирования их «веса» можно получить бесконечное количество портфелей. Если для каждого из портфелей определить ожидаемую доходность и стандартное отклонение, отложить их на графике (рис. 4), то получим совокупность точек – зону, определяющую все возможные портфели для выбранного количества ценных бумаг.

Ключ к решению проблемы выбора оптимального портфеля лежит в теореме о существовании эффективного набора портфелей [9], так называемой границы эффективности. Суть теоремы сводится к выводу о том, что любой инвестор должен выбрать из всего бесконечного набора портфелей такой портфель, который обеспечивает максимальную ожидаемую доходность при каждом уровне риска и минимальный риск для каждой величины ожидаемой доходности.

Если инвестор выбрал *n* ценных бумаг со своими характеристиками $[E(r_i); \sigma_i; \sigma_{ij}; \rho_{ij}]$, где $i, j = 1, 2, \dots, n$, то найдется только одна комбинация ценных бумаг в портфеле, минимизирующая риск портфеля при каждом заданном значении ожидаемой доходности портфеля. Если обратиться к рис. 4, то вывод теоремы сводится к тому, что какую бы величину ожидаемой доходности не определил инвестор (например, $E(r_m)$ на

рис. 4), всегда путем перебора весов ценных бумаг портфеля можно найти такой портфель, при котором уровень риска достигает минимального значения (на рис. 4 – точка *M*) [10].

Набор портфелей, которые минимизируют уровень риска при каждой величине ожидаемой доходности, образует так называемую границу эффективности – на рис. 4 это линия *R*. Как видно из данного рисунка, при перемещении по границе вверх – вправо величины $E(r)$ и σ увеличиваются, а при движении вниз – влево – уменьшаются.

Эффективный портфель – это портфель, который обеспечивает минимальный риск при заданной величине $E(r)$ и максимальную отдачу при заданном уровне риска.

На риск портфеля основное влияние оказывает степень корреляции доходностей входящих в портфель ценных бумаг – чем ниже уровень корреляции, то есть чем ближе коэффициент корреляции приближается к (-1) , тем ниже риск портфеля. Путем диверсификации – изменения количества входящих в портфель ценных бумаг и их весов – инвестор способен снизить уровень риска портфеля, не изменяя при этом его ожидаемой доходности.

Если портфель состоит более чем из 2 ценных бумаг, то для любого заданного уровня доходности существует бесконечное число портфелей, т. е. можно сформулировать бесконечное количество портфелей, имеющих одну и ту же доходность.

Тогда задача инвестора сводится к следующему: из всего бесконечного набора портфелей с ожидаемой нормой отдачи $E(r_n)$ необходимо найти такой, который обеспечивал бы минимальный уровень риска. Иными словами, можно задачу инвестора свести к следующему [10]: необходимо найти минимальное значение дисперсий портфеля

$$\sigma_n^2 = \sum_{i=1}^n W_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n W_i W_j \rho_{i,j} \sigma_i \sigma_j \quad (1)$$

при заданных начальных условиях:

$$E(r \text{ портфеля}) = \sum_i^n W_i E(r_i), \quad (2)$$

$$\sum_{i=0}^n W_i = 1. \quad (3)$$

Для решения задачи нахождения оптимального портфеля, содержащего *n* ценных бумаг, необходимо первоначально вычислить:

- а) *n* значений ожидаемой доходности $E(r_i)$ (где $i = 1, 2, \dots, n$) каждой ценной бумаги в портфеле;
- б) *n* значений дисперсий σ_i^2 каждой ценной бумаги;

в) $n(n - 1)/2$ значений ковариации $\sigma_{i,j}$, где $i, j = 1, 2, \dots, n$.

Способы их вычисления приведены ранее. Если подставить значения $E(r_i)$, σ_i и $\sigma_{i,j}$ в уравнения (1)–(3), то выясняется, что в этих уравнениях неизвестными оказываются только величины W_i – «веса» каждой ценной бумаги в портфеле. Следовательно, задача формирования оптимального портфеля из n акций сводится к следующему: для выбранной величины доходности E^* инвестор должен найти такие значения W_i , при которых риск инвестиционного портфеля становится минимальным. Иначе говоря, для выбранного значения E^* инвестор должен определить, какие суммы инвестиционных затрат необходимо направить на приобретение той или иной ценной бумаги, чтобы риск инвестиционного портфеля оказался минимальным.

В теории Марковица инвесторы стремятся сформировать портфель ценных бумаг, чтобы максимизировать получаемую полезность. Каждый инвестор желает таким образом сформировать портфель, чтобы сочетание ожидаемой доходности $E(r)$ и уровня риска σ портфеля приносило бы ему максимальное удовлетворение потребностей и минимизировало риск при желаемой доходности. Разные инвесторы имеют различные друг от друга мнения об оптимальности сочетания $E(r)$ и σ , т. к. отношение одного инвестора к риску не похоже на желание рисковать другого инвестора. Поэтому, говоря об оптимальном портфеле, надо иметь в виду, что эта категория сугубо индивидуальна и оптимальные портфели разных инвесторов теоретически отличаются друг от друга. Тем не менее, каждый оптимальный портфель непременно является эффективным, то есть инвесторы выбирают удовлетворяющий их (оптимальный) портфель из эффективных портфелей.

4 ОПТИМИЗАЦИЯ ИНВЕСТИЦИОННОГО ПОРТФЕЛЯ ПО МОДЕЛИ ШАРПА

Выведенные Марковицем правила построения границы эффективных портфелей позволяют находить оптимальный (с точки зрения инвестора) портфель для любого количества ценных бумаг. Основной сложностью применения метода Марковица является большой объем вычислений, необходимый для определения весов W_i каждой ценной бумаги. Действительно, если портфель объединяет n ценных бумаг, то для построения границы эффективных портфелей необходимо предварительно вычислить n значений ожидаемых (средних арифметических) доходностей $E(r_i)$ для каждой ценной бумаги, n величин σ_i дисперсий всех норм отдачи, и $n(n - 1)/2$ выражений попарных ковариаций $\sigma_{i,j}$ ценных бумаг в портфеле.

В 1963 г. американский экономист У. Шарп (William Sharpe) предложил новый метод построения границы эффективных портфелей, позволяющий существенно

сократить объемы необходимых вычислений. В дальнейшем этот метод модифицировался и в настоящее время известен как одноиндексная модель Шарпа (Sharpe single-index model) [9].

В основе модели Шарпа лежит метод линейного регрессионного анализа, позволяющий связать две случайные переменные величины – независимую X и зависимую Y линейным выражением типа $Y = \alpha + \beta X$. Сам Шарп в качестве независимой переменной рассматривал доходность r_m . В качестве зависимой переменной берется доходность r_i i -й ценной бумаги. Обычно модель Шарпа называют рыночной моделью (Market Model), а доходность r_m – доходностью рыночного портфеля.

Пусть доходность r_m принимает случайные значения и в течение N шагов расчета наблюдаются величины $r_{m1}, r_{m2}, \dots, r_{mN}$. При этом доходность r_i i -й ценной бумаги имеет значения $r_{i1}, r_{i2}, \dots, r_{iN}$. В таком случае линейная регрессионная модель позволяет представить взаимосвязь между величинами r_m и r_i в любой наблюдаемый момент времени в виде:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t},$$

где $r_{i,t}$ – доходность i -й ценной бумаги в момент времени t ; α_i – параметр, постоянная составляющая линейной регрессии, показывающая, какая часть доходности i -й ценной бумаги не связана с изменениями доходности рынка ценных бумаг r_m ; β_i – параметр линейной регрессии, показывающий чувствительность доходности i -й ценной бумаги к изменениям рыночной доходности; $r_{m,t}$ – доходность рыночного портфеля в момент t ; $\varepsilon_{i,t}$ – случайная ошибка, свидетельствующая о том, что реальные, действующие значения $r_{i,t}$ и $r_{m,t}$ порою отклоняются от линейной зависимости.

Особое значение необходимо уделить параметру β_i , поскольку он определяет чувствительность доходности i -й ценной бумаги к изменениям рыночной доходности. Если $\beta_i > 1$, то доходность данной ценной бумаги более чувствительна, подвержена большим колебаниям, чем рыночная доходность r_m . Соответственно, при $\beta_j < 1$ ценная бумага имеет меньший размах отклонений доходности r_j от средней арифметической (ожидаемой) величины $E(r_j)$, чем рыночная доходность. Для большинства ценных бумаг $\beta > 0$, хотя могут встретиться ценные бумаги и с отрицательной величиной β .

Для нахождения параметров α_i и β_i по результатам наблюдений используется метод наименьших квадратов (МНК). Используя этот метод в качестве параметров α_i и β_i берутся такие значения, которые минимизируют сумму квадратов ошибок ε . Если провести необходимые вычисления, то окажется, что параметры α_i и β_i принимают следующие значения:

$$\alpha_i = E(r_i) - \beta_i YE(r_m),$$

$$\beta_i = \frac{\sigma_{i,m}}{\sigma_m^2} \text{ или } \beta_i = \frac{P_{i,m} \cdot \sigma_i}{\sigma_m}.$$

Параметры α_i и β_i регрессионной модели дают представление об общих тенденциях взаимосвязей между изменениями рыночного показателя r_m и нормой отдачи r_i . Однако величины α_i и β_i не позволяют давать однозначный ответ о степени подобной взаимосвязи. На точность регрессионной модели оказывают значительное влияние ошибки ε_i . Значит, точность регрессионной модели, степень взаимосвязи r_m и r_i , определяется разбросом случайных ошибок ε_i , который можно оценить с помощью дисперсии случайной ошибки. Кроме того, точность регрессии можно определить, оценивая, сколь точно регрессионная модель определяет дисперсию ценных бумаг, для которых составляется регрессионная модель: $\sigma_{\varepsilon_i}^2 \sigma_i^2$.

Дисперсию i -й ценной бумаги можно представить в виде двух слагаемых:

$$\sigma_i^2 = \beta_i^2 \cdot \sigma_m^2 + \sigma_{\varepsilon_i}^2, \tag{4}$$

$$\sum_{i=1}^n W_i \beta_i = W_{n+1}. \tag{5}$$

Ниже приведены основные этапы, которые необходимо выполнить для построения границы эффективных портфелей в модели Шарпа:

- 1) выбрать n ценных бумаг, из которых формирует-ся портфель, и определить исторический промежуток в N шагов расчета, за которым будут наблюдаться значения доходности $r_{i,t}$ каждой ценной бумаги;
- 2) по рыночному индексу вычислить рыночные до-ходности $r_{m,t}$ для того же промежутка времени;
- 3) определить величины β_i :

$$\beta_i = \frac{\sigma_{i,m}}{\sigma_m^2}; \tag{6}$$

- 4) найти параметр α_i :

$$\alpha_i = E(r_i) - \beta_i E(r_m);$$

- 5) вычислить дисперсии ошибок регрессионной мо-дели $\sigma_{\varepsilon_i}^2$;
- 6) подставить эти значения в уравнения (4)–(6).

После такой подстановки выяснится, что неизвест-ными величинами являются веса W_i ценных бумаг. Выбрав определенную величину ожидаемой доходнос-ти портфеля E^* , можно найти веса ценных бумаг в портфеле, построить границу эффективных портфе-лей и определить оптимальный портфель.

5 ПОРТФЕЛЬ ТОБИНА

Через несколько лет после исследования Марковица другой крупнейший американский экономист Д. Тобин

заметил, что если на рынке есть безрисковые бумаги, то решение задачи об оптимальном портфеле сильно упрощается.

Связь между ожидаемой доходностью $E(r)$ и рис-ком σ представляется в виде такой зависимости [9]:

$$E(r) = r_\delta + \frac{E(r_p) - r_\delta}{\sigma_p},$$

где r_p – безрисковая ставка доходности (эфффектив-ность безрисковых бумаг); $E(r_p)$ – ожидаемая ставка доходности рискованного актива; $\sigma_r \sqrt{D_r}$ – стандартное отклонение доходности рискованного актива; D_r – дис-персия (вариация) рисковой части портфеля, вариация портфеля равна $V_\Pi = (I - X_0)^2 D$, и риск портфеля равен $\sigma_\Pi = (1 - X_0) \sigma_r$.

Если X_0 – доля капитала, вложенного в безриско-вую часть портфеля, а $(1 - X_0)$ – безрисковая доля портфеля, то задача Марковица об оптимальном порт-феле в этом случае такова:

$$\sum_i \sum_j X_i X_j \rightarrow \min, \quad E_\Pi = x_0 r_0 + \sum_i x_i E_i, \\ x_0 + \sum_i x_i = 1.$$

Пусть Ω – матрица ковариаций рисковов видов цен-ных бумаг, $X = (x_i)$, $V = (v_i)$ – вектор-столбцы долей x капитала, вкладываемых в i -й вид рисковов ценных бумаг и ожидаемых эфффективностей этого вида, $i = 1, \dots, n$. Пусть также I – n -мерный вектор-столбец, компоненты которого равны 1. Тогда оптимальное зна-чение долей x_i есть

$$X^* = \frac{E(r_p) - r_\delta}{(V - r_\delta I) \Omega^{-1} \cdot (V - r_\delta I)} \Omega^{-1} (V - r_\delta I). \tag{7}$$

Здесь Ω^{-1} – матрица, обратная к Ω . В числителе дроби стоит число, в знаменателе, если выполнить все действия, тоже получится число, причем константа, определяемая рынком и не зависящая от инвестора, $\Omega - 1(V - r_\delta I)$ – вектор-столбец размерности n . Дан-ный вектор не зависит от эфффективности портфеля $E(r_p)$. Таким образом, вектор долей рисковов видов ценных бумаг, пропорциональный этому вектору, так-же не зависит от $E(r_p)$. Следовательно, структура рис-ковой части портфеля не зависит от $E(r_p)$. Однако сумма компонент вектора X^* зависит от $E(r_p)$, а имен-но, компоненты вектора X^* пропорционально увеличи-ваются с ростом $E(r_p)$, поэтому доля x_0 безрисковов вложений будет при этом сокращаться.

Выразим риск оптимального портфеля в зависимо-сти от его доходности. Для этого в формулу вариации портфеля $V_\Pi = X^T \Omega X$ подставим оптимальный вектор

X^* из формулы (7), обозначив знаменатель формулы (7) через d^2 . Получим:

$$V_{\Pi} = \frac{(E(r_p) - r_{\delta})^2}{d^4} [\Omega^{-1}(V - r_{\delta}I)]^T \Omega [\Omega^{-1}(V - r_{\delta}I)] = \frac{(E(r_p) - r_{\delta})^2}{d^2}.$$

Окончательно:

$$V_{\Pi} = \frac{(E(r_p) - r_{\delta})^2}{d^2} \text{ или } \sigma_{\Pi} = \frac{E(r_p) - r_{\delta}}{d}.$$

Можно также написать выражение эффективности оптимального портфеля от его риска:

$$E(r_p) - (r_{\delta}) = d\sigma_{\Pi} \text{ или } E(r_p) = r_{\delta} + d\sigma_{\Pi}.$$

Полученный оптимальный портфель называется портфелем Тобина минимального риска, т. е. портфель Тобина – это портфель Марковица при наличии на рынке безрисковых ценных бумаг.

6 МОДЕЛЬ БЛЭКА

Пусть инвестор желая увеличить свой инвестируемый капитал P^0 , находит дополнительную сумму P^g , тогда при покупке различных активов на суммы P_i^0, \dots, P_n^0 будем иметь [9] $P^0 + P^g = \sum_i P_i^0$ или после деления обеих частей этого равенства на $P_0, 1 + Y^g = \sum_i x_i$, где $Y^g = \frac{P^g}{P^0}$. Пусть $x_{n+1} = -Y^g$, тогда получаем $\sum_{i=1}^{n+1} x_i = 1$, но одна из долей средств, вкладываемых в актив i -го типа, а именно величина x_i уже отрицательная.

Понятно, что в более сложных ситуациях отрицательных компонент, отвечающих заемным средствам, может быть больше одной. Доходность портфеля в этом случае вычисляется в виде

$$E_{\Pi} = \frac{P^K - P^0 - P^g}{P^0 + P^g},$$

где P^K – стоимость актива в конце периода; P^0 – стоимость актива в начале периода; P^g – дополнительный (заемный) актив.

На большинстве фондовых бирж Запады действия, которые математически формализуются в виде $x_i < 0$,

допустимы и часто используются. Но ввиду их особой рискованности обычно есть дополнительные ограничения на такие действия, а по некоторым видам ценных бумаг и полный запрет. Портфели, удовлетворяющие условиям данного рынка, называются допустимыми. В модели Блэка допустимы любые портфели, т. е. они имеют единственное ограничение $\sum_i x_i = 1$.

Особенностью модели Блэка является то, что оказывается возможным реализовать любую, сколь угодно большую доходность (но за счет быстро растущего риска!).

Пусть есть два актива с ожидаемыми доходностями $e_1 = 1$ и $e_2 = -1$. Для портфеля $x_1 = 1 + v$, $x_2 = -v$ доходность

$$E_{\Pi} = 1 \cdot (1 + v) + (-1) \cdot (-v) = 1 + 2v \rightarrow \infty \text{ при } v \rightarrow \infty.$$

7 МОДЕЛЬ ТОБИНА – ШАРПА – ЛИНТНЕРА (ТШЛ)

В данной модели предполагается, что есть безрисковый актив, доходность которого не зависит от состояния рынка (обычно это – государственные ценные бумаги или вклады в большие банки). Если доходность безрискового актива (пусть он на рынке один, его номер – ноль) r_{δ} , то ожидаемая доходность $E(r_{\delta}) = r_{\delta}$, $\delta^2(r_{\delta}) = 0$ и $\text{cov}(r_{\delta}, r_i) = 0$ для всех $i \neq 0$, последнее означает, что в ковариационной матрице рынка есть нулевая строка и нулевой столбец. Все активы, кроме нулевого, – рискованные, то есть $\delta^2(r_i) > 0$ для $i = 1, \dots, n$.

Предположим, пусть существует портфель, который характеризуется вектором активов $\bar{x} = (x_0, x_1, \dots, x_n)$ и при $x_0 \neq 1$ его можно представить в виде линейной комбинации безрискового и рискованного портфелей:

$$x = x_0 e_0 + (1 - x_0) y_0,$$

где $e_0 = (1, 0, 0, \dots, 0)$ – это безрисковый портфель, совпадающий с безрисковым активом; $Y_0 \left(0, \frac{x_1}{1 - x_0}, \dots, \frac{x_n}{1 - x_0} \right)$ – чисто рискованная портфель.

Такое разложение играет важную роль при оценке фиксированных активов.

Рассмотрим рынок двух активов, описываемый вектором ожидаемых доходностей $E = (E_1, E_2)$ и матрицей ковариации

$$\Omega = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix},$$

где $\rho = \rho_{12}$ – коэффициент корреляции доходностей активов, σ_1, σ_2 – стандартные отклонения.

Для модели Блэка, когда допустимы любые значения x_1 и x_2 , лишь бы $x_1 + x_2 = 1$, имеем в двухмерном случае прямую на плоскости x_1, x_2 , которая составлена из множества допустимых пар. Удобно представить эту прямую в параметрическом виде: $x_1 = t, x_2 = 1 - t$, тогда каждый портфель описывается так: $x = (t, 1 - t)$, t – принимает любые вещественные значения (в том числе и отрицательные).

Доходность портфеля $E_{\Pi} = E_1 t + E_2 (1 - t) = E_2 + (E_1 - E_2)t$, риск портфеля

$$\sigma_{\Pi}^2 = \sum_{i=1}^2 \sum_{j=1}^2 x_i x_j \text{cov}(r_i, r_j) = (\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)t^2 + 2(\rho\sigma_1\sigma_2 - \sigma_2^2)t + \sigma_2^2.$$

Оценкой портфеля называют ряд чисел $(\sigma^2(x), E(x))$, которую можно изобразить точкой на плоскости $\sigma^2 E_0$.

Плоскость (σ^2, E) называют критериальной. Меняя портфель, то есть меняя вектор \bar{x} , получают различные оценки, а для них разные точки на критериальной плоскости. Множество всех оценок (то есть множество пар (σ^2, E) , а не множество портфелей) допустимых портфелей называют критериальным множеством. Если критериальное множество не сводится к одной точке, то возникает проблема выбора.

8 ПОСТРОЕНИЕ ГРАНИЦ ЭФФЕКТИВНОСТИ ПОРТФЕЛЯ С ПОМОЩЬЮ МОДЕЛИ МАРКОВИЦА

Модель, первоначально разработанная Марковицем и до сих пор широко применяемая, использует общий риск каждого отдельного актива. Следовательно, при построении портфелей и определении общего риска портфеля должны рассматриваться ковариации в каждой паре потенциальных для портфеля активов.

Когда доходы по рискованному активу являются случайными переменными, доходы по портфелю – это

взвешенная по стоимости средняя доходов по отдельным активам [10], т. е.

$$E(r_{\Pi}) = \sum_{i=1}^n V_i r_i.$$

Если коэффициент корреляции в парах активов меньше чем 1,0, то диверсификация может улучшить взаимосвязь между ожидаемым риском портфеля и ожидаемым доходом по портфелю. Это происходит потому, что, если переменная доходности является линейной функцией средней доходности, то фактор риска представляет собой квадратическую функцию дисперсии доходов по ценным бумагам. Степень улучшения портфеля зависит от весов, которые каждый из активов имеет в портфеле, и от корреляции этих активов.

Лучший способ продемонстрировать это – пример с двумя активами. Рассмотрим данные табл. 1 – различные среднеквадратические отклонения портфеля, составленного из двух рискованных активов, при допущениях, что корреляция $\rho_{1,2}$ равна 0,6 или 0,9 и что доли каждого актива в портфеле меняются на 10 %.

Рис. 5 – это диаграмма границ эффективности, относящихся к портфелям, построенным с учетом предположенных $\rho_{1,2} = 0,60$ и $\rho_{1,2} = 0,90$. Актив 1 имеет ожидаемый доход 8 % со среднеквадратическим отклонением 12 %, а актив 2 – ожидаемый доход 12 % со среднеквадратическим отклонением 16 %.

Среднеквадратическое отклонения портфеля вычисляется по формуле (8).

Среднеквадратическое отклонение портфеля из двух активов [11]:

$$\sigma_{\Pi} = \sqrt{V_1^2 \sigma_1^2 + V_2^2 \sigma_2^2 + 2V_1 V_2 (\rho_{12} \sigma_1 \sigma_2)}, \quad (8)$$

где σ_{Π} – среднеквадратическое отклонение портфеля; V_1 и V_2 – веса активов 1 и 2 в портфеле; σ_1^2 и σ_2^2 – дисперсии доходов по активам 1 и 2; ρ_{12} – корреляция

Таблица 1 – Значения весов и среднеквадратических отклонений портфеля

Вес V_2	Вес V_1	Доход $E(r_n)$	σ_{Π} при $\rho_{12} = 0,6$	σ_{Π} при $\rho_{12} = 0,9$
0	1,0	8,0	12	12
0,1	0,9	8,4	11,82	12,26
0,2	0,8	8,8	11,80	12,56
0,3	0,7	9,2	11,91	12,89
0,4	0,6	9,6	12,17	13,26
0,5	0,5	10,0	12,55	13,65
0,6	0,4	10,4	13,06	14,07
0,7	0,3	10,8	13,67	14,52
0,8	0,2	11,2	14,37	15,0
0,9	0,1	11,6	15,15	15,49
1,0	0	12,0	16,0	16,0

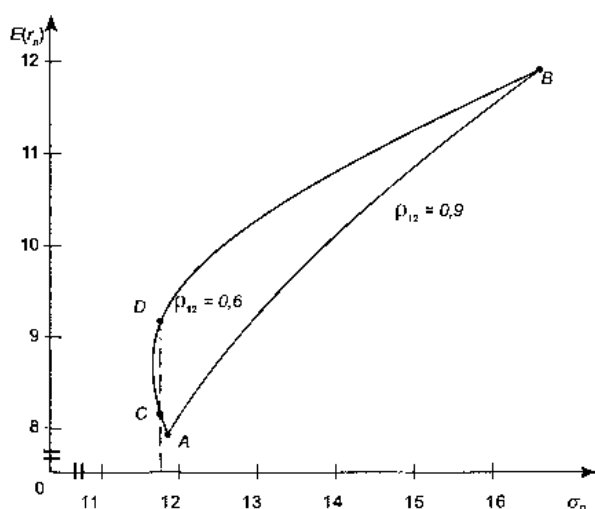


Рисунок 5 – Границы эффективности портфелей

доходов по активам 1 и 2; σ_1 и σ_2 – среднее квадратическое отклонение доходов по 1 и 2; $(\rho_{12}\sigma_1\sigma_2)$ – ковариации доходов по активам 1 и 2.

Для предположенной степени корреляции среднее квадратическое отклонение рассчитано для некоторых различных портфелей, которые могут быть построены из этих двух активов и нанесены на диаграмму (рис. 5).

Сначала рассмотрим данные в столбце табл. 1 для $\rho_{1,2} = 0,6$ и график на рис. 5 для $\rho_{1,2} = 0,6$, отражающие выгоды от диверсификации для случая, когда активы умеренно коррелированы. Данные и график, обозначенные $\rho_{1,2} = 0,9$, показывают, что диверсификация имеет благотворное влияние на соотношение риск – доход, даже когда активы высоко, но не полностью коррелированы. В обоих случаях граница эффективности вогнута. Чем больше степень вогнутости, тем больше выгоды от диверсификации. Однако не все точки на границе эффективны, а эффективна только верхняя часть каждой вогнутой границы (обозначенных AB на рис. 5).

Верхняя часть каждой из линий AB представляет границу эффективности возможных портфелей, так как на границе невозможно достичь большего дохода без несения большего риска. Выше линии находится область недостижимых комбинаций риска и дохода из-за ограниченности характеристик ценных бумаг 1 и 2. Ниже линии находятся худшие комбинации риска и дохода, которые могут быть улучшены просто перемещением в любую точку на линии AB . Это достигается продажей существующих активов и покупкой 1 или 2. Например, портфель C располагается на нижней части границы, помеченной $\rho_{1,2} = 0,6$. Инвестор может повысить свою полезность продажей этого портфеля и покупкой комбинации 1 и 2, представленной любой из точек на границе эффективности. Например, переме-

щаясь в точку D , инвестор несет тот же уровень риска, но получает более высокий доход, чем в C .

Нужно отметить, что не существует единственного наилучшего портфеля. Жирные линии указывают на многие «эффективные портфели». Граница эффективна, потому что невозможно повысить доход без увеличения риска или снизить риск без снижения дохода. Возможная комбинация риска и дохода будет зависеть от целевой функции (функция полезности для инвестора).

ВЫВОДЫ

Проведенный анализ структуры оптимального портфеля позволил выяснить, какие риски, связанные со стохастически меняющимися инвестиционными возможностями, следует хеджировать. Анализ показывает, что при конкретной динамике производных ценных бумаг предложенные модели позволяют получать различные виды портфелей инвестиций при различных функциях полезности и различных уровнях относительного неприятия риска инвестора. Не существует «единственно верной» стратегии выбора инвестиционного портфеля, которая одинаково подходила бы всем инвесторам без исключения. При принятии решений о составе портфеля инвестор достигнет более высокой ожидаемой (средней) доходности, только если согласится на более высокую степень риска.

Можно снизить степень риска инвестиций, не снижая ожидаемой доходности, за счет более полной диверсификации как в пределах одного класса активов, так и среди нескольких разных классов активов.

Способность за счет диверсификации снизить риск-ванность портфеля инвестора зависит от корреляции между активами, составляющими портфель. На практике подавляющее большинство активов имеет между собой положительную корреляцию, потому что на них влияют одни и те же экономические факторы. Следовательно, возможность снижения риска за счет диверсификации среди рискованных активов без снижения ожидаемого уровня доходности ограничена.

Чем дольше период, в течение которого инвестор собирается владеть акциями, тем меньше стандартное отклонение доходности акций, взятое в годовом исчислении, и тем меньше вероятность того, что ставка доходности акций окажется ниже соответствующей процентной ставки для безрисковых облигаций.

ПЕРЕЧЕНЬ ССЫЛОК

1. Башарин Г. В. Начала финансовой математики. – М.: Мир, 2003. – 189 с.
2. Крушвиц С. К. Финансирование и инвестиции. – СПб.: Питер, 2000. – 368 с.

3. Количественные методы финансового анализа / Под ред. Стивена Дж. Брауна и Марка П. Кричмена: Пер. с англ. – М.: ИНФРА-М, 1996. – 352 с.
4. Наталуха И. Г. Моделирование спекулятивного бума на финансовом рынке с учетом психологии инвесторов // Материалы VI Всеросс. симпозиума «Математическое моделирование и компьютерные технологии». – Кисловодск, 2004. – Т. 2. – С. 7–8.
5. Христановский В. В., Щербина В. П. Функция полезности: теория и анализ: Учебное пособие. – Х.: ИД «ИНЖЕК», 2006. – 120 с.
6. Черкасова В. А., Батенкова А. А. Влияние стратегических рисков на финансовые результаты компании // Корпоративные финансы. – 2007. – № 3. – С. 64–76.
7. Четыркин Е. М. Финансовая математика. – М.: Дело, 2002. – 540 с.
8. Шарп У., Александер Г., Бейли Д. Инвестиционный менеджмент. – М.: ИНФРА-М, 2003. – 257 с.
9. Шапки А. С. Экономические и финансовые риски. Оценка, управление, портфель инвестиций: Монография. – М.: Издательско-торговая корпорация «Дашков и К^о», 2003. – 544 с.
10. Эддоус М., Стэнфилд Р. Методы принятия решений. – М.: Мир, 2003. – 123 с.
11. Якушков В. К. Функция полезности для инвестирования. – М.: Дело, 2002. – 167 с.

Надійшла 25.04.2008

Розглянуто основні складові оптимального портфеля (очікуваний прибуток портфеля і стандартне відхилення як міра ризику), які дозволяють агенту фінансового ринку неперервно реструктурувати портфель (роблячи максимальною користь проміжного використання і (або) остаточного капіталу) у відповідності з інвестиційними можливостями, які змінюються стохастично. Проаналізовані методи оптимізації (метод Марковіца і Шарпа), а також розглянута роль функції корисності для формування інвестиційного портфеля.

The basic components of optimal portfolio were viewed: expected portfolio's yield and standard deviation as the risk's measure. These components allow the agent of the financial market to restructure portfolio uninterruptedly (maximizing profit of intermediate consumption and (or) final capital) according to stochastic varying investment opportunities. Methods of optimization (H. Markowitz's and W. Sharpe's method) were analysed, and also the role of utility function in forming investment portfolio was viewed.

УДК 004.658.3

А. Б. Кунгурцев, С. Л. Зиноватная

ИМИТАЦИОННАЯ МОДЕЛЬ ДЛЯ ИССЛЕДОВАНИЯ ЭФФЕКТИВНОСТИ ДЕНОРМАЛИЗАЦИИ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ В ИНФОРМАЦИОННОЙ СИСТЕМЕ

Рекомендуется выполнять денормализацию схемы отношений после тщательного тестирования производительности информационной системы. Предложена имитационная модель для тестирования поведения системы в условиях применения различных вариантов денормализации. На основе модели реализован программный инструмент, позволяющий исследовать базы данных, созданные с использованием различных систем управления базами данных.

ВВЕДЕНИЕ

Под денормализацией понимается процесс намеренного введения избыточности в нормализованных таблицах в целях увеличения производительности информационной системы (ИС). Производительность может быть повышена благодаря уменьшению времени выполнения запроса к БД, вызванному устранением операции соединения между отношениями либо сокращением размера отношений, участвующих в запросе. Необходимость такого изменения структуры базы данных (БД) становится очевидной лишь на этапе проектирования приложений [1]. Однако при определенных

обстоятельствах изменение структуры отношений не приводит к ожидаемому повышению производительности ИС. Например, в [2] описаны случаи, в которых устранение операции соединения в запросе, соответствующее методу нисходящей денормализации, не приводит к уменьшению времени выполнения запроса. Кроме того, сокращение времени выполнения одних запросов путем внедрения денормализации может быть нивелировано увеличением времени выполнения других запросов, использующих отношения, которые были реструктурированы, так что в целом производительность ИС может ухудшиться.

ПОСТАНОВКА ЗАДАЧИ

Существуют программные продукты, позволяющие анализировать эффективность запросов к БД. Например, Query Analyzer [3] позволяет просматривать план выполнения запроса, сгенерированный оптимизатором запросов MS SQL Server. Однако не существует универсальных инструментов, которые дают возмож-

© Кунгурцев А. Б., Зиноватная С. Л., 2008

ность проанализировать поведение системы после выполнения денормализации, в частности проверить влияние произведенных изменений в структуре отношений БД на время выполнения запросов различного типа.

Для решения данной задачи предложена имитационная модель для экспериментального исследования ИС в случае применения различных методов денормализации.

Применение модели производится после исследования ИС на предмет выявления возможных эффективных вариантов денормализации [4]. Однако реальная эффективность предложенных вариантов зависит от многих факторов, например, от используемой системы управления БД, от аппаратного обеспечения ИС, от загрузки системы, поэтому следует уточнить полученное значение прогнозируемой эффективности реструктуризации БД путем проведения эксперимента в условиях реальной ИС.

На основе модели для исследования эффективности денормализации БД реализован программный продукт, который позволяет решить следующие задачи:

1) сформировать условия для проведения эксперимента: создавать требуемую структуру отношений; задавать исходные данные для проведения эксперимента; заполнять отношения данными; изменять структуру таблиц в процессе проведения эксперимента;

2) выполнить набор запросов для определения изменений времени их выполнения для проведения сравнительного анализа поведения системы при различных исходных условиях: сканирования отношения до и после изменений в его структуре; устранения соединения отношений; соединения денормализованного отношения с другими отношениями;

3) просмотреть результатов работы программы: в табличном виде; в графическом виде.

Параметры модели, позволяющие изменять размер исследуемых отношений, включают следующее:

– исходное количество кортежей в исследуемых отношениях pC_0^R (pC_0^1 в родительском и pC_0^2 в подчиненном отношении соответственно);

– множества коэффициентов pKs (pKs^1 для изменения количества кортежей в родительском отношении, pKs^2 для изменения количества кортежей в подчиненном отношении и pKs^h для изменения размера кортежа реструктурируемого отношения соответственно).

Для проведения исследования эффективности денормализации необходимо указать значение входного параметра PCS , который задает текст строки соединения с исследуемой.

Указанная пользователем строка соединения используется для подключения к исследуемой БД, получения метаданных БД, формирования структуры тестовых отношений, изменения их состояния и выполнения запросов для проверки их поведения в различных условиях.

В исследуемой БД создаются отношения, предназначенные для проведения эксперимента. Структура отношений (количество, тип и размер атрибутов) зависит от рассматриваемых вариантов денормализации.

Созданные для проведения исследования отношения заполняются данными в количестве, указанном в исходных параметрах. Данные формируются в соответствии с типом заполняемого атрибута отношения. Тип атрибута определяется автоматически на основании метаданных. Содержимое экспериментальных отношений может быть сформировано с учетом характеристик данных в отношениях, эффективность денормализации которых исследуется.

К заполненным таким образом отношениям отправляется запрос на сканирование единственного отношения и/или соединение отношений. Запрос выполняется многократно, чтобы устранить возможность влияния случайных факторов, например, выполнения других задач в системе, на результат эксперимента. Полученные результаты усредняются.

Результаты эксперимента фиксируются в специально разработанной служебной БД. Это позволяет в дальнейшем пользователю просмотреть и проанализировать полученные данные.

В процессе работы программы пользователь может проводить любое количество экспериментов, а также работать с различными БД. Модель предоставляет возможность изменять параметры эксперимента, сохраняя при этом различные версии набора параметров.

СТРУКТУРА ЭКСПЕРИМЕНТАЛЬНЫХ ОТНОШЕНИЙ

Родительское отношение T1. Имеет начальный размер кортежа h_0^1 , совпадающий с размером кортежа исследуемого отношения.

Атрибутами отношения T1 являются:

– ключевой атрибут id числового типа;

– набор неключевых атрибутов $nonkey1_1, \dots, nonkey1_{NT1}$ строкового типа фиксированного размера, что позволяет обеспечить необходимый размер кортежа отношения.

Подчиненное отношение T2. Имеет начальный размер кортежа h_0^2 , совпадающий с размером кортежа исследуемого отношения.

Атрибутами отношения T2 являются:

– ключевой атрибут id числового типа;

– внешний ключевой атрибут $fkt1$ числового типа, служит для обеспечения связи с родительским отношением T1;

– набор неключевых атрибутов $nonkey1_1, \dots, nonkey1_{NT2}$ строкового типа фиксированного размера, что позволяет обеспечить необходимый размер кортежа отношения.

Родительское отношение T1D. В начале исследования имеет структуру, совпадающую со структурой отношения T1. В процессе проведения эксперимента размер кортежа этого отношения h_i^1 может изменяться в соответствии с условиями проведения эксперимента.

Атрибутами отношения T1D являются:

- ключевой атрибут id числового типа;
- набор неключевых атрибутов nonkeyt1_1, ..., nonkeyt1_NT1;

- добавляемый в отношение во время проведения эксперимента набор неключевых атрибутов AddP1, ..., AddPNT1D строкового типа фиксированного размера, что позволяет обеспечить необходимый размер кортежа отношения h_i^1 ; размер каждого такого атрибута и их общее количество NT1D изменяется в процессе проведения эксперимента в соответствии с заданными параметрами.

Подчиненное отношение T2D. В начале исследования имеет структуру, совпадающую со структурой отношения T2. В процессе проведения эксперимента размер кортежа этого отношения h_j^2 может изменяться в соответствии с условиями проведения эксперимента.

Атрибутами отношения T2D являются:

- ключевой атрибут id числового типа;
- внешний ключевой атрибут fkt1 числового типа, служит для обеспечения связи с родительским отношением T1;
- набор неключевых атрибутов nonkeyt2_1, ..., nonkeyt2_NT2 строкового типа фиксированного размера;

- добавляемый в отношение во время проведения эксперимента набор неключевых атрибутов AddP1, ..., AddPNT2D строкового типа фиксированного размера; размер каждого такого атрибута и их общее количество NT2D изменяется в процессе проведения эксперимента в соответствии с заданными параметрами.

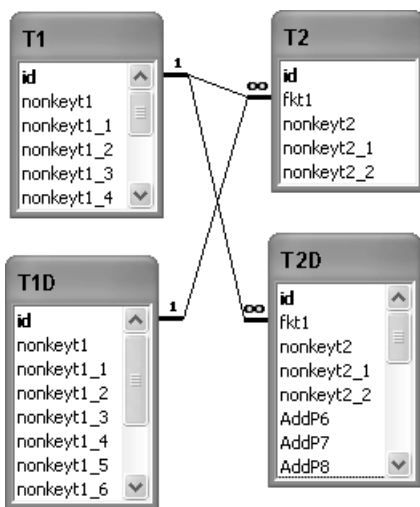


Рисунок 1 – Схема данных экспериментальных отношений

На рис. 1 приведена схема данных для экспериментального множества отношений.

СТРУКТУРА СЛУЖЕБНОЙ БД

В результате проведения эксперимента необходимо сохранять следующую информацию:

- идентификатор сеанса эксперимента (под сеансом понимается один цикл прохождения по всем коэффициентам изменения количества кортежей родительского отношения, подчиненного отношения и размера кортежа реструктурируемого отношения);
- строку соединения с исследуемой БД;
- дату и время проведения эксперимента;
- параметры проведения эксперимента (начальные значения для количества кортежей, коэффициенты их изменения, коэффициенты изменения размера кортежа исследуемого отношения, режим проведения эксперимента);
- время выполнения запросов к экспериментальным отношениям.

В процессе проведения эксперимента в течение одного сеанса происходит выполнение следующих вложенных циклов.

1) Цикл по количеству кортежей в родительском отношении:

$$i = C1_Start \text{ to } C1_End \text{ step } k1,$$

где $C1_Start$ – исходное количество кортежей в отношении T1, соответствует pC_0^1 ; $C1_End$ – конечное значение количества кортежей в отношении T1, определяется в зависимости от содержимого массива коэффициентов увеличения количества кортежей в родительском отношении; $k1$ – шаг изменения, предполагает увеличение текущего i -го значения количества кортежей в отношении в pKs_i^1 раз, где pKs_i^1 – i -е значение соответствующего множества коэффициентов pKs_i^1 ; определяется параметрами проведения эксперимента; может быть различным в разных итерациях цикла.

Количество кортежей в родительском отношении на i -й итерации цикла определяется следующим образом:

$$C_i^1 = pC_0^1 \cdot pKs_i^1.$$

2) Цикл по количеству кортежей в подчиненном отношении, выполняется на каждой итерации цикла для родительского отношения:

$$j = C2_Start \text{ to } C2_End \text{ step } k2,$$

где $C2_Start$ – исходное количество кортежей в отношении T2, соответствует pC_0^2 ; $C2_End$ – конечное значение количества кортежей в отношении T1, определя-

ется в зависимости от содержимого множества коэффициентов увеличения количества кортежей в подчиненном отношении; $k2$ – шаг увеличения, предполагает увеличение текущего j -го значения количества кортежей в отношении в pKs_i^2 раз, где pKs_i^2 – j -е значение соответствующего множества коэффициентов pKs^2 , которое определяется параметрами проведения эксперимента; может быть различным в разных итерациях цикла.

Количество кортежей в подчиненном отношении на j -й итерации цикла определяется следующим образом:

$$C_j^2 = pC_0^2 \cdot pKs_j^2.$$

3) Цикл, задающий увеличение размера кортежа исследуемого отношения, выполняется на каждой итерации цикла для исследуемого отношения:

$$k = k_Start \text{ to } k_End,$$

где k_Start – исходное значение коэффициента изменения размера кортежа pKs_0^h ; k_End – конечное значение коэффициента изменения размера кортежа.

Текущее значение pKs_k^h определяет, во сколько раз размер добавляемых атрибутов исследуемого отношения изменяется по сравнению с начальным значением (определенным при создании отношения).

Размер кортежа h_k^R реструктурируемого отношения R на k -й итерации цикла определяется по формуле:

$$h_k^R = h_0^R \cdot (1 + pKs_k^h),$$

где h_0^R – начальный размер кортежа.

На каждой итерации цикла измеряется время выполнения запросов к исследуемой БД.

Предложенная модель включает следующие варианты сравнения полученных в результате исследования показателей:

TTD – соотношение времени сканирования исходного отношения и этого же отношения с измененным размером кортежа;

T12T1D – соотношение времени сканирования соединения родительского и подчиненного отношений и родительского денормализованного отношения.

T12T2D – соотношение времени сканирования соединения родительского и подчиненного отношений и подчиненного денормализованного отношения.

Для хранения информации, использованной и полученной в результате исследования, необходимы следующие объекты:

O_CEAHC – для хранения информации о сеансе эксперимента;

O_C2 – для хранения данных о значениях количества кортежей в подчиненном отношении во время кон-

кретной итерации цикла по увеличению количества кортежей в родительском отношении;

O_C1 – для хранения данных о значениях количества кортежей в родительском отношении во время конкретного сеанса;

O_K – для хранения данных о значениях коэффициента изменения размера кортежа исследуемого отношения и времени выполнения соответствующих запросов на конкретной итерации цикла по увеличению количества кортежей в подчиненном отношении.

Между перечисленными объектами существуют следующие взаимосвязи:

– объект O_CEAHC находится во взаимосвязи «один ко многим» с объектом O_C2 , так как во время эксперимента может быть получено несколько значений O_C2 для одного сеанса;

– объект O_C2 находится во взаимосвязи «один ко многим» с объектом O_C1 , так как во время эксперимента может быть получено несколько значений O_C1 для одного O_C2 ;

– объект O_C1 находится во взаимосвязи «один ко многим» с объектом O_K , так как во время эксперимента может быть получено несколько значений O_K для одного O_C1 .

На основании приведенных отношений между объектами создана служебная БД dbExp.

Отношение Seans соответствует объекту O_CEAHC , имеет ключевой атрибут, являющийся уникальным идентификатором сеанса.

Отношение SeansC2 соответствует объекту O_C2 , имеет ключевой атрибут; внешний ключ для связи с отношением Seans; атрибут qc2 хранит текущее значение количества кортежей в отношении T2; атрибут tm2 хранит время сканирования отношения T2 при текущем количестве кортежей qc2.

Отношение SeansC1 соответствует объекту O_C1 , имеет ключевой атрибут; внешний ключ для связи с отношением SeansC2; атрибут qc1 хранит текущее значение количества кортежей в отношении T1; атрибут tm1 хранит время сканирования отношения T1 при текущем количестве кортежей qc1.

Отношение SeansK соответствует объекту O_K , имеет ключевой атрибут; внешний ключ для связи с отношением SeansC1; атрибут vk хранит текущее значение коэффициента изменения размера кортежа реструктурируемого отношения; атрибут tm2d хранит время сканирования отношения T2D при текущем количестве кортежей qc2 и текущем значении размера кортежа отношения T2D; атрибут tm12 хранит время соединения отношений T1 и T2 при текущем количестве кортежей qc2 и qc1; атрибут tm1d хранит время сканирования отношения T1 при текущем количестве кортежей qc1 и текущем значении размера кортежа отношения T1.

Схема данных dbExp приведена на рис. 2.

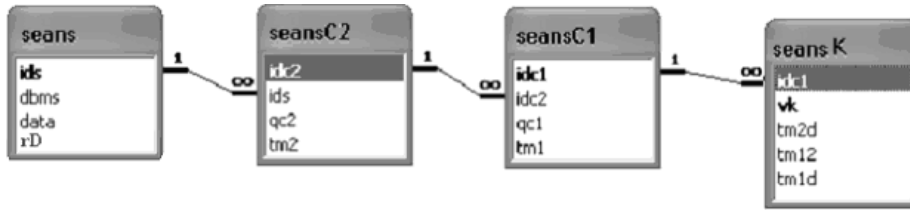


Рисунок 2 – Схема данных для служебной БД

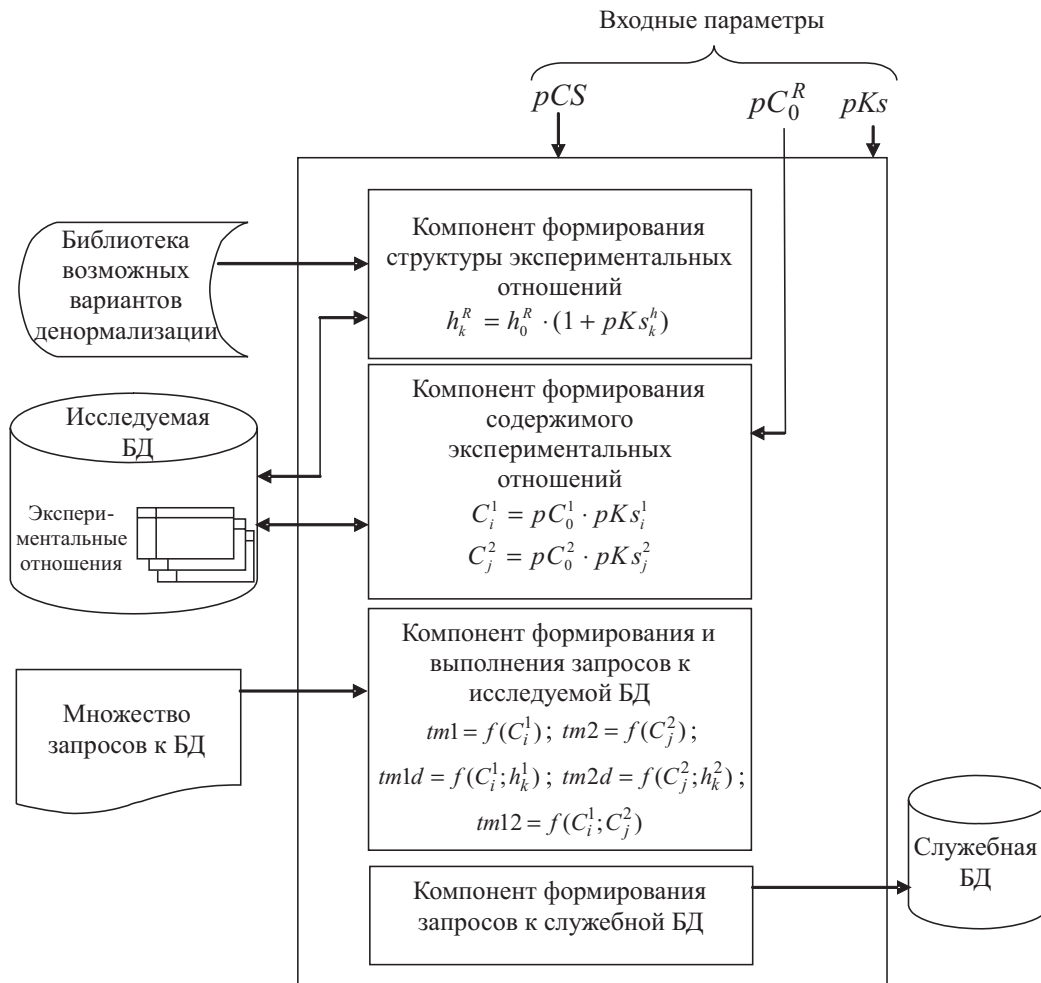


Рисунок 3 – Общая схема имитационной модели

Общая схема имитационной модели для исследования эффективности денормализации реляционной БД приведена на рис. 3.

СТРУКТУРА ФАЙЛА С ПАРАМЕТРАМИ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТА

Файл eff.ini (по умолчанию) является обычным текстовым файлом и может быть создан с использованием интерфейса программы или в текстовом редакторе.

Файл должен иметь следующую структуру. Для каждого входного параметра модели должны присутствовать пары строк. Первая строка в каждой паре содержит имя параметра в квадратных скобках. Определены следующие имена параметров:

- C1 – исходное количество кортежей в родительском отношении;
- C2 – исходное количество кортежей в подчиненном отношении;
- K1 – множество коэффициентов увеличения количества кортежей в родительском отношении;



Рисунок 4 – Заполненный файл параметров эксперимента

K2 – множество коэффициентов увеличения количества кортежей в отношении подчиненном;

K – множество коэффициентов изменения размера кортежа исследуемого отношения.

Вторая строка содержит одно или несколько числовых значений, соответствующих набору параметров. Коэффициенты должны располагаться в порядке возрастания значений. Количество коэффициентов в каждом множестве не ограничено.

Пары строк могут быть расположены в файле в произвольном порядке.

На рис. 4 приведен образец заполненного файла параметров проведения эксперимента. Из рисунка видно, что эксперимент проводится при двух значениях количества кортежей в отношении T1 и двух значениях количества кортежей в отношении T2, то есть рассматриваются четыре комбинации размера исследуемых отношений T1 и T2 соответственно: 10^5-10^5 , 10^5-10^6 , 10^6-10^5 , 10^6-10^6 . Для каждой комбинации выполняет-

ся изменение размера кортежа исследуемого отношения в 1,1 раза, в 2 раза, в 3 раза, в 6 раз.

АНАЛИЗ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА С ПРИМЕНЕНИЕМ ПРЕДЛОЖЕННОЙ ИМИТАЦИОННОЙ МОДЕЛИ

Ниже приведено описание эксперимента для различных методов денормализации.

Можно выделить следующие методы денормализации: нисходящая денормализация; восходящая денормализация; ввод повторяющихся групп; горизонтальная денормализация; вертикальная денормализация; ввод производных данных; ввод суррогатного ключа; устранение связи 1:1 [5].

Требования к определению входных параметров и режим работы программы для различных методов денормализации приведены в табл. 1.

В зависимости от выбранного режима работы программы происходит изменение размера кортежа исходного или подчиненного отношения для методов денормализации, предусматривающих устранение операции соединения между отношениями, или изменение размера кортежа единственного отношения.

1. Нисходящая денормализация. Уменьшение времени выполнения запросов может быть получено за счет устранения операции соединения между отношениями. При этом увеличивается размер кортежа подчиненного отношения, что приводит к увеличению общего объема памяти, занимаемого этим отношением. Следовательно, вероятно, увеличивается время сканирования этого отношения (зависит от размера добавляемого атрибута и количества кортежей в отношении).

Эксперимент позволяет измерить время сканирования исходного подчиненного отношения tm_2 и денор-

Таблица 1 – Правила определения входных параметров для различных методов денормализации

Метод денормализации	Режим	C1	C2	K1	K2	K
Нисходящая	T12T2D	#1	#2	#5	#5	#5
Восходящая	T12T1D	#1	#2	#5	#5	#5
Ввод повторяющихся групп	T12T1D	#1	#2	#5	#5	#5
Горизонтальная	TTD	#3	#4	#6	не задается	#5
Вертикальная	TTD	#3	#4	#6	не задается	#5
Ввод производных значений	TTD	#3	#4	#5	не задается	#5
Ввод суррогатного ключа	T12T2D	#1	#2	#5	#5	#5
Устранение связи «1:1»	T12T1D	#1	#2	#5	#5	#5
Ссылки: #1 выбирается, исходя из существующего или возможного значения количества кортежей в родительском отношении #2 выбирается, исходя из существующего или возможного значения количества кортежей в подчиненном отношении #3 выбирается, исходя из существующего или возможного значения количества кортежей в исследуемом отношении #4 выбирается произвольно (не влияет на результат эксперимента) #5 может содержать различные коэффициенты #6 может содержать различные коэффициенты, позволяющие проверить поведение исследуемого отношения при разном размере его фрагментов						

мализованного подчиненного отношения $tm2d$, а также время соединения отношений $tm12$.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения нисходящей денормализации от количества кортежей в родительском и подчиненном отношениях и от размера добавляемого в подчиненное отношение атрибута.

2. Восходящая денормализация. Уменьшение времени выполнения запросов может быть получено за счет устранения операции соединения между отношениями, подобно нисходящей денормализации. При этом увеличивается размер кортежа родительского отношения, что приводит к увеличению общего объема памяти, занимаемого этим отношением. Следовательно, вероятно, увеличивается время сканирования этого отношения (зависит от размера добавляемого атрибута и количества кортежей в отношении).

Эксперимент позволяет измерить время сканирования исходного родительского отношения $tm1$ и денормализованного подчиненного отношения $tm1d$, а также время соединения отношений $tm12$.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения восходящей денормализации от количества кортежей в родительском и подчиненном отношениях и от размера добавляемого в родительское отношение атрибута.

3. Ввод повторяющихся групп. Уменьшение времени выполнения запросов может быть получено за счет устранения операции соединения между отношениями, подобно нисходящей и восходящей денормализации. При этом увеличивается размер кортежа родительского отношения, что приводит к увеличению общего объема памяти, занимаемого этим отношением. Следовательно, вероятно, увеличивается время сканирования этого отношения (зависит от размера добавляемых повторяющихся групп и количества кортежей в отношении).

Эксперимент позволяет измерить время сканирования исходного родительского отношения $tm1$ и денормализованного родительского отношения $tm1d$, а также время соединения таблиц $tm12$.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения денормализации вводом повторяющихся групп от количества кортежей в родительском и подчиненном отношениях и от размера добавляемых в родительское отношение атрибутов.

4. Горизонтальная денормализация. Уменьшение времени выполнения запросов может быть получено за счет уменьшения объема памяти, занимаемого каждой частью денормализуемого отношения.

Эксперимент позволяет измерить время сканирования исходного отношения $tm1$ при различных значениях количества кортежей в отношении.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения горизонтальной денормализации от количества кортежей в отношении.

5. Вертикальная денормализация. Уменьшение времени выполнения запросов может быть получено за счет уменьшения объема памяти, занимаемого каждой частью денормализуемого отношения.

Эксперимент позволяет измерить время сканирования исходного отношения $tm1$ и денормализованного отношения $tm1d$ при различных значениях размера кортежа в отношении и различных значениях количества кортежей.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения вертикальной денормализации от размера кортежа в отношении и от общего размера отношения.

6. Ввод производных значений. Уменьшение времени выполнения запросов может быть получено за счет сокращения времени расчета производного значения в момент выполнения запроса, но может произойти увеличение времени выполнения запроса из-за увеличения объема памяти, занимаемого отношением после денормализации.

Эксперимент позволяет измерить время сканирования исходного отношения $tm1$ и денормализованного отношения $tm1d$ при различных значениях размера кортежа в отношении.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения производных значений от размера кортежа в отношении (зависит от размера добавляемого атрибута) и от общего размера отношения.

7. Ввод суррогатного ключа. Уменьшение времени выполнения запросов может быть получено за счет уменьшения объема памяти, занимаемого подчиненным отношением после денормализации.

Эксперимент позволяет измерить время сканирования исходного подчиненного отношения $tm1$ и денормализованного отношения $tm1d$, а также время соединения отношений $tm12$.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности внедрения ввода суррогатного ключа от размера кортежа в отношении и от общего размера отношений.

8. Устранение связи «один к одному». Уменьшение времени выполнения запросов может быть получено за счет устранения операции соединения между отношениями.

Эксперимент позволяет измерить время сканирования исходного родительского отношения $tm1$ и денормализованного отношения $tm1d$ и время соединения отношений $tm12$.

Используя полученные в ходе эксперимента показатели, можно получить функцию зависимости эффективности устранения связи «один к одному» от размера кортежа в отношениях и от общего размера отношений.

Предложенная имитационная модель исследования эффективности денормализации реляционной БД позволяет проверить поведение ИС при внедрении основных методов денормализации. Это обеспечивается за счет гибкости структуры входных параметров. Режимы работы реализованной на основе данной модели программы отличаются незначительно и обеспечиваются использованием сходных алгоритмов. В результате проведения эксперимента можно получить зависимости в табличном и графическом виде. Необходимые для получения зависимостей данные могут быть получены за счет формулировки различных запросов к служебной БД.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

В табл. 2 представлены частичные данные о поведении БД, созданной в системе управления базами данных MS Access, полученные в результате проведения серии экспериментов в режиме T12T2D.

Эксперимент проводился многократно при различных количествах кортежей в исходном и подчиненном отношении и при различном размере добавляемого

атрибута. Табл. 2 содержит сведения о количестве кортежей в исходном отношении qc1 и в подчиненном отношении qc2 соответственно, о коэффициенте vk, показывающем, во сколько раз размер добавленного атрибута больше размера кортежа исходного отношения, а также усредненные значения для времени сканирования денормализованного подчиненного отношения tm2d, времени сканирования исходного подчиненного отношения tm2 и времени выполнения операции соединения между исходными отношениями и tm12. Данные получены с помощью запроса к БД dbExp.

```
SELECT qc2, qc1, vk, Avg (tm2), Avg (tm2d), Avg (tm12)
FROM (seans INNER JOIN (seansC2 INNER JOIN
seansC1 ON seansC2.idc2 = seansC1.idc2)
ON seans.ids = seansC2.ids) INNER JOIN seansK ON
seansC1.idc1 = seansK.idc1
GROUP BY qc2, qc1, vk
```

В качестве примера результата анализа поведения ИС после реструктуризации отношений БД представлены графики (см. рис. 5), построенные на основе данных из табл. 2 для исследования эффективности нисходящей денормализации.

Введем понятие эффективности устранения операции соединения в единственном запросе E_{join} как отношение времени выполнения запроса с операцией соединения ко времени выполнения запроса со сканированием

Таблица 2 – Результаты эксперимента

.qc1	qc2	vk	Ср. знач. tm2d	Ср. знач. tm2	Ср. знач. tm12	qc1	qc2	vk	Ср. знач. tm2d	Ср. знач. tm2	Ср. знач. tm12
1000	1000	0.25	0.0132	0.0155	0.0141	10000	50000	1.25	1.5852	0.7110	8.8610
1000	1000	0.5	0.0141	0.0155	0.0133	10000	50000	1.5	1.5852	0.7110	8.8610
1000	1000	0.75	0.0156	0.0155	0.0141	10000	50000	1.75	1.5844	0.7110	8.1882
1000	1000	1	0.0157	0.0155	0.0164	10000	50000	2	1.5890	0.7110	9.2336
1000	1000	1.25	0.0140	0.0155	0.0179	10000	100000	0.25	2.5016	1.5160	18.4320
1000	1000	1.5	0.0140	0.0155	0.0204	10000	100000	0.5	4.3391	1.4380	19.7679
1000	1000	1.75	0.0133	0.0155	0.0172	10000	100000	0.75	5.5125	1.5160	16.6859
1000	1000	2	0.0172	0.0155	0.0195	10000	100000	1	8.2703	1.4327	22.7860
1000	10000	0.25	0.1547	0.1250	0.8590	10000	100000	1.25	8.4898	1.5160	22.1125
1000	10000	0.5	0.1984	0.1250	1.0310	10000	100000	1.5	8.7898	1.4380	20.0258
1000	10000	0.75	0.1953	0.1250	0.7970	10000	100000	1.75	8.7703	1.5160	18.4008
1000	10000	1	0.3109	0.1250	0.9530	10000	100000	2	8.2797	1.4380	19.6336
1000	10000	1.25	0.3509	0.1250	0.7190	50000	100000	0.25	2.9640	1.7627	105.7517
1000	10000	1.5	0.3725	0.1250	0.6410	50000	100000	0.5	4.3750	1.7369	118.5603
1000	10000	1.75	0.3809	0.1250	0.7660	50000	100000	0.75	4.4805	1.7762	99.6186
1000	10000	2	0.3410	0.1250	0.9680	50000	100000	1	6.3235	1.7521	114.1672
10000	50000	0.25	0.8586	0.7110	10.3102	50000	100000	1.25	7.5875	1.7875	98.0344
10000	50000	0.5	0.9436	0.7110	10.3102	50000	100000	1.5	7.9870	1.7003	115.3261
10000	50000	0.75	1.0500	0.7110	8.3633	50000	100000	1.75	8.0757	1.6719	104.0023
10000	50000	1	1.2500	0.7110	8.3633	50000	100000	2	8.1937	1.6973	113.3766

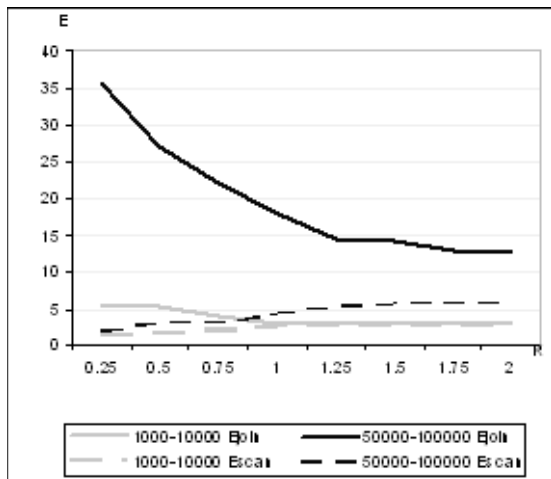


Рисунок 5 – Соотношение времени выполнения запросов

реструктурированного подчиненного отношения

$$E_{\text{join}} = \frac{tm12}{tm2d}$$

Введем также понятие эффективности изменения структуры отношения для единственного запроса E_{scan} как отношение времени выполнения запроса на сканирование исходного отношения ко времени выполнения запроса со сканированием реструктурированного отношения $E_{\text{scan}} = \frac{tm2d}{tm2}$.

Графики на рис. 5 представляют зависимость E_{join} (сплошная линия) и E_{scan} (пунктирная линия) от значения коэффициента K , определяющего во сколько раз размер (в байтах) добавленных атрибутов больше размера исходного кортежа, при различных количествах кортежей в исследуемых отношениях (10^3 в родительском отношении и 10^4 в подчиненном отношении – светлая линия, $5 \cdot 10^4$ в родительском отношении и 10^5 в подчиненном отношении – темная линия).

Результаты проведения эксперимента с использованием предложенной имитационной модели показывают, что эффект от применения нисходящей денормализации в значительной степени зависит от размера отношения определяется количеством записей и размером кортежа. Если размер отношений сопоставим с размером оперативной памяти компьютера, на котором выполняются запросы с использованием операции соединения, можно получить значительный эффект повышения производительности ИС. Однако эффективность снижается при увеличении размера добавляемых в подчиненное отношение атрибутов.

Из графика видно, что при малых значениях количества кортежей в отношениях и при размере добавленных атрибутов, близком к исходному размеру кор-

тежа, величина сокращения времени выполнения запросов путем устранения операции соединения, становится сопоставимой с величиной увеличения времени выполнения запросов на сканирование реструктурированного отношения.

Реальный эффект от проведения денормализации с использованием исследуемого варианта определяется при учете количества запросов с использованием соединения отношений и запросов со сканированием денормализованного отношения за период исследования ИС. Так, например, при малом количестве кортежей в отношениях, если размер добавленного атрибута составляет половину размера подчиненного отношения, то уже при двукратном превышении количества запросов, требующих сканирования подчиненного отношения, по сравнению с количеством запросов, использующих операцию соединения с данным отношением, в системе наблюдается снижение производительности.

Имея данные о частоте использования разных типов запросов, можно провести дальнейшее экспериментальное исследование для получения конкретного результата, соответствующего текущему состоянию БД.

ВЫВОДЫ

Предложенная имитационная модель для исследования эффективности денормализации БД позволяет разработчику получить информацию о целесообразности проведения реструктуризации отношений БД. Универсальность разработанной на основе этой модели программной системы состоит в том, что она позволяет проводить эксперименты с различными системами управления базами данных, с использованием различного аппаратного обеспечения, а также для различных методов денормализации. Накапливаемая в процессе проведения эксперимента информация дает возможность путем формирования соответствующих запросов к служебной БД сопоставить измеренные величины для получения оценок поведения системы в случае применения различных методов денормализации.

ПЕРЕЧЕНЬ ССЫЛОК

1. Принятие решения о денормализации. – Режим доступа: <http://lavsoft.webhost.ru/oracle/book/dezdb/04/04.html>. – Название с экрана.
2. Кунгурцев А. Б. Анализ целесообразности реструктуризации базы данных методом введения нисходящей денормализации / Кунгурцев А. Б., Зиноватная С. Л. – Тр. Одесск. политехн. ун-та. Одесса, 2006. – 1(25). – С. 104–108.
3. М. Ф. Гарсия. Справочник администратор Microsoft SQL Server 2000. 2-е изд. / Пер с англ. / М. Ф. Гарсия, Дж. Реддинг, Э. Уолен, С. А. ДеЛюк. – М.: Издательство «СП ЭКОМ», 2004. – 976 с.
4. Зиноватная С. Л. Анализ информационной системы для принятия решения о выполнении восходящей денормализации // Холодильная техника и технологии. Сборник научных трудов IV семинара «Информационные системы и технологии». – Одесса, 2006. – С. 14–18.

5. Кунгурцев А. Б. Модель реструктуризації реляційної бази даних путем денормалізації схеми / Кунгурцев А. Б., Зиноватная С. А. – Тр. Одесск. политехн. ун-та. – Одесса, 2006. – 2(26). – С. 105–111.

Надійшла 27.11.2007
Після доробки 19.03.2008

Рекомендується виконувати денормалізацію схеми відносин після ретельного тестування продуктивності інформаційної системи. Запропоновано імітаційну модель для тестування поведінки системи в умовах застосування різних варіантів денормалізації. На основі моделі реалізований програмний інструмент, що дозволяє дослід-

жувати бази даних, створені з використанням різних систем керування базами даних.

It is recommended to execute denormalization of relations scheme after the careful testing of the information system productivity. A simulation model for testing of the system conduct in the conditions of application of different denormalization variants is offered. On the basis of this model a program instrument is realized to allow to probe databases, created with the use of the different database management systems.

681.51:622.788

В. Й. Лобов, С. А. Рубан

РОЗРОБКА ПРИНЦИПІВ КЕРУВАННЯ ТЕМПЕРАТУРНИМ РЕЖИМОМ ПРОЦЕСУ ВИПАЛЮВАННЯ КОТУНІВ З ВИКОРИСТАННЯМ ПРОГНОЗУЮЧИХ ANFIS-МОДЕЛЕЙ

Запропоновано принцип керування температурним режимом процесу випалювання котунів на конвеєрній машині з використанням прогнозуючої ANFIS-моделі.

ВСТУП

Відомим є факт, що збільшення продуктивності, підвищення якості готової продукції та зменшення питомої витрати енергоносіїв можна досягти за рахунок удосконалення автоматизованих систем управління технологічними процесами (АСУ ТП). Одним з найперспективніших напрямків розвитку АСУ ТП гірничо-збагачувальних комбінатів є використання теорії керування на основі прогнозуючих моделей (Model Predictive Control, MPC), що забезпечують високі адаптивні властивості та можливість квазіоптимального керування нелінійними об'єктами в режимі реально часу з урахуванням обмежень на керуючі і вихідні змінні [1, 2]. В базовому варіанті MPC в якості моделі об'єкту керування використовується лінійна дискретна модель у просторі станів [2]. Однак відомо, що для більшості технологічних процесів гірничо-збагачувальних комбінатів, зокрема і для процесу температурної обробки котунів на конвеєрній випалювальній машині, достатньо складно отримати адекватну математичну модель. Це пов'язано з багатомірністю та стохастичними властивостями об'єктів керування, нестационарністю, наявністю великих часових запізнь, неможливістю безперервного контролю окремих технологічних показників [3, 4]. Тому для вирішення цієї проблеми можуть бути використані нелінійні багатомірні математичні моделі, зокрема нейромережеві та нечіткі.

Так, в роботі [1] в схемі MPC використовується лінійна дискретна модель у просторі станів, яка знаходиться з наслідків нечітких правил типу Такаґі – Сугено прямим способом. В роботі [5] запропоновано для прогнозування показників процесу збагачення використовувати нейромережеві моделі NNARX (нейромережева авторегресійна модель, екзогенний тип сигналів) та NNOE (нейромережева модель помилки виходу).

Для отримання високих показників якості котунів при забезпеченні ресурсозберігаючої технології необхідно забезпечити бажаний розподіл температур уздовж випалювальної машини шляхом стабілізації температури в кожній газоповітряній камері. Регулювання температури котунів в різних зонах випалювальної машини здійснюється шляхом змінення кількості теплоносія, що фільтрується через шар. Наприклад, розглянемо зону сушки 2. В цю зону газ-теплоносії з температурою 900–1000 °С надходить прямим перетоком із зони охолодження 1. Задане значення температури на вході в шар (350–450 °С) підтримується за рахунок розбавлення теплоносія холодним повітрям із зони охолодження 2, що подається через бічні форкамери. Змінення витрати холодного повітря дозволяє також компенсувати збурення, пов'язані зі змінами висоти шару котунів, швидкості візків випалювальної машини, температури та тиску теплоносія, що надходить в зону сушки 2 із зони охолодження прямим перетоком [6].

В існуючих системах автоматизації стабілізація температури в газоповітряних камерах зони сушки 2 здійснюється шляхом змінення витрати холодного по-

вітря $Q_{\text{хп}}$ за допомогою дроселів на основі результатів вимірювання значення температури над шаром котунів у відповідній камері T_{c2} . Як вказано в роботі [6], передаточна функція об'єкту по каналу «температура в горні зони сушки 2 – витрата холодного повітря» має вигляд

$$W(p) = \frac{T_{c2}(p)}{Q_{\text{хп}}(p)} = \frac{K \cdot e^{-\tau \cdot p}}{T \cdot p + 1}, \quad (1)$$

де коефіцієнт підсилення K , постійна часу T та величина запізнення τ змінюються в часі під впливом діючих на об'єкт збурень. Об'єкти регулювання температури в інших технологічних зонах випалювальної машини можуть бути описані аналогічними передаточними функціями. В якості регуляторів використовуються типові ПІД-регулятори з постійними коефіцієнтами. Але з урахуванням нестационарності об'єкту подібні системи в цілому не задовольняють вимогам до точності та швидкодії керування температурним режимом.

ПОСТАНОВКА ЗАВДАННЯ

З урахуванням вказаних особливостей технологічного процесу температурної обробки котунів на випалювальної машині автори вважають перспективним використання математичної моделі, яка реалізується на базі адаптивної нейро-нечіткої системи (ANFIS). ANFIS-системи поєднують переваги нейронних мереж та нечітких моделей – вони мають високу апроксимуючу здатність, можуть навчатися в процесі роботи, дозволяють формалізувати та включити в модель нечітку інформацію, отриману від експертів і виражену в лінгвістичній формі, дозволяють пояснити отримані результати.

Метою даної статті є розробка принципів керування температурним режимом процесу випалювання котунів на конвеєрній машині з використанням прогнозуючої ANFIS-моделі.

РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

Підвищення якості керування температурним режимом обробки котунів можна досягти за рахунок використання системи керування з прогнозуючою моделлю, реалізованою на базі адаптивної нейро-нечіткої системи. Структурна схема відповідної системи автоматичного керування температурним режимом в зоні сушки 2 наведена на рис. 1. ANFIS-модель апроксимує нелінійну залежність керованої величини (температури T_{c2} в горні зони сушки 2) від керуючої дії (витрати холодного повітря через бічні форкамери $Q_{\text{хп}}$) та діючих на об'єкт координатних збурень – висоти шару котунів $H_{\text{ш}}$, швидкості візків випалювальної машини $V_{\text{ом}}$, температури T_{c1} в останній газоповітряній камері попередньої технологічної зони (зони сушки 1), температури

$T_{\text{гп}}$ та тиску $P_{\text{гп}}$ теплоносія, що надходить в зону сушки 2 через колектор прямої течії. Використання моделі дозволяє спрогнозувати динаміку змінення температури в горні зони на певному інтервалі часу при заданих початкових умовах та певних значеннях керуючої дії. При повторенні процедури навчання ANFIS-моделі з періодом, меншим за інтервал квазістационарності об'єкту керування, забезпечується також адаптація моделі до параметричних збурень (погіршення газощільності газопроводів і колекторів тощо). Для вимірювання поточного значення температури в горні зони сушки 2 використовується вимірювач температури BT . Адаптивний регулятор здійснює змінення керуючої дії з метою підтримання екстремального значення обраного критерію керування в залежності від змін зовнішніх впливів та властивостей об'єкту. Внутрішній контур (контур витрати лівої сторони) призначений для підтримання заданої витрати повітря за результатами вимірювань її фактичного значення. Контур складається з клапана холодного повітря K , виконуючого пристрою $ВП$, витратоміра $ВВ$ та регулятора витрати $PВ$. Контур витрати правої сторони є веденим і автоматично слідує за змінням витрати через пальник ведучої (лівої) сторони. Вигляд та параметри передаточних функцій клапана, виконуючого пристрою та витратоміру залежать від вибору технічних засобів автоматизації для реалізації системи. В якості регулятора витрати може бути використаний типовий цифровий регулятор (ПІ або ПІД), що забезпечує налаштування контуру на технічний оптимум.

Аналогічний підхід може бути використаний для реалізації контурів керування температурним режимом в газоповітряних камерах зон нагріву, випалювання та охолодження, контуру керування тепловим навантаженням у випалюванні. Для різних зон будуть відрізнятися керуючі дії та вхідні параметри прогнозуючих моделей.

Входами прогнозуючої моделі на кроці $k + 1$ є послідовність поточного та m попередніх значень керуючої дії

$$Q_{\text{хп}}(k+1)_0^m = [Q_{\text{хп}}(k+1), Q_{\text{хп}}(k), \dots, Q_{\text{хп}}(k-m+1)],$$

послідовності n попередніх значень вимірюваних збурень

$$\begin{aligned} \{H_{\text{ш}}(k)\}_1^n &= [H_{\text{ш}}(k), H_{\text{ш}}(k-1), \dots, H_{\text{ш}}(k-n+1)], \\ \{V_{\text{ом}}(k)\}_1^n &= [V_{\text{ом}}(k), V_{\text{ом}}(k-1), \dots, V_{\text{ом}}(k-n+1)], \\ \{T_{c1}(k)\}_1^n &= [T_{c1}(k), T_{c1}(k-1), \dots, T_{c1}(k-n+1)], \\ \{T_{\text{гп}}(k)\}_1^n &= [T_{\text{гп}}(k), T_{\text{гп}}(k-1), \dots, T_{\text{гп}}(k-n+1)], \\ \{P_{\text{гп}}(k)\}_1^n &= [P_{\text{гп}}(k), P_{\text{гп}}(k-1), \dots, P_{\text{гп}}(k-n+1)] \end{aligned}$$

та послідовність n попередніх виміряних значень керованої величини об'єкту

$$\{T_{c2}(k)\}_1^n = [T_{c2}(k), T_{c2}(k-1), \dots, T_{c2}(k-n+1)].$$

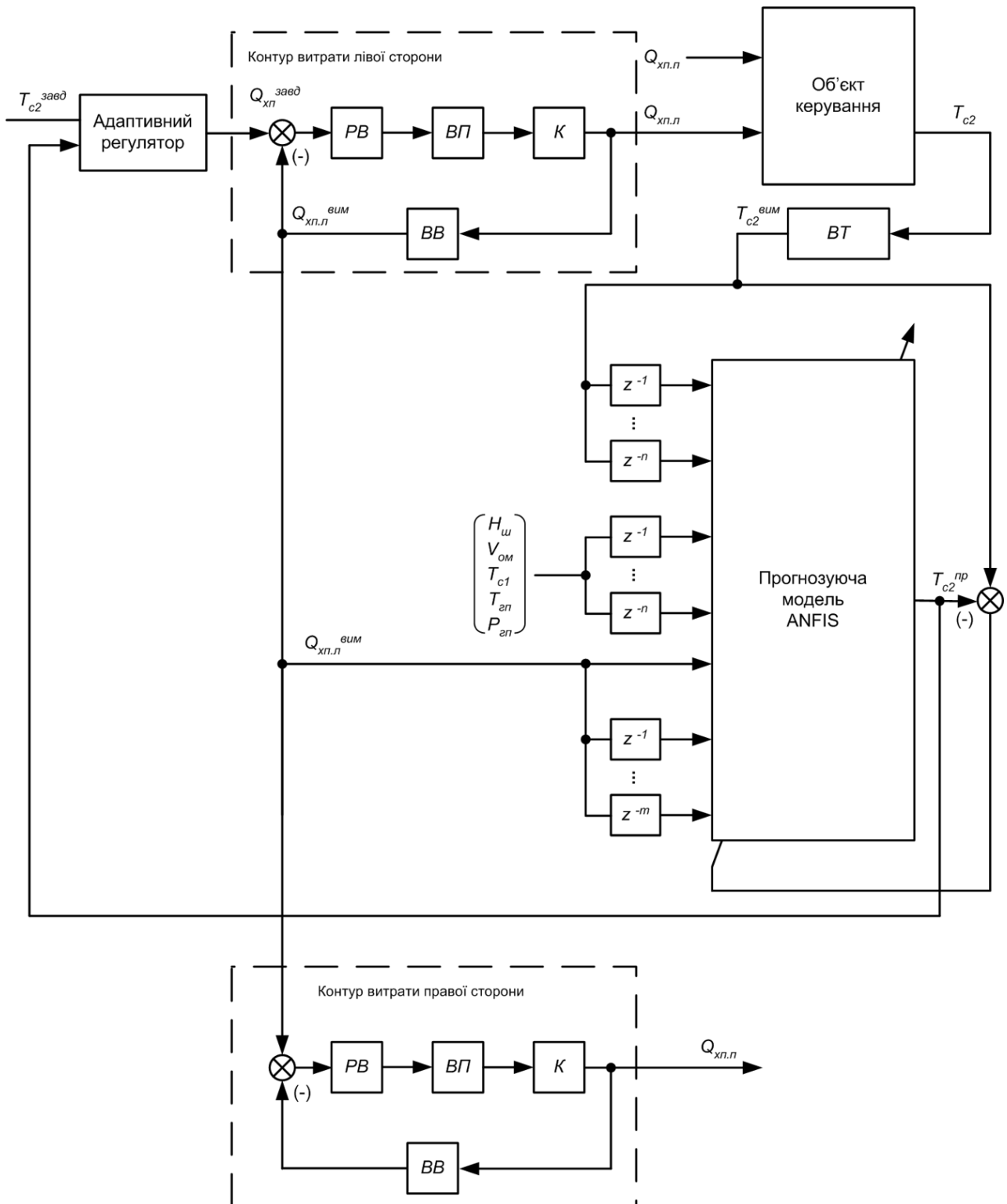


Рисунок 1 – Структурна схема системи керування температурним режимом в горні зони сушки 2 послідовності п попередніх значень вимірюваних збурень

В загальному випадку глибина затримки по кожній вхідній та вихідній змінній може бути різною.

Результат прогнозування значення температури в горні зони сушки 2 на один крок вперед визначається за формулою

$$\begin{aligned} \hat{T}_{c2}(k+1) = & \\ = NPM(\{T_{c2}(k)\}_1^n, \{H_{ш}(k)\}_1^n, \{V_{ом}(k)\}_1^n, \{T_{c1}(k)\}_1^n, & \\ \{T_{гн}(k)\}_1^n, \{P_{гн}(k)\}_1^n, \{Q_{хп}(k+1)\}_0^m), & \end{aligned} \quad (2)$$

де $NPM(\bullet)$ – оператор нелінійного перетворення типу «вхід – вихід», яке виконується ANFIS-моделлю.

За аналогією з роботами [1, 2], при керуванні температурним режимом випалювання котунів метою керування є виконання рівностей

$$\begin{aligned} \lim_{t \rightarrow \infty} \|T(t) - T^{завд}(t)\| = 0, \\ \lim_{t \rightarrow \infty} \|Q(t) - Q^{завд}(t)\| = 0, \end{aligned} \quad (3)$$

де $T(t)$ – температура в певній газоповітряній камері, $Q(t)$ – керуюча дія, а функції $T^{завд}(t)$ і $Q^{завд}$ визначають бажаний рух об'єкту.

Задача оптимального керування температурним режимом процесу випалювання може бути сформульована як пошук керуючої дії, що забезпечує досягнення заданої мети керування (3) та мінімізує обраний функціонал якості з урахуванням обмежень на вхідні та вихідні змінні.

При використанні MPC-методу модель на основі поточного стану при заданій керуючій дії дозволяє отримати прогноз траєкторії руху вихідних змінних об'єкту на інтервалі прогнозування P . На основі отриманого прогнозу адаптивний регулятор вирішує задачу про вибір оптимального керування на інтервалі P , та подає його на об'єкт на інтервалі керування C , $C \leq P$. Далі процес повторюється.

Так, для оцінки якості керування температурним режимом обробки котунів в зоні сушки 2 на інтервалі прогнозування P доцільно використовувати квадратичний функціонал вигляду

$$\begin{aligned} J_k = J_k(\bar{T}_{c2}(k+1), \bar{Q}_{хп}(k+1), P) = & \\ = \sum_{j=1}^P [(\hat{T}_{c2}(k+j) - r_x(k+j)) \cdot R_{k+j} \times & \\ \times (\hat{T}_{c2}(k+j) - r_x(k+j)) + & \\ + Q_{хп}(k+j) \cdot S_{k+j} \cdot Q_{хп}(k+j)], & \end{aligned} \quad (4)$$

де R_{k+j} – вагові коефіцієнти, які оцінюють ступінь небажаності відхилення температури на кроці $k+j$ від

заданого значення, S_{k+j} – коефіцієнти, що враховують вартість енергії керуючої дії, а

$$\begin{aligned} \bar{T}_{c2}(k+1) = [\hat{T}_{c2}(k+1), \hat{T}_{c2}(k+2), \dots, \hat{T}_{c2}(k+P)]^T, \\ \bar{Q}_{хп}(k+1) = [Q_{хп}(k+1), Q_{хп}(k+2), \dots, Q_{хп}(k+P)]^T \end{aligned}$$

– допоміжні вектори.

Авторами виконано комп'ютерне моделювання запропонованої системи керування температурним режимом обробки котунів в середовищі MATLAB 7.0. Для апроксимації залежності температури в зоні від витрати холодного повітря, яка наближено може бути описана передаточною функцією (1) зі змінними параметрами, використовувалася ANFIS-модель, входами якої було обрано поточне значення витрати $Q_{хп}$ та послідовність з трьох попередніх вимірних значень температури над шаром котунів. Для кожної вхідної змінної були визначені по дві функції приналежності узагальненого дзвоноподібного типу. База правил ANFIS-моделі містила 16 правил типу Такагі – Сугено, які описували всі можливі комбінації функцій приналежності вхідних величин. Наступне значення температури T_{c2} визначалося за формулою

$$\begin{aligned} \hat{T}_{c2}(k+1) = & \\ = NPM(T_{c2}(k), T_{c2}(k-1), T_{c2}(k-2), Q_{хп}(k+1)). & \end{aligned} \quad (5)$$

Результати тестування ANFIS-моделі при прогнозуванні значення температури в горні зони сушки 2 на один крок вперед наведені на рис. 2. Так, на рис. 2, а наведено

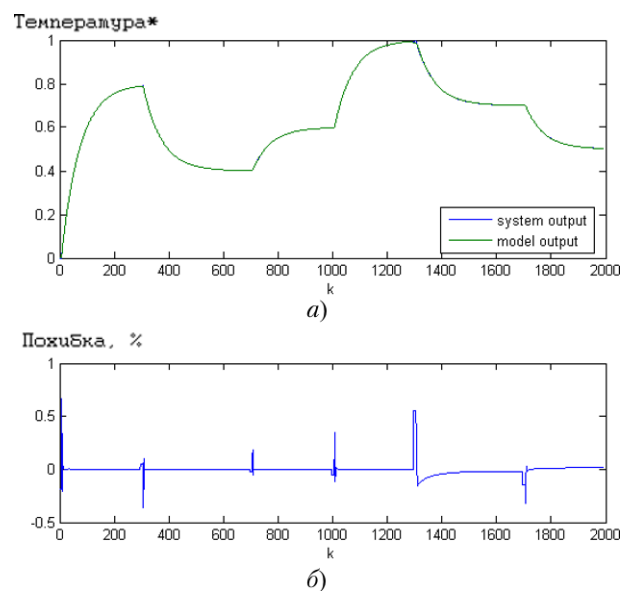


Рисунок 2 – Результати тестування ANFIS-моделі при прогнозуванні значення температури в горні зони сушки 2

графік реакції об'єкту (system output) на тестовий сигнал, а також графік змінення вихідного значення ANFIS-моделі (model output). На рис. 2, б наведено графік змінення похибки прогнозування.

Аналіз наведених графіків свідчить, що при використанні ANFIS-моделі похибка прогнозування значення температури на один крок вперед не перевищує 0,5 % абс.

Для визначення допустимого значення інтервалу прогнозування авторами було виконано дослідження можливості прогнозування температури в горні зони сушки на декілька кроків вперед. Аналіз отриманих результатів дозволяє стверджувати, що при прогнозуванні на 5–6 кроків вперед похибка не перевищує 1 % абс. Отже, при використанні ANFIS-моделі у складі системи керування температурним режимом випалювання котунів можливий інтервал прогнозування P дорівнює 5 крокам.

Задача пошуку оптимального на інтервалі P керування для прогнозуючої моделі (5) полягала у знаходженні векторної функції

$$\bar{Q}_{\text{хп}}(k+1) = [Q_{\text{хп}}(k+1), Q_{\text{хп}}(k+2), \dots, Q_{\text{хп}}(k+P)]^T,$$

що мінімізує функціонал (4)

$$J_k(\bar{T}_{c2}(k+1), \bar{Q}_{\text{хп}}(\bullet), P) \rightarrow \min_{\bar{Q} \in M}, \quad (6)$$

при обмеженнях на кожному кроці

$$Q_{\text{хп}}^{\min}(k+1) \leq Q_{\text{хп}}(k+1) \leq Q_{\text{хп}}^{\max}(k+1),$$

де $Q_{\text{хп}}^{\min}(k+1)$, $Q_{\text{хп}}^{\max}(k+1)$ – відповідно мінімальне та максимальне значення керуючої дії; M – допустима множина керувань.

Оскільки залежність (5) температури T_{c2} в горні зони сушки 2 від керуючої дії $Q_{\text{хп}}$ є нелінійною, для знаходження розв'язку задачі оптимізації (6) використовувався метод прямого пошуку.

На рис. 3 наведені результати комп'ютерного моделювання системи керування температурним режимом випалювання котунів з використанням прогнозуючої ANFIS-моделі. Так, на рис. 3, а наведено графік перехідного процесу у вихідній системі при поданні на вхід ступінчастої дії (T2), а також графік перехідного процесу при поданні на вхід об'єкту керуючої дії, що знаходилася як розв'язок задачі (6) при горизонті прогнозування $P = 5$ (T1). При моделюванні величина інтервалу керування дорівнювала величині інтервалу прогнозування ($C = P$). Відповідний графік змінення керуючої дії (витрати повітря) наведено на рис. 3, б.

Аналіз отриманих результатів дозволяє стверджувати, що при використанні MPC-методу з прогнозуючою

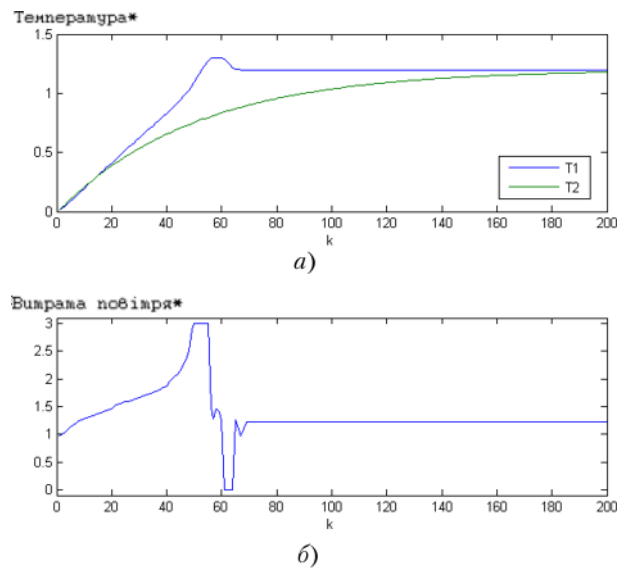


Рисунок 3 – Результати комп'ютерного моделювання системи керування температурним режимом випалювання котунів в зоні сушки 2 з використанням прогнозуючої ANFIS-моделі

ANFIS-моделлю швидкодія системи керування температурою в горні зони сушки 2 збільшується на 5–10 % при наявності перерегулювання в межах 5 %, статична похибка відсутня.

ВИСНОВОК

Для керування температурним режимом випалювання котунів на конвеєрній машині доцільно застосовувати MPC-метод з прогнозуючою ANFIS-моделлю, що дозволяє виконувати ідентифікацію об'єкту та здійснювати в режимі реального часу оптимальне керування продуктивністю та якістю продукції при раціональному використанні енергоносіїв.

ПЕРЕЛІК ПОСИЛАНЬ

1. Поркуян О. В. Принципы прогнозирующего управления технологическим процессом обогащения железных руд на основе нечетких моделей // Вісник Криворізького технічного університету. – Вип. 18. – Кривий Ріг: КТУ. – 2007. – С. 178–184.
2. Веремей Е. И., Еремеев В. В., Сотникова М. В. Пособие "Model Predictive Control Toolbox" [Электронный ресурс]. – Режим доступа: <http://matlab.exponenta.ru/modelpredict/Book1/index.php>. – Загл. с экрана.
3. Купин А. И., Рубан С. А. Исследование инверсных моделей нейроконтроллера для систем интеллектуального управления ТП горнорудных предприятий // Вісник Криворізького технічного університету. – Вип. 18. – Кривий Ріг: КТУ, 2007. – С. 157–161.
4. Рубан С. А., Лобов В. Й. Использование нечетких регуляторов для автоматизации технологического процесса обжига окатышей на конвейерных обжиговых машинах // Разраб. рудн. месторожд. – Вип. 91. – Кривий Ріг: КТУ, 2007. – С. 188–193.

5. Купин А. И. Нейросетевое прогнозирование показателей обогащения магнетитовых кварцитов // Вісник НТУ «ХПІ». – № 26. – Харків: НТУ «ХПІ». – 2006. – С. 23–31.
6. Ксендзовский В. Р. Автоматизация процессов производства окатышей. – М.: Металлургия, 1971. – 216 с.

Надійшла 9.01.2008

Предложен принцип управления температурным режимом процесса обжига окатышей на конвейерной машине с использованием прогнозирующей ANFIS-модели.

The control principle of temperature conditions of rolled briquettes sintering process on conveyor machine based on the predictive ANFIS model is suggested.

УДК 681.5.015.73:621.34

В. І. Мороз

АНАЛІЗ РАЦІОНАЛЬНОГО ПОРЯДКУ АПРОКСИМАЦІЙНОГО ПОЛІНОМА ДЛЯ ВІДНОВЛЕННЯ ІНФОРМАЦІЇ ЗА ЇЇ ДИСКРЕТНИМИ ВІДЛІКАМИ

З використанням частотних характеристик виконано дослідження раціонального порядку апроксимаційного полінома для відтворення сигналу за його дискретними відліками.

ПОСТАНОВКА ПРОБЛЕМИ

Однією з основних задач у цифрових системах керування або обробки інформації є відтворення неперервного сигналу за його дискретними відліками. Потрібно зауважити, що дана проблема може бути вирішена лише алгоритмічним шляхом у процесі формування сигналів керування, у той же час фізична реалізація можлива тільки для фіксатора нульового порядку, найпростішим варіантом якого в цифрових системах є звичайний багаторозрядний регістр. Відомо, що фіксатор нульового порядку вносить як амплітудну, так і, найголовніше, фазну похибку, що пов'язані з його півперіодним запізненням [1, 2]. Одним зі способів корекції такого запізнення в цифрових системах є використання апроксимацій сигналу вищих порядків для оцінки внесеної процесом дискретизації похибки та її компенсації.

Зрозуміло, що будь-яке відтворення сигналу за його дискретними відліками є наближенням і відображає первинний сигнал з певною похибкою. Таким чином, постає задача аналізу раціонального порядку апроксимаційного полінома, який би забезпечував необхідну точність відновлення неперервної інформації за умови найпростішого апроксимуючого виразу, тобто, за мінімальних програмно-апаратних витрат.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Огляд існуючих літературних джерел [3–6] показує, що в галузі розробки математичних основ циф-

рових систем домінує класичний підхід, закладений ще його основоположниками [1, 2]. Нічого не змінилося і в розробці практичних цифрових систем, коли навіть дуже складні сучасні алгоритми керування базуються на числових методах і способах дискретизації неперервних систем, які мають щонайменше піввікову історію [3, 4], а основна увага надається системам керування верхнього рівня ієрархії, паралельним і розподіленим системам [6]. Паралельно продовжують розвиватися числово-аналітичні підходи до синтезу цифрових систем [7].

Метою досліджень є визначення раціонального порядку апроксимаційного полінома для відтворення неперервного сигналу за його дискретними відліками та оцінка похибки, яка внесена процесом дискретизації.

Як відомо, у дискретних системах сигнали подаються своїми відліками, що найчастіше рівномірно розподілені в часі. Їхня наявність створює одну з основних проблем, які потребують свого вирішення в процесі реалізації цифрових систем – відновлення значення сигналу в проміжках між відліками. Як правило, це здійснюється шляхом апроксимації за кількома послідовними відліками за допомогою полінома відповідного порядку з наступною, за необхідністю, інтерполяцією. При цьому мета такої операції залишається незмінною – знаходження діючого значення неперервного сигналу, який відновлено за його дискретними значеннями.

Зрозуміло, що збільшення кількості послідовних відліків для побудови апроксимаційного полінома підвищує точність процедури відновлення, проте існує розумна межа, вище якої подальше збільшення відліків не дає відчутного підвищення точності. Раціональний порядок апроксимаційного поліному для сиг-

© Мороз В. І., 2008

налу $x(t)$ обмежується п'ятим порядком з кількох міркувань [8]:

1) поганою обумовленістю апроксимацій високого порядку [8] – невеликі зміни в коефіцієнтах полінома (наприклад, через обмежену розрядність їхнього подання в цифровій системі) можуть призводити до значних змін у поведінці апроксимуючої функції;

2) появою осциляцій (коливань) апроксимуючого полінома високого порядку (вище п'ятого) на ділянці апроксимації та поза нею [8, 9];

3) поведінка реальної системи високого порядку найчастіше незначно відрізняється від поведінки системи нижчого порядку [8, 10]; в роботі [10] на прикладі показано, що лінійна модель 5-го порядку забезпечує практично таку ж точність відтворення динамічних процесів, як і нелінійна модель 11-го порядку.

Похибка від процесу апроксимації сигналу $x(t)$ визначається різницею між двома його середніми значеннями на проміжку між часовими відліками (рис. 1):

– апроксимованим (наближеним) значенням, яке знаходиться інтегруванням апроксимуючої функції $x^*(t)$ в межах $t_i \leq t < t_{i+1}$;

$$X_i^* = \frac{1}{h} \int_{t_i}^{t_{i+1}} x^*(t) dt;$$

– істинним (точним) значенням, яке знаходиться інтегруванням аналітичної функції $x(t)$ в межах $t_i \leq t < t_{i+1}$:

$$X_i = \frac{1}{h} \int_{t_i}^{t_{i+1}} x(t) dt.$$

Таким чином, похибка апроксимації сигналу на i -му проміжку $t_i \leq t < t_{i+1}$ між двома його відліками визначатиметься

$$\begin{aligned} \varepsilon_i &= X_i - X_i^* = \frac{1}{h} \int_{t_i}^{t_{i+1}} x(t) dt - \frac{1}{h} \int_{t_i}^{t_{i+1}} x^*(t) dt = \\ &= \frac{1}{h} \int_{t_i}^{t_{i+1}} (x(t) - x^*(t)) dt. \end{aligned}$$

Недоліком такого способу є неможливість правильної оцінки похибки для довільного сигналу $x(t)$, оскільки в такому випадку для нього не існує аналітичного виразу, що унеможлиблює безпосереднє обчислення похибки. Вирішити дану задачу дає змогу широко використовуваний в теорії автоматичного керування аналіз частотних характеристик (амплітудних – АЧХ і фазних – ФЧХ) використовуваних апроксимацій як дискретних фільтрів [1, 2]. При цьому потрібно врахувати, що виразу $X_i = \frac{1}{h} \int_{t_i}^{t_{i+1}} x(t) dt$ відповідає операція

аналітичного інтегрування, а виразу $X_i^* = \frac{1}{h} \int_{t_i}^{t_{i+1}} x^*(t) dt$ –

інтеграл від апроксимуючого поліному. Таким чином, частотний аналіз похибок від застосування поліноміальних апроксимацій зводиться до аналізу графіків частотних похибок між операціями ідеального інтегрування та інтегрування апроксимуючого полінома $x_{(n)}^*(t) = a_n t^n + \dots + a_2 t^2 + a_1 t + a_0$.

Знаходження інтегралу від апроксимуючого полінома можливе для двох випадків побудови обчислювальної схеми: явної та неявної. Пошук невідомих коефіцієнтів для апроксимуючих формул відбувається за відомою методикою – за n рівновіддаленими на крок h точками будують апроксимаційний поліном, коефіцієнти якого знаходять для двох згаданих випадків зі системи лінійних рівнянь, яка складена для моменту часу t_i :

1) для неявної схеми побудови апроксимаційного полінома, що допускається у випадку комп'ютерного

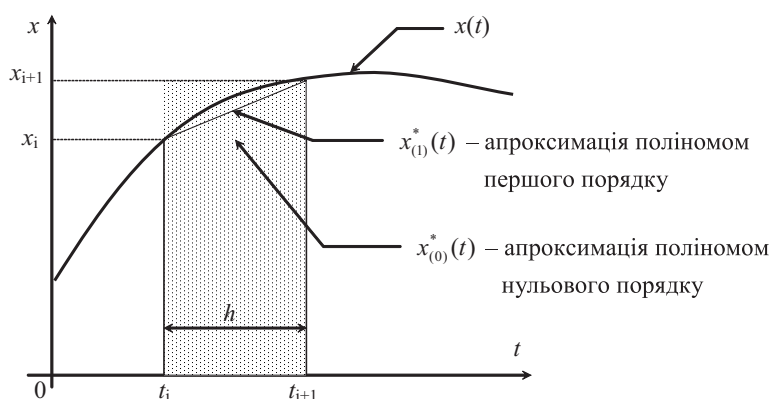


Рисунок 1 – Ілюстрація процесу визначення похибки апроксимації

моделювання, бо може бути вирішена за допомогою процедури розв'язування відповідної системи алгебричних рівнянь:

$$\left\{ \begin{array}{l} a_n \cdot h^n + a_{n-1} \cdot h^{n-1} + \dots + a_2 \cdot h^2 + a_1 \cdot h + a_0 = x_{i+1}; \\ a_0 = x_i; \\ a_n \cdot (-h)^n + a_{n-1} \cdot (-h)^{n-1} + \dots + a_2 \cdot (-h)^2 - a_1 \cdot h + a_0 = x_{i-1}; \\ \vdots \\ a_n \cdot ((n-1)h)^n + a_{n-1} \cdot ((n-1)h)^{n-1} + \dots + a_2 \cdot ((n-1)h)^2 - a_1 \cdot ((n-1)h) + a_0 = x_{i-(n-1)}. \end{array} \right.$$

2) для явної схеми побудови апроксимаційного полінома – лише така розрахункова схема допускає фізичну реалізацію, бо не передбачає використання у момент часу t_i ще невідомого $i+1$ -го значення координати; застосування даного способу апроксимації придатне для цифрових систем керування:

$$\left\{ \begin{array}{l} a_0 = x_i; \\ a_n \cdot (-h)^n + a_{n-1} \cdot (-h)^{n-1} + \dots + a_2 \cdot (-h)^2 - a_1 \cdot h + a_0 = x_{i-1}; \\ \vdots \\ a_n \cdot ((n-1)h)^n + a_{n-1} \cdot ((n-1)h)^{n-1} + \dots + a_2 \cdot ((n-1)h)^2 - a_1 \cdot ((n-1)h) + a_0 = x_{i-(n-1)}; \\ a_n \cdot (n \cdot h)^n + a_{n-1} \cdot (n \cdot h)^{n-1} + \dots + a_2 \cdot (n \cdot h)^2 - a_1 \cdot (n \cdot h) + a_0 = x_{i-n}. \end{array} \right.$$

Для ілюстрації нижче на прикладі показано процес знаходження коефіцієнтів апроксимаційних поліномів першого порядку (за двома точками) і другого порядку (використовуються три точки).

ПОБУДОВА АПРОКСИМАЦІЙНОГО ПОЛІНОМА ПЕРШОГО ПОРЯДКУ ЗА НЕЯВНОЮ СХЕМОЮ

Коефіцієнти апроксимації для заданого апроксимаційного полінома першого порядку $x^*(t) = a_1 t + a_0$ за двома відліками сигналу x_i і x_{i+1} в моменти часу t_i і t_{i+1} з інтервалом дискретності h знаходяться зі системи лінійних алгебричних рівнянь другого порядку:

$$\left\{ \begin{array}{l} a_1 h + a_0 = x_{i+1}; \\ a_0 = x_i; \end{array} \right. \Rightarrow \left\{ \begin{array}{l} a_0 = x_i; \\ a_1 = \frac{x_{i+1} - x_i}{h}. \end{array} \right.$$

ПОБУДОВА АПРОКСИМАЦІЙНОГО ПОЛІНОМА ДРУГОГО ПОРЯДКУ ЗА НЕЯВНОЮ СХЕМОЮ

Для побудови полінома $x^*(t) = a_2 t^2 + a_1 t + a_0$ використовуються три відліки сигналу x_{i-1} , x_i і x_{i+1} в моменти часу t_{i-1} , t_i і t_{i+1} з інтервалом дискретності h . Коефіцієнти апроксимації знаходяться з відповідної системи лінійних алгебричних рівнянь третього порядку:

$$\left\{ \begin{array}{l} a_2 h^2 + a_1 h + a_0 = x_{i+1}; \\ a_0 = x_i; \\ a_2 h^2 - a_1 h + a_0 = x_{i-1}; \end{array} \right. \Rightarrow \left\{ \begin{array}{l} a_1 = \frac{x_{i+1} - x_{i-1}}{2h}; \\ a_2 = \frac{x_{i+1} - 2x_i + x_{i-1}}{2h^2}. \end{array} \right.$$

У випадку потреби застосування явної апроксимаційної схеми зі системи виключається рівняння для точки t_{i+1} , що призводить до відповідних змін у коефіцієнтах:

$$\left\{ \begin{array}{l} a_0 = x_i; \\ a_2 h^2 - a_1 h + a_0 = x_{i-1}; \\ a_2 (2h)^2 - a_1 (2h) + a_0 = x_{i-2}; \end{array} \right. \Rightarrow \left\{ \begin{array}{l} a_1 = \frac{3x_i - 4x_{i-1} + x_{i-2}}{2h}; \\ a_2 = \frac{x_i - 2x_{i-1} + x_{i-2}}{2h^2}. \end{array} \right.$$

Аналогічна процедура застосовується для знаходження коефіцієнтів апроксимації поліномом довільного порядку. Знайдені коефіцієнти апроксимуючого полінома підставляються у відповідний вираз $x^*(t)$ для виконання процедури аналітичного інтегрування. За отриманими виразами будуються дискретні передатні функції для інтегралів апроксимуючих поліномів [1, 2].

Побудова графіків частотних похибок для знайдених дискретних передатних функцій відбувається в діапазоні до $1/10$ частоти квантування ω_0 згідно теорему відліків Шеннона-Котельникова – такий діапазон є робочим практично для всіх цифрових систем

[3–5]. Дана процедура є відомою [1, 2] і просто реалізується в середовищі будь-якого математичного пакету (наприклад, MathCAD чи MATLAB). Отримані графіки частотних похибок поліноміальних апроксимацій до п'ятого порядку включно показано на рис. 2 (для явних схем) і рис. 3 (для неявних схем). Для наочності графіки побудовані у тривимірному масштабі, що дає змогу одночасно оцінити амплітудну і фазну частотні похибки.

Аналіз частотних похибок поліноміальних апроксимацій показує безсумнівні переваги неявної обчислювальної схеми з точки зору амплітудних і, найголовніше, фазних частотних похибок, які є визначальними для замкнених систем автоматичного регулювання. Ще одним цікавим результатом є те, що відчутне зменшення частотних похибок у нижній частині робочого діапазону (нижче $1/20$ частоти квантування) практично припиняється після другого порядку апроксимаційного полінома.

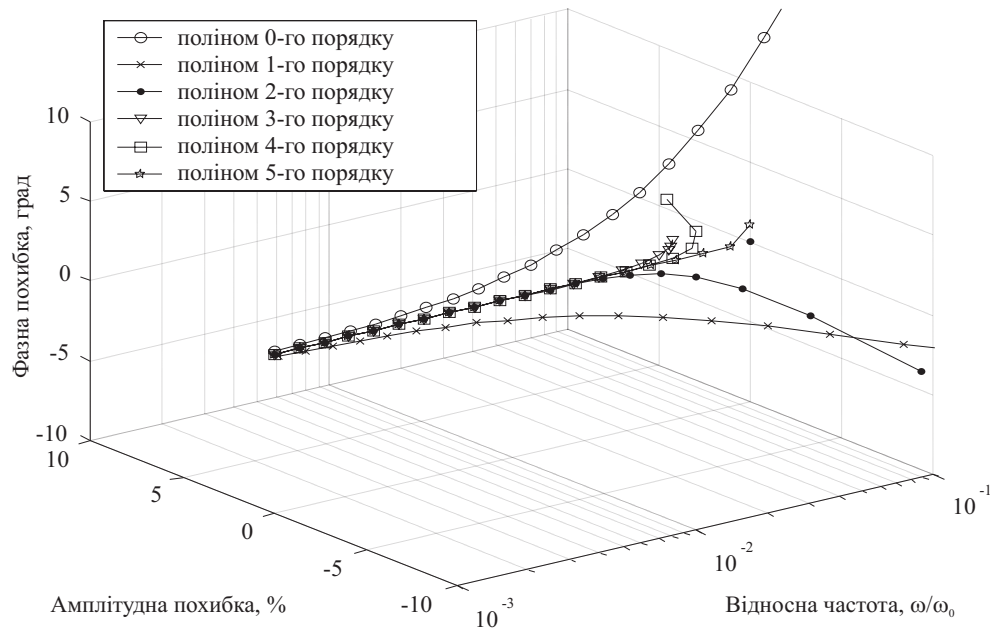


Рисунок 2 – Частотні похибки апроксимацій за явною схемою

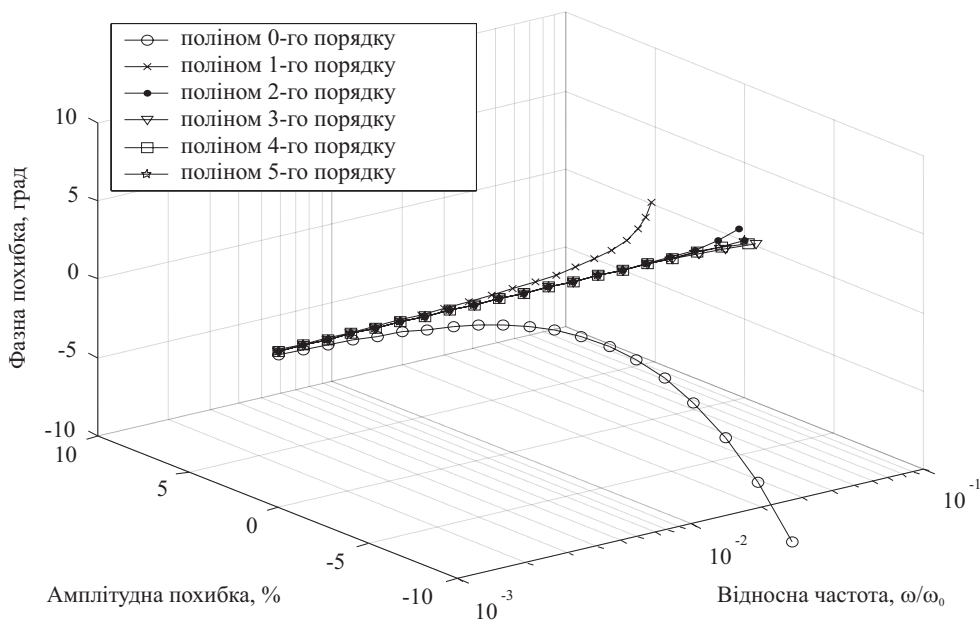


Рисунок 3 – Частотні похибки апроксимацій за неявною схемою

ВИСНОВКИ

Проведений аналіз частотних похибок апроксимаційних поліномів для процесу відновлення дискретного сигналу за його відліками показав:

1) відтворення сигналу за його дискретними відліками без фазних похибок забезпечує лише апроксимація поліномом першого порядку за неявною схемою (трапеціями);

2) у разі можливості вибору перевагу слід надавати лише неявним обчислювальним схемам побудови апроксимаційного полінома, які забезпечують значно менший рівень амплітудних і фазних частотних похибок;

3) порядок апроксимаційного полінома для забезпечення достатньо точного (амплітудна похибка не більше 5 %, фазна – не перевищує 2°) відтворення сигналу може бути обмежений другим-третьим порядком. Застосування апроксимацій вищих порядків невиправдане через:

– незначне зростання точності в нижній частині робочого діапазону (до 1/20 частоти квантування);

– різке зростання похибок у верхній частині робочого діапазону (1/20...1/2 частоти квантування), що може спричинити незадовільну роботу замкнених цифрових систем автоматичного регулювання.

ПЕРЕЛІК ПОСИЛАНЬ

1. Джури Э. Импульсные системы автоматического регулирования. – М.: Физматгиз, 1963. – 456 с.
2. Ту Ю. Цифровые и импульсные системы автоматического управления. – М.: Машиностроение, 1964. – 703 с.
3. Изерман Р. Цифровые системы управления: Пер. с англ. – М.: Мир, 1984. – 541 с.
4. Куо Б. Теория и проектирование цифровых систем управления. – М.: Машиностроение, 1986. – 448 с.
5. Олссон Г., Пиани Дж. Цифровые системы автоматизации и управления. – СПб.: Невский Диалект, 2001. – 557 с.
6. Sanz R., Ārzén K.-E. Trends in Software and Control // IEEE Control Systems Magazine. – 2003, June. – Pp. 12–15.
7. Мороз В. Застосування інтегралу згортки для синтезу цифрових систем // Вісник Хмельницького національного університету. Т. 2. Технічні науки. – 2007. – № 2. – С. 75–78.
8. Гришина Т. Ф. Определение передаточной функции линейной системы по кривой переходного процесса // Изв. вузов. Электромеханика. – 1969. – № 7. – С. 762–768.
9. Фильц Р. В. Математические основы теории электромеханических преобразователей. – К.: Наукова думка, 1979. – 208 с.
10. A. Alleyne, S. Brennan, B. Rasmussen, R. Zhang, Y. Zhang. Controls and Experiments: Lessons Learned // IEEE Control Systems Magazine. – 2003. – October. – Pp. 20–34.

Надійшла 20.03.2008

С использованием частотных характеристик проведено исследование рационального порядка аппроксимационного полинома для восстановления сигнала за его отсчетами.

The rational polynomial approximation order of sampling signal reconstruction was analyzed using Bode plots.

УДК 519.876.5

А. Г. Овский, В. А. Толок

ИСПОЛЬЗОВАНИЕ СИСТЕМЫ КОМПЬЮТЕРНОЙ МАТЕМАТИКИ MAPLE ДЛЯ ДОКАЗАТЕЛЬСТВА ЗАКОНА ОРТОГОНАЛЬНОСТИ МАТРИЦ ПРЯМОГО И ОБРАТНОГО ПРЕОБРАЗОВАНИЙ, СОСТАВЛЕННЫХ ВЛАСОВЫМ В. З.

Авторы анализируют полученное В. З. Власовым [1] общее операторное решение трехмерных уравнений теории упругости, проверяют его основные свойства в системе программирования Maple. В разработанной программе доказывается закон ортогональности матриц прямого и обратного преобразований, составленных из полученных операторов. В работе используется упрощающая символическая запись в виде трансцендентных операционных формул, которая позволяет применять ЭВМ для построения математических моделей задач теории упругости.

© Овский А. Г., Толок В. А., 2008

ВВЕДЕНИЕ

В настоящее время, в связи с появлением новых конструктивных материалов, возникает необходимость в росте прочности машин и конструкций с параллельным снижением их себестоимости и расходов на их обслуживание. В связи с этим, остро встает вопрос о разработке новых математических методов для расчета напряженно-деформируемого состояния различных тел и конструкций [2].

В рассматриваемой работе показана возможность реализации символики Власова В. З. в системе программирования Maple [3]. Проверены свойства линейных дифференциальных операторов, которые получаются в результате работы созданной авторами программы. С помощью этих операторов строится общее решение трехмерных дифференциальных уравнений теории упругости. В своей работе [1] В. З. Власов предполагал ортогональность матриц прямого и обратного преобразований исходя из физических характеристик изотропного упругого тела. Цель данной работы, с помощью аналитических преобразований, которые реализуются в системе MAPLE, доказать это предположение. Удовлетворение линейных дифференциальных операторов этому закону служит оценкой правильности получаемого в программе общего решения трехмерных дифференциальных уравнений теории упругости.

ПОСТАНОВКА И РЕШЕНИЕ ЗАДАЧИ

Рассмотрим общую задачу о равновесии твердого изотропного упругого тела, испытывающего малые деформации. Эта задача описывается в трехмерной прямоугольной Декартовой системе координат x, y, z известными дифференциальными уравнениями:

$$\begin{cases} \frac{\partial U}{\partial z} = -\frac{\partial W}{\partial x} + X, \\ \frac{\partial V}{\partial z} = \frac{\partial W}{\partial y} + Y, \\ \frac{\partial W}{\partial z} = \frac{\nu}{1-\nu} \left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} \right) + \frac{1-2\nu}{2(1-\nu)} Z, \\ \frac{\partial Z}{z} = \frac{\partial X}{\partial x} - \frac{\partial Y}{y} - c, \\ \frac{\partial Y}{\partial z} = \frac{1+\nu}{1-\nu} \frac{\partial^2 U}{\partial x \partial y} - \left(\frac{\partial^2 V}{\partial x^2} + \frac{2}{1-\nu} \frac{\partial^2 V}{\partial y^2} \right) - \frac{\nu}{1-\nu} \frac{\partial Z}{\partial y} - b, \\ \frac{\partial X}{\partial z} = \frac{1+\nu}{1-\nu} \frac{\partial^2 V}{\partial x \partial y} - \left(\frac{\partial^2 U}{\partial y^2} + \frac{2}{1-\nu} \frac{\partial^2 U}{\partial x^2} \right) - \frac{\nu}{1-\nu} \frac{\partial Z}{\partial x} - a, \end{cases} \quad (1)$$

$$\begin{cases} \sigma_x = \frac{2}{1-2\nu} \left[(1-\nu) \frac{\partial U}{\partial x} + \nu \left(\frac{\partial V}{\partial y} + \frac{\partial W}{\partial z} \right) \right], \\ \sigma_y = \frac{2}{1-2\nu} \left[(1-\nu) \frac{\partial V}{\partial y} + \nu \left(\frac{\partial U}{\partial x} + \frac{\partial W}{\partial z} \right) \right], \\ \tau_{xy} = \tau_{yx} = \frac{\partial U}{\partial y} + \frac{\partial V}{\partial x}, \end{cases} \quad (2)$$

где ν – коэффициент Пуассона тела; $U = Gu$, $V = Gv$, $W = Gw$ – пропорциональные величины для перемещений $u = u(x, y, z)$, $v = v(x, y, z)$, $w = w(x, y, z)$; $X = \tau_{xz}$, $Y = \tau_{yz}$, $Z = \sigma_z$ – обозначения напряжений.

Используем математический аппарат Власова [1], принимаем за искомые основные функции: U, V, W и напряжения X, Y, Z .

Решение системы (1) будем искать в виде бигармонических уравнений:

$$\begin{cases} U = U_0(x, y) \cos(\gamma z) + U_1(x, y) \sin(\gamma z) + \\ + U_2 z \cos(\gamma z) + U_3 z \sin(\gamma z), \\ V = V_0(x, y) \cos(\gamma z) + V_1(x, y) \sin(\gamma z) + \\ + V_2 z \cos(\gamma z) + V_3 z \sin(\gamma z), \\ W = W_0(x, y) \cos(\gamma z) + W_1(x, y) \sin(\gamma z) + \\ + W_2 z \cos(\gamma z) + W_3 z \sin(\gamma z), \\ X = X_0(x, y) \cos(\gamma z) + X_1(x, y) \sin(\gamma z) + \\ + X_2 z \cos(\gamma z) + X_3 z \sin(\gamma z), \\ Y = Y_0(x, y) \cos(\gamma z) + Y_1(x, y) \sin(\gamma z) + \\ + Y_2 z \cos(\gamma z) + Y_3 z \sin(\gamma z), \\ Z = Z_0(x, y) \cos(\gamma z) + Z_1(x, y) \sin(\gamma z) + \\ + Z_2 z \cos(\gamma z) + Z_3 z \sin(\gamma z), \end{cases} \quad (3)$$

где $U_0(x, y), V_0(x, y), W_0(x, y), X_0(x, y), Y_0(x, y), Z_0(x, y)$ – функции, задаваемые на плоскости $z = 0$ (начальные);

$\gamma = \sqrt{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}}$ – дифференциальный оператор от функций;

$\gamma z = z \sqrt{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}}$ – произведение переменной z на оператор;

$\sin(\gamma z) = \gamma z - \frac{\gamma^3 z^3}{3!} + \frac{\gamma^5 z^5}{5!} - \dots$ – бесконечный операционный ряд, разложение тригонометрической функции \sin в ряд Маклорена;

$\cos(\gamma z) = 1 - \frac{\gamma^2 z^2}{2!} + \frac{\gamma^4 z^4}{4!} - \dots$ – бесконечный операционный ряд, разложение тригонометрической функции \cos в ряд Маклорена.

Вводим упрощающую символику, операторы дифференцирования $\frac{\partial}{\partial x}$ и $\frac{\partial}{\partial y}$ заменяем на α и β . Далее используя результаты работы [4] получаем решение системы (1) в MAPLE:

$$\begin{cases} U = L_{UU}U_0 + L_{UV}V_0 + \dots + L_{UZ}Z_0, \\ V = L_{VU}U_0 + L_{VV}V_0 + \dots + L_{VZ}Z_0, \\ \dots \\ Z = L_{ZU}U_0 + L_{ZV}V_0 + \dots + L_{ZZ}Z_0, \end{cases} \quad (4)$$

где

$$\begin{aligned}
 L_{UU} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1z\sin(\gamma z)\alpha^2}{2(-1+v)\gamma}, \\
 L_{VU} &:= \frac{1\alpha\beta z\sin(\gamma z)}{2\gamma(-1+v)}, \\
 L_{WU} &:= \frac{1z\cos(\gamma z)\alpha}{2(-1+v)} + \frac{1(-1+2v)\alpha\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{XU} &:= \frac{\alpha^2 z\cos(\gamma z)}{-1+v} - \frac{(-\alpha^2 + \beta^2 v - \beta^2)\sin(\gamma z)}{(-1+v)\gamma}, \\
 L_{YU} &:= \frac{z\cos(\gamma z)\alpha\beta}{-1+v} + \frac{\alpha\beta\sin(\gamma z)v}{(-1+v)\gamma}, \\
 L_{ZU} &:= -\frac{\gamma\sin(\gamma z)z\alpha}{-1+v}, \\
 L_{UV} &:= \frac{1\alpha\beta z\sin(\gamma z)}{2\gamma(-1+v)}, \\
 L_{VV} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1\beta^2 z\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{WV} &:= \frac{1z\cos(\gamma z)\beta}{2(-1+v)} + \frac{1(-1+2v)\beta\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{XV} &:= \frac{z\cos(\gamma z)\alpha\beta}{-1+v} + \frac{\alpha\beta\sin(\gamma z)v}{(-1+v)\gamma}, \\
 L_{YV} &:= \frac{\beta^2 z\cos(\gamma z)}{-1+v} - \frac{(-\alpha^2 + v\alpha^2 - \beta^2)\sin(\gamma z)}{(-1+v)\gamma}, \\
 L_{ZV} &:= -\frac{\gamma\sin(\gamma z)z\beta}{-1+v}, \\
 L_{UW} &:= \frac{1z\cos(\gamma z)\alpha}{2(-1+v)} - \frac{1(-1+2v)\alpha\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{VW} &:= \frac{1z\cos(\gamma z)\beta}{2(-1+v)} - \frac{1(-1+2v)\alpha\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{WW} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1(-z\beta^2 - z\alpha^2)\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{XW} &:= -\frac{\gamma\sin(\gamma z)z\alpha}{-1+v}, \\
 L_{YW} &:= -\frac{\gamma\sin(\gamma z)z\beta}{-1+v}, \\
 L_{ZW} &:= -\frac{(z\alpha^2 + z\beta^2)\cos(\gamma z)}{-1+v} - \frac{(-\beta^2 - \alpha^2)\sin(\gamma z)}{(-1+v)\gamma}, \\
 L_{UX} &:= -\frac{1\alpha^2 z\cos(\gamma z)}{4(-1+v)\gamma^2} + \frac{1(4v\alpha^2 + 4\beta^2 v - 4\beta^2 - 3\alpha^2)\sin(\gamma z)}{4(-1+v)\gamma^3}, \\
 L_{VX} &:= \frac{1\alpha\beta\sin(\gamma z)}{4\gamma^3(-1+v)} - \frac{1\alpha\beta z\cos(\gamma z)}{4\gamma^2(-1+v)}, \\
 L_{WX} &:= \frac{1\alpha z\sin(\gamma z)}{4\gamma(-1+v)}, \\
 L_{XX} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1z\sin(\gamma z)\alpha^2}{2(-1+v)\gamma}, \\
 L_{YX} &:= \frac{1\alpha\beta z\sin(\gamma z)}{2\gamma(-1+v)}, \\
 L_{ZX} &:= \frac{1z\cos(\gamma z)\alpha}{2(-1+v)} - \frac{1(-1+2v)\alpha\sin(\gamma z)}{2(-1+v)\gamma},
 \end{aligned}$$

$$\begin{aligned}
 L_{UY} &:= \frac{1\alpha\beta\sin(\gamma z)}{4(-1+v)} - \frac{1\alpha\beta z\cos(\gamma z)}{4\gamma^2(-1+v)}, \\
 L_{VY} &:= -\frac{1\beta^2 z\cos(\gamma z)}{4(-1+v)\gamma^2} + \frac{1(4\beta^2 v + 4v\alpha^2 - 4\alpha^2 - 3\beta^2)\sin(\gamma z)}{4(-1+v)\gamma^3}, \\
 L_{WY} &:= \frac{1\beta z\sin(\gamma z)}{4\gamma(-1+v)}, \\
 L_{XY} &:= \frac{1\alpha\beta z\sin(\gamma z)}{2\gamma(-1+v)}, \\
 L_{YY} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1\beta^2 z\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{ZY} &:= \frac{1z\cos(\gamma z)\beta}{2(-1+v)} - \frac{1(-1+2v)\beta\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{UZ} &:= \frac{1\alpha z\sin(\gamma z)}{4\gamma(-1+v)}, \\
 L_{VZ} &:= \frac{1\beta z\sin(\gamma z)}{4\gamma(-1+v)}, \\
 L_{WZ} &:= \frac{1z\cos(\gamma z)}{4(-1+v)} + \frac{1(4v-3)\sin(\gamma z)}{4(-1+v)\gamma}, \\
 L_{XZ} &:= \frac{1z\cos(\gamma z)\alpha}{2(-1+v)} + \frac{1(-1+2v)\alpha\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{YZ} &:= \frac{1z\cos(\gamma z)\beta}{2(-1+v)} + \frac{1(-1+2v)\beta\sin(\gamma z)}{2(-1+v)\gamma}, \\
 L_{ZZ} &:= \frac{1(-2+2v)\cos(\gamma z)}{2(-1+v)} + \frac{1(-z\beta^2 - z\alpha^2)\sin(\gamma z)}{2(-1+v)\gamma}. \quad (5)
 \end{aligned}$$

Операторы $L_{UU}, L_{UV}, \dots, L_{ZZ}$ называются линейными дифференциальными операторами, которые относятся к начальным функциям $U_0(x, y), V_0(x, y), W_0(x, y), X_0(x, y), Y_0(x, y), Z_0(x, y)$.

Множество из 36 операторов $L_{UU}, L_{UV}, \dots, L_{ZZ}$ определяет матрицу прямого линейного дифференциального преобразования, предложенную Власовым В. З.:

$$\begin{pmatrix}
 L_{UU} & L_{UV} & L_{UW} & L_{UX} & L_{UY} & L_{UZ} \\
 L_{VU} & L_{VV} & L_{VW} & L_{VX} & L_{VY} & L_{VZ} \\
 L_{WU} & L_{WV} & L_{WW} & L_{WX} & L_{WY} & L_{WZ} \\
 L_{XU} & L_{XV} & L_{XW} & L_{XX} & L_{XY} & L_{XZ} \\
 L_{YU} & L_{YV} & L_{YW} & L_{YX} & L_{YY} & L_{YZ} \\
 L_{ZU} & L_{ZV} & L_{ZW} & L_{ZX} & L_{ZY} & L_{ZZ}
 \end{pmatrix}. \quad (6)$$

Умножив матрицу (6) на вектор $\begin{pmatrix} U_0 \\ V_0 \\ W_0 \\ X_0 \\ Y_0 \\ Z_0 \end{pmatrix}$, получим

формулы (4).

Если в формулах (4) считать U, V, W, X, Y, Z заданными, а $U_0, V_0, W_0, X_0, Y_0, Z_0$ искомыми, будем иметь обратное преобразование. В этом случае задача сводится к интегрированию системы из шести совместных дифференциальных уравнений в частных производных. Это сложная задача, но ее можно решить иначе, если выходить из физического содержания проблемы.

Принимая плоскость $z = \text{const}$ за начальную, функции U, V, W, X, Y, Z за заданные, а функции $U_0, V_0, W_0, X_0, Y_0, Z_0$ – за искомые и предоставляя координате z отрицательное значение и исходя из формул операторов (одни из них изменяют знак при подстановке, а другие нет), учитывая физический смысл задачи, получим:

$$\begin{cases} U_0 = L_{UU}U + L_{UV}V + L_{UW}W - L_{UX}X - L_{UY}Y + L_{UZ}Z, \\ V_0 = L_{VU}U + L_{VV}V - L_{VW}W - L_{VX}X - L_{VY}Y + L_{VZ}Z, \\ W_0 = -L_{WU}U - L_{WV}V + L_{WW}W + L_{WX}X + L_{WY}Y - L_{WZ}Z, \\ X_0 = -L_{XU}U - L_{XV}V + L_{XW}W + L_{XX}X + L_{XY}Y - L_{XZ}Z, \\ Y_0 = -L_{YU}U - L_{YV}V + L_{YW}W + L_{YX}X + L_{YY}Y - L_{YZ}Z, \\ Z_0 = L_{ZU}U + L_{ZV}V - L_{ZW}W - L_{ZX}X - L_{ZY}Y + L_{ZZ}Z. \end{cases} \quad (7)$$

Эти формулы позволяют определить матрицу обратного линейного дифференциального преобразования, предложенную Власовым В. З. [1]. Вид матрицы обратного линейного дифференциального преобразования аналогичен виду матрицы (7) за исключением знаков перед операторами $L_{UU}, L_{UV}, \dots, L_{ZZ}$:

$$\begin{pmatrix} L_{UU} & L_{UV} & -L_{UW} & -L_{UX} & -L_{UY} & L_{UZ} \\ L_{VU} & L_{VV} & -L_{VW} & -L_{VX} & -L_{VY} & L_{VZ} \\ -L_{WU} & -L_{WV} & L_{WW} & L_{WX} & L_{WY} & -L_{WZ} \\ -L_{XU} & -L_{XV} & L_{XW} & L_{XX} & L_{XY} & -L_{XZ} \\ -L_{YU} & -L_{YV} & L_{YW} & L_{YX} & L_{YY} & -L_{YZ} \\ L_{ZU} & L_{ZV} & -L_{ZW} & -L_{ZX} & -L_{ZY} & L_{ZZ} \end{pmatrix}. \quad (8)$$

В книге Власова [1] на странице 369 авторами была исправлена опечатка: оператор L_{XY} брался со знаком «-», в то время как нужно его брать со знаком «+». Благодаря доказательству закона ортогональности и соответствующему выводу программы, она была исправлена. Подставляя функции $U_0, V_0, W_0, X_0, Y_0, Z_0$, которые определяются формулами (7) в правые части равенств прямого преобразования (4) получим тождественные уравнения. Это свойство выполняется и наоборот: для функций U, V, W, X, Y, Z . Чтобы убедиться в этом, необходимо формулы (4) подставить в (7). Отсюда вытекает, что преобразования (4) и (7) ортогональные. Убедиться в правильности обратного

линейного дифференциального преобразования можно с помощью умножения двух матриц (6) и (8). В результате получим матрицу с единичной главной диагональю. Определители матриц прямого и обратного преобразований равны 1.

Операторы преобразования (4) являются взаимными (в этом можно убедиться, проанализировав результаты работы программы), т. е.:

$$L_{VU} = L_{XY}, \quad L_{UV} = L_{YX} \text{ и т. д.} \quad (9)$$

Поэтому матрицы преобразований имеют симметричную структуру с диагоналями симметрии, которые проходят через правый верхний угол и левый нижний угол каждой из матриц [1].

Взаимность операторов L_{VU} и L_{UV} и симметричных с ними операторов L_{XY} и L_{YX} преобразований (4) и (7) обуславливается изотропностью упругого тела относительно оси z . Мы получаем таким образом для операторов прямого и обратного преобразований следующие выражения:

$$\begin{cases} L_{YU} = L_{XV}, L_{ZU} = L_{XW}, L_{WU} = L_{XZ}, \\ L_{VU} = L_{XY}, L_{UU} = L_{XX}, L_{UV} = L_{YX}, \\ L_{UW} = L_{ZX}, L_{UZ} = L_{WX}, L_{UY} = L_{VX}. \end{cases} \quad (10)$$

РЕЗУЛЬТАТЫ

Результатом работы программы являются: построенные и аналитически выведенные дифференциальные операторы Власова в виде трансцендентных операционных формул (5). А также полностью аналитически проверенное свойство ортогональности матриц прямого и обратного преобразований. Результат проверки ортогональности для начальной функции напряжения X_0 показан ниже, рис. 1. По подобной схеме производятся проверки для остальных функций. Выражения вывода программы представлены в машинно-аналитической форме. К сожалению, система не производит полное приведение подобных членов, поэтому форма представления результатов усложнена.

Программа автоматически выполняет также перемножение матриц прямого и обратного преобразований и вычисляет определители каждой из матриц. Результаты этих операций в статье не приведены, в статье показана основная схема, по которой происходит аналитическое упрощение для каждой из функций.

Результаты на рис. 1 требуют дополнительного анализа, так как мы имеем дело с операторно-символическим способом записи. Однако согласно с методом начальных функций Власова В. З. [1] в формулах (4) и (5) при разложении в ряд Маклорена операционных функций выполняется умножение оператора на соответствующую начальную функцию, а потом производится

«Проверка свойства ортогональности»

$$\begin{aligned}
X_0 := & \left(\left(\frac{2v}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} - \frac{2}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} \right) z \alpha^3 + \left(\frac{2v}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} - \frac{2}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} \right) z \beta^2 + \right. \\
& + \left. \left(\frac{2\sqrt{\alpha^2 + \beta^2}}{(-1+v)^2} - \frac{2\sqrt{\alpha^2 + \beta^2}v}{(-1+v)^2} \right) z \right) \alpha \sin(\sqrt{\alpha^2 + \beta^2}z) \cos(\sqrt{\alpha^2 + \beta^2}z) \omega_0 + \\
& + \left(\left(\left(-\frac{\alpha^2}{4(-1+v)^2} + \frac{\alpha^2 \beta^2}{4(\alpha^2 + \beta^2)(-1+v)^2} + \frac{\alpha^4}{4(-1+v)^2(\alpha^2 + \beta^2)} \right) z^2 + \right. \right. \\
& + \left. \left. \frac{v^2}{(-1+v)^2} + \frac{1}{(-1+v)^2} - \frac{2v}{(-1+v)^2} \right) \cos(\sqrt{\alpha^2 + \beta^2}z)^2 + \right. \\
& + \left. \left(\left(\frac{1}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} - \frac{v}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} \right) z \alpha^4 + \right. \right. \\
& + \left. \left. \left(\frac{1}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} - \frac{v}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} \right) z \beta^2 + \right. \right. \\
& + \left. \left. \left(\frac{v}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} - \frac{1}{(-1+v)^2 \sqrt{\alpha^2 + \beta^2}} \right) z \right) \alpha^2 \sin(\sqrt{\alpha^2 + \beta^2}z) \cos(\sqrt{\alpha^2 + \beta^2}z) + \right. \\
& + \left. \left(\left(-\frac{\alpha^2}{4(-1+v)^2} + \frac{\alpha^2 \beta^2}{4(\alpha^2 + \beta^2)(-1+v)^2} + \frac{\alpha^4}{4(-1+v)^2(\alpha^2 + \beta^2)} \right) z^2 - \right. \right. \\
& - \frac{2\beta^4 v}{(-1+v)^2(\alpha^2 + \beta^2)^2} - \frac{3\alpha^2 \beta^2 v}{(\alpha^2 + \beta^2)^2(-1+v)^2} + \frac{7\beta^2 \alpha^2}{4(-1+v)^2(\alpha^2 + \beta^2)^2} - \frac{\alpha^2 v}{(-1+v)^2(\alpha^2 + \beta^2)} - \\
& - \frac{v\alpha^4}{(-1+v)^2(\alpha^2 + \beta^2)^2} + \frac{v^2 \alpha^2 \beta^2}{(-1+v)^2(\alpha^2 + \beta^2)^2} + \frac{3\alpha^4}{4(-1+v)^2(\alpha^2 + \beta^2)^2} \left. \right) \sin(\sqrt{\alpha^2 + \beta^2}z)^2 x x_0 + \\
& + \left(\left(-\frac{\beta \alpha}{4(-1+v)^2} + \frac{\beta^3 \alpha}{4(-1+v)^2(\alpha^2 + \beta^2)} + \frac{\alpha^3 \beta}{4(\alpha^2 + \beta^2)(-1+v)^2} \right) z^2 \cos(\sqrt{\alpha^2 + \beta^2}z)^2 + \right. \\
& + \left. \left(\left(\frac{1}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} - \frac{v}{(\alpha^2 + \beta^2)^{(3/2)}(-1+v)^2} \right) z \beta \alpha^3 + \right. \right.
\end{aligned}$$

Рисунок 1 – Проверка свойств ортогональности для начальных функций

$$\begin{aligned}
 & + \left(\frac{1}{(\alpha^2 + \beta^2)^{(3/2)}(-1 + v)^2} - \frac{v}{(\alpha^2 + \beta^2)^{(3/2)}(-1 + v)^2} \right) z \beta^3 + \\
 & + \left(\frac{v}{(-1 + v)^2 \sqrt{\alpha^2 + \beta^2}} - \frac{1}{(-1 + v)^2 \sqrt{\alpha^2 + \beta^2}} \right) z \beta \alpha \sin(\sqrt{\alpha^2 + \beta^2} z) \cos(\sqrt{\alpha^2 + \beta^2} z) + \\
 & + \left(\left(-\frac{\beta \alpha}{4(-1 + v)^2} + \frac{\beta^3 \alpha}{4(-1 + v)^2(\alpha^2 + \beta^2)} + \frac{\alpha^3 \beta}{4(\alpha^2 + \beta^2)(-1 + v)^2} \right) z^2 - \frac{\alpha \beta v}{(-1 + v)^2(\alpha^2 + \beta^2)} - \frac{\alpha^3 \beta}{4(\alpha^2 + \beta^2)^2(-1 + v)^2} + \right. \\
 & \left. + \frac{\alpha^3 \beta v}{(-1 + v)^2(\alpha^2 + \beta^2)^2} - \frac{\beta^3 v^2 \alpha}{(-1 + v)^2(\alpha^2 + \beta^2)^2} - \frac{\beta^3 \alpha}{4(-1 + v)^2(\alpha^2 + \beta^2)^2} \right) \sin(\sqrt{\alpha^2 + \beta^2} z)^2 y y_0 + \\
 & + \left(\left(\frac{z^2 \alpha^3}{4(-1 + v)^2 \sqrt{\alpha^2 + \beta^2}} + \left(\frac{z^2 \beta^2}{4(-1 + v)^2 \sqrt{\alpha^2 + \beta^2}} - \frac{z^2 \sqrt{\alpha^2 + \beta^2}}{4(-1 + v)^2} \right) \alpha \right) + \frac{\beta^3 \alpha v}{4(-1 + v)^2(\alpha^2 + \beta^2)^2} - \frac{v^2 \alpha^3 \beta}{(-1 + v)^2(\alpha^2 + \beta^2)^2} + \right. \\
 & \left. + \frac{\alpha \beta}{4(-1 + v)^2(\alpha^2 + \beta^2)} + \frac{\alpha v^2 \beta}{(-1 + v)^2(\alpha^2 + \beta^2)} \sin(\sqrt{\alpha^2 + \beta^2} z) \cos(\sqrt{\alpha^2 + \beta^2} z) + \left(\frac{\beta^2 \alpha v}{(-1 + v)^2(\alpha^2 + \beta^2)} - \frac{3 \alpha \beta^2}{4(\alpha^2 + \beta^2)(-1 + v)^2} - \right. \right. \\
 & \left. \left. - \frac{v \alpha}{(-1 + v)^2} + \frac{3 \alpha}{4(-1 + v)^2} + \frac{\alpha^3 v}{(-1 + v)^2(\alpha^2 + \beta^2)} - \frac{3 \alpha^3}{4(-1 + v)^2(\alpha^2 + \beta^2)} \right) z \sin(\sqrt{\alpha^2 + \beta^2} z)^2 \right) z z_0
 \end{aligned}$$

«После упрощений»

$$U_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) u_0$$

$$V_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) v_0$$

$$W_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) w_0$$

$$X_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) x x_0$$

$$Y_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) y y_0$$

$$Z_0 := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) z z_0$$

«В итоге»

$$U_0 := u_0 \quad V_0 := v_0 \quad W_0 := w_0 \quad X_0 := x x_0 \quad Y_0 := y y_0 \quad Z_0 := z z_0$$

Продолжение рисунка 1

$$\text{Res} := \left(\sin(\text{gam } z)^2 + \cos(\text{gam } z)^2 \right) u_0(x, y)$$

Прямое преобразование Фурье по переменной z :

$$f_{2\sin} := \frac{1}{4} u_0 \left(-\text{fourier}(e^{-2I\text{gam } z}, z, \text{sig}) - \text{fourier}(e^{2I\text{gam } z}, z, \text{sig}) + 4\pi \text{Dirac}(\text{sig}) \right)$$

$$f_{2\cos} := \frac{1}{4} u_0 \left(\text{fourier}(e^{-2I\text{gam } z}, z, \text{sig}) + \text{fourier}(e^{2I\text{gam } z}, z, \text{sig}) + 4\pi \text{Dirac}(\text{sig}) \right)$$

Сумма преобразований

$$\begin{aligned} \text{Res} := & \frac{1}{4} u_0 \left(-\text{fourier}(e^{-2I\text{gam } z}, z, \text{sig}) - \text{fourier}(e^{2I\text{gam } z}, z, \text{sig}) + 4\pi \text{Dirac}(\text{sig}) \right) + \\ & + \frac{1}{4} u_0 \left(\text{fourier}(e^{-2I\text{gam } z}, z, \text{sig}) + \text{fourier}(e^{2I\text{gam } z}, z, \text{sig}) + 4\pi \text{Dirac}(\text{sig}) \right) \end{aligned}$$

$$\text{Res} := 2u_0 \pi \text{Dirac}(\text{sig})$$

Обратное преобразование Фурье по z

$$gt_0 = u_0$$

Рисунок 2 – Доказательство закона ортогональности

«После упрощений»

$$U_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) U$$

$$V_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) V$$

$$W_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) W$$

$$X_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) X$$

$$Y_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) Y$$

$$Z_- := \left(\sin(\sqrt{\alpha^2 + \beta^2} z)^2 + \cos(\sqrt{\alpha^2 + \beta^2} z)^2 \right) Z$$

«В итоге»

$$U_- := U \quad V_- := V \quad W_- := W \quad X_- := X \quad Y_- := Y \quad Z_- := Z$$

Рисунок 3 – Проверка свойств ортогональности для функций, получаемых через начальные функции

замена на дифференциалы. В силу этого, с помощью преобразования Фурье для любого z в созданной программе доказываются равенства из рис. 1. Все это выполняет программа, результаты доказательства для U_0 на рис. 2.

Как видно на результат, существенно влияет знак, стоящий перед преобразованием Фурье каждой из операционных функций, а форма самого преобразования не учитывается, так как все преобразования взаимно упрощаются. Остается лишь преобразование Фурье для функции u_0 .

Аналогично для U (рис. 3).

ЗАКЛЮЧЕНИЕ

В результате проведенного анализа было проверено свойство ортогональности матриц прямого и обратного преобразований. Как следствия из этого, установлены свойства дифференциальных операторов, представимых в виде тригонометрических операционных функций. Эти функции разлагаются в символические бесконечные ряды Маклорена с производными высоких порядков.

Благодаря проверенным выше свойствам в задачах теории упругости можно производить операции с разлагаемыми операционными трансцендентными функциями, как с обычными функциями. На этой основе можно построить новую теорию операторно-символического исчисления для задач теории упругости в двумерной и трехмерной постановке. Но основной смысл предложенного к рассмотрению доказательства состоит в следующем: после проверки ортогональности операторных формул (7) и (8) можно переходить к рассмотрению вариационного принципа минимума по-

тенциальной энергии. Этот принцип, вместе с использованием формул для перемещений и напряжений за методом Власова позволит решать более широкий класс задач теории упругости на системе программирования Maple.

ПЕРЕЧЕНЬ ССЫЛОК

1. Власов В. З., Леонтьев Н. Н. балки плиты и оболочки на упругом основании. – М.: ФИЗМАТГИЗ, 1960. – 491 с.
2. Горшков А. Г., Старовойтов Э. И., Талаковский Д. В. Теория упругости и пластичности. – М.: ФИЗМАТЛИТ, 2002. – с.
3. Дьяконов В. П. Maple 9.5/10 в математике, физике и образовании. – М.: СОЛОН-Пресс, 2006. – 720 с.: ил.
4. Толок В. А., Шапар В. В. Операторно-символьные ряды Власова В. З. в решении задач теории упругости в системе Maple // Гідроакустичний журнал. – 2006. – № 3. – С. 66–74.

Надійшла 1.04.2008

Автори аналізують отримане Власовим В. З. [1] загальне операторне рішення трьохвимірних рівнянь теорії пружності, перевіряють його основні властивості в системі програмування Maple. У розробленій програмі доводиться закон ортогональності матриць прямого і зворотного перетворень, складених з отриманих операторів. У роботі використовується спрощуючий символічний запис у вигляді трансцендентних операційних формул, який дозволяє застосовувати ЕОМ для побудови математичних моделей задач теорії пружності.

Authors analyse got Vlasov V. Z. [1] general statement decision of freedimension equations of theory of resiliency, check up his basic properties in the system of programming of Maple. Law of ortogonal of matrices of direct and reverse transformations, made in obedience to work of Vlasov V. Z. is proved in the developed program. A simplifying symbolic record as transcendent operating formulas, which allows to apply COMPUTER for the construction of mathematical models of tasks of theory of resiliency, is in-process utilized.

УДК 539.3

Е. С. Решевская, С. Н. Гребенюк

МОДЕЛИРОВАНИЕ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ ТРЕХМЕРНЫХ КОНСТРУКЦИЙ НА ОСНОВЕ МЕТОДА КОНЕЧНЫХ ЭЛЕМЕНТОВ С ИНТЕРПОЛИРУЮЩИМ ПОЛИНОМОМ ЭРМИТА

Описана методика построения конечного элемента на основе интерполяционного полинома Эрмита, реализованная в подсистеме «КОЭРМА» программного комплекса «МІРЕЛА+», предназначенного для решения задач теории упругости методом конечных элементов. Для проверки достоверности результатов представлен тестовый пример в качестве нагружаемой плиты, закрепленной по контуру.

© Решевская Е. С., Гребенюк С. Н., 2008

ВВЕДЕНИЕ

Как известно, применение метода конечных элементов в механике деформируемого твердого тела позволяет решать самые разнообразные и сложные задачи теории упругости. Построение решения по данному

методу является достаточно громоздким, особенно в двух и трехмерных случаях. В связи с чем необходимо применение ЭВМ. В настоящее время для облегчения данного процесса разработано ряд программных комплексов, позволяющих автоматизировать процесс построения решения задач методом конечных элементов. Среди них выделяют: PLAXIS 3D Foundation [1], ЛИРА-Windows [2], COSMOSWORKS [3], MSC.Nastran [4], ИСПА [5], ANSYS [6], ASKA [7] и др. Большинство из выше перечисленных комплексов являются широко специализированными дорогостоящими коммерческими системами. Во всем мире большой объем задач решается по данным программам. Однако для правильности проведения расчетов и получения приемлемых результатов необходимо обладать специализированными навыками. Что связано с большой затратой времени и средств.

В настоящее время, в связи с широким применением в машиностроении резиновых виброизолирующих элементов, является актуальной задача расчета напряженно-деформированного состояния эластомерных конструкций. Материалы данного класса обладают отличительной особенностью – слабой сжимаемостью (коэффициент Пуассона имеет значения в интервале $0,49 \leq \nu < 0,5$). Учет данной особенности при решении задачи методом конечных элементов приводит к вырожденности матрицы системы уравнений. Возникает необходимость в модификации метода конечных элементов, позволяющей обойти данную особенность.

В широко специализированных вычислительных комплексах, таких как ANSYS, MSC.Nastran, COSMOSWORKS, для решения возникающей проблемы предложены различные вариации для выражения потенциала упругой энергии деформации для эластомера с различными формами добавки члена, учитывающего слабую сжимаемость. Однако, в большинстве из существующих систем, основанных на методе конечных элементов, все же не предусмотрен учет слабой сжимаемости эластомеров.

Система «МИРЕЛА+» [8], разработанная на базе вычислительного комплекса КОДЕТОМ, является специализированной системой по расчету эластомерных конструкций на прочность, разрушение и долговечность, предназначена для расчета напряженно-деформированного состояния эластомерных конструкций. Программный комплекс обладает возможностью автоматизированного задания дискретной области исследуемого трехмерного объекта, физико-механических характеристик материала. Комплекс основан на единой методике расчета эластомерных элементов методом конечных элементов, позволяющей учитывать жесткие смещения, эффект ложного сдвига и слабую сжимаемость данного материала.

Данная система позволяет производить расчеты эластомерных конструкций различной геометрической

формы. Так, в работе [9] реализован расчет жесткостных характеристик эластомерных виброизоляторов сложной формы. Одним из недавних разработок является реализация расчета контактных задач. Моделирование контактных взаимодействий осуществляется путем задания условий непроникновения контактирующих тел. Решение контактных задач взаимодействия эластомерных виброизоляторов и абсолютно жесткой плиты изложено в работах [10,11].

В рамках системы «МИРЕЛА+» была создана подсистема «КОЭРМА», в которой реализован подход решения задач теории упругости методом конечных элементов, основанном на интерполяционном полиноме Эрмита, учитывающем перемещения и их производные.

Взаимосвязь программ системы КОЭРМА с частями комплекса МИРЕЛА+ схематически изображена на рис. 1.

В приведенной системе реализован расчет конструкций методом конечных элементов. Процесс расчета



Рисунок 1

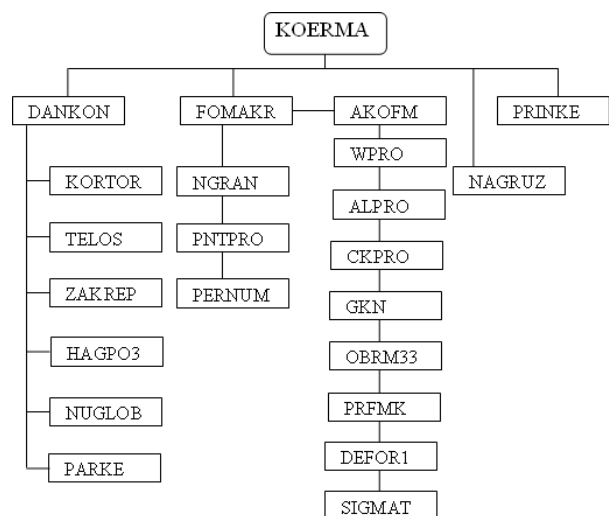


Рисунок 2

состоит из трех взаимосвязанных этапов. На первом этапе задается конечно-элементная дискретизация расчетной схемы, топология и граничные условия исследуемого объекта, физические характеристики материала. На втором этапе рассчитываются локальные матрицы жесткости конечных элементов, строится глобальная матрица жесткости конструкции, составляется разрешающая система уравнений, находится ее решение. Определяются поля деформаций, напряжений, перемещений и их производных и их значения заносятся в файл. Данная функциональная схема представлена на рис. 2.

МЕТОДИКА ПОСТРОЕНИЯ ИНТЕРПОЛЯЦИОННОГО ПОЛИНОМА ЭРМИТА

Метод конечных элементов основан на введении аппроксимирующей функции максимально приближенной к искомой. Данная функция зависит от формы элемента и числа узловых точек. Наиболее распространенным типом аппроксимации являются полиномы, число членов которых равно числу степеней свободы данного конечного элемента.

Интерполирующие функции, обеспечивающие непрерывность поля перемещений относятся к интерполирующим функциям нулевого уровня, если же в элементе непрерывны и первые производные, то обеспечивается C^1 -непрерывность, и, соответственно, при учете непрерывности производных высших порядков – C^r непрерывность. Среди полиномов первого типа следует выделить следующие полиномы: Лагранжа [9], Ньютона [13], Серендипа [9] и др. Кроме того, используется широкий класс ортогональных полиномов, которые вводятся как разложения некоторых функций в обобщенные ряды Фурье [14]: Якоби, Гегенбауэра, Чебышева, Лежандра, Лагера.

Конечно-элементные соотношения, построенные на основе перечисленных многочленов, имеют ряд положительных свойств. Одним из таких свойств является непрерывность функций формы при переходе через границу между элементами. Однако данные элементы не учитывают значения производных интерполируемых величин в узлах и, как следствие, внутри элемента. С целью получения интерполяции, учитывающей производные, для вывода функций форм используется аппроксимирующий полином Эрмита. Так, в работах [15, 16] интерполяционная формула Эрмита представлена через квадраты коэффициентов полинома Лагранжа. Для двумерных прямоугольных конечных элементов, называемых эрмитовыми бикубическими элементами, были получены соотношения для коэффициентов полинома Эрмита [17], удовлетворяющего непрерывности типа C_1 . С этой целью были за-

даны в узлах элемента значения $u(x, y)$, $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial^2 u}{\partial x \partial y}$.

В работе [17] построен интерполирующий полином Эрмита для плоского треугольного конечного элемента, обеспечивающего непрерывность первых производных при переходе от элемента к элементу.

Для построения трехмерной эрмитовой интерполяции необходимо учитывать значения перемещений и семи их частных производных в каждом из восьми узлов объемного параллелепипедного элемента. При использовании данного полинома Эрмита в методе конечных элементов строится матрица жесткости. Для одного конечного элемента размерность 512×512 . Это приводит к большим затратам машинного времени при решении системы разрешающих уравнений, построенных на основе глобальной матрицы жесткости, особенно при расчетах конструкций сложной геометрической формы.

Для уменьшения размерности матрицы жесткости без существенной потери точности решения в качестве параметров эрмитовой интерполяции были приняты узловые перемещения и значения частной производной перемещений. Построен аппроксимирующий многочлен Эрмита, обеспечивающий непрерывность функций $u(x, y, z)$, $\frac{\partial u(x, y, z)}{\partial z}$:

$$P_3(x, y, z) = \sum_{i=0}^8 \left(u_i N_{000i}^{1,1,1} + \left(\frac{\partial u}{\partial z} \right)_i N_{001i}^{1,1,1} \right). \quad (1)$$

Рассмотрен специальный конечный элемент эрмитового семейства параллелепипедной формы. Начало местной системы координат (ξ, η, μ) помещено в центр тяжести элемента, а оси координат расположены параллельно ребрам параллелепипеда. Размеры конечного элемента примем равными $1 \times 1 \times 1$.

На основании принятой аппроксимации полиномом Эрмита (1) были выведены основные соотношения моментной схемы метода конечных элементов для построения матрицы жесткости.

РАСЧЕТ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ ЭЛАСТОМЕРНЫХ КОНСТРУКЦИЙ В СИСТЕМЕ «КОЭРМА»

Для численного исследования точности результатов, полученных при применении вышеуказанной методики, рассмотрим задачи, решения которых были получены аналитически и сравним данные решения со значениями, полученными традиционной и моментной схемой лагранжевого [19] и эрмитового элементов.

Задача 1. Расчет плиты квадратной формы, которая защемлена по контуру, под действием равномерно распределенной поверхностной нагрузки.

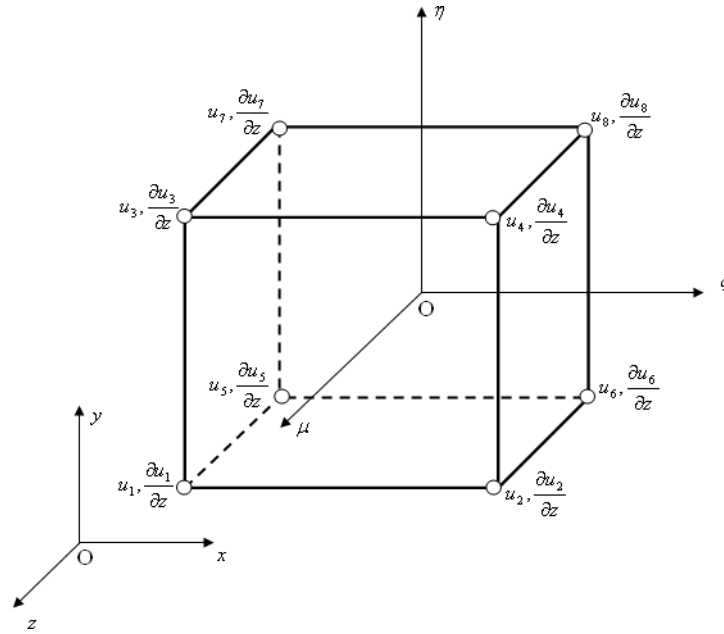


Рисунок 3

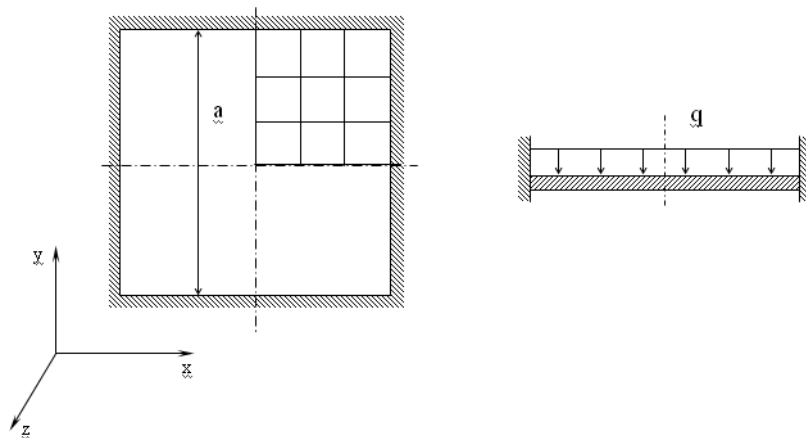


Рисунок 4

Размеры плиты: $a = 0,2$ м, толщина плиты $t = 0,01$ м, поверхностная нагрузка $q = 0,002$ МПа, модуль упругости материала балки $E = 2$ МПа. Расчетная схема изображена на рис. 4.

В связи с тем, что в расчетах кроме аппроксимации перемещений была применена также аппроксимация производной искомой функции по направлению оси z , исследуемая конструкция размещена так, чтобы направление нагрузки, действующей на плиту, совпадало с направлением оси z . Таким образом, будет задаваться производная функции перемещений по направлению наибольших перемещений, что позволит получить более точный результат.

Полученные данные были сравнены со следующими аналитическими решениями: максимальный прогиб в центре плиты [20] при $\nu = 0,3 - w = 2,42 \times 10^{-2}$ м, при $\nu = 0,49 - w = 1,87 \times 10^{-2}$ м, при $\nu = 0,499 - w = 1,85 \times 10^{-2}$ м, при $\nu = 0,4999 - w = 1,848 \times 10^{-2}$ м.

На рис. 5 приведены графики зависимости величины прогиба плиты от коэффициента Пуассона.

Распределение перемещений, возникающих в плите, защемленной по контуру, под действием равномерно распределенной нагрузки приведено на рис. 6.

В табл. 1 приведены результаты расчетов методом конечных элементов при интерполяции искомой величины полиномом Лагранжа и Эрмита.

Таблица 1 – Результаты расчета прямоугольной плиты, защемленной по контуру

Сетка разбиения	Коэффициент Пуассона, ν	МКЭ на основе аппроксимирующего полинома Лагранжа		МКЭ на основе аппроксимирующего полинома Эрмита	
		$w \times 10^{-2}$, м	ε , %	$w \times 10^{-2}$, м	ε , %
5×5×3	0,3	0,34	85,1	0,34	85,1
7×7×5		0,61	73,2	0,65	71,4
8×8×5		0,7	69,2	0,76	66,6
11×11×7		1,13	50,3	1,37	39,8
13×13×9		1,96	13,8	2,0	12,1
5×5×3	0,49	0,37	80,5	0,4	78,9
7×7×5		0,62	67,4	0,8	57,9
8×8×5		0,62	67,4	0,96	49,5
11×11×7		1,05	44,7	1,82	4,2
13×13×9		1,79	5,8	1,84	3,1
5×5×3	0,499	0,37	80,3	0,4	78,7
7×7×5		0,62	67,0	0,8	57,4
8×8×5		0,66	64,8	0,96	48,9
11×11×7		1,06	43,5	1,81	3,6
13×13×9		1,66	11,6	1,82	3,1
5×5×3	0,4999	0,37	80,3	0,4	78,7
7×7×5		0,62	66,9	0,8	57,3
8×8×5		0,67	64,3	0,96	48,8
11×11×7		1,06	43,5	1,82	2,9
13×13×9		1,59	15,2	1,83	2,4

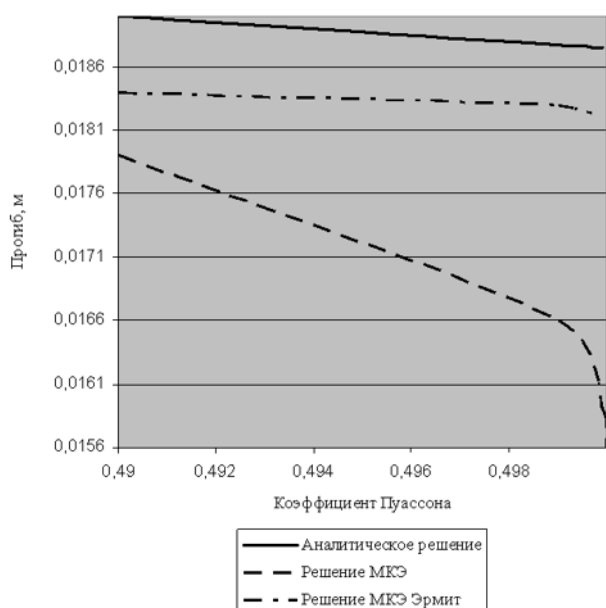


Рисунок 5

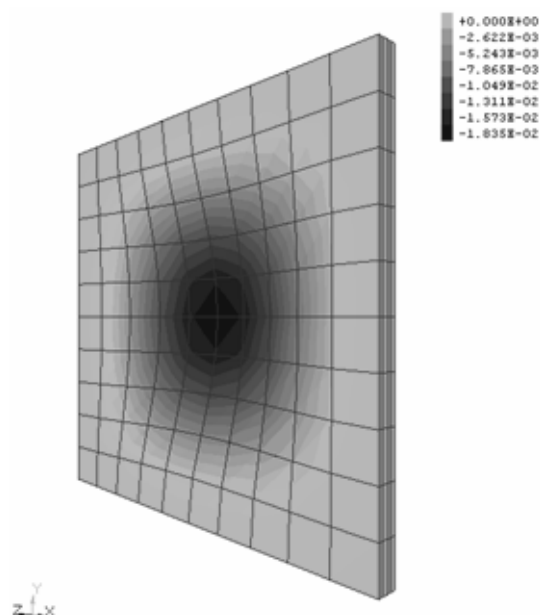


Рисунок 6

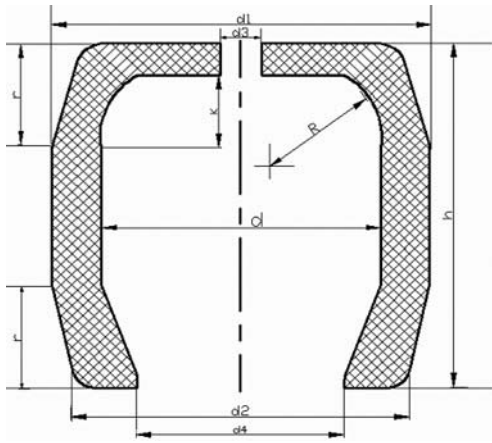


Рисунок 7

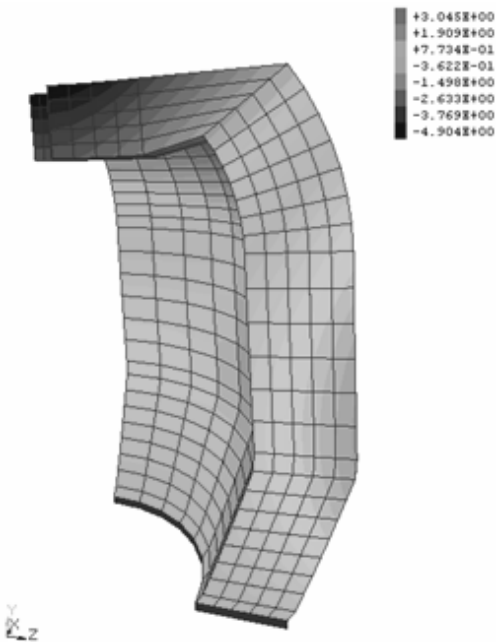


Рисунок 8

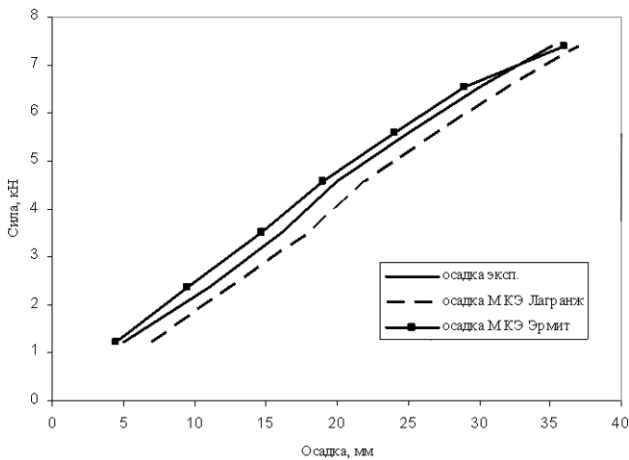


Рисунок 9

Задача 2. Проведен расчет значений осадок эластомерного виброизолятора типа ВР-203 (рис. 7, 8) при различных значениях нагрузок. Марка резины – 2959, модуль сдвига $G_0 = 1,76 \times 10^6$ Па, $\nu = 0,49$. Размеры: $r = 52$ мм, $d = 130$ мм, $d_1 = 200$ мм, $d_2 = 170$ мм, $d_3 = 32$ мм, $d_4 = 90$ мм, $h = 180$ мм, $R = 60$ мм, $k = 38$ мм.

Полученные результаты были сопоставлены с экспериментальными данными (рис. 9). Сравнение значений осадок, рассчитанных при различных усилиях приложенных к виброизолятору, показывает, что метод конечных элементов на основе полинома Эрмита дает лучший результат, чем конечный элемент, построенный с применением полинома Лагранжа.

ВЫВОДЫ

Проведенные расчеты в подсистеме «КОЭРМА» показывают, что метод конечных элементов, основанный на аппроксимации полей перемещений полиномом Эрмита, дает лучший результат, чем при аппроксимации полиномом Лагранжа. В том числе и для слабосжимаемых материалов.

Вышеуказанное утверждение было подтверждено расчетами, проведенными для эластомерной конструкции сложной формы, которые были сравнены с экспериментальными.

ПЕРЕЧЕНЬ ССЫЛОК

1. Программный комплекс конечно-элементных расчетов [Электронный ресурс]: Plaxis представляет собой пакет конечно-элементных программ для двумерных и трехмерных расчетов напряженно-деформированного состояния. – Электрон. дан. – Режим доступа: http://www.pinfor.ru/software/plaxis8/plaxis_foundation.htm – Загл. с экрана.
2. Прокопович А. А. Сопrotивление изгибу железобетонных конструкций с различными условиями сцепления продольной арматуры с бетоном. – Самара: НВФ «Сенсоры. Модули. Системы», 2000. – 567с.
3. COSMOSWorks [Электронный ресурс]: мощный и простой в использовании программный комплекс для проведения инженерных расчетов. – Электрон. дан. – Режим доступа: <http://www.solidworks.ru/products/cosmos>. – Загл. с экрана.
4. MSC.NASTRAN [Электронный ресурс]: расчет и оптимизация конструкций. – Электрон. дан. – Режим доступа: <http://www.bee-pitron.com.ua/cae/nastran.htm>. – Загл. с экрана.
5. Интегрированная Система Прочностного Анализа (ИСПА) [Электронный ресурс]: систему расчета напряженно-деформированного сосконструкций. – Электрон. дан. – Режим доступа: <http://www.ispa-soft.ru/statxi/statxq1.htm>. – Загл. с экрана.
6. Чигарев А. В., Кравчук А. С., Смалюк А. Ф. ANSYS для инженеров. Справ. пособие. – М.: Машиностроение-1, 2004. – 512 с.
7. Скрим Э., Рой Дж. Р. Автоматическая система кинематического анализа // Расчет упругих конструкций с использованием ЭВМ/ Пер. англ. под ред. А. П. Филина: В 2-х т. – Л.: Судостроение, 1974. – Т. 2. – С. 36–67.
8. Метод конечных элементов в вычислительном комплексе «МИРЕЛА+» / Киричевский В. В., Дохняк Б. М., Козуб Ю. Г., Гоменюк С. И., Киричевский Р. В., Гребенюк С. Н. – К.: Наук. думка, 2005. – 403 с.

9. Гребенюк С. Н., Лисица Н. Н., Решевская Е. С., Тархова В. М., Киричевский Р. В. Определение напряженно-деформированного состояния эластомерных виброизоляторов // Геотехнічна механіка: Міжвід. зб. наук. праць. Ін-т геотехнічної механіки ім. М. С. Полякова НАН України. – Дніпропетровськ, 2006. – Вип. 63. – С. 123–130.
10. Киричевський Р. В., Решевська Е. С., Тархова В. М., Прокопенко Е. В., Киричевський В. В. Исследование задач механики деформируемого тела в вычислительном комплексе «МИРЕЛА+» // Радіоелектроніка. Інформатика. Управління. – 2006. – № 2. – С. 54–60.
11. Решевская Е. С., Тархова В. М., Киричевский Р. В. Исследование контактного взаимодействия конического эластомерного виброизолятора // Геотехнічна механіка: Міжвід. зб. наук. праць. Ін-т геотехнічної механіки ім. М. С. Полякова НАН України. – Дніпропетровськ, 2007. – Вип. 71. – С. 222–226.
12. Зенкевич О. Метод конечных элементов в технике. – М.: Мир, 1975. – 541 с.
13. Демидович Б. П., Марон И. А. Основы вычислительной математики. – М.: Наука, 1966. – 522 с.
14. Кузнецов Д. С. Специальные функции. – М.: Высшая школа, 1965. – 641 с.
15. Березин И. С., Жидков Н. П. Методы вычислений. – М.: Наука, 1966. – 346 с.
16. Гончаров В. Л. Теория приближения и интерполирования функций. – М.: Гостехтеориздат, 1954. – 576 с.
17. Стренг Г., Фикс Дж. Теория метода конечных элементов. – М.: Мир, 1977. – 456 с.
18. Коупер Г., Коско Р., Лимберг Г., Олсон М. Применение высокоточных треугольных элементов изгибаемых пластин в статических и динамических задачах // Ракетная техника и космонавтика. – 1969. – № 10. – С. 74–81.
19. Лурье А. И. Пространственные задачи теории упругости. – М.: Гостехиздат, 1955. – 492 с.
20. Самуль В. И. Основы теории упругости и пластичности: Учеб. пособие для инж.-строит. Специальностей вузов. – М.: Высш. школа, 1970. – 288 с.

Надійшла 24.12.2007
Після доробки 13.05.2008

Описано методику побудови скінченного елемента на основі інтерполяційного поліному Ерміту, яка реалізована в підсистемі «КОЭРМА» програмного комплексу «МИРЕЛА+», призначеного для розв'язку задач теорії пружності методом скінчених елементів. Для перевірки вірогідності результатів представлено тестовий приклад в якості плити, що нагружується, закріпленої по контуру.

The technique of construction of a finite element on a basis of interpolational polynomial of Hermite, realized in a subsystem «КОЭРМА» of a program complex «МИРЕЛА+», intended for the decision of tasks of the theory of elasticity by a method of finite elements is described. For check of reliability of results the test example is submitted as a loaded plate fixed on a contour.

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 004.032.26

Е. В. Бодянский, Е. А. Викторов

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ КАСКАДНОЙ ОРТОГОНАЛЬНОЙ НЕЙРОННОЙ СЕТИ

В статье рассматривается новая нетрадиционная нейросетевая архитектура – каскадная ортогональная нейронная сеть, алгоритм ее обучения в пакетном режиме и в режиме реального времени, а также применение этой архитектуры для решения задач прогнозирования и аппроксимации.

ВВЕДЕНИЕ

В настоящее время для решения широкого класса задач, связанных с обработкой информации, заданной в виде временных рядов или массивов числовых данных, порождаемых нестационарными, хаотическими или стохастическими системами, широкое распространение получили искусственные нейронные сети, благодаря своим аппроксимирующим возможностям и способности к обучению. Традиционно под обучением понимается процесс настройки весовых параметров сети в соответствии с той или иной процедурой оптимизации принятого критерия обучения [1, 2].

Улучшить качество получаемого результата можно, настраивая не только синаптические веса, но и архитектуру (количество узлов) самой нейронной сети путем либо увеличения числа нейронов относительно ис-

ходной простой архитектуры – конструктивный подход [3–5], либо путем уменьшения их количества в исходной сложной архитектуре – деструктивный подход [6–8].

С вычислительной точки зрения более привлекательным представляется конструктивный подход, в рамках которого можно выделить каскадные нейронные сети [9–11], наиболее характерным и эффективным представителем которых является Cascade-Correlation Learning Architecture (CasCorLA) [9]. Эти сети, стартуя с простой архитектуры, состоящей из единственного нейрона, в процессе обучения добавляют один за другим новые нейроны, образуя многослойную структуру. При этом в течение каждой эпохи обучения настраивается только один нейрон последнего каскада, все же предыдущие нейроны обрабатывают информацию с «замороженными» весами. Авторы CasCorLA Фальман и Лебьер отмечают высокую скорость обучения и хорошие аппроксимирующие свойства этой сети. Вместе с тем, следует отметить, что в качестве узлов в этой архитектуре используются элементарные перцептроны Розеллатта с функциями активации типа гиперболического тангенса так, что выходной сигнал каждого нейрона

© Бодянский Е. В., Викторов Е. А., 2008

нелинейно зависит от синаптических весов. Отсюда следует, что обучение должно производиться на основе дельта-правила и его модификаций, представляющих собой, по сути, градиентные алгоритмы оптимизации. Очевидно, что говорить об оптимизации скорости обучения в этом случае затруднительно. В связи с этим представляется целесообразным синтезировать каскадную архитектуру, использующую такие искусственные нейроны, в которых выходной сигнал будет линейно зависеть от синаптических весов, что позволит оптимизировать скорость обучения и сократить размер обучающей выборки.

1 ОРТО-НЕЙРОН

Среди множества функциональных структур, используемых для аппроксимации нелинейных зависимостей, особого внимания заслуживают ортогональные полиномы [12, 13]. Они обладают весьма привлекательными свойствами как с вычислительной точки зрения и обеспечиваемой ими точности, так и с точки зрения нейросетевой реализации [14–21].

При этом элементарная одномерная система, описываемая в пространстве «вход – выход» некоторой неизвестной функциональной зависимости $y(x)$ со сколь угодно высокой точностью, может быть представлена следующей суммой:

$$\hat{y} = \hat{f}(x) = w_0\phi_0(x) + w_1\phi_1(x) + \dots + w_h\phi_h(x) = \sum_{j=0}^h w_j\phi_j(x), \quad (1)$$

где x и $y(x)$ – входные и выходная переменные оцениваемого процесса, $\phi_j(x)$ – ортогональный полином j -го порядка ($j = 0, 1, 2, \dots, h$), j, q – неотрицательные целые числа, $k = 1, 2, \dots, N$ – дискретное время или порядковый номер элемента в выборке.

Выражение (1) может быть представлено с помощью элементарной схемы, приведенной на рис. 1 и названной нами орто-синапс.

На рис. 1 x_i – i -я компонента ($i = 1, 2, \dots, n$) сигнала $x = (x_1, x_2, \dots, x_n)^T$, w_{ji} ($j = 1, 2, \dots, h$) – синапти-

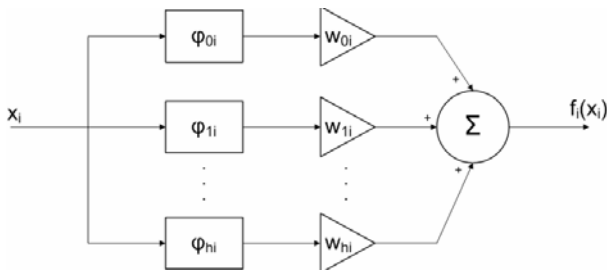


Рисунок 1 – Орто-синапс – OS

ческие веса, подлежащие определению, $f_i(x_i)$ – выходной сигнал орто-синапса, который может быть записан следующим образом:

$$f_i(x_i) = \sum_{j=0}^h w_{ji}\phi_{ji}(x_i). \quad (2)$$

Различные системы ортогональных полиномов могут быть использованы в качестве активационных функций в орто-синапсах. Выбор системы ортогональных полиномов зависит от специфики решаемой задачи. Если входные данные нормализованы на гиперкуб $[-1, 1]^n$, целесообразным представляется использовать систему ортогональных полиномов Лежандра, ортогональных на интервале $[-1, 1]$ с весом $\gamma(x) = 1$:

$$\phi_{ji}^L(x_i) = 2^{-j} \sum_{p=0}^{[j/2]} (-1)^p \frac{(2j-2p)!}{p!(j-p)!(j-2p)!} x_i^{j-2p}. \quad (3)$$

Также для упрощения вычислений может быть использована рекуррентная формула:

$$\phi_{j+1,i}^L(x_i) = \frac{2j+1}{j+1} x_i P_j(x_i) - \frac{j}{j+1} P_{j-1}(x_i). \quad (4)$$

Система ортогональных полиномов Лежандра лучше всего подходит в том случае, когда до построения нейромодели нам известен интервал, в котором изменяются данные. В ином случае может быть использована система ортогональных полиномов Эрмита:

$$H_l(u) = l! \sum_{p=1}^{[l/2]} (-1)^p \frac{(2u)^{l-2p}}{p!(l-2p)!}. \quad (5)$$

Эта система ортогональна на интервале $(-\infty, +\infty)$ с весовой функцией $h(u) = e^{-u^2}$ и дает возможность уменьшить влияние на результат точек, лежащих далеко от начала координат.

Также несложно заметить, что орто-синапс имеет идентичную архитектуру с нелинейным синапсом неофаззи-нейрона [22–24], однако реализует более гладкое полиномиальное приближение с помощью ортогональных полиномов, вместо кусочно-линейной аппроксимации.

Мы использовали орто-синапс в качестве строительного блока для архитектуры, названной нами орто-нейроном и приведенной на рис. 2.

Орто-нейрон имеет такую же архитектуру, как и неофаззи-нейрон, реализующая отображение

$$\hat{y} = \sum_{i=1}^n f_i(x_i) = \sum_{i=1}^n \sum_{j=0}^h w_{ji}\phi_{ji}(x_i), \quad (6)$$

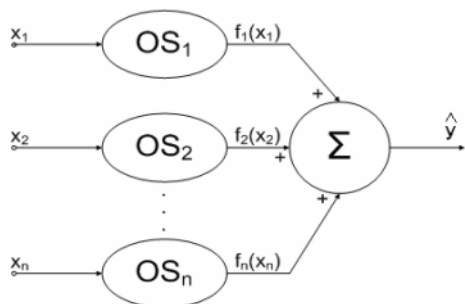


Рисунок 2 – Орто-нейрон – ON

и позволяет добиться высокой точности аппроксимации и экстраполяции существенно нелинейных сигналов и процессов [16, 17, 19–21, 25]. Однако, как будет показано далее, орто-нейрон будет использован нами в качестве элементарного узла в архитектуре, названной каскадная ортогональная нейронная сеть (CONN – Cascade Orthogonal Neural Network).

2 КАСКАДНАЯ ОРТОГОНАЛЬНАЯ НЕЙРОННАЯ СЕТЬ

Архитектура каскадной ортогональной нейронной сети приведена на рис. 3.

Представленная конструкция реализует отображение следующего вида:

– нейрон первого каскада

$$\hat{y}_1 == \sum_{i=1}^n \sum_{j=0}^h w_{ji}^{[1]} \varphi_{ji}(x_i); \tag{7}$$

– нейрон второго каскада

$$\hat{y}_2 == \sum_{i=1}^n \sum_{j=0}^h w_{ji}^{[2]} \varphi_{ji}(x_i) + \sum_{j=0}^h w_{j,n+1}^{[2]} \varphi_{j,n+1}(\hat{y}_1); \tag{8}$$

– нейрон третьего каскада

$$\hat{y}_3 == \sum_{i=1}^n \sum_{j=0}^h w_{ji}^{[3]} \varphi_{ji}(x_i) + \sum_{j=0}^h w_{j,n+1}^{[3]} \varphi_{j,n+1}(\hat{y}_1) + \sum_{j=0}^h w_{j,n+2}^{[3]} \varphi_{j,n+2}(\hat{y}_2); \tag{9}$$

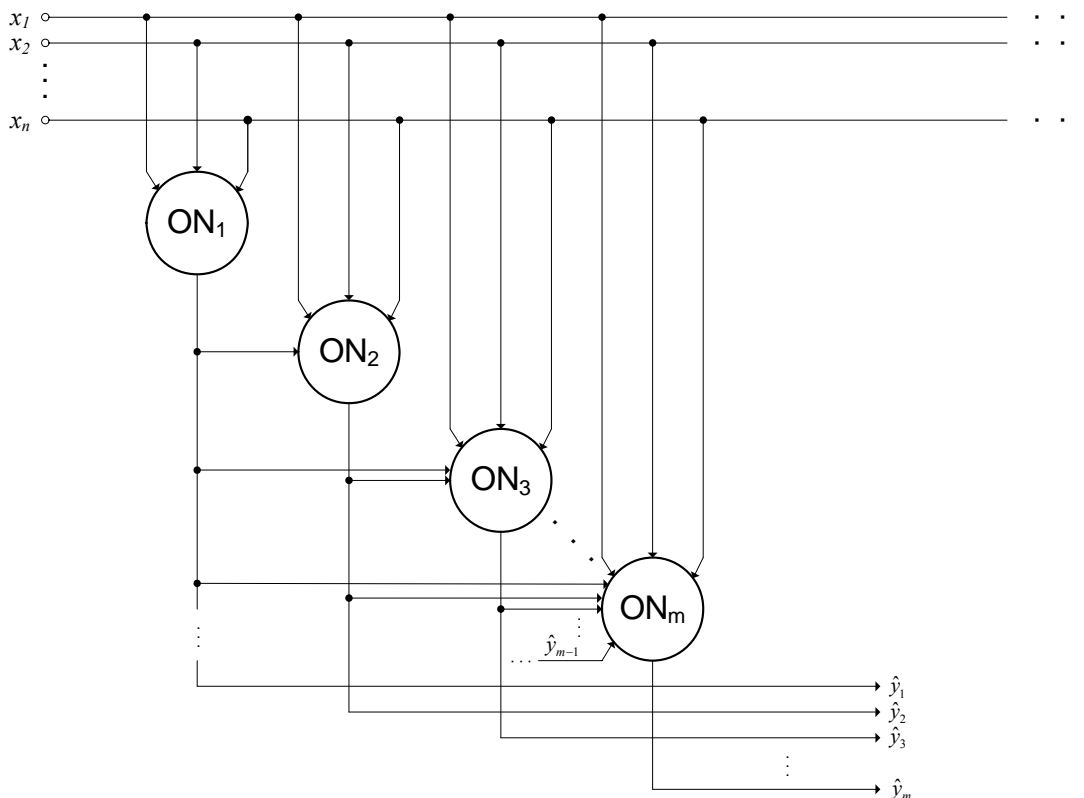


Рисунок 3 – Каскадная ортогональная нейронная сеть

– нейрон m -го каскада

$$\hat{y}_m = \sum_{i=1}^n \sum_{j=0}^h w_{ji}^{[m]} \varphi_{ji}(x_i) + \sum_{i=n+1}^{n+m-1} \sum_{j=0}^h w_{ji}^{[m]} \varphi_{ji}(\hat{y}_{l-n}), \quad (10)$$

где \hat{y}_m – выход m -го каскада нейронной сети; $w_{ji}^{[m]}$ – j -й синаптический весовой коэффициент в i -м орто-синапсе орто-нейрона m -го каскада; φ_{ji} – ортогональный полином i -го порядка в j -м орто-синапсе; h – количество ортогональных полиномов в орто-синапсах, по которым происходит разложение входных сигналов; n – количество орто-синапсов в ортонейроне первого каскада; $(n+m-1)$ – количество орто-синапсов в ортонейроне m -го каскада.

Каскадная ортогональная нейронная сеть содержит

$$(h+1) \left(n + \sum_{l=n+1}^{n+m-1} l \right)$$

настраиваемых параметров и, что

очень важно, все они линейно входят в описание (10).

Вводя далее вектор ортогональных полиномов $(h+1)(n+m-1) \times 1$ m -го орто-нейрона $\varphi^{[m]} = (\varphi_{01}(x_1), \varphi_{11}(x_1), \dots, \varphi_{h1}(x_1), \varphi_{02}(x_2), \dots, \varphi_{h2}(x_2), \dots, \varphi_{ji}(x_i), \dots, \varphi_{hn}(x_n), \varphi_{0,n+1}(\hat{y}_1), \dots, \varphi_{h,n+1}(\hat{y}_1), \dots, \varphi_{h,n+m-1}(\hat{y}_{m-1}))^T$ и соответствующий ему вектор синаптических весов $w^{[m]} = (w_{01}^{[m]}, w_{11}^{[m]}, \dots, w_{h1}^{[m]}, w_{02}^{[m]}, \dots, w_{h2}^{[m]}, \dots, w_{ji}^{[m]}, \dots, w_{hn}^{[m]}, w_{0,n+1}^{[m]}, \dots, w_{h,n+1}^{[m]}, \dots, w_{h,n+m-1}^{[m]})^T$ той же размерности, можно представить (10) в более компактной форме:

$$\hat{y}_m = w^{[m]T} \varphi^{[m]}. \quad (11)$$

3 ОБУЧЕНИЕ КАСКАДНОЙ ОРТОГОНАЛЬНОЙ НЕЙРОННОЙ СЕТИ

Обучение каскадной нейронной сети производится в пакетном режиме с использованием всех элементов обучающей выборки $x(1), y(1); x(2), y(2); \dots; x(k), y(k); \dots; x(N), y(N)$. Сначала вычисляется набор значений ортогональных функций $\varphi^{[1]}(1), \varphi^{[1]}(2), \dots, \varphi^{[1]}(N)$ для каждого образа из обучающей выборки, после чего путем прямой минимизации критерия обучения

$$E_N^{[1]} = \frac{1}{2} \sum_{k=1}^N e_1(k)^2 = \frac{1}{2} \sum_{k=1}^N (y(k) - \hat{y}_1(k))^2, \quad (12)$$

вычисляется вектор синаптических весов

$$w^{[1]}(N) = \left(\sum_{k=1}^N \varphi^{[1]}(k) \varphi^{[1]T}(k) \right)^+ \sum_{k=1}^N \varphi^{[1]}(k) y(k) = P^{[1]}(N) \sum_{k=1}^N \varphi^{[1]}(k) y(k). \quad (13)$$

Если размер обучающей выборки достаточно велик, удобнее использовать процедуру (13) в форме рекуррентного метода наименьших квадратов с последовательной обработкой элементов обучающей выборки:

$$\begin{cases} w^{[1]}(k+1) = w^{[1]}(k) + \frac{P^{[1]}(k)(y(k+1) - w^{[1]T}(k)\varphi^{[1]}(k+1))}{1 + \varphi^{[1]T}(k+1)P^{[1]}(k)\varphi^{[1]}(k+1)} \varphi^{[1]}(k+1), \\ P^{[1]}(k+1) = P^{[1]}(k) - \frac{P^{[1]}(k)\varphi^{[1]}(k+1)\varphi^{[1]T}(k+1)P^{[1]}(k)}{1 + \varphi^{[1]T}(k+1)P^{[1]}(k)\varphi^{[1]}(k+1)}. \end{cases} \quad (14)$$

Необходимо отметить, что использование процедур (13), (14) позволяет существенно сократить время необходимое для настройки весовых коэффициентов, по сравнению с градиентными алгоритмами, в основе которых лежит дельта-правило. Также ортогональность активационных функций обеспечивает численную устойчивость в процессе обращения матриц.

После завершения обучения первого каскада сети, синаптические веса орто-нейрона ON_1 «замораживаются», и генерируется второй каскад, состоящий из орто-нейрона ON_2 , содержащего один дополнительный вход для выходного сигнала предыдущего каскада. Затем процедуры (13), (14) снова применяются для настройки вектора весовых коэффициентов $w^{[2]}$, который имеет размерность $(h+1)(n+1) \times 1$.

Процесс роста нейронной сети (увеличения количества каскадов) продолжается до тех пор, пока не будет получена требуемая точность рассматриваемой задачи. При этом, для оценки весовых коэффициентов последнего m -го каскада используются следующие выражения:

$$w^{[m]}(N) = \left(\sum_{k=1}^N \varphi^{[m]}(k) \varphi^{[m]T}(k) \right)^+ \sum_{k=1}^N \varphi^{[m]}(k) y(k) = P^{[m]}(N) \sum_{k=1}^N \varphi^{[m]}(k) y(k) \quad (15)$$

или

$$\left\{ \begin{aligned} w^{[m]}(k+1) &= w^{[m]}(k) + \\ &+ \frac{P^{[m]}(k)(y(k+1) - w^{[m]T}(k)\phi^{[m]}(k+1))}{1 + \phi^{[m]T}(k+1)P^{[m]}(k)\phi^{[m]}(k+1)}\phi^{[m]}(k+1), \\ P^{[m]}(k+1) &= P^{[m]}(k) - \\ &- \frac{P^{[m]}(k)\phi^{[m]}(k+1)\phi^{[m]T}(k+1)P^{[m]}(k)}{1 + \phi^{[m]T}(k+1)P^{[m]}(k)\phi^{[m]}(k+1)}. \end{aligned} \right. \quad (16)$$

Основным недостатком CasCorLA является возможность обучения только в пакетном режиме, когда вся обучающая выборка задана априорно. Благодаря максимально возможной (квадратичной) скорости сходимости алгоритма (16), каскадная ортогональная нейронная сеть может обучаться и в реальном режиме времени. При этом сразу формируется архитектура, состоящая из m каскадов, каждый из которых обучается с помощью собственного алгоритма. Поскольку для m -го каскада дополнительными входами являются выходы предыдущих орто-нейронов, алгоритм (7) фактически реализует рекуррентный метод ошибки прогноза [26], получивший широкое распространение в теории адаптивной идентификации. Изменение числа каскадов в процессе обучения также не представляет никаких трудностей.

4 РЕЗУЛЬТАТЫ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

При апробации предложенной архитектуры в качестве тестовой решалась задача прогнозирования временного ряда Мэки – Гласса. Этот ряд представляет собой хаотическую, детерминированную последовательность, определяемую следующим дифференциальным уравнением:

$$y'(t) = \frac{0,2t(t-\tau)}{1+y^{10}(t-\tau)} - 0,1y(t). \quad (17)$$

Квантование сигнала было проведено с шагом 0,1. Под обучающую выборку был отведен участок сигнала, содержащий 500 элементов. Целью являлось про-

гнозирование сигнала на шесть шагов вперед по его предыстории: значениям элементов в моменты времени k , $(k-6)$, $(k-12)$ и $(k-18)$. Тестовая выборка содержала 9500 элементов последовательности – значения сигнала с 501 до 10000.

Для оценки полученного прогноза использовалась среднеквадратическая ошибка (18)

$$MSE(k, N) = \frac{\sum_{q=1}^N e_q^2}{N}. \quad (18)$$

Перед тем, как начинать обучение каскадной ортогональной нейронной сети необходимо определить три опциональных параметра: 1) тип системы ортогональных полиномов в каждом орто-синапсе; 2) количество ортогональных полиномов в каждом орто-синапсе; 3) максимальное количество каскадов. Поскольку входные данные были нормализованы на интервал $[-1, 1]$, нами использовалась система ортогональных полиномов Чебышева 1-го рода, что позволило избежать неограниченного роста весовых коэффициентов с увеличением количества каскадов. Основываясь на результатах предыдущих экспериментов, в орто-синапсах использовалось по 3–5 полиномов. А также, для того чтобы предотвратить потерю возможности обобщения сетью, количество каскадов было ограничено до 10.

Результаты прогнозирования временного ряда Мэки – Гласса приведены в табл. 1. Для сравнения эта же задача решалась при помощи Cascade-Correlation Learning Architecture и традиционного многослойного персептрона (MLP), который имел архитектуру 4–7–1 и обучался по алгоритму Левенберга – Марквардта в течение 100 эпох.

Как показывают результаты, каскадная ортогональная сеть предоставляет хорошее качество аппроксимации и прогнозирования существенно нелинейных процессов. Полученная ошибка прогнозирования сравнима с ошибкой, даваемой многослойным персептроном, и намного меньше, чем ошибка CasCorLA. При этом необходимо отметить, что процедура обучения каскадной

Таблица 1 – Ошибки при прогнозировании ряда Мэки – Гласса

Архитектура ИНС	Ошибка на обуч. выборке / Номер каскада										Ошибка на тест. выборке	
	1	2	3	4	5	6	7	8	9	10		
CONN	0,0005	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0003
CasCorLA	0,11	0,1	0,098	0,095	0,092	0,089	0,088	0,084	0,083	0,083		0,055
MLP	Ошибка на обучающей выборке / Номер эпохи										Ошибка на тест. выборке	
	4	8	12	16	20	24	28	32	36	40		
	0,8	0,3	0,09	0,06	0,04	0,03	0,01	0,003	0,001	0,0009		0,0006

ортогональной нейронной сети требует значительно меньше времени и вычислительных ресурсов, чем обратное распространение ошибки или процедура Левенберга – Марквардта. Также, используя каскадную ортогональную нейронную сеть, удается избежать двух существенных недостатков, свойственных CasCorLA и многослойному персептрону: неповторимости полученных результатов и использование нелинейных процедур обучения.

ЗАКЛЮЧЕНИЕ

В статье предложена каскадная ортогональная нейронная сеть, отличающаяся от своего прототипа – Cascade-Correlation Learning Architecture – повышенным быстродействием и точностью, численной устойчивостью и возможностью обработки данных в реальном времени по мере поступления новой информации. Теоретическое обоснование и результаты экспериментов подтверждают эффективность развиваемого подхода к синтезу каскадных нейронных сетей.

ПЕРЕЧЕНЬ ССЫЛОК

1. Cichocki A. Neural Networks for Optimization and Signal Processing / A. Cichocki, R. Unbehauen. – Stuttgart: Teubner, 1993. – 544 p.
2. Haykin S. Neural Networks: A Comprehensive Foundation / S. Haykin. – Upper Saddle River, N. J.: Prentice Hall. – 1999. – 864 p.
3. Platt J. A resource allocating network for function interpolation / J. Platt // Neural Computation. – 1991. – 3. – P. 213–225.
4. Nag A. Flexible resource allocating network for noisy data / A. Nag, J. Ghosh // Proc. SPIE Conf. on Applications and Science of Computational Intelligence. – 1998. – P. 551–559.
5. Yingwei L. Performance evaluation of a sequential minimal radial basis function (RBF) neural network learning algorithm / L. Yingwei, N. Sundararajan, P. Saratchandran // IEEE Trans. on Neural Networks. – 1998. – 9. – P. 308–318.
6. Cun Y. L. Optimal brain damage / Y.L. Cun, J.S. Denker, S.A. Solla // Advances in Neural Information Processing Systems. – San Mateo, CA: Morgan Kaufman, 1990. – 2. – P. 598–605.
7. Hassibi B. Second-order derivatives for network pruning: Optimal brain surgeon / B. Hassibi, D.G. Stork // Advances in Neural Information Processing Systems. Ed. Hanson et al. – San Mateo, CA: Morgan Kaufman, 1993. – P. 551–559.
8. Prechelt L. Connection pruning with static and adaptive pruning schedules / L. Prechelt // Neurocomputing. – 1997. – 16. – P. 49–61.
9. Fahlman S. E. The cascade-correlation learning architecture / S. E. Fahlman, C. Lebiere // Advances in Neural Information Processing Systems. Ed. D. S. Touretzky. – San Mateo, CA: Morgan Kaufman, 1990. – P. 524–532.
10. Schalkoff R. J. Artificial Neural Networks / R. J. Schalkoff. – N.Y.: The McGraw-Hill Comp., 1997. – 528 p.
11. Аведьян Э. Д. Каскадные нейронные сети / Э. Д. Аведьян, Г. В. Баркан, И. К. Левин // Автоматика и телемеханика. – 1999. – № 3. – С. 38–55.
12. Bateman H., Erdelyi A. Higher Transcendental Functions. Vol. 2. / H. Bateman, A. Erdelyi. – N. Y.: The McGraw-Hill Comp. – 1953. – 396 p.
13. Graupe D. Identification of Systems / D. Graupe. – Huntington, N.Y.: Robert E. Kreiger Publishing Comp. – 1976. – 276 p.
14. Scott I., Mulgrew B. Orthonormal function neural network for nonlinear system modeling / I. Scott, B. Mulgrew // Proceedings of the International Conference on Neural Networks. – 1996. – Vol. 4. – P. 1847–1852.
15. Patra J. C. Nonlinear dynamic system identification using Chebyshev functional link artificial neural network / J. C. Patra, A. C. Kot // IEEE Trans. on System, Man and Cybernetics. – 2002. – 32. Part B. – P. 505–511.
16. Bodyanskiy Ye. Artificial neural network with orthogonal activation functions for dynamic system identification / Ye. Bodyanskiy, V. Kolodyazhnyi, O. Slipchenko // Synergies between Information Processing and Automation. Ed. O. Sawodny and P. Scharff – Aachen: Shaker Verlag. – 2004. – P. 24–30.
17. Bodyanskiy Ye. Structural and synaptic adaptation in the artificial neural networks with orthogonal activation functions / Ye. Bodyanskiy, V. Kolodyazhnyi, O. Slipchenko // Sci. Proc. of Riga Technical University. Comp. Sci., Inf. Technology and Management Sci. – 2004. – 20. – P. 69–76.
18. Liyang M. Constructive feedforward neural network using Hermite polynomial activation functions / M. Liyang, K. Khorasani // IEEE Trans. on Neural Networks. – 2005. – 4. – P. 821–833.
19. Bodyanskiy Ye. Growing neural networks based on orthogonal activation functions / Ye. Bodyanskiy, I. Pliss, O. Slipchenko // Proc. XII-th Int. Conf. “Knowledge – Dialog – Solution”. – Varna, 2006. – P. 84–89.
20. Bodyanskiy Ye. Ontogenic neural networks using orthogonal activation functions / Ye. Bodyanskiy, O. Slipchenko // Prace naukowe Akademii Ekonomicznej we Wroclawiu. – 2006. – 21. – P. 13–20.
21. Bodyanskiy Ye. Growing neural network using nonconventional activation functions / Ye. Bodyanskiy, I. Pliss, O. Slipchenko // Int. J. Information Theories & Applications. – 2007. – 14. – P. 275–281.
22. Бодянский Е. В. Ортоинапс, ортонейроны и нейропредиктор на их основе / Е. В. Бодянский, Е. А. Викторов, А. Н. Слипченко // Системи обробки інформації. – Х.: ХУПС, 2007. – Вип. 4(62). – С. 139–143.
23. Yamakawa T. A neo fuzzy neuron and its applications to system identification and prediction of the system behavior / T. Yamakawa, E. Uchino, T. Miki, H. Kusanagi // Proc. 2-nd Int. Conf. on Fuzzy Logic and Neural Networks (LIZUKA-92). – Lizuka, 1992. – P. 477–483.
24. Uchino E. Soft computing based signal prediction, restoration and filtering / E. Uchino, T. Yamakawa // Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks and Genetic Algorithms. Ed. Da Ruan. – Boston: Kluwer Academic Publisher. – 1997. – P. 331–349.
25. Miki T. Analog implementation of neo-fuzzy neuron and its on-board learning / T. Miki, T. Yamakawa // Computational Intelligence and Applications. Ed. N.E. Mastorakis. – Piraeus: WSES Press. – 1999. – P. 144–149.
26. Льюнг Л. Идентификация систем. Теория для пользователя / Л. Льюнг; пер. с англ. Под ред. Я. З. Цыпкина. – М.: Наука, 1991. – 432 с.

Надійшла 14.04.2008

У статті розглянуто нову нетрадиційну нейромережеву архітектуру – каскадну ортогональну нейронну мережу, алгоритм її навчання у пакетному режимі і в режимі реального часу, а також застосування цієї архітектури для вирішення задач прогнозування й апроксимації.

In the article new non-conventional architecture called Cascade Orthogonal Neural Network is considered. Learning algorithms which can operate in batch or real-time mode are given. Also application of this architecture for solving forecasting and approximation problems is proposed.

Т. В. Киприч, В. И. Дубровин

ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МЕТОДИКИ САМОГРАНАЗУЮЩИХСЯ КАРТ КОХОНЕНА ДЛЯ ДИАГНОСТИРОВАНИЯ СОСТОЯНИЙ ПРОЦЕССОВ И СИСТЕМ

Для анализа состояний процессов и систем рассматриваются возможности программных средств, реализующих методику самоорганизующихся карт Кохонена. Разработанное программное обеспечение «Control & Diagnostics System» ориентировано на диагностирование внешних и ошибочных ситуаций, возникающих в ходе работы оборудования. На его основе проведено исследование процесса газодинамической неустойчивости турбокомпрессора по данным стендовых испытаний газотурбинного двигателя.

ВВЕДЕНИЕ

Трудности, возникающие в процессе анализа, моделирования и контроля нелинейных систем, состоят в нахождении характерных состояний или кластеров состояний, которые определяют поведение системы и отображают имеющиеся измерения. Кроме того, при моделировании поведение системы должно быть описано таким образом, чтобы предсказать ее поведение в будущем [1].

Принципы обучения искусственных нейронных сетей (ИНС) позволяют анализировать те состояния систем и процессов, которые затруднительно определить или интерпретировать стандартными средствами математической статистики. Среди ИНС самоорганизующиеся карты Кохонена (СОК) являются средством не-

линейного проецирования, позволяющим отображать характерные состояния и кластера, строить модели, управляемые данными, а также диагностировать недопустимые отклонения в работе машинного оборудования [2].

Существующие программные средства для обработки, кластеризации и визуального представления данных с помощью СОК можно разделить на свободно распространяемое программное обеспечение (ПО); программы собственного производства и коммерческие программные пакеты [3].

В данной работе рассматриваются свободно распространяемые программы, реализующие методику СОК и предназначенные для научно-исследовательских работ.

Среди свободно распространяемого ПО, предназначенного для обработки СОК (табл. 1), широко используются следующие программные пакеты (ПП):

- 1) SOM_PAK – профессиональный пакет для обширных исследований [4];
- 2) NeNet – прост в применении, однако охватывает небольшой объем задач;
- 3) SOM Toolbox – обладает гибкостью и простотой в применении, при этом вычислительные сложности данного пакета невелики [3].

Таблица 1 – Основные особенности свободно распространяемых программных пакетов [3]

Особенности СОК	Свободно распространяемое ПО		
	SOM_PAK	SOM Toolbox	NeNet
Визуализация	U-матрица, плоскости компонент, траектории	U-матрица, плоскости компонент, траектории, гистограммы соответствия	
Качество мониторинга	Ошибка квантования, фиксированные веса Саммона		
Интерфейс пользователя	Командная строка, программирование на C, текстовые файлы	Matlab и GUI	GUI, C++, файлы.txt
Алгоритм	SEQ SOM	SEQ, Batch SOM	SEQ SOM
Топология	Прямоугольная или гексагональная		
Нанесение меток	Автоматическое или ручное		
Размер карты	Неограничен		

Данные ПП охватывают широкий спектр функций работы с СОК, позволяющих упростить выполнение основных операций при создании СОК. Однако при этом ПП не адаптированы для определенных исследовательских задач.

Таким образом, целью работы является описание ПО «Control & Diagnostics System», разработанного непосредственно для диагностирования состояний процессов и систем на основании имеющихся измерений. Продемонстрировать эффективность данного ПО на примере обнаружения помпажных явлений для газотурбинного двигателя (ГТД).

1 ПОДГОТОВКА ДАННЫХ И ПОСТРОЕНИЕ МОДЕЛИ

Разработанное ПО «Control & Diagnostics System» предназначено диагностики состояний процессов и систем на основе СОК. В режиме обучения СОК данный комплекс реализует следующие функции:

1) подготовка данных: ввод примеров для обучения (рис. 1–2), задание параметров исследования (рис. 3), нормировка данных.

В процессе обучения карта настраивается таким образом, чтобы распознать только те состояния процесса, которые представлены имеющимися измерениями [1].

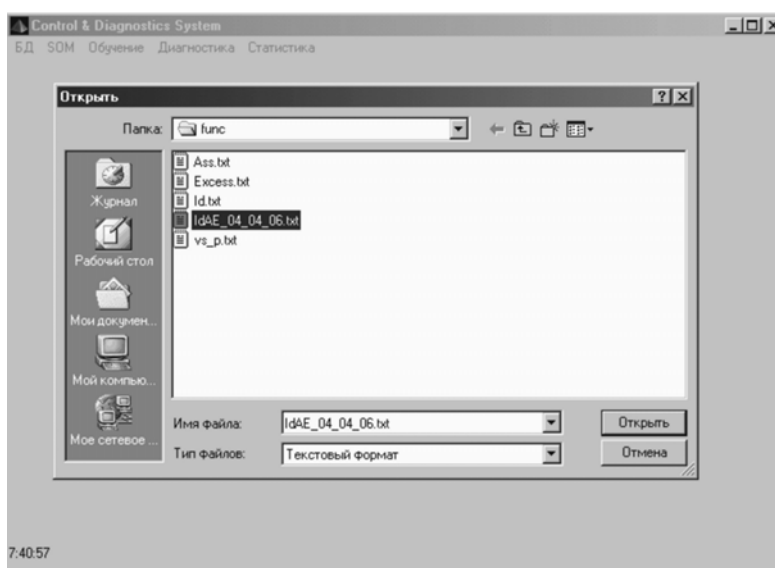


Рисунок 1 – Выбор примеров для обучения

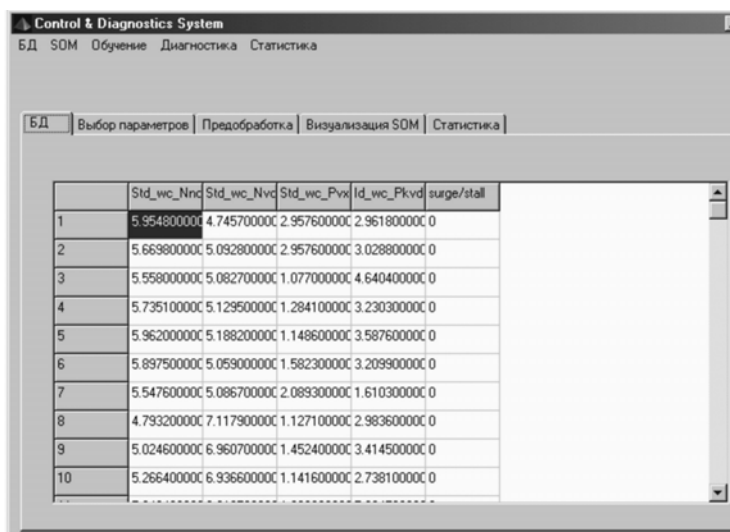


Рисунок 2 – Содержимое файла с примерами для обучения СОК

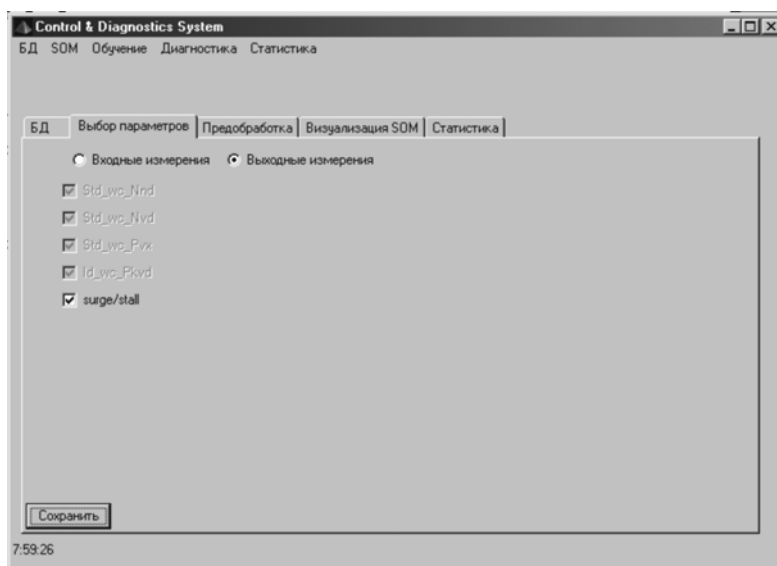
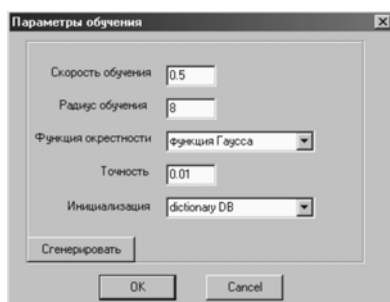
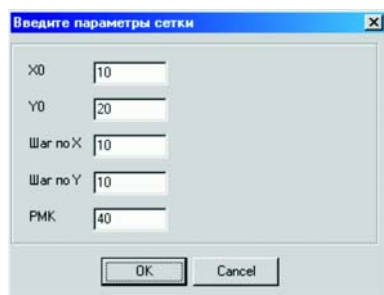


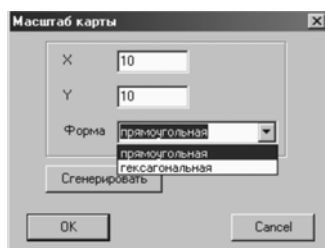
Рисунок 3 – Выбор исходных параметров для обучения СОК



а)



б)



в)

Рисунок 4 – Подготовка СОК: создание и настройка графических параметров:

а – настройка параметров обучения СОК; б – настройка графических размеров СОК; в – выбор конфигурации СОК

Исходными параметрами были выбраны данные, используемые в работе [5] при диагностировании помпажа и вращающегося срыва газотурбинного двигателя (ГТД) (рис. 1):

- среднеквадратические отклонения (σ) для вейвлет-коэффициентов (ВК) (уровень разложения 10, 'db2') следующих сигналов: частоты вращения ротора низкого давления; частоты вращения ротора высокого давления; давление воздуха на входе в компрессор;

- S-дискриминант для ВК (уровень разложения 7) сигнала давления воздуха за компрессором;

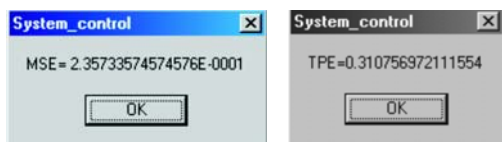
- на выходе модели снимается информация о газодинамическом состоянии двигателя: 0 – нормальный режим работы; 1 – вращающийся срыв, 2 – помпаж.

2) подготовка СОК: выбор конфигурации СОК, настройка параметров обучения, настройка разрешения цветowych карт параметров исследуемого процесса (рис. 4);

На рис. 4, б: X0, Y0 – координаты левого верхнего угла; PMK – расстояние между картами на форме вывода.

3) обучение СОК: получение цветowych карт распределения параметров процесса, а также выдача статистики: ошибок обобщения MSE и топографического упорядочивания нейронов TPE (рис. 5).

В результате обучения были получены цветowe карты исследуемых параметров процесса газодинамической неустойчивости (ГДН) и карта состояния, содержащая две «запрещенные зоны»: помпажа (2) и вращающегося срыва (1) (рис. 6).



а) б)

Рисунок 5 – Погрешности обучения СОК:

а – ошибка обобщения, MSE; б – погрешность топографического упорядочивания нейронов, TPE

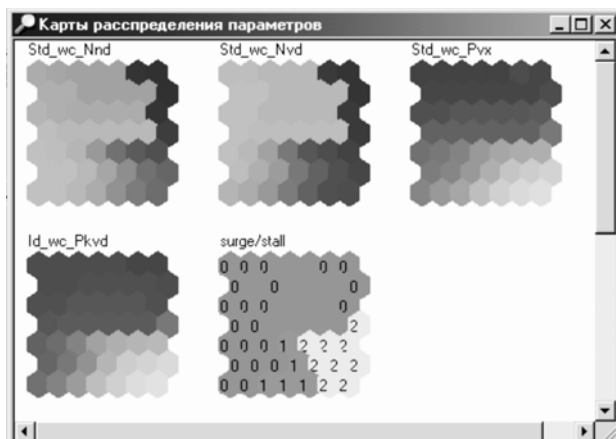


Рисунок 6 – Карты распределения параметров процесса ГДН

2 РЕЖИМ ДИАГНОСТИРОВАНИЯ

При диагностировании ошибочных состояний системы карта может быть использована в двух случаях: (1)

чтобы определить и (2) идентифицировать ошибку. ПО «Control & Diagnostics System» предоставляет возможность выбора соответствующего метода диагностирования (рис. 7): метод слежения за изменением параметра MSE и за траекторией перемещения рабочей точки по картам параметров процесса.

На рис. 8 приведен пример слежения за траекторией рабочей точки процесса по картам, соответствующим изменению параметра σ для ВК сигнала $P_{вх}$ (Std_wc_Pvx) и состояния процесса ГДН (surge/stall). При попадании рабочей точки в область вращающегося срыва можно говорить о начале помпажных явлений ГТД.

В случае, когда обучающая выборка не включает измерения, описывающие ошибочные ситуации, пространство состояний на карте соответствует нормальному функционированию системы. В дальнейшем произошедший сбой может быть определен, исходя из мониторинга погрешности аппроксимации СОК [2]. Большое значение данного параметра, превышающее определенный порог, укажет, что процесс вышел из нормального состояния.

На рис. 9 представлена кривая изменения погрешности аппроксимации MSE для данных по стендовым испытаниям ГТД. Карта обучалась только на тех примерах, для которых было характерно отсутствие состояния помпажа. Из рис. 9 следует, что увеличение исследуемой величины происходит на 205-м измерении и постепенно возрастает. Сопоставление данных результатов с реальными значениями показывает, что с 210-го образца в выборке присутствовал сигнал об обнаружении помпажа.

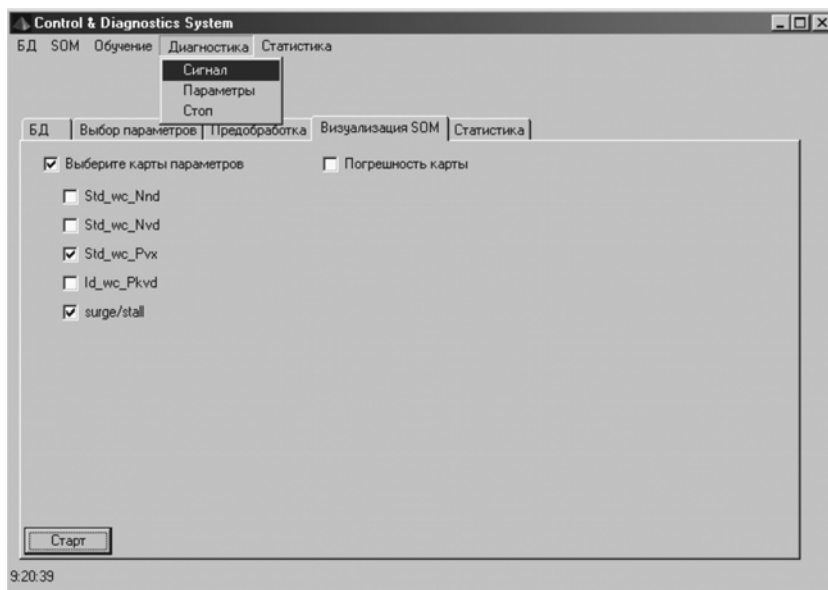


Рисунок 7 – Выбор метода диагностирования процесса на основе СОК

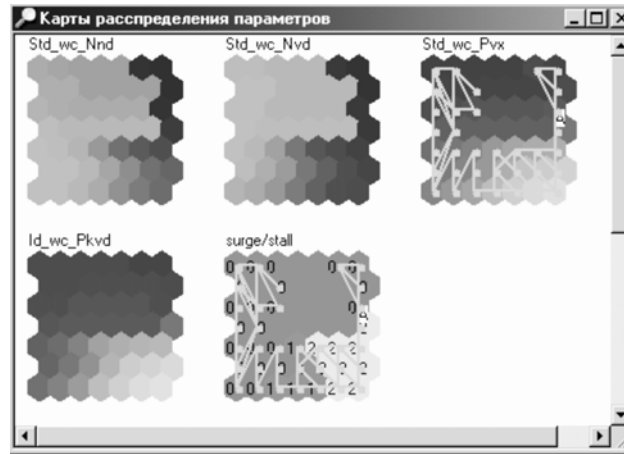


Рисунок 8 – Диагностирование помпажных явлений на основании слежения за траекторией рабочей точки процесса

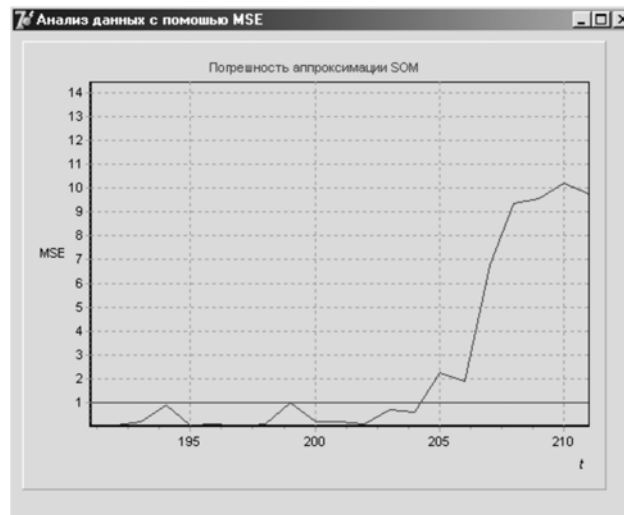


Рисунок 9 – Диагностика состояний процесса с помощью MSE

ВЫВОДЫ

Для диагностирования состояний процессов и систем представлена разработка ПО «Control & Diagnostics System», использующая методику самоорганизующихся карт Кохонена. Возможности данного программного продукта в режиме обучения СОК позволяют выбирать топологию карты, вводить графические настройки размера изображения, адекватно оценивать результаты обучения по погрешностям топографического упорядочивания нейронов ТРЕ и аппроксимации MSE. С помощью построения набора цветowych карт параметров исследуемого процесса значения нескольких показателей системы и их взаимодействие могут быть просмотрены одновременно. В режиме диагностирования ПО «Control & Diagnostics System», позволяет следить в режиме реального времени за траек-

торией перемещения рабочей точки по выбранным картам параметров процесса и определять возникновение внештатной ситуации с помощью среднеквадратической погрешности аппроксимации СОК. Разработанное ПО было использовано при исследовании процесса ГДН турбокомпрессора по данным стендовых испытаний ГТД.

ПЕРЕЧЕНЬ ССЫЛОК

1. Esa Alhoniemi, Jaako Hollmen, Olli Simula. Process Monitoring and Modeling using the Self-Organizing Map // Integrated Computer Aided Engineering. – 1999. – V 6, № 1. – P. 3–14.
2. O. Simula, Esa Alhoniemi, Jaako Hollmen. Monitoring and modeling of complex process using hierarchical self-organizing maps // Proceedings of the IEEE International on Circuits and Systems (ISCAS'96), volume Supplement. – 1996. – P. 73–76.

3. Гвидо Дебок, Тейво Кохонен. Анализ финансовых данных с помощью самоорганизующихся карт. Пер. с англ. – М.: Издательский дом «Альпина», 2001. – 316 с.
4. Teuvo Kohonen, Jussi Hynnien, Jari Kangas, Jorma Laaksonen. SOM_PAK the Self-organizing map program package. Version 3.1 – Finland: Helsinki University of Technology, 1997. – 27 p.
5. Киприч Т. В., Дубровин В. И. Методика контроля помпажных явлений ГТД на основе вейвлет-анализа и дискриминантных признаков // Вестник двигателестроения. – 2008. – № 1. – С. 168–169.

Надійшла 25.04.2008

Для аналізу станів процесів та систем розглядаються можливості програмних засобів, що реалізують методику власнеорганізованих карт Кохонена. Розроблене програмне забезпечення «Control & Diagnostics System» зорієнто-

вано на діагностування позаштатних та помилкових ситуацій, що виникають у ході роботи обладнання. На його основі проведено дослідження газодинамічної нестійкості турбокомпресору по даним стендових випробувань газотурбінного двигуна.

Possibilities of the software tools to analyze states of the processes and systems, implemented the methodic of self-organizing Kohonen's maps, are considered. Developed «Control & Diagnostics System» software is directed to diagnostics of contingencies and error events, that occur in the equipment work process. On the base of it the research of the turbo compressor gas-dynamic instability under data of the gas-turbine block test was carried out.

УДК 519.876.5

Б. Б. Нестеренко, М. А. Новотарський

МОДЕЛЮВАННЯ ДИСКРЕТНИХ КЛІТКОВИХ НЕЙРОННИХ МЕРЕЖ

В роботі дано обґрунтування перспективності застосування дискретних кліткових нейронних мереж для розв'язування крайових задач математичної фізики локально-асинхронними методами. Коротко описано основні елементи АПРО-мереж та запропонована дискретна кліткова мережа, яка представлена у вигляді дворівневої АПРО-мережі. Розглянуто базові принципи функціонування даної мережі, що визначають характер еволюції послідовних кроків. Описана методика формування реального робочого навантаження та основні підходи до навчання.

ВСТУП

Протягом останнього часу чітко окреслились області ефективного застосування штучних нейронних мереж. Значних успіхів досягнуто, зокрема, при реалізації алгоритмів апроксимації функцій та регресійного аналізу, включаючи передбачення часових послідовностей. Широкого застосування штучні нейронні мережі набули при вирішенні класифікаційних задач, а саме: при розпізнаванні образів та визначенні новизни сигналів. Разом з тим, велика кількість публікацій, присвячених проблемі реалізації обчислювальних алгоритмів на штучних нейронних мережах, дає впевненість у тому, що цей напрямок також має великі перспективи. Обчислювальні алгоритми можуть бути реалізовані на традиційних нейронних структурах. Наприклад, в [1, 2] для розв'язування крайових задач математичної фізики використовуються штучні нейронні мережі прямого поширення. Альтернативний підхід полягає у застосуванні кліткових нейронних мереж [3, 4] і є більш органічним при розв'язуванні

крайових задач, оскільки структура області обчислення корелюється зі структурою кліткової мережі. Але обидва підходи мають загальний недолік, пов'язаний з можливістю застосування лише прямих обчислювальних методів. Крім того, алгоритми навчання таких мереж характеризуються значною трудомісткістю і не завжди гарантують успішність виконання процедури навчання.

В даній роботі запропоновано використання дискретних кліткових мереж, які реалізують алгоритми чисельних методів розв'язування крайових задач математичної фізики, що значно розширює коло проблем, які можуть бути вирішені за допомогою нейронних мереж. Також запропоновано підходи до навчання, які базуються не тільки на зміні вагових коефіцієнтів нейронів, а й на властивості пластичності, що виражена у модифікації структурних зв'язків між нейронами.

1 СТРУКТУРА МОДЕЛІ ДИСКРЕТНОЇ КЛІТКОВОЇ МЕРЕЖІ

Будемо розглядати дискретну кліткову мережу у вигляді імітаційної моделі складної дискретної системи, що дає можливість реалізувати її на сучасних обчислювальних системах. Застосувавши АПРО-мережі [5] як інструмент формального опису, представимо дискретну кліткову мережу коротцем:

$$\Phi = (P, T, F, M, V), \quad (1)$$

де $P = \{p_i\}_{i=1}^n$ – скінченна множина позицій,

$T = \{t_j\}_{j=1}^q$ – скінченна множина переходів,
 $F = (P \times T) \cup (T \times P)$ – множина ребер між переходами та позиціями,

$M = \left\{ \left(p_k, \{\mu_l\}_{l=1}^{\text{Max}_k p_k} \right) \right\}_{k=1}^n$ – скінченна множина маркувань,
 $V = (\Delta, \Psi, \Lambda)$ – множина глобальних змінних.

Подібно до графічних позначень, що застосовуються в мережах Петрі, позиція APRO-мереж позначається колом, простий перехід – лінією, ребро – лінією зі стрілкою, а мітка – крапкою. Додатково введено позначення, які пов'язані з елементами операторного переходу. Сам операторний перехід будемо зображати у вигляді прямокутника, а входи і виходи операторного переходу представимо у вигляді лівого і правого півкіл. Найпростішу APRO-мережа, що включає всі згадані елементи, наведено на рис. 1.

Позиції APRO-мережі: $p_i = \{d_i, q_i\}$, де d_i – множина параметрів позиції, q_i – множина міток, розміщених на даній позиції. Переходи APRO-мережі включають два класи переходів: $t = \{\tau, Ne\}$, де τ – клас простих переходів, Ne – клас операторних переходів. Простий перехід описує множина елементів $\tau_j = \{\chi_j, N_j\}$, де χ_j – множина параметрів переходу, $N_j = \{\rho_j, \pi_j, \gamma_j, \omega_j, A_j, O_j\}$ – функціональне ядро переходу, що складається з таких елементів: ρ_j – процедура активації переходу, π_j – процедура обслуговування переходу, γ_j – процедура деактивації переходу, ω_j – процедура очікування, A_j – частково впорядкована послідовність активностей, O_j – частково впорядкована послідовність вихідних міток.

Операторний перехід задає APRO-мережу нижнього рівня і може бути описаний коротцем:

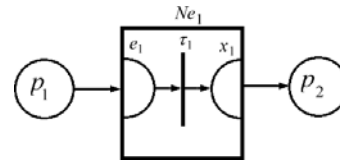


Рисунок 1 – Найпростіша APRO-мережа

$$Ne = (P, T, E, X, F), \tag{2}$$

де $E = \{e_c\}_{c=1}^h$ – скінченна множина входів операторного переходу,

$X = \{x_d\}_{d=1}^r$ – скінченна множина виходів операторного переходу.

Ребра мережі задають матрицею інцидентності I з елементами:

$$I(p_i, t_j) = \begin{cases} -1, & (p_i, t_j) \in \mathbf{F}, \\ +1, & (p_i, t_j) \in \mathbf{F}^{-1}, \\ 0, & (p_i, t_j) \notin \mathbf{F}, (p_i, t_j) \notin \mathbf{F}^{-1}, \end{cases} \tag{3}$$

$1 \leq i \leq n, \quad 1 \leq j \leq q.$

Мітки APRO-мережі: $\mu_k = \{\lambda_k, \alpha_k\}$, де λ_k – множина параметрів мітки, α_k – множина атрибутів мітки.

Множина глобальних змінних: $V = (\Delta, \Psi, \Lambda)$, де Δ – підмножина показників продуктивності, Ψ – підмножина показників реактивності, Λ – підмножина показників використання.

Модель дискретної кліткової мережі, кожний нейрон якої представлений у вигляді операторного переходу APRO-мережі, показана на рис. 2. Вона має дво-

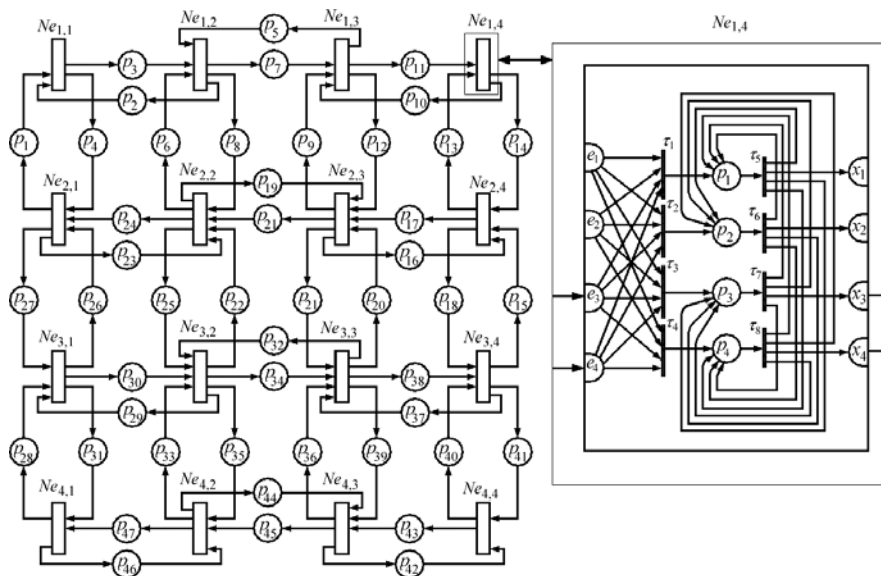


Рисунок 2 – Модель двовимірної дискретної кліткової мережі

рівневу структуру, представлену окремими АПРО-мережами. Верхній рівень відображає двовимірне однорідне середовище, у якому обчислювальні функції зосереджені в операторних переходах $T = \{Ne_{1,1}, \dots, Ne_{4,4}\}$, а комунікаційні функції реалізують позиції $P = \{p_1, \dots, p_{47}\}$. Матриця інцидентності має вигляд:

$$I = \left\{ \begin{array}{cccccccc} & p_1 & p_2 & p_3 & p_4 & p_5 & \dots & p_{47} \\ Ne_{1,1} & -1 & -1 & +1 & +1 & 0 & \dots & 0 \\ Ne_{1,2} & 0 & 1 & -1 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Ne_{4,4} & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{array} \right\}.$$

Структура моделі нижнього рівня описує роботу кліткового нейрона, у якій зв'язок з моделлю верхнього рівня відбувається за допомогою входів $E(Ne_{i,j}) = \{e_1, \dots, e_4\}$ та виходів $X(Ne_{i,j}) = \{x_1, \dots, x_4\}$ довільного операторного переходу $Ne_{i,j}$. Група переходів $\{\tau_1, \dots, \tau_4\} \subset T(Ne_{i,j})$ забезпечує виконання арифметичних операцій для урахування впливу вагових коефіцієнтів, пов'язаних з міжнейронними зв'язками. Переходи $\{\tau_5, \dots, \tau_8\} \subset T(Ne_{i,j})$ виконують обчислення активаційної функції. Матриця інцидентності АПРО-мережі довільного кліткового нейрона $Ne_{i,j}$:

$$I(Ne_{i,j}) = \left\{ \begin{array}{cccccccccccc} e_1 & e_2 & e_3 & e_4 & p_1 & p_2 & p_3 & p_4 & x_1 & x_2 & x_3 & x_4 \\ \tau_1 & -1 & -1 & -1 & -1 & +1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \tau_2 & -1 & -1 & -1 & -1 & 0 & +1 & 0 & 0 & 0 & 0 & 0 \\ \tau_3 & -1 & -1 & -1 & -1 & 0 & 0 & +1 & 0 & 0 & 0 & 0 \\ \tau_4 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & +1 & 0 & 0 & 0 \\ \tau_5 & 0 & 0 & 0 & 0 & -1 & +1 & +1 & +1 & +1 & 0 & 0 \\ \tau_6 & 0 & 0 & 0 & 0 & +1 & -1 & +1 & +1 & 0 & +1 & 0 \\ \tau_7 & 0 & 0 & 0 & 0 & +1 & +1 & -1 & +1 & 0 & 0 & +1 \\ \tau_8 & 0 & 0 & 0 & 0 & +1 & +1 & +1 & -1 & 0 & 0 & +1 \end{array} \right\}.$$

Структура розглянутої дискретної кліткової мережі складається з матриці кліткових нейронів розміром 4×4 , але може бути розширена до довільних розмірів без зміни принципів її опису.

2 ФУНКЦІОНУВАННЯ МОДЕЛІ ДИСКРЕТНОЇ КЛІТКОВОЇ МЕРЕЖІ

Завдяки однорідності структури дискретної кліткової мережі функціонування верхнього рівня моделі базується на визначенні правил взаємодії між довільним операторним переходом $Ne_{i,j}$ і множиною сусідніх з ним операторних переходів $\{Ne_{i,j-1}, Ne_{i,j+1}, Ne_{i-1,j}, Ne_{i+1,j}\}$, що утворюють хрестоподібний обчислювальний шаблон. Взаємодія між сусідніми переходами шаблону має асинхронний характер. Тому повний цикл

обміну інформацією для $Ne_{i,j}$ утворює множину кроків $\{Ne_{i,j-1}, Ne_{i,j+1}, Ne_{i-1,j}, Ne_{i+1,j}\}_{i,j}$ на заданому обчислювальному шаблоні.

Мережа здійснює довільний крок U тільки за умови спрацювання переходів, які пов'язані з даним кроком. Зміну маркування в результаті спрацювання кроку U часто позначають як $M|U\rangle M'$. Враховуючи, що для випадку дискретної кліткової мережі з кожним кроком пов'язане спрацювання відповідного операторного переходу, еволюція маркування $\Delta_{i,j}$ під дією послідовності кроків може бути представлена варіантами, кількість яких дорівнює кількості перестановок $P(4,4) = 4!$:

$$\left\{ \begin{array}{l} M_{i,j}|\Delta_{i,j}\rangle M'_{i,j} = \\ = M_{i,j}|U_{i,j-1}\rangle M_1|U_{i,j+1}\rangle M_2|U_{i-1,j}\rangle M_3|U_{i+1,j}\rangle M'_{i,j}, \\ \dots \\ M_{i,j}|\Delta_{i,j}\rangle M'_{i,j} = \\ = M_{i,j}|U_{i+1,j}\rangle M_3|U_{i-1,j}\rangle M_2|U_{i,j+1}\rangle M_1|U_{i,j-1}\rangle M'_{i,j}. \end{array} \right. \quad (4)$$

Загальна еволюція маркування Δ дискретної кліткової мережі (рис. 2) розмірністю 4×4 визначається кількістю перестановок $P(4^2, 4^2) = 16!$:

$$\left\{ \begin{array}{l} M|\Delta\rangle M' = M|\Delta_{1,1}\rangle M_{1,1} \dots M_{4,3}|\Delta_{4,4}\rangle M', \\ \dots \\ M|\Delta\rangle M' = M|\Delta_{4,4}\rangle M_{4,4} \dots M_{1,2}|\Delta_{1,1}\rangle M'. \end{array} \right. \quad (5)$$

Функціонування нижнього рівня моделі відповідає еволюції АПРО-мережі кліткового нейрона, представленого складним кроком $U_{i,j} = u_1 \circ u_2 \circ u_3 \circ u_4$, який відображає суперпозицію груп переходів:

$$\begin{aligned} u_1 &= \{(e_1, \beta_{e_1}), \dots, (e_4, \beta_{e_4})\}, \\ u_2 &= \{(\tau_1, \beta_{\tau_1}), \dots, (\tau_4, \beta_{\tau_4})\}, \\ u_3 &= \{(\tau_5, \beta_{\tau_5}), \dots, (\tau_8, \beta_{\tau_8})\}, \\ u_4 &= \{(x_1, \beta_{x_1}), \dots, (x_4, \beta_{x_4})\}. \end{aligned}$$

Параметри β вказують на кількість входжень даного переходу при виконанні відповідного кроку.

Еволюція маркування $\delta_{i,j}$ мережі нижнього рівня:

$$m_{i,j}|\delta_{i,j}\rangle m'_{i,j} = m_{i,j}|u_1\rangle m_1|u_2\rangle m_2|u_3\rangle m_3|u_4\rangle m'_{i,j}. \quad (6)$$

Отже, поєднання еволюції маркування на рівні кліткового нейрона (6) з еволюцією маркування на обчислювальному шаблоні (4) та загальною еволюцією маркування (5) дозволяє охопити всі можливі стани моделі дискретної кліткової мережі в ході її функціонування.

3 РОБОЧЕ НАВАНТАЖЕННЯ МОДЕЛІ

Описана модель дискретної кліткової мережі безумовно відноситься до складних систем, дослідження яких у аналітичному вигляді викликає значні труднощі. Тому для одержання коректних результатів доцільно використовувати імітаційне моделювання, яке у даному випадку має ряд особливостей. Головна з цих особливостей полягає у тому, що імітаційна модель кліткової мережі використовує реальне робоче навантаження у вигляді алгоритмів, побудованих за спеціально розробленими для таких мереж чисельними локально-асинхронними методами [6] розв'язування систем алгебраїчних рівнянь:

$$Ov = \Phi, \quad (7)$$

де O – різницевий оператор, що задає відображення $O:R^n \rightarrow R^n$; v – шукана різницева функція; Φ – функція правої частини.

Ітераційний процес у таких мережах починається з надходження на вхід кожного нейрона мережі початкового вектора даних $(v_{i,j-1}(0), v_{i,j+1}(0), v_{i-1,j}(0), v_{i+1,j}(0))$. Кожен нейрон реалізує свою частину $O_{i,j}$ загального оператора O , використовуючи свої дані та ті, які були одержані від інших нейронів. Асинхронною ітераційною послідовністю, що відповідає оператору O , будемо називати послідовність $\{v(\alpha)\}_{\alpha=1}^{\infty}$ векторів $v(\alpha) \in R^n$, що визначається рекурсивно за ітераційною схемою:

$$v_{i,j}(\alpha) = \begin{cases} v_{i,j}(\alpha-1), & (i,j) \notin A_{\alpha}, \\ O_{i,j}[v_{i-1,j}(s_{i-1,j}(\alpha-1)), \dots, v_{i,j+1}(s_{i,j+1}(\alpha-1))], & (i,j) \in A_{\alpha}. \end{cases} \quad (8)$$

Сукупність непустих підмножин A_{α} множини $A = \{(1, 1), (1, 2), \dots, (i, j), \dots, (4, 4)\}$ утворює хаотичну послідовність $\Omega = \{A_{\alpha}\}_{\alpha=1}^{\infty}$, а затримки $s_{i,j}(\alpha)$ об'єднані в послідовність затримок $S = \{s_{i,j}(\alpha)\}_{\alpha=1}^{\infty}$, $(i,j) \in A_{\alpha}$, що відповідає умовам: $0 \leq s_{i,j}(\alpha) \leq \alpha$ при $\alpha = 1, 2, \dots$; $\lim_{\alpha \rightarrow \infty} (s_{i,j}(\alpha)) = \infty$. Затримки $s_{i,j}(\alpha)$ дають можливість використання довільних компонентів векторів попередніх ітерацій при обчисленні вектора поточної ітерації. Однак для досягнення прогресу в обчисленнях усі компоненти вектора ітерації повинні періодично оновлюватися, що накладає певні обмеження на хаотичну послідовність Ω . Найкращого результату збіжності методу можна досягти, якщо на кожному черговому ітераційному кроці у вузлах сіткової області будуть використовуватись тільки оновлені компоненти ітераційного вектора. Однак використання хаотичної підмножини A_{α} для формування чергового вектора

ітерації дає можливість організації обчислень із довільними алгоритмами оновлення попередніх компонентів. Тому немає необхідності очікувати оновлення всіх відповідних компонентів вектора ітерації перед продовженням ітераційного процесу в даному клітковому нейроні мережі. Такий принцип організації обчислень є базовою властивістю асинхронності розглянутого методу. У сукупності із властивістю локальності, що спричиняє використання тільки деякої підмножини компонент вектора ітерації для успішного продовження ітераційного процесу в кожному вузлі мережі, одержуємо метод, що дає можливість застосування кліткових нейронних мереж для розв'язування крайових задач математичної фізики з високим ступенем дискретизації.

4 НАВЧАННЯ ДИСКРЕТНОЇ КЛІТКОВОЇ МЕРЕЖІ

Здатність до навчання є фундаментальною властивістю штучних нейронних мереж. Процес навчання можна розглядати як модифікацію міжнейронних зв'язків та налаштування вагових коефіцієнтів для ефективного виконання поставленої задачі. Існує велика кількість правил та процедур навчання, які залежать від типу штучних нейронних мереж.

Для даного типу дискретних кліткових мереж суть навчання полягає у задаванні міжнейронних зв'язків, які виникають в ході розв'язування крайових задач математичної фізики, а також вагових коефіцієнтів, які підвищують збіжність ітераційних процесів у нейронах. Використання різних міжнейронних зв'язків зумовлено застосуванням мультисіткових методів [6], які реалізують обчислення на ієрархічній послідовності сіток $G^{(0)} \supset G^{(1)} \supset \dots \supset G^{(z)}$, для кожної з яких нейрону $Ne_{i,j}$ відповідає своя множина сусідніх нейронів. Особливість розглянутого підходу до зміни міжнейронних зв'язків полягає у тому, що така зміна відбувається без модифікації структури нейронної мережі. Механізм транзитних пересилок забезпечує зв'язки між нейронами, які утворюють сітки $G^{(z)}$, $z > 0$. Завдяки цьому механізму виникає можливість організації обчислень на віртуальних сітках, структура яких не співпадає з фізичною структурою кліткової мережі. Суть мультисіткових методів полягає у підвищенні ефективності розв'язування крайових задач математичної фізики локально-асинхронним методом за рахунок згладжування низькочастотної складової нев'язки на грубих сітках. Ефективність такого згладжування залежить від послідовності застосування W -циклів та V -циклів і є індивідуальною для кожної конкретної крайової задачі. Отже, вибір конкретного алгоритму мультисіткового методу можна розглядати як навчання кліткової нейронної мережі.

Тепер розглянемо методи навчання дискретних кліткових нейронних мереж, що забезпечують покращення показників розв'язування крайових задач шляхом модифікації вагових коефіцієнтів нейронних зв'язків. Відомо, що швидкість збіжності довільного ітераційного методу залежить від спектральних властивостей різницевого оператора. Отже, існує можливість модифікації системи різницевих рівнянь (7) до еквівалентної системи, яка має той же розв'язок при кращих спектральних властивостях різницевого оператора O . Нехай оператор O заданий у вигляді матриці коефіцієнтів системи різницевих рівнянь:

$$Or(g) = \Phi(g), \quad g \in G. \quad (9)$$

Тоді припустимо існування деякої матриці W такої, що модифікована система різницевих рівнянь

$$W^{-1}Or(g) = W^{-1}\Phi(g) \quad (10)$$

має той же розв'язок, що і система (9), але спектральні властивості матриці коефіцієнтів $W^{-1}O$ кращі, ніж матриці коефіцієнтів O . Існує велика кількість підходів до визначення матриці W в залежності від вибору методу прискорення. Як один з прикладів застосування методів прискорення до навчання кліткових нейронних мереж може бути застосоване поліноміальне прискорення [7].

ВИСНОВОК

Інтерес до розширення сфери застосування штучних нейронних мереж став причиною збільшення кількості успішних реалізацій обчислювальних алгоритмів на даних структурах. Практика показала, що при розв'язуванні крайових задач математичної фізики найкраще проявили себе кліткові нейронні мережі, оскільки у даному випадку область обчислень крайової задачі може бути співвіднесена зі структурою мережі. Загальним недоліком застосування відомих штучних нейронних мереж є алгоритмічні обмеження, що зводяться до можливості застосування тільки прямих методів розв'язування крайових задач. В роботі запропоновано підхід до побудови моделі дискретної кліткової нейронної мережі, яка дозволяє застосування широкого спектру чисельних методів, що значно розширює коло доступних до розв'язування задач та зменшує похибку при одержанні результатів. Представлення моделі у формалізованому вигляді за допомогою АПРО-мереж дає можливість реалізації її в обчислювальних середовищах різного типу, включаючи паралельні розподілені обчислювальні системи. Специфіка представлення даних в АПРО-мережах забезпечує можливість

використання на моделі реального робочого навантаження. Застосування спеціальних локально-асинхронних методів, в основі яких лежить забезпечення незалежних обчислювальних процесів на кожному з вузлових нейронів кліткової мережі, запропоновано для повного використання властивостей кліткових нейронних мереж. Дискретний характер обчислень в клітковій мережі вимагає перегляду підходів до навчання таких структур. У роботі запропоновано два можливих алгоритми навчання. Один з них базується на використанні властивості пластичності і дозволяє модифікувати характер обчислень таким чином, щоб використовувати транзитні пересилки даних від віддалених нейронів. Другий підхід використовує традиційну модифікацію вагових коефіцієнтів для міжнейронних зв'язків з метою підвищення швидкості збіжності ітераційних процесів у кожному з нейронів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Lagaris I. E., Likas A., Fotiadis D. I. Artificial Neural Networks for Solving Ordinary and Partial Differential Equations // IEEE Trans. on Neural Networks. – 1998. – Vol. 9, No. 5. – P. 987–1000.
2. Lagaris I. E., Likas A., Papageorgiou D. G. Neural Networks Methods for Boundary Value Problems with Irregular Boundaries // IEEE Trans. on Neural Networks. – 2000. – Vol. 11, No. 5. – P. 1041–1049.
3. Chua L. O., Yang L. Cellular Neural Networks: Theory // IEEE Trans. on Circuits and Systems. – 1988. – Vol. 35, No. 10. – P. 1257–1272.
4. Chua L. O., Yang L. Cellular Neural Networks: Applications // IEEE Trans. on Circuits and Systems. – 1988. – Vol. 35, No. 10. – P. 1273–1290.
5. Новотарський М. А. Мережі для моделювання складних систем // Радіоелектроніка. Інформатика. Управління. – 2006. – № 2. – С. 60–66.
6. Новотарський М. А., Нестеренко Б. Б. Штучні нейронні мережі: обчислення. – Київ: Ін-т математики, 2004. – 408 с.
7. Хейгеман Л., Янг Д. Прикладные итерационные методы. – М.: Мир, 1986. – 448 с.

Надійшла 9.04.2008

В работе дано обоснование перспективности применения дискретных клеточных нейронных сетей для решения краевых задач математической физики локально-асинхронными методами. Коротко описаны основные элементы АПРО-сетей и предложена дискретная клеточная сеть, представленная в виде двухуровневой АПРО-сети. Рассмотрены базовые принципы функционирования данной сети, определяющие характер эволюции последовательных шагов. Описана методика формирования реальной рабочей нагрузки и основные подходы к обучению.

In the paper the substantiation of availability of application discrete cellular neural networks for solving of boundary value problems of mathematical physics by local – asynchronous methods is given. Basic elements of АПРО-nets are shortly described and the discrete cellular network submitted as a two-level АПРО-net is offered. Base principles of functioning of the given network which determine evolution of consecutive steps are considered. The technique of formation of real working loading and the basic approaches to training is described.

И. В. Новосельцев, Н. Г. Аксак

ПРИМЕНЕНИЕ МНОГОПРОЦЕССОРНЫХ СИСТЕМ ДЛЯ РЕАЛИЗАЦИИ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА

В данной работе предложена модель реализации МП на многопроцессорной системе при ограниченном количестве процессоров. В качестве топологии вычислительной системы предложена организация процессоров в виде звезды. Данный подход позволяет значительно снизить время обучения НС, и дает возможность получения максимального ускорения, ограниченного числом имеющихся процессоров.

ВВЕДЕНИЕ

Сегодня по-прежнему остается актуальной задача распознавания изображения. Применение нейронных сетей (НС) является перспективным для решения задач в этой области [1]. Однако сложность подбора оптимальной архитектуры сети для решения конкретной задачи, а также большие вычислительные затраты, требуемые для обучения нейросетей, сдерживают их массовое применение.

Время обучения нейронных сетей, особенно при использовании стандартных персональных компьютеров и больших объемах исходных данных, может быть очень велико. Проблему высокой вычислительной сложности обучения нейросетей чаще всего решают, используя более мощный компьютер или специализированный аппаратный нейроускоритель. Однако такой путь доступен далеко не всем.

Снизить вычислительную сложность нейросетевых алгоритмов можно путем распараллеливания однотипных операций, которые выполняются в процессе функционирования нейронной сети. Подобные подходы позволяют обучать нейронные сети с меньшими затратами времени и дают возможность получения максимального ускорения, ограничиваемого числом имеющихся процессоров.

В данной работе рассмотрен метод распараллеливания при ограниченном количестве процессоров [2]. В качестве топологии вычислительной системы предложена организация процессоров в виде звезды. Управляющий процессор в подобной топологии может использоваться для загрузки вычислительных процессоров исходными данными и для приема результатов выполненных вычислений.

1 АРХИТЕКТУРА НЕЙРОННОЙ СЕТИ

Для решения задачи распознавания образов применяется многослойный персептрон (МП) с архитектурой $N_i - N_j - N_k$, где N_i – количество нейронов во входном слое, N_j – количество нейронов в скрытом слое, N_k – количество нейронов в выходном слое (рис. 1) [3–5].

Обучающий пример представляется парой $\{X = [x_1, x_2, \dots, x_i], D = [d_1, d_2, \dots, d_k]\}$, где X – входной массив, D – вектор цели.

Функционирование нейрона определяется следующим соотношением:

$$y_j^{(1)}(t) = \Phi \left(\sum_{l=1}^i w_{lj}^{(1)}(t)x_l(t) + b_j^{(1)} \right) \quad \text{для скрытого слоя,} \quad (1)$$

$$y_k^{(2)}(t) = \Phi \left(\sum_{l=1}^j w_{lk}^{(2)}(t)y_l^{(1)}(t) + b_k^{(2)} \right) \quad \text{для выходного слоя,} \quad (2)$$

где $w_{ij}^{(1)}$ – синаптический вес связи i -го входа с j -м нейроном, $b_j^{(1)}$ – порог j -го нейрона скрытого слоя, $w_{jk}^{(2)}$ – синаптический вес связи j -го выхода нейрона скрытого слоя с k -м нейроном выходного слоя, $b_k^{(2)}$ – порог k -го нейрона выходного слоя, $\Phi(\bullet)$ – функция активации нейрона.

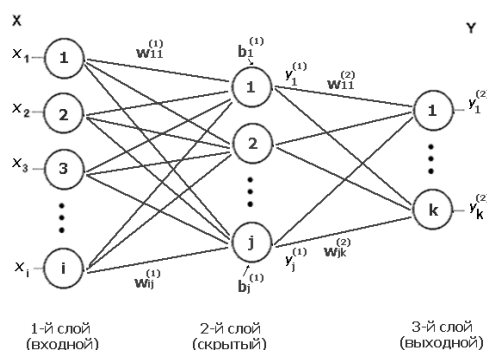


Рисунок 1 – Архитектура многослойного персептрона

В качестве функции активации для всех слоев выбрана сигмоидальная функция

$$\varphi(y^{(L)}(t)) = \frac{1}{1 + \exp(-ay^{(L)}(t))}, \quad (3)$$

где a – параметр наклона сигмоидальной функции ($a > 0$).

Обучение сводится к минимизации среднеквадратической ошибки. Подстройка весов $w_{ij}^{(1)}$ и $w_{jk}^{(2)}$ проводится после подачи каждого примера (последовательный режим обучения). Настройка весов слоя L осуществляется в соответствии

$$w_{ij}^{(L)}(t+1) = \alpha w_{ij}^{(L)}(t) + \eta \delta_j^{(L)}(t) y_j^{(L)}(t), \quad (4)$$

где η – параметр, отвечающий за скорость обучения, α – постоянная момента ($\alpha = 1$), $\delta_j^{(L)}$ – локальный градиент j -го нейрона слоя L . Для выходного слоя локальный градиент вычисляется следующим образом:

$$\delta_k^{(2)}(t) = e_k(t), \quad (5)$$

где $e_k(t)$ – ошибка k -го нейрона выходного слоя.

А для скрытого слоя:

$$\delta_j^{(1)}(t) = \sum_{l=1}^k \delta_l^{(2)}(t) w_{jl}^{(2)}(t) y_j^{(1)}(t). \quad (6)$$

На однопроцессорной системе данные вычисления выполняются последовательно, т. е. при формировании выходного вектора скрытого слоя $Y^{(1)}$ последовательно вычисляются $y_1^{(1)}, y_2^{(1)}, \dots, y_j^{(1)}$, затем аналогично вычисляются значения выходного слоя.

2 РЕАЛИЗАЦИЯ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА НА МНОГОПРОЦЕССОРНОЙ СИСТЕМЕ

Реализация МП (рис. 1) на многопроцессорной системе может быть представлена схемой, приведенной на рис. 2.

Соотношение (1) можно представить в матричном виде как умножение матрицы весовых коэффициентов $W^{(1)}$ скрытого слоя на вектор входа X :

$$Y^{(1)} = \varphi(W^{(1)T}X + B^{(1)}) = \varphi \left(\begin{bmatrix} W_1^{(1)} \\ W_2^{(1)} \\ \dots \\ W_j^{(1)} \end{bmatrix} \times X + B^{(1)} \right) =$$

$$= \varphi \left(\begin{bmatrix} w_{11}^{(1)} & w_{21}^{(1)} & \dots & w_{i1}^{(1)} \\ w_{12}^{(1)} & w_{22}^{(1)} & \dots & w_{i2}^{(1)} \\ \dots & \dots & \dots & \dots \\ w_{1j}^{(1)} & w_{2j}^{(1)} & \dots & w_{ij}^{(1)} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_i \end{bmatrix} + \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \dots \\ b_j^{(1)} \end{bmatrix} \right). \quad (7)$$

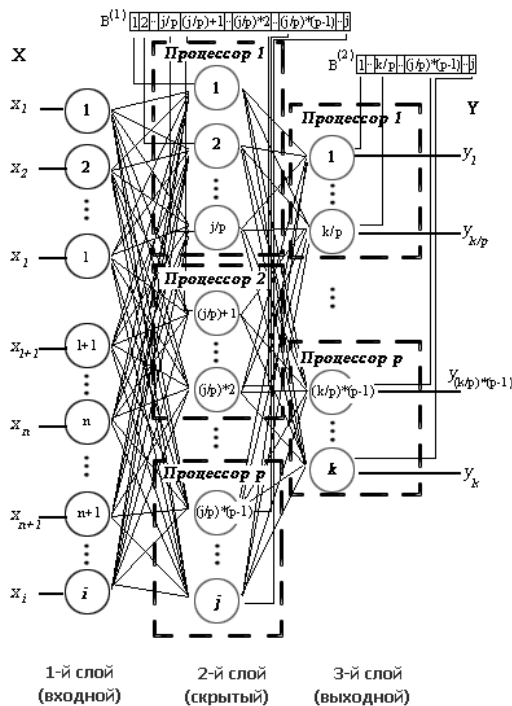


Рисунок 2 – Схема реализации многослойного персептрона

Соотношение (2) можно представить в виде умножения матрицы весовых коэффициентов $W^{(2)}$ выходного слоя на вектор $Y^{(1)}$ (выход скрытого слоя), полученный в результате операции (1):

$$Y^{(2)} = \varphi(W^{(2)T} Y^{(1)} + B^{(2)}) = \varphi \left(\begin{bmatrix} W_1^{(2)} \\ W_2^{(2)} \\ \dots \\ W_j^{(2)} \end{bmatrix} \times Y^{(1)} + B^{(2)} \right) =$$

$$= \varphi \left(\begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} & \dots & w_{j1}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} & \dots & w_{j2}^{(2)} \\ \dots & \dots & \dots & \dots \\ w_{1k}^{(2)} & w_{2k}^{(2)} & \dots & w_{jk}^{(2)} \end{bmatrix} \times \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \dots \\ y_k^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ \dots \\ b_k^{(2)} \end{bmatrix} \right). \quad (8)$$

Соотношение (6) – умножение матрицы весовых коэффициентов $W^{(2)}$ скрытого слоя на вектор локальных градиентов выходного слоя:

$$\delta_j^{(1)} = \begin{bmatrix} \delta_1^{(2)} w_{11}^{(2)} \delta_2^{(2)} w_{12}^{(2)} \dots \delta_k^{(2)} w_{1k}^{(2)} \\ \delta_1^{(2)} w_{21}^{(2)} \delta_2^{(2)} w_{22}^{(2)} \dots \delta_k^{(2)} w_{2k}^{(2)} \\ \dots \\ \delta_1^{(2)} w_{j1}^{(2)} \delta_2^{(2)} w_{j2}^{(2)} \dots \delta_k^{(2)} w_{jk}^{(2)} \end{bmatrix} \times \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \dots \\ y_j^{(1)} \end{bmatrix}. \quad (9)$$

Тем самым, получение результирующего вектора $Y^{(2)}$ на каждом шаге обучения предполагает повторение $n = (2j + k)$ однотипных операций по умножению строк матриц на векторы. Выполнение каждой такой операции включает поэлементное умножение элементов строки матрицы на векторы и последующее суммирование полученных произведений. Общее количество необходимых скалярных операций необходимых для обучения НС оценивается величиной:

$$T_1 = 2n^2 M, \quad (10)$$

где M – количество эпох, n – количество однотипных операций.

Если, например, размерность изображения 100×100 , обучающая выборка состоит из 1000 примеров и количество эпох обучения равно 500, тогда $T_1 = 2(2 \times 100 \times 100 + 4)^2 \times 500 \times 1000 = 400160016000000$.

При использовании количества процессоров $p \leq n$ параллельная вычислительная схема умножения матрицы на вектор может быть конкретизирована следующим образом:

– на каждый из имеющихся управляющих процессоров пересылается вектор X и $s = n/p$ строк матрицы;

– выполнение операции умножения строк матрицы на вектор выполняется при помощи последовательного алгоритма.

3 ВЫБОР ТОПОЛОГИИ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ

В соответствии с характером выполняемых межпроцессорных взаимодействий, для предложенной вычислительной схемы в качестве топологии выбрана организация процессоров в виде звезды. На рис. 3 представлен фрагмент распараллеливания соотношения (7).

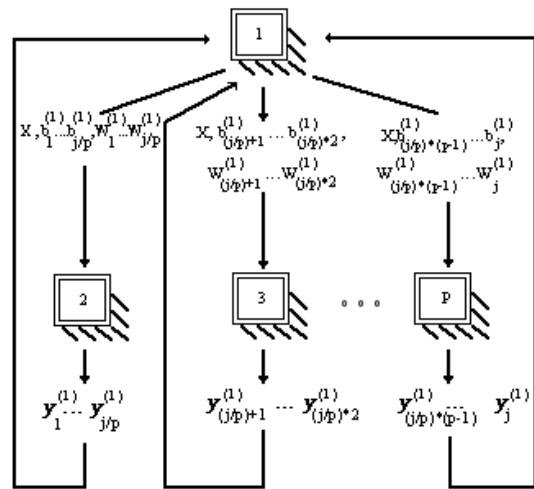


Рисунок 3 – Топология вычислительной системы

Управляющий процессор формирует данные для обработки вычислительными процессорами (рис. 3):

- посылает входной вектор X всем процессорам;
- распределяет матрицу весовых коэффициентов $W^{(1)}$ и вектор смещений $B^{(1)}$ между процессорами, а именно: строки $W_1^{(1)}, W_2^{(1)}, \dots, W_{j/p}^{(1)}, b_1^{(1)}, b_2^{(1)}, \dots, b_{j/p}^{(1)}$ – первому процессору, строки $W_{(j/p)+1}^{(1)}, \dots, W_{(j/p)+2}^{(1)}, b_{(j/p)+1}^{(1)}, \dots, b_{(j/p)+2}^{(1)}$ – второму процессору, ..., строки $W_{(j/p)*(p-1)}^{(1)}, \dots, W_j^{(1)}, b_{(j/p)*(p-1)}^{(1)}, \dots, b_j^{(1)}$ – p -му процессору.

После выполнения операций, управляющий процессор осуществляет прием результатов выполнения вычислений. Таким образом формируется выход первого слоя НС – $Y^{(1)}$. Аналогично производится вычисление соотношений (8) и (9).

ЗАКЛЮЧЕНИЕ

С целью снижения времени обучения НС в работе решена актуальная задача распределения НС по процессорам.

Научная новизна работы заключается в том, что впервые предложена новая модель реализации МП на многопроцессорной системе, которая дает возможность получения максимального ускорения, ограниченного числом имеющихся процессоров, за счет распределения однотипных операций умножения, осуществляющихся в процессе функционирования НС.

Практическая ценность работы состоит в том, что проведенные эксперименты, реализующие предложенную модель, показали высокую эффективность. Данная модель реализована с помощью технологии параллельного программирования MPI.

ПЕРЕЧЕНЬ ССЫЛОК

1. Richard O. Duda, Peter E. Hart, David G. Stork. Pattern classification: Wiley-Interscience; 2nd edition, 2000. – 680 p.

2. Гергель В. П., Стронгин Р. Г. Основы параллельных вычислений для многопроцессорных вычислительных систем: Учебное пособие Нижегородский госуниверситета. – Нижний Новгород, 2003. – 184 с.
3. Саймон Хайкин. Нейронные сети: полный курс, 2-е издание. М.: Издательский дом «Вильямс», 2006. – 1104 с.
4. Миркес Е. М. Учебное пособие по курсу НЕЙРОИНФОРМАТИКА. – Красноярск, 2002. – 347 с.
5. Ben Krose, Patrick van der Smagt. An Introduction to Neural Networks: University of Amsterdam, 1996. – 154 p.

Надійшла 2.03.2007

In this work the MP model realization on multiprocessing system is offered. The processors organization as a star topology is offered. The given approach allows considerably to lower time of training neural network and enables receptions of the maximal acceleration limited to number of available processors.

В даній роботі запропонована модель реалізації багатопроцесорного перцептронів на багатопроцесорній системі при обмеженій кількості процесорів. В якості топології обчислювальної системи запропонована організація процесорів у вигляді зірки. Даний підхід дозволяє значно знизити час навчання нейронної мережі та дає можливість одержання максимального прискорення, що обмежений числом наявних процесорів.

УДК 519.7:004.93

С. О. Субботін, А. О. Олійник

СТРУКТУРНИЙ СИНТЕЗ НЕЙРОМОДЕЛЕЙ НА ОСНОВІ ПОЛІМОДАЛЬНОГО ЕВОЛЮЦІЙНОГО ПОШУКУ

Досліджено проблему структурного синтезу нейромережних моделей. Розроблено метод полімодального еволюційного пошуку з кластеризацією хромосом, що дозволяє одержати різні структури нейромереж, підвищуючи ймовірність знаходження моделі, адекватної розв'язуваній задачі. Проведено експерименти по синтезі моделей залежності стану здоров'я населення від забруднення навколишнього середовища.

ВСТУП

Етап вибору оптимальної архітектури нейромоделі є одним з найважливіших завдань при синтезі нейромереж [1], оскільки на цьому етапі формується топологія зв'язків та обираються функції активації нейронів, які надалі визначають принцип функціонування мережі й її ефективність для вирішення досліджуваної задачі. Так, нейромережі, що мають невелику кількість нейронів і лінійні функції активації, як правило, через свої обмежені апроксимаційні здатності не дозволяють вирішувати реальні практичні задачі. У той же час

вибір надлишкової кількості нейронів у мережі призводить до проблеми перенавчання й втрати апроксимаційних властивостей нейромоделі [2, 3].

У наш час структурний синтез нейромережних моделей відбувається за участю експерта в предметній області розв'язуваної задачі, що призводить до значного впливу рівня досвіду й знань експерта на ефективність побудованої нейромоделі [4].

Існуючі методи автоматичного пошуку оптимальної структури нейромережних моделей використовують жадібну стратегію пошуку [5]. Так конструктивні (constructive) методи починають пошук з мінімально можливою архітектурою мережі (нейромережа із мінімальною кількістю шарів, нейронів і міжнейронних зв'язків) і по-слідовно на кожній ітерації додають нові шари, нейрони й міжнейронні зв'язки. При використанні деструктивних методів (destructive) на початковій ітерації оцінюється ефективність нейромоделі, що містить максимально припустиму кількість шарів, ней-

ронів і міжнейронних зв'язків, потім у процесі пошуку структура такої моделі скорочується до найбільш прийнятної.

Однак такі методи внаслідок застосування жадібної стратегії досліджують незначну частину простору всіх можливих структур нейромоделей і схильні до влучення й «застрявання» у локальних оптимумах цільової функції.

Методи еволюційної оптимізації [6] є випадково спрямованими, не використовують жадібну стратегію пошуку й не схильні до зациклення в областях локальних екстремумів, що дозволяє ефективно їх застосовувати для пошуку оптимальної структури нейромережних моделей. Крім того, синтез структури нейромоделі за допомогою методів еволюційної оптимізації дозволяє адаптувати (налагодити, вибрати) топологію мережі до вирішення різних задач без необхідності участі експерта в цьому етапі.

Однак результатом оптимізації за допомогою класичних еволюційних методів є набір рівних між собою рішень або таких рішень, що мало відрізняються, у результаті чого оптимальна структура нейромоделі може не бути знайдена, оскільки класичні еволюційні методи можуть нерівномірно покривати простір пошуку, і значні за розміром області в просторі змінних можуть виявитися недослідженими за обмежену кількість ітерацій.

Метою даної роботи є створення полімодального методу еволюційного пошуку, який підвищує розмаїтість популяції й дозволяє більш рівномірно покривати простір пошуку, результатом якого є не єдине рішення (структура нейромережі), а множина різних рішень, що дозволяє вибрати архітектуру нейромоделі, яка якнайкраще задовольняє зовнішнім критеріям.

1 ПОСТАНОВКА ЗАВДАННЯ

Нехай задано максимально припустиму кількість нейронів A у нейромережі, що використовується для апроксимації залежності за вибіркою вихідних даних $\langle X, Y \rangle$, де $X = \{X_i\}$ – набір значень ознак, що характеризують досліджуваний об'єкт або процес; $Y = \{y_p\}$ – масив значень вихідного параметру в заданій вибірці; $X_i = \{x_{ip}\}$ – i -та ознака у вибірці, $i = 1, 2, \dots, L$; x_{ip} – значення i -ї ознаки для p -го екземпляру вибірки, $p = 1, 2, \dots, m$; y_p – значення прогнозованого параметра для p -го екземпляру; L – загальна кількість ознак у вихідному наборі; m – кількість екземплярів вибірки.

Тоді задача структурного синтезу нейромоделі полягає у пошуку структури моделі вигляду $HM = HM(C)$, для якої $\xi(HM, X, Y) \rightarrow \min$, де $C = C(L, A)$ – матриця, що визначає наявність синаптичних зв'язків між елементами мережі (рецепторами, нейронами); $\xi(HM, X, Y)$ – критерій, що визначає ефективність

використання нейромережної моделі НМ для апроксимації залежності між набором вхідних параметрів X і відповідним йому вектором значень вихідного параметру Y . Як правило, за критерій оптимальності нейромоделі обирають середньоквадратичну помилку:

$$\xi = \sum_{p=1}^m (y_p - y(HM, X_p))^2, \text{ де } X_p - \text{ набір значень оз-}$$

нак для p -го екземпляру; $y(HM, X_p)$ – значення виходу нейромоделі НМ, обчислене для набору значень X_p .

2 СТРУКТУРНИЙ СИНТЕЗ НЕЙРОМЕРЕЖНИХ МОДЕЛЕЙ НА ОСНОВІ ЕВОЛЮЦІЙНОЇ ОПТИМІЗАЦІЇ

При використанні еволюційного пошуку для синтезу структури нейромереж необхідно визначити спосіб подання структури мережі в хромосомі й вибрати фітнес-функцію для оцінювання хромосом [5, 7].

Існують такі методи кодування інформації про структуру нейромоделі в хромосомах [5]: пряме кодування, параметричне подання, комбіноване подання, подання у вигляді породжуючих правил, фрактальне подання, популяційне подання, дерево подання.

Найбільш ефективним є пряме кодування, при якому наявність кожного можливого міжнейронного зв'язку безпосередньо описується в бінарній матриці зв'язків C нейромережі, у якій одиничне значення в комірці c_{ij} відповідає наявності зв'язку від i -го до j -го нейрона. Таким чином, нейронна мережа подається у вигляді матриці суміжності.

Хромосома при прямому кодуванні подається бітовим рядком, що містить інформацію про наявність міжнейронних зв'язків.

При цьому довжина хромосоми дорівнює K^2 , де K – максимально припустима кількість вузлів (сума загальної кількості ознак у навчальній вибірці даних і максимально припустимої кількості нейронів) у нейромоделі. У випадку структурного синтезу нейромереж прямого поширення всі значення елементів матриці зв'язків, що знаходяться на головній діагоналі й нижче її, дорівнюють нулю, тому хромосому можна спростити, залишивши в ній тільки елементи матриці зв'язків, що перебувають вище головної діагоналі (рис. 1), у результаті чого кількість генів у хромосомі визначається за формулою: $K \times (K - 1) / 2$.

Як видно, при такому поданні структури мережі розмір хромосоми є пропорційним квадрату кількості нейронів мережі, внаслідок чого збільшення кількості нейронів у базовій мережі призводить до значного збільшення часу, необхідного на еволюційний пошук.

Перетворення хромосоми в структуру нейромоделі (декодування) відбувається у такий спосіб.

Крок 1. Сформувати матрицю зв'язків нейромережі, що відповідає хромосомі (рис. 1, б).

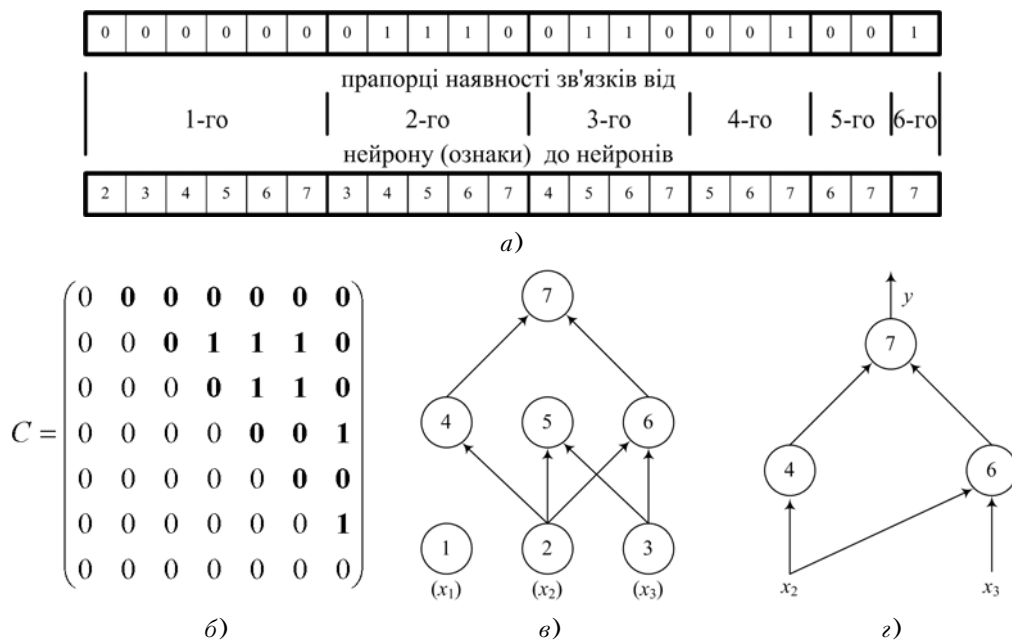


Рисунок 1 – Приклад хромосоми й відповідної їй нейромережі при структурному синтезі:

а – хромосома; б – матриця зв'язків нейромережі, що відповідає хромосомі;
в – побудований на основі матриці зв'язків граф; г – синтезована нейромережа (без надлишкових нейронів і зв'язків)

Крок 2. Побудувати на основі матриці зв'язків граф (рис. 1, в).

Крок 3. Синтезувати нейромережу на основі побудованого на попередньому кроці графа, видаливши при цьому нейрони, які не мають вихідних зв'язків до нейронів наступних шарів (рис. 1, г).

Для зменшення розміру хромосоми можна використовувати апріорну інформацію про структуру мережі. Так, наприклад, якщо відомо, що між двома сусідніми шарами мережі використовуються повнозв'язні з'єднання, то в такому випадку в хромосомі може бути подана лише інформація про кількість схованих шарів і кількість нейронів на кожному шарі.

У випадку необхідності вибору функції активації нейронів при структурному синтезі в хромосому можна ввести додаткові гени, що містять інформацію про вид функції активації для кожного нейрону.

У фітнес-функції при структурному синтезі, як правило, враховуються:

- складність синтезованої мережі, що відповідає оцінюваній хромосомі. У цьому критерії враховуються кількість шарів мережі, нейронів на кожному шарі, а також характер, топологія й кількість міжнейронних зв'язків;

- помилка прогнозування або класифікації за синтезованою моделлю.

Структурний синтез нейромоделі на основі еволюційного підходу може бути виконаний як така послідовність кроків [5, 7–8].

Крок 1. Сформувані початкове покоління хромосом, які містять інформацію про структуру мережі.

Крок 2. Виконати оцінювання хромосом поточної популяції.

Крок 2.1. Декодувати кожну хромосому в популяції в архітектуру відповідної нейронної мережі.

Крок 2.2. Навчити кожну нейронну мережу за обраним методом (наприклад, за допомогою методу Левенберга-Марквардта [1]) за допомогою даних з навчальної вибірки (виконати параметричний синтез нейромережі). Початкові значення вагових коефіцієнтів при навчанні задаються залежно від обраного методу параметричного синтезу або випадковим чином.

Крок 2.3. Обчислити значення фітнес-функції хромосом, що враховує помилку нейромережі й складність її архітектури (кількість шарів, нейронів, міжнейронних зв'язків).

Крок 3. Перевірити критерії закінчення пошуку. У випадку задоволення таких критеріїв, виконати перехід до кроку 7.

Крок 4. Вибрати найбільш пристосовані хромосоми для виконання над ними еволюційних операторів схрещування й мутації.

Крок 5. Виконати оператори схрещування й мутації над відібраними раніше хромосомами.

Крок 6. Створити нове покоління з отриманих на попередньому кроці хромосом-нащадків і найбільш

приспособлених хромосом поточного покоління. Виконати перехід до кроку 2.

Крок 7. Зупинення.

З ПОЛІМОДАЛЬНИЙ ЕВОЛЮЦІЙНИЙ МЕТОД СТРУКТУРНОГО СИНТЕЗУ НЕЙРОМОДЕЛЕЙ

При вирішенні завдання структурного синтезу нейромережних моделей цільова функція, як правило, має декілька оптимумів [1-4], розташованих у різних областях простору пошуку. При використанні класичних методів еволюційної оптимізації результатом пошуку є популяція розв'язків, які мало відрізняються один від одного, у результаті чого знайдене рішення може виявитися локальним оптимумом багатоекстремальної функції. Такий розв'язок (структура нейромоделі), як правило, є неефективним при його використанні на практиці, оскільки не забезпечує достатню точність класифікації для даних, які не входять до навчаючої вибірки.

Тому для структурного синтезу нейромоделей доцільним є використання еволюційних методів, здатних до пошуку декількох субоптимальних розв'язків. При використанні традиційних методів еволюційного пошуку до оптимізації полімодальних функцій існує проблема передчасної збіжності до локального оптимуму. Для подолання даної проблеми розроблено дві групи методів: методи відхилення від передчасної збіжності та методи відновлення [9].

Методи першої групи зменшують генетичну збіжність популяції з метою забезпечення її розмаїтості й недопущення зациклення в локальних оптимумах. Однак при такому підході використовуються штрафні функції для зменшення ймовірності появи в популяції схожих розв'язків, що викликає необхідність розрахунку значень таких функцій для кожної хромосоми й, отже, значно сповільнює процес еволюційного пошуку.

Методи відновлення спрямовані на збільшення розмаїтості в популяції шляхом використання стратегії перезавантаження у випадках недостатньої розмаїтості в популяції, що також вимагає значних часових витрат.

У розробленому методі полімодального еволюційного пошуку з кластеризацією хромосом пропонується групувати рішення (хромосоми) у кластери за їхнім розташуванням в просторі пошуку.

Пропонований метод у процесі еволюційного пошуку визначає групи близьких (подібних) хромосом і підвищує розмаїтість популяції шляхом погіршення значень фітнес-функцій хромосом в залежності від близькості до центру їхньої групи, не вимагаючи при цьому обчислення значень штрафних функцій і не використовуючи процедури перезавантаження.

Розроблений полімодальний еволюційний пошук з кластеризацією хромосом передбачає виконання таких кроків.

Крок 1. Задати: кількість оптимумів (оптимальних структур нейромоделей) k , які необхідно знайти в ході еволюційного пошуку; N – кількість хромосом у популяції, $N \gg k$.

Крок 2. Встановити лічильник ітерацій: $t = 1$.

Крок 3. Встановити кількість елітних особин (хромосом): $k_e = k$.

Крок 4. Ініціалізувати початкову популяцію у вигляді хромосом H_j , $j = 1, 2, \dots, N$.

Крок 5. Обчислити значення фітнес-функції $f(H_j)$ для кожної хромосоми H_j .

Крок 6. Згрупувати хромосоми в k кластерів за значенням їхніх фітнес-функцій і розташуванням у просторі пошуку.

Крок 6.1. Для кожної хромосоми H_j обчислити відстань Хемінга (кількість незбіжних генів в однакових позиціях хромосом) до всіх інших хромосом у популяції. Відстань Хемінга $d(H_j; H_l)$ між хромосомами H_j та H_l розраховується за формулою:

$$d(H_j; H_l) = \sum_{i=1}^L |h_{ij} - h_{il}|,$$

де L – розмір хромосом; h_{ij} та h_{il} – значення i -х бітів хромосом H_j та H_l , відповідно.

Крок 6.2. Встановити лічильник сформованих кластерів: $m = 1$.

Крок 6.3. Вибрати хромосому з найкращим значенням фітнес-функції як центр m -го кластеру. При цьому розглядаються хромосоми, ще не згруповані по кластерах.

Шаг 6.4. Ввести в кластер $(N/k - 1)$ хромосом, найближчих за відстанню Хемінга до хромосоми, що є центром поточного m -го кластеру. У випадку, якщо (N/k) – неціле, то для визначення необхідної кількості хромосом в m -му кластері k_m (не враховуючи вже присутню хромосому, що є центром кластеру) пропонується використовувати формулу $k_m = \text{Round}(m/k) - (m - 1) \cdot \text{Round}(N/k) - 1$, де $\text{Round}(p)$ – округлене значення p .

Крок 6.5. Якщо всі кластери сформовані ($m = k$), тоді виконати перехід до кроку 7.

Крок 6.6. Встановити: $m = m + 1$. Виконати перехід до кроку 6.3.

Крок 7. Збільшити значення фітнес-функцій хромосом, що не є кращими в кластері, за формулою:

$$f_n(H_j) = \left(\frac{d(H_j; H_{\max, j})}{d(H_j; H_{c, j})} \right)^s f_j,$$

де $f_n(H_j)$ – нове значення фітнес-функції j -ї хромосоми; $f(H_j)$ – значення фітнес-функції до зміни j -ї хромосоми; $d(H_j; H_{c,j})$ – відстань Хемінга від j -ї хромосоми до центра її групи; $d(H_j; H_{\max,j})$ – максимальна відстань Хемінга в групі j -ї хромосоми; s – параметр, що визначає ступінь погіршення значень фітнес-функцій хромосом, що не є центрами кластерів, $s \geq 1$.

Крок 8. Відібрати хромосоми для схрещування й мутації.

Крок 9. Застосувати оператори схрещування й мутації до відібраних на попередньому кроці хромосом. Як такі оператори пропонується використовувати класичні еволюційні оператори (наприклад, точкове, однорідне або рівномірне схрещування й просту мутацію).

Крок 10. Сформувати нове покоління. При цьому кращі (елітні) хромосоми в кожному кластері гарантовано переходять у нове покоління.

Крок 11. Якщо $t = T$, де T – максимально задана кількість ітерацій, тоді виконати перехід до кроку 14.

Крок 12. Встановити: $t = t + 1$.

Крок 13. Виконати перехід до кроку 5.

Крок 14. Оцінити кожну з k хромосом, що є центрами кластерів, за допомогою даних тестової вибірки. Вибрати найкращу хромосому. Нейромодель, що відповідає такій хромосомі, приймається як рішення.

Крок 15. Зупинення.

Розроблений метод полімодального еволюційного пошуку з кластеризацією хромосом підвищує розмаїтість популяції й дозволяє більш рівномірно покрити простір пошуку, підвищуючи в такий спосіб можливість відшукування глобального оптимуму й збільшуючи ймовірність успішного виконання процедури оцінки знайдених рішень за допомогою зовнішніх критеріїв на тестовій вибірці, оскільки результатом пошуку є не єдина структура нейромережі, а множина різних розв'язків, що дозволяє вибрати таку нейромережну модель, що якнайкраще задовольняє зовнішнім критеріям оцінювання ефективності нейромоделей.

4 ЕКСПЕРИМЕНТИ Й РЕЗУЛЬТАТИ ПО МОДЕЛЮВАННЮ ПОКАЗНИКА ЗДОРОВ'Я ДІТЕЙ

З метою дослідження ефективності запропонованого методу полімодального еволюційного пошуку з кластеризацією хромосом для структурного синтезу нейромоделей вирішувалася практична задача побудови моделей залежності стану здоров'я населення від забруднення навколишнього середовища.

У великих промислових містах України в умовах несприятливої екологічної ситуації актуальним є вивчення впливу показників забруднення навколишнього середовища на стан здоров'я населення й, у першу

чергу, здоров'я дітей, оскільки діти є більш сприйнятливими до впливу екологічних факторів.

У ході проведеного вибіркового дослідження стану здоров'я дітей шкільного віку [10], які проживають у трьох різних за рівнем техногенного навантаження районах міста Запоріжжя (контрольному й двох дослідних) у дітей дослідних районів були виявлені статистично достовірні відхилення показників здоров'я в порівнянні з контрольним районом. Однак необхідно брати до уваги низьку специфічність виявлених змін стану здоров'я. За даними літератури [11] подібні ефекти можуть викликатися крім екологічних й іншими факторами, у першу чергу медико-біологічного й соціального характеру.

Тому актуальним є встановлення кількісного внеску екологічних і інших факторів у формування здоров'я дитячого населення.

Вихідна вибірка складалася з екземплярів, кожний з яких являв собою набір значень характеристик для відповідної дитини, і складалася з 48 ознак, що характеризують екологічні, медико-біологічні й соціально-побутові фактори, які впливають на стан здоров'я дітей.

Перед синтезом моделі залежності стану здоров'я населення від забруднення навколишнього середовища було проведене виділення інформативного набору ознак. Для відбору ознак використовувався еволюційний метод з фіксацією частини простору пошуку [12]. У результаті скорочення розмірності вихідного ознакового набору було залишено 19 ознак: x_1 – район; x_2 – вік, років; x_3 – стать; x_4 – маса, кг; x_5 – рівень фізичного розвитку; x_6 – пульс до навантаження; x_7 – кількість років проживання дитини в даному районі; x_8 – час щоденного перегляду телевізора; x_9 – тривалість нічного сну, годин; x_{10} – чи займається дитина спортом; x_{11} – вік матері при народженні дитини, років; x_{12} – чи була пов'язана робота матері із професійними хворобами до народження дитини; x_{13} – тип родини (повна, неповна); x_{14} – протікання вагітності (без ускладнень, з ускладненнями); x_{15} – чи страждає мати хронічними захворюваннями; x_{16} – чи страждає батько хронічними захворюваннями; x_{17} – чи курить батько; x_{18} – чи курить мати; x_{19} – середній дохід на одного члена родини на місяць, грн; y – індекс напруги, що визначає стан здоров'я дитини, у. о.

Для структурного синтезу нейромоделей використовувалися: метод додавання нейронів, метод видалення нейронів, класичний еволюційний пошук і запропонований еволюційний метод полімодального еволюційного пошуку з кластеризацією хромосом, який був програмно реалізований мовою пакета Matlab. Максимально припустима кількість нейронів для всіх методів структурного синтезу – 15, при цьому останній шар нейромоделі містив один нейрон, оскільки модель має один вихід, всі нейрони мали сигмоїдні функції акти-

вації, а як дискримінантні функції використовувалися зважені суми.

Початкові значення параметрів еволюційних методів встановлювалися такими: оператор відбору – з використанням рулетки, оператор схрещування – рівномірний, оператор мутації – точковий, кількість особин у популяції $N = 100$, ймовірність схрещування $p_{скр} = 0,8$, максимальна кількість ітерацій $T = 100$, кількість елітних особин $N_e = 4$.

Як критерій оцінювання синтезованих нейромоделей пропонується використовувати вираз:

$$f = E + \frac{N_w}{N_{w\max}}$$

де E – помилка прогнозування за синтезованою моделлю; N_w – кількість синаптичних зв'язків у мережі; $N_{w\max}$ – максимально можлива кількість зв'язків у мережі. Як помилка прогнозування приймалася середньоквадратична помилка нейромоделі, параметричний синтез якої виконувався за допомогою методу зворотного поширення помилки [1–2].

Таким чином, запропонований критерій дозволяє враховувати як помилку прогнозування за синтезованою нейромоделлю, так і її складність.

У зв'язку з імовірнісним характером роботи еволюційних методів для дослідження їхніх властивостей побудова моделей за допомогою таких методів виконувалася 100 разів для одержання більш об'єктивних усереднених результатів.

У табл. 1 наведені результати синтезу нейромережних моделей залежності стану здоров'я населення від екологічних, медико-біологічних і соціально-побутових факторів.

У методах додавання нейронів і видалення нейронів результатом є нейромережі, що містять всі можливі прямо спрямовані зв'язки від нейронів поточного шару до нейронів наступного шару. Тому, як видно з табл. 1 (стовпець «Кількість зв'язків»), нейромоделі, отримані в результаті структурного синтезу за допомогою еволюційних методів, є більш простими, оскільки містять значно меншу кількість зв'язків. Це викликано тим, що методи еволюційного пошуку, на відміну від методів додавання й видалення нейронів, як оптимізовані параметри використовують не кількість нейронів у мережі, а наявність міжнейронних зв'язків.

Як видно з табл. 1, помилка прогнозування за більшістю синтезованих за допомогою різних методів нейромоделей значно погіршується при використанні даних тестової вибірки. Тому доцільним є застосування розробленого методу полімодального еволюційного пошуку з кластеризацією хромосом, результатом якого є не єдиний, а декілька різних розв'язків (структур нейромереж), деякі з яких забезпечують досить гарні результати й для даних тестової вибірки. Так середньоквадратична помилка нейромоделі зі структурою 4–2–1, розрахована за даними тестової вибірки, становить 0,0097, що незначно гірше в порівнянні з помилкою, розрахованою для даних навчаючої вибірки.

У табл. 2 наведено структуру нейромережі, що забезпечує достатню точність прогнозування як на даних навчаючої вибірки (0,0085), так і на даних тестової вибірки (0,0097).

Результати побудови нейромережних моделей, отримані за допомогою застосування різних методів структурного синтезу, показали, що розроблений метод полімодального еволюційного пошуку з кластеризацією хромосом дозволяє одержати різні структури нейромоделей, підвищуючи в такий спосіб імовірність синтезу

Таблиця 1 – Результати структурного синтезу нейромоделей

№	Назва методу	Середній час, с	Параметри кращої нейромережної моделі				
			Структура	Кількість зв'язків	Значення критерія оцінювання f	Середньоквадратична помилка	
						за навчаючою вибіркою	за тестовою вибіркою
1	Метод додавання нейронів	181,2	5–2–1	107	0,3914	0,0093	0,0207
2	Метод видалення нейронів	327,6	4–2–1	86	0,3158	0,0087	0,0193
3	Класичний еволюційний пошук	285,9	4–3–1	53	0,1919	0,0096	0,0162
4	Метод полімодального еволюційного пошуку з кластеризацією хромосом	296,3	4–2–1	47	0,1764	0,0085	0,0097
			5–3–1	56	0,2083	0,0083	0,0178
			3–2–1	42	0,1592	0,0092	0,0232
			4–3–1	49	0,1844	0,0094	0,0114

Таблиця 2 – Структура оптимальної нейромоделі

Номер шару	Номер нейрона в шарі	Входи нейрона (та значення вагових коефіцієнтів)
1	1	ознаки № 1 (0,462), 2 (-7,228), 4 (-0,824), 5 (-1,733), 8 (3,065), 12 (-2,751), 14 (5,488), 16 (4,524)
	2	ознаки № 1 (-0,174), 2 (1,068), 3 (-2,571), 4 (-0,171), 6 (2,353), 7 (-1,384), 8 (-3,205), 11 (-1,518), 12 (1,005), 14 (-7,203), 15 (8,507), 16 (-0,807), 18 (2,13), 19 (-1,031)
	3	ознаки № 1 (3,412), 4 (-2,183), 5 (-3,343), 6 (-1,419), 9 (2,81), 10 (-1,723), 11 (-4,665), 12 (-1,176), 14 (-1,029), 15 (2,073), 16 (-4,616), 17 (-1,152)
	4	ознаки № 4 (1,827), 7 (-3,163), 12 (5,49), 17 (-0,733), 18 (-4,41), 19 (0,161)
2	1	нейрони 1-го шару № 1 (-2,091), 2 (10,55), 3 (2,361)
	2	нейрони 1-го шару № 3 (-0,242), 4 (3,751)
3	1	нейрони 2-го шару № 1 (-0,558), 2 (2,382)

нейромереж, що мають прийнятні апроксимаційні здатності як для даних навчаючої, так і для даних тестової вибірок.

ВИСНОВОК

У роботі вирішено завдання структурного синтезу нейронних мереж за допомогою методів еволюційного пошуку.

Наукова новизна роботи полягає в тому, що для структурного синтезу нейромоделей розроблено метод полімодального еволюційного пошуку з кластеризацією хромосом, який рівномірно покриває простір пошуку і підвищує можливість відшукування глобального оптимуму, збільшуючи ймовірність успішного виконання процедури оцінки знайдених рішень за допомогою зовнішніх критеріїв на тестовій вибірці.

Практична цінність результатів роботи полягає в тому, що:

– створено програмне забезпечення, яке реалізує запропонований метод структурного синтезу нейромоделей;

– вирішено завдання синтезу нейромережних моделей залежності стану здоров'я населення від екологічних, медико-біологічних і соціально-побутових факторів.

ПЕРЕЛІК ПОСИЛАНЬ

1. Омату С., Халид М., Юсоф Р. Нейроуправление и его приложения / Пер. с англ. под ред. А. И. Галушкина, В. А. Птичкина. – М.: ИПРЖР, 2000. – 272 с.
2. Круглов В. В., Борисов В. В. Искусственные нейронные сети: Теория и практика. – М.: Горячая линия – Телеком, 2001. – 382 с.
3. Руденко О. Г., Бодянский Е. В. Основы теории искусственных нейронных сетей. – Харьков: Телетех, 2002. – 317 с.
4. Минаев Ю. Н., Филимонова О. Ю., Лиес Б. Методы и алгоритмы идентификации и прогнозирования в ус-

- ловиях неопределенности в нейросетевом логическом базисе. – М.: Горячая линия – Телеком, 2003. – 205 с.
5. Yao X. Evolving Artificial Neural Network // Proceedings of the IEEE. – 1999. – № 9(87). – P. 1423–1447.
 6. The Practical Handbook of Genetic Algorithms. Volume II. New Frontiers / Ed. L. D. Chambers. – Florida: CRC Press, 2000. – 421 p.
 7. Maniezzo V. Genetic Evolution of the Topology and Weight Distribution of Neural Networks // IEEE transactions on neural networks. – 1994. – № 1(5). – P. 39–53.
 8. Romaniuk S. G. Applying Crossover Operators to Automatic Neural Network Construction // IEEE transactions on neural networks. – 1994. – № 4. – P. 750–752.
 9. Eiben A.E., Hintering R., Michalewicz Z. Parameter Control in Evolutionary Algorithms // IEEE Transactions on Evolutionary Computation. – 1999. – № 3. – P. 124–141.
 10. Кирсанова Е. В., Субботин С. А. Обобщенный метод кластер-регрессионной аппроксимации в задаче моделирования показателя здоровья детей // Радиоэлектроника. Информатика. Управление. – 2004. – № 1. – С. 62–67.
 11. Вельтищев Ю. Е. Экологически детерминированная патология детского возраста // Российский вестник периматологии и педиатрии. – 1996. – № 2. – С. 5–12.
 12. Subbotin S., Oleynik A. The feature selection method based on the evolutionary approach with a fixation of a search space // Modern problems of radio engineering, telecommunications and computer science: Proceedings of the IX International Conference TCSET'2006 (21–25 February 2006). – Lviv: Publishing house of Lviv Polytechnic, 2006. – P. 574–575.

Надійшла 5.03.2008

Исследована проблема структурного синтеза нейросетевых моделей. Разработан метод полимодального эволюционного поиска с кластеризацией хромосом, позволяющий получить различные структуры нейросетей, повышая вероятность нахождения модели, адекватной решаемой задаче. Проведены эксперименты по синтезу моделей зависимости состояния здоровья населения от загрязнения окружающей среды.

The problem of structural synthesis of neural network models is considered. The polymodal evolutionary search method with a chromosomes clustering is developed, allowing to develop various structures of neural networks with good computational capabilities. Experiments on model synthesis of health state dependence from the environmental population are lead.

М. Ю. Юрич, А. Н. Щербаков, А. Ф. Камышанов

АРХИТЕКТУРА ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ НЕЙРОУПРАВЛЕНИЯ ЛЕТАТЕЛЬНЫМ ОБЪЕКТОМ С ЭВОЛЮЦИОННЫМ ПРОГНОЗОМ ПРЕДПОЧТИТЕЛЬНОГО ПОВЕДЕНИЯ

В данной статье рассматривается архитектура интеллектуальной системы управления, основанной на нейронной сети и эволюционных алгоритмах для прогноза и выбора предпочтительного поведения летательных аппаратов. Сложная технология прикладного программирования, особенно в условиях информативной неопределенности, заменяется относительно простой технологией обучения нейроконтроллера правильно реагировать на изменения параметров на входе системы управления.

ВВЕДЕНИЕ

В последнее время попытки применить искусственный интеллект в той или иной области науки и техники стали более частыми. Свойство интеллектуальных систем принимать решения в условиях изменяющегося окружения позволяет обеспечить управление летательными аппаратами (ЛА).

Управление современными ЛА осуществляется при помощи мультиструктурного управляющего комплекса в результате передачи функций управления ЛА, ранее осуществляемых человеком, системе управления без вмешательства человека.

Анализ математических методов моделирования сложных моделей интеллектуальных систем управления (ИСУ) ЛА показал, что одним из наиболее перспективных подходов к синтезу ИСУ ЛА может быть подход, основанный на теории эволюционного моделировании функциональных систем [1].

Классические методы синтеза систем управления базируются на хорошо развитом аппарате интегро-дифференциального исчисления. Нейронные сети представляют собой альтернативное, существующее всего несколько десятков лет, направление в теории автоматического управления, предлагающее иной способ отражения и преобразования действительности, в котором можно обнаружить и сходные, и различные черты с классической парадигмой.

Термин «нейроуправление» впервые появился в работах Вербоса уже в 1976 году, однако решающую роль во внедрении искусственных нейронных сетей

(ИНС) в сферу управленческих задач сыграли работы Нарендры с соавторами [2].

В настоящее время в системах управления ИНС применяются нейроконтроллеры и нейроэмуляторы, имитирующие динамическое поведение объекта управления в целом или описывающие его отдельные характеристики, трудно поддающиеся математическому моделированию.

ПОСТАНОВКА ЗАДАЧИ

В данной работе ставится цель разработать архитектуру интеллектуальной системы управления, построенной на таких принципах, алгоритмах и программах эволюционного моделирования, которые дают возможность отбора лучших моделей ситуационного управления ЛА, а также проведение их селекции с помощью генетического алгоритма (ГА), формирование из лучших моделей-потомков шаблонов для обучения нейроконтроллера, способного заменить сложный вычислительный узел в системе управления. А также формирование банка знаний из шаблонов обучения для типичных ситуаций, которые можно использовать в целях распознавания ситуации, прогнозирования и выбора алгоритма ситуационного управления.

РАЗРАБОТКА АРХИТЕКТУРЫ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ЛА

Используя модель интеллектуальной системы с виртуальным обучением, рассмотренную в работе [2], разрабатываем концептуальное решение ИСУ ЛА на базе ИНС с блоком обучения, реализующим ГА для оптимального управления, добавляя такие блоки как выбор типовых сценариев, бортовая база знаний ИСУ, наземная база знаний ИСУ, оценка текущей обстановки. Получив в виртуальном тракте модель управления, близкую к желаемой, ИНС обучает нейроконтроллер виртуального тракта, проверяя его работу с помощью нейроэмулятора по значению ошибки расхождения его

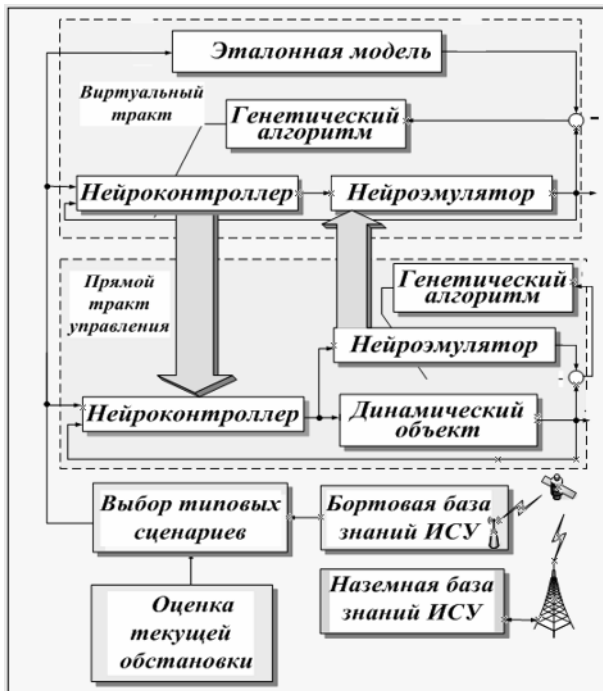


Рисунок 1 – Структура ИСАУ ЛА с виртуальным обучением

выхода с эталонной моделью. Как только ошибка станет минимальной, шаблоны фенотипов передаются на нейроконтроллер прямого тракта управления для обучения и управления динамическим объектом. Структура интеллектуальной системы автоматического управления (ИСАУ) ЛА с виртуальным обучением представлена на рис. 1.

В данной системе изучается объект и постепенно совершенствуются способности к управлению.

Динамические свойства объектов управления и системы могут изменяться случайным образом, что проявляется в существенном возрастании ошибки прогноза динамического состояния. Следовательно, ошибочными будут и управляющие воздействия, сформированные на основе результатов прогноза, что может ухудшить динамическое качество системы и снизить точность управления. Поэтому все алгоритмы адаптивного управления, основанные на использовании прогнозирующих моделей ГА плюс ИНС, требуют применения процедур их параметрического оценивания. Использование прогнозирующих моделей ГА плюс ИНС в режиме реального времени позволяет в неявной форме решить задачу параметрического оценивания и реализовать адаптивную настройку параметров регулятора.

Как указано в [3], при внедрении в бортовых системах управления элементов искусственного интеллекта ставят задачу реализовать их на уровне выбора способа достижения поставленной цели и реализации спосо-

ба разрешения текущей проблемы субситуации. Для каждой типовой ситуации должны работать соответствующая бортовая оперативно-советующая экспертная система (БОСЭС), экспертные системы бортовых измерительных и исполнительных устройств, которые обеспечивают получение необходимой информации в текущий момент времени и точное исполнение принятых решений.

Эффективность ИСУ и целесообразность ее разработки зависят от структуры, полноты и адекватности базы знаний.

Для каждой субситуации необходимо построить соответствующую базу знаний. Определенная субситуация представляет собой совокупность математических моделей, которая описывает пространственное расположение участников субситуации, прогнозирует изменение расположения участников и определяет моменты времени возможного наступления значимых событий.

В практических приложениях часто требуется предварительное исследование субситуации с использованием различных математических моделей, которые формулируются в форме математических задач оптимального управления, теории игр, задач принятия решений. Полученное или подобранное при имитационном математическом моделировании рациональное разрешение субситуации закладывается в математические модели (ММ), которые являются одним из блоков базы знаний в БОСЭС. Анализ внешней и бортовой обстановки, а также выработка соответствующих рекомендаций управления закладываются в базу знаний БОСЭС. Знания, необходимые для работы БОСЭС, используются в виде правил, ММ и группируются по сценариям.

Технология разработки базы знаний предполагает разработку системы критериев выбора (четких таблиц), на основе которых эксперты могут определять принадлежность всех возможных сценариев для конкретной ситуации. Неопределенность оценок экспертов выражается нечеткими переменными. Все возможные сценарии должны быть предварительно промоделированы в реальном времени и оценены опытными экспертами. Проверка и отработка моделей сценариев проводится на наземном комплексном стенде с участием макетов ЛА. Сценарии и критерии выбора определяются из ситуационного моделирования.

В системе измерений параметров и обстановки внешней среды отличительной чертой любого результата, который способствует достижению цели, является то, что он получается на основе принципа саморегуляции и обладает одинаковыми узловыми механизмами: синтезом цели, принятием решения к действию, сравнением параметров полученного результата с прогнозированными.

В системе выбора типовых сценариев управления ЛА акцептор действия реализуется с помощью алгоритма самоорганизации, который позволяет построить прогнозирующую модель при минимуме априорной информации бортовых параметров и панорамной обстановки. Информация о соответствии результатов действия и прогноза передается в динамическую экспертную систему (ДЭС) и в систему выбора типовых сценариев.

Таким образом, при отсутствии соответствия результатов действия и прогноза вырабатывается новая экспертная оценка, принимается другое решение и реализуется новое управляющее воздействие. Если соответствия достичь невозможно, то изменяется цель функционирования ИС. Если результаты действия соответствуют прогнозу, то управляющее воздействие выбрано правильно и цель ИС достигается.

Система адаптивной настройки регулятора прямого тракта управления предназначена для формирования сигналов, подаваемых на исполнительные механизмы динамического объекта ЛА.

Система ручного управления позволяет осуществлять навигацию в ручном режиме с использованием консультативного варианта ИСУ. Консультации, сове-

ты, подсказки, выводятся на блок бортовой индикации и на наземный комплексный стенд.

При формировании функций ИС, часто возникают проблемы недостатка информативности базы знаний, ограниченной объемом памяти. Возникают вопросы и при реализации модели на содержательном и конструктивном уровнях, что требует дополнительной обработки ситуации. Все эти вопросы решаются на наземном комплексном стенде (НКС). ДЭС НКС наряду с задачами расчета, оптимизации, моделирования будущих результатов должна обеспечивать высокое быстродействие. Теоретические результаты, полученные при разработке алгоритмов принятия решений, и выработки управления на НКС, можно использовать дистанционно в ИС. Реализация алгоритмического обеспечения ИС требует параллельного выполнения ряда сходных алгоритмов и их проверки на оптимальность для практической реализации.

На рис. 2 представлена блок-схема описанной выше интеллектуальной системы управления ЛА.

Комплекс ИСАУ ЛА состоит из бортового аппаратно-программного комплекса (БАПК) и наземного комплексного стенда в состав которого входит действующая модель ЛА, БАПК, мощный комплекс вычислительных средств с программным обеспечением.

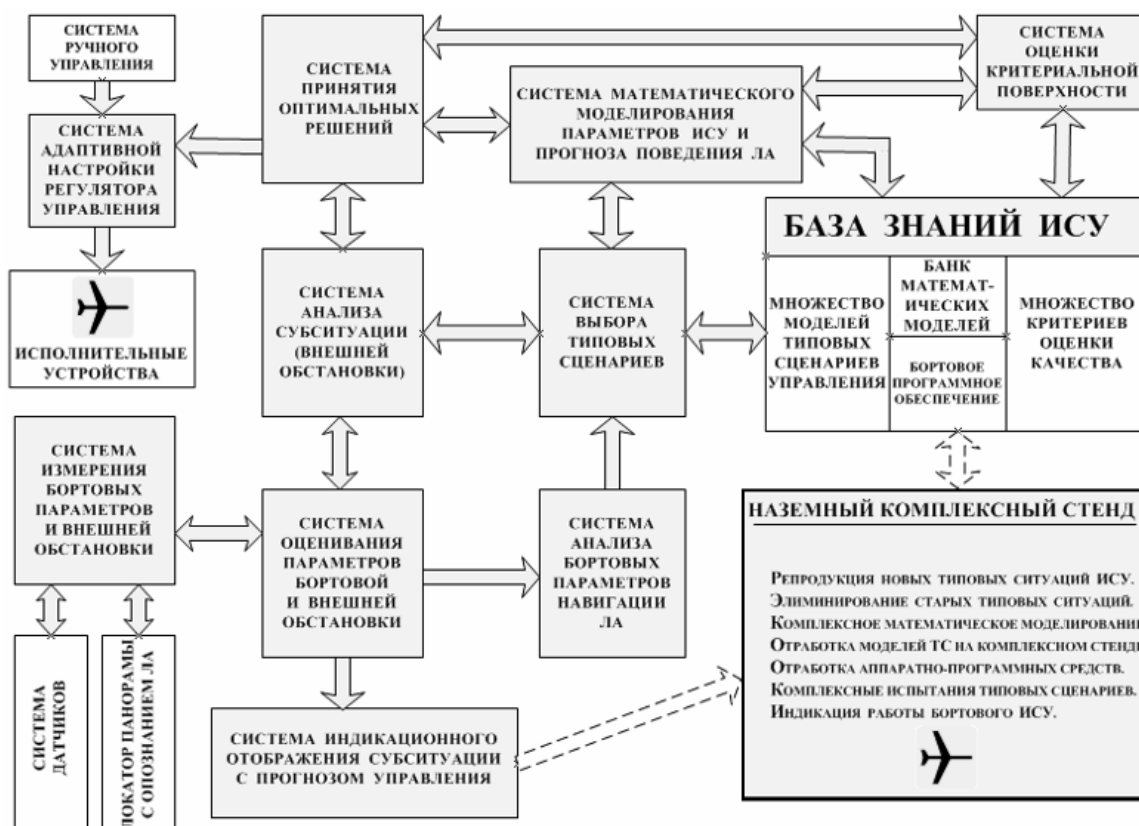


Рисунок 2 – Блок-схема интеллектуальной системы управления ЛА

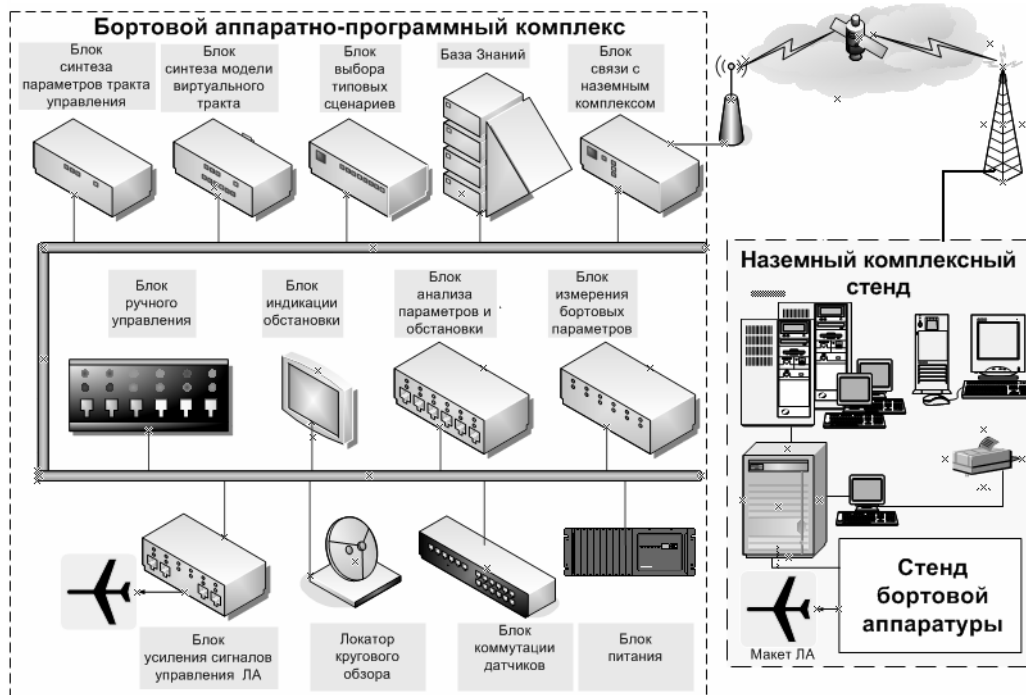


Рисунок 3 – Комплекс ИСАУ ЛА

Бортовой аппаратно-программный комплекс размещается на ЛА и связан с НКС через радиоканал спутниковой связи. Этот канал дает возможность НКС дистанционно контролировать состояние БАПК перед полетом и во время полета. Состав аппаратных средств ИСАУ приведен на рис. 3.

ВЫВОДЫ

В данной статье представлена разработанная архитектура интеллектуальной системы управления летательными аппаратами. Найдено концептуальное решение разрабатываемой системы управления на базе ИНС с блоком обучения, реализующим ГА для поиска оптимального управления. А также спроектирована блок-схема интеллектуальной системы управления ЛА. Подобная структура ИСАУ может быть использована при синтезе других систем управления динамическими объектами.

Полученные результаты могут быть использованы для создания конкурентоспособных технологий управления в организациях, выполняющих исследовательские работы в области интеллектуального управления, космической и авиационной техники, в области управления технологическими объектами промышленности.

В дальнейшем планируется использовать предложенную архитектуру интеллектуальной системы для исследований и разработок в области управления динамических объектов.

ПЕРЕЧЕНЬ ССЫЛОК

1. Редько В. Г. От моделей поведения к искусственному интеллекту. Серия «Науки об искусственном» / Под ред. Редько В. Г., ИОНТ РАН. – М.: УРСС, 2006. – 456 с.
2. Вороновский Г. К. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности / Г. К. Вороновский, К. В. Махотило, С. Н. Петрашев, С. А. Сергеев. – Х.: ОСНОВА, 1997. – 112 с.
3. Кэ Фан, Фам Суан Фанг, Ким Дже Су. Исследование интеллектуальных систем управления летательными аппаратами // Автоматизация и современные технологии. – 2006. – № 7. – Стр. 31–34.

Надійшла 14.12.2007
Після доробки 6.02.2008

У даній статті розглянуто архітектуру інтелектуальної системи управління, яка базується на нейронній мережі та еволюційних алгоритмах для прогнозування та вибору задовольняючою поведінки літальних апаратів. Складна технологія прикладного програмування, особливо в умовах інформативної невизначеності, замінюється відносно простою технологією навчання нейроконтролера правильно реагувати на зміну параметрів на вході системи управління.

In this paper are showed creation and architecture of intellectual control system (ICS) based on neural network and evolutionary algorithms for forecasting and selection of behavior for aircraft. Complex technology of the applied programming, especially in conditions of informative uncertainty, was replaced by rather simple technology of training of neural controller to react correctly to change of parameters on a control system entrance.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 004.451

С. А. Зайцев, С. К. Корниенко

ОПЕРАЦИОННАЯ СИСТЕМА ДЛЯ ВСТРАИВАЕМЫХ УСТРОЙСТВ

Проведен анализ рынка современных операционных систем для встраиваемых устройств. Разработано ядро многозадачной операционной системы, предназначенной для управления встраиваемым устройством. Описаны основные принципы функционирования встраиваемой операционной системы и предложены модели разработки встраиваемых приложений и драйверов.

ВВЕДЕНИЕ

Быстрый рост рынка портативных компьютеров, интеллектуальных устройств, сетевых компьютеров и т. д. стимулирует интерес к встраиваемым системам. Встраиваемая система (*embedded system*) – это специализированная компьютерная система, в которой сам компьютер обычно встроено в устройство, которым он управляет. Примерами встраиваемых систем могут служить банкоматы, авионика, карманных персональных компьютеров (КПК), телекоммуникационное оборудование и тому подобные устройства.

Целью данной работы является разработка операционной системы (ОС) для встраиваемых решений, основанных на процессорах семейства Advanced RISC Machines (ARM).

1 ОС ВСТРАИВАЕМОЙ СИСТЕМЫ

Программным обеспечением для встраиваемой системы является микропрограмма (англ. *firmware*). Около 75 % встраиваемых решений используют какие-либо ОС, и лишь остальные 25 % реализованы в виде самостоятельной программы, управляющей ресурсами системы [1]. Как правило, операционная система в таких случаях не является самостоятельным продуктом, а лишь частью монолитной микропрограммы устройства.

Более половины решений, работающих под управлением ОС – 51 % (или 36,3 % от общего числа систем) используют коммерческие ОС. В следующей по численности группе используются ОС собственной разработки – 21 % или 15,1 % от всех систем. Решения на базе условно «бесплатных» ОС с открытым кодом и на коммерчески распространяемых системах с открытым кодом занимают 16 % и 11,8 % соответственно (11,5 % и 8,4 % общего числа систем) [2].

Таким образом, если создать бесплатную ОС с открытым кодом, которая обеспечит такую же гибкость, что и ОС собственной разработки, то это удовлетворит преобладающую часть рынка.

© Зайцев С. А., Корниенко С. К., 2008

Из-за накладываемых ограничений на подобные ОС встраиваемых систем их называют ОС реального времени.

ОС реального времени (далее – realtime OS, RTOS) – это многозадачные ОС, выполняющие задачи в реальном времени [3]. Основные подходы к реализации RTOS базируются либо на реакциях на события (event-driven RTOS), либо на распределении времени выполнения задач (переключении задач по прерыванию от таймера). Второй подход выполняет переключение задач чаще, чем требуется, но при этом задачи выполняются более «гладко» и создается впечатление, что задачи выполняются параллельно.

Существующие встраиваемые ОС можно разделить на две группы. Первые представляют собой UNIX-подобные ОС (QNX, eCos, VxWorks), вторые же являются самостоятельными продуктами и представляют собой специфичную реализацию (uC/OS, TNKernel, FreeRTOS). Они обладают полезными особенностями, например ОС uC/OS внедрила поддержку сопроцедур (co-routines) – программ с минимальными требованиями к ресурсам, работающих параллельно с основными задачами ОС.

Практически все встраиваемые ОС поддерживают многозадачность, но лишь небольшое их количество поддерживает возможность выбора алгоритма планировщика (подробнее об алгоритмах планировщика смотрите ниже). В большинство ОС включены примитивы синхронизации – семафоры, очереди сообщений и т. д. Как правило, минимальный размер таких ОС – от 2 до 4 Кб.

В то же время в некоторых сферах задач существующие ОС обладают и недостатками – многие решения имеют высокую цену (цена лицензии ОС eCos оставляет около 12000\$). Большинство ОС ориентированы на разработку только под конкретной гостевой ОС, что ограничивает разработчика (ОС ThreadX ориентирована на разработку только из-под ОС Windows), к тому же в состав многих ОС не входит загрузчик. Из-за этого многие ОС представляют собой только набор базовых функций (FreeRTOS, TNKernel). Сами по себе без пользовательских приложений такие ОС работать не могут. Относительно высокие требования развитых встраиваемых ОС не позволяют применять их в простых встраиваемых системах (uClinux нуждается в 512 Кб RAM [3,4], eCos – в 32 Кб, в то время как дешевые встраиваемые решения обладают примерно 64 Кб RAM, которая расходуется на прикладные задачи в большем количестве, чем на ОС).

Таким образом, рынок встраиваемых ОС представлен широким ассортиментом, но не для всех практических задач (в особенности – для малых встраиваемых систем) можно сделать однозначный выбор ОС.

2 ОБЗОР АРХИТЕКТУРЫ ARM

Хотя в дальнейшем планируется поддержка других архитектур, в данный момент разрабатываемая ОС ориентирована на семейство процессоров ARM. Такой выбор платформы вполне очевиден.

ARM Ltd. (от Advanced RISC Machines) – британская корпорация, являющаяся одним из крупнейших разработчиков и распространителей лицензий современной архитектуры 32-разрядных RISC-процессоров, специально ориентированных для использования в портативных и мобильных устройствах (таких как мобильные телефоны, персональные органайзеры, пр.).

Технология ARM оказалась весьма успешной и в настоящее время является доминирующей микропроцессорной архитектурой для портативных цифровых устройств. ARM Ltd. утверждает, что общий объем микропроцессоров, произведенных по их лицензии, превышает 2,5 миллиарда штук. Семейство процессоров ARM используется примерно в 75 % встраиваемых решений.

Архитектура ARM представляет значительный интерес для разработчиков встраиваемых систем. ARM оперирует 15 рабочими регистрами и работает в 7 режимах – пользовательских (USR/SYS), супервизорном (SVC), режимах прерываний (IRQ/FIQ), режимах исключений (ABT/UND).

Ассемблер процессоров ARM во многом схож с Intel 80286 и Motorola 68020, но имеет несколько характерных отличий: двухуровневая система прерываний с банкируемыми регистрами (IRQ, FIQ); мощные режимы адресации; операции сдвига в связке с арифметическими, не снижающие производительность, и пр. К тому же каждая инструкция содержит 4-битный префикс условия, т. е. каждая инструкция может быть условной. Дальнейшее развитие ветки процессоров ARM добавило дополнительные преимущества – ускоренное выполнение Java-инструкций, сокращенный набор инструкций Thumb и пр [5].

3 РАЗРАБОТКА ОС

В основе тестовой платформы для разработки ОС использовался процессор AT91SAM7X256. Он обладает 256 Кб flash-памяти и 64 Кб RAM.

Отличительными характеристиками данного чипа является наличие USB и Ethernet, что может оказаться полезным в дальнейшей разработке. Чип содержит два последовательных порта ввода – вывода и порт отладки (синхронный последовательный), поддерживает интерфейс JTAG. ОС разрабатывалась таким образом, чтобы без значительных модификаций использоваться на любом чипе серии AT91SAM7. Планируется поддержка и других чипов.

Разработка ОС велась в среде ОС Linux, при этом использовались компиляторы семейства GCC (gnuparm 3.4.3). Используемый отладчик – GDB 6.4 + OpenOCD [3, 6–8].

Исходные тексты ОС доступны для более детального ознакомления в дискуссионной группе http://groups.google.com/group/avos_devel. На использование и распространение исходных текстов ОС никаких ограничений не накладывается.

Проект ОС состоит из следующих частей:

- архитектурно-зависимая часть (arch) включает в себя загрузчик, а также реализацию базовых функций, необходимых для работы ОС и зависящих от ресурсов и архитектуры платформы (функции ввода/вывода, обработка прерываний, функции для работы с таймером и т. д.);

- ядро (kernel) – основные функции ядра системы;

- заголовочные файлы (include) – содержат заголовки основных функций ядра системы;

- библиотечные функции (lib) – с целью минимизации размера ОС стандартная библиотека C была исключена, из-за этого некоторые функции пришлось реализовать в самой ОС;

- пользовательские ресурсы (usr) – здесь должны храниться пользовательские приложения. На данный момент поддерживаются пользовательские приложения и драйвера. Подробнее о написании таких приложений читайте далее;

- вспомогательные утилиты (utils) содержат конфигурационные файлы для программ, использующихся при разработке – отладчика, программатора, терминала, On-Chip отладчика. Также здесь содержатся скрипты оболочки, которые используются для динамической генерации некоторых файлов.

3.1 Загрузчик

Разработка ОС начиналась с загрузчика. Детальное описание он не заслуживает, так как является типичным для архитектуры ARM. Загрузчик написан на ассемблере с использованием препроцессора языка C. Микропрограмма начинается с обработки прерывания сброса (reset). Чип стартует в режиме SVC. В этом режиме устанавливается временный стек (от вершины адресного пространства) и вызывается функция `lowlevel_init()`, отвечающая за инициализацию аппаратных ресурсов. Затем загрузчик устанавливает стеки для каждого режима процессора. Все стеки, кроме пользовательского имеют фиксированный размер и занимают память от вершины адресного пространства. Далее инициализируются секции памяти (копируются из Flash-памяти в ОЗУ чипа) и происходит переход на функцию `main()`. С этого момента начинается работа ОС.

3.2 Архитектурно-зависимая часть

Некоторые функции являются архитектурно-зависимыми. Чтобы портировать ОС на другие платформы (или даже на другие чипы, отличные от используемого) необходимо будет изменить архитектурно-зависимую часть ОС.

Для функционирования ОС требуется реализация следующих функций:

- `lowlevel_init()` – должна выполнять действия по инициализации ресурсов системы (установка режимов портов ввода – вывода, установка обработчиков прерываний, если это необходимо и т. д.);

- `sys_reset()` – функция программной перезагрузки системы;

- `serial_init()`, `serial_tx(unsigned char)`, `serial_rx()` – в асинхронном режиме эта функция вызывается из обработчика прерывания соответствующего порта. Полученный байт функция должна передать в функцию ядра `serial_interrupt(unsigned char)`;

- `timer_init()`, `timer_handler()` – инициализация таймера и обработчик его прерываний. Последний должен вызывать `do_timer()`, отвечающий за многозадачность;

- `Ram.ld` – скрипт линковщика GNU ld.

В данной версии ОС, эти функции реализованы применительно к вышеупомянутому чипу AT91SAM7X256.

3.3 Контекст в архитектуре ARM. Переключение контекста

Чтобы перейти к многозадачности необходимо понимать, что является контекстом задачи. Обычно под контекстом подразумевают виртуальное адресное пространство, в котором выполняется программа, и состояние процессора в данный момент времени.

В данной ОС для архитектуры ARM контекстом считается состояние всех рабочих регистров, включая банкируемые, и текущее содержимое стека задачи. Кроме этого, нужно запоминать остальные данные о задаче в соответствующей структуре (приоритет, состояние, начальная вершина стека и т. д.) [9].

3.4 Алгоритмы многозадачности

Принято использовать следующие типы псевдопараллельной многозадачности:

Кооперативная многозадачность – вид многозадачности, при котором фоновые задачи выполняются только во время простоя основного процесса и только в том случае, если на это получено разрешение основного процесса. Этот вид удобно использовать для тесно связанного между собой небольшого количества задач. При кооперативной многозадачности время работы расходуется наиболее эффективно.

Вытесняющая многозадачность – вид многозадачности, при котором операционная система сама передает управление от одной выполняемой программы другой. Распределение процессорного времени осуществляется планировщиком процессов. Этот вид многозадачности обеспечивает более быстрый отклик на действия пользователя. В то же время, квота, выделяемая на каждую задачу, часто может оказаться избыточной.

Планировщик рассматриваемой ОС использует вытесняющую многозадачность.

В ранних версиях ОС Linux использовался довольно простой, но эффективный алгоритм работы планировщика:

1. Каждой задаче задается значение приоритета. Например, есть три задачи с приоритетами $p_1 = 2$, $p_2 = 5$, $p_3 = 8$. Также вводится понятие суммарного приоритета $q_1 = q_2 = q_3 = 0$.

2. С каждым переключением задачи накапливается приоритет всех задач.

3. Задача с максимальным приоритетом становится активной, и ее суммарный приоритет устанавливается равным нулю.

Рассмотрим несколько итераций такого алгоритма:

- | | | | |
|---------------|------------|------------|--|
| 1. $q_1 = 0$ | $q_2 = 0$ | $q_3 = 0$ | Нет активных задач,
$p_1 = 2, p_2 = 5, p_3 = 8$ |
| 2. $q_1 = 2$ | $q_2 = 5$ | $q_3 = 8$ | Активная задача 3, $q_3 = 0$ |
| 3. $q_1 = 4$ | $q_2 = 10$ | $q_3 = 8$ | Активная задача 2,
$q_2 = 0$ |
| 4. $q_1 = 6$ | $q_2 = 5$ | $q_3 = 16$ | Активная задача 3,
$q_3 = 0$ |
| 5. $q_1 = 8$ | $q_2 = 10$ | $q_3 = 8$ | Активная задача 2,
$q_2 = 0$ |
| 6. $q_1 = 10$ | $q_2 = 5$ | $q_3 = 16$ | Активная задача 3,
$q_3 = 0$ |
| 7. $q_1 = 12$ | $q_2 = 10$ | $q_3 = 8$ | Активная задача 1,
$q_1 = 0$ |
| 8. $q_1 = 2$ | $q_2 = 15$ | $q_3 = 16$ | Активная задача 3,
$q_3 = 0$ |

Как видно из рассмотренного примера, квота времени, выделенная на задачу не прямо пропорциональна приоритету, но сохраняется зависимость $q_1 < q_2 < q_3$ при $p_1 < p_2 < p_3$, где q_i – квота времени, выделенная на задачу i .

Этот алгоритм реализован и в разрабатываемой ОС в связи со своей простотой и наглядностью. Существует возможность переключения задачи с помощью системного вызова (реализована кооперативная многозадачность наряду с вытесняющей). В ближайшей перспективе планируется опциональное включение различных алгоритмов переключения задач и реализация примитивов синхронизации [9–11].

3.5 Системные вызовы

Некоторые функции ОС не должны прерываться – это может повлиять на их выполнение (например, если во время записи данных в порт задача будет прервана и порт будет захвачен конкурирующей задачей, то данные будут отправлены некорректно).

В ОС реализованы следующие системные вызовы:

- fork () – создание дочернего процесса;
- exit (int status) – завершение текущего процесса;
- sleep (int sec) – приостановка текущего процесса на указанное время;
- setpriority (int new_priority) – установка нового приоритета текущего процесса;
- reset () – программная перезагрузка системы;
- wait (pid_t pid) – ожидание, пока какой-либо процесс не получит сигнал;
- waitpid (pid_t pid, int *status, int options) – ожидание, пока указанный процесс не получит сигнал;
- execve (const char *filename, const char *argv [], const char *envp []) – выполнение указанной программы (замещает текущий процесс);
- malloc (size_t len) – выделение блока памяти;
- free (void *ptr) – освобождение блока памяти.

Большая часть этих функций относится к управлению задачами, execve () обеспечивает интерфейс для запуска пользовательских программ, а работа последних двух функций касается динамической памяти.

3.6 Менеджер памяти

Стек задачи выделяется при ее создании. В стеке хранятся все локальные переменные процедур, а это часто приводит к избыточному использованию стека и преждевременному его переполнению. Чтобы избежать этого, необходимо использовать динамическое выделение памяти. Поскольку встраиваемые устройства ограничены в объеме ОЗУ, то второй способ является более предпочтительным.

Менеджер памяти выполняет следующие действия: обнаруживает свободный участок памяти, помечает его как занятый, возвращает адрес его начала. При освобождении памяти помечает снова блок как незанятый.

В ОС Linux используется следующий принцип выделения памяти. В оперативной памяти резервируется место под блоки различного размера. При выделении памяти ОС находит минимальный блок, подходящий запрошенному размеру. На практике этот способ весьма эффективен, хотя некоторая часть пространства памяти остается незанятой из-за избыточности блоков памяти (каждый блок несколько превышает запрошенный размер).

В разрабатываемой ОС используется простейшее последовательное выделение памяти. Т.е. блоки изначально не выделены, а создаются при необходимости в первом найденном свободном участке памяти. При

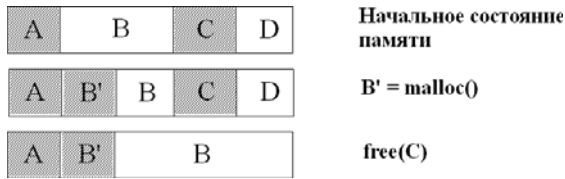


Рисунок 1 – Пример работы менеджера памяти

этом каждый блок указывает на начало следующего (работа алгоритма показана на рис. 1).

Недостатком этого алгоритма является чрезмерная фрагментация памяти, поэтому использование динамической памяти в данной ОС оправдано только для больших блоков.

3.7 Написание приложений для ОС

Разрабатываемая ОС поддерживает такие виды пользовательских задач как приложения и драйвера. В данный момент различие между ними заключается только в интерфейсах обращения к ним (у приложений только одна точка входа – main (), а у драйверов – функции загрузки, выгрузки, реакция на открытие/закрытие устройства и обращение к нему). В дальнейшем планируется выполнение драйверов на системном уровне, в то время как приложения выполняются на пользовательском.

Все программы и драйверы хранятся в папках usr/app и usr/drv соответственно. При сборке проекта происходит этап «bootstrap». На этом этапе специальный скрипт анализирует, какие программы и драйвера должны быть включены в сборку, и изменяет соответствующим образом заголовочные файлы applist.h и drvlist.h. На содержимом этих файлов основан запуск пользовательских программ и работа с драйверами.

Поскольку ОС находится на стадии разработки, то технология написания приложений и драйверов для данной ОС не является достаточно удобной для конечного пользователя. В дальнейшем возможны существенные изменения в этой области.

ВЫВОДЫ

Подводя итоги проделанной работы, видим, что разработана основа операционной системы и реализованы базовые функции, достаточные для реализации простых встраиваемых систем.

Удалось запустить под управлением ОС простую текстовую оболочку, позволяющую выполнять приложения и проводить элементарный мониторинг ресурсов

системы. ОС обладает достаточной стабильностью. Минимальная сборка ОС занимает менее 3 Кб, что оставляет достаточно памяти для реализации прикладного программного обеспечения.

В то же время в ОС следует реализовать поддержку синхронизации (семафоров, мьютексов, критических секций, очередей сообщений, событий), файловой системы, максимально приблизить ОС к стандартам POSIX, портировать ОС на другие платформы. При этом одним из основных критериев должна оставаться гибкость системы (возможность легко включать/исключать ее компоненты) и простота в использовании.

ПЕРЕЧЕНЬ ССЫЛОК

1. Jim Turley. Operating systems on the rise // Embedded Systems Design. – 2006. – No. 6 – P. 21.
2. Таненбаум А. П. Операционные системы: Разработка и реализация. – СПб.: Питер, 2007 – 704 с.
3. Лав Р. Разработка ядра Linux – К.: Диалектика/Вильямс, 2006 – 448 с.
4. eCos vs. uClinux: Which is best for your embedded target? (Электронный ресурс). – Rob Wehrli. – Электрон. дан. – New York, 2002. – Режим доступа: <http://www.linuxdevices.com/articles/AT3393503683.html> свободный. – Загл. с экрана.
5. Программная платформа для телекоммуникаций: открытые системы (Электронный ресурс). – Александр Трофимов. – Электрон. дан. – М., 2008. – Режим доступа: <http://www.osp.ru/os/2008/02/4924564> свободный. – Загл. с экрана.
6. Keith E. Curtis Embedded Multitasking – NEWNES, 2006.– 416 p.
7. Статьи по электронным компонентам: компоненты и технологии (Электронный ресурс). – Электрон. журнал. – М., 2008. – Режим доступа: <http://com-pitech.ru/> свободный. – Загл. с экрана.
8. ARM System Developer's Guide: Designing and Optimizing System Software (The Morgan Kaufmann Series in Computer Architecture and Design) – Morgan Kaufmann, 2004. – 689 p.
9. Michael Barr, Anthony Massa. Programming Embedded Systems: With C and GNU Development Tools, 2nd Edition. – O'Reilly Media, 2006. – 301 p.
10. David Seal. ARM Architecture Reference Manual, Second Edition – Addison-Wesley, 1996. – 807 p.
11. L. Sha, R. Rajkumar, J. Lehoczky. Priority Inheritance Protocols: An Approach to Real-Time Synchronization // IEEE Transactions on Computers. – 1990. – Vol. 39, No.9. – P. 61–64.

Надійшла 23.05.2008

Проведено аналіз ринку сучасних операційних систем для вбудованих пристроїв. Розроблено ядро багатозадачної вбудованої операційної системи, що призначена для керування вбудованим пристроєм. Описано основні принципи функціонування вбудованої операційної системи та моделі розробки вбудованих додатків та драйверів.

Modern embedded operating system market analysis has been performed. Multitasking operating system kernel for embedded device control has been developed. Major embedded operating system functioning concepts has been described, embedded applications and driver programming model has been suggested.

А. В. Неласая, Г. Л. Козина, Н. А. Молдовян

ПРОТОКОЛЫ КОЛЛЕКТИВНОЙ ЦИФРОВОЙ ПОДПИСИ НА ЭЛЛИПТИЧЕСКИХ И ГИПЕРЭЛЛИПТИЧЕСКИХ КРИВЫХ

В статье представлены новые протоколы коллективной цифровой подписи на эллиптических и гиперэллиптических кривых. Вычислительные схемы проиллюстрированы примерами.

ВВЕДЕНИЕ

Развитие технологий электронного документооборота требует новых механизмов обеспечения юридической силы коллективных электронных документов. В частности, при разработке коллективных проектов важной проблемой является использование протоколов [1–4], обеспечивающих реализацию коллективной электронной цифровой подписи.

Известные в настоящее время протоколы электронной цифровой подписи позволяют осуществить реализацию кратной подписи (директор, главный бухгалтер, ведущий инженер и т. д.), но при этом, в силу последовательной реализации кратной подписи, возникают следующие проблемы. При подписании электронного документа важна последовательность формирования подписей каждого из участников, при проверке подписей также важна проверяющая последовательность проверки подписей участников. Кроме этого, размер подписи увеличивается пропорционально числу участников, подписавших электронный документ.

Для устранения указанных недостатков были предложены новые протоколы [2–4] формирования и проверки подлинности коллективной электронной цифровой подписи. В этих протоколах используется общий (коллективный) открытый ключ, который формируется на основе индивидуальных открытых ключей группы пользователей. Применяемые на практике системы электронной цифровой подписи предоставляют возможность использования (доступности через Internet) стандартных справочников открытых ключей и/или типовых сертификатов открытых ключей, что благоприятствует практическому применению нового подхода генерации коллективной электронной цифровой подписи.

Важен также вопрос минимизации размера коллективной электронной цифровой подписи при необходимости ее записи в виде штрих-кода на бумажных носителях, например, в методах защиты от подделки документов с помощью электронной цифровой подписи

[4]. В данном аспекте представляет интерес изучение вопроса о возможности реализации новых протоколов коллективной электронной цифровой подписи как с использованием процедур проверки электронной цифровой подписи, специфицируемых стандартами подписи, так и новых протоколов, позволяющих уменьшить размер подписи.

В настоящей работе представлены новые протоколы коллективной цифровой подписи на эллиптических и гиперэллиптических кривых на основе стандарта электронной цифровой подписи ДСТУ 4145-2002 и предложенного в [6] протокола электронной цифровой подписи с предвычислениями ЕСРР. Вычислительные схемы протоколов проиллюстрированы примерами.

1 ГИПЕРЭЛЛИПТИЧЕСКИЕ КРИВЫЕ

В качестве источника абелевой группы для протокола коллективной подписи на основе ДСТУ-4145, предложенного в [4], можно взять группу дивизоров гиперэллиптической кривой. Основное преимущество использования гиперэллиптических кривых состоит в том, что размер основного поля, над которым определена кривая, уменьшается пропорционально роду кривой без потери стойкости, хотя сама формула группового сложения выглядит более громоздко.

Пусть F – конечное поле и пусть \bar{F} – алгебраическое замыкание поля F . Гиперэллиптическая кривая C рода $g \geq 1$ над F представляет собой [7] набор решений $(x, y) \in F \times F$ уравнения

$$C: y^2 + h(x)y = f(x), \quad (1)$$

где $h(x) \in F[x]$ – полином степени не более g , $f(x) \in F[x]$ – нормированный полином степени $2g + 1$ и не существует решений $(x, y) \in \bar{F} \times \bar{F}$, которые бы одновременно удовлетворяли уравнению (1) и уравнениям $2y + h(x) = 0$ и $h'(x)y - f'(x) = 0$. Считаем, что бесконечно удаленная точка ∞ также принадлежит кривой.

Согласно [7], в качестве групповой структуры в случае гиперэллиптических кривых рассматривается якобиан кривой C . Каждый элемент якобиана – это класс эквивалентности дивизоров, который может быть

представлен унікально приведеним дивизором в формі пари поліномів в формі Мамфорда. На якобіані визначені групові операції додавання і дублювання дивизорів.

Згідно [8] порядок якобіана гіперелліптичної кривої обмежений інтервалом Хассе – Вейля

$$|(\sqrt{q}-1)^{2g}| \leq \#J/F_q \leq [(\sqrt{q}+1)^{2g}],$$

де q – характеристика поля, над яким визначена крива, g – род кривої. Будемо вважати, що порядок якобіана

$$\#J/F_q \approx q^g.$$

Більшість криптографічних застосувань базуються на еліптичних або гіперелліптичних кривих з довжиною ключа не менше 160 біт, тобто з порядком групи не менше 2^{160} . Отже, для криптосистем на гіперелліптичних кривих над полем F_q повинно виконуватися як мінімум

$$g \cdot \log_2 q \approx 160.$$

В частині, для кривої роду 2 необхідно вибрати основне поле F_q з $|F_q| \approx 2^{80}$, з довжиною операндів 80 біт. Для кривої роду 3 потужність основного поля $|F_q| \approx 2^{54}$, для кривої роду 4 $|F_q| \approx 2^{40}$.

Оскільки елементи підпису належать основному полю, над яким визначена крива, в разі гіперелліптичних кривих розмір кінцевої колективної підпису зменшується пропорційно роду кривої. Так, при використанні гіперелліптичної кривої другого роду розмір колективної підпису буде приблизно в два рази менше, ніж при використанні еліптичної кривої, яка має аналогічний рівень криптостійкості. Відповідно, при використанні кривої третього роду розмір колективної підпису зменшиться приблизно в три рази і т. д.

В криптографічних цілях використовуються гіперелліптичні криві роду 2 і 3. Криві вищого роду не є стійкими.

2 ПРОТОКОЛ КОЛЕКТИВНОЇ ПОДПИСИ НА ОСНОВЕ СТАНДАРТА ДСТУ 4145-2002 НА ГІПЕРЕЛЛІПТИЧЕСКИХ КРИВЫХ

Введемо позначення:

D – базовий дивизор гіперелліптичної кривої;

l – кількість користувачів;

n – порядок циклічної підгрупи якобіана гіперелліптичної кривої;

d_i – секретний ключ i -го користувача;

h – хеш-образ повідомлення.

$\psi(R)$ – функція перетворення дивизора в елемент основного поля. Авторами пропонується наступне перетворення: коефіцієнти першого полінома дивизора R представимо в формі числа в системі числення з основою, рівною модулю основного поля, над яким визначена крива (в разі простого поля). А потім переведемо це представлення в десятичну систему числення.

Генерація відкритого колективного ключа.

1. Кожен i -й користувач ($i = 1 \dots l$) формує відкритий ключ виду

$$Q_i = -d_i D.$$

2. Колективний відкритий ключ обчислюється як сума відкритих ключів групи з l користувачів

$$Q = \sum_{i=1}^l Q_i = \sum_{i=1}^l -d_i D.$$

Формування колективної підпису.

1. Кожен користувач обчислює дивизор R_i наступним чином:

а) вибирає випадковий параметр k_i , $1 < k_i < n$;

б) обчислює $R_i = k_i D$.

2. По представленим користувачами дивизорам R_i обчислюється загальний дивизор

$$R = \sum_{i=1}^l R_i.$$

3. Обчислюється значення функції $\psi(R)$.

4. Перша частина підпису визначається формулою $r = h\psi(R) \bmod n$.

5. Кожен користувач обчислює свій параметр s_i

$$s_i = (k_i + d_i r) \bmod n.$$

6. Друга частина підпису визначається формулою $s = \sum_{i=1}^l s_i \bmod n$.

Колективним підписом є пара чисел (r, s) .

Перевірка колективної підпису.

1. Перевіраючий обчислює хеш-образ h' загального повідомлення.

2. Використовуючи відкритий колективний ключ Q , обчислює дивизор $R' = sD + rQ$ і

3. значення $v = h'\psi(R')$.

4. Якщо $v = r$, то підпис вважається дійсним.

Пример 1. Проиллюстрируем протокол коллективной подписи, используя гиперэллиптическую кривую рода 2 над полем F_7 :

$$y^2 = x^5 + 2x^2 + x + 3 \pmod{7}.$$

Как показано в [9], якобиан этой кривой содержит 34 дивизора. Дивизор $D = \langle x + 4, 1 \rangle$ формирует подгруппу порядка $n = 17$.

Генерация открытого коллективного ключа.

Пусть число пользователей $l = 3$ и их секретные ключи соответственно равны:

$$d_1 = 5, \quad d_2 = 7, \quad d_3 = 11.$$

Тогда открытыми ключами пользователей являются:

$$\begin{aligned} Q_1 &= -5D = -\langle x^2 + 5x + 2, 2x \rangle = \langle x^2 + 5x + 2, 5x \rangle, \\ Q_2 &= -7D = -\langle x^2 + 5x + 5, 5 \rangle = \langle x^2 + 5x + 5, 2 \rangle, \\ Q_3 &= -11D = -\langle x^2 + 4x + 6, x + 4 \rangle = \\ &= \langle x^2 + 4x + 6, 6x + 3 \rangle. \end{aligned}$$

Общий открытый ключ группы пользователей равен:

$$Q = Q_1 + Q_2 + Q_3 = \langle x^2 + 4x + 6, x + 4 \rangle.$$

Формирование коллективной подписи.

Каждый пользователь генерирует случайный параметр k_i :

$$k_1 = 5, \quad k_2 = 9, \quad k_3 = 12$$

и вычисляет дивизор $R_i = k_i D$:

$$\begin{aligned} R_1 &= 5\langle x + 4, 1 \rangle = \langle x^2 + 5x + 2, 2x \rangle, \\ R_2 &= 9\langle x + 4, 1 \rangle = \langle x^2 + 3x + 5, 6x + 5 \rangle, \\ R_3 &= 12\langle x + 4, 1 \rangle = \langle x^2 + 5x + 2, 5x \rangle. \end{aligned}$$

Вычисленные дивизоры R_i предоставляются для вычисления общего дивизора

$$R = \sum_{i=1}^3 R_i = \langle x^2 + 3x + 5, 6x + 5 \rangle.$$

По дивизору R вычисляется функция $\psi(R)$:

$$\psi(R) = 135_7 \pmod{17} = (7^2 + 3 \cdot 7 + 5) \pmod{17} = 7.$$

Пусть хэш-образ h общего для группы пользователей сообщения равен 15: $h = 15$. Тогда первая часть коллективной подписи r вычисляется по формуле:

$$r = h\psi(R) = 15 \cdot 7 \pmod{17} = 3.$$

Далее каждый пользователь по своему секретному ключу d_i , значению k_i и общему значению r вычисляет свою долю второй части подписи:

$$\begin{aligned} s &= 5 + 5 \cdot 3 \pmod{17} = 3, \quad s = 9 + 7 \cdot 3 \pmod{17} = 13, \\ s &= 12 + 11 \cdot 3 \pmod{17} = 11. \end{aligned}$$

Вторая часть коллективной подписи равна:

$$s = \sum_{i=1}^3 s_i = (3 + 13 + 11) \pmod{17} = 10.$$

Таким образом, коллективная подпись группы из трех пользователей под общим сообщением есть $(r, s) = (3, 10)$.

Проверка коллективной подписи.

Проверяющий вычисляет хэш-образ $h' = 15$ общего сообщения.

Используя открытый коллективный ключ Q , вычисляет дивизор

$$\begin{aligned} R' &= sD + rQ = 10D + 3Q = \\ &= 10\langle x + 4, 1 \rangle + 3\langle x^2 + 4x + 6, x + 4 \rangle = \\ &= \langle x^2 + 5x + 5, 2 \rangle + \langle x + 4, 6 \rangle = \\ &= \langle x^2 + 3x + 5, 6x + 5 \rangle \end{aligned}$$

и находит значение

$$\begin{aligned} \psi(R') &= \psi(\langle x^2 + 3x + 5, 6x + 5 \rangle) = \\ &= 135_7 \pmod{17} = 7 + 3 \cdot 7 + 5 = 7. \end{aligned}$$

Поскольку

$$v = h'\psi(R') = 15 \cdot 7 \pmod{17} = 3$$

совпадает с r , то подпись признается подлинной.

3 ПРОТОКОЛ КОЛЛЕКТИВНОЙ ПОДПИСИ НА ОСНОВЕ ПРОТОКОЛА ЕСРР НА ЭЛЛИПТИЧЕСКИХ КРИВЫХ

Протокол электронной цифровой подписи с предвычислениями ЕСРР [6] был предложен с целью уменьшить трудоемкость операции верификации подписи в корпоративной сети за счет умножения только на базовую точку, которое можно выполнить с предвычис-

лениями. Модифицируем этот протокол для реализации коллективной подписи.

Введем обозначения:

P – базовая точка эллиптической кривой;

l – количество пользователей;

n – порядок циклической подгруппы точек эллиптической кривой;

d_i – секретный ключ i -го пользователя;

h – хэш-образ сообщения;

$\pi(R) = X_R \bmod n$ – выделение x -координаты точки

$R = (X_R, Y_R)$ эллиптической кривой.

Генерация открытого коллективного ключа.

1. Каждый i -й пользователь ($i = 1 \dots l$) формирует открытый ключ вида

$$Q_i = -d_i P.$$

2. Коллективный открытый ключ вычисляется как сумма открытых ключей группы из l пользователей

$$Q = \sum_{i=1}^l Q_i = \sum_{i=1}^l -d_i P.$$

Формирование коллективной подписи.

1. Каждый i -й пользователь ($i = 1 \dots l$) рассчитывает точку R_i следующим образом:

а) выбирает случайный параметр k_i , $1 < k_i < n$;

б) вычисляет значение $t_i = \frac{k_i}{h} \bmod n$;

в) и точку $R_i = t_i P$.

2. По представленным пользователями точкам R_i вычисляется общая точка

$$R = \sum_{i=1}^l R_i = (X_R, Y_R)$$

3. и значение $w = \pi(R) = X_R \bmod n$.

4. Формируется точка $wR = (x, y)$ и

5. первая часть коллективной подписи

$$r = \pi(x, y) = x \bmod n.$$

6. Каждый пользователь вычисляет свой параметр s_i

$$s_i = (wk_i + hd_i) \bmod n$$

7. и предоставляет его для вычисления второй части коллективной подписи

$$s = \sum_{i=1}^l s_i.$$

Коллективной подписью является пара чисел – (r, s) .

Проверка коллективной подписи.

1. Проверяющий вычисляет хэш-образ h' общего сообщения

2. и значение $t = \frac{s}{h'} \bmod n$.

3. Используя открытый коллективный ключ Q , формирует точку $tP + Q = (x, y)$

4. и вычисляет значение $v = \pi(x, y) = x \bmod n$.

5. Если $v = r$, то подпись признается подлинной.

Обоснование корректности представленного протокола

Поскольку при формировании подписи значение r

$$\text{определяется формулой } r = \pi(wR) = \pi\left(w \sum_{i=1}^l \frac{k_i}{h} P\right),$$

а при проверке подписи проверочное выражение $tP + Q$ дает точку wR

$$\begin{aligned} tP + Q &= \frac{s}{h} P - \sum_{i=1}^l d_i P = \frac{\sum_{i=1}^l (wk_i + hd_i)}{h} P - \sum_{i=1}^l d_i P = \\ &= w \sum_{i=1}^l \frac{k_i}{h} P + \sum_{i=1}^l d_i P - \sum_{i=1}^l d_i P = w \sum_{i=1}^l \frac{k_i}{h} P = wR, \end{aligned}$$

в итоге имеем:

$$v = \pi(tP + Q) = \pi(wR),$$

что соответствует r при формировании подписи.

Пример 2. Проиллюстрируем протокол коллективной подписи, используя эллиптическую кривую над полем F_{79} :

$$y^2 = x^3 + x + 1 \pmod{79}.$$

Базовая точка $P(5,62)$ этой эллиптической кривой образует циклическую подгруппу порядка 43.

Генерация открытого коллективного ключа.

Пусть число пользователей $l = 3$ и их секретные ключи соответственно равны:

$$d_1 = 11, \quad d_2 = 26, \quad d_3 = 38.$$

Тогда открытыми ключами пользователей являются:

$$Q_1 = (30, 48), \quad Q_2 = (15, 28), \quad Q_3 = (32, 4).$$

Общий открытый ключ группы пользователей равен:

$$Q = Q_1 + Q_2 + Q_3 = (30, 31).$$

Формирование коллективной подписи.

Каждый пользователь генерирует случайный параметр k_i :

$$k_1 = 5, \quad k_2 = 17, \quad k_3 = 23.$$

Следуя протоколу, каждый пользователь рассчитывает значение t_i :

$$t_1 = \frac{5}{37} \bmod 43 = 35, \quad t_2 = \frac{17}{37} \bmod 43 = 33, \\ t_3 = \frac{23}{37} \bmod 43 = 32$$

и находит точку R_i :

$$R_1 = 35P = (29, 18), \quad R_2 = 33P = (16, 20), \\ R_3 = 32P = (30, 48).$$

Вычисленные точки R_i предоставляются для вычисления общей точки

$$R = R_1 + R_2 + R_3 = (34, 32).$$

По точке R вычисляется функция $\pi(R)$

$$w = \pi(R) = 34 \bmod 43 = 34.$$

С использованием полученного значения определяется точка

$$wR = 34R = (31, 35)$$

и первая часть коллективной подписи r

$$r = \pi(31, 35) = 31 \bmod 43 = 31.$$

Пусть хэш-образ h общего для группы пользователей сообщения равен 37: $h = 37$.

Далее каждый пользователь по своему секретному ключу d_i , значению k_i и общим значениям w и h вычисляет свою долю второй части подписи:

$$s_1 = (43*5 + 37*11) \bmod 43 = 18, \\ s_2 = (34*17 + 37*26) \bmod 43 = 35, \\ s_3 = (34*23 + 37*38) \bmod 43 = 38.$$

Вторая часть коллективной подписи равна:

$$s = (s_1 + s_2 + s_3) \bmod n = (18 + 35 + 38) \bmod 43 = 5.$$

Таким образом, коллективная подпись группы из трех пользователей под общим сообщением есть $(r, s) = (31, 5)$.

Проверка коллективной подписи.

Проверяющий вычисляет хэш-образ $h' = 37$ общего сообщения.

Используя хэш-образ сообщения и открытый коллективный ключ Q , вычисляет значение

$$t = \frac{s}{h} \bmod n = \frac{5}{37} \bmod 43 = 35$$

и точку

$$tP + Q = 35*(5, 62) + (30, 31) = \\ = (29, 18) + (30, 31) = (31, 35).$$

Поскольку

$$v = 31 \bmod 43 = 31,$$

то есть $v = r$, то подпись признается подлинной.

4 ПРОТОКОЛ КОЛЛЕКТИВНОЙ ПОДПИСИ НА ОСНОВЕ ПРОТОКОЛА ЕСРР НА ГИПЕРЭЛЛИПТИЧЕСКИХ КРИВЫХ

В качестве источника абелевой группы для предложенного протокола можно также взять группу дивизоров гиперэллиптической кривой, как это было сделано выше для ДСТУ-4145. В этом случае базовой точке, открытым ключам пользователей, и промежуточной точке R соответствуют дивизоры гиперэллиптической кривой.

Генерация открытого коллективного ключа.

1. Каждый i -й пользователь ($i = 1 \dots l$) формирует открытый ключ вида

$$Q_i = -d_i D.$$

2. Коллективный открытый ключ вычисляется как сумма открытых ключей группы из l пользователей

$$Q = \sum_{i=1}^l Q_i = \sum_{i=1}^l -d_i D.$$

Формирование коллективной подписи.

1. Каждый пользователь рассчитывает дивизор R_i следующим образом:

а) выбирает случайный параметр k_i , $1 < k_i < n$;

б) вычисляет $t_i = \frac{k_i}{h} \bmod n$

в) и $R_i = t_i D$.

2. По представленным пользователями дивизорам R_i вычисляется общий дивизор

$$R = \sum_{i=1}^l R_i$$

3. и значение $w = \psi(R)$.
4. Формируется дивизор wR и
5. первая часть коллективной подписи $r = \psi(wR)$.
6. Каждый пользователь вычисляет свой параметр s_i

$$s_i = (wk_i + hd_i) \bmod n$$

7. и предоставляет его для вычисления второй части коллективной подписи

$$s = \sum_{i=1}^l s_i.$$

Коллективной подписью является пара чисел – (r, s) .

Проверка коллективной подписи.

1. Проверяющий вычисляет хэш-образ h' общего сообщения
2. и значение $t = \frac{s}{h'} \bmod n$.
3. Используя открытый коллективный ключ Q , формирует дивизор $tD + Q$
4. и вычисляет значение $v = \psi(tD + Q)$.
5. Если $v = r$, то подпись признается подлинной.

Пример 3. Проиллюстрируем протокол коллективной подписи, используя гиперэллиптическую кривую, секретные ключи, открытый коллективный ключ, хэш-образ сообщения и случайные параметры k_i из примера 1.

Формирование коллективной подписи.

Следуя протоколу, каждый пользователь рассчитывает значения t_i :

$$t_1 = \frac{5}{15} \bmod 17 = 6, \quad t_2 = \frac{9}{15} \bmod 17 = 4, \\ t = \frac{12}{15} \bmod 17 = 11$$

и находит дивизор R_i :

$$R_1 = 6D = \langle x^2 + 4x + 6, 6x + 3 \rangle, \\ R_2 = 4D = \langle x^2 + 5x + 3, 5x \rangle, \\ R_3 = 11D = \langle x^2 + 4x + 6, x + 4 \rangle.$$

Вычисленные дивизоры R_i предоставляются для вычисления общего дивизора

$$R = R_1 + R_2 + R_3 = \langle x^5 + 5x + 3, 5x \rangle.$$

По дивизору R вычисляется функция $\psi(R)$:

$$w = \psi(R) = 153_7 \bmod 17 = 7^2 + 5*7 + 3 \bmod 17 = 2.$$

С использованием полученного значения определяется дивизор

$$wR = 2R = \langle x^2 + 3x + 5, x + 2 \rangle$$

и первая часть коллективной подписи r

$$r = \pi(x^2 + 3x + 5, x + 2) = 135_7 \bmod 17 = \\ = (7^2 + 3*7 + 5) \bmod 17 = 7.$$

Далее каждый пользователь по своему секретному ключу d_i , значению k_i и общим значениям w и h вычисляет свою долю второй части подписи:

$$s_1 = (2*5 + 15*5) \bmod 17 = 0, \\ s_2 = (2*9 + 15*7) \bmod 17 = 4, \\ s_3 = (2*12 + 15*11) \bmod 17 = 2.$$

Вторая часть коллективной подписи равна:

$$s = (s_1 + s_2 + s_3) \bmod n = (0 + 4 + 2) \bmod 17 = 6.$$

Таким образом, коллективная подпись группы из трех пользователей под общим сообщением есть $(r, s) = (7, 6)$.

Проверка коллективной подписи.

Проверяющий вычисляет хэш-образ $h' = 15$ общего сообщения.

Используя хэш-образ сообщения и открытый коллективный ключ Q , вычисляет

$$t = \frac{s}{h'} \bmod n = \frac{6}{15} \bmod 17 = 14$$

и

$$tD + Q = 14 * \langle x + 4, 1 \rangle + \langle x^2 + 4x + 6, x + 4 \rangle = \\ = \langle x^2 + x + 6, 6x + 1 \rangle + \langle x^2 + 4x + 6, x + 4 \rangle = \\ = \langle x^2 + 3x + 5, x + 2 \rangle.$$

Поскольку

$$v = \psi(\langle x^2 + 3x + 5 \rangle) = 135_7 \bmod 17 = \\ = (7^2 + 3*7 + 5) \bmod 17 = 7,$$

то есть $v = r$, то подпись признается подлинной.

Таким образом, действие предложенных протоколов наглядно показано на примерах с небольшими размерами параметров. На практике для обеспечения достаточной стойкости порядок группы, для которой определен протокол, должен превышать 2^{160} .

ЗАКЛЮЧЕНИЕ

Представленные в статье протоколы основаны на предложенном недавно способе формирования и проверки подлинности коллективной цифровой подписи,

базирующейся на понятии общего (коллективного) открытого ключа. Они обладают тем качеством, что размер подписи не увеличивается пропорционально числу подписавших участников. В дальнейшем необходимо рассмотреть вопросы стойкости предложенных схем к различным типам атак.

ПЕРЕЧЕНЬ ССЫЛОК

1. Min-Shiang Hwang, Cheng-Chi Le. Research issues and challenges for multiple digital signature // Int. J. of Network Security. – 2005. – Vol. 1, No 1. – P. 1–7.
2. Молдовян Н. А., Молдовян П. А. Новые протоколы слепой подписи // Безопасность информационных технологий. – М.: МИФИ. – 2007. – № 3. – С. 17–21.
3. Артамонов А. В., Маховенко Е. Б. Применение алгоритма Шнора в протоколе коллективной подписи // Материалы XIV Всероссийской научной конференции «Проблемы информационной безопасности в системе высшей школы». – 2007. – С. 17–18.
4. Гортинская Л. В., Молдовян Н. А., Козина Г. Л. Реализация протоколов коллективной подписи на основе стандартов ГОСТ 34.310–95 и ДСТУ 4145-2002 // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні. – Киев: НТУУ «КПІ». – 2008. – № 1. – С. 21–25.
5. Карякин Ю. Д. Технология «AXIS-2000» защиты материальных объектов от подделки // Управление защитой информации. – Минск: Институт технической кибернетики АН Белоруссии. – 1997. – Т. 1, № 2. – С. 90–97.
6. Anna Nelasa, Victor Dolgov, Anatolij Pogorily. Digital Signature Protocol for corporate network // Proceedings of International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2008). – Lviv-Slavsko (Ukraine). – 2008. – Pp. 396–397.
7. A. Menezes, Y. Wu, R. Zuccherato. An Elementary Introduction to Hyperelliptic Curves – Springer-Verlag, Berlin (Germany), 1998. – 31 p.
8. D. G. Cantor. Computing in Jacobian of a Hyperelliptic Curve // In Mathematics of Computation, volume 48 (177). – January 1987. – P. 95–101.
9. Неласая А. В. Протокол цифровой подписи на гиперэллиптических кривых // Радіоелектроніка. Інформатика. Управління. – Запоріжжя: ЗНТУ. – 2006. – № 1. – С. 113–118.

Після доробки 17.03.2008

В статті пропонуються нові протоколи колективного цифрового підпису на еліптичних та гіпереліптичних кривих. Обчислювальні схеми проілюстровані на прикладах.

New collective digital signature protocols on elliptic and hyperelliptic curves are proposed. Computational algorithms are illustrated with numerical examples.

УДК 65.012.8(043)

В. И. Слепцов, Л. М. Карпуков

ПРАВОВАЯ ПОДГОТОВКА КАДРОВ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Рассматриваются некоторые вопросы правовой подготовки специалистов по направлениям, входящим в отрасль знаний «Информационная безопасность», с учетом специфики их будущей профессиональной деятельности, учебных задач, законодательства Украины и опыта работы кафедры защиты информации Запорожского национального технического университета.

ВВЕДЕНИЕ

Уровень информационной безопасности, как в государственных структурах, так и в сфере хозяйственной деятельности, во многом определяется качеством подготовки работающих там специалистов, получивших образование по различным направлениям, объединяемым областью знаний «Информационная безопасность». Эффективность их усилий, направленных на защиту интересов субъектов информационных отношений зависит, прежде всего, от умения выявлять и оценивать угрозы, определять состояние защищенности информации, обоснованно выбирать способы ее защи-

ты от совокупности реальных угроз, разрабатывать и внедрять системы защиты на основе требований законодательства Украины.

Базовые теоретические знания с получением необходимых практических навыков студенты, обучающиеся в ЗНТУ по специальностям «Защита информации в компьютерных системах и сетях» и «Системы защиты от несанкционированного доступа», приобретают при изучении дисциплин: «Основы информационной безопасности», «Физико-технические методы защиты информации», «Организационно-техническое обеспечение систем защиты информации», «Криптографические методы защиты информации» и др. Не возникает особых проблем при определении объема лекционного материала, методологии преподавания, если речь идет о традиционных вопросах, связанных с изучением технических каналов утечки информации и методов их закрытия, программно-аппаратных способов получения несанкционированного доступа к защищаемой информации, обрабатываемой в информационно-теле-

коммуникационных системах, и механизмах защиты от НСД, которые рассматриваются в этих курсах. В то же время, вследствие отсутствия принятой на государственном уровне программы подготовки по защитным специальностям до настоящего времени у специалистов и преподавателей вузов не сложилось единого представления, во-первых, о необходимости включения в рабочие программы обучения дисциплины, включающей в себя вопросы правового обеспечения деятельности в области защиты информации, и, во-вторых, о тематике курса. Содержание изданных учебных пособий [1,2] так же не дает возможность сформировать хотя бы концептуально общий подход к объему и характеру сведений в области правовой защиты информации, необходимых для качественной подготовки бакалавров, специалистов и магистров.

На кафедре защиты информации ЗНТУ с первого года подготовки специалистов в сфере защиты информации в рабочие программы включено изучение дисциплины «Правовые основы защиты информации». При этом авторы программы исходили из неоспоримости утверждения, что легитимность (а в некоторых случаях и действенность) использования любых организационных и технических средств защиты определяется их соответствием нормативно-правовой базе, сложившейся в государстве. Следовательно, без знания системы и основных норм информационного законодательства Украины любая деятельность в сфере защиты информации зачастую приобретает не правовой характер с вытекающими отсюда негативными последствиями. Ниже изложены те подходы к формированию содержания указанного курса и методологии преподавания, которые использованы в ЗНТУ и которые, как нам видится, могут быть предметом обсуждения в вузовской и инженерной среде.

ОБОСНОВАНИЕ ПОДХОДА К ФОРМИРОВАНИЮ ПРОГРАММЫ ДИСЦИПЛИНЫ

Исходя из достаточно универсального подхода к изучению любой отрасли знаний: идти от общего к частному, рабочая программа дисциплины «Правовые основы защиты информации» предполагает рассмотрение в порядке очередности следующих базовых тем:

1. Общая характеристика информации и принципов правового обеспечения информационной безопасности.
2. Информационное законодательство Украины.
3. Правовое регулирование информационных отношений между субъектами в информационных системах.
4. Законодательство Украины в сфере защиты государственной тайны.
5. Правовое регулирование электронного документооборота.

6. Лицензирование хозяйственной деятельности в сфере защиты информации.

7. Сертификация средств ТЗИ и КЗИ на соответствие требованиям обеспечения безопасности информации.

8. Правовое регулирование отношений, которые касаются творческой деятельности.

Какие же задачи ставились кафедрой при наполнении каждой из названных тем конкретным содержанием, и на какой методологической основе это осуществлялось?

Тема № 1. Поскольку дисциплина изучается в рамках первого года подготовки специалистов указанных выше специальностей, в этой части курса студентам, которые впервые сталкиваются с научным подходом к постановке проблем обеспечения информационной безопасности и возможностями их решения, даются не только сведения, характеризующие информацию, как объект правовой защиты, но и характеристику: существующих в информационном поле взаимодействия субъектов угроз; способов и методов их нейтрализации; понятия «информационная безопасность»; критериев и категорий защищенности информации. При этом упор делается на усвоении будущими специалистами базового подхода к обеспечению информационной безопасности, основывающегося на том, что защита информации не является самоцелью, а следствием осознания субъектами информационных отношений необходимости защиты своих прав от возможных посягательств на них в виде реализации угроз различного характера.

Вторым важным постулатом, которому, по нашему мнению, необходимо уделить особое внимание, – это признание главенствующей роли государства в защите прав всех участников информационного взаимодействия, не выходящего за пределы правового поля. Такая роль, как известно, основывается на разработанной государственной политике обеспечения информационной безопасности, основные положения которой рассматриваются в рамках курса. Отсюда вытекает необходимость характеристики организационной государственной структуры – системы ТЗИ в Украине, обеспечивающей реализацию этой политики. В результате рассмотрения указанной темы у студентов должно сложиться четкое представление о проблемах, существующих в сфере обеспечения информационной безопасности, путях их решения, роли государства в защите прав законных участников информационных процессов.

При изучении темы № 2 основное внимание, по нашему мнению, следует уделить обоснованию приоритетности решения задачи защиты интересов субъектов информационных отношений на основе нормативно-правового регулирования этой деятельности во всех ее аспектах с общей характеристикой той правовой базы,

которая сложилась на данный момент в Украине. В этом контексте рассматриваются избранный в нашем государстве способ классификации информационного законодательства, его преимущества и недостатки, а так же источники информационного права. При этом подчеркивается роль Основного закона Украины – Конституции, конкретные нормы которой прямо защищают основные права участников информационных процессов и определяют условия, при которых возможно ограничение этих прав. При рассмотрении этой темы внимание студентов также обращается на существующие пробелы в информационном законодательстве Украины, его противоречивость в отдельных аспектах, на существующие возможности его реформирования [1–3].

Отдельного рассмотрения требуют нормы закона Украины «Об информации», являющегося базовым в информационном праве, который заложил правовые основы защиты информации с ограниченным доступом. Здесь основное внимание уделяется таким вопросам как: характеристика структуры информационных отношений и принципы их правового регулирования; виды информационных ресурсов и право собственности на них; виды режима доступа к информации и ее классификация в соответствии с режимом доступа; основания возникновения ответственности за нарушения законодательства и виды такой ответственности. Поскольку наличие уголовной ответственности за отдельные нарушения является наиболее эффективной мерой защиты законных интересов владельцев и пользователей информационных ресурсов, в рамках курса рассматриваются те виды преступлений (предусмотренных криминальным кодексом Украины), предметом которых является информация или поддерживающая ее инфраструктура [2]. Это, в свою очередь, требует формирования у студентов некоторых представлений, связанных с уголовным правом, в частности, понятия преступления и субъекта преступления (с разделением субъектов на виды), состава преступления и виды составов, основания привлечения лица к уголовной ответственности и формы вины. Практика показала, что наиболее эффективной формой изучения конкретных составов преступлений, которые могут совершаться в информационной сфере, является разбор студентами этих составов на практических занятиях с использованием демонстрационных материалов в виде плакатов или слайдов, на которые выведены диспозиция и санкции конкретной статьи КК Украины.

Необходимое внимание уделяется получению студентами представлений о правах и обязанностях спецслужб при осуществлении скрытого доступа к интересующей их информации, принципах оперативно-розыскной деятельности (ОРД), гарантиях законности во время осуществления такой деятельности [4]. Акцент при изучении этих вопросов делается на недопустимости, с точки зрения требований законодательства Ук-

раины, использования специальных методов и средств ОРД не уполномоченными на то структурами или физическими лицами [5].

Давая в целом положительную оценку закону Украины «Об информации» и характеризуя его роль в вопросах правового регулирования информационных отношений, в лекционном курсе одновременно подчеркивается, что этот законодательный акт не в полной мере отвечает современному уровню развития отношений в сфере информационной деятельности. В частности, регулирование отношений в таких сферах, как информация о личности, информация для служебного пользования требует дальнейшего развития на основе разработки и принятия соответствующих законодательных актов [6]. Неопределенными до настоящего времени являются положения действующего законодательства в отношении секретной информации, не составляющей государственной тайны. Накопленный опыт свидетельствует, что если «не уходить» от подобных вопросов, а ставить их перед студентами в рамках изучаемого курса, это поднимает их интерес к правовой проблематике обеспечения информационной безопасности, способствует более глубокому усвоению материала.

Основным средством, обеспечивающим на современном этапе развития общества эффективное информационное взаимодействие между субъектами, являются информационно-телекоммуникационные технологии, обеспечивающие практически все процессы создания, распространения и использования информации. Это влечет за собой необходимость разработки правовых норм, определяющих правила работы с информационными ресурсами, ответственность за их создание и сохранность, за обеспечение их достоверности, а также механизмы пресечения общественно-опасного поведения в их использовании. Правовые нормы указанных типов относятся к проблематике информационного права, и их достаточно детальное рассмотрение является, по нашему мнению, насущной необходимостью. При этом важно указать, что главной особенностью правового регулирования в этой сфере является его опора на совершенно новые для теории права и государства понятия («компьютер», «информационные ресурсы», «база данных», «вычислительная сеть», «электронная цифровая подпись» и т. д.) [7]. Рассмотрение вопросов правового регулирования отношений между субъектами, являющимися законными пользователями автоматизированных систем обработки информации (АС), осуществляется на основе Украины «О защите информации в АС». Основные положения этого закона, касающиеся характеристике угроз информационной безопасности АС, объема прав различных категорий пользователей АС, требований к обеспечению защиты информации, которая обрабатывается в АС, правовой основы создания службы защиты ин-

формации и определения ответственности за правонарушения в сфере использования компьютерных технологий излагаются в лекционном курсе, опираясь на тот объем сведений в сфере компьютерных технологий, которыми располагают студенты в соответствии с программой их обучения. Затем, привязываясь к названному закону, рассматривается порядок защиты государственных информационных ресурсов в информационно-телекоммуникационных системах, определенных соответствующим НД ТЗИ.

Перед изложением базовых положений закона Украины «О государственной тайне» обосновывается особая важность эффективной организации защиты такого вида информации ограниченного доступа, как государственная тайна, утечка которой предполагает нанесения вреда национальным интересам Украины. Неоднократно подчеркивается, что защитные мероприятия должны реализовываться в строгом соответствии с нормами этого закона, достаточно четко регламентирующими все процессы, происходящие в этой сфере, начиная от отнесения информации к категории государственной тайны и заканчивая контролем за соблюдением режима секретности. Этим обуславливается необходимость усвоения студентами таких вопросов как: порядок отнесения информации к государственной тайне; порядок засекречивания и рассекречивания материальных носителей секретной информации; получение прав допуска и доступа к секретной информацией или к работам, с ней связанным; понятие режима секретности и меры по его обеспечению.

Все более широкое использование абонентами компьютерных сетей технологии обмена документами в электронной форме требует знания от специалистов в области обеспечения информационной безопасности юридических аспектов электронного документооборота и возможности использования для подтверждения неизменности и аутентичности электронных документов электронной цифровой подписи (ЭЦП). По этой причине необходимо внимание уделяется изложению основных положений законов Украины «Об электронных документах и электронном документообороте», «Об электронной цифровой подписи». Раскрываются правовые понятия электронного документа и ЭЦП, их статус. Дается характеристика действующей в Украине системе сертификации открытых ключей, особенностей использования ЭЦП в зависимости от статуса сертификата открытого ключа. В то же время, на наш взгляд, нельзя не упомянуть о наличии существенных отклонений названных законов в части определения базовых терминов от законодательства и Директивы ЕС, что может привести к недоразумениям и к неопределенности при использовании этих законодательных актов и к возможным правовым коллизиям на межгосударственном уровне.

Поскольку практически вся хозяйственная деятельность в области технической и криптографической защиты информации подлежит согласно действующему законодательству государственному регулированию, не вызывает сомнений необходимость изложения некоторых норм Закона Украины «О лицензировании определенных видов хозяйственной деятельности», выделяя такие вопросы как: характеристика государственной системы лицензирования; порядок лицензирования; контроль в сфере лицензирования. Отталкиваясь от этих базовых вопросов, далее дается достаточно подробная характеристика системы государственного лицензирования деятельности в области защиты информации, являющейся необходимым элементом национальной системы ТЗИ. Излагаются требования нормативных документов, определяющих перечень видов деятельности в сфере защиты информации, подлежащих лицензированию, особенности лицензирования, характеристика лицензионных условий и порядок контроля за их соблюдением.

Освоение студентами темы, связанной с правовыми основами проведения сертификации технических и криптографических средств защиты информации, предполагает так же ознакомление их с общими правилами подтверждения соответствия, сформулированными в Законе Украины «О подтверждении соответствия» и в декрете КМ Украины «О стандартизации и сертификации». На этой базе далее рассматриваются требования действующих нормативных документов в отношении сертификации средств ТЗИ и КЗИ, а именно: распределение ответственности в системе УкрСЕПРО; организационная структура этой системы; порядок подготовки и проведения сертификации средств КЗИ и ТЗИ; права и обязанности органов сертификации и заявителей; перечень средств ТЗИ и КЗИ, которые подлежат обязательной сертификации; виды документов, подтверждающих соответствие; правовые последствия использования не сертифицированных или не прошедших государственную экспертизу средств обработки и защиты информации.

Важнейшим видом информации на современном уровне развития общества является информация, полученная в результате творческой деятельности, поскольку результаты, например, научно-технической деятельности во многом определяют темпы развития цивилизации, обеспечивают при наличии соответствующих условий достижение высокого уровня экономического развития страны и благосостояния ее граждан. Отсюда вытекает необходимость четкого правового регулирования отношений, возникающих в сфере научно-технического творчества и использования его результатов. Так как эти результаты всегда представляют собой конкретные документированные сведения (описание изобретений или полезных моделей) или формализованное изложение выводов, полученных в ходе мысли-

тельной деятельности, патентное и авторское право в определенной части можно считать частью информационного права, по крайней мере, если мы говорим о защите неимущественных прав авторов. На основе таких соображений в рамках дисциплины рассматриваются основные положения Законов Украины «Об авторских и смежных правах» и «Об охране прав на изобретения и полезные модели». При этом акцентируется внимание студентов на принципиальном различии объектов правового регулирования в авторском и патентном праве при схожести подходов к защите прав авторов как произведений, так и результатов изобретательской деятельности.

Проведение практических занятий, с целью надежного закрепления знаний, полученных студентами в ходе слушания лекций по данной дисциплине и в процессе самостоятельной подготовки, осуществляется в форме заслушивания рефератов, подготовленных студентами по темам, требующим более обстоятельного рассмотрения, чем это сделано (ввиду недостатка времени) при чтении лекций, и путем решения студентами практических задач, вытекающих из рассмотренных вопросов.

ВЫВОДЫ

Востребованность специалистов в области защиты информации определяется уровнем их подготовки, позволяющей решать (в числе прочих) практические задачи по предупреждению или нейтрализации существующих угроз информационной безопасности. Достаточный уровень подготовки таких специалистов может быть достигнут только с учетом необходимости усвоения ими базовых знаний в области правового регулирования многообразной деятельности, конечной целью которой является защита интересов законных участников информационных процессов. Это является гарантией не только законности такой деятельности, но, что не менее важно, и ее эффективности.

Предлагаемый подход к формированию программы и методики обучения, разумеется, не является единственно возможным. Авторы предполагают возможность существования других концепций, на основе которых возможно построения курса, посвященного правовым основам защиты информации. Прежде всего, речь, по-видимому, может идти о другом видении про-

граммы дисциплины, ее структуры, акцентах. Возможно, большее внимание следует уделить таким вопросам, как: формирование информационной культуры, как объекта информационного права; международным аспектам информационного права; роли информации в общественных отношениях на разных исторических этапах развития общества и т. д. [1], оставляя некоторые из названных выше тем для самостоятельного изучения студентами. Цель, которая преследовалась авторами при написании этой работы, будет достигнута, если данная статья послужит одним из поводов для активизации дискуссии по данной проблематике.

ПЕРЕЧЕНЬ ССЫЛОК

1. Цимбалюк В. С., Гавловський В. Д., Гриценко В. В. та ін. Основи інформаційного права України.: Навч. посіб. / О-72. – К.: Знання, 2004. – 274 с.
2. Голубев В. І., Гавловський В. Д., Цимбалюк В. С. Інформаційна безпека: проблеми боротьби зі злочинами у сфері використання комп'ютерних технологій.: монографія. – Запоріжжя: Просвіта, 2001. – 252 с.
3. Гавловський В. Д., Гуцалюк М. В., Цимбалюк В. С. Електронний бізнес та концепція реформування системи інформаційного законодавства України // Бизнес и безопасность. – 2001. – № 4. – С. 2–4.
4. Планов С. А. Негласное получение информации (нормативно-правовое эссе) // Бизнес и безопасность. – 2003. – № 4 – С. 9–12.
5. Пиза Д. М., Слепцов В. И. Информационная безопасность. Вопросы подготовки кадров // Радиоелектроніка, інформатика, управління. – 2005. – № 2.
6. Ботвінкін О. В., Ворожко В. П. Інформація з обмеженим доступом, що не є державною таємницею, в законодавстві України. Аналітичний огляд. – Київ.: НА СБ України, 2006. – 96 с.
7. Загородников С. Н., Шмелев А. А. Организационное и правовое обеспечение информационной безопасности. Часть 1 // Приложение к журналу «Информационные технологии». – 2005. – № 12. – С. 5–7.

Надійшла 5.06.2008
Після доробки 9.06.2008

Розглядаються деякі питання правової підготовки фахівців з напрямків, що входять в галузь знань «Інформаційна безпека», з урахуванням специфіки їх майбутньої професійної діяльності, учбових завдань, законодавства України і досвіду роботи кафедри захисту інформації Запорізького національного технічного університету.

Some questions of legal preparation of specialists are examined on directions, to included in the field of knowledges «Informative safety», taking into account the specific of their future professional activity, educational tasks, legislation of Ukraine and experience department of priv the Zaporozhia national technical university.

КЕРУВАННЯ У ТЕХНІЧНИХ СИСТЕМАХ

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УДК 681.513

А. Е. Казурова, Н. В. Ивочка

ОБЕСПЕЧЕНИЕ ЗАДАННОГО СООТНОШЕНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ЭЛЕКТРОТЕХНИЧЕСКОЙ СИСТЕМЫ С НЕПОЛНЫМ ИЗМЕРЕНИЕМ ВЕКТОРА СОСТОЯНИЯ

Рассматривается управление упругой электромеханической системой, особенностью которой является наличие упругости и значительного нелинейного трения как на объекте управления, так и на двигателе. В качестве измерений выступают ток якорной цепи двигателя, скорость ротора двигателя и угол поворота и скорость объекта управления. Для этой системы синтезирован оптимальный по быстродействию робастный комбинированный регулятор с наблюдателем неопределенностей и их компенсацией. Компьютерное моделирование подтвердило задаваемые динамические характеристики синтезированной системы управления.

Рассматривается электромеханическая система (ЭМС), состоящая из электродвигателя постоянного тока (ДПТ) или синхронного двигателя с векторным управлением, редуктора, обладающего упругостью и объекта управления (ОУ). Редуктор обладает ограниченной жесткостью, вследствие чего в нем при работе возникают упругие деформации. ОУ представляет собой твердое тело с неизвестными массой и моментом инерции, к которому приложены ступенчато изменяющиеся воздействия. Работа двигателя с редуктором и объекта управления сопровождается нелинейным трением. Измеряются ток якоря (в случае ДПТ), скорость ротора дви-

гателя и перемещение ОУ. Такая система является составной частью систем управления роботов, металлообрабатывающих станков, телескопов, радиолокаторов, стационарных и на подвижном основании систем автоматического прицеливания стрелкового оружия, систем управления перемещением электродов в электроплавильных печах и т. п.

Традиционным методом управления указанными системами является использование ПИ- и ПИД-регуляторов. Однако этот метод управления обладает недостаточной робастностью, плохо противодействует ступенчато изменяющимся внешним воздействиям и нелинейному трению с провалами характеристики. Этих недостатков лишены системы со скользящими режимами (ССР). К сожалению, скользящие режимы сопровождаются вибрациями, которые ухудшают надежность системы, повышают ее электропотребление, вызывают появление акустических шумов. Вибрации еще больше усиливаются при наличии высокочастотных помех. Кроме того, ССР могут возбуждать высокочастотную паразитную динамику (в данном случае, упругость) вплоть до потери устойчивости. Хорошей робастностью к неопределенностям, быстрым реагированием

© Казурова А. Е., Ивочка Н. В., 2008

на ступенчато изменяющиеся воздействия и одновременно при «мягком» управлении без скользящих режимов обладают комбинированные системы управления с оценкой и компенсацией неопределенностей динамики [1–3]. Благодаря отсутствию скользящих режимов, эти системы в меньшей степени возбуждают паразитную высокочастотную динамику.

Целью данной статьи является обеспечение высокой точности и быстродействия ЭМС с выполнением заранее заданных соотношений показателей качества переходных процессов.

1 ПОСТАНОВКА ЗАДАЧИ

Система описывается следующими уравнениями:

$$I_c \ddot{\varphi}_c = c(n^{-1}\varphi_m - \varphi_c) + f_c - f_l, \quad (1)$$

$$I_m \ddot{\varphi}_m = -n^{-1}c(n^{-1}\varphi_m - \varphi_c) + c_m i + f_m, \quad (2)$$

$$L \dot{i} + R i = u - c_e \dot{\varphi}_m, \quad (3)$$

где φ_c , φ_m – углы поворота ОУ и ротора двигателя; i , u – ток и напряжение якорной цепи; I_c , I_m – моменты инерции ОУ и ротора двигателя соответственно; R , L – активное сопротивление и индуктивность якорной цепи; c_m , c_e – моментный и скоростной коэффициенты двигателя постоянного тока (ДПТ); n – передаточное число редуктора; c – коэффициент жесткости упругой части; f_c – нелинейное трение в ОУ; f_m – нелинейное трение в двигателе; f_l – момент нагрузки.

Для достижения поставленной цели необходимо синтезировать комбинированный робастный регулятор с максимальным быстродействием при заданном соотношении показателей качества.

2 СИНТЕЗ РОБАСТНОГО КОМПЕНСАТОРА

Рассмотрим управление ЭМС (1)–(3) со следующим составом измерителей: датчик тока якорной цепи ДПТ, тахогенератор, измеряющий скорость ротора двигателя, и датчики положения и скорости ОУ. Таким образом, в системе пятого порядка измеряются 4 координаты: i , $\dot{\varphi}_m$, φ_c , $\dot{\varphi}_c$. В системе (1)–(3) неизвестными являются I_c , f_c , f_l , f_m . Представим управляющее напряжение в виде

$$u = u_i + u_m + u_c, \quad (4)$$

где u_i , u_m , u_c – управляющие напряжения соответственно для контуров тока, двигателя и объекта управления.

Пусть

$$u_i = -Rk_i i, \quad (5)$$

где k_i – постоянный коэффициент. Тогда уравнение (3) примет вид

$$T_e \dot{i} + (1 + k_i)i = R^{-1}(u_m + u_c - c_e \dot{\varphi}_m). \quad (6)$$

Подставим i из (6) в уравнение (2). Получим

$$I_m \ddot{\varphi}_m = -n^{-1}c(n^{-1}\varphi_m - \varphi_c) + c_{m1}[R^{-1}(u_m + u_c - c_e \dot{\varphi}_m) - T_e \dot{i}] - f_m, \quad (7)$$

где

$$c_{m1} = c_m(1 + k_i)^{-1}. \quad (8)$$

Управляющее напряжение двигателя представим в виде

$$u_m = -k_m \dot{\varphi}_m, \quad (9)$$

где k_m – постоянный коэффициент. Подставим уравнение (9) в уравнение (7), затем умножим полученное выражение на n и сложим с уравнением (1), в результате чего получим

$$I_c \ddot{\varphi}_c + nI_m \ddot{\varphi}_m = nc_{m1}[R^{-1}(u_c - (k_m + c_e)\dot{\varphi}_m) - T_e \dot{i}] - f_c - f_l - nf_m. \quad (10)$$

Представим момент инерции ОУ в следующем виде:

$$I_c = I_{c0} + I_{c\delta}, \quad (11)$$

где I_{c0} – известное номинальное значение момента инерции ОУ, $I_{c\delta}$ – его неизвестное отклонение. С учетом обозначения (11) уравнение (10) переписывается в виде

$$I_{c0} \ddot{\varphi}_c + nI_m \ddot{\varphi}_m = nc_{m1}[R^{-1}(u_c - (k_m + c_e)\dot{\varphi}_m) - T_e \dot{i}] - f_c - f_l - nf_m - I_{c\delta} \ddot{\varphi}_c. \quad (12)$$

Сформируем неопределенность

$$f_{c\Sigma} = -f_c - f_l - nf_m - I_{c\delta} \ddot{\varphi}_c, \quad (13)$$

с учетом которой уравнение (12) принимает вид

$$I_{c0} \ddot{\varphi}_c + nI_m \ddot{\varphi}_m = nc_{m1}[R^{-1}(u_c - (k_m + c_e)\dot{\varphi}_m) - T_e \dot{i}] + f_{c\Sigma}. \quad (14)$$

Управляющее напряжение объектом управления представим в виде

$$u_c = u_{cs} + u_{cc}, \quad (15)$$

где u_{cS} , u_{cC} – управляющие напряжения стабилизации и компенсации неопределенности, которые сформируем в виде

$$u_{cC} = -(nc_{m1})^{-1}R\hat{f}_{c\Sigma}, \quad (16)$$

$$u_{cS} = -k_{c1}(\varphi_c - \varphi_{cp}) - k_{c2}(\dot{\varphi}_c - \dot{\varphi}_{cp}), \quad (17)$$

где k_{c1} , k_{c2} – постоянные коэффициенты, $\hat{f}_{c\Sigma}$ – оценка $f_{c\Sigma}$, которую можно получить с помощью наблюдателя неопределенности.

Для нахождения оценки неопределенности сформируем эту неопределенность, но с учетом того, что на управляющее напряжение накладывается ограничение, т. е.

$$u = \text{sat}(u_i + u_m + u_c). \quad (18)$$

Найдем ток из уравнения (3), он будет иметь вид

$$i = R^{-1}(u - c_e \dot{\varphi}_m) - T_e \dot{i}. \quad (19)$$

Последовательная подстановка уравнения (19) в уравнение (2), умножение полученного выражения на n и сложение с уравнением (1) дает

$$I_c \ddot{\varphi}_c + nI_m \ddot{\varphi}_m = nc_m[R^{-1}(u - c_e \dot{\varphi}_m) - T_e \dot{i}] - f_c - f_l - nf_m, \quad (20)$$

или с учетом обозначения (11)

$$I_{c0} \ddot{\varphi}_c + nI_m \ddot{\varphi}_m - nc_m[R^{-1}(u - c_e \dot{\varphi}_m) - T_e \dot{i}] = -f_c - f_l - nf_m - I_{c\delta} \ddot{\varphi}_c. \quad (21)$$

Неопределенность представлена выражением (13). Тогда уравнение (21) переписывается в виде

$$f_{c\Sigma} = I_{c0} \ddot{\varphi}_c + nI_m \ddot{\varphi}_m - nc_m[R^{-1}(u - c_e \dot{\varphi}_m) - T_e \dot{i}]. \quad (22)$$

Поскольку дифференцировать ток i нежелательно и нет информации об ускорениях двигателя и ОУ, то избавимся от необходимости получения производной тока и указанных ускорений следующим образом. Будем полагать быстрое действие наблюдателя $f_{c\Sigma}$ настолько большим, что в масштабе времени наблюдателя можно считать

$$\dot{f}_{c\Sigma} = 0. \quad (23)$$

По уравнениям (23), (24) построим наблюдатель для оценки $f_{c\Sigma}$ в виде

$$\dot{\hat{f}}_{c\Sigma} = l(\hat{f}_{c\Sigma} - f_{c\Sigma}), \quad (24)$$

где l_f – коэффициент передачи наблюдателя. Подстановка (22) в (24) дает

$$\dot{\hat{f}}_{c\Sigma} = l_f(\hat{f}_{c\Sigma} + nc_m[R^{-1}(u - c_e \dot{\varphi}_m) - T_e \dot{i}] - I_{c0} \ddot{\varphi}_c - nI_m \ddot{\varphi}_m). \quad (25)$$

Введем обозначение

$$z := \hat{f}_{c\Sigma} + l_f(I_{c0} \dot{\varphi}_c + nI_m \dot{\varphi}_m + nc_m T_e \dot{i}) \Rightarrow \Rightarrow \dot{\hat{f}}_{c\Sigma} = z - l_f(I_{c0} \dot{\varphi}_c + nI_m \dot{\varphi}_m + nc_m T_e \dot{i}), \quad (26)$$

с учетом которого уравнение (26) принимает вид

$$\dot{z} = l_f(\hat{f}_{c\Sigma} + nc_m R^{-1}(u - c_e \dot{\varphi}_m)). \quad (27)$$

$\hat{f}_{c\Sigma}$ определяется по результату интегрирования уравнения (28) и второму выражению в (27). Пусть ошибка наблюдателя

$$\tilde{f}_{c\Sigma} = \hat{f}_{c\Sigma} - f_{c\Sigma}. \quad (28)$$

Вычитая из уравнения (24) уравнение (23), получим уравнение $\dot{\tilde{f}}_{c\Sigma} = l_f \tilde{f}_{c\Sigma}$ с решением

$$\tilde{f}_{c\Sigma}(t) = \tilde{f}_{c\Sigma}(0) \exp(l_f t), \quad (29)$$

где ε – зона, при заходе в которую переходный процесс считается закончившимся, t_t – заданное время переходного процесса. Из (29) следует, что при $l_f < 0$ решение $\tilde{f}_{c\Sigma} = 0$ является асимптотически устойчивым, поэтому на основании (29) можно записать $\hat{f}_{c\Sigma} = f_{c\Sigma}$ при $t \rightarrow \infty$. Из выражения (30) по заданному времени переходного процесса можно определить коэффициент l_f

$$l_f = \frac{1}{t_t} \ln \varepsilon. \quad (30)$$

Для нахождения коэффициентов закона управления (4) составим характеристическое уравнение для системы уравнений (1)–(3), вектор состояния которой $x = [\varphi_c \ \dot{\varphi}_c \ \varphi_m \ \dot{\varphi}_m \ i]^T$. Определитель системы (1)–(3) имеет вид

$$\Delta = \begin{bmatrix} I_{c0} p^2 + c & -cn^{-1} & 0 \\ -n^{-1}c & I_m p^2 + n^{-2}c & -c_m \\ R^{-1}(k_{c1} + k_{c2}p) & R^{-1}(k_m + c_e)p & T_e p + 1 + k_i \end{bmatrix}, \quad (31)$$

откуда получается характеристическое уравнение

$$p^5 + T_e^{-1}(1 + k_i)p^4 + (c(I_m^{-1}n^{-2} + I_{c0}^{-1}) + c_m(RT_e I_m)^{-1}(k_m + c_e))p^3 + T_e^{-1}c(I_m^{-1}n^{-2} + I_{c0}^{-1})(1 + k_i)p^2 + c_m c(RT_e I_m I_{c0})^{-1}(k_m + c_e + k_{c2}n^{-1})p + c(nRT_e I_m I_{c0})^{-1}c_m k_{c1} = 0, \quad (32)$$

где p – оператор дифференцирования. Сопоставляя уравнение (32) с эталонным характеристическим уравнением (33)

$$p^5 + 5\omega_0 p^4 + 10\omega_0^2 p^3 + 10\omega_0^3 p^2 + 5\omega_0^4 p + \omega_0^5 = 0, \quad (33)$$

где ω_0 характеризует быстродействие управляемого контура, можно получить значения коэффициентов закона управления и оптимальное значение ω_0 в виде

$$\omega_0 = \sqrt{0,5(I_{c0}I_m)^{-1}(I_{c0}n^{-2}c + I_m c)}, \\ k_i = 5T_e\omega_0 - 1,$$

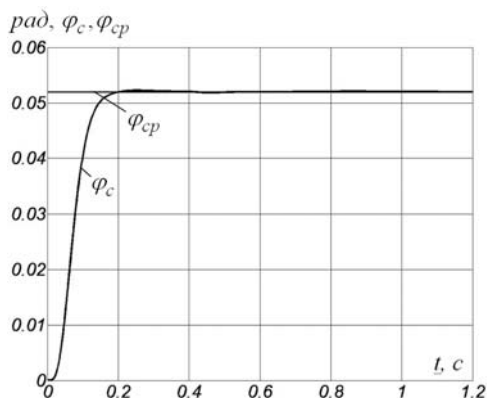


Рисунок 1

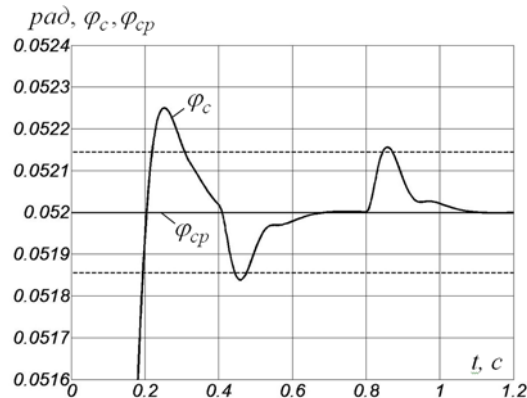


Рисунок 2

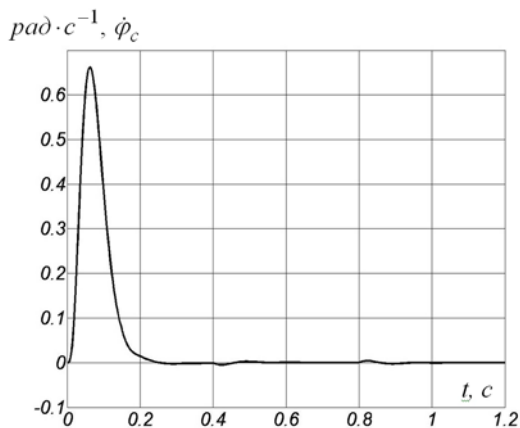


Рисунок 3

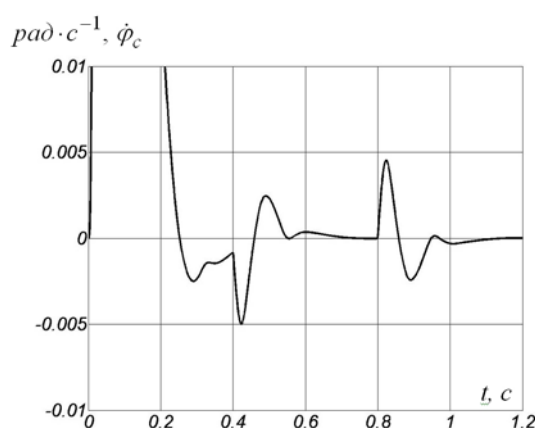


Рисунок 4

$$k_m = \frac{RT_e(10I_{c0}I_m\omega_0^2 - c(I_{c0}n^{-2} + I_m))}{c_m I_{c0}} - c_e, \\ k_{c1} = I_{c0}I_m T_e c^{-1} n c_m^{-1} R \omega_0^5, \\ k_{c2} = \left(\frac{5I_{c0}I_m T_e R \omega_0^4}{c_m c} - k_m - c_e \right) n. \quad (34)$$

3 РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

При моделировании принимались следующие параметры: $I_{c0} = 250 \text{ кг} \cdot \text{м}^2$, $I_c = (0, 67 \dots 1, 5)I_{c0}$, $I_m = 27 \times 10^{-5} \text{ кг} \cdot \text{м}^2$, $R = 0,075 \text{ Ом}$, $T_e = 4,5 \cdot 10^{-3} \text{ с}$, $L = 3,375 \cdot 10^{-4} \text{ Гн}$, $c_m = 0,062 \text{ Н} \cdot \text{м} \cdot \text{А}^{-1}$, $c_e = 0,062 \text{ В} \cdot \text{с} \cdot \text{рад}^{-1}$, $n = 377$, $c = 3 \cdot 10^5 \text{ Н} \cdot \text{м} \cdot \text{рад}^{-1}$, $f_l = 0 \dots 80 \text{ Н} \cdot \text{м}$, $f_c = 200 \text{ Н} \cdot \text{м}$, $f_m = 0,15 \text{ Н} \cdot \text{м}$, пусковой ток 360 А, напряжение питания двигателя 27 В.

На рис. 1–9 показаны переходные процессы, возникающие при подаче ступенчатого программного сигнала величиной 3° ($\varphi_{cp} = 0,05236 \text{ рад}$) при номинальном моменте инерции ОУ. Нагрузка $f_l = 80 \text{ Н} \cdot \text{м}$ в виде

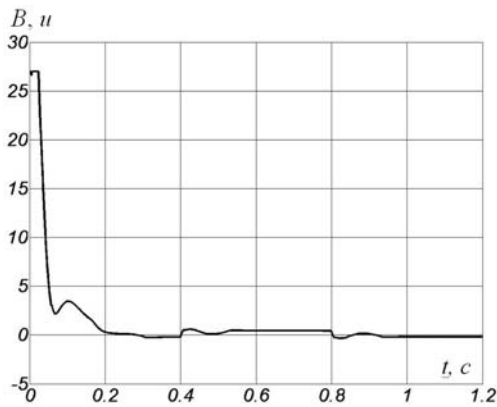


Рисунок 5

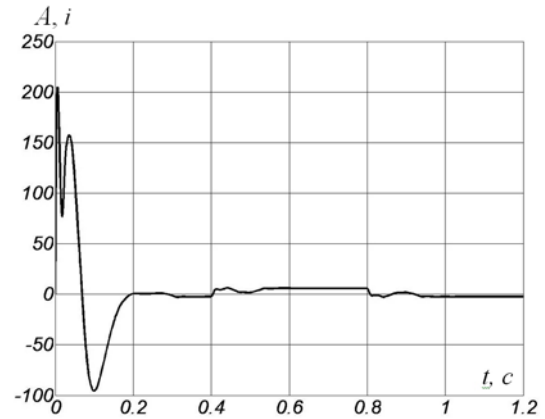


Рисунок 6

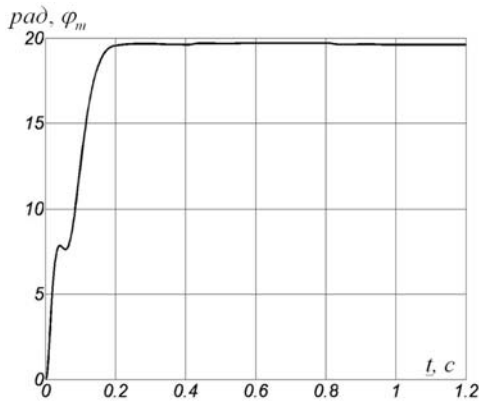


Рисунок 7

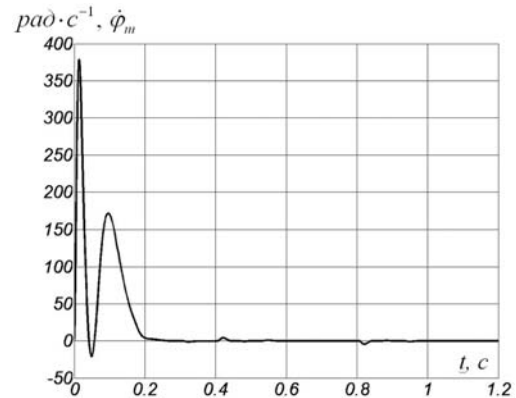


Рисунок 8

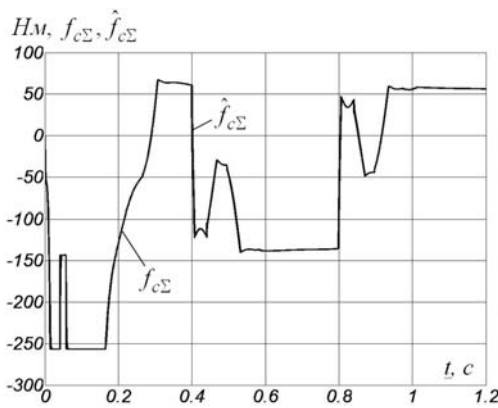


Рисунок 9

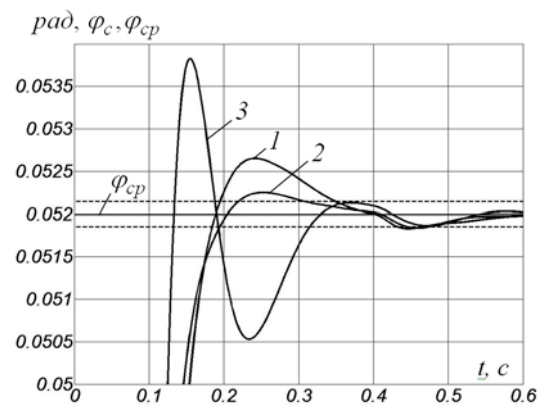


Рисунок 10

ступенчатого сигнала прикладывалась на 0,4 с и снималась на 0,8 с. На рис. 1 показан весь переходный процесс, а на рис. 2 – его фрагмент. На рис. 2 пунктирными линиями ограничена зона ± 30 угл. с. Анализ графиков показывает, что синтезированный алгоритм управления обеспечивает входение в зону ± 30 угл. с за 0,31 с. Рис. 2 свидетельствует, что система является

астигматической по отношению к нагрузке. На рис. 3, 4 представлена скорость ОУ. Рис. 4 является фрагментом рис. 3. Графики напряжения и тока изображены на рис. 5, 6. На рис. 7, 8 представлены графики угла и скорости поворота ротора двигателя. Графики неопределенности и ее оценки показаны на рис. 9. Рис. 10 демонстрирует фрагменты переходных процес-

сов при значениях момента инерции ОУ 1) $0,67I_{c0}$, 2) I_{c0} , 3) $1,5I_{c0}$, подтверждающие работоспособность системы при неизвестных отклонениях момента инерции от его номинального значения в указанном пределе.

ЗАКЛЮЧЕНИЕ

Рассмотрено управление упругой электромеханической системой со следующим составом управляющей системы: ДПТ, датчик тока его якорной цепи, тахогенератор, измеряющий скорость ротора двигателя, и датчики положения и скорости ОУ. Таким образом, в системе пятого порядка измеряются 4 координаты. Для такого соотношения размерностей векторов состояния и измерения предложен метод синтеза оптимальных по быстродействию комбинированных алгоритмов управления с наблюдателем неопределенностей и их компенсацией. Особенностью данной ЭМС является наличие значительного нелинейного трения как на объекте управления, так и на двигателе. Примененное комбинированное управление практически устраняет влияние нелинейного трения. Как показали результаты моделирования, синтезированный алгоритм управления обеспечивает входжение в зону ± 30 угл. с за $0,31$ с. Система является астатической по отношению к нагрузке. Она работоспособна при отклонениях момента инерции ОУ в пределах $0,67 \dots 1,5$ от его номинального значения.

УДК 681.527

Е. В. Душинова, В. И. Левыкина, Е. В. Васильева

ПРОСТОЙ МЕТОД ОЦЕНКИ ПОТОКОСЦЕПЛЕНИЙ АСИНХРОННОГО ДВИГАТЕЛЯ

Разработан новый алгоритм для оценки потокосцеплений, соответствующих им синхронных скоростей и модулей, работоспособный при всех частотах. Предложенный интегратор двухфазных сигналов прост в вычислительном отношении и обеспечивает высокую точность оценки указанных переменных.

ВВЕДЕНИЕ

В современных системах управления асинхронными двигателями (АД) (векторное управление и прямое управление моментом и потокосцеплением) ключевыми координатами их вектора состояния являются проекции векторов потокосцеплений ротора или статора на

ПЕРЕЧЕНЬ ССЫЛОК

1. Потапенко Е. М. Сравнительная оценка робастных систем управления с различными типами наблюдателей // Изв. РАН. Теория и системы управления. – 1995. – № 1. – С. 109–116.
2. Потапенко Е. М. Робастные комбинированные системы управления с наблюдателями // Проблемы управления и информатики (Киев, НАНУ). – 1995. – № 2. – С. 36–44.
3. Потапенко Е. М. Исследование робастности систем управления с наблюдателями // Изв. РАН. Теория и системы управления. (Москва) – 1996. – № 2. – С. 104–108.

Надійшла 8.04.2008

Розглядається керування пружною електромеханічною системою, особливістю якої є наявність пружності та значного нелінійного трення як на об'єкті керування, так і на двигуні. В якості вимірювань виступають струм якорного кола двигуна, швидкість ротора двигуна й кут повороту та швидкість об'єкта керування. Для цієї системи синтезовано оптимальний за швидкодією робастний комбінований регулятор зі спостережником невизначеностей та їх компенсацією. Комп'ютерне моделювання підтвердило динамічні характеристики синтезованої системи керування.

The flexible electromechanical system control, the specificity of which is the presence of elasticity and significant nonlinear friction on both control object and motor, is considered. Motor armature current, motor rotor speed and control object rotation angle and speed are measured. For this system the optimal in quick-action robust combined regulator with the observer of uncertainty and its compensation is synthesized. Computer simulation confirms the dynamic characteristics of the synthesized control system.

оси различных базисов и скорости вращения этих векторов. Известно, что потокосцепление можно найти интегрированием ЭДС. Будем различать приближенное и формальное с математической точки зрения («чистое») интегрирование. При «чистом» интегрировании возникают ошибки из-за неизвестных начальных условий и интегралов от постоянных погрешностей датчиков тока и напряжения, формирующих ЭДС. Для устранения ошибок от начальных условий и ограничения ошибок от смещения нулей датчиков предложено вместо звеньев «чистого» интегрирования устанавливать инерционные звенья первого порядка [1, 2]. Однако при низких частотах такая замена приводит

к недопустимо большим погрешностям амплитуды и фазы потокосцепления. К настоящему времени предложено большое количество методов оценки указанных переменных. Большинство из них неработоспособны в окрестности нулевой скорости ротора. Исключение составляют методы $Hu - Wu$ [3], Ohtany [4], Holtz [5, 6]. Алгоритмы методов [3–6] хорошо работают только при малых начальных отклонениях $\hat{\psi}$ от истинного вектора потокосцепления ψ и имеют слабую сходимостъ при больших начальных отклонениях. Поэтому алгоритмы [3–6] не годятся при включении в режиме подхвата (включение двигателя с разогнанным ротором). Этим методам свойственны длительные переходные процессы.

Для оценки сопротивления ротора R_r необходимо изменение модуля вектора потокосцепления ротора $|\psi|$. С этой целью в намагничивающую составляющую тока статора инжектируют низкочастотную синусоидальную составляющую с малой амплитудой. Это накладывает дополнительное требование на методы оценки ψ , а именно, достаточно точное воспроизведение инжектированной составляющей в $|\psi|$ в оценке $|\hat{\psi}|$. Коррекция алгоритмов работ [4–6] осуществляется путем минимизации разности $(\hat{\psi} - \psi_p)$, где ψ_p – программное значение потокосцепления в статорной системе координат, не содержащее инжектируемой составляющей. Поэтому при достаточно больших коэффициентах передачи алгоритма, необходимых для требуемого быстродействия, $\hat{\psi}$ стремится к ψ_p , вследствие чего оценка $\hat{\psi}$ также не будет содержать инжектируемой составляющей. По этой причине коэффициент при невязке $(\hat{\psi} - \psi_p)$ следует уменьшать, а это уменьшает быстродействие алгоритма. Предложенный в работе [2] метод интегрирования лишен этих недостатков. Однако вследствие того, что участвующая при оценке вектора потокосцепления его скорость вращения сама оценивается по проекциям искомого вектора потокосцепления, длительность переходных процессов велика.

Цель данной статьи – модификация и исследование модифицированного метода оценки потокосцепления, работоспособного при любых скоростях ротора, включая близкие к нулевой скорости, обладающего короткими переходными процессами.

ПОСТАНОВКА ЗАДАЧИ

Оценку вектора потокосцепления статора ψ_s можно получить путем интегрирования ЭДС статора e_s в соответствии с уравнением

$$\dot{\psi}_s = e_s. \tag{1}$$

В свою очередь,

$$e_s = u - R_s i. \tag{2}$$

Вектор потокосцепления ротора ψ_r связан с вектором ψ_s соотношением

$$\psi_r = L_{rm}(\psi_s - \sigma L_s i). \tag{3}$$

В (2) и (3) L_s, L_r, L_m – индуктивности статора, ротора и взаимная индуктивность статора и ротора, $L_{rm} = L_r L_m^{-1}$, R_s – активное сопротивление статора, σ – коэффициент рассеивания, u, i – векторы напряжения и тока статора. Векторы ψ, u, i, e_s заданы в статорном базисе (α, β) двумя проекциями.

Недостаток уравнения (1) заключается в том, что для получения ψ_s необходимо использовать чистое интегрирование. При чистом интегрировании возникают ошибки из-за неизвестности начальных условий, а также появляются ошибки из-за интегрирования возможных постоянных неизвестных смещений нулей, возникающих при измерении токов и напряжений. На рис. 1 дано сравнение истинной составляющей потокосцепления ψ_α с ее оценкой $\hat{\psi}_\alpha$, полученной чистым интегрированием.

Одним из методов устранения ошибок чистого интегрирования является замена интегратора инерционным звеном первого порядка с постоянной времени T [1, 7]. В этом случае оценка потокосцепления осуществляется по зависимости

$$T \dot{\hat{\psi}}_s + \hat{\psi}_s = T e_s. \tag{4}$$

Уравнение (4) соответствует пропуску сигнала ЭДС через инерционное звено. При такой оценке возникают ошибки оценки амплитуды и фазы, особенно когда двигатель работает на частоте, ниже частоты сопряжения (T^{-1}) инерционного звена. Привода, использующие инерционные звенья, обычно имеют ограниченный диапазон регулирования, обычно 1:10 при требовании, предъявляемом к современным электроприводам, 1:150...40000.

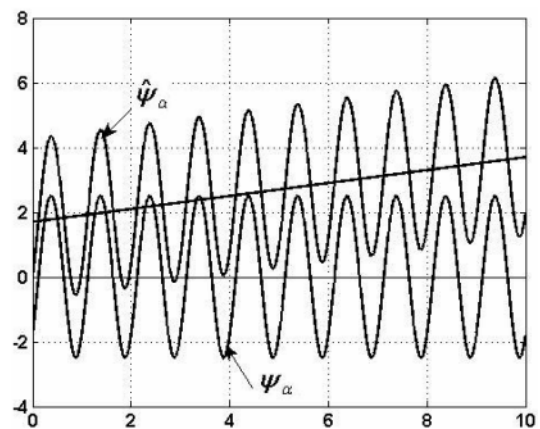


Рисунок 1 – Сравнение истинной составляющей потокосцепления ψ_α и ее оценки $\hat{\psi}_\alpha$, полученной с помощью чистого интегрирования

ОСНОВНОЙ РЕЗУЛЬТАТ

Наиболее привлекательным методом оценки потокосцепления является метод, описанный в работе [2]. Этот метод так же, как и (4), базируется на инерционном звене, но с переменной постоянной времени. Для увеличения рабочего диапазона рабочих скоростей с уменьшением синхронной скорости должна возрастать постоянная времени. Поэтому принимается $T = (\varepsilon|\omega_{0s}|)^{-1}$, где ω_{0s} – синхронная статорная скорость, ε – малый положительный скаляр. С учетом сказанного для оценки потокосцепления статора уравнение (4) примет вид

$$\dot{\hat{\psi}}_s + \varepsilon|\omega_{0s}|\hat{\psi}_s = e_s. \quad (5)$$

В работе [2] по проекциям вектора $\hat{\psi}_s$ находится угол ϑ между вектором $\hat{\psi}_s$ и осью α статорного базиса, а затем дифференцированием находится оценка $\hat{\omega}_{0s}$ синхронной скорости ω_{0s} . Для точного воспроизведения интегрирования необходимо ε делать как можно меньше. С другой стороны, для быстрого устранения начальных условий и уменьшения накопления ошибки за счет интегрирования возможного постоянного смещения в сигнале e_s надо увеличивать ε . Для смягчения этого противоречия в работе [2] предлагается приблизить динамические свойства звена (5) к динамическим свойствам интегратора за счет коррекции вектора e_s . Это достигается следующим образом.

Передаточные функции интегрирующего и инерционного звеньев соответственно равны

$$\frac{1}{s}, \quad \frac{1}{s + T^{-1}},$$

где постоянная времени T определяется уравнением

$$T^{-1} = \varepsilon|\omega_{0s}|. \quad (6)$$

Тогда АФЧХ указанных звеньев определяются выражениями

$$\frac{1}{j\omega_{0s}}, \quad \frac{1}{j\omega_{0s} + \varepsilon|\omega_{0s}|}.$$

АФЧХ инерционного звена представим в виде

$$W(j\omega) = \frac{1}{j\omega + \varepsilon|\omega|} = \frac{1}{j\omega(1 - j\varepsilon\text{sign}\omega)}. \quad (7)$$

С учетом вида АФЧХ (7) будем подавать на вход инерционного звена сигнал

$$\begin{aligned} e_\alpha^* + je_\beta^* &= (e_\alpha + je_\beta)(1 - j\varepsilon\text{sign}\omega) = \\ &= (e_\alpha + e_\beta\varepsilon\text{sign}\omega) + j(e_\beta - e_\alpha\varepsilon\text{sign}\omega). \end{aligned} \quad (8)$$

Тогда с учетом (7) и (8) будем иметь

$$\begin{aligned} (e_\alpha^* + je_\beta^*) \frac{1}{j\omega + \varepsilon|\omega|} &= \\ = (e_\alpha + je_\beta)(1 - j\varepsilon\text{sign}\omega) \frac{1}{(1 - j\varepsilon\text{sign}\omega)} &= \\ = (e_\alpha + je_\beta) \frac{1}{j\omega}. \end{aligned}$$

Пусть сигналы e_α , e_β содержат постоянные составляющие e_α^0 , e_β^0 . Тогда при $\text{sign}\omega = \text{const}$ постоянные сигналы будут пропускаться через инерционное звено.

$$\dot{\hat{\psi}} + \varepsilon|\omega|\hat{\psi} = e^0.$$

В этом случае оценка потокосцепления будет иметь постоянную ошибку

$$\hat{\psi}_0 = \frac{e^0}{\varepsilon|\omega|},$$

которая при $\omega \neq 0$ будет ограниченной.

Для того чтобы выходной сигнал инерционного звена совпадал с выходным сигналом интегрирующего звена, скорректируем в соответствии с (8) входной сигнал инерционного звена $e^* = (e_\alpha^*, e_\beta^*)^T$ следующим образом:

$$e^* = (E - J\varepsilon\text{sign}\omega_{0s})e = \begin{bmatrix} e_\alpha + e_\beta\varepsilon\text{sign}\omega_{0s} \\ e_\beta - e_\alpha\varepsilon\text{sign}\omega_{0s} \end{bmatrix}, \quad (9)$$

где J , E – кососимметрическая и симметрическая матрицы соответственно, определенные выражениями

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Уравнение фильтра (5) со скорректированным входным сигналом имеет вид

$$\dot{\hat{\psi}}_s + \varepsilon|\hat{\omega}_{0s}|\hat{\psi}_s = e_s^* = (E - J\varepsilon\text{sign}\hat{\omega}_{0s})e_s, \quad (10)$$

где скорость ω_{0s} заменена ее оценкой $\hat{\omega}_{0s}$. Структурная схема фильтра (10), предложенного в работе [2], изображена на рис. 2.

Метод оценки скорости ω_{0s} , изображенный на рис. 2, соответствует операции

$$\hat{\omega}_{0s} = -\frac{\hat{\psi}_s^T J \dot{\hat{\psi}}_s}{|\hat{\psi}_s|^2}. \quad (11)$$

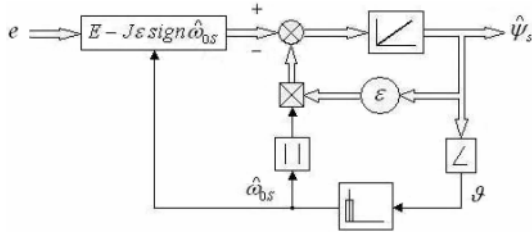


Рисунок 2

Как видно на рис. 2, оценка $\hat{\psi}_s$ зависит от $\hat{\omega}_{0s}$, то есть оценка $\hat{\omega}_{0s}$ осуществляется в замкнутом контуре, что приводит к длительным переходным процессам. Для устранения этих недостатков предлагается вместо (11) использовать выражение

$$\hat{\omega}_{0s} = -\frac{e_s^T J \dot{e}_s}{|e_s|^2}. \quad (12)$$

Для оценки потокосцепления ротора воспользуемся соотношением (3). В результате потокосцепления статора и ротора предлагается оценивать с помощью схемы, представленной на рис. 3.

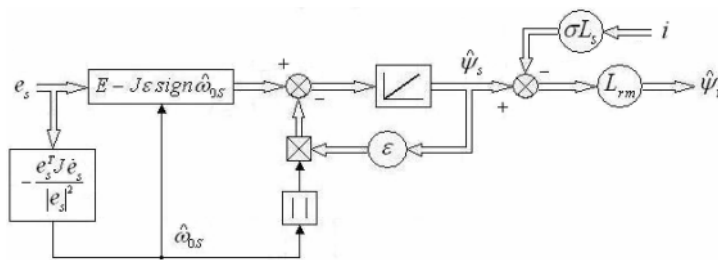


Рисунок 3

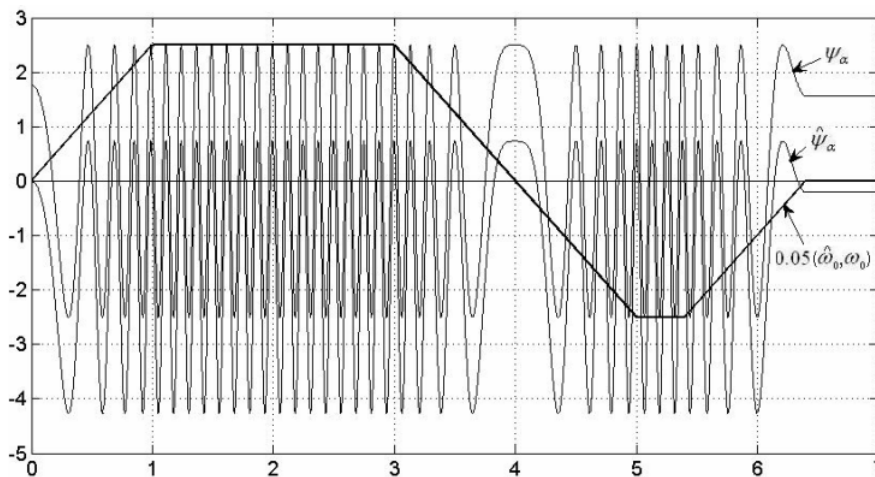


Рисунок 4 – Результаты моделирования чистого интегрирования

Следует отметить, что угловые скорости ротора и статора связаны соотношениями $\omega_{0s} \approx \omega_{0r}$ в переходных процессах и $\omega_{0s} = \omega_{0r}$ в стационарных процессах.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

На рис. 4–6 моделировалась следующая циклограмма работы: разгон в течение первой секунды до скорости 50 с^{-1} , на третьей секунде включался реверс до скорости -50 с^{-1} , затем осуществлялось торможение до нулевой скорости. На рис. 4 представлены результаты чистого интегрирования ЭДС. Рисунок свидетельствует о наличии в оценке потокосцепления постоянного смещения. Тем не менее, осуществляется точная оценка скорости. На рис. 5 показаны результаты моделирования процесса оценки скорости и вектора потокосцепления по соотношениям (10), (12) (на рисунке показаны только составляющие ψ_α , $\hat{\psi}_\alpha$ полностью характеризующие точность оценки всего вектора) и скорости его вращения. Переходный процесс оценки потокосцепления заканчивается через 0,5 с. Как видно на рисунке, во всех режимах работы, в том числе и при нулевой скорости, осуществляется точная оценка потокосцепления и его скорости.

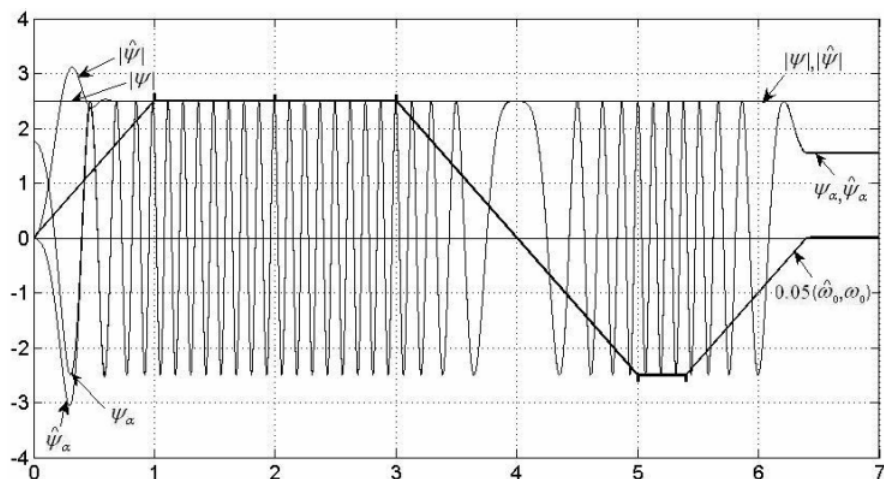


Рисунок 5 – Оценка скорости и потокосцепления при $\varepsilon = 0,5$

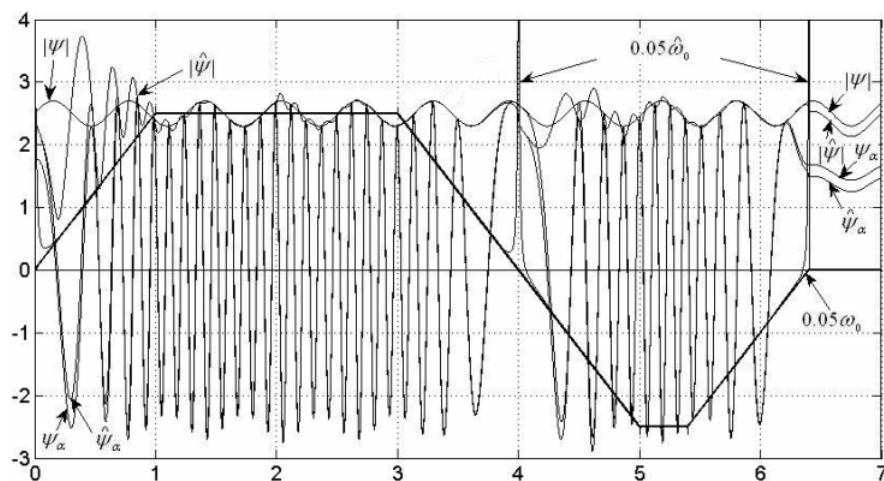


Рисунок 6 – Результаты моделирования оценки потокосцепления и его скорости вращения по выражениям (10), (12) при наличии инъекции гармонического сигнала в модуль потокосцепления

На рис. 6 показан процесс оценки вектора потокосцепления по соотношениям (10), (12) при наличии инъекции в модуль потокосцепления. В начальный момент коэффициент ε принимался равным 0,1, а на третьей секунде задавался равным 0,05. После переходного процесса длительностью одна секунда осуществляется точная оценка потокосцепления с инжектируемой составляющей. Высокая точность оценки скорости нарушается только при ее прохождении через нулевое значение.

ЗАКЛЮЧЕНИЕ

Предложен простой в вычислительном отношении, интегратор двухфазных сигналов, работоспособный при всех частотах, включая нулевую частоту. Этот метод применен для оценки потокосцеплений, соответ-

ствующих им синхронных скоростей и модулей потокосцеплений. Компьютерное моделирование показало высокую точность оценки указанных переменных при любых значениях скоростей и вполне приемлемое быстродействие. Данный метод точно воспроизводит переменную составляющую потокосцепления, обусловленную инъекцией синусоидального сигнала в намагничивающую составляющую статорного тока. Сравнение указанного метода с известными методами того же назначения показывает, что данный метод, будучи самым простым, по всем показателям превосходит известные методы.

ПЕРЕЧЕНЬ ССЫЛОК

1. Xu X., De Doncker R. and Novotny D. W. Stator Flux Orientation Control of Induction Machines in the Field Weake-

- ning Region. // IEE-IAS Annual Meeting Conf. Rec. (Pittsburgh, PA), Oct. – 1988. – P. 437–443.
2. *Hinkkanen M., Luomi J.* Modified Integrator for voltage model flux estimation of induction motors // IEEE Trans. on Industrial Electronics. – 2003. – V. 50, No. 4. – Pp. 818–820.
 3. *Jun Hu and Bin Wu.* New Integration Algorithms for Estimating Motor Flux over a Wide Speed Range // IEEE Transactions on Power Electronics. – 1998. – V. 13, No. 5. – Pp. 969–977.
 4. *Ohtani T., Tacada N. and Tanaka K.* Vector Control of Induction Motor without Shaft Encoder // IEEE Trans. Indust. Applic. – 1992. – V. 28, No. 1. – Pp. 157–164.
 5. *Holtz J.* Methods for Speed Sensorless Control of AC Drives // IEEE PCC–Yokohama. – 1993. – Pp. 415–420.
 6. *Holtz J., Quan J.* Drift- and parameter-compensated flux estimator for persistent zero-stator-frequency operation of sensorless-controlled induction motors // IEEE Trans.

on Industry Application. – 2003. – Vol. 39, No. 4. – Pp. 1052–1060.

Надійшла 2.04.2008

Розроблено новий алгоритм для оцінки потокозчеплень, відповідних їм синхронних швидкостей та модулів, працездатний при всіх частотах. Наведений інтегратор двофазних сигналів простий для розрахунку та забезпечує високу точність оцінки вказаних змінних.

New algorithm for flux estimations, corresponding they synchronous speeds and modules, which able to work on the all frequency was developed. Proposing integrator of two phase signals is the simple with respect to calculation and provide high precision of showing variable estimation.

УДК 681.527.2

А. В. Соломаха

ФОРМИРОВАНИЕ СИСТЕМЫ ВЕКТОРНОГО УПРАВЛЕНИЯ АСИНХРОННЫМ ПРИВОДОМ С КОМПЕНСАЦИЕЙ НЕОПРЕДЕЛЕННОСТЕЙ

Разработана система прямого векторного управления асинхронным приводом с компенсацией неопределенностей, работоспособная при всех скоростях ротора (включая нулевую). Численное моделирование подтвердило работоспособность предложенной системы.

ВВЕДЕНИЕ

Для качественного векторного управления (ВУ) асинхронным двигателем с короткозамкнутым ротором (АД) необходима информация о его параметрах. Параметрами модели АД являются индуктивности статора и ротора, взаимная индуктивность, активные сопротивления статора и ротора и момент инерции ротора. Наряду с активным сопротивлением ротора часто рассматривается постоянная времени ротора. Сопротивления в зависимости от температуры статора и ротора могут изменяться в два раза в большую и меньшую стороны по отношению к номинальным значениям. Данный вопрос рассматривается в работах [1–9]. Для оценки сопротивления ротора R_r в штатном режиме необходимо изменение модуля вектора потокосцепления ротора. С этой целью в намагничивающую составляющую тока статора инжектируют низкочастотную синусоидальную составляющую с малой амплитудой. При этом в потокосцеплении появляется переменная гармоническая составляющая, которая участвует в идентификации. Это накладывает дополнительное требование к методам оценки потокосцепления, а именно, до-

статочно точное воспроизведение инжектированной составляющей в оценке [10–13]. Сильным изменениям подвержен также момент инерции ротора, чему посвящены работы [14, 15]. Существенную роль также имеет неизвестная нагрузка АД и погрешности датчиков тока и напряжения [16–18]. Задача значительно усложняется, когда скорость ротора и потокосцепление АД не измеряются, а измеряются лишь клеммные токи и напряжения [1–13]. Названные неопределенности ухудшают динамические и статические характеристики систем управления и могут привести к потере их работоспособности.

В связи со спецификой данной работы сравнение методов одного и того же назначения приводятся ниже.

Целью данной статьи является формирование структуры системы ВУ асинхронным приводом с компенсацией неопределенностей, работоспособной во всем рабочем диапазоне скоростей ротора.

ВЫБОР БЛОКОВ СИСТЕМЫ ВУ

Для достижения простоты и удобства синтеза системы ВУ предпочтительно применять блочный метод синтеза. Разделение системы управления на блоки позволяет строить всю систему в каждом конкретном случае из различных блоков, исходя из требований точности, простоты алгоритмов, электропотребления, диапазона регулирования и т. п.

© Соломаха А. В., 2008

Формирователи электромагнитного момента и программных токов. Как показали исследования [14, 15], для формирования электромагнитного момента предпочтительным является использование компенсатора, состоящего из наблюдателя неопределенности и комбинированного регулятора. Согласно этим работам, неточность знания параметров АД, неидеальности ПЧ, наличие неизвестных перекрестных связей между каналами и другие неопределенности объединяются в одно слагаемое, которое компенсируется комбинированным регулятором. Компенсатор по сравнению с классическим ПИ-регулятором обеспечивает большее быстродействие при отсутствии перерегулирования, и является предпочтительным для системы ВУ.

Регулятор тока. В настоящее время в качестве регуляторов тока наибольшее распространение получили ПИ-регуляторы, установленные в трактах обоих контуров тока, описанных уравнениями в синхронном базисе, связанном с потокосцеплением ротора [19]. В работе [20] предложена система векторного регулирования статорного тока релейного типа, базирующаяся на регулировании в установленных токовых «коридорах», вследствие чего ток является пульсирующим. Система [20] сложна в вычислительном отношении, так как используются обратные тригонометрические функции и извлечение квадратного корня. Как показано в работах [21, 22], в замкнутых контурах тока с ПИ-регуляторами приходится находить компромисс между быстродействием, точностью, с одной стороны, и перерегулированием, с другой, вследствие чего быстродействие токовых контуров ограничено. Это является одним из основных недостатков, приписываемых ВУ по сравнению с прямым управлением моментом [1]. Неопределенности электродвигателя и преобразователя частоты существенно ухудшают динамические характеристики контуров тока с ПИ-регуляторами. Кроме того, между контурами токов имеют место перекрестные связи, пропорциональные не измеряемой синхронной скорости вектора потокосцепления ротора. В системах с ПИ-регуляторами для устранения влияния перекрестных связей приходится вводить компенсирующие перекрестные связи, усложняющие алгоритм управления. Этих недостатков лишены робастные компенсаторы токовых контуров [22, 23], состоящие из наблюдателя неопределенности и комбинированного регулятора. Именно их предпочтительно использовать в системе ВУ.

Компенсатор фильтра выходного напряжения ПЧ. Крутые фронты напряжения с ШИМ вызывают старение изоляции двигателя, питающих кабелей и ухудшение их диэлектрических свойств. Это также приводит к дополнительному нагреву обмоток двигателя и появлению высокочастотного акустического шума. Устранение вышеназванных эффектов особенно актуально для мощных приводов [24, 25]. Поэтому на выходе пре-

образователя частоты для сглаживания выходного напряжения устанавливают силовые LC-фильтры [26]. Пассивные фильтры улучшают гармонический состав тока и напряжения, уменьшают акустический шум, имеют простую конструкцию и относительно дешевые в изготовлении. Ограничение скорости изменения напряжения и снижение перенапряжений позволяет отказаться от увеличения номиналов инверторов за счет учета явления обратной волны в кабелях. Напряжения, получаемые после LC-фильтров, искажены по фазе и амплитуде относительно главных гармоник выходных напряжений ПЧ. Это приводит к ухудшению динамики электропривода. Для восстановления качества регулирования, с участием автора предложено использовать упреждающий компенсатор фильтра, описанный в работе [27].

Оценка угловой скорости ω и постоянной времени ротора T_r . В работах [1–6] с помощью адаптивных наблюдателей и фильтров Калмана седьмого порядка одновременно оцениваются вектор потокосцепления ротора ψ , токи статора i , угловая скорость ротора АД ω , активные сопротивления статора и ротора R_s, R_r . Однако, эти алгоритмы сложны. Это выдвигает повышенные требования к объему памяти и быстродействию вычислителя. Предпочтительным является метод работы [7] (выполненной с участием автора), работоспособный во всех режимах работы АД с любой нагрузкой во всем диапазоне скоростей ротора. Время идентификации T_r не превышает 10 с.

Оценка сопротивления статора R_s . Для качественного векторного управления асинхронным двигателем необходимо знание его параметров, в особенности, переменных вследствие изменения температуры АД сопротивлений ротора (R_r) и статора (R_s). В публикации [1] оцениваются ω и R_s . В работах [1–3, 6] оцениваются ω, R_s и R_r . Во всех упомянутых методах оценки, наряду с перечисленными параметрами, оцениваются необходимые для управления векторы потокосцепления ротора или статора. Рассмотренные в работах [1–6] методы оценки, в особенности те, которые оценивают ω, R_s и R_r , сложны. В частности, методы оценки работ [2, 4], где оцениваются ($i_\alpha, i_\beta, \psi_\alpha, \psi_\beta, \omega, R_s, R_r$) с помощью адаптивного наблюдателя седьмого порядка. Наблюдатель первого порядка для оценки R_s работы [9] отличается простотой, так как оценивает только сопротивление статора, поэтому является предпочтительным для системы ВУ.

Наблюдатель потокосцепления ротора. Как показали исследования [10], работоспособностью и удовлетворительным воспроизведением инжектированной составляющей отличаются методы оценки потокосцепления работ [11–13]. Исходя из анализа работы [10] и простоты реализации, предпочтительным является метод [12].

Оценка угловой скорости вектора потокосцепления ротора ω_0 . Метод оценки скорости, описанный в работе [9], основан на использовании реактивной мощности Q . Он показал хорошие результаты на низких скоростях. Известно, что противоЭДС двигателя определяется выражением

$$e_r = u - R_s i - \sigma L_s \dot{i}, \quad (1)$$

где u – вектор напряжения статора, i – вектор ток статора, σ – коэффициент рассеивания, L_s – индуктивность статора. При больших скоростях из-за противоЭДС напряжение растет, а ток уменьшается. В этом случае на основании (1) можно полагать $e_r \approx u \approx e_s$. Тогда в соответствии с известным выражением

$$\hat{\omega}_0(u) = \frac{u_\alpha \dot{u}_\beta - u_\beta \dot{u}_\alpha}{u_\alpha^2 + u_\beta^2} \quad (2)$$

можно полагать $\omega_{0s} \approx \omega_0(u) \approx \omega_{0r}$, где ω_{0s} , ω_{0r} – угловые скорости векторов потокосцепления статора и ротора соответственно. Поскольку $\omega_0 \approx \omega_0(u) \approx \omega_0(i)$ в переходных режимах и $\omega_0 = \omega_0(u) = \omega_0(i)$ в установившихся, то выражение (2) можно отождествлять со скоростью вращения вектора потокосцепления ротора. Выражение (2) хорошо известно, отличается высокой точностью определения ω_0 в диапазоне высоких и средних скоростей ротора двигателя.

Совместим достоинства этих методов путем их сопряжения с помощью весового коэффициента k . Будем угловую скорость вектора потокосцепления определять выражением

$$\hat{\omega}_0 = \hat{\omega}_0(Q) \cdot (1 - k) + \hat{\omega}_0(u) \cdot k, \quad (3)$$

где $\hat{\omega}_0(Q)$ – оценка, полученная из выражения реактивной мощности Q работы [9], k – весовой коэффициент, который определяется из условий

$$k = \begin{cases} 0, & |\hat{\omega}_0| < \omega_{low}, \\ 1, & |\hat{\omega}_0| > \omega_{high}, \\ \frac{|\hat{\omega}_0| - \omega_{low}}{\omega_{high} - \omega_{low}}, & \end{cases} \quad (4)$$

где ω_{low} , ω_{high} – нижняя и верхняя границы диапазона сопряжения соответственно.

Оценка постоянных смещений. Цепи электропривода, состоящие из датчиков сигналов, аналоговых фильтров и преобразователей «аналог – код», могут привносить в измеряемые сигналы неизвестные постоянные ошибки, которые ухудшают качество регулирования, а порой могут вызвать неустойчивость всей системы в целом. В работе [16], выполненной с участием автора, предложены методы получения гармонического

сигнала без постоянной составляющей, которые используем в системе управления.

Компенсаторы искажений фильтров измеряемых сигналов. Установка аналоговых RC-фильтров в каналах измерения приводит к искажению амплитуд и фаз главных гармоник. Поэтому в методах управления используют программные значения напряжений, по сигналам которых осуществляется управление блоком ШИМ. Однако, из-за запаздываний открытия и закрытия транзисторов блока ШИМ, падений напряжений на этих транзисторах, а также мертвого времени возникают дополнительные погрешности в коэффициентах усиления блока ШИМ, которые в методах не учитываются. Все это существенно ухудшает показатели качества управления, в особенности, на низких частотах. Фильтры [17], выделяя главные гармоники, не дают искажений их фаз и амплитуд. Практические испытания и анализ результатов работы этих фильтров приведены в работе [18].

На рис. 1 представлена функциональная схема ВУ асинхронным приводом с компенсацией неопределенностей. Собственно система ВУ состоит из следующих блоков: формирователь электромагнитного момента и программных токов, регулятор тока, компенсатор пассивного силового фильтра (ПСФ), блок оценки угловой скорости ω и постоянной времени ротора T_r , блок оценки сопротивления статора, наблюдатель потокосцепления ротора, блок оценки угловой скорости вектора потокосцепления ротора, блок оценки постоянных смещений, фильтры с компенсаторами. Задача состоит в выборе элементов схемы, исходя из критериев простоты алгоритмов, диапазона, в котором метод работоспособен.

ОПИСАНИЕ РАБОТЫ СИСТЕМЫ ВУ

Сигналы с датчиков тока и напряжения, установленных в двух фазах, пропускаются через аналоговые фильтры 1 и АЦП. Полученные сигналы проходят через блок 2, где компенсируются постоянные составляющие, и подаются в блок 3. В блоке 3 осуществляется необходимая коррекция сигналов по фазе и амплитуде, после чего, на выходе блока получаем неискаженные главные гармоники статорных тока и напряжения. С помощью преобразований [28] трехфазные токи и напряжения приводятся к двухфазным, которые и используются в системе ВУ, а именно в блоках 4–7, 11. Далее рассчитываются модуль и угловая скорость вектора потокосцепления ротора в блоках 4 и 5 соответственно. При этом информация о сопротивлении статора рассчитывается в штатном режиме работы привода в блоке 6 и участвует в расчете модуля потокосцепления. Оценку скорости ротора можно получать с помощью адаптивного алгоритма оценки скорости вращения и постоянной времени ротора (блок 7) пря-

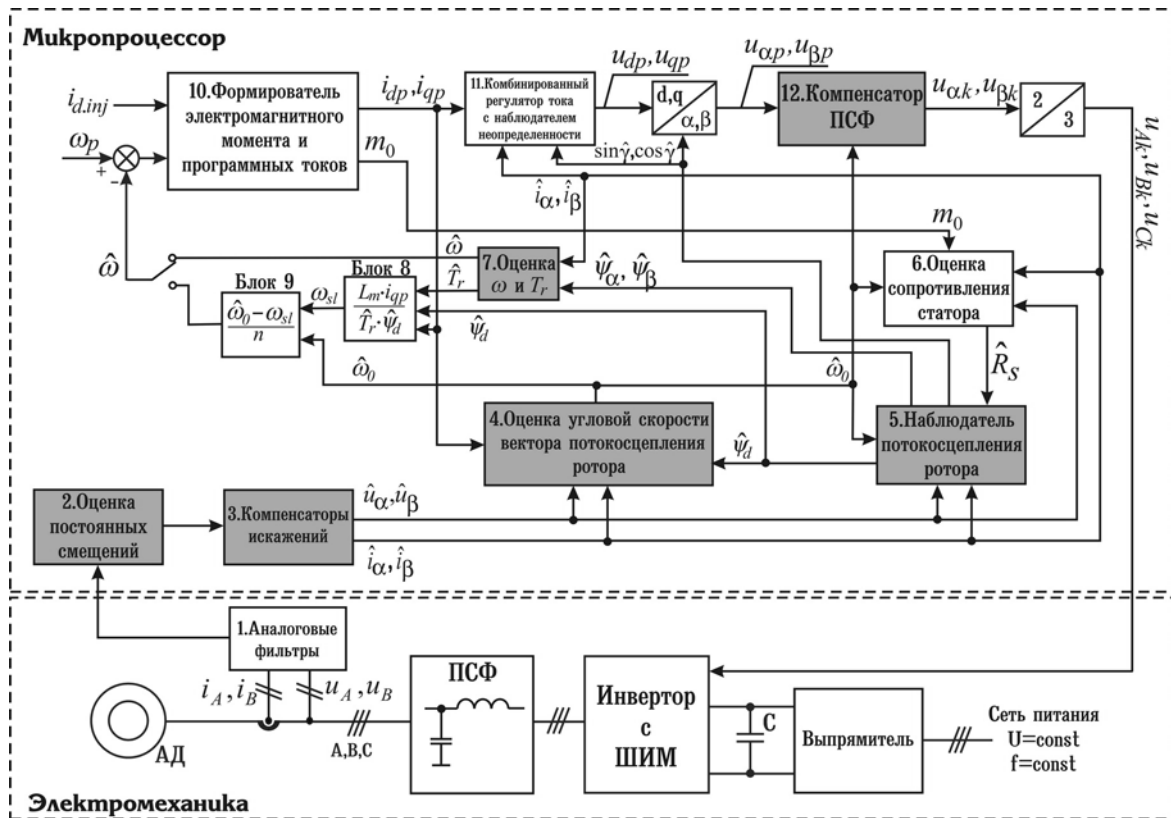


Рисунок 1 – Функциональная схема ВУ асинхронным приводом с компенсацией неопределенностей

мо или через блоки 7–9. Расчет робастного по отношению к изменению момента инерции электромагнитного момента двигателя и программных составляющих стартовых токов в стационарном базисе с учетом задания программной угловой скорости вращения ω_p происходит в блоке 10. Далее программные токи и ток инжекции подаются в блок 11 – на комбинированный регулятор тока с наблюдателем неопределенности. Неточность знания параметров АД, неидеальности ПЧ, наличие неизвестных перекрестных связей между каналами и другие неопределенности объединяются в одно слабое, которое компенсируется комбинированным регулятором. Компенсатор по сравнению с классическим ПИ-регулятором обеспечивает большее быстродействие при отсутствии перерегулирования. Полученные программные сигналы напряжения пропускаются через упреждающий компенсатор пассивного фильтра (блок 12), при наличии такового на выходе ПЧ. С помощью компенсатора сигналы корректируются с учетом вносимых фильтром искажений. Далее двухфазные программные сигналы напряжения преобразуются в трехфазные блоком «2/3» и подаются в блок формирования ШИМ в качестве управляющих сигналов.

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

При моделировании динамических процессов на ПЭВМ рассматривался АД типа А2-81-4 (с номинальной скоростью ротора 1460 об/мин (около 150 рад/с) мощностью 32 кВт и номинальным моментом нагрузки 262 Нм) и рассмотренная выше система ВУ. В алгоритмы были введены ограничения тока и воспроизведены эффекты ШИМ. Алгоритмы управления моделировались в дискретном виде с тактом счета 200 мкс. На рис. 2, 3 показаны процессы управления только по клеммным измерениям (без использования датчиков скорости и потокосцепления).

Моделировались следующие режимы работы. Рис. 2, а иллюстрирует режим отработки приводом заданной скорости ω_p , а рис. 2, б – нулевой. В период от пуска до 2 с осуществлялось намагничивание двигателя, после чего осуществлялся разгон до заданной скорости. С 4-й по 8-ю секунды к двигателю был приложен номинальный момент нагрузки. На 10-й секунде включался реверс скорости. С 14-й по 18-ю секунды к двигателю был приложен момент, однонаправленный с моментом двигателя (генераторный режим), равный 100 Н·м (0,38 $M_{ном}$). Задания на скорость составляли 150, 50, 1 рад/с (соответствуют 100 %, 33,3 %, 0,67 % от номинальной скорости) и характеризуют высокие,

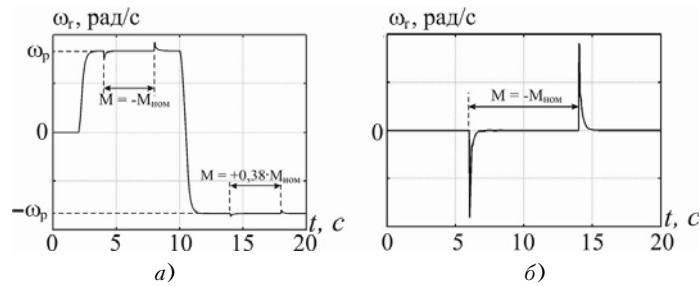


Рисунок 2 – Режимы работы привода

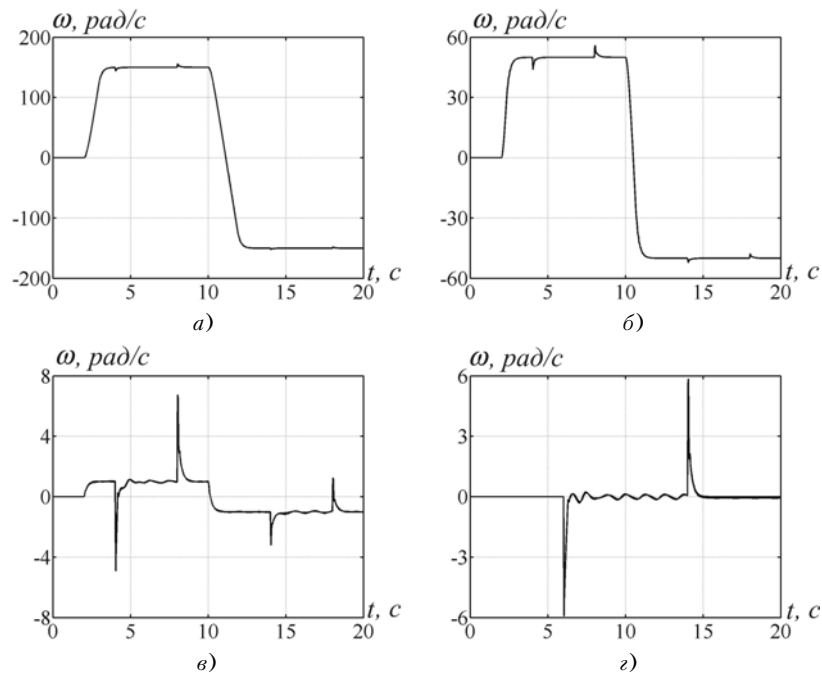


Рисунок 3 – Результаты моделирования

средние и низкие частоты вращения ротора двигателя. Результаты работы электропривода с предложенной системой векторного управления показаны на рис. 3, а, б, в соответственно. Для работы при заданной нулевой скорости вращения ротора был задан следующий режим. В период до 2 с осуществлялось намагничивание двигателя, затем с 6 по 14 секунды был приложен номинальный момент. Результаты моделирования для этого режима показаны на рис. 3, г. Как видно из рисунков, оценка и полученные из модели АД угловые скорости ротора практически совпадают. Это говорит о хорошем качестве регулирования.

ЗАКЛЮЧЕНИЕ

Разработана система прямого векторного управления асинхронным приводом с компенсацией неопределенностей сопротивлений статора, постоянной времени

ротора, момента инерции привода. Разработанная система компенсирует постоянные составляющие в сигналах датчиков, искажения, вносимые аналоговыми фильтрами датчиков и пассивным LC-фильтром на выходе ПЧ. Численное моделирование подтвердило работоспособность и показало высокое качество регулирования предложенной системы при всех скоростях ротора (включая нулевую).

ПЕРЕЧЕНЬ ССЫЛОК

1. Vas P. Sensorless vector and direct torque control. – Oxford, New York, Tokyo: Oxford University Press, 1998. – 729 с.
2. Tajima H., Guidi G., Umida H. Consideration about Problems and Solutions of Speed Estimation Method and Parameter Tuning for Speed-Sensorless Vector Control of Induction Motor Drives // IEEE Trans. Industry Applications. – 2002. – Vol. 38, No. 5. – Pp. 1282–1289.
3. Kubota H., Matsuse K. Speed Sensorless Field-Oriented Control of Induction Motor with Rotor Resistance Adapta-

- tion // IEEE Trans. Industry Applications. – 1994. – Vol. 30, No. 5. – Pp. 1219–1224.
4. Matsuse K., Taniguchi Sh., Yoshizumi T., Namiki K. A Speed-Sensorless Vector Control of Induction Motor Operating at High Efficiency Taking Core Loss into Account // IEEE Trans. Ind. Appl. – 2001. – Vol. 37, No. 2. – P. 548–557.
 5. Kubota H., Sato I., Tomura Y., Matsuse K., Ohta H., Hori Y. Regenerating-Mode Low-Speed Operation of Sensorless Induction Motor Drive with Adaptive Observer // IEEE Trans. Ind. Applications. – 2002. – Vol. 38, No. 4. – Pp. 1081–1086.
 6. Ha I.-J., Lee S.-H. An Online Identification Method for both Stator and Rotor Resistances of Induction Motor without Rotational Transducers // IEEE Trans. On Industrial Electronics. – 2000. – Vol. 47, No. 4. – Pp.842–852.
 7. Потапенко Е. М., Потапенко Е. Е., Соломаха А. В. Определение скорости и постоянной времени ротора асинхронного двигателя с помощью адаптивного наблюдателя // Матеріали 12-ї міжнародної конференції з автоматичного управління «Автоматика-2005», 30 травня – 3 червня 2005 р. Т. другий. – Харків: НТУ «ХПІ». – 2005. – С. 123–124.
 8. Beguenane R., Ouhrouche M. MRAC-IFO induction motor control with simultaneous velocity and rotor-inverse time constant estimation // Proceeding of IASTED International Conference PES'2003. – 2003. – 8 pp.
 9. Потапенко Е. М., Потапенко Е. Е. Синтез инвариантных и адаптивных к изменению сопротивления статора алгоритмов векторного управления асинхронным двигателем // Проблемы управления и информатики. – 2007. – № 2. – С. 16–29.
 10. Потапенко Е. М., Соломаха А. В. Сравнение перспективных методов оценки вектора потокоцепления асинхронного двигателя // Сборник научных трудов Днепропетровского государственного технического университета (технические науки). Тематический выпуск «Проблемы автоматизированного электропривода. Теория и практика»/ Днепропетровск: ДГТУ, 2007. – С. 415–417.
 11. Hu J., Wu B. New Integration Algorithms for Estimating Motor Flux over a Wide Speed Range // IEEE Transactions on Power Electronics. – 1998. – Vol. 13, No. 5. – P. 969–977.
 12. Ohtani T., Tacada N., Tanaka K. Vector Control of Induction Motor without Shaft Encoder // IEEE Trans. Indust. Applic. – 1992. – Vol. 28, No. 1. – P. 157–164.
 13. Holtz J., Quan J. Sensorless Vector Control of Induction Motors at Very Low Speed Using a Nonlinear Inverter Model and Parameter Identification // IEEE Trans. Indust. Applic. – 2002. – Vol. 38, No. 4. – P. 1087–1095.
 14. Потапенко Е. Е., Потапенко Е. М. Энергосберегающее робастное управление асинхронными приводами // Механіка та машинобудування. – 1999. – № 1. – С. 106–111.
 15. Потапенко Е. Е., Потапенко Е. М. Синтез экстремального робастного управления асинхронным приводом // Технічна електродинаміка. Тематичний випуск: Проблеми сучасної електротехніки. – Ч. 6. – Київ: ІЕ НАНУ. – 2000. – С. 34–37.
 16. Потапенко Е. М., Потапенко Е. Е., Соломаха А. В. Синтез и анализ методов фильтрации многофазных неопределенных сигналов // Проблемы управления и информатики. – 2007. – № 6. – С. 1–10.
 17. Потапенко Е. Е., Потапенко Е. М. Синтез и анализ компенсационных фильтров многофазных неопределенных сигналов // Вісник Національного технічного університету «ХПІ». Тематичний випуск «Проблеми автоматизованого електропривода. Теорія і практика». Харків: НТУ «ХПІ». – 2003. – № 10, т. 2. – С. 342–344.
 18. Потапенко Е. М., Соломаха А. В., Савранская А. В. Экспериментальное исследование аналого-цифровых адаптивных фильтров трехфазных сигналов // Електромашинобудування та електрообладнання. – 2006. – № 66. – С. 271, 272.
 19. Novotny D. W., Lipo T. A. Vector control and dynamics of AC drives. – Oxford: Oxford University Press, 1996. – 440 p.
 20. Волков А. В., Антонов Н. Л. Быстродействующее векторное регулирование статорного тока в асинхронных электроприводах с двухзвенным непосредственным преобразователем частоты // Электромашиностроение и электрооборудование. – К.: Техніка. – 2004. – № 62. – С. 21–24.
 21. Leonhard W. Control of electrical drives. – Berlin: Springer, 1996. – 420 p.
 22. Потапенко Е. М., Даниличев Д. С., Потапенко Е. Е., Би чай В. Г., Васильева Е. В. Оптимизация и исследование динамики токовых контуров системы векторного управления асинхронным двигателем // Електротехніка і електроенергетика. – 2003. – № 1. – С. 42–47.
 23. Потапенко Е. М., Потапенко Е. Е. Синтез робастных комбинированных регуляторов токовых контуров при векторном управлении двигателями переменного тока // Проблемы управления и информатики. – 2006. – № 6. – С. 69–78.
 24. Das J. C. Passive Filters – Potentialities and Limitation // IEEE Trans. Ind. Applicat. – Jan./Febr. 2004. – Vol. 40, No. 1. – Pp. 232–241.
 25. Kojima M., Hirabayashi K., Kawabata Y., Ejiogu E. C., Kawabata T. Novel Vector Control System Using Dead-beat-Controlled PWM Inverter With Output LC Filter // IEEE Trans. Ind. Applicat. – Jan./Febr. 2004. – Vol. 40, No. 1. – Pp. 162–169.
 26. Salomaki J., Hinkkanen M., Luomi J. Sensorless Control of Induction Motor Drives Equipped With Inverter Output Filter // IEEE Trans. on Ind. Electronics. – 2006. – Vol. 53, No. 4. – Pp. 1188–1197.
 27. Потапенко Е. М., Соломаха А. В. Синтез и анализ компенсатора пассивного LC-фильтра на выходе преобразователя частоты // Радиоэлектроника. Информатика. Управление. – 2007. – № 2. – С. 12–17.
 28. Чиликин М. Г., Ключев В. И., Сандлер А. С. Теория автоматизированного электропривода. – М.: Энергия, 1979. – 616 с.

Надійшла 23.11.2007

Розроблена система прямого векторного керування асинхронним приводом з компенсацією невизначеностей, яка працездатна при всіх швидкостях ротора (включаючи нульову). Чисельне моделювання підтвердило працездатність запропонованої системи.

The system of direct vector control by induction drive with indeterminacy compensation, efficient in all rotor speed range (including zero speed) is made. Numerical modelling of serviceability of the offered system has confirmed.

АВТОРЫ НОМЕРА

Аксак Наталья Георгиевна

кандидат технических наук, старший научный сотрудник Харьковского национального университета радиоэлектроники

Антропов Олег Сергеевич

старший преподаватель Днепропетровского национального университета

Артюшенко Богдан Андреевич

аспирант Запорожского национального технического университета

Белая Наталья Ивановна

кандидат физико-математических наук, доцент Запорожского национального технического университета

Бодянский Евгений Владимирович

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Бондарев Виктор Павлович

кандидат физико-математических наук, доцент Запорожского национального технического университета

Бондаренко Лариса Алексеевна

старший преподаватель Запорожского национального технического университета

Борулько Валентин Федорович

кандидат физико-математических наук, старший научный сотрудник Днепропетровского национального университета

Васильева Евгения Владимировна

старший преподаватель Запорожского национального технического университета

Викторов Евгений Александрович

аспирант Харьковского национального университета радиоэлектроники

Вовк Сергей Михайлович

кандидат физико-математических наук, заведующий кафедрой Днепропетровского национального университета

Гапоненко Николай Прокофьевич

кандидат технических наук, доцент Запорожского национального технического университета

Гребенюк Сергей Николаевич

кандидат технических наук, доцент Запорожского национального университета

Довбня Анатолий Николаевич

доктор физико-математических наук, директор Института физики высоких энергий и ядерной физики ННЦ «Харьковский физико-технический институт»

Дубровин Валерий Иванович

кандидат технических наук, доцент Запорожского национального технического университета

Душинова Евгения Викторовна

студентка Запорожского национального технического университета

Зайцев Сергей Алексеевич

студент Запорожского национального технического университета

Зиноватная Светлана Леонидовна

старший преподаватель Одесского национального политехнического университета

Ивочка Наталья Васильевна

студентка Запорожского национального технического университета

Казурова Алина Евгеньевна

аспирант Запорожского национального технического университета

:

Камышанов Анатолий Федорович

доцент Запорожского национального технического университета

Карпуков Леонид Матвеевич

доктор технических наук, заведующий кафедрой Запорожского национального технического университета

Киприч Татьяна Витальевна

аспирант Запорожского национального технического университета

Коваленко Дарья Аркадьевна

аспирант Запорожского национального технического университета

Козина Галина Леонидовна

кандидат физико-математических наук, доцент Запорожского национального технического университета

Корниенко Сергей Константинович,

кандидат технических наук, доцент Запорожского национального технического университета

Корольков Роман Юрьевич

ассистент Запорожского национального технического университета

Кравивной Александр Викторович

кандидат физико-математических наук, доцент Запорожского национального технического университета

Крисан Алексей Александрович

инженер ОАО НИИ «Преобразователь», г. Запорожье

Крисан Юрий Алексеевич

кандидат технических наук, доцент Запорожского национального технического университета

Кришук Владимир Николаевич

кандидат технических наук, заведующий кафедрой Запорожского национального технического университета

Кунгурцев Алексей Борисович

кандидат технических наук, профессор Одесского национального политехнического университета

Левыкина Виктория Ивановна

ассистент Запорожского национального технического университета

Лобов Вячеслав Иосифович

кандидат технических наук, доцент Криворожского технического университета

Молдовян Николай Андреевич

доктор технических наук, главный научный сотрудник Научного филиала ФГУП НИИ «Вектор» – Специализированного центра программных систем «СПЕКТР», г. Санкт-Петербург, Россия

Мороз Владимир Иванович

кандидат технических наук, докторант Национального университета «Львовская политехника»

Неласая Анна Викторовна

старший преподаватель Запорожского национального технического университета

Нестеренко Борис Борисович

доктор технических наук, заместитель директора Института математики НАН Украины, г. Киев

Нечипоренко Нина Александровна

кандидат физико-математических наук, доцент Запорожского национального технического университета

Новосельцев Игорь Валериевич

аспирант Харьковского национального университета радиоэлектроники

Новотарский Михаил Анатольевич

кандидат технических наук, старший научный сотрудник Института математики НАН Украины, г. Киев

Овский Александр Геннадьевич

аспирант Запорожского национального университета

Олейник Андрей Александрович

аспирант Запорожского национального технического университета

Пица Дмитрий Макарович

доктор технических наук, проректор Запорожского национального технического университета

Прохорец Иван Михайлович

кандидат физико-математических наук, старший научный сотрудник ННЦ «Харьковский физико-технический институт»

Решевская Екатерина Сергеевна

ассистент Запорожского национального университета

Рубан Сергей Анатольевич

ассистент Криворожского технического университета

Самойлик Сергей Сергеевич

аспирант Запорожского национального технического университета

Слепцов Валерий Иванович

старший преподаватель Запорожского национального технического университета

Соломаха Александр Викторович

аспирант Запорожского национального технического университета

Степаненко Александр Алексеевич

ассистент Запорожского национального технического университета

Субботин Сергей Александрович

кандидат технических наук, доцент Запорожского национального технического университета

Толок Вячеслав Александрович

доктор технических наук, заведующий кафедрой Запорожского национального университета

Удовенко Сергей Григорьевич

доктор технических наук, профессор Харьковского национального университета радиоэлектроники

Чумаченко Виталий Павлович

доктор физико-математических наук, заведующий кафедрой Запорожского национального технического университета

Чумаченко Ярослав Виталиевич

кандидат физико-математических наук, доцент Запорожского национального технического университета

Шамраев Анатолий Анатольевич

кандидат технических наук, старший преподаватель Харьковского национального университета радиоэлектроники

Шило Галина Николаевна

кандидат технических наук, доцент Запорожского национального технического университета

Щербаков Адольф Николаевич

кандидат технических наук, доцент Запорожского национального технического университета

Юрич Мария Юрьевна

аспирант Запорожского национального технического университета

Юськив Олеся Игоревна

студентка Запорожского национального технического университета

ДЛЯ СВЕДЕНИЯ АВТОРОВ

Журнал «Радиоэлектроника. Информатика. Управління» предназначен для публикации наиболее значимых научных и практических результатов исследований ученых высших учебных заведений и научных организаций.

Журнал включен в перечень научных изданий Украины, в которых могут публиковаться результаты диссертационных работ на соискание ученых степеней доктора и кандидата технических наук.

Статьи, опубликованные в журнале, реферируются в реферативных журналах и базах данных ВИНИТИ (Россия) и «Джерело» (Украина). Рефераты статей размещаются на web-сайте Запорожского национального технического университета (<http://www.zntu.edu.ua/base/news2005/ric.htm>).

Журнал издается два раза в год и распространяется по подписке в Украине. Подписной индекс 22914.

Последовательность размещения материала статьи: индекс УДК, название статьи, инициалы и фамилия автора(ов), аннотация, текст, перечень ссылок. Рукопись статьи должна быть подписана всеми авторами.

Текст статьи

Принимаются статьи на русском, украинском и английском языках. Размер статьи до 0,5 авторского листа.

Структура статьи

Редакционная коллегия принимает к рассмотрению только научные статьи, которые содержат такие необходимые элементы: постановка проблемы в общем виде и ее связь с важными научными или практическими заданиями; анализ последних исследований и публикаций, в которых начато решение данной проблемы, и на которые опирается автор; выделение нерешенных ранее частей общей проблемы, которым посвящается статья; формулирование целей статьи (постановка задачи); изложение основного материала исследования с полным обоснованием полученных научных результатов, выводы по данному исследованию и перспективы дальнейших исследований в данном направлении.

Требования к оформлению

Статьи принимаются набранные в редакторе Word for Windows (v.6 и выше).

Параметры страницы:

- размер бумаги – А4 (210x297);
- ориентация – книжная;
- шрифт – Times New Roman Cyr, размер – 12 pt;
- междустрочный интервал – полуторный;
- верхнее поле – 20мм, нижнее – 20 мм, левое – 25 мм, правое – 15 мм.

Аннотация

Аннотация является обязательной и подается в начале статьи на русском, украинском (для граждан Ук-

раины) и английском языках. Объем аннотации не должен превышать 40 слов.

Рисунки

Рисунки помещаются в тексте и подаются в отдельных файлах (формат.TIF с разрешением не менее 200 dpi). Рисунки нумеруют и подписывают внизу.

Таблицы

Заголовки таблиц размещаются по центру.

Формулы

Формулы выполняются с помощью встроенного в Word for Windows редактора Microsoft Equation. Формулы нумеруют в круглых скобках справа:

$$Z(\Theta) = 101 \log \left(\frac{\bar{y}^2}{s^2} \right) \quad (3)$$

Формулы большого размера записываются в несколько строк.

Перечень ссылок

Перечень ссылок в конце рукописи подается языком оригинала и составляется в порядке упоминания ссылок в тексте и в соответствии со стандартом на библиографическое описание. Ссылки на литературу в тексте обозначаются цифрой в квадратных скобках.

В сведениях об авторах необходимо привести фамилию, имя, отчество (полностью), место работы, должность, ученую степень, рабочий и домашний адрес, телефоны, e-mail.

В редакцию журнала необходимо представить:

- 1) распечатанный текст статьи в 2-х экземплярах;
- 2) экспертное обоснование о возможности опубликования;
- 3) сведения об авторах;
- 4) сопроводительное письмо-ходатайство из организации, где была выполнена работа;
- 5) название статьи, фамилии, имена и отчества авторов на английском языке;
- 6) дискета 3,5' с текстом документа, файлами рисунков, данными о статье на английском языке и сведениями об авторах в формате Word for Windows (v.6 и выше). Файлы можно выслать электронной почтой в виде заархивированных (ZIP-архиватором) файлов.

Статьи, которые не отвечают указанным требованиям, не принимаются к рассмотрению.

Рукописи и дискеты не возвращаются.

Адрес редакции: 69063, г. Запорожье, ул. Жуковского, 64, ЗНТУ, редакция журнала "PIU"

Тел.:

(061) 7-698-2-96 – редакционно-издательский отдел

(061) 7-698-3-13 – главный редактор

(061) 7-698-2-96 – зам. главного редактора

Факс: (0612) 64-21-41

E-mail: rvv@zntu.edu.ua, TQM@zntu.edu.ua

Наукове видання

**Радіоелектроніка.
Інформатика.
Управління**

№ 1 / 2008

Науковий журнал

Головний редактор – д. т. н. Потапенко Є. М.
Заст. головного редактора – к. т. н. Дубровін В. І.

Комп'ютерне моделювання та верстання

Новікова К. О.

Оригінал-макет підготовлено у редакційно-видавничому відділі ЗНТУ

*Підписано до друку 01.07.2008. Формат 60x84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 18,4.
Тираж 300 прим. Зам. № 1169.*

69063 м. Запоріжжя, ЗНТУ, друкарня, вул. Жуковського, 64