

АКАДЕМИЯ НАУК УКРАИНСКОЙ ССР
ИНСТИТУТ КИБЕРНЕТИКИ
им. В. М. ГЛУШКОВА

*A.Г.Ивахненко
B.С.Стешко*

**ПОМЕХО-
УСТОЙЧИВОСТЬ
МОДЕЛИРОВАНИЯ**

**Помехоустойчивость моделирования / Ивахненко А. Г., Степашко В. С. - Киев:
Наук. думка, 1985. - 216 с.**

В монографии рассмотрена проблема моделирования реальных объектов и процессов по данным наблюдений в условиях помех. Основное внимание уделено задаче повышения помехоустойчивости и эффективности алгоритмов структурного моделирования, построенных на принципе самоорганизации моделей на ЭВМ. Изложены основные структурные особенности и способы организации вычислений в этих алгоритмах. Исследована помехоустойчивость одно- и двухуровневых алгоритмов комбинаторного типа с применением различных внешних критериев селекции моделей. Описаны алгоритмы самоорганизации моделей с повышенной помехоустойчивостью, реализующие идеи общей теории связи.

Для специалистов в области моделирования сложных систем.

Программы и статьи по МГУА доступны на <http://www.GMDH.net>

E-mail авторов: lvakhnenko@gmdh.net, Stepashko@gmdh.net

Редакция физико-математической литературы

и 1502000000-087
M221(04)-85 124-85

© Издательство «Наукова думка», 1985

ОГЛАВЛЕНИЕ

Предисловие	6
Основные обозначения	9
Глава 1	
Развитие помехоустойчивых методов моделирования по данным наблюдений	11
1. Методы моделирования, порожденные подходами теорий информации и автоматического управления	11
2. Самоорганизация моделей на ЭВМ. Метод группового учета аргументов	13
3. Обзор развития МГУА	15
4. Общая характеристика структуры алгоритмов МГУА	22
4.1. Однорядные (комбинаторные) алгоритмы	24
4.2. Многорядные алгоритмы без вычисления остатков	27
4.3. Многорядные алгоритмы МГУА с вычислением остатков	31
4.4. Объективный системный анализ в задачах, где объект описывается системой уравнений	33
5. Особенности МГУА, объясняющие его эффективность в задачах моделирования и долгосрочного количественного прогноза	33
6. Области применения различных алгоритмов МГУА	40
7. Основные направления будущей работы	42
Глава 2	
Самоорганизация моделей на ЭВМ в терминах общей теории связи (теории информации)	45
1. Аналогия структурных схем системы связи, вычислительного эксперимента и системы самоорганизации модели	46
2. Основные понятия теории информации и их аналоги в теории самоорганизации моделей	48
3. Вторая (предельная) теорема Шеннона	53
4. Применение геометрического построения Шеннона при самоорганизации моделей для распознавания образов, долгосрочного прогноза и векторной оптимизации	54
5. Линейная обработка сигналов и помехоустойчивость связи	60
6. Самоорганизация в терминах ЛОС	61
7. Новые критерии селекции, построенные на основе ЛОС	62
8. Конструктивность аналогии задач теории информации и самоорганизации	65

Глава 3

Теория комбинаторного алгоритма МГУА

1. Задача структурного моделирования, решаемая по МГУА.....	67
2. Структура комбинаторного алгоритма.....	67
3. Уменьшение затрат памяти и времени счета в алгоритмах МГУА с применением матриц нормальных уравнений.....	69
4. Рекуррентное вычисление коэффициентов моделей в комбинаторном алгоритме.....	73
5. Конечная селекционная процедура сокращения объема полного перебора моделей.....	78
6. Универсальная схема перебора моделей на основе управляемого комбинаторного алгоритма.....	84
7. Применение комбинаторных алгоритмов МГУА.....	90

Глава 4

Потенциальная помехоустойчивость моделирования по комбинаторному алгоритму МГУА

1. Задача оценки помехоустойчивости моделирования.....	95
2. Особенности моделирования в условиях помех.....	95
3. Идеальный внешний критерий.....	96
4. Селективные свойства идеального внешнего критерия.....	99
5. Потенциальная помехоустойчивость моделирования.....	102
6. Обсуждение результатов исследования помехоустойчивости.....	111
7. Приложения.....	114

Глава 5

Внешние критерии селекции и исследование помехоустойчивости моделирования

1. О внутренних и внешних критериях.....	118
2. Классификация основных внешних критериев.....	118
3. Связь внешних критериев.....	119
4. Каноническая форма квадратичных критериев.....	127
5. Анализ помехоустойчивости индивидуальных критериев селекции.....	130
6. Численное исследование помехоустойчивости одно- и двухкритериальной селекции моделей.....	132
7. Обсуждение результатов сравнительного исследования внешних критериев.....	138

Глава 6

Помехоустойчивое моделирование при наличии информации о керреляционных, частотных и временных характеристиках помех

1. Помехоустойчивый алгоритм самоорганизации моделей по принципу корреляционного приемника.....	149
2. Сведение задачи обратного преобразования к решению системы нелинейных уравнений.....	149
3. Влияние частотных помех на результат моделирования.....	152
4. Предварительная фильтрация частотных помех с помощью скользящего усреднения.....	154
5. Учет информации о распределении помех во времени.....	160
6. Эффективность моделирования при учете характеристик помех.....	165

Глава 7

Помехоустойчивость двухуровневых алгоритмов долгосрочного прогнозирования

1. Многоуровневое моделирование на нескольких языках различной степени общности.....	170
2. Основные формы критерия баланса прогнозов.....	170

3. Помехоустойчивость выбора моделей по критерию баланса прогнозов.....	177
4. Структура двухуровневого алгоритма самоорганизации разностных прогнозирующих моделей.....	186
5. Особенности двухуровневых прогнозирующих моделей, построенных по критерию баланса.....	192
6. Анализ устойчивости двухуровневых разностных прогнозирующих моделей.....	196
7. Обсуждение результатов анализа алгоритмов двухуровневого прогнозирования.....	201

Приложение.....	203
Список литературы.....	206

Во многих отраслях науки ученые сталкиваются с необходимостью коренного пересмотра своих взглядов и устремлений. Наиболее известным примером является «неожиданность» простой мысли о том, что наша планета и ее ресурсы имеют ограниченные размеры и, следовательно, пора подумать об оптимальном сценарии их расходования и, возможно, частичного обновления. При этом надежды возлагаются на математическое моделирование глобальных экологических и экономических процессов, которое должно создать общественное понимание (по выражению В. И. Вернадского) указанной проблемы.

Но и в самом моделировании требуется существенный пересмотр взглядов. До сих пор при так называемом системном анализе без всяких сомнений полагали, что всемерное расширение границ моделируемой системы и учет разного рода уточнений внутри нее являются основным путем повышения точности прогнозирования будущего. Так, Дж. Форрестер в своей модели динамики мира рассматривал только шесть переменных. Последователи его метода довели число переменных уже до ста, и дальнейшее усложнение продолжается, хотя оно пренебречь все более сомнительные и часто просто неубозримые результаты.

Теория самоорганизации моделей, которая рассматривается в настоящей монографии, категорически отвергает путь расширения и усложнения модели и увеличения исходного объема информации об объекте и постулирует существование *оптимального, ограниченного размера области моделирования и единственной модели оптимальной сложности*. Их можно найти при помощи самоорганизации, т. е. перебора многих моделей-претендентов по целесообразно выбранным внешним критериям селекции моделей. Оптимизация модели по некоторому ансамблю критериев определяет достижимые при заданных уровнях помех и объеме наблюдений результаты моделирования.

В моделях прогноза погоды, глобальных моделях динамики мира, эконометрии следует прежде всего сократить число уравнений, оставив только самые непротиворечивые уравнения (т. е. равноточные на разных выборках данных), и тем самым перейти к рассмотрению малого числа хорошо прогнозируемых «будущих» характеристических переменных. Каждый ведущий фактор содержит информацию о целом множестве коррелированных с ним факторов. Все другие переменные прогнозируются в функции ведущих переменных и времени. Такая возможность сокращения объема задачи объясняется тем, что многие переменные взаимосвязаны. ЭВМ, выбирая ведущие переменные (иногда только одну или две из них), косвенно учитывает эти связи, хотя они и остаются неизвестными. Выбирая наиболее непротиворечивые уравнения и эффективные факторы, ЭВМ как бы сама увеличивает объем того, что не учитывается (т. е. помех), до тех пор, пока это не начинает ухудшать результат моделирования — минимум критерия соответствует модели оптимальных размеров (числу учитываемых уравнений) и оптимальной сложности описаний (числу слагаемых и степени полинома в каждом из них).

Такой новый взгляд на задачи моделирования возник не сразу после постановки экспериментов по самоорганизации моделей. Потребовались годы, пока

была замечена удивительная помехоустойчивость моделей оптимальной сложности, позволяющая успешно решать задачи моделирования в условиях помех и при отсутствии учета многих факторов.

Помехоустойчивые алгоритмы самоорганизации моделей при специальном выборе вида, формы и последовательности применения критерии позволяют уже сейчас получать весьма точные прогнозы в условиях, когда мощность помех превышает мощность полезного сигнала. Объяснение этого эффекта связано с отказом от принципа «чем больше и сложнее, тем точнее» и признанием принципа существования оптимального размера и сложности модели. Этот процесс становления новых взглядов подробно описан в первой главе книги. А тот факт, что с увеличением уровня помех модели вырабатываются машиной все более и более простыми, позволил перебросить мост от теории самоорганизации к общей теории связи, где вторая теорема Шеннона в канале с шумом объясняет именно такой характер пропускной способности линий связи.

Возник вопрос: не поможет ли теория связи повысить помехоустойчивость моделирования? Оказалось, что восстановление сигналов аналогично задаче моделирования по экспериментальным данным, и предельные возможности самоорганизации моделей могут быть оценены подобно тому, как это делается в теории связи [97, 172] и теории потенциальной эффективности [160]. Эти вопросы рассматриваются во второй главе. Таким образом, подход самоорганизации оказался новым только в области экспериментального моделирования, а общая теория связи может рассматриваться как дополнительное обоснование этого подхода. Нельзя сказать, что перенесение идей теории связи в теорию самоорганизации моделей уже закончено, скорее — оно только начато. Например, возможно прямое применение методов кодирования сигналов для планирования экспериментов, обеспечивающих помехоустойчивое моделирование. Обе указанные теории требуют обогащения за счет использования арсенала средств теории распознавания образцов. Но это дело ближайшего будущего.

Рассматриваемая аналогия задач и методов теорий связи и самоорганизации моделей относится прежде всего к однорядным (комбинаторным) алгоритмам самоорганизации, основанным на полном переборе моделей различной сложности в выбранном базисе (в котором, как предполагается, содержится истинная или физическая модель). Третья глава посвящена описанию структуры комбинаторных алгоритмов метода группового учета аргументов (МГУА) и способов оптимизации вычислений при реализации этих алгоритмов на ЭВМ. Здесь же описан эффективный селекционно-комбинаторный алгоритм, позволяющий существенно сократить объем полного перебора моделей и тем самым расширить область освещаемости моделирования по однорядным алгоритмам.

Для комбинаторных алгоритмов МГУА решена задача теоретической оценки потенциальной (предельно достижимой) помехоустойчивости моделирования для случая нормального распределения помехи. Это результаты, изложенные в четвертой главе, качественно соответствуют второй теореме К. Шеннона и теории идеального приемника В. А. Котельникова [97, 112], однако отражают специфику задач моделирования. Указанные оценки получены на основе анализа предельно помехоустойчивого (идеального) критерия, дающего количественную оценку эффективности моделирования по экспериментальным данным. Разработка аналитических методов оценки помехоустойчивости многорядных алгоритмов МГУА (на основе теории связи или теории чувствительности) еще далеко не закончена.

При решении практических задач моделирования и с помощью специальных вычислительных экспериментов установлено, что помехоустойчивость моделирования зависит от выбора алгоритма самоорганизации и критерии селекции, применяемых для получения модели оптимальной сложности. Классификация внешних критериев селекции и результаты оценки помехоустойчивости моделирования с применением некоторых из них описаны в пятой главе, причем здесь результаты получены как аналитически, так и с помощью вычислительных экспериментов.

Следуя известным направлениям в общей теории связи, в шестой главе рассматриваются вопросы повышения помехоустойчивости самоорганизации моделей за счет использования некоторой информации о помехах. Рекомендуемые здесь алгоритмы весьма эффективны для помехоустойчивой обработки экспери-

ментальных данных, но имеют ограниченное применение — в случаях наличия дополнительной информации о корреляционных, частотных или временных характеристиках помех.

В седьмой главе исследуется помехоустойчивость двухуровневых (или двухмасштабных) алгоритмов долгосрочного количественного прогноза. Подробно рассмотрены свойства критерия баланса прогнозов — главного в таких алгоритмах, определены условия его корректного применения. Новым результатом является доказательство сходимости (устойчивости) этого алгоритма. По мнению авторов, долгосрочный количественный прогноз можно получить только при помощи двухуровневых алгоритмов. Исключением может быть одноуровневый гармонический алгоритм МГУА в случае моделирования колебательных процессов. Если это так, то становится понятным, почему до сих пор нет удачных долгосрочных количественных прогнозов — потому, что помехоустойчивые двухуровневые алгоритмы еще не были применены.

Значение повышения помехоустойчивости моделирования нельзя переоценить. Мы не располагаем точными количественными долгосрочными прогнозами в метеорологии, экономике, экологии, землеведении и других сложных науках только из-за малой помехоустойчивости методов моделирования. Дальнейшее повышение помехоустойчивости моделирования при помощи усовершенствования методов самоорганизации моделей вплоть до предельно достижимых значений должно привести к эффективному решению указанных важных практических задач.

Полные, легко интерпретируемые физические модели, необходимые для по-знавательных целей, могут быть получены как в результате глубокого изучения объекта моделирования, так и с помощью алгоритмов МГУА по экспериментальным данным. Однако помехоустойчивые нефизические модели оптимальной сложности могут быть найдены только при помощи перебора вариантов по целесообразным критериям, заданным экспертами. Это определяет путь дальнейшего развития математического моделирования для количественного долгосрочного прогнозирования сложных процессов и систем. Он неизбежно лежит через самоорганизацию моделей на ЭВМ.

- N — общая длина выборки (таблицы) наблюдений за моделируемым объектом или процессом (число точек исходных данных)
- A — первая обучающая подвыборка (последовательность) данных наблюдений
- B — вторая обучающая подвыборка данных
- C — экзаменационная часть выборки
- $W = A \cup B$ — объединение подвыборок A и B
- G — произвольная часть исходной выборки (может соответствовать A , B , W или C)
- N_A , N_B , N_W , N_C — длины соответствующих частей выборки ($N_A + N_B = N_W$; $N_W + N_C = N$)
- $x_i = (x_{i1}, \dots, x_{iN})^T$ — вектор наблюдений i -й независимой (входной) переменной
- $u = (u_1, \dots, u_N)^T$ — вектор наблюдений зависимой (выходной) переменной в отсутствие шума (точный сигнал)
- $\xi = (\xi_1, \dots, \xi_N)^T$ — вектор реализации произвольных помех
- $y = (y_1, \dots, y_N)^T$ — вектор наблюдений зашумленной выходной переменной ($y = u + \xi$, зашумленный сигнал)
- n — общее число независимых переменных (аргументов)
- $X = (x_1 | x_2 | \dots | x_n)$ — полная матрица наблюдений аргументов
- X_G, u_G, y_G, ξ_G — матрица и векторы, соответствующие подвыборке G исходных данных
- θ — уровень помех (отношение мощности помех к мощности сигнала)
- s — сложность частной модели (число оцениваемых коэффициентов)
- s^0 — сложность истинной модели
- s^* — число коэффициентов в модели оптимальной сложности
- F — свобода выбора (число лучших моделей, отбираемых на некотором этапе алгоритма самоорганизации)
- $\hat{b} = (\hat{b}_1, \dots, \hat{b}_s)^T$ — оценка вектора коэффициентов частной модели по точным данным
- $\hat{a} = (\hat{a}_1, \dots, \hat{a}_s)^T$ — оценка вектора коэффициентов частной модели по зашумленным данным
- \hat{u}, \hat{y} — оценки зависимых переменных (выходы частных моделей с коэффициентами \hat{b} и \hat{a})
- δ^2 — квадрат ошибки модели при точных данных (остаточная сумма квадратов)

ε^2 — квадрат ошибки модели при зашумленных данных
 η^2 — квадрат теоретической ошибки восстановления точных данных (вектора u) по модели, построенной по зашумленным данным
 I — произвольный внешний критерий селекции
 $\Delta^2(B)$ — критерий регулярности
 $\Delta^2(C)$ — критерий точности прогноза
 n_{cm}^2 — критерий минимума смещения
 d^2 — симметричный критерий регулярности
 S^2 — критерий стабильности
 V^2 — критерий абсолютной помехоустойчивости
 $\Delta_{cp}^2(W)$ — усредненный критерий регулярности
 $i^2(W)$ — интегральный критерий
 k^2 — комбинированный критерий (например, $k^2 = n_{cm}^2 + \Delta^2(C)$)
 R^2 — идеальный внешний критерий
 k — индекс такта дискретного времени
 L — длина периода циклического процесса
 t, T — индексы двухмерного (двухмасштабного) отсчета времени
 q_{tT} — переменная нижнего уровня алгоритма двухуровневого долгосрочного прогнозирования
 Q_T — переменная верхнего уровня алгоритма
 \hat{q}_{tT}, \hat{Q}_T — оценки соответствующих переменных
 B_L^2 — критерий баланса прогнозов

Глава 1

РАЗВИТИЕ ПОМЕХОУСТОЙЧИВЫХ МЕТОДОВ МОДЕЛИРОВАНИЯ ПО ДАННЫМ НАБЛЮДЕНИЙ

1. МЕТОДЫ МОДЕЛИРОВАНИЯ, ПОРОЖДЕННЫЕ ПОДХОДАМИ ТЕОРИИ ИНФОРМАЦИИ И АВТОМАТИЧЕСКОГО УПРАВЛЕНИЯ

Математическое моделирование в настоящее время применяется для решения самых разнообразных задач кибернетики и является необходимым звеном, объединяющим эти задачи. Основными целями моделирования являются: 1) системный анализ взаимодействий переменных в сложных объектах для определения множества характеристических переменных и разделения его на подмножества входных и выходных (независимых и зависимых) переменных; 2) структурная и параметрическая идентификация моделируемых объектов с уже известными системными взаимодействиями; 3) прогнозирование процессов с различной заглобо-временностью, в том числе качественное и количественное; 4) управление, планирование и принятие решений.

На примере решения задачи долгосрочного количественного прогнозирования яснее, чем на других задачах, видно, что существует некоторое принципиальное различие двух подходов к моделированию. Только опыт наблюдений за действием объекта или сравнение ряда расчетов может быть основой построения его математической модели. Однако такой опыт может быть представлен в виде сложившихся представлений автора модели об объекте или в виде таблицы опытных данных — результатов активного или пассивного эксперимента. Первый вид опыта приводит к имитационному моделированию, второй — к экспериментальному методу теории самоорганизации моделей. Классическим примером имитационного моделирования является известная модель динамики мира американского ученого Дж. Форрестера [161], который называет свой метод динамическим. Это подтверждает связь его метода с идеями теории автоматического управления.

Для этой теории характерны несколько преувеличенное внимание к динамике и устойчивости процессов и недостаточная для практики разработка вопросов квазистатики (трендов). Само понятие «тренд» остается в этой теории неопределенным.

В отличие от теории автоматического управления в теории информации (или связи) рассматривается прежде всего квазистатическая составляющая процесса, называемая оптимальным прини-

маемым сигналом, т. е. тренд сигнала, при некотором упрощенном описании его динамической составляющей. Последнюю чаще всего рассматривают как помеху типа белого шума и называют флюктуационной. Идеи теории информации породили в моделировании ряд направлений. По-видимому, первым было «моделирование матрицы вход — выход», связанное с именем Р. Стоуна — автора известной модели экономики Англии [151].

Не менее полно подход теории информации отражен в теории самоорганизации моделей, реализуемой в алгоритмах МГУА. Самоорганизующиеся модели также можно называть информационными.

Указанные два подхода теорий управления и информации являются антиподами, но отнюдь не врагами: начиная с исходной идеи Н. Винера, никто никогда не противопоставлял теорию автоматического управления и теорию информации, несмотря на различие их подходов. Комбинированный метод, использующий совместно оба подхода, является наиболее эффективным.

В ближайшем будущем кибернетику можно представить себе как некоторую метанавуку о моделировании, использующую совместное применение всех подходов: имитационного, вероятностного и самоорганизации.

Уязвимым местом имитационных методов является то, что автор общей модели вынужден указать закономерности для всех элементов изучаемой системы, в том числе и для тех, которые он знает хуже других или которые вообще в меньшей степени поддаются имитации.

При комбинированном методе модели таких плохо изученных элементов получаются при помощи экспериментальных методов самоорганизации моделей. Известные элементы задаются непосредственно автором модели.

Особенности комбинированного метода построения физической модели. Укажем основные этапы построения динамических моделей физических процессов:

1) регистрация данных наблюдений на объекте (измерения желательно производить через равные интервалы времени и равные шаги в пространстве; максимально допустимый шаг таблицы следует определять по теореме отсчетов Котельникова — Шеннона [132]);

2) выбор физического закона, который целесообразно положить в основу модели (элемент имитационного моделирования);

3) замена непрерывного математического описания его дискретным аналогом (разностным уравнением);

4) оценка и адаптация (т. е. пересчет по всем данным) коэффициентов разностного уравнения (элемент самоорганизации модели).

Особенности индуктивного метода самоорганизации как физических, так и нефизических моделей [49]. Для случая, когда трудно положить в основу моделирования тот или иной физический закон, применяется так называемый индуктивный подход, при ко-

тором структура модели оптимальной сложности подбирается при помощи опробования многих моделей-претендентов по критериям селекции. В пространстве критериев каждой модели соответствует определенная характеристическая точка. В соответствии с принципом самоорганизации выбирается некоторое количество точек (и моделей), ближайших к началу координат. Процедура доопределения модели состоит в том, что из отобранных лучших моделей выбирается единственная модель по главному критерию.

2. САМООРГАНИЗАЦИЯ МОДЕЛЕЙ НА ЭВМ. МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Самоорганизация моделей подобно регрессионному анализу относится к экспериментальным методам моделирования, так как основана на обработке таблицы данных наблюдений, полученной в режиме пассивного или активного эксперимента. Пассивный эксперимент представляет собой наблюдение за нормальной работой объекта (обычно при ручном, не автоматическом, управлении). В активном эксперименте выполняется изменение внешних воздействий по специальной программе (планирование эксперимента).

При самоорганизации моделей в отличие от регрессионного анализа, где структура модели задается, основное внимание уделяется определению структуры модели оптимальной сложности. Согласно принципу самоорганизации, при постепенном усложнении структуры модели (например, при увеличении числа членов и степени полиномиальной модели) значения внешних критериев сначала уменьшаются, а затем возрастают, т. е. имеется минимум, определяющий модель оптимальной сложности.

Если бы в нашем распоряжении имелась аналитическая зависимость значения критерия от показателей сложности, можно было бы поставить типовую задачу нелинейного программирования и решить ее известными дедуктивными (математическими) методами, такими, как градиентный метод, метод наискорейшего спуска и др. Однако аналитическая зависимость величины критерия от сложности модели не известна, и поэтому задача поиска минимума критерия решается методом индукции, т. е. при помощи полного или неполного направленного перебора многих моделей-претендентов. Этот перебор часто называется селекцией, так как напоминает известные процессы селекции в агробиологии («гипотеза селекции» [43]).

В программах, реализующих самоорганизацию модели, предусматриваются блок генератора множества моделей-претендентов, блок вычисления значения критерия и выходной блок выбора модели оптимальной сложности по минимуму критерия.

В случае точных исходных данных выбор оптимальной модели можно сделать по любому одному ее признаку, т. е. по внешним или внутренним критериям. Однако при неточных, зашумленных данных требуется применение выбора модели по многим критериям. Здесь вступает в действие принцип внешнего дополнения

(по терминологии теоремы неполноты Геделя [125]), по которому только внешние критерии (т. е. рассчитанные на «свежих» данных, не использованных для модели) при увеличении сложности модели проходят через минимумы. Таким образом, в основе самоорганизации моделей лежат неотделимые друг от друга два принципа: самоорганизации и внешнего дополнения.

Алгоритмы самоорганизации (перебора) моделей могут быть однорядными или многорядными. И те и другие реализуют метод группового учета аргументов.

Метод полной математической индукции (т. е. полного перебора всех вариантов модели, которые можно получить из заданного полного описания) реализуется в однорядных (комбинаторных) алгоритмах МГУА, применяемых для решения определенных и переопределенных задач моделирования, в которых число членов полного описания n (чаще всего полинома) меньше числа точек в таблице опытных данных N .

Для решения недоопределенных задач (при $N < n$) разработаны многорядные алгоритмы МГУА. Многорядность является основным средством рационализации перебора, при котором число опробуемых моделей резко уменьшается с весьма малым риском потери самой оптимальной модели. Число моделей, перебираемых на машине, растет не в степенной, а в полиномиальной зависимости от числа переменных.

Однорядные комбинаторные алгоритмы практически применимы для решения задач с числом переменных не более двадцати ($n < 20$), а многорядные — не более одной тысячи ($n < 1000$). При переходе от однорядных алгоритмов к многорядным возможна как потеря самой оптимальной модели, так и небольшая «ошибка многорядности». Для полиномиальных алгоритмов ошибку многорядности можно определить в виде разности прогнозов (решений), полученных по однорядному и многорядному алгоритмам. В работе [139] предложены алгоритмы с возвратом к исходным переменным на каждом ряду, не имеющие ошибки многорядности.

В многорядных алгоритмах селекции модели, а также в случае последовательного использования нескольких критерии отбора по очереди, кроме указанных двух принципов, в алгоритмах самоорганизации реализуется еще принцип неокончательных решений Гabora [140]: из ряда в ряд селекции передается не одна, а несколько лучших по критериям моделей, или, другими словами, на каждом ряду сохраняется «свобода выбора»; только на последнем ряду выбирается единственная (для каждого ансамбля критериев) модель оптимальной сложности.

Выбор критериев, участвующих в ансамбле, все еще является большим искусством, в связи с чем метод самоорганизации называют эвристическим. Для выбора множества критериев, по-видимому, можно использовать аналитические методы, подобные методам выбора ансамбля признаков для распознавания образов. Есть критерии, включение которых в ансамбль обеспечивает самоорганизацию физической модели (с легко интерпретируемыми коэффи-

циентами), например критерий минимума смещения. Другие критерии применимы только для селекции нефизических прогнозирующих моделей, которые при наличии зашумленных исходных данных значительно отличаются от физических: критерии баланса прогнозов и устойчивости вероятностных характеристик. При неточных, зашумленных опытных данных в теории самоорганизации рекомендуется как одно из главных средств повышения помехоустойчивости многокритериальный выбор модели оптимальной сложности.

3. ОБЗОР РАЗВИТИЯ МГУА

Метод группового учета аргументов, применяемый для математического моделирования сложных систем, является существенной частью комбинированного метода и полностью соответствует индуктивному методу моделирования. Вместе с расширением применения алгоритмов МГУА развивается математическая теория самоорганизации моделей на ЭВМ, которая является обоснованием метода. Историю возникновения и развития МГУА легко проследить по журналу «Автоматика» начиная с 1968 г. Почти в каждом номере можно найти статьи, относящиеся к тематике МГУА. Кроме того, в [42] было помещено приложение, посвященное новому методу. В приложении указывалась аналогия схемы перцептрона Ф. Розенблата [130] и многорядных алгоритмов МГУА.

Одно из первых исследований МГУА за рубежом выполнил Р. Шанкар [210]. Он сравнил МГУА и регрессионный анализ по точности моделирования для короткой выборки исходных данных. Позже различные варианты алгоритмов МГУА были опубликованы японскими (обычно с некоторыми предложениями по их модификации) [189, 200, 208, 212] и польскими учеными [202]. Вывод работы [202] был таков: МГУА является лучшим методом для решения задач идентификации и краткосрочного прогноза. К сожалению, это исследование не было полным, потому что его автор не рассмотрел возможностей гармонического и многоуровневого алгоритмов МГУА с критерием баланса прогнозов, иначе он пришел бы к выводу, что и для долгосрочного прогноза МГУА является наилучшим и, по-видимому, единственным методом, обеспечивающим точность прогноза при большом времени его упреждениям. Математическая теория МГУА показала, что регрессионный анализ является частным случаем МГУА [174, 175].

В настоящее время в разработку теории и применений МГУА включились десятки (а может, и сотни) ученых. Теперь это достояние не только автора метода, но и многих других. Тем не менее целесообразно коротко рассмотреть, что было сделано, что нужно сделать и что можно ожидать от МГУА в ближайшем будущем.

Период 1968–1971 гг. характерен применением одного критерия селекции при полном пренебрежении к помехоустойчивости моделей [41–43, 55, 63, 64, 66, 68, 69, 190–192].

В 1972—1975 гг. была предпринята первая попытка повысить помехоустойчивость: появился многокритериальный выбор моделей и метод априорного задания ограниченного круга гипотез, из которых одна заведомо верна (пример — гармонический алгоритм МГУА) [44, 45, 56, 58, 61, 65, 85, 130, 169, 193, 207, 210, 211].

В 1976—1979 гг. получены существенные результаты в математическом обосновании МГУА (только для задачи идентификации) [46, 54, 57, 59, 152, 189, 194—196, 202, 204, 208, 212, 215].

В 1980—1982 гг. появилось помехоустойчивое, многоуровневое, долгосрочное, количественное прогнозирование и создан объективный системный анализ сложных объектов [47—59, 70—77, 79—81, 116, 144, 174—180, 205].

В настоящее время ищутся новые пути повышения помехоустойчивости алгоритмов МГУА за счет привлечения идей теории информации (связи) [60], предстоит также дальнейшее развитие теории МГУА (задача прогнозирования). Рассмотрим сказанное подробнее.

Период 1968—1971 гг. На ранней стадии развития МГУА применялся для решения задач распознавания образов, идентификации характеристик и краткосрочного прогноза при помощи перебора моделей по критерию регулярности:

$$\Delta^2(B) = \frac{\sum_{\Phi} (q_{\Phi} - \hat{q}_m)^2}{\sum_{\Phi} q_{\Phi}^2} \rightarrow \min,$$

где q_{Φ} — фактические (табличные) данные; \hat{q}_m — выход модели; $N = N_A + N_B$ — множество точек исходных данных, разделенное на две части: $N_A \sim 70\%$ и $N_B \sim 30\%$.

Было сделано много предложений относительно разделения данных на две части [128]. Например, предлагалось ранжировать данные по дисперсии и разделять полученный ряд. Экспериментально было установлено, что при этом достигается минимум рядов селекции в многорядных алгоритмах МГУА, т. е. получается самая простая модель. В тот период рассматривались только многорядные алгоритмы МГУА (см. «параболу Тодуа» [69]). Другие решения задачи разделения данных были найдены значительно позже [20, 178]. Критерий регулярности использовался также в форме коэффициента корреляции между переменными q_{Φ} и q_m внутри интервала данных B или в форме индекса корреляции (для нелинейных моделей).

В качестве опорных функций использовались полиномы, логические сети и размытые множества Заде. Для трендов применялись алгебраические (полиномиальные) уравнения, а для динамической составляющей процессов — уравнения в конечных разностях. Были также разработаны алгоритмы МГУА, в которых

в качестве опорных функций использовались формулы Байеса (так называемые вероятностные алгоритмы МГУА, или многорядная теория статистических решений [43, 57]). Эти алгоритмы были успешно применены для решения экологических задач.

Авторы первых работ были поражены высокой точностью краткосрочных прогнозов. В [41] говорилось: «Автор считает точность, достигнутую в примере прогноза реального случайного процесса, наибольшим чудом, которое он встретил за всю свою жизнь».

При этом использовались только физические модели, имеющие полное множество аргументов (полный информационный базис). Очень низкий уровень помех (около 5—10 %) допускал выбор физических моделей.

Краткость интервала точного прогнозирования (упреждения прогноза), очень малая помехоустойчивость и чувствительность к нарушению информационного базиса быстро охладили восторги. Наоборот, появилась тревога — может быть, МГУА, не работает на очень зашумленных данных?

Период 1972—1975 гг. Решалась проблема моделирования при зашумленных данных и неполном информационном базисе. Были предложены два способа повышения помехоустойчивости алгоритмов МГУА: 1) многокритериальный выбор модели при специальной форме критериев селекции [45, 54]; 2) использование дополнительной априорной информации (комбинированный метод).

Основные идеи многокритериального выбора близки к идеям теории распознавания образов [67], которую можно применить для нахождения множества наиболее эффективных критериев.

Были предложены критерии минимума смещения

$$n_{cm}^2 = \frac{\sum_{\Phi} (\hat{q}_A - \hat{q}_B)^2}{\sum_{\Phi} q_{\Phi}^2} \rightarrow \min,$$

где \hat{q}_A и \hat{q}_B — выходы моделей одинаковой структуры, коэффициенты которых вычислены на частях A и B выборки, и критерий баланса переменных, например, вида

$$B = \sum_{\Phi} b_i^2 \rightarrow \min, \quad b_i = f_1(q_i) - f_2(q_i),$$

где b_i — рассогласование предполагаемой связи двух функций f_1 и f_2 .

Первый критерий рекомендуется в случаях, когда требуется найти физическую модель [92], второй — для решения задачи доопределения модели и долгосрочного прогноза [58]. Предложены многие другие математические формы критериев: точности краткосрочного прогноза $\Delta^2(C) \rightarrow \min$, точности пошагового прогноза $i^2(N) \rightarrow \min$, устойчивости вероятностных характеристик долгосроч-

ного прогноза (устойчивость нижних или верхних моментов $\rho_{1-2}(m) \rightarrow \min$, $\rho_{2-3}(m) \rightarrow \min$, корреляционной функции $\rho(K(\tau)) \rightarrow \min$, распределений и др.). Основным в совершенствовании методов моделирования является оценка моделей по критериям минимума смещения и баланса переменных вместе с решением задачи помехоустойчивости.

Использование достоверной априорной информации не только сокращает объем перебора, но, что важнее, повышает точность и время упреждения прогноза. Например, задание общего вида корней характеристического уравнения разностной модели уменьшает число всех гипотез. Свойства решений для определенного вида корней известны, поэтому многие из них можно сразу отбросить и не рассматривать. Таким образом, остается организовать перебор нескольких гипотез о характере корней и найти (по формулам Виетта) ограничения на коэффициенты полиномов. Очень часто мы приходим к так называемому гармоническому алгоритму МГУА [22, 115].

Можно привести еще один пример реализации этой общей идеи в обобщенном алгоритме МГУА [45], где выполняется перебор гипотез о взаимодействии переменных состояния. Например, для четырех переменных имеем:

1-я гипотеза — переменные не взаимодействуют:

$$q = a_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4);$$

2-я гипотеза — первая переменная не взаимодействует, а остальные взаимодействуют между собой:

$$q = a_0 + f_5(x_1) + f_6(x_2, x_3, x_4);$$

3-я гипотеза — взаимодействуют первая со второй и третья с четвертой:

$$q = a_0 + f_7(x_1, x_2) + f_8(x_3, x_4) \text{ и т. д}$$

Здесь f_1, f_2, f_3, f_4, f_5 — полиномы первой, f_7, f_8 — второй, f_6 — третьей степени.

Таким образом, цель оптимизации структуры модели достигается. Комбинаторный алгоритм МГУА также можно рассматривать как полный перебор ряда гипотез, одна из которых верна.

Дополнительная априорная информация, внесенная в такой обобщенной форме, резко повышает помехоустойчивость алгоритмов МГУА (гармонический и обобщенный алгоритм МГУА показали наивысшую помехоустойчивость), хотя практически желательно повысить ее еще в десятки раз. Для исследования свойства помехоустойчивости алгоритмов МГУА были разработаны специальные схемы «вычислительного эксперимента» [2, 23, 54]. Оказалось, что многокритериальный выбор действительно намного лучше, чем выбор модели по одному критерию. Лучшие эксперименты показали, что шум может быть в десять раз больше сигнала. Но этого совершенно недостаточно для практических задач прогнози-

рования погоды или в экологии. Дальнейший путь улучшения помехоустойчивости алгоритмов МГУА связан с использованием общей теории связи (см. гл. 2).

Исходные многорядные алгоритмы МГУА были основаны на так называемой гипотезе селекции, используемой в агротехнической практике. Эти алгоритмы не требовали длинной выборки данных: число точек данных могло быть в несколько раз меньше числа слагаемых полного полинома ($n > N$). В 1972 г. был предложен однорядный, или комбинаторный, алгоритм МГУА [94], который можно применять для решения определенных задач, где $n < N$.

Комбинаторный алгоритм не имеет «ошибки многорядности» и не может потерять оптимальную модель потому, что обеспечивает полный перебор всех возможных (при заданной опорной функции) моделей. Практически его можно применить в случаях, когда число слагаемых полного полинома меньше двадцати (это соответствует трем часам счета на БЭСМ-6). Каждая дополнительная переменная увеличивает время счета в два раза («проклятие многомерности», по Беллману). Для преодоления «проклятия многомерности» можно использовать: статистические методы уменьшения числа переменных (прежде всего ранговую корреляцию); многорядные алгоритмы МГУА; способ «усечения треугольника перебора».

Поясним последний способ. Полиномы, используемые в комбинаторном алгоритме МГУА, можно записать в треугольной форме: в первой строке — полный полином, во второй — более короткий полином и так далее, пока не будут получены все полиномы, имеющие два члена. Перебору подлежат только полиномы, имеющие менее 20 слагаемых. Обоснование метода весьма простое: законы природы всегда прости и обозримы, можно не рассматривать слишком сложные многокомпонентные модели — они всегда невозможны. А если вспомнить, что нефизические оптимальные модели (получаемые при увеличении помех) всегда проще физических, то ясно, что усечение треугольника перебора можно применять для самоорганизации как физических, так и нефизических моделей. Более эффективный способ сокращения перебора изложен в гл. 3.

Практически теперь многорядные алгоритмы МГУА используются довольно редко, если модель содержит только одно уравнение. Они применяются для выбора системы уравнений в системном анализе, а также в случае, если мы располагаем очень малым числом точек данных (что встречается, например, при моделировании быстроизменяющихся объектов).

Период 1976—1979 гг. Развитие теории самоорганизации в рассматриваемый период отражено в двух книгах [45, 57]. Было показано, что многие многорядные алгоритмы имеют некоторую «ошибку многорядности», иногда существенную. Эта ошибка аналогична статической ошибке систем управления. Было показано, что многорядный алгоритм подобно системе стабилизации сходится к состоянию равновесия — «точке сгущения».

Типы сходимости позже исследовались более тщательно, и были предложены новые астатические многорядные алгоритмы, которые вообще не имеют ошибки многорядности [45, 174]. В случае, когда таблица входных данных изменяется во времени (появляются добавочные точки), вполне уместно использовать в многорядных алгоритмах МГУА идеи теории инвариантности систем управления. Но до сих пор исследованы только статические задачи при неизменной таблице входных данных. Теория устойчивости следящих систем с динамическим подходом должна быть применена для исследования алгоритмов долгосрочного прогноза при постепенно изменяющейся таблице входных данных. Здесь вместо сходимости алгоритма к «точке сгущения» необходимо исследовать сходимость к трендам процесса (см. гл. 7).

В 1977 г. были предложены многорядные алгоритмы МГУА для объективного решения задачи системного анализа. Последний может быть выполнен на уровне трендов (по алгебраическим уравнениям) или дифференциальных уравнений либо их конечно-разностных аналогов [152, 70].

Наиболее общий системный анализ основан на уравнениях вида

$$q_0 = f_1(t, T, U) + f_2(q_{-1}, q_{-2}, \dots, q_{-\tau}),$$

где f_1 — функция, называемая в математической физике функцией источника [49]. Для упрощения общего исследования уравнение делится на две части:

- а) анализ трендов $q_{1(0)} = f_1(t, T, U)$;
- б) анализ динамики $q_{2(0)} = f_2(q_{-1}, q_{-2}, \dots, q_{-\tau})$ (здесь $q_0 \neq q_{1(0)} + q_{2(0)}$; чтобы получить $q_0 = q_{1(0)} + q_{2(0)}$, необходимо использовать таблицу отклонений от трендов для второго уравнения).

До публикации этих работ объективного системного анализа не существовало. Оказалось, что перебор по ансамблю критериев (включая критерий минимума смещения) позволяет выбрать единственную оптимальную систему уравнений (в конечно-разностной или алгебраической форме) и, следовательно, указать элементы сложного объекта, их входные переменные и переменные состояния (выходные).

Теперь мы имеем в своем распоряжении мощный аппарат моделирования, чтобы объективно ответить на важные вопросы, которые стоят в экологии, экономике, социальных и других научных областях. Компьютер становится объективным судьей для решения научных споров, если ансамбль критериев согласован между экспертами, принимающими участие в дискуссии.

Новая форма системного анализа совершенно изменила и значительно повысила эффективность существующих глобальных экономических моделей (таких, как модель ЛИНК [30] и др.). Например, в блоке «Великобритания» модели ЛИНК используется 227 полиномиальных уравнений регрессии. Но чего стоят эти уравнения с точки зрения долгосрочного количественно-

го прогноза, если они сильно зависят от выбора места таблицы исходных данных, на котором они получены? Без анализа по ансамблю критериев ими пользоваться просто нельзя! Алгоритм для объективного системного анализа выбрал для этого блока только четыре уравнения — четыре выходных переменных [70]. Они являются своеобразным сердцем системы, и только они удовлетворяют критерию минимума смещения (т. е. эти уравнения остаются все время почти без перемен). Все остальные уравнения непрерывно изменяются как по структуре, так и по оценкам коэффициентов: на одной части таблицы данных они одни, а на другой — другие. Следовательно, прогнозировать нужно только существенные, или ведущие, четыре переменные. Прогнозы остальных интересующих нас переменных можно найти потом, в функции ведущих переменных, пользуясь комбинаторным алгоритмом МГУА и перебором по критерию баланса прогнозов как главным судьей.

Период 1980—1983 гг. В этот период были получены многие важные результаты. Прежде всего стало ясно, что физические модели не могут быть использованы для долгосрочного количественного прогноза. Больше шум в исходных данных — более простые модели выбирают алгоритмы МГУА. Таким образом, физические модели пригодны лишь для задачи идентификации и краткосрочного прогноза, и то в случае, когда шум мал, а информационный базис модели полный, что показала практика самоорганизации моделей. Основные результаты этого периода отражены в [79].

Единственный сигнал из отдаленного будущего, которым мы располагаем, — это сигнал о том, что там, в будущем, будет выполняться критерий баланса прогнозов. Поэтому для долгосрочных прогнозов мы должны организовать доопределение модели по критерию баланса. Мы говорим «доопределение», так как, чтобы избежать многозначности выбора, по критерию баланса должны сравниваться модели, уже предварительно отобранные по другим вспомогательным критериям. Критерий баланса становится главным в выборе модели оптимальной сложности.

Двухуровневые алгоритмы МГУА для долгосрочного прогноза осуществляют сравнение нескольких прогнозов, полученных при двух различных опорных функциях, по критерию баланса прогнозов. Например, сравниваются прогнозы годовых и сезонных значений и выбирается лучшая пара прогнозов. Одновременное использование двух различных опорных функций является совершенно новым приемом в моделировании. Двухуровневый долгосрочный прогноз был показан на примере решения экстраполяционных задач [82] и на прогнозировании экологической системы оз. Байкал [75]. Прогнозы были выданы с упреждением на двадцать лет, и они уже удовлетворительно выполняются за последние четыре года.

Очень важно найти ведущую переменную в множестве выходных переменных объекта. Ведущей называется переменная, которая прогнозируется лучше, более точно, чем другие.

Для прогноза годовых значений ведущей переменной рекомендуется гармонический (или экспоненциально-гармонический) или обобщенный алгоритмы МГУА, так как они обладают наиболее высокими свойствами помехоустойчивости [45]. Эти алгоритмы лучше других среди всех одноуровневых алгоритмов. Зная прогноз ведущей переменной, легко получить по критерию баланса прогноз всех других переменных системы. В работах, начиная с [73], в качестве опорной функции широко применяются полиномы с переменными коэффициентами. Например, имея три управляющие переменные U_1 , U_2 и U_3 и две системные переменные q_1 , q_2 , мы можем записать следующие полные полиномы для комбинаторного алгоритма МГУА:

для алгебраических уравнений (выражающих тренды или «функции источника»)

$$\begin{aligned} q_1 &= (a_0 + a_1t + a_2T + a_3T^2 + a_4U_1 + a_5U_2 + a_6U_3) 1 + \\ &+ (b_0 + b_1t + b_2T + b_3T^2 + b_4U_1 + b_5U_2 + b_6U_3) q_2, \\ q_2 &= (c_0 + c_1t + c_2T + c_3T^2 + c_4U_1 + c_5U_2 + c_6U_3) 1 + \\ &+ (d_0 + d_1t + d_2T + d_3T^2 + d_4U_1 + d_5U_2 + d_6U_3) q_1; \end{aligned}$$

для конечно-разностных уравнений (при учете двух запаздываний)

$$\begin{aligned} q_{1(0)} &= (a_0 + a_1t + a_2T + a_3U_1 + a_4U_2 + a_5U_3) 1 + (c_0 + c_1t + c_2T + \\ &+ c_3U_1 + c_4U_2 + c_5U_3) q_{1(-1)} + (d_0 + d_1t + d_2T + d_3U_1 + d_4U_2 + \\ &+ d_5U_3) q_{1(-2)} + (b_0 + b_1t + b_2T + b_3U_1 + b_4U_2 + b_5U_3) q_{2(0)} + \\ &+ (f_0 + f_1t + f_2T + f_3U_1 + f_4U_2 + f_5U_3) q_{2(-1)} + (g_0 + g_1t + g_2T + \\ &+ g_3U_1 + g_4U_2 + g_5U_3) q_{2(-2)}, \end{aligned}$$

где $q_{1(-1)}$, $q_{1(-2)}$, $q_{2(-1)}$, $q_{2(-2)}$ — запаздывающие аргументы. Аналогичное уравнение составляется и для переменной $q_{2(0)}$. Уравнения с переменными коэффициентами помогли решить задачу исследования стабильности инфляции в Англии, рассмотренную в [70]. Если компьютер допускает, то оба уравнения могут быть объединены в одно уравнение более общего вида.

4. ОБЩАЯ ХАРАКТЕРИСТИКА СТРУКТУРЫ АЛГОРИТМОВ МГУА

Особенность алгоритмов МГУА состоит в том, что вид опорной функции, класс уравнений и структура моделей устанавливаются объективным способом при помощи перебора вариантов по целесообразно выбранному ансамблю критериев (индуктивный метод). Способ введения критериев (находится минимум их свертки) обеспечивает объективное нахождение структуры единственной модели оптимальной сложности при высокой помехоустойчивости метода.

Допустимое отношение шум/сигнал может достигать значения единицы и больше (без учета информации о помехах).

Наиболее существенными признаками, отличающими алгоритмы МГУА, являются: число рядов селекции, наличие или отсутствие вычисления остатка, число уравнений в системе. Рассмотрим четыре соответствующие группы алгоритмов, в каждой из которых выделим типичные схемы перебора, общие как для полиномиальных, так и для тригонометрических базисных (опорных) функций. Все изложение ведется только для одномерных задач.

Структуры алгоритмов МГУА остаются аналогичными друг другу для различных опорных функций. Так, для полиномиальных и гармонических алгоритмов можно указать три основных вида структуры: однорядные (комбинаторные); многорядные, без вычисления остатков после каждого ряда селекции; многорядные, с вычислением остатков.

Однорядные алгоритмы МГУА предназначены для решения определенных задач самоорганизации моделей по опытным данным, в которых число точек измерения равно или превышает некоторый (различный для разных типов моделей) алгебраический минимум.

Многорядные алгоритмы МГУА применяются для решения некорректных или недоопределенных задач моделирования, т. е. в случае, когда число точек N в таблице опытных данных меньше числа аргументов n , входящих в синтезируемую модель ($N < n$). Методы регрессионного анализа в этом случае неприменимы, так как не дают возможности построения единственной модели, адекватной процессу за пределами интервала интерполяции.

В зависимости от априорных знаний о процессе в алгоритмах МГУА используются различные опорные (базисные) функции. Например, многорядные алгоритмы МГУА с полиномиальными опорными функциями позволяют получать модели в виде полиномов, число членов в которых значительно больше, чем в моделях, построенных по однорядным алгоритмам (в том числе регрессионным), при одинаковом ограниченном объеме выборки.

Для моделирования колебательных процессов при малом числе данных применяются гармонические и экспоненциально-гармонические многорядные алгоритмы МГУА. Алгоритмы этого класса при определенных условиях также позволяют получить число гармонических составляющих, которое может значительно превышать число гармоник, определяемых теоремой Котельникова для выборки заданной длины ($n \leq \frac{N}{2}$). Например, в задаче моделирования экосистемы Кауховского водохранилища [73] при выборке в 22 точки по однорядному алгоритму можно определить не более 11 гармоник. В то же время оптимальная (по критерию регулярности) модель, полученная по многорядному гармоническому алгоритму, содержит 14 гармоник, т. е. по сложности превышает ограничения указанной теоремы.

Применение многорядных алгоритмов МГУА не ограничивается указанной областью некорректных (недоопределенных) задач. Они успешно используются и в случаях, когда исходных данных достаточно для применения однорядных алгоритмов. Однако и при этом многорядные алгоритмы часто предпочтительнее — прежде всего тогда, когда объем перебора по однорядному (комбинаторному) алгоритму превосходит возможности ЭВМ. Например, комбинаторный алгоритм [146] эффективен при $n \leq 18$, а алгоритм многорядного типа — даже при $n < 1000$.

4.1. Однорядные (комбинаторные) алгоритмы

В комбинаторных алгоритмах выполняется перебор всевозможных частных моделей из заданного базиса (полиномиального или гармонического) с выбором лучшей из этих моделей по заданному критерию селекции (или по комбинации критерии). При переборе сложность частных моделей, т. е. число аргументов или частот, постепенно наращивается от 1 до максимального числа n (числа аргументов базисного набора функций). Таким образом, общая схема комбинаторного алгоритма включает следующие операции: по методу наименьших квадратов (МНК) определяются коэффициенты всех частных моделей при сложности $s=1, s=2, \dots, s=n$; для каждой из них вычисляется значение внешнего индивидуального или комбинированного критерия селекции; единственная модель оптимальной сложности выбирается по минимальному значению критерия. Можно сказать, что комбинаторный алгоритм МГУА основан на методе полной математической индукции, так как при этом не пропускается ни один из возможных вариантов модели, заложенных в исходном полном базисе.

Рассмотрим, как эта общая схема реализуется в полиномиальных и гармонических алгоритмах.

1. **Полиномиальный алгоритм.** Здесь опорным набором аргументов являются n членов некоторого полинома заданной максимальной степени от заданного числа переменных. Например, для случая трех переменных полный квадратичный полином имеет вид

$$q = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_2^2 + a_6x_3^2 + a_7x_1x_2 + \\ + a_8x_1x_3 + a_9x_2x_3.$$

В этом полиноме 10 членов ($n=10$). Иногда в состав аргументов требуется ввести обратные величины $1/x_i$, их степени или другие нелинейные функции (целесообразность введения новых аргументов определяется по уменьшению минимального значения основного критерия). В любом случае полный полином является линейным по коэффициентам, для определения которых применяется МНК. При этом степени и ковариации исходных переменных рассматриваются как эквивалентные «переобозначенные» линейные аргументы.

В соответствии с приведенной схемой алгоритм работает так. Сначала определяются все модели при $s=1$, т. е. состоящие лишь из одного аргумента:

$$q_1 = a_0, q_2 = a_1x_1, q_3 = a_2x_2, \dots, q_{10} = a_9x_2x_3$$

(таких моделей $C_{10}^{10}=10!$). Далее рассматриваются всевозможные модели при $s=2$, состоящие из двух аргументов:

$$q_{11} = a_0 + a_1x_1, q_{12} = a_0 + a_2x_2, \dots,$$

$$q_i = a_0 + a_1x_2x_3, \dots, q_j = a_1x_1 + a_2x_2, \dots,$$

$$q_k = a_1x_1 + a_2x_2x_3, \dots, q_{45} = a_8x_1x_3 + a_9x_2x_3,$$

которых всего будет $C_{10}^2 = 45$. Аналогично строятся C_{10}^3 частных моделей при $s=3$, C_{10}^4 при $s=4$ и т. д. до $C_{10}^{10} = 1$ модели при $s=n$, т. е. до полного полинома.

Таким образом, общее число p_n всевозможных частных моделей, построенных из n аргументов (слагаемых) полного полинома методом полного комбинаторного перебора, вычисляется по формуле

$$p_n = \sum_{s=1}^n C_n^s = 2^n - 1.$$

Очевидно, что p_n при увеличении n очень быстро возрастает (например, $p_{10}=1024$, а $p_{15}=32\,768$). Поэтому возможности комбинаторного перебора ограничены: например, алгоритм [146], где изменение сложности частных моделей выполняется не последовательно, а по более универсальной схеме, основанной на двоичном n -разрядном счетчике, за приемлемое время перебирает все модели при $n \leq 18$. Добавление одного аргумента увеличивает время счета примерно вдвое.

Существуют некоторые способы ускорения счета при полном переборе. Например, оптимальная схема перебора [148], основанная на вычислении коэффициентов последовательно получаемых частных моделей с помощью метода окаймления, позволяет увеличить n примерно до 23.

При большем n приходится применять целесообразное усечение перебора. Например, известен способ «усечения треугольника перебора»: рассматриваются все модели со сложностью от $s=1$ до $s_{\max}=m$, а модели с большим числом аргументов исключаются из перебора. При этом объем перебора определяется формулой

$$p_m = \sum_{s=1}^m C_n^s \ll 2^n - 1.$$

Максимальная сложность s_{\max} задается, исходя из возможностей ЭВМ (например, в алгоритме [146] так, чтобы $p_m < 2^{18}$). Это ограничение может определяться также числом точек исходных

данных, точнее, длиной обучающей последовательности: если $N_A < n$, то $m = N_A$.

2. Гармонический алгоритм с дискретным набором гармоник. Точный аналог полиномиального алгоритма может быть построен на основе [69]. При этом интервал главных значений частот иско-мых гармоник длиной 2π разбивается (вообще говоря, произвольно) на n промежутков. Каждому полученному дискретному значению частоты ω_i ставится в соответствие «переобозначенный гармонический аргумент» вида

$$x_i = C_i + A_i \sin \omega_i t + B_i \cos \omega_i t, \quad i = \overline{1, n},$$

коэффициенты C_i, A_i, B_i определяются по МНК на всех N точках исходных данных. Далее, в полном соответствии с предыдущим алгоритмом выполняется полный перебор частных моделей различной сложности: при $s=1$ C^1_n моделей содержат одну гармонику (с уже фактически вычисленными коэффициентами C_i, A_i, B_i), при $s=2$ C^2_n моделей — две различные гармоники, и так до $s=n$.

При этом, начиная с $s=2$, определяются коэффициенты моделей по отношению к уже вычисленным «гармоническим аргументам»:

$$q_i = a_0 + a_1 x_i + a_2 x_j.$$

Очевидно, что объем перебора $p_n = 2^n - 1$, и для такого алгоритма справедливы те же ограничения, что и выше. Кроме того, из-за дискретности задания частот истинные частоты колебательного процесса могут быть пропущены [22], причем увеличение числа дискрет n , как правило, не помогает (время счета при этом резко возрастает).

3. Гармонический алгоритм с аналитическим определением частот процесса. Поскольку такой алгоритм был предложен и подробно описан в [22], здесь напомним только его идею. В нем использовано специальное балансовое соотношение ординат колебательного процесса (которое является свойством такого процесса), позволяющее разделить задачи определения частот и амплитуд модели процесса в виде суммы гармоник с некратными частотами:

$$q(t) = a_0 + \sum_{i=1}^s (A_i \sin \omega_i t + B_i \cos \omega_i t).$$

При этом частоты определяются независимо от амплитуд и однозначно (а не задаются, как ранее) из специального нелинейного алгебраического уравнения, а коэффициенты (амплитуды) вычисляются после этого по МНК.

Таким образом, комбинаторный гармонический алгоритм в этом случае работает по такой схеме: находятся частные модели при постепенном увеличении сложности — с одной гармоникой ($s=1$), с двумя гармониками и т. д. до заданного числа n гармо-

ник ($s=n$). Следовательно, в результате аналитического определения частот гармонических составляющих полное число частных моделей $p=n$. Из полученных моделей отбирается лучшая по заданному комбинированному критерию.

Несмотря на небольшой объем перебора моделей, возможности этого алгоритма также ограничены как затратами времени счета на ЭВМ (операция определения частот гармоник является весьма трудоемкой), так и количеством исходных данных: для синтеза модели с n гармониками требуется число точек $N \geq 3n+1$, что даже превосходит требования теоремы Котельникова ($N \geq 2n$). Однако в пределах ограничений алгоритм позволяет восстановить точную модель колебательного процесса, состоящего из суммы гармонических составляющих.

4. Гармонический алгоритм с перебором частот, определенных аналитически. На основе предыдущего алгоритма можно предложить еще один точный аналог полиномиального комбинаторного алгоритма, свободный от недостатков алгоритма с дискретным набором гармоник. Допустим, что по алгоритму 3 получены l моделей, содержащих $1, 2, 3, \dots, l$ гармонических составляющих с различными частотами. Всего в этих l моделях, таким образом, имеется $n = 1 + 2 + \dots + l = \frac{1}{2}l(l+1)$ гармоник.

Переобозначив, как и в алгоритме 2, все эти составляющие набором n эквивалентных «гармонических аргументов» x_1, x_2, \dots, x_n , получим возможность обращаться с ними, как и в обычном полиномиальном алгоритме. Но этот алгоритм принципиально отличается от алгоритма 2: там частоты фактически назначены, а здесь определены аналитически, т. е. они максимально близки к истинным гармоникам колебательного процесса.

4.2. Многорядные алгоритмы без вычисления остатков

Структурная схема многорядных алгоритмов (без вычисления остатков) приведена на рис. 1. Считаем, что исходный состав аргументов, с которого начинается процедура многорядной селекции модели процесса, строится на так называемом нулевом ряду алгоритма, который организуется по-разному в полиномиальных и гармонических алгоритмах. Например, в классе алгебраических функций наиболее общей моделью является полином Колмогорова — Габора от k переменных:

$$q = a_0 + \sum_{i=1}^k a_i z_i + \sum_{i=1}^k \sum_{j=1}^k a_{ij} z_i z_j + \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k a_{ijl} z_i z_j z_l + \dots,$$

представляющий собой сумму линейных, квадратичных, кубических и других членов. После переобозначения всех имеющихся в нем слагаемых получим линейный полином

$$q = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n.$$

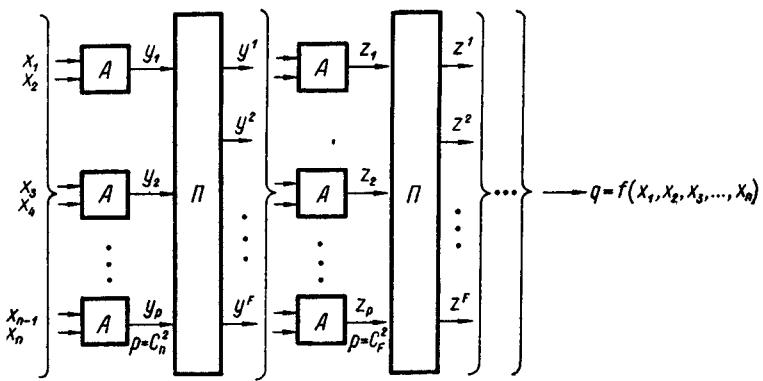


Рис. 1. Многорядный алгоритм МГУА без вычисления остатков (Π — пороговый отбор F лучших моделей; A — алгоритм попарной обработки аргументов).

Члены этого полинома и составляют исходный набор аргументов, построенный на нулевом ряду многорядного алгоритма.

В случае разностных моделей к аналогичному виду сводится модель, полученная с помощью переобозначения всех исходных переменных, их запаздываний и заданных нелинейных функций от них.

Для многорядных гармонических алгоритмов (как с вычислением остатков, так и без него) исходный состав аргументов должен строиться на нулевом ряду так же, как и в описанном комбинаторном гармоническом алгоритме 4: сначала получаем l моделей в виде суммы одной, двух, трех и т. д. до l гармоник с некратными частотами, а затем все имеющиеся в этих моделях $n = \frac{1}{2}l(l+1)$ гармонических составляющих переобозначаем в виде новых аргументов:

$$x_i = A_i \sin \omega_i t + B_i \cos \omega_i t, \quad i = \overline{1, n}.$$

После указанных переобозначений получаем однотипный базисный набор исходных аргументов как для полиномиальных, так и для гармонических алгоритмов. Поэтому ниже будем характеризовать многорядные схемы перебора общего вида, указывая только на специфику интерпретации синтезированных моделей в полиномиальном и гармоническом случаях.

Многорядные алгоритмы, как правило, работают по следующей общей схеме:

1-й ряд — на основе данных таблицы наблюдений строятся частные описания от всех попарных комбинаций исходных (переобозначенных) аргументов, приближающие по МНК выходную переменную q

$$y_1 = f_1(x_1, x_2), \quad y_2 = f_2(x_1, x_3), \dots, \quad y_k = f_k(x_{n-1}, x_n)$$

(из этих $k = C_n^2$ моделей отбирается некоторое число F_1 лучших по внешнему критерию);

2-й ряд — полученные переменные принимаются в качестве аргументов 2-го ряда, и снова строятся все частные модели от двух аргументов

$$z_1 = \varphi_1(y_1, y_2), \quad z_2 = \varphi_2(y_1, y_3), \dots, \quad z_l = \varphi_l(z_{F_1-1}, z_{F_1}), \quad l = C_{F_1}^2.$$

Из них по внешнему критерию отбираются F_2 лучших моделей в качестве переменных следующего ряда и т. д. Ряды наращиваются до тех пор, пока снижается значение внешнего критерия.

Каждое частное описание может быть линейной

$$f = a_0 + a_1 x_i + a_2 x_k$$

или нелинейной

$$f = a_0 + a_1 x_i + a_2 x_k + a_3 x_i x_k,$$

$$f = a_0 + a_1 x_i + a_2 x_k + a_3 x_i x_k + a_4 x_i^2 + a_5 x_k^2$$

функцией от двух переменных, коэффициенты которых можно определить по МНК, имея соответственно 3, 4 или 6 точек наблюдений в обучающей последовательности. Исключив промежуточные переменные после останова алгоритма, получим модель, число коэффициентов в которой значительно превосходит число точек.

Таким образом, в многорядных алгоритмах МГУА воспроизводится схема массовой селекции. В них есть генератор усложняющихся из ряда в ряд частных описаний («скрещивание») и пороговые самоотборы лучших из них («селекция»), т. е. реализуется «гипотеза селекции» [43].

Опишем коротко основные схемы перебора в многорядных алгоритмах МГУА, отметив, что для гармонических алгоритмов все они являются новыми.

Алгоритм с линейными частными описаниями. В качестве частного описания здесь используется полином

$$y = a_0 + a_1 x_i + a_2 x_j.$$

При этом синтезированная модель может содержать переменные (гармоники или члены полинома) только из исходного базиса, полученного после переобозначения на нулевом ряду.

Алгоритм с нелинейными частными описаниями. При использовании даже простейшего нелинейного частного описания

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j,$$

синтезированная модель, «развернутая» (при помощи исключения промежуточных переменных) после останова алгоритма, может содержать члены, отсутствующие в исходном базисе. Такой алгоритм автоматически наращивает сложность модели (в том числе при помощи постоянного расширения базиса от ряда к ряду), «руководствуясь» только условием убывания внешнего критерия.

Совершенно новые возможности в этом случае открываются для гармонических алгоритмов. Учитывая, что аргументами являются тригонометрические функции, видим, что члену $x_i x_j$ в частном описании соответствует сумма произведений синусов и косинусов вида

$$x_i x_j = \sin \omega_i \sin \omega_j + \sin \omega_i \cos \omega_j + \cos \omega_i \sin \omega_j + \cos \omega_i \cos \omega_j,$$

(для простоты коэффициенты опускаем).

Известные формулы тригонометрии

$$\sin \omega_i \sin \omega_j = \frac{1}{2} [\cos(\omega_i - \omega_j) + \cos(\omega_i + \omega_j)],$$

$$\sin \omega_i \cos \omega_j = \frac{1}{2} [\sin(\omega_i - \omega_j) - \cos(\omega_i + \omega_j)]$$

показывают, что в этом случае мы приходим к новым гармоническим составляющим, содержащим суммы и разности исходных частот.

Кроме того, применение частного описания вида

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2$$

приводит к появлению кратных частот $2\omega_i$, $2\omega_j$, так как

$$\sin^2 \omega_i = \frac{1}{2} (1 - \cos 2\omega_i).$$

Таким образом, нелинейный многорядный алгоритм расширяет возможности успешного моделирования колебательных процессов, поскольку наращивает число синтезируемых частот (даже при малом числе точек). Алгоритм вводит или отсеивает новые «размножающиеся» частоты, руководствуясь условием уменьшения внешнего критерия.

Модифицированный гармонический алгоритм с переобозначением переменных. Так как при большом числе рядов могут возникнуть трудности при «разворачивании» синтезированной модели, можно модифицировать нулевой ряд линейного алгоритма следующим образом. Получив n гармонических составляющих, соответствующих l моделям в виде сумм разного числа гармоник (число l определяется объемом имеющихся данных), существенно увеличим полученный базисный набор частот, введя новые гармоники с помощью «размножения» частот

$$\omega'_i = \omega_i - \omega_j, \quad \omega'_j = \omega_i + \omega_j,$$

$$\omega^*_i = 2\omega_i, \quad \omega^*_j = 2\omega_j,$$

$$\omega''_i = \omega'_i - \omega'_j, \quad \omega''_j = \omega'_i + \omega'_j,$$

$$\omega^{**}_i = 2\omega^*_i = 4\omega_i, \quad \omega^{**}_j = 2\omega^*_j = 4\omega_j,$$

$$\omega'''_i = \omega''_i - \omega''_j, \quad \omega'''_j = \omega''_i + \omega''_j,$$

и т. д., ориентируясь только на возможности ЭВМ работать с расширенным гармоническим базисом при линейных частных описаниях.

Алгоритм с использованием исходных базисных переменных. Данный алгоритм отличается от предыдущих только тем, что в качестве переменных r -го ряда используются как F_{r-1} лучших моделей предыдущего ряда, так и n исходных аргументов. Он может работать с линейными и нелинейными частными описаниями. Целесообразность «отхода» от исходного базиса или «приближения» к нему определяется по условию уменьшения внешнего критерия (процедуры такого типа описаны в [174, 175]).

Алгоритм с передачей в следующий ряд списка переменных. Такой алгоритм предназначен прежде всего для отсева несущественных аргументов из исходного базиса и отбора заданного числа наиболее значимых переменных или гармоник. На первом ряду селекции используется линейное частное описание

$$y = b_0 + b_1 x_i + b_2 x_j, \quad i, j = \overline{1, n}, \quad i \neq j.$$

Всевозможные частные описания идентифицируются по МНК и оцениваются на плоскости двух критериев: а) для динамических (разностных) моделей ($n_{cm}^2 - i^2(N)$); б) для статических (алгебраических) моделей ($n_{cm}^2 - \Delta^2(C)$).

Из всего числа частных моделей, равного $F_0 = C^2 n$, выбираются F_1 лучших по указанным критериям моделей (ближайших к началу плоскости).

На втором ряду селекции составляется список эффективных переменных, попавших в число аргументов моделей, отобранных на первом ряду. Это единственная информация, используемая по данным первого ряда. Модели первого ряда «забываются» (кроме списка эффективных переменных). Снова составляется линейный полином

$$q = c_0 + c_1 x_1 + \dots + c_{F_1} x_{F_1},$$

где x_1, \dots, x_{F_1} — переменные, указанные в списке, причем $F_1 \ll n$. Частное описание второго ряда $y = d_0 + d_1 x_i + d_2 x_j$ позволяет отобрать по указанным критериям F_2 лучших моделей и составить второй, еще более сокращенный список переменных.

Нарашивание рядов селекции продолжается с заданным постоянным «коэффициентом редукции» до тех пор, пока количество эффективных переменных в последнем списке не уменьшится до заданного числа.

Другой алгоритм с передачей списков разработан в [139].

4.3. Многорядные алгоритмы МГУА с вычислением остатков

Структура многорядных алгоритмов МГУА с вычислением остатков показана на рис. 2, где 1 — таблица исходных данных; 2 — нулевой ряд (здесь по комбинаторному алгоритму МГУА выбира-

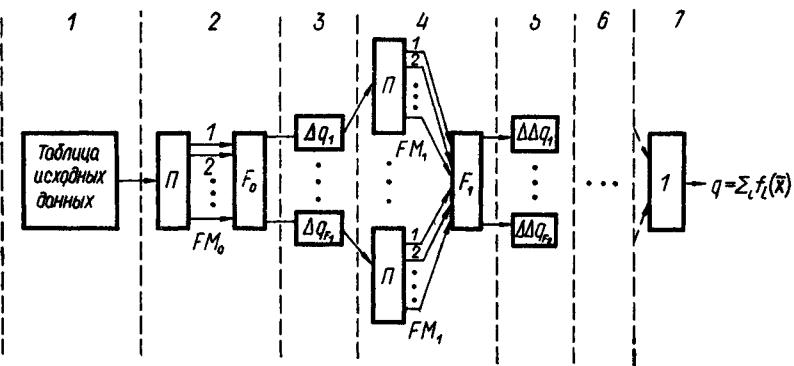


Рис. 2. Многорядный алгоритм МГУА с вычислением остатков.

ются F_0 лучших моделей); 3 — вычисление первых разностей (остатков) $\Delta q = q - q^*$ (q — фактические данные таблицы; q^* — выход модели), составление F_0 таблиц данных первых остатков; 4 — первый ряд селекции (здесь по комбинаторному алгоритму МГУА находятся F_{0p_n} моделей, из которых выбираются F_1 лучших); 5 — вычисление вторых разностей (остатков) $\Delta^2 q = \Delta q - \Delta q^*$ (Δq — фактическое значение первого остатка; Δq^* — выход модели); 6 — последующие ряды селекции и вычисления остатков; 7 — последний ряд с выбором модели оптимальной сложности.

При наращивании рядов селекции такого алгоритма комбинированный критерий селекции $\rho^2 = n^2_{\text{см}} + \Delta^2(C) \rightarrow \min$ монотонно уменьшается. Поэтому сигналом для останова счета служит достаточно малое его уменьшение (по правилу «левого угла» [90]). Объективный характер выбора модели теряется.

Алгоритмы с вычислением остатка также позволяют получить описания, число членов в которых значительно превышает алгебраический минимум точек (для полиномов) или число гармоник, разрешаемое теоремой Котельникова (для гармонических моделей).

Если строятся модели для долгосрочного прогноза, то вместо пары критериев $n^2_{\text{см}}$, $\Delta^2(C)$ следует использовать пару критериев $n^2_{\text{см}}$, $i^2(N)$. Последнее возможно только для полиномиальных разностных моделей, так как аналог критерия $i^2(N)$ для алгебраических и гармонических моделей пока не предложен.

Алгоритмы МГУА с вычислением остатков разработаны как для полиномиальных (разностных) [77], так и гармонических [22] моделей. Как отмечалось, эти алгоритмы могут давать идентичные результаты решения задач моделирования и прогнозирования, если в полиномиальном (разностном) алгоритме учесть ряд ограничений на коэффициенты синтезируемой модели. Эти ограничения подробно обсуждались в [115]. При их отсутствии разностный многорядный алгоритм с вычислением остатка является фактически экспоненциально-гармоническим алгоритмом [77].

4.4 Объективный системный анализ в задачах, где объект описывается системой уравнений

До сих пор мы проводили параллельное рассмотрение полиномиальных и гармонических многорядных алгоритмов МГУА в предположении, что модель может быть представлена одним уравнением (полиномом или суммой гармоник).

В ряде задач объект имеет разветвленную структуру из нескольких блоков, каждый из которых должен быть описан своим уравнением. В этом случае модель должна быть представлена системой уравнений, отражающих закономерности отдельных блоков объекта.

ЭВМ при помощи перебора большого числа вариантов может объективно найти все блоки и их закономерности, скрытые в данной суммарной таблице опытных данных, взятых для всей системы в целом, если воспользоваться следующим определением: «закономерностью» условимся называть систему уравнений, которая: а) остается действительной на любой части таблицы опытных данных («критерий минимума смещения модели» $n_{\text{см}} \rightarrow \min$) и б) на «хорошо» ($\Delta(C) \leq 0,5$) или «удовлетворительно» ($\Delta(C) \leq \leq 0,8$) обеспечивает краткосрочный прогноз («критерий точности краткосрочного прогнозирования» $\Delta(C) \rightarrow \min$).

Объективное определение множества выходных переменных и числа уравнений, описывающих объект, сводится к перебору всех вариантов по указанным критериям при постепенном увеличении числа учитываемых строк таблицы исходных данных. Нарашивание рядов селекции прекращается, как только системный критерий начал повышаться [70].

Вблизи минимума кривой размещается целая последовательность характеристических точек, отвечающих лучшим моделям. Не обязательно выбирать модель, соответствующую самой нижней точке. По вторичному критерию (например, по критерию точности прогноза или баланса переменных и др.) мы можем сравнить несколько лучших моделей и таким образом выбрать модель по двум критериям.

5. ОСОБЕННОСТИ МГУА, ОБЪЯСНЯЮЩИЕ ЕГО ЭФФЕКТИВНОСТЬ В ЗАДАЧАХ МОДЕЛИРОВАНИЯ И ДОЛГОСРОЧНОГО КОЛИЧЕСТВЕННОГО ПРОГНОЗА

Математическое прогнозирование происходящих в мире процессов является одним из главных средств для нахождения путей «сбалансированного роста» — основной цели работ многих общественных и научных организаций, начало которым положил известный Римский клуб [131], объединивший многих видных ученых и общественных деятелей.

Прогнозы динамики развития мира, данные Дж. Форрестером [161] и его учениками, прозвучали как набат о приближающихся со многих сторон опасностей. Однако, как пишет А. Печеи, «пора

перейти уже от стадии обычного шока — который был необходим, чтобы привлечь внимание к близящейся опасности, — к новому этапу позитивного взгляда на то, чего реально может достигнуть в обозримом будущем человечество» [131].

Позитивный взгляд требует привлечения методов точного количественного долгосрочного прогноза. Информация качественного характера, например такая, как «тропические леса уничтожаются со скоростью 20 га в минуту», «производство ядохимикатов увеличивается за каждые четыре года в два раза» и т. п., может только привлечь внимание, но не способствует принятию деловых решений. Тому пример быстрое уменьшение (в течение всего нескольких лет!) рождаемости или фактическое прекращение роста процента научных работников во многих странах мира [154]. Сработали какие-то не вполне ясные для нас внутренние механизмы, ограничившие рост указанных двух переменных величин.

Достаточно быстрое изучение и раскрытие этих механизмов в форме уравнений просто невозможно и не требуется для точного прогноза и принятия решений. Необходимо расширить арсенал используемых сейчас методов прогнозирования, в первую очередь за счет привлечения (наряду с использованием уже известных) новых методов самоорганизации прогнозирующих моделей.

Современные модели для долгосрочного количественного прогноза случайных процессов реализуют в основном идеи имитационного моделирования, при котором использование физических моделей заводит в тупик, так как долгосрочные прогнозы жестко сходятся к средней линии (тренду). Класс уравнений и вид опорных функций априорно назначаются автором модели, как и множество выходных (зависимых) и входных (независимых) переменных. Это приводит к слишком большому числу негибких и смещенных моделей, вид и оценки коэффициентов которых резко изменяются с появлением каждой новой точки данных.

В итоге в отраслях науки, особенно требующих долгосрочных прогнозов (например, в метеорологии, климатологии, экологии, экономике и др.), нет успешных количественных прогнозов. Можно говорить только о некотором успехе размытых, долгосрочных качественных (синоптических) прогнозов, таких, как «похолодаёт», «возможные грозы» и т. п. Все известные количественные долгосрочные прогнозы (в том числе прогнозы Римского клуба) не выполняются.

Это послужило появлению утверждений о непредсказуемости окружающего нас мира, будто бы мир устроен так, что долгосрочное количественное прогнозирование будущего принципиально невозможно. Знать будущее нам не дано.

В отличие от методов имитационного моделирования, имеющих (как и породившая их математическая физика) дедуктивный характер, МГУА основан на индуктивном подходе, другими словами — на переборе моделей по внешним критериям, согласованным между экспертами. Самоорганизацией моделей называется нахождение математического описания сложного объекта при помощи

перебора многих вариантов по некоторому ансамблю внешних критериев.

Самоорганизация нефизических моделей и применение схем двухуровневого прогнозирования позволяют поручить выбор модели ЭВМ, что значительно повышает время упреждения прогноза. Точность, заблаговременность и объективный характер долгосрочных прогнозов по МГУА позволяет ускорить формирование «общественного понимания» процессов, происходящих в природе и обществе, и, главное, дадут возможность избежать ложных субъективных решений. Теория самоорганизации решает задачи rationalизации перебора моделей, с тем чтобы он был однозначным, помехоустойчивым и реально осуществимым на ЭВМ.

МГУА не устраивает экспертов, а отводит им особое место: они указывают ЭВМ критерии выбора весьма общего вида и интерпретируют модели, рекомендуемые машиной как лучшие (т. е. участвуют в процессе доопределения модели оптимальной сложности). Эксперты могут влиять на результат моделирования только через постановку новых критериев. При этом МГУА становится объективным судьей при решении научных споров.

При известных затруднениях других методов количественного долгосрочного прогноза МГУА претендует на решение этой важной задачи. Ниже указаны шесть особенностей МГУА, позволяющих улучшить прогнозирование моделей сложных объектов и придать им объективный характер.

Первая особенность: самоорганизация физической модели для познания объекта исследований и нескольких нефизических — для долгосрочного прогноза.

Критерий минимума смещения модели требует, чтобы модели, полученные на двух разных частях таблицы данных наблюдений, были одинаковыми. Самоорганизация модели по критерию минимума смещения приводит к физической модели. Так называется модель, изоморфная механизму моделируемого объекта и адекватная ему.

Человек, глубоко изучивший объект, может предложить только физическую его модель. Но оказывается, что для долгосрочного прогноза физические модели не подходят. Исследователи Кругиус провели 20 лет на берегах оз. Дальнего (Камчатка) [114]. С максимально возможной точностью и тщательностью они сами в любую погоду измеряли все необходимые для моделирования экологической системы озера данные. В результате получили исчерпывающую информацию об основных механизмах системы. Математики помогли formalизовать выводы наблюдений. Так была получена физическая модель оз. Дальнего, одна из первых и лучших имитационных моделей сложной системы. Предполагалось, что модель позволит получить долгосрочный количественный прогноз продуктивности озера (урова рыбы). Однако оказалось, что модель хотя и имеет большую познавательную ценность, но пригодна только для краткосрочного прогноза с заблаговременностью не более трех лет. Это легко объяснимо. Экологическая система озе-

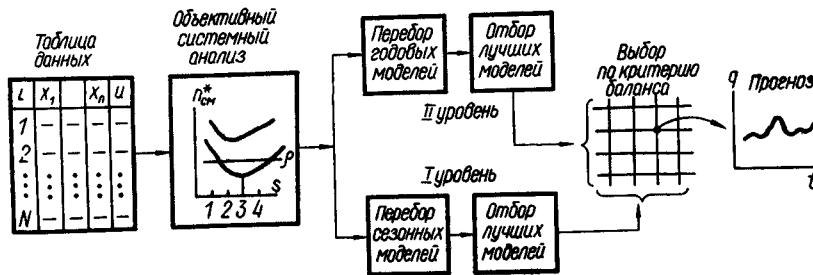


Рис. 3. Структура алгоритма двухуровневого долгосрочного прогнозирования.

ра в интервале наблюдений устойчива — иначе она бы не существовала столь продолжительное время. Устойчивому объекту соответствует физическая модель (система дифференциальных и алгебраических уравнений), у которой корни характеристического уравнения устойчивы, т. е. имеют отрицательную действительную часть. Внешние возмущения задаются по календарным данным. При пошаговом интегрировании разностных аналогов устойчивых уравнений прогнозы обязательно жестко сходятся к тренду. Модель как бы показывает, что все колебательные процессы в озере с течением времени замирают, а это не соответствует действительности [122]. Таким образом, физическая модель устойчивого объекта не пригодна для долгосрочного количественного прогноза.

В МГУА самоорганизация физической модели используется только для того, чтобы выделить состав (список) множества выходных переменных, которые подлежат прогнозированию. Зная это множество, переходят к самоорганизации прогнозирующих нефизических моделей.

Нефизические модели отличаются от физических как по информационному базису (по составу аргументов), так и по критериям селекции: важное значение приобретает критерий точности пошагового прогноза. Нефизические модели могут быть построены на различных математических языках, отличающихся по степени размытия (измеряемой временем корреляции [62]) (например, для среднегодовых, сезонных или месячных данных). Наконец, нефизические модели могут содержать как годовые, так и сезонные или месячные данные одновременно, т. е. использовать многомерный отсчет времени. Модель становится двухмерной или даже трехмерной (пространственной). Все это важно потому, что предел информативности прогноза определяется степенью размытия математического языка моделирования. Более размытый язык (например, модели для среднегодовых значений) соответствует большему времени корреляции и, следовательно, более длительному пределу заблаговременности (информационности) прогноза.

Нефизические модели используются в МГУА для организации многоуровневого прогноза (рис. 3), при котором критерий баланса прогнозов позволяет увеличить время заблаговременности де-

タルных (сезонных) прогнозов до величины времени заблаговременности более размытых (годовых) прогнозов. Предварительная самоорганизация годовых и сезонных прогнозов по вспомогательным критериям необходима для обеспечения однозначности выбора прогнозов по критерию баланса и сокращения объема перебора. Принципиальным является то, что критерий баланса используется для выбора не только сезонных, но и более размытых годовых прогнозов. Критерий как бы является сигналом из отдаленного будущего для выбора как тех, так и других прогнозов.

Вторая особенность: выбор класса уравнений и вида опорной функции поручается ЭВМ, которая перебирает многие варианты по критериям выбора модели. Обычно для сложных систем машина выбирает разностные нелинейные модели с переменными коэффициентами, гибкие решения которых наиболее адекватны богатству процессов, реально наблюдаемых в системе.

Априорное назначение класса уравнений и вида опорных функций приводит к малоинформативным, «скучным» прогнозам. В современных моделях сложных объектов (в том числе в самых совершенных) класс уравнений и вид опорных функций априори назначаются человеком — автором модели. Так было в моделях динамики мира Дж. Форрестера [161], так обстоит дело и в экономической модели ЛИНК [30]. В аналогичных разработках, например в экологии [16], та же картина: используются разностные уравнения с одним запаздывающим аргументом при постоянных коэффициентах, т. е. такого же вида, как в указанных выше моделях. Непрерывным аналогом подобных уравнений является линейное дифференциальное уравнение первого порядка, решение которого — экспонента. Представим себе систему дифференциальных уравнений первого порядка, почти не связанных между собой. Что можно выбрать в качестве прогноза в решениях такой системы? Не удивительно, что прогнозы таких процессов по имитационным моделям никого не удовлетворяли. Бедность возможностей математического описания и в том и в другом случае привели к малоинформативным прогнозам. Как пояснялось, свободная составляющая процессов быстро затухает и долгосрочный прогноз определяется заданным законом изменения внешнего воздействия.

Третья особенность: выбор множества выходных и входных переменных, а также «ведущей» переменной поручается ЭВМ.

Априорное назначение выходных (зависимых) и входных (независимых) переменных приводит к смещенным уравнениям. В имитационных моделях выходные (зависимые) переменные моделей назначаются произвольно, так же как и независимые (аргументы). Например, в блоке «Великобритания» модели ЛИНК используется 227 уравнений, только частично взаимосвязанных между собой. Слишком большое число уравнений не позволяет выбрать только те из них, которые являются несмещенными, т. е. не зависят от интервала времени, где они идентифицируются. МГУА позволяет объективно выбрать наименее смещенную систему

му уравнений при помощи перебора вариантов по системному критерию минимума смещения.

Согласно этому критерию только те уравнения «заслуживают» право войти в состав модели, которые остаются неизменными при приходе новых данных или изменении интервала (участка) таблицы. Алгоритм МГУА для объективного системного анализа позволяет перебрать по очереди все возможные системы уравнений, состоящие из одного, двух, трех и т. д. уравнений, и выбрать самую несмещенную физическую модель. Оптимальная по критерию минимума смещения система уравнений определяет собой структуру объекта и множество выходных переменных.

В модель экономики Англии, синтезированную по МГУА [70], вошло только небольшое число переменных. Из тридцати исходных переменных в качестве выходных величин алгоритм выбрал следующие шесть: доходы частных лиц; капиталовложения (общие); индекс розничных цен; экспорт; отношение «аналоги на компании/доходы компаний»; расходы правительства (общие). Только эти переменные дают несмещенные модели. Остальные 24 переменные исследуются как «ведомые», т. е. зависимые от значений ведущих переменных и времени.

Четвертая особенность: плохо прогнозирующиеся (ведомые) переменные прогнозируются дополнительно, во вторую очередь.

Интересующие нас переменные могут не войти во множество выходных переменных, отобранных машиной на этапе самоорганизации физической модели. Тогда их следует прогнозировать дополнительно в функции выходных переменных и времени, прогноз которых получается по описанному выше двухуровневому алгоритму. Так, число безработных в Англии прогнозируется моделью в виде функции времени и указанных выше шести ведущих переменных, прогнозы которых уже получены ранее [49, 70].

Пятая особенность: возможность прогнозирования при неполном информационном базисе.

МГУА позволяет получать точные прогнозы при неполном информационном базисе, т. е. без измерения многих существенных аргументов. Такой факт противоречит обычно принятым взглядам. Например, при моделировании экономических систем часто высказывается уверенность, что создание большой вычислительной системы, моделирующей экономику всех стран мира (т. е. построение автономной всемирной системы уравнений), может решить проблему долгосрочного прогноза. Это неверно.

В метеорологии уже существуют и действуют подобные «всемирные» системы. Вся атмосфера планеты разделена на боксы размером 5° . Для каждого из них составлена физическая модель в виде системы дифференциальных уравнений. Необходимая информация вводится в программу почти непрерывно. И что же? Оказывается, что подобная система может рассчитывать погоду только на 5—7 суток вперед.

Опубликованные математические исследования показывают, что предел заблаговременности прогноза по дифференциальным

уравнениям погоды принципиально не может превысить 15 суток [19]. Для каждого математического языка моделирования существует свой предел заблаговременности прогноза.

Таким образом, вычислительные системы, работающие на любом одном языке, не решают задачи количественного долгосрочного прогноза, хотя и обеспечивают краткосрочный прогноз (в пределах заблаговременности своего языка моделирования).

Моделирование по МГУА при неполном информационном базисе является полной противоположностью идеи увеличения информационного базиса модели до учета максимального числа всех воздействий. МГУА часто демонстрирует успех такого моделирования. Например, для прогноза урожая пшеницы требуется ввести в модель более 50 факторов (аргументов): удобрения, способ обработки, сроки полива и т. п. [166]. Все аргументы являются весьма существенными. И тем не менее в результате самоорганизации модели в ней могут участвовать только два-три аргумента при достаточной точности прогноза. Дело в том, что аргументы взаимосвязаны, и хотя эти связи остаются нам неизвестными, самоорганизация моделей позволяет учесть их. Минимальное число аргументов, достаточное для точного прогноза, определяется разностью [49] $s > n - m$, где n — полное число аргументов; s — число аргументов модели; m — число внутренних связей (неизвестных нам) в объекте.

Шестая особенность: самоорганизация физической и прогнозирующих моделей возможна при сильно зашумленных исходных данных.

Алгоритмы МГУА подчинены законам, которые действуют в системах связи. В частности, в них действует предельная теорема Шеннона для каналов с шумом [172]. Например, геометрическое место точек минимумов критерия регулярности (по которому выбирается модель оптимальной сложности) соответствует указанной теореме Шеннона. С увеличением помех полоса канала (как и сложность модели) должна уменьшаться с целью достижения предельно возможной пропускной способности (минимума критерия) (подробнее см. гл. 2).

В настоящее время алгоритмы МГУА позволяют восстановить физическую модель объекта в случае, когда помехи измерения данных в несколько раз превосходят регулярный сигнал [86]. Перенесение опыта теории связи и распознавания образов позволяет поднять помехоустойчивость МГУА еще во много раз. Тогда можно будет снизить требования к точности информации, необходимой для моделирования и прогнозирования. Точные модели можно получить при неполном информационном базисе и при неточных данных так же, как связисты восстанавливают точный сигнал при наличии помех и разнородных искажений.

МГУА является объективным, так как все вопросы решаются в нем по минимуму целесообразно выбранных и согласованных критериев коллекции моделей. Результат самоорганизации может

не совпадать с представлениями автора модели о моделируемом объекте.

Убедительная объективная идентификация точной физической модели, определяющая множество хорошо прогнозируемых выходных переменных по критерию смещения, и последующая самоорганизация нефизических прогнозирующих моделей по критерию баланса квартальных и годовых прогнозов являются основным результатом МГУА, способствующим решению проблемы точного количественного долгосрочного прогноза и тем самым созданию объективного общественного понимания исследуемых систем.

6. ОБЛАСТИ ПРИМЕНЕНИЯ РАЗЛИЧНЫХ АЛГОРИТМОВ МГУА

МГУА является эмпирическим методом и предназначен для решения задач самоорганизации математических моделей по данным, представленным в таблице наблюдений.

Модели необходимы для идентификации объекта и краткосрочного или долгосрочного прогноза. Без дополнительной априорной информации задача является некорректной, т. е. требует регуляризации. Обычные способы регуляризации (задание структуры полинома, критерий гладкости Тихонова, минимум собственного вектора матрицы в методе псевдообратных матриц и др.) представляются как математически изящные, но не всегда эффективные внешние дополнения.

В теории самоорганизации предложен целый ряд более целесообразных критериев (регулярности, минимума смещения, точности пошагового прогноза и др., см. гл. 5), направленных на удовлетворение практических заказчиков моделей. Эти критерии применяются последовательно, один за другим, что исключает трудности нормировки (выбора коэффициентов веса) критерии. Многокритериальный выбор модели — одна из основ помехоустойчивости МГУА.

Геделевский подход состоит в данном случае в том, что вопрос о выборе ансамбля критериев, его состава и последовательности применения решается при помощи сравнения вариантов по критериям второго (верхнего) уровня. Такими критериями обычно служат:

- 1) критерий потенциальной помехоустойчивости $\Theta_1 \sim \frac{\text{помеха}}{\text{сигнал}} \rightarrow \text{так}$;
- 2) критерий упреждения прогноза: $\Theta_2 \sim \frac{\text{время упреждения}}{\text{время наблюдения}} \rightarrow \text{так}$;
- 3) критерий объема вычислений: $\Theta_3 \sim \text{время работы ЭВМ} \rightarrow \text{так}$.

Тот ансамбль критериев или тот алгоритм МГУА лучше, который обеспечивает требуемое значение пары критериев Θ_1 и Θ_2 . Объективность МГУА опирается на то, что выбор множества критериев второго уровня очевиден и обычно не вызывает возражений экспертов.

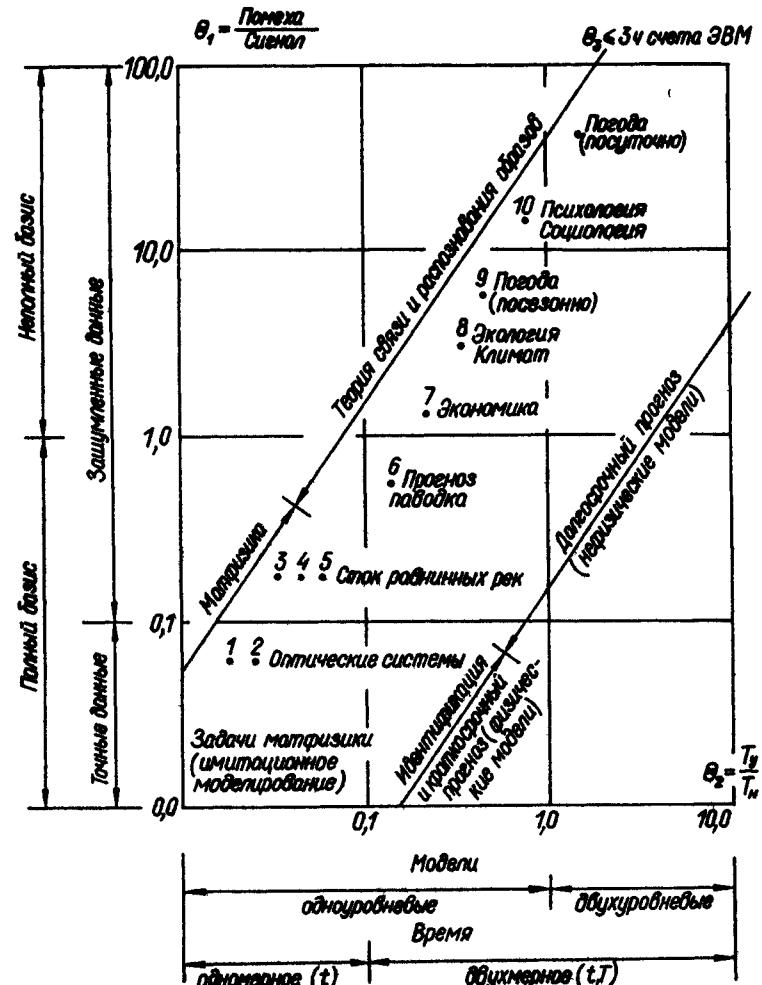


Рис. 4. Области применения различных алгоритмов МГУА.

На рис. 4 показаны области применения различных алгоритмов МГУА в зависимости от значений Θ_1 и Θ_2 (Θ_3 ограничен тремя часами счета на БЭСМ-6: $\Theta_3 \leq 3$ ч).

Имитационные задачи, как и задачи математической физики, требуют применения одноуровневых моделей с обычным одномерным отсчетом времени. Исходные данные в этих задачах — точные и информационный базис — полный (ни один аргумент не забыт). Примером могут служить модели оптических систем [49, 179].

Двухмерное время (в сезонах и годах) при зашумленных данных требуется в задаче прогноза стока рек, особенно прогноза весеннего паводка [90, 22, 85, 87].

Двухуровневый долгосрочный прогноз при двухмерном отсчете времени и неполном информационном базисе применяется для эконометрических моделей, моделирования изменений климата и построения моделей экологических систем [70, 75, 82]. Здесь возникают новые теоретические вопросы (в частности, сходимость к «точке сгущения» должна быть заменена сходимостью к тренду и др.).

Более сложные задачи моделирования при $\Theta_1 > 10$ и $\Theta_2 > 1$ требуют разработки новых помехоустойчивых алгоритмов МГУА, использующих как специальные формулы для расчета критериев, так и специальные виды корреляционных алгоритмов, один из которых описан в [60].

Здесь открывается возможность перенесения в теорию самоорганизации моделей достижений общей теории связи и теории распознавания образов [60, 67, 130].

7. ОСНОВНЫЕ НАПРАВЛЕНИЯ БУДУЩЕЙ РАБОТЫ

Задачи прогноза самых сложных процессов в экологии, метеорологии, экономике и других областях требуют повышения помехоустойчивости алгоритмов МГУА приблизительно в 10—100 раз. Это может быть достигнуто при помощи перенесения в теорию самоорганизации результатов общей теории связи (включая теории кодирования и распознавания образов), где имеется большой опыт повышения помехоустойчивости [97]. Располагая весьма помехоустойчивыми алгоритмами МГУА, можно будет решать задачи долгосрочного прогноза даже при весьма неполном информационном базисе (что иногда уже применяется [49]).

Для повышения помехоустойчивости при помощи корреляционного принципа радиоприема разрабатываются методы обратного перехода от вероятностных характеристик к исходной реализации случайного процесса [80, 81]. Обратный переход обещает много важных применений, таких, как точное предсказание землетрясений, прогнозирование местонахождения «пятен» загрязнения водной поверхности озера или моря и решение других задач. Какова участь Байкала и других озер и морей? Повышается ли уровень мирового океана? Изменяется ли климат Земли? Каков прогноз урожая сельскохозяйственных культур? Куда приведет процесс инфляции? На подобные вопросы объективно можно ответить, воспользовавшись алгоритмами МГУА для самоорганизации моделей по натурным данным. Моделируемые процессы соответствуют определенным расположениям точек в гильбертовом пространстве, широко используемом в теории связи. Теория проектирования векторов (включающая МНК) также работает в пространстве такого рода, размерность которого увеличивается на единицу с приходом каждой новой точки измерений. Необходимость сравнения обеих теорий очевидна.

Идеальный приемник по теории Котельникова — Шеннона рассчитывает расстояние между входным сигналом и опорными точ-

ками эталонов в гильбертовом пространстве и находит его минимум. Теория распознавания предлагает более развитые способы нахождения оптимального решения, которые обеспечивают получение наиболее помехоустойчивого алгоритма самоорганизации для идентификации, краткосрочного и долгосрочного прогнозов. Последние два будут отличаться друг от друга по числу измерений пространства, используемого для самоорганизации модели.

Но сейчас нет нужды ждать, пока особо помехоустойчивые алгоритмы будут разработаны. Если все упомянутые выше меры приняты во внимание, то существующие алгоритмы МГУА с учетом указанных выше мер часто достаточно помехоустойчивы. Поэтому следует расширить фронт работ, с тем чтобы их результаты повлияли на формирование общественного мнения и, таким образом, на ход событий в мире. В. И. Вернадский говорил, что для решения многих проблем, стоящих перед человечеством, прежде всего нужно организовать «общественное понимание» вокруг данных проблем [15]. Например, системный анализ физической модели и последующая самоорганизация долгосрочного прогноза по нефизическим моделям были применены для выяснения механизма процессов инфляции в Англии [70]. Было показано, что введение так называемых компараторов (известных в теории автоматического управления) может остановить ее. Никто, конечно, не рассчитывал на то, что такие компараторы будут реализованы немедленно, но точный объективный результат, несомненно, повлияет на общественное понимание необходимости структурных изменений экономической системы, что, возможно, приведет к важным переменам в сознании людей.

Современные задачи моделирования сложных объектов в области метеорологии, экологии, экономики и производства требуют разработки особых методов помехоустойчивого моделирования по сравнительно коротким рядам наблюдений.

МГУА, реализующий принцип самоорганизации моделей, возможен и развивается как метод помехоустойчивого моделирования. Объективный системный анализ (выбор множества выходных переменных) и выделение подмножества хорошо прогнозируемых, ведущих переменных (выбираемых по системному критерию минимума смещения) при многокритериальном выборе для них моделей оптимальной сложности обеспечивают значительную помехоустойчивость алгоритмов МГУА.

Подобно тому как в системах связи (в соответствии с основной теоремой Шеннона для канала с шумом) при увеличении помех следует уменьшить полосу пропускания приемника, так и в результате самоорганизации по критерию регулярности получаются нефизические модели, более простые, чем физические. Физическая модель (используемая при имитационном моделировании) является моделью оптимальной сложности только при отсутствии помех, т. е. в нереальных условиях.

Дальнейшее перенесение идей теории информации в моделирование позволит еще более поднять помехоустойчивость МГУА и,

таким образом, решить возникшие в настоящее время задачи моделирования сложных объектов в указанных выше областях с целью идентификации характеристик, долгосрочного прогнозирования и многокритериального управления с оптимизацией прогноза.

Из указанных важных и перспективных направлений работы в области экспериментального моделирования в настоящей монографии рассматриваются вопросы исследования помехоустойчивости алгоритмов и критериев МГУА. При этом используются три основных инструмента исследований: применение аналогии между схемами обработки информации в системах связи и алгоритмах самоорганизации; аналитическое изучение влияния помех в экспериментальных данных на процесс самоорганизации модели оптимальной сложности; специальные вычислительные эксперименты на ЭВМ.

В связи с указанной теоретической направленностью данной книги в ней отсутствуют примеры решения практических задач моделирования по экспериментальным данным. Однако этот недостаток легко восполнить, обратившись к монографии [49], где рассмотрено много реальных задач, в основном из области экологии. Весьма содержательны в этом плане также новые книги, вышедшие в ГДР [198] и в СССР [79], где приведены разнообразные экономические и экологические задачи, решенные с применением алгоритмов МГУА, в том числе новых.

Из последних публикаций по тематике МГУА следует еще отметить специальный сборник [209], в который включены оригинальные работы теоретического и прикладного содержания ряда известных ученых-статистиков из разных стран.

Глава 2

САМООРГАНИЗАЦИЯ МОДЕЛЕЙ НА ЭВМ В ТЕРМИНАХ ОБЩЕЙ ТЕОРИИ СВЯЗИ (ТЕОРИИ ИНФОРМАЦИИ)

Значение теории информации не исчерпывается открытием принципиально новых возможностей для техники связи; она оказывает благотворное влияние и на другие отрасли науки и техники, в частности на теорию самоорганизации моделей на ЭВМ.

Покажем возможность и целесообразность взаимного обмена основных идей теории информации и теории самоорганизации. Будем полагать, что речь идет о простой амплитудной модуляции (АМ), применяемой в телефонии, и о простейших, например, алгебраических линейных полиномиальных моделях вида $q = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$. Такие модели соответствуют нелинейным при использовании переобозначения переменных. Например, модель вида $q = b_0 + b_1t + b_2t^2$ можно записать в виде $q = a_0 + a_1x_1 + a_2x_2$, если обозначить $x_1 = t$, $x_2 = t^2$.

В теории связи, как и при самоорганизации, в той или иной степени применяется перебор гипотез — например перебор эталонных сигналов. Для упрощения проведения аналогий считаем, что при моделировании используется простейший перебор — прямое наращивание сложности полинома с помощью последовательного добавления членов, например a_0 , $a_0 + a_1t$, $a_0 + a_1t + a_2t^2$ и т. д. Более подробно способы перебора, применяемые при самоорганизации моделей методом полной индукции, описаны в гл. 3.

При рассмотрении структурных аналогий теорий самоорганизации и связи условимся различать систему:

- 1) связи, рассматриваемую в общей теории связи;
- 2) вычислительного эксперимента, в которой исследуются методы и критерии самоорганизации при известных характеристиках шумов и заданной истинной модели (заданном точном сигнале);
- 3) самоорганизации, решающую задачу построения модели по зашумленным данным таблицы наблюдений (по зашумленному сигналу).

В данной главе в системе вычислительного эксперимента и моделирования анализируется только случай решения задачи идентификации модели. Самоорганизация прогнозирующих моделей не рассматривается, хотя в теории связи существует метод пред-

сказаний, используемый для уменьшения избыточности сигнала. Это не помешает проведению достаточно содержательных аналогий между системами связи и самоорганизации моделей.

Другое ограничение данного рассмотрения состоит в том, что в качестве примера используется только одномерная модель $q(x)$. Это ограничение можно снять и распространить результаты на двухмерные $q(x_1, x_2)$ и многомерные модели $q(x_1, x_2, \dots, x_n)$. В теории информации двухмерные и трехмерные сигналы (модели) рассматриваются в телевидении.

1. АНАЛОГИЯ СТРУКТУРНЫХ СХЕМ СИСТЕМЫ СВЯЗИ, ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА И СИСТЕМЫ САМООРГАНИЗАЦИИ МОДЕЛИ

Опишем функции элементов сравниваемых между собой структурных схем (рис. 5).

Источник информации выбирает частный вид сообщения из совокупности возможных сообщений. В схеме вычислительного эксперимента источником информации является полином (например, $q = a_0 + a_1 x = 10 - 0,1 t^2$), в системе самоорганизации модели — сам исследуемый объект (например, экосистема), «передающий» весьма продолжительное время только одно «сообщение».

Передатчик отображает пространство сообщений в пространство незашумленных его сигналов: $q_1(t) \rightarrow q_1$. Например, полином $q_1(t)$ представляется в виде таблицы данных со столбцами t и q_1 . В системе моделирования таблицы истинных, фактических данных скрыты в самом объекте.

Канал связи — звено системы, где вносится помеха; на выходе его получается копия сигнала — таблица зашумленных данных со столбцами t и $q_2 = q_1 + \xi$.

Приемник отображает пространство принимаемых сигналов в пространство принимаемых сообщений: $q_2 \rightarrow q_2(t)$. Таблица со столбцами t и q_2 преобразуется в полином $q_2(t)$, называемый физической моделью. Приемник соответствует алгоритму самоорганизации моделей на ЭВМ.

Адресат или заказчик — лицо, для которого предназначено принимаемое сообщение (модель).

Теория информации изучает сигнал на выходе канала связи, теория самоорганизации — таблицу опытных данных, полученную на выходе объекта исследования. Легко установить, что наиболее важная выходная часть, начиная от канала связи и до адресата или заказчика, у всех трех систем совершенно идентична (см. рис. 5).

Теории информации и самоорганизации уделяют основное внимание квазистатической части процесса, называемой сигналом или трендом. Динамическая составляющая — это помеха или шум, наложенные на полезный (истинный) входной сигнал. Обе теории исходят из предположения, что в обрабатываемых данных содержится информация об истинном входном сигнале или что в них

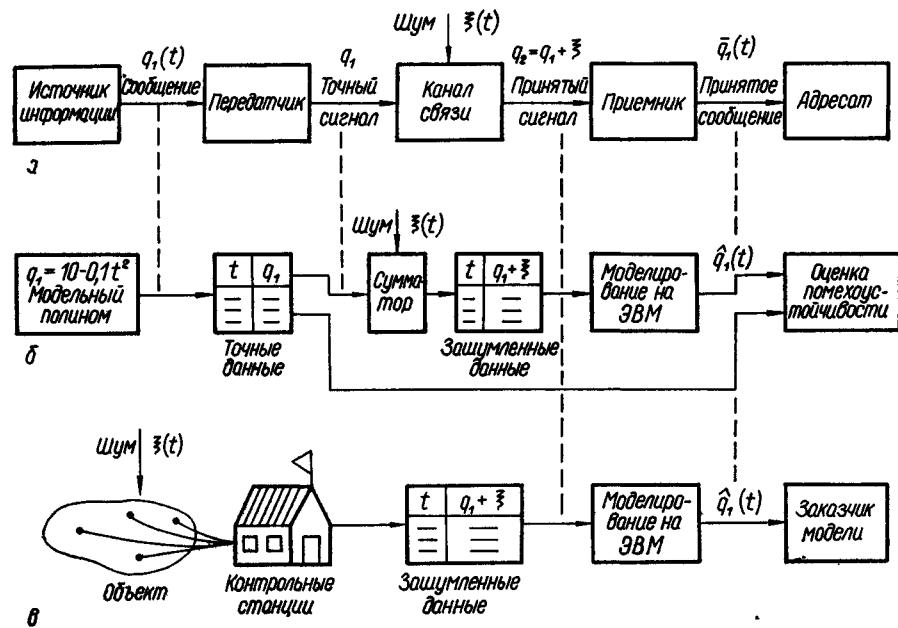


Рис. 5. Структурные схемы систем:
а — связи; б — вычислительного эксперимента; в — самоорганизации модели.

скрыта информация об истинной закономерности, свойственной объекту наблюдения. Задача приемного устройства (алгоритма моделирования) состоит в том, чтобы восстановить, по возможности правильно, первоначальный неискаженный сигнал (закономерность).

Теория информации полагает, что сигнал, поступающий на вход линии связи, урезан по частоте и, кроме того, на него наложена аддитивная помеха — шум (шум может быть в тысячу раз больше регулярного сигнала). В теории самоорганизации учитывают, что в исходной таблице опытных данных представлена только малая выборка, а на выходную величину также наложена аддитивная помеха.

Понятия «истинный входной сигнал» $q_1(t)$ и «помеха» $\xi(t)$ определяются прагматически: в сигнал входит часть выходного напряжения, способствующая передаче сообщения. Все остальное — помеха. Например, пусть на выходе радиоприемника слышна смесь передач на французском и русском языках. Тогда для француза полезным сигналом будет часть напряжения, воспроизводящая французскую речь, для русского — русскую и т. п. Таким образом, что называть сигналом, а что помехой, зависит от адресата или заказчика. При моделировании — полезная часть данных определяется как часть, способствующая идентификации или прогнозу в зависимости от задачи моделирования. Все остальное — помеха или шум. Помеха — это слагаемое, мешающее вы-

полнению задачи моделирования, снижающее глубину минимума критерия селекции моделей.

В теории информации чаще всего (например, при доказательстве предельной теоремы Шеннона) шум подразумевается независимым от сигнала, аддитивным и распределенным по нормальному закону (т. е. гауссовским). Поэтому, если при самоорганизации модели считать помехи независимыми, то теория информации применима непосредственно, а если помехи зависят от сигнала, то она применима только для ортогонализированных алгоритмов МГУА [169].

2. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ИНФОРМАЦИИ И ИХ АНАЛОГИ В ТЕОРИИ САМООРГАНИЗАЦИИ МОДЕЛЕЙ

В теории информации сигнал на входе канала связи характеризуется следующими величинами [97, 172]: амплитудой $q_1(t)$, мощностью $P_1(t)$, полосой частот сигнала W_1 , предельной частотой передатчика W_{\max} , уровнем отношения сигнал/помеха $H_1 = \log_2 \frac{P_1}{M}$, объемом сигнала $V_1 = H_1 W_1 T_1$, длительностью сигнала T_1 .

Сигнал на выходе канала связи определяется амплитудой сигнала $q_2(t) = q_1(t) + \xi(t)$, мощностью $P_2 = P_1 + M$, полосой частот W_2 , уровнем отношения сигнал/помеха $H_2 = \log_2 \frac{P_2}{M}$, объемом канала $V_2 = H_2 W_2 T_2$.

Длительность сигнала T аналогична продолжительности наблюдения объекта моделирования, указанного в таблице опытных данных. Максимально допустимый шаг таблицы данных определяется по теореме отсчетов Котельникова: деления таблицы данных должны быть не шире $\Delta = \frac{1}{2W_1}$, где W_1 — полоса частот. Следовательно, время передачи сигнала (минимальная длина таблицы данных) равно:

$$\text{при отсутствии помех} — T_1 = \frac{A_1}{2W_1} \text{ с;}$$

$$\text{при наличии помех} — T_2 = \frac{A_2}{2W_1} \text{ с, причем, } T_2 \leq T_1,$$

где A_1, A_2 — алгебраические минимумы точек, необходимые для самоорганизации моделей.

Для полиномиальных моделей число точек равно числу слагаемых частных полиномов, для гармонических — утроенному числу гармоник модели $A_1 = 3F$ и т. п.

Пропускная способность C системы связи по Хартли есть двоичный логарифм числа сообщений, которые можно передать за единицу времени при заданной точности.

Оптимальная (предельно достижимая) система связи обладает пропускной способностью, вычисляемой по формуле

$$C = W_2 \log_2 \frac{P + M}{M} \text{ бит/с,}$$

где C — пропускная способность (скорость передачи); W_2 — полоса частот канала; $P = q_1^2$ — мощность сигнала; $M = \xi^2$ — мощность помехи.

За время T по системе связи можно передать $J = CT$ бит информации. Формула показывает, что при равной информации, т. е. при $J = \text{const}$, мощность сигнала P_1 можно «обменять» на ширину полосы W или продолжительность передачи T и др.

При самоорганизации модели решается более узкая задача: если ограничиться стационарными моделями с постоянными коэффициентами, то требуется передать всего одно сообщение, т. е. построить одну модель. Оптимальная система самоорганизации линейной модели при отсутствии помех потребует для этого число измерений, равное алгебраическому минимуму точек A_1 .

Аналогом пропускной способности системы можно считать величину, обратную минимуму критерия селекции модели: $C = k/\Delta(B)_{\min}$. С увеличением помех глубина минимума критерия уменьшается, т. е. «пропускная способность» падает (рис. 6).

Помехоустойчивость системы связи определяется минимальным предельно допустимым значением отношения сигнал/помеха H_{\max} , при котором еще можно принять сигнал, т. е. восстановить его, отфильтровав помеху.

При самоорганизации моделей используются два предела:

1) в вычислительном эксперименте помеха увеличивается до тех пор, пока критерий точности RR не станет равным единице:

$$RR = \frac{\sum_{i=1}^{A_{\max}} (q_i - \bar{q}_i)^2}{\sum_{i=1}^{A_{\max}} (q_i - \bar{q}_i)^2} \leq 1,0;$$

2) пока не будет потеряна правильная структура модели, т. е. пока не изменится структура полинома.

Эффективность системы связи \mathcal{E} прямо пропорциональна пропускной способности C и помехоустойчивости H_{\max} и обратно пропорциональна времени сигнала T . Эффективность алгоритма само-

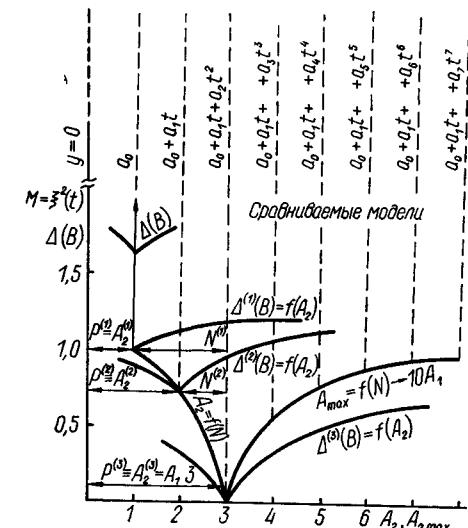


Рис. 6. Изменение оптимальной сложности модели A_2 и длины таблицы данных A_{\max} при увеличении мощности помех M (при $A_1=3$).

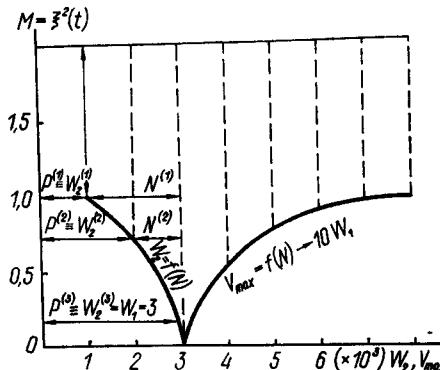


Рис. 7. Изменение полосы канала W_2 и полосы передатчика W_{\max} при увеличении мощности помех M (при $W=3 \cdot 10^3$).

зует возможность передачи сообщений по каналам связи с узкой полосой при малой затрате энергии, а эффективность моделирования — возможность построения достаточно точной модели по малому числу точек при малой затрате времени на измерение, сбор и обработку данных.

Указанные оценки из теории информации непосредственно применимы для одномерных задач моделирования (например, для самоорганизации модели стока реки как функции времени). Двухмерные модели потребуют введения двух полос частот (W_1' , W_1''), подобно тому как это делается в теории телевизионных сигналов.

Закон сохранения избыточности. Свойства системы связи и самоорганизации определяются значением избыточности. Различают свойства широкополосных и узкополосных систем.

В широкополосных системах связи избыточность больше нуля, объем канала превышает объем сигнала: $V_{\max} > V_1$ или $W_{\max} > W_1$.

В широкополосных системах самоорганизации моделей на перебор по ансамблю критериев подаются модели-претенденты от самых простых до таких, сложность которых значительно превосходит сложность истинной или физической модели. Аналогами широкополосных линий связи в моделировании могут быть алгебраические модели. Для них редуцирование таблицы данных (увеличение числа точек с последующим сглаживанием) бесполезно.

В узкополосных системах связи объем канала меньше объема сигнала, избыточность отсутствует: $V_{\max} < V_1$ или $W_{\max} < W_1$. Оптимальные соотношения предельной теоремы Шеннона (см. рис. 6, 7) нарушаются. При этом оказывается полезным редуцирование сигнала. Например, если линия связи пропускает только частоты до $W_{\max}=3000$ Гц, а нужно передать запись на магнитной ленте, содержащую частоты до $W_1=12000$ Гц, то достаточно снизить скорость передачи сигнала $1/T_1$ в четыре раза, с тем чтобы при воспроизведении снова увеличить скорость $1/T_2$ также в четыре

раза. Таким способом можно почти без искажений передать высококачественную запись по узкополосной линии [100].

В узкополосных системах самоорганизации перебираются модели, сложность которых меньше или не превосходит сложность истинной модели: $A_{\max} \leq A_1$. Примером могут быть конечно-разностные модели, сложность которых ниже оптимальной. Такие модели являются аналогом дифференциальных уравнений только при малых шагах таблицы данных. Здесь последовательность операций «снятие малого числа точек на объекте — редуцирование таблицы на большее число точек — самоорганизация модели оптимальной сложности — детальная идентификация (сезонные данные) — сглаженная идентификация (годовые данные)» может быть весьма эффективной. Без редуцирования оптимальная сложность модели не достигается из-за недостаточного числа точек исходных данных.

Так, в работе по идентификации поля загрязнения Невской Губы [72] исходная таблица данных представляла собой матрицу размерностью 3×8 точек. При помощи сплайнов она была редуцирована до размера 6×16 точек, что дало возможность получить модели оптимальной сложности.

Сложность модели — аналог полосы пропускания канала связи. В теории самоорганизации моделей часто используется термин «сложность модели». Для линейных полиномиальных моделей сложность определяется числом слагаемых правой части или, что одно и то же, алгебраическим минимумом точек.

При самоорганизации сложность модели (а следовательно, и алгебраический минимум точек) постепенно увеличивается до тех пор, пока будет найден минимум заданного внешнего критерия. Сложность моделей, получаемых по алгоритму, изменяется в пределах от нуля до A_{\max} и проходит искомое значение A_2 . В связи с этим при самоорганизации моделей удобно рассматривать следующие величины: A_1 — алгебраический минимум точек, необходимый для самоорганизации истинной (физической) модели $A_1 = -2W_1 T_1$; A_2 — алгебраический минимум точек, необходимый для модели оптимальной сложности, получаемой по алгоритмам самоорганизации, $A_2 = 2W_2 T_2$; A_{\max} — алгебраический минимум точек, необходимый для самой сложной модели, которую можно получить в результате самоорганизации, $A_{\max} = 2W_{\max} T_{\max}$, или число точек, реально представленных в таблице данных.

При самоорганизации моделей действуют следующие законы (см. рис. 6):

1) при отсутствии помех, начиная с некоторой сложности, равной сложности истинной модели A_1 , дальнейшее ее повышение для решения задачи моделирования уже не требуется: при $M = \xi^2(t) = 0$ $A_2 = A_1$, $A_{\max} \geq A_1$;

2) при наличии помех оптимальная сложность модели наступает раньше; алгебраический минимум точек (сложность оптимальной модели) уменьшается при $M = \xi^2(t) > 0$ $A_2 < A_1$, $A_{\max} \geq A_2$.

В теории информации известны аналогичные законы (см. рис. 7):

1) для точной передачи сообщения необходимо, чтобы объем канала был равен или превышал объем сигнала: при $M = \xi^2(t) = 0$ $V_2 = V_1$, $V_{\max} \geq V_1$;

2) при наличии помех оптимальный объем канала несколько меньше объема сигнала: при $M = \xi^2(t) > 0$ $V_2 < V_1$, $V_{\max} \geq V_2$.

Например, полоса пропускания предназначенных для работы в условиях помех специальных приемников уже, чем у широкополосных приемников для художественного воспроизведения музыки при малых помехах. Таким образом, полоса канала связи оказывается аналогом сложности модели, оцениваемой по алгебраическому минимуму точек

$$V_1 = H_1 W_1 T_1 = H_1 \frac{A_1}{2}, \quad V_2 = H_2 W_2 T_2 = H_2 \frac{A_2}{2}. \quad (2.4)$$

Влияние помех на точность модели можно в некоторой степени преодолеть увеличением длины таблицы данных. Однако при чрезмерной длине таблицы точность и помехоустойчивость модели снижаются. Существует оптимальная длина таблицы данных как для стационарных, так и для нестационарных сигналов. С учетом необходимости уменьшения влияния помехи длина таблицы выбирается приблизительно в 10 раз больше алгебраического минимума точек $T_\xi = 10 T_1$ (см. рис. 6). За это время система сможет собрать $J = CT_\xi = T_\xi W \log_2 \frac{P+M}{M}$ бит информации. При этом также можно рассматривать вопрос об «обмене» одного показателя на другой.

Аналогия изменения оптимальной сложности моделей для алгоритма самоорганизации и полосы пропускания для системы связи показана на рис. 6 и 7, где A_1 — сложность физической модели; A_2 — сложность нефизической модели оптимальной сложности; A_{\max} — оптимальный диапазон сложности моделей-претендентов; W_1 — полоса истинного сигнала; W_2 — оптимальная полоса приемника; W_{\max} — оптимальный объем сигнала передатчика.

Закон компромисса. Важным результатом исследований, выполненных в теории информации, является установление связи между пропускной способностью и помехоустойчивостью. Повышение помехоустойчивости влечет за собой уменьшение пропускной способности. При этом варьируются параметры системы связи, например полоса частот W_2 .

При самоорганизации моделей по критерию регулярности действует аналогичный закон: увеличение мощности (амплитуды) помех приводит к выбору более простых помехоустойчивых моделей, для которых алгебраический минимум точек меньше, чем для модели объекта, полученной в условиях отсутствия шума. При этом варьируются параметры модели, например алгебраический минимум точек (сложность модели) $A_2 = \text{var}$.

3. ВТОРАЯ (ПРЕДЕЛЬНАЯ) ТЕОРЕМА ШЕННОНА

Сформулируем вторую теорему Шеннона в следующем виде [132]. Пусть P — средняя мощность сигнала, а помехой является независимый белый шум мощностью M в полосе частот W . Оптимальная (предельно достижимая) скорость передачи определяется соотношением

$$C_{\max} = W_2 \log_2 \frac{P+M}{M} = W_2 \log_2 \frac{W_1}{W_1 - W_2}. \quad (2.5)$$

Чем больше мощность сигнала P по сравнению с мощностью помех M , тем больше и предельно достижимая пропускная способность. Теорема, таким образом, устанавливает предел пропускной способности системы связи, достигаемый при оптимальном выборе способа кодирования и полосы канала W_2 (полоса сигнала W_1 считается заданной) (см. рис. 7).

Для самоорганизации моделей теорема позволяет выбрать оптимальную сложность модели A_2 (сложность объекта моделирования A_1 задана). Чем больше шум, тем меньше глубина минимума критерия селекции и тем проще модель (см. рис. 6). Теорема указывает оптимальные (предельно достижимые) значения полосы сигнала (и сложности моделей) и таким образом объясняет, почему при наличии помех необходимо перейти к использованию нефизических моделей (физические модели соответствуют случаю отсутствия помех).

Геометрическое построение Шеннона [172]. Оно является геометрической интерпретацией изложенной выше теоремы о пределе

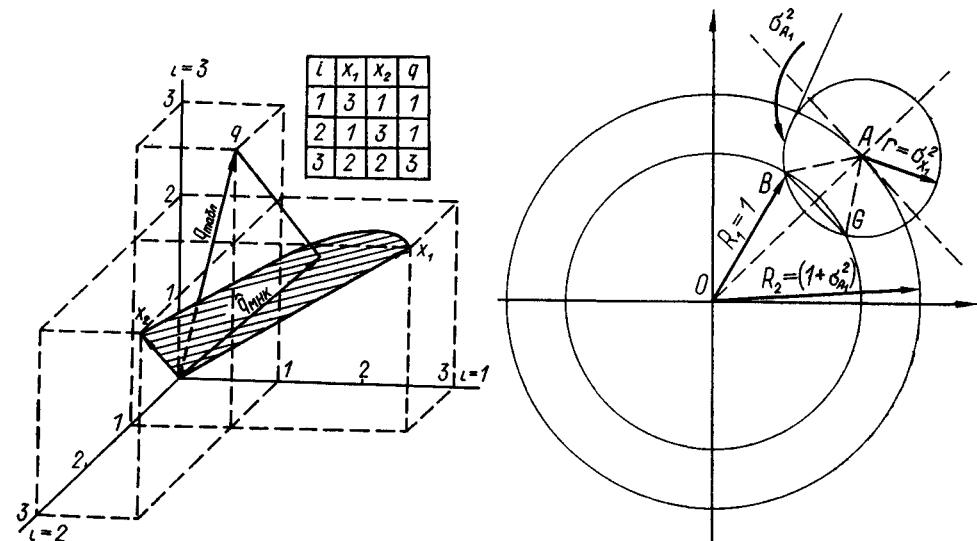


Рис. 8 Геометрическая интерпретация метода наименьших квадратов (МНК) при трех измерениях.

Рис. 9. Геометрическая интерпретация предельной теоремы Шеннона.

пропускной способности системы связи. Входной сигнал (как и таблица исходных данных при моделировании) все время пополняется новыми точками с шагом $\Delta = 1/2 W_1$. Так же как в теории проекторов (рис. 8), являющейся развитием МНК, сигнал — функция времени $q(t)$ — может быть представлен точкой в гиперпространстве с числом измерений $2WT$. С приходом каждой новой точки число измерений гиперпространства увеличивается на единицу. Однако среднее значение сигнала (принимаемое равным единице) стабильно, и это изображается внутренней гиперсферой с центром O (рис. 9). Помеха принимается равной дисперсии отклонений сигнала от среднего значения: она изображается гиперсферой с центром в точке A .

Число моделей при отсутствии помех бесконечное. Все они соответствуют точке $(3,0)$ на рис. 6 и 7 (и потому часто объединяются в одну «физическую модель»), а в гиперпространстве Шеннона лежат на гиперсфере единичного радиуса $R_1 = OB$ (см. рис. 9).

При наличии помех число моделей, называемых нефизическими, конечно. Эти модели соответствуют точкам $(1; 1)$ и $(2; 0,75)$ на рис. 6 и 7 или лежат на гиперсфере радиуса $R_2 = OA$ (см. рис. 9), что удовлетворяет соотношениям предельной теоремы Шеннона:

$$A_1 = (P + M); \quad A_2 = P; \quad A_1 - A_2 = M \quad \text{для рис. 6};$$

$$W_1 = (P + M); \quad W_2 = P; \quad W_1 - W_2 = M \quad \text{для рис. 7},$$

откуда

$$\frac{A_2}{A_1} = \frac{P}{P + M}, \quad \text{или} \quad \frac{W_2}{W_1} = \frac{P}{P + M}. \quad (2.6)$$

Если мощность помехи M задана, а физическая модель A_1 известна, то теорема Шеннона позволяет найти модель оптимальной сложности дедуктивным способом, т. е. без перебора моделей-претендентов.

Теорема относится только к самоорганизации модели по критериям, характеризующим ее точность (регулярности, точности прогноза и др.).

4. ПРИМЕНЕНИЕ ГЕОМЕТРИЧЕСКОГО ПОСТРОЕНИЯ ШЕННОНА ПРИ САМООРГАНИЗАЦИИ МОДЕЛЕЙ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ, ДОЛГОСРОЧНОГО ПРОГНОЗА И ВЕКТОРНОЙ ОПТИМИЗАЦИИ

Основная цель построения Шеннона — вычисление площади или длины хорды лунки BG (см. рис. 9). Площадь лунки находится в прямой зависимости от длины хорды BG , равной

$$h = \sqrt{2WT \frac{PM}{P+M}}, \quad (2.7)$$

где T — продолжительность сигнала; W — полоса частот; P — мощность сигнала; M — мощность помехи.

В работе [168] показано, что апостериорная вероятность принадлежности зашумленного сигнала к «абсолютному эталону класса» зависит от площади лунки BG (пропорциональной хорде BG) следующим образом:

$$p(\bar{x} | l) = \left(1 - \left(\frac{V_{BG}}{V_C}\right)^{MT}\right) \approx 1 - \left(\frac{h}{V_C}\right)^{MT}. \quad (2.8)$$

Следовательно, чем меньше хорда BG (или свертка хорд нескольких построений), тем вероятнее отнесение данного сигнала к данному абсолютному эталону. Трудности нахождения вероятности другими способами известны. Практически многомерные задачи при зависимых признаках не решаются.

В соответствии с этим хорда BG является, таким образом, одним из видов меры близости идеального и реального (входного) сигналов.

Построение Шеннона позволяет найти удобный для самоорганизации моделей критерий устойчивости свертки хорд BG , по которому можно решать задачи распознавания образов, долгосрочного прогноза и векторной оптимизации. Для нескольких переменных потребуется свертка хорд.

Распознавание образов [168]. Для конкретности покажем алгоритм на примере распознавания трех образов A , B и C по трем признакам x_1 , x_2 , x_3 . Исходные таблицы данных имеют следующий вид:

$$\text{образ } A\{x_1, x_2, x_3\}, \quad \text{образ } B\{x_1, x_2, x_3\}, \quad \text{образ } C\{x_1, x_2, x_3\}.$$

В них представлены значения признаков. Строки этих таблиц (числом n) называются изображениями. Кроме того, задана таблица данных неизвестного объекта: $X\{x_1, x_2, x_3\}$. Задача состоит в том, чтобы определить, к какому классу относится этот неизвестный объект.

Чтобы приблизиться к понятиям теории информации, возведем значения признаков в квадрат, т. е. получим таблицы мощности сигналов

$$A\{P_1, P_2, P_3\}, \quad B\{P_1, P_2, P_3\}, \quad C\{P_1, P_2, P_3\}, \quad X\{P_1, P_2, P_3\}.$$

Нормируя мощности по среднему значению $\Delta P_i = \frac{P_i - \bar{P}_i}{\bar{P}_i}$, $i = 1, 2, 3$, получаем таблицы отклонений от среднего

$$A\{1 + \Delta P_1; 1 + \Delta P_2; 1 + \Delta P_3\}, \quad B\{1 + \Delta P_1; 1 + \Delta P_2; 1 + \Delta P_3\},$$

$$C\{1 + \Delta P_1; 1 + \Delta P_2; 1 + \Delta P_3\}, \quad X\{1 + \Delta P_1; 1 + \Delta P_2; 1 + \Delta P_3\}.$$

Усредняя отклонения по столбцам таблиц, находим дисперсии

$$\sigma_{A_1}^2 = \frac{1}{n} \sum_{i=1}^n \Delta P_{A_1 i}^2, \dots, \sigma_{A_3}^2 = \frac{1}{n} \sum_{i=1}^n \Delta P_{A_3 i}^2,$$

$$\sigma_{B_1}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{B_1 t}^2, \dots, \sigma_{B_s}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{B_s t}^2,$$

$$\sigma_{C_1}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{C_1 t}^2, \dots, \sigma_{C_s}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{C_s t}^2,$$

$$\sigma_{X_1}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{X_1 t}^2, \dots, \sigma_{X_s}^2 = \frac{1}{n} \sum_{t=1}^n \Delta P_{X_s t}^2.$$

Далее следует получить девять построений Шеннона для образов A , B и C и неизвестного объекта X и для каждого признака соответственно. Пример показан на рис. 9 для первого признака x_1 первого образа A . Внутренняя окружность единичного радиуса с центром O отвечает среднему значению всех сигналов $R_1=1$. Внешняя окружность соответствует зашумленному сигналу, т. е. имеет радиус $R_2 = 1 + \sigma_{A_1}^2$. Малая окружность с центром в точке A отвечает дисперсии отклонений неизвестного образа X : $r = \sigma_{X_1}^2$. Целью построения является определение длины хорды

$$h_{A_1} = BG \begin{cases} h_{A_1} > 0, \sigma_{X_1} > \sigma_{A_1}, \\ h_{A_1} = 0, \sigma_{X_1} \leqslant \sigma_{A_1}, \end{cases}$$

находящейся в обратной зависимости от апостериорной вероятности принадлежности неизвестного объекта X к образу A по признаку x_1 . Аналогично можно получить значения длин хорд.

Практически вместо геометрических построений можно воспользоваться формулами для определения длины хорды, полученными приравниванием вычисленных двумя способами площадей треугольника AOB :

$$h_{A_1} = \frac{2\sigma_{A_1}^2}{1 + \sigma_{X_1}^2}, \dots, h_{A_s} = \frac{2\sigma_{A_s}^2}{1 + \sigma_{X_s}^2},$$

$$h_{B_1} = \frac{2\sigma_{B_1}^2}{1 + \sigma_{X_1}^2}, \dots, h_{B_s} = \frac{2\sigma_{B_s}^2}{1 + \sigma_{X_s}^2},$$

$$h_{C_1} = \frac{2\sigma_{C_1}^2}{1 + \sigma_{X_1}^2}, \dots, h_{C_s} = \frac{2\sigma_{C_s}^2}{1 + \sigma_{X_s}^2}.$$

Длины хорд рассматриваются как составляющие трехмерных векторов:

$$h_A = \sqrt{\frac{1}{3}(h_{A_1}^2 + h_{A_2}^2 + h_{A_3}^2)}, \quad h_B = \sqrt{\frac{1}{3}(h_{B_1}^2 + h_{B_2}^2 + h_{B_3}^2)}, \quad (2.9)$$

$$h_C = \sqrt{\frac{1}{3}(h_{C_1}^2 + h_{C_2}^2 + h_{C_3}^2)}.$$

Объект X при $h_A < h_B$, $h_A < h_C$ принадлежит классу A ; при $h_B < h_A$, $h_B < h_C$ — классу B ; при $h_C < h_A$ и $h_C < h_B$ — классу C .

Долгосрочный прогноз. В качестве примера рассмотрим выбор одного оптимального прогноза из десяти претендентов. Таблица исходных данных процесса, по которым выполнена самоорганизация десяти прогнозирующих моделей, может быть представлена в следующем виде:

ИД $\{x_1, x_2, \dots, x_{n_1}\}$ (n_1 точек).

Десять сравниваемых прогнозов можно записать в виде десяти подобных таблиц (они включают исходные данные $n_2 > n_1$; интервал прогноза $n_2 - n_1$):

$$\begin{array}{lll} 1\text{-й прогноз} & 2\text{-й прогноз} & \dots 10\text{-й прогноз} \\ 1\{x_1, x_2, \dots, x_{n_1}\} & 2\{x_1, x_2, \dots, x_{n_1}\} & \dots 10\{x_1, x_2, \dots, x_{n_1}\}. \end{array}$$

Возведя значение переменных в квадрат, получим таблицы мощности:

$$\begin{array}{l} \text{ИД } \{P_1, P_2, \dots, P_{n_1}\}, \quad 1\{P_1, P_2, \dots, P_{n_1}\}, \\ 2\{P_1, P_2, \dots, P_{n_1}\}, \dots, 10\{P_1, P_2, \dots, P_{n_1}\}. \end{array}$$

Нормируя мощности по среднему значению $\Delta P_i = \frac{P_i - \bar{P}_i}{\bar{P}_i}$, получаем таблицы, указывающие отклонения от среднего:

$$\begin{array}{l} \text{ИД } \{1 + \Delta P_1, 1 + \Delta P_2, \dots, 1 + \Delta P_{n_1}\}, \\ 1\{1 + \Delta P_1, 1 + \Delta P_2, \dots, 1 + \Delta P_{n_1}\}, \\ \dots \dots \dots \dots \dots \dots \\ 10\{1 + \Delta P_1, 1 + \Delta P_2, \dots, 1 + \Delta P_{n_1}\}. \end{array}$$

Усреднив отклонения по столбцам таблиц, получим дисперсии: для исходных данных

$$\begin{aligned} \sigma_{\text{ИД-1}}^2 &= \frac{1}{n_1} \sum_{t=1}^{n_1} \Delta P_{\text{ИД-1 } t}^2, \quad \sigma_{\text{ИД-2}}^2 = \frac{1}{n_1} \sum_{t=1}^{n_1} \Delta P_{\text{ИД-2 } t}^2, \dots \\ &\dots, \sigma_{\text{ИД-10}}^2 = \frac{1}{n_1} \sum_{t=1}^{n_1} \Delta P_{\text{ИД-10 } t}^2; \end{aligned}$$

для прогнозов

$$\begin{aligned} N_1 &= 1 \quad \sigma_{1-1}^2 = \frac{1}{n_2} \sum_{t=1}^{n_2} \Delta P_{1-1 t}^2, \dots \\ &\dots, \sigma_{1-10}^2 = \frac{1}{n_2} \sum_{t=1}^{n_2} \Delta P_{1-10 t}^2, \end{aligned}$$

$$N_2 = 2 \quad \sigma_{2-1}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \Delta P_{2-1}^2 i, \dots,$$

$$\dots, \sigma_{2-10}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \Delta P_{2-10}^2 i,$$

.....

$$N_{10} = 10 \quad \sigma_{10-1}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \Delta P_{10-1}^2 i, \dots$$

$$\dots, \sigma_{10-10}^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \Delta P_{10-10}^2 i.$$

Графически при помощи построений Шеннона или по аналитическим формулам вида

$$h_{(N,i)} = \frac{2\sigma_{N,i}^2}{1 + \sigma_i^2} \quad (2.10)$$

(где N — номер прогноза, $N=1, 10$; i — номер признака, $i=1, n$) находим длины хорд BG и их свертки (табл. 1). Выбираем тот прогноз, для которого свертка хорд меньше. Например, если $h_{10} < h_N$, $N=1, 9$, то оптимальным по критерию устойчивости длины хорды построений Шеннона является десятый прогноз.

Таблица 1

Номер прогноза	Длина хорды для i -го прогноза			Средняя длина хорды (свертка)
	$i=1$...	$i=10$	
1	h_{1-1}	...	h_{1-10}	$h_1 = \sqrt{\frac{1}{10}(h_{1-1}^2 + \dots + h_{1-10}^2)}$
2	h_{2-1}	...	h_{2-10}	$h_2 = \sqrt{\frac{1}{10}(h_{2-1}^2 + \dots + h_{2-10}^2)}$
...
10	h_{10-1}	...	h_{10-10}	$h_{10} = \sqrt{\frac{1}{10}(h_{10-1}^2 + \dots + h_{10-10}^2)}$

Векторная оптимизация. При решении задач многокритериального выбора экспертам подаются таблицы дискретного ряда вариантов, относящихся к так называемой области эффективных решений Парето, из которых требуется выбрать одно оптимальное решение. ЭВМ в этой области, без дополнительной информации бессильна что-либо выбрать, поскольку критерии противоречат друг другу (один увеличивается, другой уменьшается и т. п.).

В то же время если эксперты равноправны (нет предпочтения критериев), то совет выбирает некоторое среднее решение. Это среднее решение машина может найти, пользуясь критериями устойчивости определителя матрицы множественной корреляции [133] или, что более просто, при помощи построений Шеннона.

Для конкретности рассмотрим задачу двухкритериального выбора: требуется выбрать режим работы гидроэлектростанции, причем увеличение выработки электроэнергии \mathcal{E} уменьшает улов рыбы P . Исходная таблица векторной оптимизации содержит шесть режимов. Исходные данные: $\mathcal{E}\{x_{\mathcal{E}}, x_P\}$. Следуя алгоритму, предложенному в [133], составим еще шесть таблиц (по пять точек каждая), в которых по очереди пропустим один исследуемый режим: 1) пропуск 1-го режима $1\{x_{\mathcal{E}}, x_P\}$; 2) пропуск 2-го режима $2\{x_{\mathcal{E}}, x_P\}$; ... 6) пропуск 6-го режима $6\{x_{\mathcal{E}}, x_P\}$.

Будем считать режим наилучшим, оптимальным, если исключение его меньше, чем исключение других, влияет на свертку хорд построений Шеннона, т. е. на «абсолютный эталон класса» [168].

Переходя к мощностям, получим таблицы:

$$\cdot \quad \mathcal{E}\{P_{\mathcal{E}}, P_P\}, 1\{P_{\mathcal{E}_1}, P_{P_1}\}, \dots, 6\{P_{\mathcal{E}_6}, P_{P_6}\}.$$

Далее, нормируя по среднему значению, получаем таблицы в отклонениях:

$$\mathcal{E}\{1 + \Delta P_{\mathcal{E}}, 1 + \Delta P_P\}, 1\{1 + \Delta P_{\mathcal{E}_1}, 1 + \Delta P_{P_1}\}, \\ 2\{1 + \Delta P_{\mathcal{E}_2}, 1 + \Delta P_{P_2}\}, \dots, 6\{1 + \Delta P_{\mathcal{E}_6}, 1 + \Delta P_{P_6}\}.$$

Усредняя отклонения по столбцам таблицы, находим дисперсии:

$$\sigma_{\mathcal{E}}^2 = \frac{1}{6} \sum_{i=1}^6 \Delta P_{\mathcal{E},i}^2, \quad \sigma_P^2 = \frac{1}{6} \sum_{i=1}^6 \Delta P_{P,i}^2,$$

$$\sigma_{1-1}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta P_{\mathcal{E}_1,i}^2, \quad \sigma_{1-2}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta P_{P_1,i}^2, \dots$$

$$\dots, \sigma_{6-1}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta P_{\mathcal{E}_6,i}^2, \quad \sigma_{6-2}^2 = \frac{1}{5} \sum_{i=1}^5 \Delta P_{P_6,i}^2$$

Далее выполняем двенадцать построений Шеннона или аналитически находим длины хорд BG и их свертки.

В качестве оптимального выбираем режим, для которого средний абсолютный эталон — свертка хорд построения Шеннона — меньше, чем для других режимов. Например, если $h_6 < h_N$, $N=1, 5$, то оптимальным (по критерию устойчивости свертки длины хорд построений Шеннона) является шестой режим, и именно его выберут эксперты в случае отсутствия у них предпочтений.

Подобная задача решается также по критерию устойчивости определителя матрицы множественной корреляции [133]. Оба параллельных друг другу алгоритма на задачах векторной оптимизации еще не сравнивались между собой.

5. ЛИНЕЙНАЯ ОБРАБОТКА СИГНАЛОВ И ПОМЕХОУСТОЙЧИВОСТЬ СВЯЗИ

В теории связи [97] хорошо изучены задачи оценки и повышения помехоустойчивости передачи сообщений для случаев, когда задан способ передачи и известны характеристики сигналов. Самой простой из таких задач является задача *обнаружения* сигнала $c(t)$ на фоне аддитивной помехи $\xi(t)$, причем форма и все параметры сигнала полностью известны. При этом на выходе схемы обработки сигнала приемник должен дать ответ: был ли на входе сигнал с помехой или только помеха? Если приемник успешно решает такую задачу обнаружения, то он также может решать и задачу *различения*: какой из m известных сигналов $c_l(t)$, $l=1, \dots, m$, был передан.

В помехоустойчивых системах связи широко применяется метод *линейной обработки сигналов* (ЛОС) [97]. Общим видом ЛОС является интегрирование:

$$E_{zx} = \int_0^T z(t)x(t) dt, \quad (2.11)$$

где $z(t) = c(t) + \xi(t)$ — принятый (зашумленный) сигнал; $x(t)$ — неизвестный сигнал, который должен обнаружить приемник; T — длительность переданного (известного) сигнала $c(t)$.

Видом $x(t)$ определяются различные частные случаи ЛОС:

- 1) когерентный прием, когда $x(t) = c(t)$;
- 2) интегральный прием, когда $x(t) = 1$;
- 3) корреляционный прием, когда $c(t)$ не известен, и $x(t) = z(t-\tau)$;
- 4) прием на согласованный фильтр с импульсной характеристикой $x(t) = c(t-\tau)$.

В работе [97] показано, что оптимальным когерентным обнаружителем является схема, реализующая условие

$$E_{zc} > \frac{E}{2}, \quad E = \int_0^T c^2(t) dt, \quad (2.12)$$

где E — энергия переданного сигнала $c(t)$. При выполнении неравенства принимается решение о появлении сигнала на входе схемы.

Очевидно, что при решении задачи различения двух сигналов $c_1(t)$ и $c_2(t)$ приемник должен проверять условие, следующее из предыдущего неравенства:

$$E_{zc_1} - E_{zc_2} > \frac{1}{2}(E_1 - E_2),$$

и если $E_1 - E_2 > 0$, приходим к сравнению степени коррелированности принятого сигнала с двумя известными (эталонными):

$$E_{zc_1} > E_{zc_2}.$$

При выполнении неравенства передавался сигнал $c_1(t)$, при невыполнении — сигнал $c_2(t)$. Ясно, что при m эталонных сигналах приемник, который можно назвать и когерентным и корреляционным, принимает решение в пользу наиболее коррелированного сигнала:

$$E_{zc_l} \rightarrow \max_l, \quad l = 1, \dots, m. \quad (2.13)$$

Оказывается, что и прием на согласованный фильтр фактически решает ту же задачу, поскольку решение принимается в момент окончания сигнала T [97]:

$$E_{zc_l} = \int_0^T z(T-\tau)c_l(T-\tau)d\tau \equiv \int_0^T z(t)c_l(t)dt.$$

Отличаются эти способы приема только схемными реализациями, целесообразными в тех или иных случаях.

В [97] показано, что перечисленные способы приема являются частными случаями *идеального приемника Котельникова*, работающего по алгоритму

$$K_l = \int_0^T [z(t) - c_l(t)]^2 dt \rightarrow \min_l. \quad (2.14)$$

Такой приемник обеспечивает максимальную (предельно достижимую) или *потенциальную помехоустойчивость*, так как минимизирует полную вероятность ошибки приема. Все реальные приемники имеют помехоустойчивость ниже потенциальной.

Идеальный приемник Котельникова основан на схеме ЛОС и определяет в гильбертовом пространстве сигналов расстояние между принятым сигналом и всеми сигналами из ансамбля передаваемых, или эталонных. Решение принимается в пользу сигнала, ближайшего к переданному.

6. САМООРГАНИЗАЦИЯ В ТЕРМИНАХ ЛОС

Нетрудно переложить термины ЛОС на язык самоорганизации моделей по МГУА (с учетом дискретности наблюдений): здесь интервал T соответствует длине (числу точек) таблицы наблюдений N , принятый сигнал $z(t)$ — таблице данных зашумленных наблюдений за выходом объекта q_i , $i=1, \dots, N$, сигнал $c(t)$ — таблице данных c_i , получаемых на выходе испытуемой (т. е. выданной генератором и проверяемой по критерию) модели.

Однако следует обратить внимание на одно существенное различие: в схеме ЛОС на обработку подается один из эталонных,

т. е. известных точно, сигналов $c(t)$, а в схеме самоорганизации — «сигнал» от одной из испытываемых моделей c_i , который является оценкой выхода объекта, полученной по зашумленным данным, т. е. $c_i = \hat{q}_i$, $i = \overline{1, N}$.

Таким образом, в идеальном приемнике Котельникова набор эталонных сигналов является независимой информацией по отношению к принятому сигналу, или внешним дополнением (притом идеальным), т. е. критерий (2.14) является внешним, что и обеспечивает его потенциальную помехоустойчивость. Указанное же выше различие в сущности сигналов $c(t)$ и $c_i = \hat{q}_i$ не позволяет применять непосредственно критерий, аналогичный (2.14):

$$\sum_{i=1}^N (q_i - c_i)^2 = \sum_{i=1}^N (q_i - \hat{q}_i)^2 \rightarrow \min,$$

который уже не будет внешним — он просто совпадает, например, с критерием минимума ошибки аппроксимации по МНК, который при увеличении сложности всегда выбирает самую сложную модель.

Итак, в системе самоорганизации аналог критерия Котельникова можно применять только на проверочной последовательности:

$$\Delta^2(B) = \sum_{i \in B} (q_i - \hat{q}_i)^2 \rightarrow \min;$$

при этом он является обычным критерием регулярности.

В то же время в системе вычислительного эксперимента прямой аналог критерия (2.14) возможен, так как известна таблица незашумленных данных q^0 :

$$R^2(N) = \sum_{i=1}^N (q_i^0 - \hat{q}_i)^2 \rightarrow \min, \quad (2.15)$$

где \hat{q}_i — оценки, полученные по зашумленным данным. Этот критерий является уже внешним, но совпадает с критерием Котельникова «с точностью до наоборот»: по идеальному критерию (2.14) один зашумленный сигнал $z(t)$ сравнивается с набором эталонных $c(t)$, а по идеальному критерию (2.15) один эталонный (незашумленный) сигнал (таблица) q_i^0 сравнивается с набором зашумленных оценок (выходов моделей) $\hat{q}_i(l)$, $l = \overline{1, m}$. Выявленная аналогия позволяет решить задачу определения потенциальной помехоустойчивости самоорганизации моделей, о чём будет идти речь в гл. 4.

7. НОВЫЕ КРИТЕРИИ СЕЛЕКЦИИ, ПОСТРОЕННЫЕ НА ОСНОВЕ ЛОС

Проанализировав соответствие между критерием Котельникова (2.14) и критерием когерентного обнаружителя (2.12), предложим по аналогии новые формы критериев самоорганизации.

Представим (2.14) в следующем виде:

$$K = \int_0^T z^2(t) dt - 2 \left(\int_0^T (z(t) c(t) dt - \frac{1}{2} \int_0^T c^2(t) dt \right) = \\ = E_z - 2 \left(E_{zc} - \frac{1}{2} E \right), \quad (2.16)$$

где E — как и ранее, энергия переданного сигнала $c(t)$; E_z — энергия принятого (зашумленного) сигнала $z(t)$.

С учетом равенства $z(t) = c(t) + \xi(t)$ устанавливаем, что величина $(E_{zc} - \frac{1}{2} E)$ в (2.16) равна нулю при отсутствии сигнала $c(t)$ и может быть даже отрицательна при достаточно большом шуме (в случае $E_{zc} < 0$, $|E_{zc}| > \frac{1}{2} E$). Поэтому условием наличия полезного сигнала на входе приемника является $E_{zc} > \frac{1}{2} E$, что соответствует применению критерия (2.12). Если же передается один из m сигналов $c_l(t)$, $l = \overline{1, m}$, то он определяется по условию $K_l \rightarrow \min$, или, что то же, по условию

$$\left(E_{zc_l} - \frac{1}{2} E \right) \rightarrow \max, \quad (2.17)$$

либо просто по максимуму корреляции $E_{zc_l} \rightarrow \max$ (2.13). Таким образом, разложение (2.16) устанавливает эквивалентность идеального приемника Котельникова с когерентным (2.17) и корреляционным (2.12) приемниками.

Как было показано выше, непосредственная аналогия критерия Котельникова в системе самоорганизации приводит к критерию ошибки МНК. Разложим его по аналогии с (2.16):

$$\varepsilon^2 = \sum_{i=1}^N (q_i - \hat{q}_i)^2 = \sum_{i=1}^N q_i^2 - 2 \left(\sum_{i=1}^N q_i \hat{q}_i - \frac{1}{2} \sum_{i=1}^N \hat{q}_i^2 \right), \quad (2.18)$$

откуда видно, что условию $\varepsilon^2 \rightarrow \min$ однозначно соответствует условие

$$e_1 = \sum_{i=1}^N q_i \hat{q}_i - \frac{1}{2} \sum_{i=1}^N \hat{q}_i^2 \rightarrow \max. \quad (2.19)$$

Проверим это, предположив, что $\hat{q} = \hat{X} \hat{a}$, где X — матрица измерений независимых переменных (аргументов). Тогда

$$e_1 = q^T \hat{X} \hat{a} - \frac{1}{2} \hat{a}^T \hat{X}^T \hat{X} \hat{a},$$

и по условию $\partial e_1 / \partial \hat{a}$ получаем оценку

$$\hat{a} = (X^T X)^{-1} X^T q, \quad (2.20)$$

равную в точности оценке МНК, получаемой по $\varepsilon^2 \rightarrow \min$. При этом (2.20) соответствует условию $\varepsilon_1^2 \rightarrow \max$, так как $\frac{\partial \varepsilon_1^2}{\partial a^2} = -X^T X < 0$ (в предположении полного ранга матрицы X матрица $X^T X$ является положительно определенной).

Предложим теперь новую форму внешнего критерия регулярности, эквивалентную $\Delta^2(B)$:

$$\Delta_1(B) = \sum_{i \in B} q_i \hat{q}_i - \frac{1}{2} \sum_{i \in B} \hat{q}_i^2 \rightarrow \max. \quad (2.21)$$

Очевидно, что разложение (2.18) можно представить еще и так:

$$\varepsilon^2 = \sum_{i=1}^N q_i^2 + \left(\sum_{i=1}^N \hat{q}_i^2 - 2 \sum_{i=1}^N q_i \hat{q}_i \right).$$

При этом оценка МНК соответствует также условию

$$\varepsilon_2 = \sum_{i=1}^N \hat{q}_i^2 - 2 \sum_{i=1}^N q_i \hat{q}_i \rightarrow \min. \quad (2.22)$$

Тогда критерию регулярности типа (2.21) можно придать форму, позволяющую искать минимум:

$$\Delta_2(B) = \sum_{i \in B} \hat{q}_i^2 - 2 \sum_{i \in B} q_i \hat{q}_i \rightarrow \min. \quad (2.23)$$

Эквивалентность оценок, получаемых по критериям $\varepsilon^2 \rightarrow \min$, $\varepsilon_1 \rightarrow \max$, $\varepsilon_2 \rightarrow \min$, очевидна. По-видимому, целесообразность применения той или иной формы критерия должна определяться в каждом практическом случае отдельно.

Как известно, любой внутренний критерий, применяемый для определения оценок коэффициентов моделей, можно использовать в качестве внешнего, вычисляя его на независимой (проверочной) части таблицы данных. В то же время есть ряд внешних критериев самоорганизации моделей, которые неприменимы как внутренние.

Можно предложить два новых вида внешних критериев, используя еще одно разложение ошибки ε^2 :

$$\varepsilon^2 = \left(\sum_{i=1}^N q_i^2 - \sum_{i=1}^N \hat{q}_i^2 \right) + 2 \left(\sum_{i=1}^N \hat{q}_i^2 - \sum_{i=1}^N q_i \hat{q}_i \right). \quad (2.24)$$

Оба выражения в круглых скобках имеют максимумы, не соответствующие оценке МНК. Однако, записав их на последовательности B , получим две новые формы критерия регулярности:

$$h_1(B) = \sum_{i \in B} q_i^2 - \sum_{i \in B} \hat{q}_i^2 \rightarrow 0, \quad (2.25)$$

$$h_2(B) = \sum_{i \in B} \hat{q}_i^2 - \sum_{i \in B} q_i \hat{q}_i \rightarrow 0. \quad (2.26)$$

При переборе необходимо выбирать модель именно по условию $h_1 \rightarrow 0$, $h_2 \rightarrow 0$, так как для точной модели $\hat{q} = q$ и $h_1(B) = h_2(B) = 0$. Можно эти критерии и минимизировать, если вычислять модули выражений (2.25), (2.26).

8. КОНСТРУКТИВНОСТЬ АНАЛОГИИ ЗАДАЧ ТЕОРИИ ИНФОРМАЦИИ И САМООРГАНИЗАЦИИ

Обсуждаемая аналогия теорий информации (связи) и моделирования по данным измерений объясняется прежде всего общностью предмета исследования — напряжение на выходе приемного устройства системы связи, как и некоторый показатель состояния любого реального объекта или процесса, в равной степени могут быть названы *сигналами*. Поскольку же принятый сигнал всегда содержит ту или иную помеху (зашумлен), то и цель у этих теорий также оказывается общей — восстановить переданный сигнал или смоделировать выход объекта, т. е. задачи, решаемые приемником сигналов и алгоритмом моделирования, оказываются аналогичными. При этом не следует упускать из виду два существенных обстоятельства. Во-первых, задача моделирования является более общей, поскольку она рассматривает процессы как во времени (чем занимается и теория связи), так и в пространстве (пример — моделирование конфигурации подземного водного бассейна или вообще задача описания многомерной поверхности), а также смешанные процессы (гидродинамического типа). Во-вторых, задача моделирования, как правило, значительно сложнее: в отличие от теории связи в ней только иногда известен класс опорных функций (полиномы, дифференциальные уравнения и т. д.) и практически всегда неизвестно, какой незашумленный «сигнал» может «передавать» изучаемый объект (неизвестны его физические закономерности).

Однако и отмеченного частичного совпадения задач двух теорий достаточно, чтобы попытаться использовать некоторые несомненные достижения теории связи для улучшения методов моделирования, и прежде всего для повышения их помехоустойчивости. Поэтому в последующих главах, с одной стороны, анализируются методы самоорганизации моделей с точки зрения их помехоустойчивости, а с другой — предлагаются некоторые новые алгоритмы с использованием идей теории связи. При этом в основном исследуются алгоритмы самоорганизации, основанные, по аналогии с теорией ЛОС, на прямом переборе моделей различной сложности, «генерируемых», например, в классе полиномов — т. е. комбинаторные алгоритмы МГУА.

Одним из наиболее конструктивных результатов, «индукционных» идеями теории связи, является изложенная в гл. 4 теория потенциальной помехоустойчивости моделирования по комбинатор-

ному алгоритму в условиях аддитивных помех. В ней в рамках теории самоорганизации получены точные оценки предельно достижимой помехоустойчивости. В том числе строго доказано утверждение, обсуждавшееся выше на основе аналогии с теорией Шеннона, что при возрастании уровня шума более эффективными должны становиться более простые модели.

Возможности прямого перенесения некоторых схем помехоустойчивого приема в алгоритмы самоорганизации при учете некоторых дополнительных предположений о помехах обсуждаются в шестой главе. Это предварительная фильтрация частотных помех и использование схемы корреляционного приема. Отметим, что частично фильтрация помех выполняется всегда в алгоритмах моделирования — во-первых, усреднением повторных измерений и, во-вторых, применением методов оценивания при числе данных измерений, превышающем число оцениваемых параметров (например, МНК).

Одним из главных способов повышения помехоустойчивости приема сигналов является их кодирование еще до передачи. Главная задача теории кодирования — приблизиться к предельным показателям второй теоремы Шеннона о пропускной способности канала связи с шумом. По-видимому, прямым аналогом теории кодирования в моделировании следует считать теорию планирования эксперимента: цель кодирования — создать наилучшие условия для правильного приема сигналов; цель планирования эксперимента — создать наилучшие условия для последующего успешного моделирования изучаемого процесса [104].

Данное исследование не содержит результатов, полученных на основе аналогии задач кодирования и планирования экспериментов. Однако к простым способам кодирования относятся, например, квантование сигнала по уровням и амплитудная модуляция. Они, как известно, позволяют бороться с помехами простейшей демодуляцией в виде фильтрации и сглаживания (усреднения) принятого сигнала. Некоторыми аналогами такой демодуляции следует считать двухуровневые алгоритмы самоорганизации прогностирующих моделей, которые основаны на идее поиска наиболее согласованных моделей, построенных по данным измерений изучаемого процесса, усредненным на интервалах различной длины.

Глава 3

ТЕОРИЯ КОМБИНАТОРНОГО АЛГОРИТМА МГУА

1. ЗАДАЧА СТРУКТУРНОГО МОДЕЛИРОВАНИЯ, РЕШАЕМАЯ ПО МГУА

Различные технические объекты управления, а также природные и технологические процессы можно представить в виде некоторого многомерного объекта (статического или динамического), имеющего v входов $z[n \times 1]$ и l выходов $y[l \times 1]$. Многие из них описываются моделями, линейными по коэффициентам, но нелинейными по характеру связи вход-выход или по характеру зависимости выхода от времени. В общем виде такую модель можно представить как систему с линейной входящей матрицей коэффициентов (параметров):

$$y = A^T f(y^*, z, z^*, t),$$

где $f[n \times 1]$ — заданный вектор базисных (опорных) функций, в общем случае нелинейных (полиномиальных, тригонометрических и др.); y^* , z^* — векторы входа-выхода, взятые с запаздываниями (в случае статического объекта $y^*=0$); $A^T[l \times n]$ — матрица коэффициентов модели; t — время.

Матрица A , как правило, для большинства объектов не известна или известна не полностью. Поэтому при изучении конкретного объекта с целью управления или прогнозирования возникает задача идентификации матрицы коэффициентов по данным наблюдений за функционированием объекта. Наблюдения могут быть представлены в виде таблицы, где записаны измеренные значения входных и выходных переменных объекта и каждую из N строк которой будем называть *точкой* измерений. Информацию такой таблицы можно представить в виде матричного уравнения

$$Y = XA, \quad Y[N \times l], \quad X[N \times n], \quad (3.1)$$

которое в регрессионном анализе называется системой условных уравнений. Каждое уравнение в (3.1) имеет вид

$$y_j^T = f_j^T(y^*, z, z^*, t) A \stackrel{\Delta}{=} x_j^T A, \quad j = \overline{1, N},$$

где сложная функция $f(\dots)$ обозначена эквивалентным обобщенным аргументом $x[n \times 1]$. Далее будем рассматривать только мо-

дели линейного вида $y = Ax$, расшифровывая смысл обобщенного входного аргумента x при необходимости.

Моделируемый объект может иметь самую разную структуру связи вход-выход, т. е. различные прямые, обратные и перекрестные связи. Эти связи, а также их отсутствие можно представить с помощью того или иного сочетания нулевых и ненулевых элементов матрицы A . Каждое такое сочетание будем называть *структурой модели*.

При идентификации модели, как известно, могут решаться две существенно различные задачи: *параметрическая* идентификация, когда задана структура матрицы A и требуется получить оценки неизвестных параметров этой матрицы по данным наблюдений; *структурная* идентификация, когда заданы лишь наблюдения и требуется определить структуру и параметры модели (т. е. матрицы A). Методы самоорганизации моделей на ЭВМ дают возможность решать более трудную задачу структурной идентификации. При этом реализуется индуктивный подход: генерируется некоторое множество различных структур, из которых выбирается лучшая (оптимальная) по тому или иному целесообразно выбранному внешнему критерию или ансамблю критериев селекции.

Промежуточная задача вычисления оценок коэффициентов каждой генерируемой структуры (т. е. параметрической идентификации) в алгоритмах самоорганизации обычно решается по МНК. Для уравнения (3.1) решение по МНК имеет вид (в случае матрицы X полного ранга, для чего прежде всего необходимо $N \geq n$)

$$\hat{A} = (X^T X)^{-1} X^T Y.$$

Эту оценку можно переписать более подробно так:

$$[\hat{a}^{(1)} | \hat{a}^{(2)} | \dots | \hat{a}^{(l)}] = (X^T X)^{-1} X^T [\tilde{y}^{(1)} | \tilde{y}^{(2)} | \dots | \tilde{y}^{(l)}], \quad (3.2)$$

где $\hat{a}^{(i)} [n \times 1]$ — вектор коэффициентов, характеризующих связь i -го выхода с вектором входов x ; $\tilde{y}^{(i)} [N \times 1]$ — вектор N измерений i -го выхода.

Из (3.2) видно, что оценивание коэффициентов по МНК можно осуществлять для каждого i -го выхода в отдельности:

$$\hat{a}^{(i)} = (X^T X)^{-1} X^T \tilde{y}^{(i)}.$$

Это позволяет в дальнейшем без ограничения общности рассматривать задачу идентификации только скалярного объекта (с одним выходом, $l=1$), где $A [n \times 1]$ — вектор.

С другой стороны, (3.2) показывает, что для получения оценок любого вектора $a^{(i)}$ достаточно только один раз вычислить обратную матрицу $(X^T X)^{-1}$ или псевдообратную матрицу $(X^T X)^{-1} X^T$. Это дает возможность построить весьма экономные алгоритмы структурной идентификации моделей с многомерным выходом, причем

в различных случаях бывает выгоднее строить или обратную, или псевдообратную матрицу.

Как указывалось в гл. 1, комбинаторный алгоритм МГУА решает задачу самоорганизации с помощью полного перебора моделей с различной структурой, т. е. по методу полной математической индукции. Это означает, что на основе информации о наблюдениях, представленной системой условных уравнений

$$y = Xa, \quad y [N \times 1], \quad a [n \times 1], \quad (3.3)$$

определяется наилучшая (оптимальная) оценка структуры вектора параметров по минимуму заданного внешнего критерия I :

$$a^0 = \arg \min_{i=1, p_n} I(\hat{a}_i(X, y)). \quad (3.4)$$

Полное число возможных структур, т. е. число всех возможных сочетаний нулевых и ненулевых элементов вектора a , получается по формуле

$$p_n = \sum_{j=1}^n C_n^j = 2^n - 1. \quad (3.5)$$

Оно является *показательной* функцией числа аргументов n , что ограничивает применимость полного перебора — он эффективен примерно до $n \leq 20$. Однако в пределах своей эффективности, т. е. при числе аргументов в модели не более 20, комбинаторный алгоритм имеет несомненное преимущество в смысле помехоустойчивости моделирования по сравнению с другими алгоритмами МГУА. Это объясняется применением полного перебора: если при сравнении различных моделей проверяется модель с истинной структурой, то по внешнему критерию будет выбрана она (или наилучшее приближение к ней при большом уровне шума). Поэтому в данной главе подробно рассмотрены структура и характерные особенности комбинаторных алгоритмов.

2. СТРУКТУРА КОМБИНАТОРНОГО АЛГОРИТМА

В структуре каждого из алгоритмов МГУА можно выделить три основных блока [45]: 1) преобразования исходных данных в соответствии с выбранной системой опорных (базисных) функций, в которой ищется модель (базис может быть полиномиальным, тригонометрическим, в виде разностных уравнений и т. д.); 2) генерирования (перебора) полного или неполного множества усложняющихся частных моделей в выбранном базисе; 3) вычисления значений некоторого критерия селекции, имеющего свойства внешнего дополнения, и последовательного отбора частных моделей, лучших по этому критерию. Во всех разработанных и применяемых на практике алгоритмах МГУА используются (или в принципе могут быть использованы) одни и те же критерии селекции и одинаковые системы опорных функций. Различие же алгоритмов самоор-

ганизации определяется структурой «генератора моделей», т. е. схемой вычислений, реализованной во втором блоке. По этому признаку алгоритмы МГУА делятся на многорядные, комбинаторные и смешанные, для которых характерны специфические схемы наращивания сложности частных описаний. Общая характеристика этих алгоритмов МГУА дана в гл. 1. Здесь рассмотрим более подробно вычислительные особенности комбинаторного алгоритма по каждому из указанных основных блоков.

Блок формирования базиса. Рассмотрим случай полиномиальных опорных функций. Если заданы измерения некоторых входных переменных z_1, \dots, z_v моделируемого объекта и максимальная степень полинома, то число слагаемых n в полном полиноме степени σ_{\max} от v переменных определяется однозначно:

$$n = \prod_{j=1}^v \frac{\sigma_{\max} + j}{j} = \frac{(\sigma_{\max} + v)! - (\sigma_{\max})!}{v!}. \quad (3.6)$$

Сам полный полином записывается в следующем общем виде:

$$y = \sum_{i=1}^n a_i \prod_{j=1}^v z_j^{\sigma_{ij}} \stackrel{\Delta}{=} \sum_{i=1}^n a_i x_i, \quad (3.7)$$

где каждый обобщенный «линейный» аргумент x_i является нелинейной функцией исходных переменных z_j :

$$x_i = \prod_{j=1}^v z_j^{\sigma_{ij}}. \quad (3.8)$$

Степени σ_{ij} аргументов могут принимать целые значения $0, 1, \dots, \sigma_{\max}$, причем так, чтобы их сумма не превышала значения σ_{\max} , т. е. они должны удовлетворять ограничению

$$\sum_{j=1}^v \sigma_{ij} \leq \sigma_{\max}.$$

Для формирования степеней аргументов удобно организовать процедуру получения последовательности v -разрядных чисел $\sigma_i = (\sigma_{1i}, \dots, \sigma_{vi})$ с основаниями $1, 2, \dots, \sigma_{\max} + 1$ и отбирать те из них, которые удовлетворяют указанному ограничению. Например, при $v=2$ (два аргумента) и $\sigma_{\max}=2$ получим последовательность $n=6$ чисел σ_i , $i=1, 6$: 00, 10, 01, 20, 11, 02, что соответствует полному полиному

$$y = a_1 + a_2 z_1 + a_3 z_2 + a_4 z_1^2 + a_5 z_1 z_2 + a_6 z_2^2.$$

Итак, члены x_i полного полинома являются базисным набором опорных функций для комбинаторного алгоритма. Описанная процедура позволяет сформировать матрицу измерений обобщенных аргументов $X[N \times n]$, где N — число точек измерений. Как обычно, далее будем предполагать, что эти точки разделены (подряд) на

три последовательности: обучающую A длиной N_A , проверочную B длиной N_B и экзаменационную C длиной N_C , причем $N_A + N_B + N_C = N$.

Отметим, что для перехода к другому базису, например разностному или тригонометрическому, достаточно вычислить соответствующим образом аргументы z_j (пример приведен в гл. 1) и далее воспользоваться описанной процедурой формирования матрицы X .

Блок генерирования (перебора) частных моделей. Основными операциями, выполняемыми в этом блоке, являются: формирование структуры очередной частной модели; формирование соответствующей нормальной системы уравнений; решение полученной системы (оценка коэффициентов модели).

Формирование структур частных моделей формализуется с помощью *двоичного структурного вектора* $d = (d_1, d_2, \dots, d_n)$: если элемент d_i этого вектора принимает значение 1, то соответствующий i -й аргумент включается в частную модель, если 0 — не включается ($i=1, n$). Изменение состояний вектора d можно организовать многими способами (некоторые из них описаны в [141]). Наиболее простым по своей идее является следующий способ: получать все возможные варианты размещения в векторе d сначала одной единицы (всего $C_n^1 = n$ вариантов), затем двух (всего $C_n^2 = \frac{1}{2} n(n+1)$ вариантов) и т. д., вплоть до n единиц ($C_n^n = 1$ вариант). Общее число вариантов и составит $p_n = 2^n - 1$ (3.5) различных структур, т. е. полный перебор. Однако программная реализация этого способа (см., например, [1]) является далеко не быстрой, действующей, и такой перебор структур целесообразно применять только в алгоритмах неполного перебора (о чем будет идти речь дальше).

Наиболее компактной и универсальной, по-видимому, является схема изменения двоичного вектора по принципу работы двоичного счетчика, в последний разряд которого добавляется единица. При этом, что существенно, соблюдается взаимно однозначное соответствие между порядковым номером очередной частной модели и состоянием структурного вектора (соответствие десятичного и двоичного чисел). Этот способ алгоритмически очень прост и весьма удобен именно при полном переборе, несмотря на то что количество и состав аргументов в частных моделях все время меняются (например, при $n=3$ последовательно получаются структуры: 001, 010, 011, 100, 101, 110, 111). Более того, он позволяет даже организовать рекуррентную процедуру перебора.

Для формирования *нормальной системы* уравнений, соответствующей очередному структурному вектору, можно поступить формально: из столбцов полной матрицы X , указанных единичными элементами d , составляется частная матрица X_i , а затем вычисляются элементы нормальной матрицы $X_i^T X_i, X_i^T y$. Однако при пол-

ном переборе этот формальный путь является наихудшим, поскольку приводит к многократному вычислению одних и тех же скалярных произведений. Например, для структур 010, 011 и 110 трижды вычисляется одна и та же величина $x_2^T x_2 = \sum_{j=1}^N x_{2j}^2$. Поэтому в комбинаторном алгоритме достаточно только один раз вычислить матрицы полной нормальной системы, которая содержит элементы всех частных нормальных систем:

$$X^T X = \begin{pmatrix} x_1^T x_1 & x_1^T x_2 & \dots & x_1^T x_n \\ x_2^T x_1 & x_2^T x_2 & \dots & x_2^T x_n \\ \dots & \dots & \dots & \dots \\ x_n^T x_1 & x_n^T x_2 & \dots & x_n^T x_n \end{pmatrix}, \quad X^T y = \begin{pmatrix} x_1^T y \\ x_2^T y \\ \dots \\ x_n^T y \end{pmatrix}. \quad (3.9)$$

Для получения любой частной нормальной системы достаточно взять элементы матрицы $X^T X$, находящиеся на пересечении строк и столбцов, указанных единицами вектора d , а также соответствующие элементы вектора $X^T y$.

Для решения каждой нормальной системы $X_l^T X_l a_l = X_l^T y$, т. е. для вычисления оценок коэффициентов частной модели, можно применять любые процедуры решения систем алгебраических уравнений (с симметричной матрицей) с хорошими вычислительными свойствами (в смысле точности и устойчивости).

Блок отбора по внешним критериям. Вычисления в этом блоке, как и в предыдущем, ограничиваются с учетом применяемых внешних критериев. Если критерий несимметричный — типа регулярности $\Delta^2(B)$, то оценки коэффициентов a_l вычисляются на одной последовательности, а значение критерия — на другой (чем обеспечивается его «внешний» характер):

$$\Delta_l^2(B) = (y_B - X_{lB} \hat{a}_l(A))^T (y_B - X_{lB} \hat{a}_l(A)). \quad (3.10)$$

Если критерий симметричный, то для одной и той же структуры d_l коэффициенты вычисляются и на A , и на B , а затем вычисляется критерий — например критерий минимума смещения:

$$n_{cm}^2 = (X_l \hat{a}_l(B) - X_l \hat{a}_l(A))^T (X_l \hat{a}_l(B) - X_l \hat{a}_l(A)). \quad (3.11)$$

Селекция (отбор) моделей, лучших по заданному критерию I_l , обычно выполняется не в конце перебора (когда получены все частные модели), а в процессе его. Для этого запоминаются значения критерия для заданного числа F (свобода выбора [140]) первых моделей, а затем величина критерия каждой последующей модели I_l сравнивается с худшим I_{max} из F значений. Если $I_l < I_{max}$, то новая модель запоминается вместо худшей (запоминаются структура, оценки коэффициентов и критерий); если же $I_l \geq I_{max}$, то эта модель пропускается. После окончания перебора

оставшиеся F моделей являются лучшими из всех p_n в смысле заданного критерия. Очевидно, что при таком подходе блоки генерации и отбора алгоритмически объединяются.

В алгоритмах самоорганизации применяется также отбор по двум и более критериям. Например, критерием может быть сумма двух описанных: $k^2 = \alpha n_{cm}^2 + (1-\alpha) \Delta^2(B)$. Критерии такого типа называются комбинированными, однако использование их в виде суммы затрудняется сложностью выбора весового коэффициента α . Поэтому обычно несколько критериев применяются последовательно — по степени важности того или иного свойства моделей, которое оценивается по критерию: например, по n_{cm}^2 отбирается F_1 наименее смещенных моделей, затем из них по $\Delta^2(B) — F < F_1$ самых точных на проверочной последовательности.

Для повышения достоверности и помехоустойчивости выбора может использоваться также информация третьей части C выборки. При этом выполняется сначала этап адаптации коэффициентов — пересчет оценок F_1 моделей, отобранных на предыдущем этапе, по точкам двух первых выборок $A \cup B$. Затем из F_1 моделей отбирается $F < F_1$ по новому критерию регулярности $\Delta^2(C)$. Например, двухэтапный отбор ($n_{cm}^2 \rightarrow \Delta^2(C)$) намного повышает помехоустойчивость моделирования по комбинаторному алгоритму МГУА (пример см. в гл. 5).

Отметим, что в алгоритмах МГУА обычно выполняется еще один этап вычислений — оценки качества отобранных лучших моделей. При этом вычисляются, например, квадратическая ошибка аппроксимации $\varepsilon^2(A+B)$ (точность оценки МНК на $A \cup B$) и ошибка экстраполяции (прогноза) $\Delta^2(C)$.

Изложенные принципы организации вычислений реализованы в комбинаторном алгоритме МГУА [146, 144], имеющем название СОМВИ.

3. УМЕНЬШЕНИЕ ЗАТРАТ ПАМЯТИ И ВРЕМЕНИ СЧЕТА В АЛГОРИТМАХ МГУА С ПРИМЕНЕНИЕМ МАТРИЦ НОРМАЛЬНЫХ УРАВНЕНИЙ

Во всех разработанных алгоритмах МГУА существенная часть памяти ЭВМ отведена для так называемой системы условных уравнений. Появление новых задач моделирования процессов с распределенными параметрами (типа «полей загрязнения» [72]), для которых характерно большое число опытных данных, привело к тому, что некоторые варианты таких задач нельзя решать из-за превышения объема оперативной памяти даже таких ЭВМ, как БЭСМ-6.

Отметим, что в этих алгоритмах в память ЭВМ вводятся значения исходных переменных, по которым затем формируется матрица U преобразованных (эквивалентных) переменных, содержащая исходные значения аргументов (в том числе с запаздываниями), их попарные произведения, квадраты, функцию источника

и т. д. Рассмотрим модификацию структуры алгоритмов МГУА, не требующую запоминания системы условных уравнений.

Пусть на первом ряду селекции алгоритма МГУА сформирован полный вектор эквивалентных аргументов $x^t = (x_1, x_2, \dots, x_n)$, содержащий исходные аргументы, их степени, смешанные произведения, запаздывания и т. д. Для каждой i -й точки исходных данных можно записать следующее полное уравнение связи выходной величины y_i с вектором аргументов x_i :

$$y_i = \sum_{j=1}^n a_j x_{ij} = a^t x_i, \quad (3.12)$$

где $a^t = (a_1, a_2, \dots, a_n)$ — вектор неизвестных коэффициентов. Полная система из N условных уравнений вида (3.12) запоминается в существующих алгоритмах в виде матрицы

$$U = (X | y) = \left[\begin{array}{c|c} x_1^t & y_1 \\ x_2^t & y_2 \\ \dots & \dots \\ x_N^t & y_N \end{array} \right]. \quad (3.13)$$

Эта матрица используется для:

- а) ранжировки точек исходных данных с последующим их разделением, например, на три последовательности A , B и C длины N_A , N_B и N_C соответственно;
- б) формирования систем условных уравнений для различных частных моделей;
- в) соответствующего вычисления частных нормальных систем уравнений;
- г) вычисления величины заданного критерия селекции;
- д) формирования матрицы условных уравнений для следующего ряда селекций.

Покажем, что для всех этих операций система (3.13) в явном виде не нужна. При этом предположим, что данные об исходных значениях аргументов (без степеней, запаздываний и т. д.) содержатся в памяти ЭВМ и имеется подпрограмма для получения одной строки матрицы U .

Для ранжировки и разделения данных достаточно получить каждую «точку», т. е. строку матрицы U , вычислить ее дисперсию и расположить полученные величины в убывающем порядке, поставив каждой в соответствие номер строки. Далее будем считать, что разделение на последовательности A , B , C выполнено.

Формирование матриц условных уравнений для каждого частного описания с последующим вычислением соответствующих нормальных систем в алгоритмах МГУА в принципе является нецелесообразным, так как приводит к многократному вычислению одних и тех же величин. Действительно, матрица нормальных

уравнений для полного вектора аргументов

$$H_A = [X_A^t X_A | X_A^t y_A], \quad (3.14)$$

полученная на последовательности A , содержит матрицы всевозможных частных систем. Таким образом, для получения оценок коэффициентов частных описаний достаточно один раз построить полную матрицу (3.14) и «извлекать» из нее необходимые частные нормальные системы.

Оказывается, что матрицу H_A также можно вычислить с помощью обработки последовательно получаемых строк матрицы U . Это следует из таких очевидных формул:

$$X_A^t X_A = \sum_{i=1}^{N_A} x_{iA} x_{iA}^t, \quad X_A^t y = \sum_{i=1}^{N_A} x_{iA} y_{iA}. \quad (3.15)$$

При этом строки матрицы U вычисляются вторично с помощью нетрудоемких операций, но в память ЭВМ не вносятся. Параллельно получается матрица H_B для последовательности B .

Рассмотрим вопрос о вычислении значений критериев селекции для произвольной частной модели с вектором коэффициентов a , оценки которого уже получены на A или на B в отдельности. С учетом формул, аналогичных (3.12), можно получить выражения для критериев в следующем виде:

критерий несмещенностии (3.11)

$$n_{cm}^2 = (\hat{y}_A - \hat{y}_B)^t (\hat{y}_A - \hat{y}_B) = (\tilde{a}_A - \tilde{a}_B)^t \tilde{X}^t \tilde{X} (\tilde{a}_A - \tilde{a}_B), \quad (3.16)$$

где \hat{y}_A , \hat{y}_B — оценки выходной величины, полученные по коэффициентам \tilde{a}_A и \tilde{a}_B , вычисленным на последовательностях A и B соответственно, а матрица $\tilde{X}^t \tilde{X}$ равна сумме матриц $\tilde{X}_A^t \tilde{X}_A$ и $\tilde{X}_B^t \tilde{X}_B$;

критерий регулярности (3.10)

$$\Delta^2(B) = (y_B - \hat{y}_B^A)^t (y_B - \hat{y}_B^A) = y_B^t y_B - 2\tilde{a}_A^t \tilde{X}_B^t y_B + \tilde{a}_A^t \tilde{X}_B^t \tilde{X}_B \tilde{a}_A, \quad (3.17)$$

где y_B — исходный вектор измерений выходной величины на последовательности B ; \hat{y}_B^A — оценки выхода на B по модели \tilde{a}_A .

Величины с тильдой относятся к рассматриваемому частному описанию. Отметим, что выражение (3.16) впервые получено в [21], но применялось не для вычисления критерия n_{cm}^2 , а для исследования его свойств.

Очевидно, что для $\Delta^2(C)$ можно записать формулу, аналогичную (3.17):

$$\Delta^2(C) = y_C^t y_C - 2\tilde{a}_W^t \tilde{X}_C^t y_C + \tilde{a}_W^t \tilde{X}_C^t \tilde{X}_C \tilde{a}_W, \quad (3.18)$$

где \tilde{a}_W — оценки коэффициентов частных моделей, отобранных по n_{cm}^2 или $\Delta^2(B)$ после пересчета (адаптации) на объединенной последовательности $A \cup B = W$.

Из (3.16) — (3.18) видно, что для нахождения значения критерия, характеризующего каждую частную модель, достаточно иметь только соответствующие оценки коэффициентов. Матрицы же $\tilde{X}^T \tilde{X}$, $\tilde{X}_B^T \tilde{X}_B$, $\tilde{X}_C^T \tilde{X}_C$ — элементы соответствующих полных матриц нормальных уравнений, вычисляемых только один раз, так же как и скаляры $y_B^T y_B$, $y_C^T y_C$ и векторы $\tilde{X}_B^T y_B$, $\tilde{X}_C^T y_C$.

Очевидно, что ошибку МНК, определяемую после пересчета коэффициентов на W , также можно выразить через нормальные матрицы:

$$e^2(W) = (\hat{y}_W - y_W)^T (\hat{y}_W - y_W) = y_W^T y_W - \tilde{a}_W^T \tilde{X}_W^T y_W, \quad (3.19)$$

где учтено, что

$$\tilde{a}_W^T \tilde{X}_W^T \tilde{X}_W \tilde{a}_W = \tilde{a}_W^T \tilde{X}_W^T \tilde{X}_W (\tilde{X}_W^T \tilde{X}_W)^{-1} \tilde{X}_W^T y_W = \tilde{a}_W^T \tilde{X}_W^T y_W.$$

Формулы (3.16) — (3.19), кроме того, позволяют экономно вычислять также различные комбинированные критерии типа $n_{cm}^2 + \Delta^2(B)$ или $n_{cm}^2 + \Delta^2(C)$ и др.

Экономность полученных формул вычисления критериев состоит не только в том, что они позволяют обходиться без запоминания матриц систем условных уравнений, а и в выигрыше в скорости счета ввиду использования матриц размерностью не более $n \times n$ вместо матриц $N \times n$ при обычных вычислениях (так как, например, $\hat{y} = X\hat{a}$, $X [N \times n]$). Таким образом, переход от расчета критериев «по решениям» (с вычислением оценок вида \hat{y}_B^A и др.) к расчету «по коэффициентам» уменьшает число элементарных операций примерно в $\frac{N}{n}$ раз для n_{cm}^2 , в $\frac{N_B}{n}$ раз для $\Delta^2(B)$ и т. д. Очевидно, что для $\Delta^2(B)$ ускорение счета будет только в случае, когда число членов частной модели меньше числа точек в проверочной последовательности. Для n_{cm}^2 выигрыш будет всегда, так как необходимым условием для вычисления оценок \hat{a}_A и \hat{a}_B является $N \geq 2n$, т. е. $\frac{N}{n} \geq 2$.

Отметим, что применение (3.16) — (3.18) в тестовых примерах при точных (незашумленных) данных может давать вместо нулевых значений критериев очень малые отрицательные величины. Объясняется это, как указано, например, в [167], особенностями вычисления скалярных произведений в арифметике с плавающей запятой: при нормализации результата сложения двух близких по величине, но разных по знаку чисел появляются неверные циф-

ры. При зашумленных данных, а также в реальных задачах моделирования критерии вычисляются через нормальные матрицы без погрешностей.

Наконец, для перехода на следующий ряд селекции после отбора F лучших частных моделей первого ряда необходимо в третий раз последовательно вычислить строки матрицы U и построчно заполнить новую непреобразованную (без нелинейностей) матрицу исходных данных.

Более того, в линейных многорядных алгоритмах МГУА можно получить нормальную матрицу следующего ряда непосредственно через нормальную матрицу предыдущего. Действительно, пусть, например, на первом ряду отобрано F лучших моделей, коэффициенты которых записаны в матрице $A [n \times F] = (\hat{a}_1 | \hat{a}_2 | \dots | \hat{a}_F)$. Оценки выходов по соответствующим моделям представим матрицей

$$\hat{Y} = XA, \quad \hat{Y} [N \times F].$$

Поскольку столбцы \hat{Y} в соответствии с процедурой многорядного алгоритма являются аргументами для следующего ряда, матрица \hat{Y} является матрицей условных уравнений. Тогда полная нормальная матрица имеет вид

$$\hat{Y}^T \hat{Y} = A^T X^T X A, \quad (3.20)$$

т. е. может быть вычислена непосредственно через $X^T X$ предыдущего ряда и оценки коэффициентов F отобранных лучших моделей. Очевидно, что аналогично вычисляется и столбец правой части нормальной системы

$$\hat{Y}^T y = A^T X^T y, \quad (3.20a)$$

где $X^T y$ определено на предыдущем ряде; y — табличные значения выхода моделируемого объекта.

Таким образом, предложенная модификация алгоритмов МГУА позволяет выполнять все необходимые операции без записи в оперативную память ЭВМ полной системы условных уравнений в явном виде. При этом достигается унификация алгоритмов и существенно экономится время счета. Очевидно, что все эти предложения полностью применимы и в комбинаторном алгоритме МГУА, причем здесь выигрыш получается больше в ускорении счета. Это связано с тем, что полный перебор реально возможен, например, по алгоритму [146] примерно при $n \leq 18$ и в реальных задачах затраты памяти всегда допустимы. Однако применение описанных приемов позволяет получить скорость счета по СОМВИ, не зависящую от числа точек N исходных данных, что весьма существенно.

Что касается многорядных алгоритмов, то предложенная модификация не является полностью универсальной. Ускорение счета и «удаление» из памяти ЭВМ матрицы условных уравнений про-

исходит за счет запоминания двух матриц нормальных уравнений размерностью $n(n+1)$. Поэтому модификация эффективна для всех задач, в которых число эквивалентных (переобозначенных) аргументов позволяет записать в оперативной памяти матрицу $2n(n+1)$. Это ограничение существенно снижается, если учесть, что матрицы нормальных уравнений симметричны, и достаточно запоминать лишь диагональ и, например, нижнюю треугольную часть матриц. При этом общее число запоминаемых элементов равно $n(n+3)$. Для примера отметим, что на БЭСМ-6 реализация изложенного предложения позволит эффективно решать задачи идентификации размерностью до 150 аргументов с практической произвольным числом точек исходных данных.

4. РЕКУРРЕНТНОЕ ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ МОДЕЛЕЙ В КОМБИНАТОРНОМ АЛГОРИТМЕ

Рассмотрим задачу вычисления обратных матриц типа $(X^t X)^{-1}$ в комбинаторном алгоритме. Пусть на каком-то этапе работы алгоритма определены оценки параметров модели, содержащей s аргументов: x_1, \dots, x_s . Требуется построить эффективную процедуру оценки коэффициентов модели, полученной добавлением $(s+1)$ -го аргумента к предыдущим.

Обозначим

$$X_s a_s = y_s \quad (3.21)$$

уравнение измерений, или систему N условных уравнений для частной модели, содержащей s неизвестных коэффициентов (из полного числа n).

Расширенную матрицу нормальных уравнений, полученную по МНК для s аргументов, обозначим так:

$$[H_s | g_s] \triangleq [X_s^t X_s | X_s^t y], \quad (3.22)$$

причем здесь индекс одновременно соответствует размерности квадратной матрицы H_s и вектора g_s . Если каким-либо образом вычислена H_s^{-1} , то оценка коэффициентов имеет вид

$$\hat{a}_s = H_s^{-1} g_s. \quad (3.23)$$

В существующих алгоритмах эта оценка находится решением системы линейных уравнений по методу Гаусса (см., например, [144]).

При добавлении одного аргумента матрица нормальной системы принимает вид (с учетом ее симметрии)

$$[H_{s+1} | g_{s+1}] = \begin{bmatrix} H_s & h_{s+1} & g_s \\ h_{s+1}^t & \omega_{s+1} & \gamma_{s+1} \end{bmatrix}, \quad (3.24)$$

где h_{s+1} [$s \times 1$] — вектор; ω_{s+1} , γ_{s+1} — скаляры, значения которых определяются вектором измерений $(s+1)$ -го аргумента.

В комбинаторном алгоритме [146] оценка вектора \hat{a}_{s+1} опять находится по методу Гаусса, т. е. выполняются операции, аналогичные новому обращению матрицы $H_{s+1} : \hat{a}_{s+1} = H_{s+1}^{-1} g_{s+1}$. В то же время структура этой оценки

$$\begin{bmatrix} \hat{a}_s^* \\ \hat{\alpha}_{s+1} \end{bmatrix} = \begin{bmatrix} H_s & h_{s+1} \\ h_{s+1}^t & \omega_{s+1} \end{bmatrix}^{-1} \begin{bmatrix} g_s \\ \gamma_{s+1} \end{bmatrix} \quad (3.25)$$

предоставляет возможность выразить матрицу H_{s+1}^{-1} и соответственно новую оценку через уже известную H_s^{-1} и элементы «окаймления» h_{s+1} , ω_{s+1} , γ_{s+1} .

Действительно, обозначая

$$H_{s+1}^{-1} \triangleq Q_{s+1} = \begin{bmatrix} Q_s & q_{s+1} \\ q_{s+1}^t & \kappa_{s+1} \end{bmatrix} \quad (3.26)$$

и решая уравнение $Q_{s+1} H_{s+1} = E_{s+1}$ (или применяя известные формулы обращения блочных матриц), получаем

$$\kappa_{s+1} = \frac{1}{\omega_{s+1} - h_{s+1}^t H_s^{-1} h_{s+1}} = \frac{1}{\omega_{s+1} - h_{s+1}^t c_{s+1}}, \quad (3.27)$$

$$c_{s+1} = H_s^{-1} h_{s+1}, \quad q_{s+1} = -\kappa_{s+1} c_{s+1}, \quad Q_s = H_s^{-1} + \kappa_{s+1} c_{s+1} c_{s+1}^t.$$

Формулы (3.27), по существу, дают алгоритм вычисления H_{s+1}^{-1} . Однако записывать эту матрицу в явном виде не требуется, поскольку, как следует из (3.25) — (3.27),

$$\hat{a}_s^* = H_s^{-1} g_s + \kappa_{s+1} c_{s+1} (c_{s+1}^t g_s - \gamma_{s+1}), \quad \hat{\alpha}_{s+1} = -\kappa_{s+1} (c_{s+1}^t g_s - \gamma_{s+1}). \quad (3.28)$$

Таким образом, при добавлении к модели одного аргумента новая оценка вектора коэффициентов явно выражается через уже вычисленную (с учетом (3.23), (3.28)):

$$\hat{a}_{s+1} = \begin{bmatrix} \hat{a}_s^* \\ \hat{\alpha}_{s+1} \end{bmatrix} = \begin{bmatrix} \hat{a}_s - \hat{\alpha}_{s+1} c_{s+1} \\ \hat{\alpha}_{s+1} \end{bmatrix}. \quad (3.29)$$

Очевидно, что применение формул (3.27) — (3.29) значительно сокращает число операций для вычисления \hat{a}_{s+1} по сравнению с прямым определением матрицы H_{s+1}^{-1} . Рекуррентный алгоритм (3.26) — (3.29) является наиболее эффективным именно в алгоритмах, в которых сравниваются частные модели, получаемые последовательным добавлением одного аргумента. При этом указанные формулы не требуют специальных операций обращения, поскольку для модели с одним аргументом «обратная матрица»

Таблица 2

Номер модели	Двоичный счетчик	Счетчик Гарсайда
1	1000	1000
2	1001	1001
3	1010	1011
4	1011	1010
5	1100	1110
6	1101	1111
7	1110	1101
8	1111	1100

и единственный коэффициент определяются алгебраически через скалярные величины:

$$\hat{a}_1 = H_1^{-1} \gamma_1 = \frac{1}{\omega_1} \gamma_1,$$

где $\omega_1 = \tilde{x}_1^T \tilde{x}_1$; \tilde{x}_1 — вектор измерений первого аргумента. Более того, этот алгоритм можно использовать для решения линейных систем алгебраических уравнений (в данном случае — с симметричной матрицей). Такой подход известен в литературе под названием «метода окаймления».

Существенные результаты по оценке эффективности метода окаймления для общего случая несимметричной матрицы получены в [25], где доказано, что, во-первых, этот метод полностью совпадает по результатам и числу операций с классическим методом Гаусса — Жордана, а, во-вторых, в случае симметричной положительно определенной матрицы системы он абсолютно устойчив.

Таким образом, метод окаймления вполне равноправен по сравнению с другими методами решения линейных уравнений. В алгоритмах же самоорганизации, основанных на прямом (однорядном) переборе усложняющихся моделей, он, по-видимому, наиболее целесообразен.

Однако возникает естественный вопрос: можно ли в комбинаторном алгоритме, осуществляющем полный перебор частных моделей различной сложности, организовать универсальную схему перебора, обеспечивающую вычисление каждой обратной матрицы только один раз, при любом сочетании аргументов? Оказывается, можно — этому условию удовлетворяет схема перебора с помощью двоичного регистра сдвига (двоичного вектора), уже реализованная в СОМВИ. При ее использовании в частную модель включаются аргументы, соответствующие состоянию двоичного регистра, установившемуся после добавления единицы в последнем разряде (число разрядов n равно числу аргументов). В табл. 2 показан пример последовательности изменения состояний двоичного счетчика, которому соответствуют структуры проверяемых частных моделей. Видно, что при переходе от нечетных моделей к четным происходит добавление нового члена и можно непосредственно применить формулы (3.27), но при этом необходимо как-то строить обратные матрицы для нечетных моделей. Однако имеются другие, более полезные закономерности: «обратную матрицу» модели 1 можно использовать для моделей 2, 3 и 5; обратная матрица модели 5 (полученная по рекуррентным формулам) пригодна для моделей 6 и 7; наконец, самое существенное то, что обратная матрица для любой комбинации из s аргументов перестает использоваться после появления следующего состояния счетчика, имеющего s единиц.

Это замечательное свойство двоичного регистра и позволяет на основе алгоритма последовательного обращения матриц (3.26) — (3.29) реализовать комбинаторный алгоритм МГУА, наиболее эффективный в вычислительном плане. При этом достигается наилучшее согласование схемы перебора моделей и метода вычисления оценок коэффициентов.

Таким образом, зарезервировав в ЭВМ память для последовательности массивов размера $1 \times 1, 2 \times 2, \dots, s \times s, \dots, (n-1) \times (n-1)$ (полная обратная матрица $n \times n$ не нужна), получим возможность вычисления оценок всех $2^n - 1$ частных моделей с помощью рекуррентных формул. Естественно, нужно хранить также последовательность векторов длины 1, 2, ..., $s, \dots, (n-1)$ с записанными коэффициентами моделей с соответствующим числом членов. При появлении двоичного числа с s единицами используются обратная матрица H_{s-1}^{-1} из массива $(s-1) \times (s-1)$ и оценка коэффициентов \hat{a}_{s-1} из вектора размера $(s-1)$. Важно отметить, что при этом строить и запоминать матрицу H_s^{-1} необходимо не для каждой модели, а через одну: наличие единицы в последнем разряде двоичного вектора исключает дальнейшее добавление нового члена к полученной их комбинации.

Указанные матрицы удобно хранить в памяти ЭВМ в виде общего одномерного массива, в котором каждая квадратная матрица записывается построчно: матрица 1×1 — в первом элементе, 2×2 — в последующих четырех, 3×3 — в последующих девяти и т. д. Так как имеет место соотношение

$$\sum_{i=1}^s i^2 = \frac{1}{6} s(s+1)(2s+1),$$

то местоположение каждой обратной $s \times s$ -матрицы идентифицируется однозначно, а общая длина «цепочки» обратных матриц равна $n(n-1)(2n-1)/6$. Аналогично организуется хранение последовательных значений векторов коэффициентов с учетом того, что сумма s первых членов натурального ряда равна $s(s+1)/2$.

Предложенный способ использования рекуррентного алгоритма (3.26) — (3.29) при комбинаторном переборе моделей дает, по-видимому, максимальное ускорение вычисления коэффициентов моделей. При этом, как видно из предыдущего, существенно возрастает используемая оперативная память ЭВМ. Если хранение всех элементов каждой обратной матрицы на конкретной ЭВМ ограничивает возможности использования комбинаторного алгоритма, то затраты на память можно уменьшить практически вдвое. Для этого надо организовать специальную процедуру записи и вызова только, например, верхней треугольной части матриц (воспользовавшись их симметрией). При этом элемент (j, l) квадратной $(i \times i)$ -матрицы находится в элементе с индексом $(2i-j) \times (i-j)/2 + l$ эквивалентного одномерного массива длины $i(i+1)/2$, а местоположение каждой $s \times s$ -матрицы при последовательной

записи идентифицируется с помощью соотношения

$$\sum_{i=1}^s \frac{i(i+1)}{2} = \frac{1}{6} s(s+1)(s+2).$$

Общая длина «цепочки» обратных матриц теперь равна $n(n-1)(n+1)/6$. Пример программы с использованием «уплотненной» записи симметричной матрицы можно найти в [1] (алгоритм 676).

Сравнение с другими способами перебора. Вопросам организации полного перебора моделей в заданном базисе, или «проверки всех регрессий», посвящено много публикаций. Интересный обзор современных работ на эту тему дан в [141], где рекомендуется при переборе применять вместо последовательности двоичных чисел метод Гарсайда. В этом методе модели, следующие подряд, отличаются только одним членом. Пример сравнения счетчика Гарсайда с двоичным приведен в таблице 2. Такой счетчик удобен тем, что позволяет рекуррентно вычислять коэффициенты моделей без использования «цепочки» обратных матриц: при добавлении члена применяются формулы (3.26) — (3.29), при исключении — обратные им.

Обратную рекуррентную процедуру для вычисления H_s^{-1} , \hat{a}_s по известным H_{s+1}^{-1} , \hat{a}_{s+1} можно получить непосредственно из (3.26) — (3.29). Если известны

$$H_{s+1}^{-1} = \begin{bmatrix} Q_s & q_{s+1} \\ q_{s+1}^T & x_{s+1} \end{bmatrix}, \quad \hat{a}_{s+1} = \begin{bmatrix} \hat{a}_s^* \\ \hat{x}_{s+1} \end{bmatrix}, \quad (3.30)$$

то после исключения $(s+1)$ -го члена получим

$$H_s^{-1} = Q_s - \frac{1}{x_{s+1}} q_{s+1} q_{s+1}^T, \quad (3.31)$$

$$\hat{a}_s = \hat{a}_s^* - \frac{\alpha_{s+1}}{x_{s+1}} q_{s+1}. \quad (3.32)$$

Ясно, что применение счетчика Гарсайда совместно с прямыми и обратными рекуррентными соотношениями позволяет существенно экономить память ЭВМ — все операции выполняются в одном массиве размера $n \times n$. Однако такая процедура уступает по числу выполняемых операций (т. е. по скорости счета) предложенной здесь процедуре, использующей двоичный счетчик совместно с методом окаймления. Это объясняется, по крайней мере, двумя причинами: во-первых, в процедуре со счетчиком Гарсайда обратная матрица вычисляется по (3.31) каждый раз (для каждой модели), а в предложенной (с двоичным счетчиком) — только через раз; во-вторых, для некоторых структур обратная $s \times s$ -матрица H_s^{-1} находится через матрицу $(s+1) \times (s+1)$ с применением ис-

ключения, а при двоичном счетчике — через матрицу $(s-1) \times (s-1)$ (например, для структур 1010 и 1100, см. табл. 2).

Таким образом, предложенное усовершенствование алгоритма COMBI позволяет за счет увеличения затрат памяти добиться максимального уменьшения времени полного перебора моделей по сравнению с известными способами.

Пример оценки эффективности новой схемы перебора. Для численной оценки эффекта уменьшения времени счета для усовершенствованного комбинаторного алгоритма был выполнен ряд расчетов на разных этапах создания новой программы. При этом использовались одни и те же исходные данные, «сгенерированные» по полиному от двух переменных:

$$y = 1 - 2x_2 + 3x_1^2 - 4x_2^2 + 5x_1^3. \quad (3.33)$$

Всего было взято 100 точек, разделенных пополам на последовательности A и B . Исходные данные и коэффициенты полученных моделей здесь не приводятся, так как наша цель — показать пути ускорения вычислений. Отметим лишь, что результаты решения линейных систем уравнений, т. е. коэффициенты перебираемых частных моделей, вычисленные по стандартной процедуре (использующей метод Гаусса) и рекуррентному алгоритму, совпадают полностью.

В качестве базисных аргументов были приняты члены полного полинома четвертой степени от двух переменных при фиксированном свободном члене. При этом полное число всевозможных частных моделей (включавшее модель со структурой (3.33)) равно $2^{14} = 16\,384$. Время счета такого примера на БЭСМ-6 изменялось следующим образом: по исходному алгоритму COMBI — 11 мин 13 с; после улучшения блоков формирования двоичного вектора и отбора лучших моделей — 10 мин 23; после перехода к вычислению критериев селекции через нормальные матрицы — 6 мин 17 с; с применением рекуррентного алгоритма — 2 мин 21 с. При этом применялся критерий регулярности.

Таким образом, после реализации разработанных предложений время счета конкретного примера уменьшилось практически в пять раз. Это расхождение должно увеличиваться с ростом количества членов полного полинома: при увеличении n число элементарных операций возрастает пропорционально n^3 для метода Гаусса и n^2 для метода окаймления (в описанном выше варианте). Такой вывод подтверждается решением того же контрольного примера, но без «протекции» для свободного члена (число частных моделей при этом в два раза возрастает): время счета до усовершенствования равно 21 мин 13 с, после — 3 мин 47 с. Это означает, что теперь при использовании комбинаторного алгоритма структуру полного полинома можно увеличить примерно на 5 членов, т. е. довести до $n=23$. При этом затраты оперативной памяти ЭВМ, связанные с записью «цепочки» обратных матриц, остаются вполне допустимыми.

5. КОНЕЧНАЯ СЕЛЕКЦИОННАЯ ПРОЦЕДУРА СОКРАЩЕНИЯ ОБЪЕМА ПОЛНОГО ПЕРЕБОРА МОДЕЛЕЙ

Так как объем полного перебора p_n является показательной функцией числа базисных аргументов n , то очевидна ограниченность возможностей комбинаторных алгоритмов МГУА или подобных им (в смысле применения полного перебора моделей) процедур метода всех регрессий [141]. Даже с применением быстродействующей схемы перебора, описанной выше и требующей максимального использования оперативной памяти ЭВМ, комбинаторный алгоритм эффективен практически при числе аргументов порядка двадцати. Поэтому для решения более сложных задач необходимо искать способы сокращенного перебора, позволяющие в отличие от многорядных алгоритмов МГУА оставаться в заданном базисном множестве функций (аргументов) при высокой вероятности получения результата полного перебора.

В настоящее время часто применяется способ «сокращения треугольника перебора» [49]: сравниваются последовательно все модели с одним, двумя, тремя и т. д. аргументами, до тех пор пока время счета остается приемлемым. Однако если полученная таким образом модель содержит $s_{\max} < s_0$ аргументов, где s_0 — размерность истинной модели, то задача самоорганизации в смысле (3.4) остается нерешенной. Причем этот способ быстро себя исчерпывает: если, например, без осложнений можно перебрать $p_{17}=131\,072$ модели при $n=17$, то при $n=20$ «треугольник» от 1

до 7 аргументов содержит уже $\sum_{s=1}^7 C_{20}^s = 137\,972 > p_{17}$ моделей, а при $n=30$ даже $\sum_{s=1}^5 C_{30}^s = 184\,756 > p_{17}$.

Оригинальный способ преодоления трудностей перебора предложен в [28], где задача комбинаторного алгоритма интерпретируется как задача отыскания глобального минимума существенно многоэкстремальной одномерной функции изменения внешнего критерия, т. е. осуществляется прямая оптимизация вида (3.4).

В данном разделе предлагается новый алгоритм сокращения полного перебора моделей, являющийся, как и многорядные алгоритмы, селекционным, однако имеющий конечное число этапов селекции (не более n) и позволяющий искать модель в том же заданном функциональном базисе, который используется в комбинаторных алгоритмах. Предлагаемая схема сокращенного перебора основана на идее сглаживания многоэкстремальной функции изменения внешнего критерия в одноэкстремальную, минимум которой соответствует модели оптимальной сложности. Существенно то, что при этом сложность моделей возрастает на каждом ряду только на один аргумент.

Многоэтапный селекционно-комбинаторный алгоритм. Общая схема алгоритма имеет такой вид: сначала оцениваются все мо-

дели, содержащие только один аргумент, и выбирается несколько лучших из них; на втором этапе к этим моделям добавляются различные аргументы по одному и отбираются модели с двумя аргументами, улучшающие величину критерия, и т. д. до его ухудшения. При этом все генерируемые модели остаются в исходном базисе, причем число содержащихся в них аргументов совпадает с номером этапа, а общее число этапов не может быть больше n . Такой алгоритм назовем многоэтапным в отличие от многорядных, в которых число этапов-рядов может быть бесконечным.

Основной составной частью алгоритма является описанная выше рекуррентная процедура последовательного оценивания коэффициентов частных моделей по МНК, реализующая метод окаймления. Представим матрицу нормальных уравнений для модели с аргументами, полученную из полной нормальной $n \times (n+1)$ -матрицы $[H|g] = [X^T X | X^T y]$ в виде

$$H_r = \begin{bmatrix} H_{r-1} & h_r \\ h_r^T & \omega_r \end{bmatrix}, \quad g_r = \begin{bmatrix} g_{r-1} \\ \gamma_r \end{bmatrix} \quad (3.34)$$

Тогда оценка \hat{a}_r вычисляется с помощью следующих операций:

$$\kappa_r = 1/(\omega_r - h_r^T c_r), \quad c_r = H_{r-1}^{-1} h_r, \quad H_0^{-1} = 0, \quad (3.35)$$

$$\hat{a}_r = \begin{bmatrix} \hat{a}_{r-1} \\ \hat{\alpha}_r \end{bmatrix} = \begin{bmatrix} \hat{a}_{r-1} - \hat{\alpha}_r c_r \\ \kappa_r (\gamma_r - g_{r-1}^T c_r) \end{bmatrix}, \quad \hat{a}_0 = 0. \quad (3.36)$$

Новая обратная матрица принимает вид

$$H_r^{-1} = \begin{bmatrix} H_{r-1}^{-1} + \kappa_r c_r c_r^T & -\kappa_r c_r \\ -\kappa_r c_r^T & \kappa_r \end{bmatrix}. \quad (3.37)$$

Приведем теперь последовательность операций, выполняемых на каждом r -м этапе алгоритма, обратив предварительно внимание на отличие данной схемы формирования частных описаний от применяемых в обычных многорядных алгоритмах МГУА. В обычных алгоритмах «скрещиваются» на каждом ряду все модели предыдущего ряда [144] и, кроме того, в селекции могут участвовать также исходные (базисные) аргументы [174]. В рассматриваемом же алгоритме селекция между моделями предыдущего ряда не допускается, «скрещиваются» они только с базисными аргументами, что и позволяет все время оставаться в исходном базисе.

1. Для каждой i -й модели предыдущего этапа, $i=1, F_{r-1}$, вычисляется обратная матрица $H_{r-1}^{-1}(i)$, например, методом квадратного корня (алгоритм 666 в [1]). Для $r=1$ полагается $F_0=1$.

2. Генерируются частные модели r -го этапа упорядоченным добавлением к каждой i -й модели одного из отсутствующих в ней аргументов, соответствующих нулевым элементам двоичного струк-

турного вектора i -й модели. Следовательно, на r -м этапе генерируется максимум $q_r = (n-r)F_{r-1}$ частных моделей со сложностью (числом аргументов) r , причем q_r является завышенной оценкой, так как она учитывает возможные повторения структур, которые несложно алгоритмически избежать. Каждая l -я новая модель, $l = \overline{1, q_r}$, однозначно определяется номером i модели $(r-1)$ -го этапа, $i = \overline{1, F_{r-1}}$, и номером j добавленного аргумента, $j = \overline{1, (n-r)}$.

3. Вычисляются по (3.35), (3.36) оценки коэффициентов $\hat{a}_r(i, j)$ при каждом i на основе одной и той же матрицы $H_{r-1}^{-1}(i)$. При этом обратные матрицы $H_r^{-1}(l) = H_r^{-1}(i, j)$ для всех q_r моделей определять нет необходимости, причем по условию малости знаменателя в (3.35) легко автоматизируется отсев моделей с плохо обусловленной нормальной матрицей.

4. Вычисляются значения заданного критерия селекции для каждой из q_r частных моделей по формулам, выраженным через прямые нормальные матрицы $H_r(i, j)$.

5. Определяются F_r лучших моделей, передаваемых на следующий этап селекции, по условию улучшения минимального значения внешнего критерия I_{r-1} , полученного на предыдущем этапе:

$$F_r = \sum_{l=1}^{q_r} \delta_l, \quad \delta_l = \begin{cases} 1, & I_r(l) < I_{r-1}, \\ 0, & I_r(l) \geq I_{r-1}, \end{cases} \quad (3.38)$$

$$I_{r-1} = \min_{l=\overline{1, q_{r-1}}} I_{r-1}(l). \quad (3.39)$$

При этом следует иметь в виду, что такая процедура определения F_r применима в случае, если реально используемый внешний критерий в данном алгоритме является унимодальным. В общем случае этого доказать не удается, а численные исследования показывают, что данное упорядочение перебора для некоторых критериев может давать функцию с более чем одним экстремумом, однако их существенно меньше, чем при обычном полном переборе.

Для преодоления возможных локальных минимумов, когда при $r \ll n$ по (3.38) получаем $F_r = 0$, рекомендуется ввести ограничение снизу на величину свободы выбора, причем переменное, например $F_{r\min} = (n-r)$. При этом сохраняются наиболее перспективные направления усложнения моделей на начальных этапах перебора, что можно интерпретировать как внесение «мутаций» в процедуру направленной селекции. Можно подобрать и другой закон уменьшения $F_{r\min}$, например экспоненциальный.

С другой стороны, число F_r , полученное по (3.38), при больших n может оказаться очень большим, превосходящим возможности конкретных ЭВМ. Поэтому необходимо ввести некоторое целесообразное ограничение сверху F_{\max} . Таким образом, свобода выбора на r -м этапе определяется по (3.38) с учетом ограничения

$$F_{\min} \leq F \leq F_{\max}. \quad (3.40)$$

6. Принимается решение о прекращении процедуры или переходе на следующий этап (на п. 1). Очевидно, без учета ограничения (3.40) условием останова было бы $F_r = 0$, что означало бы достижение минимума критерия на $(r-1)$ -м этапе.

С учетом (3.40) процедуру можно пройти до конца, т. е. до $r=n$, и затем определить минимальное значение полученной сглаженной функции. Однако эффективнее применить более сложное правило останова «на скользящем интервале»: если на τ этапах подряд получаем $F_r = F_{\min}$, то функция критерия устойчиво возрастает, поэтому процедура останавливается, когда

$$\sum_{t=r}^{r-\tau+1} (F_t - F_{\min}) = 0;$$

при этом минимум пройден на этапе $r-\tau$. Обычно достаточно задать τ равным 2 или 3.

Основные свойства алгоритма. 1. Алгоритм является селекционным, но с конечным числом этапов перебора, причем свобода выбора на каждом этапе и момент останова определяются автоматически.

2. Все анализируемые модели принадлежат тому же функциональному базису, что и модели обычного комбинаторного алгоритма.

3. На каждом этапе сложность частных моделей увеличивается только на один аргумент, поэтому для оценки коэффициентов применяется экономная рекуррентная процедура. При этом затраты памяти оказываются теми же, что и в СОМВИ.

4. Алгоритм позволяет получить некоторые известные схемы перебора: при $F_r = F_{\max} = C_r n$ полный перебор (метод всех регрессий), при $F_r = F_{\min} = 1$ — метод последовательной регрессии.

5. Можно определить следующую оценку сверху общего числа моделей, сравниваемых в процессе самоорганизации модели оптимальной сложности по этому алгоритму:

$$\bar{q}_n = \sum_{r=1}^n q_{r\max} = \sum_{r=1}^n (n-r) F_{\max} = \frac{n(n-1)}{2} F_{\max}. \quad (3.41)$$

Таким образом, число проверяемых моделей \bar{q}_n является степенной функцией числа аргументов n вместо показательной зависимости p_n при полном переборе, т. е. алгоритм работоспособен и при больших n ввиду $\bar{q}_n \ll p_n$. Например, при $n=100$ и $F_{\max}=25$ оценивается не более $\bar{q}_{100}=123\,750$ моделей, что примерно эквивалентно объему полного перебора при $n=17$ ($\bar{q}_{100} < p_{17}$).

6. Если истинные аргументы сильнее коррелированы с выходом y , чем остальные (избыточные), алгоритм обеспечивает получение результата полного перебора, поскольку селекционная процедура поиска выбирает наиболее вероятные пути усложнения моделей. В сложных случаях вероятность достижения результата полного

Таблица 3

t	y	t	y	t	y
-3,0	-55,0	-0,5	0,125	1,5	-2,5
-2,5	-32,0	0,0	1,05	2,0	-2,9
-2,0	-19,5	0,5	0,39	2,5	-2,1
-1,5	-9,3	1,0	-1,03	3,0	0,97
-1,0	-2,9				

Таблица 4

Номер этапа	Структура модели	Критерий k
1	10000	0,28*
	01000	0,77*
	00100	0,92*
	00001	0,98
2	00011	0,14
	00110	0,15*
	11000	0,56
3	10110	0,03*
	00111	0,06*
4	11110	0,064
5	11111	1,03

Таблица 5

Номер этапа	Структура модели	Критерий $n_{\text{см}}$
1	00001	0,08
	00100	0,68
	10000	1,20
	01000	1,24
2	10001	0,11
	01001	0,22
	10100	0,59
3	10110	0,005
	01101	0,28
4	10111	0,59
5	11111	16,6

определения глобального минимума достаточно задать $F_1^{\text{min}}=3$, $F_2^{\text{min}}=\dots=F_5^{\text{min}}=0$. В первом случае сравнивается 27 частных моделей из 31 при $F_r=F_1^{\text{min}}$, во втором (отмечен звездочками в таблице) — только 20 при $F_1=3$, $F_2=1$, $F_3=2$, $F_4=0$ с автоматическим остановом. В данном случае комбинированный критерий k ведет себя аналогично идеальному (он описан в следующей главе).

Критерий $n_{\text{см}}$ даже в этом простом случае такой аналогии не выявил: все модели из двух аргументов оказались хуже, чем лучшая модель первого этапа (табл. 5, рис. 12). Однако применение ограничения $F_r \geq F_1^{\text{min}}$ помогло однозначно определить глобальный минимум, при этом сравнивалось 26 моделей. Восстановленная по обоим критериям модель имеет вид $y = 1,0016 - 2,954 t^2 + 1,0506 t^3$.

Отметим, что в рассмотренном примере число сравниваемых моделей q_5 оказалось близким к объему полного перебора p_5 . Это следствие того, что в описанном алгоритме практически всегда анализируются все модели с одним и двумя аргументами, и при $n=5$ уже имеем $C_6^1 + C_5^2 = 15$. При больших n , когда $C_n^2 \gg F_{\max}$, всегда справедливо соотношение $q_n \ll p_n$.

Рис. 10. Изменение критерия в процессе полного перебора моделей различной сложности.

Рис. 11. Зависимость критерия от сложности частных моделей.

перебора остается достаточно высокой, что объясняется, кроме указанного, также следующими факторами: каждая усложняемая модель проверяется на возможность включения в нее отсутствующих аргументов; для проверки единственности решения вносятся мутации за счет введения ограничения на свободу выбора F_{\min} ; после получения истинной модели (или модели оптимальной сложности при зашумленных данных) алгоритм гарантированно останавливается.

Численный пример. По описанному алгоритму была решена задача идентификации функции $y=1-3t^2+t^3$ с помощью двух критериев селекции — минимума смещения $n_{\text{см}}^2$ и комбинированного $k^2=n_{\text{см}}^2+\Delta^2(c)$. Для моделирования были взяты 13 точек, зашумленных выборкой 5 % равномерного шума (табл. 3) при разделении $N_A=N_B=5$, $N_C=3$ точки. Поиск модели осуществляется среди членов полинома четвертой степени, т. е. $n=5$, $p_5=31$.

На рис. 10 представлена функция изменения критерия k в процессе перебора по COMBI. Номерам i соответствуют модели со структурами, определяемыми единичными элементами эквивалентных двоичных чисел. Истинную модель со структурой 10110 указывает глобальный минимум этой многоэкстремальной функции в точке 22. Зависимость этих же значений критерия от числа аргументов (без учета состава) изображена на рис. 11. Видно, что огибающая снизу всех точек — одноэкстремальная функция сложности моделей r , которую можно построить по многоэтапному алгоритму.

Результаты всех этапов поиска минимума критерия k приведены в табл. 4. Минимальные значения критерия на каждом этапе k_r (рис. 12) совпадают с точками нижней огибающей на рис. 11. При этом ограничение $F_r^{\text{min}}=5-r$ оказалось избыточным — для

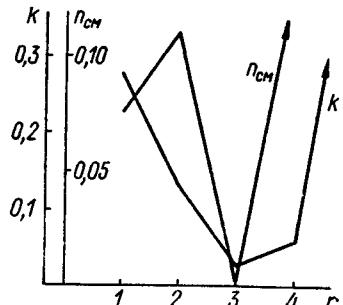


Рис. 12. Изменение лучших значений критериев на этапах селекционно - комбинаторного алгоритма.

выбора и момента останова. Получена оценка сверху для числа сравниваемых моделей, пропорциональная при больших n примерно n^3 вместо 2^n при полном переборе.

6. УНИВЕРСАЛЬНАЯ СХЕМА ПЕРЕБОРА МОДЕЛЕЙ НА ОСНОВЕ УПРАВЛЯЕМОГО КОМБИНАТОРНОГО АЛГОРИТМА

Применяемый в СОМВИ двоичный счетчик обладает достоинствами, позволяющими существенно расширить возможности полного перебора, а также построить алгоритм с универсальной структурой перебора. Например, очень просто исключить из рассмотрения модели с линейно зависимыми членами: если при использовании алгоритма последовательного обращения матриц очередное значение знаменателя x_{s+1} в (3.27) оказывается близким к нулю, т. е. матрица H_{s+1} вырождена, то соответствующий $(s+1)$ -й аргумент линейно зависит от предыдущих. Тогда нужно прекратить дальнейшее наращивание сложности частных моделей, включающих полученный набор $(s+1)$ -го аргумента, а для этого достаточно при вычислениях пропускать модели, содержащие соответствующую комбинацию единиц в двоичном векторе. Далее, чтобы обеспечить протекцию для $0 \leq p \leq n$ аргументов (членов) модели, достаточно в двоичном векторе ввести в первые p разрядов единицы и разрешить изменение состояния счетчика только в остальных разрядах. Отметим, что при этом число сравниваемых частных моделей уменьшается от $p_n = 2^n - 1$ до $p_p = 2^{n-p}$.

Наконец, двоичный регистр позволяет ввести в комбинаторный алгоритм специальное ограничение на сложность частных моделей: можно исключить из перебора модели с числом аргументов меньше заданного s_{\min} и больше s_{\max} , где $1 \leq s_{\min} \leq s_{\max} \leq n$. Очевидно, что для такого исключения достаточно следить за суммой единиц двоичного счетчика. При этом появляются новые для комбинаторного алгоритма возможности: а) учет априорной информации о том, что искомая модель оптимальной сложности не дол-

жна быть слишком простой или, наоборот, слишком громоздкой; б) идентификация моделей при числе аргументов $n > 20$: разумным заданием s_{\min} и s_{\max} можно за приемлемое время счета выполнить частичный перебор достаточно сложных моделей, посколь-

ку объем перебора уменьшается от $p_n = 2^n - 1$ до $p_s = \sum_{i=s_{\min}}^{s_{\max}} C_i$

в) контролируемый перебор, или перебор «по частям», если время полного расчета недопустимо велико: за время одного счета можно определить несколько лучших моделей с числом аргументов от $s_{\min} = 1$ до $s_{\max} = s_1$; за время второго счета — от $s_{\min} = s_1 + 1$ до $s_{\max} = s_2$; третьего — от $s_{\min} = s_2 + 1$ до $s_{\max} = n$; затем из всех полученных моделей выбрать лучшую по заданному критерию как результат полного перебора.

Однако введение ограничения на сложность частных моделей имеет также другое, еще более существенное значение — оно дает возможность создания весьма гибкого алгоритма МГУА, обобщающего ряд уже известных.

Допустим, что построен специальный (вообще говоря, многорядный) алгоритм МГУА, на каждом ряде которого вычисления выполняются по описанному выше управляемому комбинаторному алгоритму с возможностью введения ограничений на сложность s_{\min} , s_{\max} и назначения протекции ρ . Пусть также можно задавать максимальное число рядов селекции r , а базисными функциями являются полиномы, причем степень σ нелинейности задается на каждом ряде отдельно. Переход от ряда к ряду осуществляется обычным образом. Опишем возможности такого алгоритма, т. е. покажем достаточную универсальность (общность) его вычислительной схемы: при определенных сочетаниях значений r , σ , s_{\min} и s_{\max} формирование и наращивание сложности частных описаний происходит так же, как в некоторых известных алгоритмах МГУА.

Полный (комбинаторный) перебор. Очевидно, что при $r=1$ предложенный алгоритм работает как комбинаторный, причем при $s_{\min}=1$, $s_{\max}=n$ и произвольном σ выполняется, как и в СОМВИ, полный перебор всевозможных частных моделей различной сложности, структуры которых содержатся в формируемой алгоритмом структуре полинома максимальной сложности. Далее, при $\rho>0$ алгоритм обеспечивает протекцию ρ членам, а при заданных значениях s_{\min} , s_{\max} получаем комбинаторный алгоритм с контролируемым управляемым перебором, возможности которого описаны выше.

Многорядная селекция. Если задать значения $s_{\min}=s_{\max}=2$, то будет выполнен полный перебор всех частных моделей, содержащих лишь два аргумента — что в точности соответствует схеме вычислений одного ряда многорядного алгоритма МГУА. Таким образом, при числе рядов $r>1$, линейном полиноме $\sigma=1$, $s_{\min}=s_{\max}=2$ предложенный универсальный алгоритм формирует частные описания аналогично многорядному алгоритму с линейными частными описаниями (например, [64]), а для получения

аналога нелинейного многорядного алгоритма [110] достаточно задать $\sigma=2$ и соответствующие $s_{\min}=s_{\max}\geqslant 2$.

Многорядно-комбинаторная схема. Существенно иной будет совокупность генерируемых моделей в случае $s_{\min}=1$, $s_{\max}\geqslant 2$: в селекции участвуют частные описания не только в виде полного (линейного или нелинейного) полинома от двух аргументов (как при $s_{\min}=s_{\max}$), но и в виде всех структур меньшей сложности от этих же аргументов. Это соответствует генератору моделей в алгоритме смешанного типа [95] с многорядной селекцией и комбинаторной оптимизацией сложности частных описаний.

Схема с контролируемой нелинейностью. Интересные возможности наращивания сложности частных моделей можно получить, используя линейные описания, но изменяя при этом степень σ полного полинома от ряда к ряду. Например, если на первом ряде $\sigma_1>1$, а на последующих $\sigma_r=1$, $r\geqslant 2$, то будем получать модели не выше заданной степени σ_1 . Ведь при этом, как и в алгоритме МГУА «с переобозначением» [72], на первом ряде формируется полный полином степени σ_1 от всех исходных аргументов, члены которого используются затем в линейных частных описаниях на всех рядах. Можно также повышать степень нелинейности постепенно: если, например, при $\sigma_1=4$ число эквивалентных переменных (членов полного полинома) слишком велико, то достаточно задать $\sigma_1=2$ и $\sigma_2=2$, чтобы получить результирующую модель четвертой степени при существенно меньшем объеме задачи. Если же число самих исходных аргументов превышает возможности ЭВМ, то можно принять $\sigma_1=1$, и только после селекции на первом ряде приемлемого числа частных моделей вводить на следующих рядах требуемую нелинейность.

Схема с протекцией переменным. Очевидно, что в предложенном алгоритме просто учитывается необходимость протекции некоторым переменным: задав на первом ряде какое-то число ρ_1 , получим требуемые переменные во всех частных моделях, отобранных после первого ряда — а тем самым и в конечной модели. Отметим, что специальные алгоритмы МГУА с протекцией уже существуют [45].

Многорядный перебор общего вида. Исследуемый алгоритм обладает универсальностью в смысле включения в частные описания групп произвольного числа аргументов: вместо обычного попарного учета аргументов ($s_{\min}=s_{\max}=2$) естественным образом можно получить желаемый «групповой» учет — например, по одной переменной ($s_{\min}=s_{\max}=1$), по трем ($s_{\min}=s_{\max}=3$) и т. д. При этом можно не фиксировать сложность частных описаний, а оптимизировать ее, приняв $s_{\min}=1$, $2\leqslant s_{\max}\leqslant n$. Можно также применять протекцию, изменение степени нелинейности от ряда к ряду и другие возможности, упомянутые выше. Таким образом, обсуждаемая универсальная вычислительная структура является достаточно гибкой: она позволяет легко изменять способ «генерирования» моделей и использовать как ряд известных схем усложнения частных описаний, так и совершенно новые, не применявшиеся ранее.

Анализ других схем перебора. Поскольку данная структура охватывает то общее, что имеется в ряде казавшихся ранее различными основных алгоритмов самоорганизации, ее целесообразно назвать *основной структурой перебора моделей*. При этом появляется возможность глубже анализировать те многие алгоритмы МГУА, генераторы моделей которых не укладываются в основную структуру, а также создавать (или даже проектировать) принципиально новые структуры. Например, существенно другим является алгоритм МГУА для выделения гармонических трендов [22] — предложенный алгоритм не будет имитировать работу [22] даже после перехода к тригонометрическому базису.

В основную структуру не укладываются также схемы перебора в алгоритмах с ортогонализацией частных описаний [170] и с «передачей списков аргументов» [139]. Они весьма различные, но имеют существенный общий элемент: на каждом ряде аргументами являются не функции (частные описания), отобранные на предыдущем ряде, а сами исходные аргументы. При этом в алгоритме, предложенном в работе [170], на каждом ряде частное описание формируется добавлением одного аргумента, а в [139] — также, как и в обычных алгоритмах, но с вычислением оценок коэффициентов для «развернутого» описания, т. е. в [170] каждая модель содержит r членов (исходных аргументов), а в [139] — 2^r . Очевидно, что приведенную выше основную структуру можно модифицировать, обеспечив в каждом частном описании возврат к исходному базису. Тогда при $s_{\min}=s_{\max}=1$ получим некоторый аналог [170], а при $s_{\min}=s_{\max}=2$ — аналог [139], а также появляются другие возможности наращивания сложности моделей.

Отметим, что описанный в предыдущем параграфе селекционно-комбинаторный алгоритм в этой схеме можно получить, обеспечив на каждом ряде протекцию членам моделей предыдущего ряда ($\rho=r-1$) и добавление только одного аргумента из исходных ($s_{\min}=s_{\max}=1$).

Таким образом, указанные выше свойства двоичного регистра сдвига, являющегося существенным элементом комбинаторного алгоритма, предоставляют новые возможности для анализа и синтеза вычислительных схем генераторов частных моделей. Многие комбинаторные, многорядные и смешанные алгоритмы МГУА не являются принципиально различными в смысле организации перебора. Их генераторы моделей естественным образом обобщаются в некоторую основную вычислительную структуру, которую можно реализовать на основе оптимального комбинаторного алгоритма. Возможно, практическая реализация такого универсального алгоритма не является целесообразной — специализированные алгоритмы, как правило, более эффективны (например, по быстродействию). Однако рассмотренная обобщенная структура перебора моделей дает возможность однозначного сравнения существующих генераторов моделей, более глубокого понимания сходства и различия известных, а также целенаправленного конструирования новых схем перебора в алгоритмах МГУА.

Методы самоорганизации моделей на ЭВМ являются структурно-параметрическими и предназначены для решения задачи отыскания модели оптимальной сложности, соответствующей минимуму заданного внешнего критерия, которую целесообразно называть задачей структурного моделирования [49, 176].

Задача структурного моделирования решается с помощью различных алгоритмов МГУА, которые по способу организации вычислений разделяются на три основные группы: комбинаторные (однорядные), многорядные и смешанные. Все они имеют характерные особенности, которые определяют области их эффективного применения. Для задач моделирования с числом оцениваемых коэффициентов не более 20 несомненное преимущество имеют комбинаторные алгоритмы, поскольку они осуществляют перебор и сравнение всех моделей различной сложности. Кроме того, идея полного перебора весьма конструктивна для аналитического исследования свойств методов и критериев самоорганизации.

В данной главе подробно описаны общая структура комбинаторного алгоритма МГУА и особенности практической организации вычислений в основных блоках алгоритма: преобразование исходных данных; генерирование частных моделей различной сложности; вычисление значений внешних критериев и выбор оптимальной модели. Показаны способы оптимизации вычислений с помощью применения ковариационных матриц (матриц нормальных уравнений) и метода окаймления (для рекуррентного оценивания коэффициентов моделей). Описана основная структура перебора моделей, органично объединяющая комбинаторные, многорядные и смешанные схемы перебора.

Эффективность, или осуществимость (в терминах системологии [160]), комбинаторных алгоритмов МГУА существенно повышает алгоритм с селекционно-комбинаторной схемой перебора моделей, рассмотренный в данной главе. Это первый алгоритм МГУА, решающий задачу полного перебора, но с вычислительными затратами, свойственными многорядным (итерационным) алгоритмам.

Следует отметить, что в связи с простотой и однозначностью идеи полного перебора комбинаторные алгоритмы МГУА не являются столь разнообразными, как многорядные, однако спектр решаемых с их помощью задач достаточно широк. Например, алгоритм COMBI применяется для моделирования статических [73, 166] и динамических [88, 89] объектов, в том числе с распределенными параметрами [49]. Кроме того, он является необходимой составной частью комплексных алгоритмов объективного системного анализа [70, 71] сложных систем и двухуровневого долгосрочного прогнозирования [85, 87, 75, 82] циклических процессов.

Результаты последующих глав книги в основном относятся к исследованию помехоустойчивости комбинаторных алгоритмов МГУА, поскольку получены с применением COMBI или с использованием идеи полного перебора моделей.

ПОТЕНЦИАЛЬНАЯ ПОМЕХОУСТОЙЧИВОСТЬ МОДЕЛИРОВАНИЯ ПО КОМБИНАТОРНОМУ АЛГОРИТМУ МГУА

1. ЗАДАЧА ОЦЕНКИ ПОМЕХОУСТОЙЧИВОСТИ МОДЕЛИРОВАНИЯ

Целью моделирования является прежде всего как можно более точное воспроизведение, экстраполяция либо прогнозирование выхода изучаемого объекта или процесса. Для достижения этого необходимы помехоустойчивые методы моделирования, поскольку в данных наблюдений за поведением объекта практически всегда присутствуют помехи различной природы: шумы измерений, случайные неизмеряемые возмущения, отсутствие в данных какой-либо существенной переменной или, наоборот, участие в моделировании «лишней» переменной, теоретически не воздействующей на объект, но из-за ограниченности выборки имеющей значимую корреляцию с выходом или другими переменными, и т. д.

Понятие помехоустойчивости в моделировании можно определить так же, как и в теории связи: это способность достаточно точно восстанавливать переданный сигнал по зашумленному принятому сигналу, т. е. моделировать выход изучаемого объекта по данным с помехами. При этом важно знать, как изменяется качество (точность) моделирования при увеличении мощности шума или при увеличении отношения мощности шума к мощности сигнала (или просто отношения шум/сигнал).

Помехоустойчивость алгоритмов самоорганизации изучалась в основном с помощью вычислительных экспериментов (см., например, [2, 54], а также следующую главу): задается модель, по которой строится некоторая выборка (таблица) данных изменения входных и выходной переменных; затем данные зашумляются известным аддитивным шумом возрастающей дисперсии (мощности) и находится отношение шум/сигнал, при котором алгоритм, работающий с заданным критерием селекции, «отказывает», т. е. строит модели неудовлетворительной точности или теряет истинную структуру модели.

Задача аналитической оценки помехоустойчивости самоорганизации моделей по МГУА впервые была рассмотрена в [54], где исследовался так называемый абсолютно помехоустойчивый критерий, обладающий свойствами при любых шумах надежно отсеивать пере усложненные модели. Было показано, что при некотором уровне (дисперсии) шума этот критерий «теряет» истинную модель

и начинает выбирать более простую. Однако в этом исследовании не рассматривался вопрос о точности моделирования при различных уровнях шума.

Новые возможности исследования и повышения помехоустойчивости алгоритмов МГУА связаны, как показано в гл. 2, с глубокой аналогией задач теорий информации, связи и самоорганизации. При этом привлекают прежде всего полученные в рамках общей теории связи количественные оценки предельно достижимых возможностей передачи информации в условиях помех. Так, вторая теорема Шеннона [172] указывает предельную пропускную способность канала с шумом, или максимальное число различных сигналов с заданной полосой частот при известных мощностях сигнала и шума. В теории оптимального приемника Котельникова [97] показано, что наилучшим (идеальным) является приемник, работающий на принципе сравнения принятого (зашумленного) сигнала с набором эталонных (незашумленных) сигналов, один из которых передан передатчиком. Никакие реальные устройства не могут превзойти ограничения, указанные теориями Котельникова и Шеннона. Идеальный приемник сигналов, работающий на основе минимизации обсуждавшегося во второй главе квадратичного критерия, называется *потенциально помехоустойчивым* [97, 112].

Отметим, что понятие потенциальной, т. е. предельно достижимой, помехоустойчивости является в данном случае аналогом более широкого понятия *потенциальной эффективности*, рассматриваемого в системологии [160]. Потенциальной эффективностью некоторой системы является способность достичь, например, максимального числа целей при фиксированных затратах или, наоборот, достичь заданной цели при минимальных затратах. В моделировании, таким образом, потенциальной эффективностью алгоритма следует считать его способность обеспечить максимально достижимую точность модели при заданном уровне шума. Поэтому оценки потенциальной эффективности и помехоустойчивости совпадают.

В настоящей главе получены теоретические оценки потенциальной помехоустойчивости моделирования по комбинаторному алгоритму МГУА, не зависящие от конкретных применяемых критериев селекции. При этом для выяснения предельных возможностей моделирования в условиях помех эффективно использована аналогия с обсуждавшимся в гл. 2 идеальным приемником. Однако само исследование выполнено полностью в рамках методов и средств теории самоорганизации моделей.

2. ОСОБЕННОСТИ МОДЕЛИРОВАНИЯ В УСЛОВИЯХ ПОМЕХ

Пусть некоторый объект описывается точной (истинной) функциональной зависимостью вида

$$u = b^T x^0, \quad x^0 [s_0 \times 1], \quad (4.1)$$

а на его выход u наложен некоторый шум ξ с дисперсией σ , которая может изменяться:

$$y = u + \xi = u + \sigma \xi, \quad (4.2)$$

где ξ — вообще говоря, произвольный шум с постоянной дисперсией (например, единичной).

Рассмотрим задачу самоорганизации модели объекта (4.1) с помощью комбинаторного алгоритма МГУА по измерениям зашумленного выхода (4.2) и расширенного вектора входов $x [n \times 1]$, включающего как часть точный вектор x^0 , $n > s_0$. Информацию об N измерениях представим в виде системы

$$y = Xa, \quad y [N \times 1], \quad X [N \times n], \quad N > n, \quad (4.3)$$

где a — неизвестный вектор коэффициентов, подлежащий оптимизации.

При различных интенсивностях (дисперсиях) σ шума результаты моделирования могут быть разными. Для определенности интенсивность шума измеряем в единицах энергии (мощности) «точного сигнала» (вектора незашумленных измерений) $u [N \times 1]$:

$$\sigma = \theta c, \quad c^2 \stackrel{\Delta}{=} (u - \bar{u})^T (u - \bar{u}); \quad (4.4)$$

тогда $\theta = \frac{\sigma}{c}$ является уровнем шума, или отношением шум/сигнал. Предметом рассмотрения и является изучение влияния уровня шума на результаты самоорганизации модели.

Для упрощения обозначений считаем, что вся последовательность N измерений разделена на две части A и B длины N_A и N_B ($N_A + N_B = N$), а оценки коэффициентов каждой i -й частной модели определяются на A (при условии $N_A > n$):

$$\hat{a}_A(i) \stackrel{\Delta}{=} \hat{a}(i) = (X_A^T(i) X_A(i))^{-1} X_A^T(i) y_A = X_A^+(i) y_A, \quad (4.5)$$

где каждое из $i = 1, 2, \dots, 2^n - 1$ отражает структуру частной модели, причем здесь и далее предполагаем существование обратной матрицы, т. е. $\text{rank}(X_A^T(i) X_A(i))$ всегда полный. С учетом (4.2) эту оценку можно представить в виде суммы двух оценок:

$$\hat{a}(i) = X_A^+(i) u_A + \theta c X_A^+(i) \xi_A \stackrel{\Delta}{=} \hat{b}(i) + \theta c \hat{a}_{\xi}(i), \quad (4.6)$$

где $\hat{b}(i)$ — оценка коэффициентов частной модели по точному вектору измерений u_A ; $\hat{a}_{\xi}(i)$ — случайная добавка, определяющая смещение оценки $\hat{a}(i)$ и учитывающая реализацию шума ξ .

Тогда оценка выхода по i -й модели имеет вид

$$\hat{y}_G(i) = X_G(i) \hat{a}(i) = X_G(i) X_A^+(i) y_A \stackrel{\Delta}{=} P_G(i) y_A, \quad (4.7)$$

причем она записана для произвольной части данных G , которая может совпадать с A , B , $A \cup B$ или соответствовать дополнительной

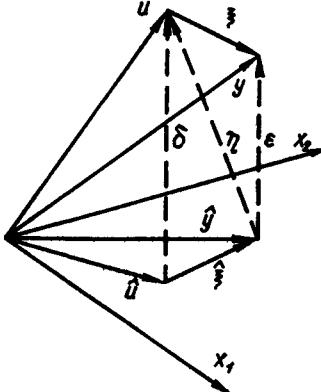


Рис. 13. Геометрическая интерпретация МНК в условиях помех (η — вектор теоретической ошибки модели).

\hat{y} и \hat{u} соответствуют применению МНК, причем векторы ошибок проектирования равны: для точного вектора u

$$\delta_i \stackrel{\Delta}{=} \delta_A(i) = u_A - \hat{u}_A(i) = (E - P_A(i)) u_A \stackrel{\Delta}{=} D_A(i) u_A,$$

для зашумленного вектора y

$$\varepsilon_i \stackrel{\Delta}{=} \varepsilon_A(i) = y_A - \hat{y}_A(i) = D_A(i) y_A.$$

Эти векторы ошибок всегда параллельны, причем связаны между собой через ошибку проектирования вектора шума ξ :

$$\varepsilon_i = \delta_i + (\xi - \hat{\xi}_A(i)) = \delta_i + D_A(i) \xi_A.$$

Однако успех моделирования в соответствии со сформулированной в начале главы целью определяется длиной вектора ошибки восстановления точного сигнала по зашумленному:

$$\eta_i \stackrel{\Delta}{=} \eta_A(i) = u_A - \hat{y}_A(i) = u_A - P_A(i)(u_A - \xi_A) = D_A(i) u_A - P_A(i) \xi_A.$$

Его следует назвать *теоретическим* вектором ошибки i -й модели. Он связан с векторами δ_i и ε_i очевидными соотношениями:

$$\eta_i = \varepsilon_i - \xi_A = \delta_i - \hat{\xi}_A(i).$$

Отметим, что точной модели соответствует случай, когда вектор u лежит в подпространстве векторов аргументов $x_1^0, \dots, x_{s_0}^0$. При этом $\delta(s_0)=0$ и η_i находится в этом же подпространстве, однако $\eta_i \neq 0$ и $\varepsilon_i \neq 0$, кроме тривиального случая ортогональности $\xi \perp u$. Очевидно, что все возможные значения случайного вектора всегда образуют некоторую сферу вокруг вектора u .

экзаменационной выборке C . С учетом (4.6) оценка (4.7) примет вид

$$\begin{aligned} \hat{y}_G(i) &= P_G(i) u_A + \theta c P_G(i) \xi_A \stackrel{\Delta}{=} \hat{u}_G(i) + \\ &+ \theta c \hat{\xi}_G(i), \end{aligned} \quad (4.8)$$

где $\hat{u}_G(i)$ — детерминированная проекция точного вектора u_A по i -й модели; $\hat{\xi}_G(i)$ — случайная добавка, учитывающая шум.

Рассмотрим геометрическую интерпретацию этих оценок для случая $\sigma=1$, $G=A$, т. е. на обучающей последовательности (рис. 13). Все три вектора u , ξ и y проектируются ортогонально на линейное подпространство, образованное векторами аргументов частной модели (в данном случае x_1, x_2). Их проекции \hat{u} , $\hat{\xi}$ и \hat{y} соответствуют применению МНК, причем векторы ошибок проектирования равны: для точного вектора u

В общем случае (для произвольных σ и G) теоретический вектор ошибки i -й модели имеет вид

$$\eta_G(i) = u_G - \hat{y}_G(i) = u_G - \hat{u}_G(i) - \theta c \hat{\xi}_G(i). \quad (4.9)$$

Он состоит из двух компонент: детерминированного вектора ошибки проектирования, вызванного неточностью структуры частной модели

$$\delta_G(i) = u_G - \hat{u}_G(i) = u_G - P_G(i) u_A, \quad (4.10)$$

и случайного вектора $\theta c \hat{\xi}_G(i)$, определяемого реализацией и уровнем шума.

Приведем еще раз некоторые из принятых обозначений:

$$X_A^+ (i) = (X_A^T(i) X_A(i))^{-1} X_A^T(i),$$

$$P_G(i) = X_G(i) X_A^+(i),$$

$$D_A(i) = E - P_A(i) = E - X_A(i) X_A^+(i),$$

где $X_A^+(i)$ — псевдообратная матрица; $P_G(i)$ — обобщенная матрица проектирования; $D_A(i)$ — матрица минимизации расстояния до подпространства, образованного столбцами $X_A(i)$. При этом $P_G(i)$ задает, вообще говоря, неортогональную проекцию — ортогональной она будет только в случае $G = A$.

3. ИДЕАЛЬНЫЙ ВНЕШНИЙ КРИТЕРИЙ

Предположим, что комбинаторный алгоритм отыскивает оптимальную модель, соответствующую минимуму следующего идеального внешнего критерия:

$$Q_G^2(i) = \frac{1}{c_G^2} (u_G - \hat{y}_G(i))^T (u_G - \hat{y}_G(i)), \quad (4.11)$$

который оценивает относительную погрешность моделирования (прогнозирования) незашумленных данных (аналогичный критерий применялся в [86] для численного изучения помехоустойчивости практических критериев самоорганизации). Знаменатель c_G^2 введен для того, чтобы величина критерия была легко интерпретируемой.

Критерий (4.11) является внешним даже на A , так как вектор u_A не используется непосредственно при вычислении оценок $\hat{a}_A(i)$ и $\hat{y}_A(i)$, что видно из (4.5) и (4.7). Он является идеальным, т. е. теоретически наилучшим, поскольку минимизирует теоретический вектор ошибки модели (4.9). Это следует из самого определения критерия: при переборе из всех $p_n=2^n-1$ моделей ни по какому другому из возможных критериев нельзя выбрать модель, лучше приближающую точные (незашумленные) данные, чем та, которая

соответствует минимуму $Q_G^2(i)$. С другой стороны, если известен точный вектор u_G , то качество (точность) модели, выбранной по любому другому критерию селекции, можно оценить по величине критерия (4.11), т. е. по степени приближения к минимуму идеального критерия. Таким образом, (4.11) является предельным случаем внешнего критерия, что и позволяет решить задачу оценки потенциальной помехоустойчивости моделирования по комбинаторному алгоритму МГУА.

С учетом (4.9), (4.10) выражение (4.11) можно представить так:

$$Q_G^2(i) = \frac{1}{c_G^2} \eta_G^\top(i) \eta_G(i) = \frac{1}{c_G^2} (\delta_G(i) - \theta c \hat{\xi}_G(i))^\top (\delta_G(i) - \theta c \hat{\xi}_G(i)), \quad (4.12)$$

или в развернутом виде

$$Q_G^2(i) = \frac{1}{c_G^2} (\delta_G^\top(i) \delta_G(i) - 2\theta c \delta_G^\top(i) \hat{\xi}_G(i) + \theta^2 c^2 \hat{\xi}_G^\top(i) \hat{\xi}_G(i)). \quad (4.13)$$

Все частные модели, генерируемые комбинаторным алгоритмом, целесообразно разделить на несколько групп: 1) *полная* (правильная) — одна, содержащая только истинные аргументы x^0 ; 2) *неточные*: а) *неполные* — содержащие только истинные аргументы, но не все; б) *избыточные* — содержащие все истинные аргументы, а также другие, не совпадающие с x^0 ; в) *смешанные* — как из истинных (но не всех), так и из других аргументов; 3) *неправильные* — без истинных аргументов.

Для всех этих моделей понятие «сложность» включает как число s аргументов, так и их состав: в алгоритме при каждом значении s сравнивается C_n^s моделей, $s = \overline{1, n}$. Однако это понятие без ограничения общности можно упростить, предположив, что при каждом s , т. е. для любого числа аргументов, учитывается только наилучшая по критерию (4.9) модель. Тогда идеальный внешний критерий примет вид

$$R_G^2(s) = \min_{i=1, C_n^s} Q_G^2(i) = \frac{1}{c_G^2} [\delta_G^\top(s) \delta_G(s) - 2\theta c \delta_G^\top(s) \hat{\xi}_G(s) + \theta^2 c^2 \hat{\xi}_G^\top(s) \hat{\xi}_G(s)]. \quad (4.14)$$

Очевидно, что $\min_{s=\overline{1, n}} R_G^2(s) = \min_{i=\overline{1, p_n}} Q_G^2(i)$, так что рассматривать только $R_G^2(s)$ вполне правомерно.

Естественно предположить, что все избыточные (лишние) аргументы менее коррелированы с точным вектором u (и с вектором \hat{y}) по сравнению с истинными аргументами, т. е. $x_j^{0\top} u \gg x_l^\top u$, $j = \overline{1, s_0}$, $l = \overline{1, n}$, $j \neq l$. Тогда при каждом s лучшей по критерию (4.9) всегда будет неполная, полная или избыточная модель и именно такие модели будут сравниваться по критерию (4.14). При этом сложность модели однозначно соответствует числу аргументов: при $s < s_0$ — модели неполные, при $s > s_0$ — избыточные, а при $s = s_0$ — модель полная.

Тогда модель оптимальной сложности определяется как модель с числом аргументов

$$s^* = \arg \min_{s=\overline{1, n}} R_G^2(s). \quad (4.15)$$

Понятие модели оптимальной сложности является фундаментальным в теории самоорганизации. В рассматриваемом случае идеального алгоритма оно соответствует модели, лучше других приближающей точные данные при оценивании параметров по зашумленным данным. Частным случаем задачи поиска модели оптимальной сложности по зашумленным данным является традиционная задача структурной идентификации, т. е. восстановления модели с правильной структурой.

Отметим еще некоторые аналогии данного исследования с общей теорией связи. Вектор точных измерений и эквивалентен в пространстве измерений точному сигналу, вектор измерений y — сигналу с шумом. Тогда задача минимизации $R_G^2(s)$ зеркально отражает задачу, решаемую идеальным приемником Котельникова: там один зашумленный сигнал сравнивается с n эталонными (точными), здесь — n «зашумленных сигналов» (выходов n частных моделей) сравниваются с одним эталонным (точным). Как видим, эта «зеркальная аналогия» содержит принципиальное различие, которое можно установить также формально: критерий Котельникова легко выводится с помощью принципа максимального правдоподобия [97] (т. е. его минимум соответствует максимуму вероятности восстановления точного сигнала), а критерий $R_G^2(s)$ — нет, даже при $G=A$.

Действительно, попытаемся установить условия, при которых вероятность восстановления точного вектора u_A максимальна при переборе n структур:

$$f(y_A^* | u_A) = \max_{s=\overline{1, n}} f(\hat{y}_A(s) | u_A).$$

Функция правдоподобия ввиду детерминированности u_A определяется «шумом» частной модели $\hat{y}_A(s)$, которым является вектор ошибки $\eta(s)$:

$$f(\hat{y}_A(s) | u_A) = f_1(\eta(s)).$$

Здесь $f_1(\eta(s))$ — функция распределения помехи $\eta(s)$. Однако сама помеха зависит от номера s (от структуры) частной модели, т. е. от s зависит также ее среднее η_s и, что существенно, дисперсия σ_s^2 (она вычисляется ниже). Поэтому максимум функции правдоподобия

$$f(\hat{y}_A(s) | u_A) = \frac{1}{(2\pi\sigma_s^2)^{N/2}} \exp \left(-\frac{1}{2\sigma_s^2} \sum_{i=1}^N \eta_i^2(s) \right)$$

не обеспечивается при достижении минимума суммы в показателе экспоненты, так как из $\partial f/\partial s = 0$ нельзя получить $\frac{\partial}{\partial s} \sum_{j=1}^N \eta_j^2(s) = 0$.

Однако указанное обстоятельство вовсе не означает, что критерий $R_G^2(s)$ неэффективен: это свидетельствует о том, что вероятность восстановления модели с истинной структурой не является максимальной, т. е. можно предположить, что при повышении уровня шума более эффективной может оказаться модель с другой структурой. Как мы уже установили, минимуму критерия (4.14) соответствует модель оптимальной сложности с наименьшей ошибкой восстановления точного сигнала u_G , а это и есть главная цель моделирования.

Итак, основные предположения о работе идеального комбинаторного алгоритма следующие: 1) известны точные значения выхода u и истинной модели и реализация шума ξ (или его характеристики); 2) оценки коэффициентов частных моделей строятся по зашумленному выходу y , но могут быть разделены на детерминированную $\hat{b}(s)$ и случайную $\theta c \hat{a}_G(s)$ части; 3) уровень шума $\theta \geq 0$ регулируется; 4) модель оптимальной сложности выбирается по условию минимума идеального критерия селекции (4.14).

В этих условиях исследуются два основных вопроса: 1) изменяется ли выбор оптимальной модели при возрастании уровня шума θ от 0 до ∞ ? 2) существует ли для каждой модели сложности s предельное значение уровня шума, при котором точность восстановления (прогнозирования) незашумленных данных еще остается удовлетворительной?

Ответ на эти вопросы и составляет цель данного исследования помехоустойчивости моделирования по комбинаторному алгоритму МГУА. Причем результаты этого исследования будут иметь достаточно общий характер, что определяется постановкой рассматриваемой в настоящей главе задачи оценки эффективности моделирования по экспериментальным данным. Сформулированные выше вопросы весьма важны не только для МГУА, но и для других методов структурной идентификации линейных моделей типа методов построения лучшей регрессии.

4. СЕЛЕКТИВНЫЕ СВОЙСТВА ИДЕАЛЬНОГО ВНЕШНЕГО КРИТЕРИЯ

Исследуем особенности выбора модели оптимальной сложности по идеальному критерию (4.14) для двух случаев: при нормальном шуме и когда задана лишь отдельная реализация произвольного шума.

Свойства математического ожидания критерия при нормальном шуме. Для случая шума с $M\xi_i=0$, $M\xi_i^2=1$ рассмотрим свойства критерия (4.14) «в среднем», т. е. исследуем его математическое ожидание:

$$r_G^2(s) = MR_G^2(s) = \frac{1}{c_G^2} M\eta_G^T(s)\eta_G(s) = \frac{1}{c_G^2} [\delta_G^T(s)\delta_G(s) + \theta^2 c^2 \text{tr}(P_G^T(s)P_G(s))]. \quad (4.16)$$

При получении (4.16) учтено, что $\delta_G(s)$ — детерминированный вектор, $M\xi_G(s) = P_G(s)M\xi_A = 0$ и что

$$M\xi_G^T(s)\hat{\xi}_G(s) = M\xi_A^T P_G^T(s)P_G(s)\xi_A = \text{tr}(P_G^T(s)P_G(s)),$$

поскольку для случайноговектора ξ с нормальным распределением компонент справедливо [32] $M\xi^T H \xi = \text{tr} H$, где $\text{tr}(H)$ — след (сумма диагональных элементов) квадратной матрицы H .

Фигурирующий в (4.16) след произведения матриц при $G=A$ равен просто s [32], т. е. числу аргументов в частной модели, а при $G \neq A$ может быть выражен через матрицы $X_A(s)$ и $X_G(s)$ в более простом виде (см. приложение 1 данной главы):

$$\text{tr}(P_G^T(s)P_G(s)) = \begin{cases} \text{rank } P_A(s) = s, & G = A, \\ \text{tr}[(X_A^T(s)X_A(s))^{-1}X_G^T(s)X_G(s)], & G \neq A. \end{cases} \quad (4.17)$$

Рассмотрим подробнее свойства критерия (4.16) на последовательности A , указав, что он зависит также от уровня шума θ (с учетом (4.17)):

$$r_A^2(\theta, s) = \frac{\delta_A^2(s)}{c_A^2} + \frac{c^2}{c_A^2} \theta^2 s = r_1^2(s) + r_2^2(\theta, s). \quad (4.18)$$

Функция $r_A^2(\theta, s)$ равна сумме двух функций: первая $r_1^2(s)$ зависит от дискретного аргумента s нелинейно и неявно, вторая $r_2^2(\theta, s)$ — линейно.

Основные свойства критерия $r_A^2(\theta, s)$ сформулируем в виде следующих утверждений.

1º. Минимум критерия (4.18) существует и единствен, причем $r_A^2(\theta, s)$ является *унимодальной* функцией дискретного аргумента s (сложности модели).

Исследуем составляющую $\delta_A^2(s)$, которая в данном случае равна квадратической ошибке оценки u_A по МНК:

$$\delta_A^2(s) = \delta_A^T(s)\delta_A(s) = u_A^T D_A(s) u_A \stackrel{\Delta}{=} u_A^T D_s u_A, \quad (4.19)$$

причем для любого s $D_s^2 = D_s$, т. е. $D_s > 0$. Установим, как изменяется эта ошибка при увеличении сложности s на единицу, т. е. при добавлении к s -й частной модели одного аргумента. С помощью формул, полученных в гл. 3 методом окаймления, можно вывести (см. приложение 2)

$$\delta_A^2(s+1) = \delta_A^2(s) - \frac{(x_A^T D_s u_A)^2}{x_A^T D_s x_A} \stackrel{\Delta}{=} \delta_A^2(s) - \gamma_A^2(s+1). \quad (4.20)$$

Здесь x_A — вектор измерений $(s+1)$ -го аргумента; $\gamma_A^2(s+1)$ — приращение ошибки за счет добавления аргумента.

Так как $D_s \stackrel{\Delta}{=} D_A(s)$ — положительно определенная матрица, то $\gamma_A^2(s+1) > 0$, т. е. ошибка МНК $\delta_A^2(s)$ монотонно убывает с рос-

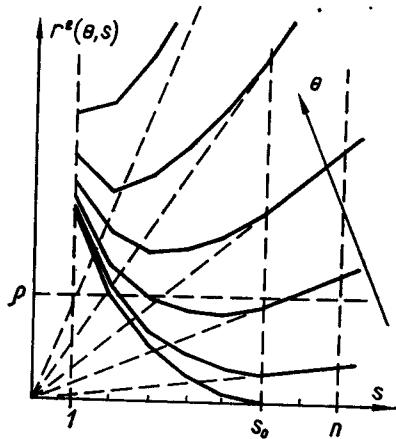


Рис. 14. Изменение выбора модели оптимальной сложности по минимуму идеального критерия $r^2(\theta, s)$ при возрастании уровня шума θ .

$r_1^2(s)$ характеризует изменение ошибки МНК модели, построенной по точным данным с помощью последовательного добавления истинных аргументов так, что при каждом $1 \leq s \leq s_0$ выбирается аргумент, дающий *наибольшее уменьшение ошибки*. Поэтому функция $r_1^2(s)$ является выпуклой вниз. Отметим еще, что максимальное значение может быть равно единице и достигается оно при выборе первой модели в виде свободного члена $b_0 = u_A$.

Функция $r_2^2(\theta, s)$ линейно возрастает по s с параметром θ . Поэтому при любом θ идеальный критерий (4.18) является строго унимодальной функцией сложности модели s (рис. 14).

2⁰. При достаточно малом уровне шума θ минимум критерия (4.18) соответствует истинной модели; при возрастании θ минимум смещается в область более простых моделей.

Поскольку $\delta_A^2(s|s>s_0)=0$, то очевидно, что минимум критерия возможен лишь при $s < s_0$, т. е. избыточные модели всегда хуже неполных и надежно отсеиваются. Из рис. 14 видно, что при увеличении θ минимум смещается влево. Это можно доказать строже.

Рассмотрим произвольную модель сложности s и запишем с учетом (4.18), (4.20) приращение (первую разность) $r_A^2(\theta, s)$ по s :

$$\Delta r_A^2(\theta, s) = r_A^2(\theta, s+1) - r_A^2(\theta, s) = \frac{1}{c_A^2} (\theta^2 c^2 - \gamma_A^2(s+1)). \quad (4.21)$$

Так как $\gamma_A^2(s+1)$ от θ не зависит, то это приращение в каждой точке s при малых θ отрицательно, при больших — положительно, т. е. с ростом θ экстремум (минимум) переходит из $(s+1)$ -й точки в s -ю (смещается влево) при $\theta > \theta_{kp}(s)$, определяе-

мом из условия экстремума $\Delta r_A^2(\theta, s) = 0$:

$$\theta_{kp}(s) = \frac{1}{c^2} \gamma_A^2(s+1). \quad (4.22)$$

3⁰. Геометрическое место точек минимумов критерия (4.18) при различных θ описывается функцией

$$r_{\min}^2(s) = r_A^2(\theta_{kp}, s) = r_1^2(s) + \frac{s}{c_A^2} \gamma_A^2(s+1). \quad (4.23)$$

Эта функция получена подстановкой (4.22) в (4.18) и является в силу выпуклости вниз $\delta_A^2(s)$ монотонно убывающей по s .

4⁰. Модель оптимальной сложности, соответствующая минимуму критерия $r_A^2(\theta, s)$, обладает наименьшей дисперсией вектора ошибки восстановления точных данных и определяется при заданном θ выражением

$$s_0^* = \begin{cases} s_\theta, & c^2 \theta^2 \leq \gamma_A^2(s+1), \\ s_\theta + 1, & c^2 \theta^2 > \gamma_A^2(s+1), \end{cases} \quad (4.24)$$

$$s_\theta = \arg \min_{s=1, s_0} |c^2 \theta^2 - \gamma_A^2(s+1)|.$$

Это условие эквивалентно (4.15), но более конструктивно: увеличивая s , отыскиваем s_0^* , при котором приращение $\Delta r_A^2(\theta, s_0^*)$ из отрицательного станет положительным, причем для этого достаточно рассматривать только одну переменную величину $\gamma_A^2(s+1)$. Алгоритм (4.24) получен на основе (4.22) с учетом дискретности изменения сложности моделей s .

Существенно подчеркнуть, что сам критерий $r_A^2(\theta, s)$ является теоретической дисперсией значений вектора ошибки $\eta_A(s)$, так что действительно модель оптимальной сложности соответствует минимуму дисперсии по определению. Более того, из (4.15), (4.16) следует, что этот факт имеет место в общем случае (на любой части G выборки данных).

5⁰. Модель с *точной структурой* будет моделью оптимальной сложности при уровнях шума, определяемых неравенством

$$0 \leq \theta^2 \leq \theta_0^2 = \gamma_A^2(s_0)/c^2. \quad (4.25)$$

Поскольку $\delta_A^2(s_0) = 0$, имеем $\gamma_A^2(s_0) = \delta_A^2(s_0-1)$. Собственно говоря, (4.25) является областью помехоустойчивости обычной задачи структурной идентификации. Это неравенство получено из (4.22) и определяется уровнем шума θ_0 , при котором происходит переход минимума критерия в точку (s_0-1) . Очевидно, что существование области помехоустойчивого выбора (4.25) вызвано дискретностью изменения сложности моделей: при непрерывном изменении сложности минимум критерия смещается влево при сколь угодно малом уровне шума.

6⁰. При $\theta > \theta_0$ модель оптимальной сложности s^* всегда является *неполной*, причем для нее существует область помехоустойчивого выбора, определяемая неравенством

$$\theta_{kp}^2(s^*) = \gamma_A^2(s^* + 1)/c^2 \leq \theta^2 < \gamma_A^2(s^*)/c^2 = \theta_{kp}^2(s^* - 1). \quad (4.26)$$

Это также некоторый «интервал нечувствительности» критерия, обусловленный дискретностью изменения сложности модели при непрерывном изменении уровня шума θ . Очевидно, что с ростом θ оптимальная модель будет иметь все меньше аргументов, а при $\theta^2 > \theta_{kp}^2(1) = \gamma_A^2(2)/c^2$ лучшей будет модель с одним аргументом (*самым существенным*).

Интересно отметить, что в некоторых задачах значения критических уровней шума могут оказаться близкими: $\theta_{kp}^2(s^*) \approx \theta_{kp}^2(s^* - 1)$. При этом соседние по сложности модели будут неразличимыми как в смысле выбора модели оптимальной сложности, так и в смысле помехоустойчивости. Этот случай «растянутого» минимума внешнего критерия иногда встречается в практических задачах. Из данного анализа следует, что причиной этого может быть, в частности, большой уровень шума или близость двух аргументов по степени коррелированности с выходом (последнее определяется значениями $\gamma_A(s^*)$ и $\gamma_A(s^* - 1)$).

7⁰. Модели со сложностью выше оптимальной, т. е. с $s > s^*$, однозначно отсеиваются критерием (4.18) при любом уровне шума θ , причем надежность отсева возрастает с ростом θ .

Для доказательства рассмотрим вопрос об избирательности критерия $r_A^2(\theta, s)$: уменьшается ли надежность определения минимума (глубина минимума) при увеличении уровня шума? Для этого получим приращение первой разности (4.21) критерия при увеличении θ , точнее, при двух значениях уровня θ_1 и θ_2 ($\theta_2 > \theta_1$):

$$\Delta r_A^2(\theta_2, s) - \Delta r_A^2(\theta_1, s) = \frac{c^2}{c_A^2} (\theta_2^2 - \theta_1^2) > 0. \quad (4.27)$$

Эта разность положительна, однако на левой ветви функции $r_A^2(\theta, s)$ приращения Δr_A^2 отрицательны, на правой (возрастающей) — положительны. Следовательно, при $s < s_0$ имеем $|\Delta r_A^2(\theta_1, s)| > |\Delta r_A^2(\theta_2, s)|$, а при $s > s_0$ получаем $\Delta r_A^2(\theta_2, s) > \Delta r_A^2(\theta_1, s)$, т. е. при росте θ крутизна критериальных функций слева от экстремума уменьшается, справа — увеличивается (см. рис. 14).

Это означает, что при любом θ надежно отсеиваются не только избыточные модели, но и все модели со сложностью больше оптимальной, причем этот эффект возрастает при увеличении θ .

Доказанные утверждения относятся к свойствам идеального критерия, определенного на последовательности A выборки наблюдений, т. е. там же, где оцениваются коэффициенты $\hat{a}_A(s)$. Справедливы ли они также на последовательности B ? Для ответа на этот вопрос необходимо рассмотреть поведение критерия (4.16) для $G=B$ при возрастании сложности s . Точнее, надо установить, как изменяются при этом значения $\delta_B^T(s) \delta_B(s)$ и $\text{tr}(P_B^T(s) P_B(s))$.

В приложении 3 к данной главе получено следующее рекуррентное соотношение для изменения ошибки на проверочной последовательности:

$$\delta_B^T(s+1) \delta_B(s+1) = \delta_B^T(s) \delta_B(s) - 2\hat{\beta}(u_B - X_B(s) \hat{b}(s))^T \times \quad (4.20a)$$

$$\times (x_B - X_B(s) X_A^\dagger x_A) + \hat{\beta}^2(x_B - X_B(s) X_A^\dagger(s) x_A)^T (x_B - X_B(s) X_A^\dagger(s) x_A).$$

Анализ этого выражения не позволяет сделать непосредственно вывод об уменьшении ошибки $\delta_B^T(s) \delta_B(s)$ при любом способе увеличения сложности частной модели, что справедливо для $\delta_A^T(s)$ (4.20). Однако, во-первых, эта ошибка положительна при всех $s < s_0$, в том числе, например, $\delta_B^T(1) \delta_B(1) > 0$ и $\delta_B^T(s_0 - 1) \delta_B(s_0 - 1) > 0$. Во-вторых, при $s = s_0$ имеем $\hat{b}(s_0) = b$ и при $s > s_0$ эта оценка сохраняется: $\hat{b}(s | s \geq s_0) = b$, поэтому для точной и избыточных структур незашумленные данные точно восстанавливаются также и на B , т. е. $\delta_B(s | s \geq s_0) = 0$. Это означает, что ошибка $\delta_B^T(s) \delta_B(s)$ на интервале $1 \leq s \leq s_0$ обязательно уменьшается до нуля, хотя в общем случае и не монотонно. Однако критерий (4.16) по построению является таким же, как и (4.14), т. е. в модель всегда при увеличении сложности вводится тот аргумент из числа истинных, который максимально уменьшает ошибку модели. При этом с необходимостью мы получаем монотонно уменьшающуюся ошибку $\delta_B^T(s) \delta_B(s)$, т. е. функция $r_{1B}^2(s)$ ведет себя аналогично $r_1^2(s)$ в (4.18).

Что касается функции $\text{tr}((X_A^\dagger X_A)^{-1} X_B^\dagger X_B)$, то она является строго монотонно возрастающей функцией сложности частной модели s . Для доказательства этого учтем, что матрицы $H_A(s) = X_A^\dagger(s) X_A(s)$ и $H_B(s) = X_B^\dagger(s) X_B(s)$ положительно определены при любом s . При этом матрица $L(s) = H_A^{-1}(s) H_B(s)$ имеет *положительные* собственные числа λ_i , $i = \overline{1, s}$, которые являются решением характеристического уравнения $\lambda H_A(s) - H_B(s) = 0$ ([118], теорема 2.15.1). Поэтому рассматриваемый след вычисляется по формуле

$$\text{tr}[H_A^{-1}(s) H_B(s)] = \sum_{i=1}^s \lambda_i(s).$$

Эта сумма положительных чисел возрастает при увеличении s , так что, действительно, функция

$$r_{2B}^2(\theta, s) = \theta c \text{tr}[(X_A^\dagger(s) X_A(s))^{-1} X_B^\dagger(s) X_B(s)]$$

является монотонно возрастающей по s , как и функция $r_2^2(\theta, s)$ в (4.18), т. е. минимум $r_B^2(\theta, s)$ существует и единствен, а с ростом θ смещается влево.

Таким образом, все доказанные выше утверждения, а также рис. 14 должны быть справедливы и для критерия $r_B^2(\theta, s)$. Однако,

строго говоря, это совпадение прежде всего качественное, поскольку точные оценки вида (4.22) — (4.26) получить не удается. Более того, минимумы функций $r_A^2(\theta, s)$ и $r_B^2(\theta, s)$ не обязательно должны совпадать при всех уровнях шума θ , т. е. выбор модели оптимальной сложности зависит и от части выборки, на которой определен критерий. По-видимому, можно сказать так: модель, соответствующая минимуму $r_A^2(\theta, s)$, больше подходит для решения задачи идентификации (фильтрации), а та, что выбрана по минимуму $r_B^2(\theta, s)$, — для прогнозирования (экстраполяции).

Завершая обсуждение свойств идеального критерия (4.16) «в среднем», отметим существенное обстоятельство: оценки коэффициентов модели точной структуры, полученные по МНК, являются теоретически *несмещанными*, т. е. $\hat{M}_A(s_0) = b^0$. Однако эта несмешанность, как мы уже установили, вовсе не гарантирует точного моделирования незашумленного вектора u : дисперсия оценок вектора ошибки моделирования, равная значению критерия $r_G^2(\theta, s_0)$, с ростом θ увеличивается, и после превышения некоторого уровня шума θ эта дисперсия оказывается меньшей у более простых, т. е. *смещенных* моделей. А эта дисперсия, т. е. $r_G^2(\theta, s)$, является мерой надежности (качества) моделирования. Таким образом, модели *оптимальной* сложности, несмотря на смешанность оценок коэффициентов, наиболее эффективно (надежно) оценивают незашумленный вектор u по зашумленному вектору измерений y . И этот вывод справедлив в смысле критерия (4.16), т. е. в среднем для всех возможных реализаций вектора шума ξ с нормальным распределением компонент. Можно ли сделать аналогичные выводы для каждой отдельно взятой реализации вектора шума?

Случай отдельной реализации произвольного шума. Рассмотрим поведение критерия $R_G^2(s)$ (4.14) при увеличении уровня шума θ , если известна только фиксированная реализация шума ξ , а его характеристики не заданы.

Подробно исследуем случай, когда G совпадает с A . При этом векторы ошибки $\delta_A(s)$ и проекции $\hat{\xi}_A(s)$ ортогональны, что следует из (4.8), (4.11):

$$\hat{\delta}_A^T(s)\hat{\xi}_A(s) = u_A^T D_s P_{ss} \hat{\xi}_A = u_A^T P_{ss} \xi - u_A^T P_{ss}^2 \xi_A = 0. \quad (4.28)$$

С учетом этого идеальный критерий примет вид

$$R_A^2(\theta, s) = \frac{1}{c_A^2} (\delta_A^2(s) + \theta^2 c^2 \hat{\xi}_A^T(s) \hat{\xi}_A(s)) \stackrel{\Delta}{=} r_1^2(s) + R_2^2(\theta, s), \quad (4.29)$$

т. е., как и в (4.18), он равен сумме двух функций — детерминированной $r_1^2(s)$ и случайной $R_2^2(\theta, s)$. Свойства $r_1^2(s)$ уже рассмотрены.

Определим характер изменения $R_2^2(\theta, s)$ при увеличении s , вычислив $R_2^2(\theta, s+1)$, точнее, $\hat{\xi}_A^T(s+1) \hat{\xi}_A(s+1)$. Для этого удобнее пользоваться формулами, записанными через оценки коэффициентов (4.5) — (4.7):

$$\hat{\xi}_A^T(s) \hat{\xi}_A(s) = \xi_A^T D_s \xi_A = \hat{a}_\xi^T(s) X_A^T(s) X_A(s) \hat{a}_\xi(s).$$

При увеличении s на единицу в матрице аргументов добавляется столбец: $X_A(s+1) = (X_A(s); x_A)$, а в векторе оценок коэффициентов — один элемент (см. приложение 2): $\hat{a}_\xi^T(s+1) = (\hat{a}_\xi^*(s); \alpha_\xi) = ((\hat{a}_\xi^*(s) - \alpha_\xi X_A^T(s) x_A)^T; \alpha_\xi)$. Поэтому с учетом $X_A^T(s) X_A(s) = E$ имеем

$$\begin{aligned} \hat{\xi}_A^T(s+1) \hat{\xi}_A(s+1) &= (\hat{a}_\xi^*(s); \alpha_\xi) \left(\frac{X_A^T(s)}{x_A^T} \right) (X_A(s); x_A) \left(\frac{\hat{a}_\xi^*(s)}{\alpha_\xi} \right) = \\ &= \hat{a}_\xi^*(s) X_A^T(s) X_A(s) \hat{a}_\xi^*(s) + 2\hat{a}_\xi^*(s) X_A^T(s) x_A \alpha_\xi + \alpha_\xi^2 x_A^T x_A = \\ &= \hat{\xi}_A^T(s) \hat{\xi}_A(s) + \alpha_\xi^2 (x_A^T x_A - x_A^T P_s x_A). \end{aligned}$$

Далее, так как $\alpha_\xi = x_A^T D_s \xi_A / x_A^T D_s x_A$ (см. приложение 2), получаем

$$\begin{aligned} \hat{\xi}_A^T(s+1) \hat{\xi}_A(s+1) &= \hat{\xi}_A^T(s) \hat{\xi}_A(s) + \\ &+ (x_A^T D_s \xi_A)^2 / (x_A^T D_s x_A) \stackrel{\Delta}{=} \hat{\xi}_A^T(s) \hat{\xi}_A(s) + \mu_A^2(s+1). \end{aligned} \quad (4.30)$$

Поскольку $D_s > 0$, выражение (4.30) доказывает, что функция $R_2^2(\theta, s)$ монотонно возрастает по s , т. е. последовательные проекции вектора помехи ξ_A на подпространства, образованные одним, двумя, тремя и т. д. векторами измерений аргументов (независимо от порядка следования), все время увеличиваются.

Таким образом, качественно поведение функций $r_2^2(\theta, s)$ в (4.18) и $R_2^2(\theta, s)$ в (4.29) аналогично и все выводы, полученные для $R_A^2(\theta, s)$ «в среднем», имеют место и в случае отдельно взятой реализации вектора шума ξ , и прежде всего *унимодальность* $R_A^2(\theta, s)$.

С учетом (4.30) несложно получить формулы, аналогичные выведенным для $r_A^2(\theta, s)$. Например, приращение критерия записывается выражением

$$\Delta R_A^2(\theta, s) = \frac{1}{c_A^2} [\theta^2 c^2 \mu_A^2(s+1) - \gamma_A^2(s+1)], \quad (4.31)$$

которое, как и (4.21), при малых θ отрицательно, при больших — положительно, что также означает смещение влево экстремума при росте θ . Модель сложности $s < s_0$ становится оптимальной при уровне шума

$$\theta_{kp}^2(s) = \frac{\gamma_A^2(s+1)}{c \mu_A^2(s+1)} = \frac{(x_A^T D_s u_A)^2}{c^2 (x_A^T D_s \xi_A)^2}. \quad (4.32)$$

С учетом (4.31), (4.32) легко написать выражения для определения областей помехоустойчивого выбора $\theta_{kp}(s+1) \leq \theta < \theta_{kp}(s)$, оптимальной сложности s_0^* и линии минимумов $R_{min}^2(s)$, аналогичные (4.23) — (4.26), а также доказать свойство отсева моделей со сложностью выше оптимальной (как в (4.27)).

Например, для определения сложности s_θ^* оптимальной модели справедлив алгоритм, аналогичный (4.24) и эквивалентный (4.15):

$$s_\theta^* = \begin{cases} s_\theta, & \Delta R_A^2(\theta, s) \leq 0, \\ s_\theta + 1, & \Delta R_A^2(\theta, s) > 0, \end{cases} \quad (4.24a)$$

$$s_\theta = \arg \min_{s=1, s_0} |\Delta R_A(\theta, s)|.$$

Область помехоустойчивого выбора, в частности, для модели точной структуры (сложности s_θ) определяется неравенством

$$0 \leq \theta^2 < \theta_0^2 = \gamma_A^2(s_0)/c^2\mu_A^2(s_0). \quad (4.25a)$$

Не составляет труда записать также аналог (4.26).

Геометрическое место точек минимумов критерия $R_A^2(\theta, s)$ описывается функцией дискретного аргумента s :

$$R_{\min}^2(s) = r_1^2(s) + \frac{\gamma_A^2(s+1)}{\mu_A^2(s+1)} \sum_{j=1}^s \mu_A^2(j), \quad (4.23a)$$

полученного подстановкой в (4.29) выражения (4.32) для $\theta_{kp}^2(s)$, а также решения рекуррентного уравнения (4.30) при $\hat{\xi}_A(0) = 0$:

$$\hat{\xi}_A^T(s) \hat{\xi}_A(s) = \sum_{j=1}^s \mu_A^2(j). \quad (4.33)$$

Таким образом, все без исключения утверждения, доказанные выше для математического ожидания $r_A^2(\theta, s)$ идеального критерия, как и рис. 14, полностью справедливы для $R_A^2(\theta, s)$ также в случае любой отдельно взятой случайной реализации вектора шума ξ , причем не нужны никакие дополнительные предположения о характере и свойствах этого шума.

Перенести все эти результаты на случай $G = B$ не удается, поскольку $R_B^2(\theta, s)$, строго говоря, не обладает свойством унимодальности. Причиной этого является неортогональность векторов ошибки и проекции шума: $\delta_B(s) \hat{\xi}_B(s) \neq 0$, вследствие чего шум на проверочной выборке $\hat{\xi}_B$ может как ухудшать, так и улучшать результаты моделирования, т. е. уменьшать теоретический вектор ошибки $\eta_B(s) = u_B - \hat{y}_B(s)$. Однако глобальный минимум критерия $R_B^2(\theta, s)$, безусловно, существует, так что модель оптимальной сложности можно найти при любом ξ , тем более что по построению критерий все равно будет или унимодальным, или максимально близким к таковому.

Отсюда следует очень важный вывод, обосновывающий применимость вычислительных экспериментов для исследования свойств алгоритмов и критериев самоорганизации: результаты отдельного эксперимента (при одной реализации шума) носят об-

щий характер, если они оцениваются на обучающей последовательности A ; для достоверных оценок на последовательности B необходимо выполнить серию экспериментов (при различных реализациях шума) и усреднить их результаты.

5. ПОТЕНЦИАЛЬНАЯ ПОМЕХОУСТОЙЧИВОСТЬ МОДЕЛИРОВАНИЯ

Полученные выше неравенства (4.24) указывают для каждой модели сложности s область изменения уровня шума, в которой эта модель устойчиво выбирается как оптимальная по идеальному критерию. Однако целесообразно ли использовать модель оптимальной сложности, полученную при очень большом θ ? Поскольку при увеличении θ растет смещение оценок коэффициентов (4.6) и соответственно выхода (4.8) модели, построенной по зашумленным данным, то этот вопрос эквивалентен следующему: до какого уровня шума модель сложности еще остается удовлетворительной по точности восстановления или прогнозирования незашумленных данных?

Так как значение $R_G^2(s)$ указывает, насколько существенно модель, построенная по зашумленным данным, отличается от простейшей «модели» в виде среднего u_G , то ясно, что удовлетворительными можно считать только модели с $R_G^2(s) < 1$. Более общим ограничением на величину критерия является

$$R_G^2(s) \leq \rho^2 < 1, \quad (4.34)$$

где «уровень доверия» ρ назначается на основе практических требований к точности моделей. Тогда задача определения потенциальной помехоустойчивости моделирования состоит в следующем: из условия $R_G^2(\theta, s) = \rho^2$ найти функцию потенциальной помехоустойчивости $\theta^*(\rho, s)$ как верхнюю границу изменений уровня шума, при которых модели еще остаются удовлетворительными по точности в смысле (4.34). Соответствующую область $0 \leq \theta \leq \theta^*(\rho, s)$ назовем областью помехоустойчивости модели сложности s .

Помехоустойчивость «в среднем». Из выражения (4.16) для математического ожидания идеального критерия по условию $r_G^2(\theta^*, s) = \rho^2$ нетрудно получить

$$\bar{\theta}_G^*(\rho, s) = \frac{c_G}{c \sqrt{\text{tr}(P_G^T(s) P_G(s))}} \sqrt{\rho^2 - \frac{1}{c_G^2} \delta_G^T(s) \delta_G(s)}, \quad (4.35)$$

причем в случае $G = A$ (из (4.18)) имеем

$$\bar{\theta}_A^*(\rho, s) = \frac{c_A}{c \sqrt{s}} \sqrt{\rho^2 - \frac{\delta_A^2(s)}{c_A^2}}. \quad (4.36)$$

Прежде всего отметим, что $(\delta_A^2(s)/c_A^2) \leq 1$, и $\theta_A^*(\rho, s)$ существует только при $(\delta_A^2(s)/c_A^2) \leq \rho^2$, т. е. при достаточно малых ρ^2 некоторые

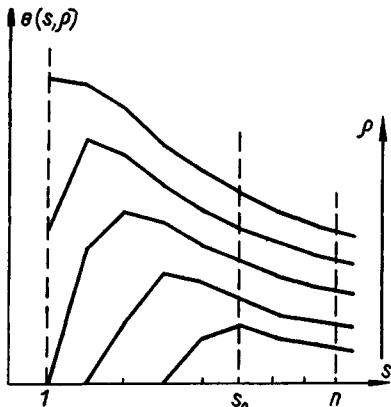


Рис. 15. Показатель потенциальной помехоустойчивости $\theta(\rho, s)$ при различных ограничениях ρ на допустимую величину идеального критерия.

стии кривой изменения критерия, которая лежит ниже прямой $r_A^2(\theta, s) = \rho^2$. Очевидно, что предельным случаем пересечения этой прямой с функцией критерия при увеличении θ является касание прямой и кривой в точке s^* минимума $r_A^2(\theta, s)$, т. е. максимум помехоустойчивости имеет модель, соответствующая точке пересечения прямой $r_A^2(\theta, s) = \rho^2$ с линией геометрического места точек минимумов идеального критерия $r_{\min}^2(s)$. Подставив в (4.36) значение $\rho^2 = r_{\min}^2(s)$ из (4.23), получим функцию геометрического места точек максимумов помехоустойчивости

$$\bar{\theta}_{A\max}^*(s) = \bar{\theta}_A^*(s, r_{\min}(s)) = \frac{1}{c} |\gamma_A(s+1)|. \quad (4.37)$$

С учетом вогнутости функции $\delta_A^2(s)$ линия максимумов (4.37) является монотонно убывающей, причем $\bar{\theta}_{\max}^*(s | s \geq s_0) = 0$. Из сравнения (4.22) и (4.37) видно, что при каждом s $\bar{\theta}_{\max}^*(s) = \theta_{kp}(s)$, т. е. наиболее помехоустойчивой при любом ρ всегда является модель оптимальной сложности.

На рис. 15 представлены типичные варианты графиков верхней границы помехоустойчивости, или функции потенциальной помехоустойчивости $\theta_A^*(\rho, s)$. При каждом ρ область помехоустойчивости под кривой $\rho = \text{const}$ на рис. 15 однозначно отображает соответствующую область удовлетворительных моделей на рис. 14. Отметим, что при $\rho = 0$ существует только одно реальное значение $\theta^*(s_0, 0) = 0$; т. е. восстановить незашумленные данные с нулевой ошибкой можно только по модели с точной структурой при незашумленных данных.

Случай произвольной реализации шума. Здесь функцию потенциальной помехоустойчивости следует искать из квадратного урав-

нения

модели небольшой сложности не только неустойчивы к помехам, но и недовольствительны по точности в смысле $r_A^2(\theta, s) > \rho^2$ даже при $\bar{\theta}_A^* = 0$. При равенстве величин под корнем $\bar{\theta}_A^* = 0$. Это видно на рис. 14: область помехоустойчивых моделей ограничена слева кривой $r_1^2(s)$ и сверху — прямой $r_A^2(\theta, s) = \rho^2$.

Так как $\delta_A^2(s | s \geq s_0) = 0$, то при $s \geq s_0$ функция $\bar{\theta}_A^*(\rho, s)$ имеет вид гиперболы, асимптотически убывающей к нулю. Поэтому ясно, что функция (4.36) при всяком $\rho > 0$ имеет максимум в области $1 \leq s \leq s_0$. Из рис. 14 следует также, что при заданном ρ удовлетворительными будут только модели, соответствующие той части кривой изменения критерия, которая лежит ниже прямой $r_A^2(\theta, s) = \rho^2$.

Поскольку

$$\theta^2 \hat{\xi}_G^T(s) \hat{\xi}_G(s) - 2\theta c \delta_G^T(s) \hat{\xi}_G(s) + \delta_G^T(s) \delta_G(s) = c_G^2 \rho^2. \quad (4.38)$$

Его решение относительно θ имеет вид

$$\theta^* = \frac{h}{p} + \frac{1}{p} \sqrt{h^2 + pq},$$

где $h = c \delta_G^T(s) \hat{\xi}_G(s)$, $p = c \hat{\xi}_G^T(s) \hat{\xi}_G(s)$, $q = c_G^2 \rho^2 - \delta_G^T(s) \delta_G(s)$. Преобразуем подкоренное выражение так:

$$h^2 + pq = c^2 c_G^2 \hat{\xi}_G^T(s) \hat{\xi}_G(s) d,$$

$$d = \rho^2 - \frac{1}{c_G^2} \delta_G^T(s) \delta_G(s) \left[1 - \frac{(\delta_G^T(s) \delta_G(s))^2}{(\delta_G^T(s) \delta_G(s)) (\hat{\xi}_G^T(s) \hat{\xi}_G(s))} \right].$$

Поскольку дробь в квадратных скобках равна $\cos^2 \phi$, где ϕ — угол между векторами $\delta_G(s)$ и $\hat{\xi}_G(s)$, и $\cos \phi \leq 1$, достаточным условием положительности детерминанта является $(\delta_G^T(s) \delta_G(s)/c_G^2) < \rho^2$, т. е. то же условие, что и при помехоустойчивости «в среднем» (определяется независимо от шума).

Итак, для общего случая можно записать

$$\theta_G^*(\rho, s) = \frac{c_G}{c \sqrt{\hat{\xi}_G^T(s) \hat{\xi}_G(s)}} \left(V d + \sqrt{\delta_G^T(s) \delta_G(s)} \frac{\cos \phi}{c_G} \right). \quad (4.39)$$

В случае $G = A$ согласно (4.28) $\delta_A^T(s) \hat{\xi}_A(s) = 0$, т. е. $\cos \phi = 0$ и (4.39) упрощается:

$$\theta_A^*(\rho, s) = \frac{c_A}{c \sqrt{\hat{\xi}_A^T(s) \hat{\xi}_A(s)}} \sqrt{\rho^2 - \delta_A^2(s)/c_A^2}, \quad (4.40)$$

или с учетом (4.38)

$$\theta_A^*(\rho, s) = \frac{c_A}{c \sqrt{\sum_{j=1}^s \mu_A^2(j)}} \sqrt{\rho^2 - \frac{\delta_A^2(s)}{c_A^2}}, \quad (4.40a)$$

причем, как указывалось, $\delta_A^2(s | s \geq s_0) = 0$.

Поскольку согласно (4.31) скаляр $\hat{\xi}_A^T(s) \hat{\xi}_A(s)$ является монотонно возрастающей функцией s , функции потенциальной помехоустойчивости (4.40) свойственны те же закономерности, что и (4.36), справедлив для нее также рис. 15. Кроме того, из равенства $R_{\min}^2(s) = \rho^2$ несложно получить уравнение линии максимумов, аналогичное (4.37):

$$\theta_{A\max}^*(s) = \frac{1}{c} |\gamma_A(s+1)/\mu_A(s+1)|. \quad (4.41)$$

Полученные выше оценки потенциальной помехоустойчивости моделирования по комбинаторному алгоритму МГУА можно суммировать в виде следующих утверждений.

- 1*. Предельно достижимый уровень шума, при котором еще соблюдается ограничение (4.34) на величину относительной ошибки модели, для любой сложности s определяется:
 - а) в смысле математического ожидания идеального критерия — по (4.35) при $A \neq G$ и по (4.36) при $A = G$;
 - б) для отдельной реализации шума — по (4.39) при $G \neq A$ и по (4.40) при $G = A$.
- 2*. Максимумы функций потенциальной помехоустойчивости $\theta_A^*(\rho, s)$ (4.36) и $\theta_A^*(\rho, s)$ (4.40) при увеличении ρ смещаются в область более простых моделей.
- 3*. Модель оптимальной сложности всегда обладает максимальной потенциальной помехоустойчивостью среди всех моделей, удовлетворительных в смысле (4.34).
- 4*. Геометрическое место точек максимумов функций потенциальной помехоустойчивости $\theta_{\max}^*(s)$ и $\theta_{\max}^*(s)$ описывается выражениями (4.37) и (4.41) при $G = A$.

6. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ ПОМЕХОУСТОЙЧИВОСТИ

При построении моделей по ограниченному ряду зашумленных данных измерений оценки коэффициентов и выходов моделей всегда являются неточными. Эта неточность тем больше, чем выше уровень шума, в том числе для модели с точной структурой.

Выполненное аналитическое исследование комбинаторного алгоритма МГУА с идеальным внешним критерием показало, что модель с точной структурой принципиально не может быть наилучшей при любых шумах. В этих условиях целью моделирования является поиск модели оптимальной сложности, которая имеет смещенную (более простую) структуру, но при этом дает наилучшее при данном уровне шума приближение к точным данным. Такая задача моделирования (прогнозирования) принципиально требует применения внешних критериев типа используемых в алгоритмах самоорганизации (они подробно обсуждаются в следующей главе). Отметим, что в терминах гл. 1 рассмотренную выше модель с точной структурой можно назвать *физической* (соответствующей «физике» объекта), а модель оптимальной сложности — *нефизической*. Тогда один из основных выводов выполненного исследования можно выразить так: при больших шумах нефизическая модель эффективнее (точнее) и помехоустойчивее физической. Этим объясняется эффективность применения алгоритмов МГУА в практических задачах с зашумленными данными.

Интересно, что обсуждаемый вывод можно сформулировать также в терминах, аналогичных следствию из второй теоремы Шеннона: при увеличении уровня (мощности) шума *точность* мо-

дели можно сохранить (в определенных пределах), уменьшив ее *сложность*.

Результаты данной главы теоретически однозначно подтверждают известный из практики моделирования по МГУА факт, который был также отмечен в гл. 2 на основе качественной аналогии теорий информации и самоорганизации: модель оптимальной сложности не может быть сложнее полной, а при повышении уровня шума она постепенно становится все более простой по числу и составу аргументов. Это объясняется постепенным исключением из модели аргументов, наименее существенно связанных с выходом (которые становятся соизмеримыми с шумом).

Нами впервые получены точные оценки потенциальной помехоустойчивости моделирования по комбинаторному алгоритму МГУА, в котором не используется никакая дополнительная информация о свойствах шумов, присутствующих в данных (типа корреляционной функции, полосы частот и др.). Эти результаты являются теоретическим обоснованием численных экспериментов по сравнению различных алгоритмов и критериев самоорганизации: даже на частном примере при единственной реализации шума можно однозначно и обоснованно установить преимущество и степень помехоустойчивости того или иного способа селекции моделей. Некоторые результаты численных исследований помехоустойчивости ряда внешних критериев приведены в следующей главе. В шестой главе рассматриваются некоторые способы учета более подробной предварительной информации о помехах.

Основной результат выполненного исследования, обобщающий доказанные выше утверждения, можно сформулировать в виде, аналогичном второй (предельной) теореме Шеннона:

предельно достижимой (потенциальной) помехоустойчивостью моделирования обладает алгоритм самоорганизации с идеальным внешним критерием (4.14). Никакими реальными критериями или их комбинациями нельзя превысить потенциальную помехоустойчивость, можно только сколь угодно точно приблизиться к ней.

Полученные оценки потенциальной помехоустойчивости указывают тот точный предел удовлетворительного моделирования по зашумленным данным с помощью комбинаторного алгоритма МГУА (или любого другого алгоритма, основанного на переборе моделей), которого можно достигнуть, применяя реальные внешние критерии селекции или их комбинации.

7. ПРИЛОЖЕНИЯ

1. Для вывода формулы (4.17) докажем, что

$$\text{tr}(VW) = \text{tr}(WV), \quad V[m \times n], \quad W[n \times m]. \quad (\Pi.1)$$

Для этого запишем указанные следы явно, имея в виду, что $V = \{v_{ij}\}$, $W = \{w_{ji}\}$, $i = 1, m$, $j = 1, n$, и что VW и WV — квадратные

матрицы размеров $m \times m$ и $n \times n$ соответственно:

$$\text{tr}(VW) = \sum_{j=1}^n v_{1j}w_{j1} + \dots + \sum_{j=1}^n v_{mj}w_{jm} = \sum_{i=1}^m \sum_{j=1}^n v_{ij}w_{ji}, \quad (\Pi.2)$$

$$\text{tr}(VV) = \sum_{i=1}^m v_{ii}v_{ii} + \dots + \sum_{i=1}^m v_{nn}v_{nn} = \sum_{i=1}^m \sum_{j=1}^m v_{ij}v_{ij}. \quad (\Pi.3)$$

Двойные суммы в (П.2) и (П.3) перестановочны, что и доказывает (П.1).

С учетом (П.1) нетрудно получить (4.17):

$$\begin{aligned} \text{tr}(P_B^T(s) P_G(s)) &= \text{tr}(P_G(s) P_G^T(s)) = \text{tr}[X_G(s) X_A^+(s) X_A^{+\top}(s) X_G^T(s)] = \\ &= \text{tr}[X_G(s) (X_A^T(s) X_A(s))^{-1} X_G^T(s)] = \text{tr}[(X_A^T(s) X_A(s))^{-1} X_G^T(s) X_G(s)]. \end{aligned} \quad (\Pi.4)$$

2. Выведем некоторые формулы, следующие из определенного в предыдущей главе рекуррентного алгоритма вычисления оценок коэффициентов моделей.

Получим сначала (4.20), для чего формулу (3.19) представим в обозначениях данной главы:

$$\delta_A^2(s) = u_A^T u_A - u_A^T X_A(s) \hat{b}(s). \quad (\Pi.5)$$

При увеличении сложности s на единицу в матрице аргументов добавляется столбец: $X_A(s+1) = (X_A(s); x_A)$, а оценка коэффициентов согласно (3.29) принимает вид

$$\hat{b}(s+1) = \left[\frac{\hat{b}(s) - \hat{\beta} H^{-1}(s) h(s+1)}{\hat{\beta}} \right] = \left[\frac{\hat{b}(s) - \hat{\beta} X_A^+(s) x_A}{\hat{\beta}} \right], \quad (\Pi.6)$$

где в соответствии с (3.27), (3.28)

$$\hat{\beta} = -\kappa(u_A^T P_s(s) x_A - x_A^T u_A) = \kappa x_A^T D_s u_A, \quad (\Pi.7)$$

$$\kappa = 1/(x_A^T x_A - x_A^T P_s x_A) = 1/x_A^T D_s x_A. \quad (\Pi.8)$$

Тогда

$$\begin{aligned} X_A(s+1) \hat{b}(s+1) &= X_A(s) \hat{b}(s) - \hat{\beta} P_s x_A + \hat{\beta} x_A = \\ &= X_A(s) \hat{b}(s) + \hat{\beta} D_s x_A. \end{aligned} \quad (\Pi.9)$$

Подставив (П.7)–(П.9) в формулу для $\delta_A^2(s+1)$, с учетом (П.5) получим (4.20).

Использованная при получении (4.30) величина a_ξ находится из (П.7), (П.8) заменой u_A на x_A :

$$a_\xi = x_A^T D_s x_A / x_A^T D_s x_A. \quad (\Pi.10)$$

3. Для получения (4.20а) заметим сначала, что

$$\delta_G^T(s) \delta_G(s) = u_G^T u_G - 2u_G^T X_G(s) \hat{b}(s) + \hat{b}^T(s) X_G^T(s) X_G(s) \hat{b}(s). \quad (\Pi.11)$$

С учетом (П.6)–(П.8) при увеличении s на единицу имеем

$$\begin{aligned} X_G(s+1) \hat{b}(s+1) &= X_G(s) \hat{b}(s) - \hat{\beta} P_G(s) x_A + \hat{\beta} x_G, \\ \hat{b}(s+1) X_G^T(s) X_G(s) \hat{b}(s+1) &= \hat{b}^T(s) X_G^T(s) X_G(s) \hat{b}(s) + \\ &+ 2\hat{\beta} \hat{b}^T(s) X_G^T(s) (x_G - P_G(s) x_A) + \hat{\beta}^2 (x_G - P_G(s) x_A)^T (x_G - P_G(s) x_A). \end{aligned}$$

Подставив это в $\delta_G^T(s+1) \delta_G(s+1)$, приходим при $G = B$ к (4.20а).

Легко проверить, что при $B = A$ из (4.20а) получаем в точности (4.20) (для этого вместо u_B и $X_B(s)$ надо подставить u_A и $X_A(s)$).

Интересно отметить, что (4.20а) можно представить в виде полного квадрата:

$$\delta_G^T(s+1) \delta_G(s+1) = \|\delta_G(s) - \hat{\beta}(x_G - P_G(s) x_A)\|^2, \quad (\Pi.12)$$

поскольку $\delta_G(s) = u_G - X_G(s) \hat{b}(s)$ (норма в (П.12) евклидова), т. е. ошибку можно рекуррентно вычислять непосредственно через своеобразную «ошибку» предсказания аргумента x_G (имеется в виду $(x_G - P_G(s) x_A)$) с помощью проектирования вектора x_A .

ВНЕШНИЕ КРИТЕРИИ СЕЛЕКЦИИ И ИССЛЕДОВАНИЯ ПОМЕХОУСТОЙЧИВОСТИ МОДЕЛИРОВАНИЯ

1. О ВНУТРЕННИХ И ВНЕШНИХ КРИТЕРИЯХ

В алгоритмах МГУА, как уже отмечалось, задачи параметрической и структурной идентификации решаются совместно: генерируются модели различной сложности (структуры), оцениваются их коэффициенты и выбирается наилучшая модель (модель оптимальной сложности). Для решения этих двух задач применяются по крайней мере два критерия: один — для оценивания параметров, другой — для выбора оптимальной структуры. Использовать один критерий принципиально нельзя (кроме идеального случая отсутствия помех): например, критерий ошибки МНК при увеличении сложности моделей монотонно уменьшается вплоть до обращения в нуль при равенстве числа коэффициентов n и числа точек N , и отдать предпочтение какой-либо модели нельзя («чем сложнее, тем точнее» [49]).

Таким образом, для оценивания параметров и выбора моделей необходимо применять критерии, содержащие различную информацию. При этом информация, заключенная в критерии выбора, является дополнительной, или *внешней*, по отношению к информации, на которой построен критерий оценивания. Такой термин основывается на понятии «внешнего дополнения», введенного С. Биром [7]. Поэтому применяемые в методах самоорганизации критерии выбора (селекции) моделей называются *внешними критериями*. Соответственно критерии для оценивания коэффициентов принято называть *внутренними*.

Основным способом разделения информации на «внутреннюю» и «внешнюю» в алгоритмах МГУА является разбиение всей имеющейся выборки исходных данных на две и более частей (последовательностей). Такое разделение позволяет построить разнообразные критерии для количественного сравнения моделей различной структуры. Кроме этого для построения внешних критериев можно также привлекать и другую априорную информацию о моделируемом объекте или процессе. Например, если известно, что несколько моделируемых переменных связаны между собой функциональной зависимостью, то эту информацию можно отразить в виде критерия баланса переменных [58]. Критерий баланса прог-

нозов [85, 87] отражает связь между значениями одной и той же переменной, вычисленными с различным интервалом усреднения (подробнее см. гл. 7). Таким образом, внешний критерий является количественной мерой степени соответствия конкретной модели тому или иному требованию, предъявляемому к ней «заказчиком». Поскольку же требования, которым должна удовлетворять лучшая модель, могут быть различными, то часто приходится при моделировании учитывать не один, а несколько внешних критериев, т. е. осуществлять многокритериальную селекцию моделей [49].

Отметим, что ряд критериев, применяемых в регрессионном анализе (например, критерий Фишера), в принципе также позволяют выбирать одну модель из некоторого их множества. При этом используются «внешние дополнения» другого рода — предположения о свойствах генеральной совокупности, уровне доверия, характере распределения случайных помех и др. [35, 141]. Однако эти предположения часто являются трудно проверяемыми на практике и, кроме того, субъективными. Внешний характер имеют также так называемые информационные критерии [182, 202, 106]. В алгоритмах МГУА статистические и информационные критерии применяются редко. Внешние критерии селекции строятся так, чтобы они отражали некоторые очевидные требования: «модель должна хорошо прогнозировать» (критерий регулярности), «коэффициенты модели должны минимально зависеть от части выборки» (критерий минимума смещения), «модель должна отражать связь между переменными» (критерий баланса). В то же время показано, что при известных характеристиках помех применение статистических и внешних критериев селекции моделей дает согласованные результаты [2, 3, 56].

В общем случае внешний критерий селекции можно определить как меру количественного сравнения моделей различной сложности, которая позволяет выделить некоторое подмножество лучших моделей из всего множества моделей, генерируемых в процессе самоорганизации. Если, кроме того, наложить требование *единственности* выбора по одному или нескольким критериям, то процедура самоорганизации дает единственную модель оптимальной сложности. Очевидно, что для отдельного внешнего критерия требование единственности является обязательным.

Отдельно взятые критерии часто называют *индивидуальными*, свертку нескольких критериев — *комбинированным*, а несколько поочередно применяемых критериев — *последовательностью критериев*.

2. КЛАССИФИКАЦИЯ ОСНОВНЫХ ВНЕШНИХ КРИТЕРИЕВ

Предположим, что все множество (выборка) точек исходных данных разделено на три непересекающихся подмножества (части выборки) A , B и C , причем отдельно обозначим объединение $W = A \cup B$. Все критерии, используемые в алгоритмах МГУА, можно выразить в том или ином виде через оценки коэффициентов моде-

лей, полученные на A , B и \mathbb{W} и через оценки выхода моделей на A , B , \mathbb{W} и C . С учетом этого опишем основные индивидуальные, комбинированные и корреляционные (иеквадратичные) критерии.

Все индивидуальные внешние критерии, являющиеся квадратичными формами, можно разделить на две основные группы: 1) критерии *точности*, выражающие ошибку проверяемой модели на различных частях выборки, 2) критерии *согласованности*, являющиеся мерой близости оценок, полученных на различных частях выборки. При этом как те, так и другие критерии могут применяться в симметричной или несимметричной форме. Симметричным называем критерий, в котором равноправно используется информация частей A и B выборки; несимметричным — в противном случае. Рассмотрим типичные квадратичные критерии, представляющие обе основные группы. Как и в предыдущих главах, предполагаемую модель представим в виде

$$y = Xa, \quad y [N \times 1], \quad X [N \times n], \quad (5.1)$$

причем оптимальный вектор a^* ищем по комбинаторному алгоритму (полным перебором моделей различной сложности) с учетом разделения $y^r = (y_A^r; y_B^r; y_C^r)$, $X^r = (X_A^r; X_B^r; X_C^r)$, а матрицы X_A и X_B имеют полный ранг.

Критерии точности. Типичным критерием этого вида (и исторически первым критерием селекции [41]) является критерий *регулярности*:

$$\begin{aligned} \Delta^2(B) &= \Delta^2(B|A) = \sum_{i \in B} (y_i - \hat{y}_i(A))^2 = \|y_B - \hat{y}_B(A)\|^2 = \\ &= (y_B - X_B \hat{a}_A)^T (y_B - X_B \hat{a}_A) = \|y_B - X_B \hat{a}_A\|^2, \end{aligned} \quad (5.2)$$

где запись $\Delta^2(B|A)$ означает «ошибка на B модели, коэффициенты которой получены на A », причем $\hat{a}_A = (X_A^T X_A)^{-1} X_A^T y$, $\hat{y}_B(A) = X_B \hat{a}_A$. Здесь и далее для упрощения обозначений не будем указывать сложность модели s и соответствующую этой модели подматрицу аргументов $X(s)$, норма всегда евклидова.

Очевидно, что критерий регулярности (5.2) несимметричен — выборка A является обучающей, B — проверочной. Поменяв A и B местами, получим еще один критерий регулярности

$$\Delta^2(A) = \Delta^2(A|B) = \|y_A - X_A \hat{a}_B\|^2. \quad (5.2a)$$

Теперь легко сконструировать *симметричный критерий регулярности* [138]

$$\begin{aligned} d^2 &= d^2(A, B|B, A) = \Delta^2(B|A) + \Delta^2(A|B) = \\ &= \|y_B - X_B \hat{a}_A\|^2 + \|y_A - X_A \hat{a}_B\|^2, \end{aligned} \quad (5.3)$$

где уже части A и B используются равноправно. Он позволяет более надежно выбирать модель по сравнению с обычным крите-

рием (5.2), так как позволяет сглаживать влияние помех на обеих частях A и B выборки.

Если требуется, чтобы лучшая модель имела достаточно хорошую точность одновременно как на обучающей, так и на проверочной последовательностях (хорошо экстраполировала и аппроксимировала), то это можно выразить критерием

$$\begin{aligned} \kappa^2 &= \kappa^2(\mathbb{W}|A) = \Delta^2(\mathbb{W}|A) = \Delta^2(A|A) + \Delta^2(B|A) = \\ &= \varepsilon^2(A) + \Delta^2(B) = \|y_{\mathbb{W}} - X_{\mathbb{W}} \hat{a}_A\|^2, \end{aligned} \quad (5.4)$$

где $\varepsilon^2(A) = \Delta^2(A|A)$ — точность МНК (остаточная сумма квадратов). Его можно назвать критерием *стабильности*, который также несимметричен. Нетрудно записать аналогичный (5.4) критерий $\kappa^2(\mathbb{W}|B)$ и, соответственно, *симметричный критерий стабильности*:

$$\begin{aligned} S^2 &= S^2(\mathbb{W}|A, B) = \kappa^2(\mathbb{W}|A) + \kappa^2(\mathbb{W}|B) = \\ &= \|y_{\mathbb{W}} - X_{\mathbb{W}} \hat{a}_A\|^2 + \|y_{\mathbb{W}} - X_{\mathbb{W}} \hat{a}_B\|^2. \end{aligned} \quad (5.5)$$

Использовать симметричные формы критериев всегда предпочтительнее, учитывая конечность выборки данных: во-первых, снижается чувствительность критериев к способу разделения данных на части, во-вторых, усредняется влияние помех (происходит своеобразная фильтрация). Другими словами, симметричные критерии обладают большей помехоустойчивостью, что и будет показано ниже.

Оригинальный критерий точности предложен в работе [212], согласно которому вычисляется среднее значение $N_{\mathbb{W}}$ критериев регулярности для каждой проверяемой частной модели в условиях, когда проверочной выборкой является поочередно каждая точка из \mathbb{W} , а обучающей — остальные $N_{\mathbb{W}} - 1$ точек. Такой усредненный критерий регулярности (УКР) имеет вид

$$\Delta_{cp}^2(\mathbb{W}) = \frac{1}{N_{\mathbb{W}}} \sum_{i=1}^{N_{\mathbb{W}}} (y_i - \hat{y}_i(\mathbb{W}_i))^2, \quad (5.6)$$

где $\hat{y}_i(\mathbb{W}_i) = x_i^T \hat{a}_{\mathbb{W}_i}$; x_i — измерения аргументов в i -й точке, $\mathbb{W}_i = \mathbb{W} \setminus i$, $N_{\mathbb{W}_i} = N_{\mathbb{W}} - 1$, $\hat{a}_{\mathbb{W}_i}$ — оценка коэффициентов на подвыборке $A = \mathbb{W}_i$. Реально оценки $\hat{a}_{\mathbb{W}_i}$ можно вычислить не непосредственно, а рекуррентно через $\hat{a}_{\mathbb{W}}$ с учетом исключаемой точки (такой алгоритм предложен в [173]).

Как видим, все критерии точности порождаются формулой квадратической ошибки аппроксимации (ошибки МНК). Внешний характер этих критериев обеспечивается учетом проверочных выборок. Другими словами, одни и те же (по виду формул) критерии точности могут использоваться в качестве внутренних или

внешних в зависимости от частей выборки (это замечание не относится к $\Delta_{\text{cp}}^2(W)$).

Рассмотренные внешние критерии точности (как и все критерии других типов) применяются для решения задач моделирования статических и динамических задач. Однако для моделирования динамики процессов, т. е. для самоорганизации моделей в виде разностных уравнений, наилучшим внешним критерием является так называемый *интегральный критерий* [21]:

$$I^2(W) = \sum_{k=1}^{N_W} (y_k - y_k^*)^2,$$

где оценка y_k^* получена с помощью «пошагового интегрирования» разностного уравнения из заданных начальных условий. Например, в случае авторегрессионного уравнения

$$y_k = a_0 + a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_\tau y_{k-\tau}$$

интегрирование сводится к рекуррентному вычислению y_1, \dots, y_{N_W} при заданных $y_0, y_{-1}, y_{-2}, \dots, y_{-\tau}$, где τ — число учитываемых запаздывающих значений. При этом коэффициенты модели оцениваются обычно по МНК, когда разностное уравнение рассматривается как простое регрессионное. Более строгий подход требует и оценивать коэффициенты по интегральному критерию, однако при этом возникают значительные трудности, связанные с *нелинейной* зависимостью критерия от оцениваемых коэффициентов. В работе [77] предложен итерационный алгоритм минимизации интегрального критерия для оценивания коэффициентов. При этом, очевидно, он становится *внутренним* — например, $I^2(A)$. Тогда для селекции моделей необходимо применять $I^2(B)$.

Критерии согласованности. Критерии этой группы не учитывают в явном виде ошибку модели и поэтому даже формально не могут применяться в качестве внутренних (для оценивания коэффициентов). Они строятся как количественное выражение какого-либо дополнительного требования к свойствам искомой модели.

Например, критерий *минимума смещения коэффициентов* (первый из критериев согласованности) отражает требование, чтобы лучшая модель имела коэффициенты, оценки которых, полученные на A и B , минимально отличались:

$$\tilde{n}_{\text{cm}}^2 = \tilde{n}_{\text{cm}}^2(A, B) = \|\hat{a}_A - \hat{a}_B\|^2.$$

Другая, более распространенная форма этого критерия — *минимум смещения решений*:

$$\begin{aligned} n_{\text{cm}}^2 &= n_{\text{cm}}^2(W | A, B) = \|\hat{y}_W(A) - \hat{y}_W(B)\|^2 = \\ &= \|X_W \hat{a}_A - X_W \hat{a}_B\|^2 = (\hat{a}_A - \hat{a}_B)^T X_W^T X_W (\hat{a}_A - \hat{a}_B), \end{aligned} \quad (5.7)$$

который отличается от $\tilde{n}_{\text{cm}}^2(A, B)$ только наличием весовой матрицы $X_W^T X_W$, однако выражает несколько другое требование — чтобы минимально отличались оценки выхода модели на W по коэффициентам, полученным на A и B .

К этой же группе относится и так называемый *абсолютно по-мехоустойчивый критерий*, предложенный и исследованный в [54]:

$$\begin{aligned} V^2 &= V^2(W | A, B, W) = (\hat{y}_W(W) - \hat{y}_W(A))^T (\hat{y}_W(B) - \hat{y}_W(W)) = \\ &= (\hat{a}_W - \hat{a}_A)^T X_W^T X_W (\hat{a}_B - \hat{a}_W). \end{aligned} \quad (5.8)$$

Он требует максимальной согласованности оценок выхода модели при коэффициентах, полученных на трех частях выборки — A , B и W . Его название связано с тем, что он теоретически удовлетворяет одному из наиболее важных условий работоспособности внешнего критерия — надежно отсеивать переусложенные (избыточные) модели при любом шуме.

Критерии (5.7), (5.8) являются симметричными, однако для них несложно записать их несимметричные формы, например, на B :

$$n_{\text{cm}}^2(B) = \|\hat{y}_B(A) - \hat{y}_B(B)\|^2 = \|X_B \hat{a}_A - X_B \hat{a}_B\|^2, \quad (5.7a)$$

$$\begin{aligned} V^2(B) &= (\hat{y}_B(W) - \hat{y}_B(A))^T (\hat{y}_B(B) - \hat{y}_B(W)) = \\ &= (\hat{a}_W - \hat{a}_A)^T X_B^T X_B (\hat{a}_B - \hat{a}_W). \end{aligned} \quad (5.8a)$$

На практике критерии согласованности в несимметричной форме применяются редко, хотя могут оказаться полезными для выяснения связи между различными внешними критериями. С учетом $X_W^T X_W = X_A^T X_A + X_B^T X_B$ нетрудно получить, что $n_{\text{cm}}^2 \equiv n_{\text{cm}}^2(A) + n_{\text{cm}}^2(B)$ и точно так же $V^2 = V^2(A) + V^2(B)$.

Отметим, что симметричные критерии точности d^2 и S^2 также отражают требование, заложенное в n_{cm}^2 : оптимальная модель должна минимально изменяться при переходе на другую часть выборки.

Комбинированные критерии. Часто в практических задачах необходимо получить модель, удовлетворяющую одновременно нескольким требованиям. Если эти требования выражены в виде соответствующих критериев, то приходим к известной проблеме многокритериального выбора, когда несколько или все критерии являются противоречивыми. Так, даже задача выбора модели, которая была бы одновременно наиболее регулярной (точной) и наименее смещенной, часто оказывается противоречивой (см., например, [45, 89]).

Обычно многокритериальная задача сводится к линейной свертке критериев — *комбинированному критерию* вида

$$k^2 = \sum_{j=1}^m \alpha_j c_j^2, \quad \sum_{j=1}^m \alpha_j = 1.$$

Меняя весовые коэффициенты с учетом ограничения на их сумму, можно получать эффективные решения задачи, принадлежащие «области Парето». Однако в задачах моделирования нас, как правило, интересует единственная лучшая модель. Для получения единственного решения необходимо целесообразно выбрать весовые коэффициенты a_j — а это не всегда удается сделать даже в случае двух критериев:

$$k^2 = \alpha c_1^2 + (1 - \alpha) c_2^2.$$

Для преодоления этой специфической проблемы в алгоритмах МГУА применяется свертка нормированных критериев:

$$k_l^2 = \bar{c}_{1l}^2 + \bar{c}_{2l}^2 = c_{1l}^2/c_{1\max}^2 + c_{2l}^2/c_{2\max}^2, \quad (5.9)$$

где l — индекс рассматриваемой модели, а максимальные значения критериев определяются среди всех F , участвующих в переборе моделей:

$$c_{1\max}^2 = \max_{l=1..F} c_{1l}^2, \quad c_{2\max}^2 = \max_{l=1..F} c_{2l}^2. \quad (5.10)$$

В комбинированных критериях, как правило, учитывается критерий минимума смещения. Например, в COMBI можно воспользоваться критериями «смещение плюс ошибка аппроксимации» (индекс модели опускаем)

$$k_1^2 = \bar{n}_{cm}^2 + \bar{\varepsilon}^2(W) \quad (5.11)$$

и «смещение плюс ошибка на экзамене»

$$k_2^2 = \bar{n}_{cm}^2 + \bar{\Delta}^2(C), \quad (5.12)$$

где $\Delta^2(C) = \Delta^2(C|W)$. Известны еще два критерия этого типа: «смещение плюс регулярность» [45]

$$k_3^2 = \bar{n}_{cm}^2 + \bar{\Delta}^2(B) \quad (5.13)$$

и «смещение плюс симметричная регулярность» [138]

$$k_4^2 = \bar{n}_{cm}^2 + \bar{d}^2. \quad (5.14)$$

Отметим общую черту критериев, участвующих в комбинации: один является критерием согласованности, другой — критерием точности. По-видимому, причина этого состоит в формальной неединственности критериев согласованности n_{cm}^2 и V^2 : первый равен нулю при любых оценках $\hat{a}_A = \hat{a}_B = var$, второй — при любых $\hat{a}_A = \hat{a}_B = var$, $\hat{a}_B = \hat{a}_W = var$. Хотя при наличии помех в данных такие случаи совпадения оценок теоретически весьма маловероятны, из практики моделирования известна тенденция критерия n_{cm}^2 выбирать более простые модели. Поэтому к нему обычно добавляют «стабилизатор» в виде какого-либо критерия точности, так как в явном виде ошибку модели сам n_{cm}^2 не учитывает.

Однако общей трудностью непосредственного применения комбинированных критериев (5.11)–(5.14) является то, что максимальное значение одного из критериев часто оказывается на несколько порядков больше другого. Чаще всего это n_{cm}^2 , и после нормировки член в k^2 , соответствующий смещению, становится несущественным — модель выбирается в основном по второму критерию.

В связи с этим в настоящее время в алгоритмах МГУА реализуется не двухкритериальный, а последовательный отбор: сначала, как правило, выбираются F моделей по n_{cm}^2 , а затем одна лучшая — по какому-либо критерию точности. Такое *последовательное применение* критериев повышает эффективность моделирования, в том числе помехоустойчивость. При этом обходится проблема нормировки, а задачу определения свободы выбора F удается решать в каждом конкретном случае, например, подбором или по опыту решения аналогичных задач. В случае, когда критерии больше двух, также применяется многоэтапный отбор, позволяющий успешно решать достаточно сложные задачи, и прежде всего задачи многоуровневого прогнозирования (см., например [87, 75, 49], а также гл. 7).

Комбинированные критерии с использованием коэффициентов веса слагаемых (частных критериев). Часто ставится специальная цель выбора комбинированного критерия. Например, можно потребовать, чтобы при увеличении мощности помех критерий приводил к выбору одной и той же модели (устойчивости выбора модели). Другим примером может быть требование такого выбора критерия, при котором с увеличением мощности помех минимум критерия (на плоскости критерий — сложность модели) перемещался согласно второй теореме Шеннона (см. гл. 2). Удовлетворить подобные требования можно при помощи коэффициентов веса, с которыми суммируются частные критерии. Например, критерий среднеквадратической ошибки смещает минимум в сторону усложнения модели при увеличении стоящего при нем коэффициента веса.

Таким образом, выбор коэффициентов веса частных критериев позволяет управлять формой геометрического места минимумов комбинированного критерия.

Корреляционные внешние критерии. Эти критерии предъявляют определенные требования к соотношению авто- и взаимокорреляционных характеристик выходов модели и объекта (таблицы данных) и поэтому формально относятся к критериям согласованности. Однако они не выражаются через нормы векторов, т. е. могут быть как положительными, так и отрицательными в отличие от перечисленных выше квадратичных критериев. Поэтому внешние критерии корреляционного типа выделены нами в специальную группу, причем их внешний характер обеспечивается тем, что они вычисляются на проверочной последовательности B по оценкам коэффициентов, полученным на обучающей последовательности A .

Первый критерий этого типа получил название *корреляционный критерий регулярности* [110]:

$$K(B) = \frac{(\bar{y}_B - \hat{\bar{y}}_B)^T (\hat{y}_B(A) - \hat{\bar{y}}_B(A))}{\|\bar{y}_B - \hat{\bar{y}}_B\| \|\hat{y}_B(A) - \hat{\bar{y}}_B(A)\|}, \quad (5.15)$$

где \bar{y}_B и $\hat{\bar{y}}_B(A)$ — средние значения выходов объекта и модели, вычисленные на B . Очевидно, что лучшая модель ищется по условию $K(B) \rightarrow 1$.

В [23] предложена еще одна форма корреляционного критерия, выведенная на основе следующего представления квадрата ошибки между табличными данными y_i и их аппроксимацией \hat{y}_i ($i=1, \dots, N$), полученной, например, по МНК:

$$\varepsilon^2 = \frac{1}{N} (y - \hat{y})^T (y - \hat{y}) = (1 - J_c^2 + J_s^2 + J_m^2) z_v^2, \quad (5.16)$$

где

$$J_c = r(\hat{v}, v) = \hat{v}^T v / N z_{\hat{v}} z_v, \quad (5.17)$$

$$J_s = z_{\hat{v}}^2 / z_v^2 - r(\hat{v}, v), \quad (5.18)$$

$$J_m = (\bar{y} - \hat{\bar{y}}) / z_v. \quad (5.19)$$

$$z_v = (v^T v / N)^{\frac{1}{2}}, \quad z_{\hat{v}} = (\hat{v}^T \hat{v} / N)^{\frac{1}{2}},$$

$$v = y - \bar{y}, \quad \hat{v} = \hat{y} - \hat{\bar{y}}.$$

Полученные компоненты J_c , J_s , J_m вектора ошибки в [23] предложено рассматривать как самостоятельные критерии селекции и вычислять на B с учетом оценки a_A . При этом, очевидно, $J_c(B) = J_c(B|A)$ в точности совпадает с критерием $K(B)$ (5.15). Компонента $J_s(B) = J_s(B|A)$ названа в [23] критерием *согласованности степени нелинейности*, применения которой следует искать модель по условию $J_s(B) \rightarrow 0$. Компонента $J_m(B) = J_m(B|A)$ названа критерием согласования средних, однако он, по-видимому, самостоятельного значения не имеет. Пример эффективного применения корреляционного критерия $J_c(B)$ приведен в [23], причем его рекомендовано использовать в виде $|J_c(B)| \rightarrow \min$. Отметим, что критерий $J_c(B) = K(B)$ также несложно привести к виду, удобному для минимизации: $\tilde{K}(B) = |1 - K^2(B)| \rightarrow \min$.

Корреляционными являются также внешние критерии (2.21), (2.22), (2.25) и (2.26), предложенные в гл. 2 на основе других представлений ошибки ε^2 , отличающихся от (5.19). Первые два из них являются разновидностью критерия регулярности $\Delta^2(B)$:

$$\begin{aligned} \Delta_1(B) &= y_B^T \hat{y}_B(A) - \frac{1}{2} \hat{y}_B^T(A) \hat{y}_B(A) = \\ &= \hat{a}_A^T X_B^T y_B - \frac{1}{2} \hat{a}_A^T X_B^T X_B \hat{a}_A, \\ \Delta_2(B) &= \hat{y}_B^T(A) \hat{y}_B(A) - 2 \hat{y}_B^T \hat{y}_B(A). \end{aligned} \quad (5.20)$$

Эти критерии эквивалентны $\Delta^2(B)$, только $\Delta_1(B)$ максимизируется, а $\Delta_2(B)$ — минимизируется, причем минимум последнего, как правило, отрицателен. Однако эта эквивалентность не означает, что $\Delta_1(B)$ и $\Delta_2(B)$ относятся к критериям точности, поскольку их значения не выражают ошибку модели.

Критерии (2.25) и (2.26) являются типичными корреляционными критериями согласованности:

$$\begin{aligned} h_1(B) &= y_B^T y_B - \hat{y}_B^T(A) \hat{y}_B(A) \rightarrow 0, \\ h_2(B) &= y_B^T \hat{y}_B(A) - \hat{y}_B^T(A) \hat{y}_B(A) \rightarrow 0. \end{aligned} \quad (5.21)$$

Они требуют равенства авто- и взаимокорреляционных функций выходов объекта и модели. Очевидно, что здесь также несложно перейти к минимизации критериев: $|h_1(B)| \rightarrow \min$, $|h_2(B)| \rightarrow \min$.

Все рассмотренные выше корреляционные внешние критерии являются несимметричными. Для симметризации достаточно к каждому из них добавить еще одно их выражение, в котором частоты выборки A и B меняются ролями.

Заканчивая обзор основных внешних критериев, применяемых в алгоритмах МГУА, следует сделать несколько замечаний. Здесь отражены не все критерии, которые в настоящее время существуют, а только их основные разновидности. В конкретных же практических задачах может возникнуть необходимость «изобрести» новый критерий, отражающий специфику задач или дополнительные требования «заказчика». Все эти индивидуальные (отдельно взятые) критерии могут участвовать в различных целесообразно выбранных последовательностях критериев, применяемых в более сложных задачах — таких, как системный анализ, моделирование многомерных объектов, одновременный прогноз нескольких связанных переменных, многоуровневый прогноз.

3. СВЯЗЬ ВНЕШНИХ КРИТЕРИЕВ

Выведем ряд формул, выражающих связь различных внешних критериев друг с другом. Рассмотрим сначала квадратичные критерии симметричного типа. Прежде всего непосредственно из (5.3), (5.5) следует

$$S^2 = d^2 + \varepsilon^2(A) + \varepsilon^2(B), \quad (5.22)$$

где $\varepsilon^2(A) = \Delta^2(A|A)$, $\varepsilon^2(B) = \Delta^2(B|B)$ — ошибки МНК на A и B .

Преобразуем теперь n_{cm}^2 (5.7):

$$n_{cm}^2 = \|(\hat{y}_W - \hat{y}_W(B)) - (\hat{y}_W - \hat{y}_W(A))\|^2 = \|\hat{y}_W - \hat{y}_W(A)\|^2 + \\ + \|\hat{y}_W - \hat{y}_W(B)\|^2 - 2(\hat{y}_W - \hat{y}_W(A))^T(\hat{y}_W - \hat{y}_W(B)).$$

С учетом того что

$$X_W^T X_W = X_A^T X_A + X_B^T X_B, \quad X_W^T y_W = X_A^T y_A + X_B^T y_B, \quad (5.23)$$

$$\hat{a}_A = (X_A^T X_A)^{-1} X_A^T y_A, \quad \hat{a}_B = (X_B^T X_B)^{-1} X_B^T y_B, \quad \hat{a}_W = (X_W^T X_W)^{-1} X_W^T y_W,$$

преобразуем последний член в разложении n_{cm}^2 :

$$(\hat{y}_W - \hat{y}_W(A))^T(\hat{y}_W - \hat{y}_W(B)) = y_A^T y_A - y_A^T X_A^T \hat{a} + \\ + y_B^T y_B - y_B^T X_B^T \hat{a}_B = \varepsilon^2(A) + \varepsilon^2(B). \quad (5.24)$$

Тогда с учетом (5.5) получаем

$$S^2 = n_{cm}^2 + 2(\varepsilon^2(A) + \varepsilon^2(B)) \quad (5.25)$$

или согласно (5.22)

$$d^2 = n_{cm}^2 + \varepsilon^2(A) + \varepsilon^2(B). \quad (5.26)$$

Последняя формула получена впервые в [23], а (5.25) — в [86]. Рассмотрим теперь V^2 (5.8):

$$V^2 = [(y_W - \hat{y}_W(A)) - (y_W - \hat{y}_W(W))]^T [(y_W - \hat{y}_W(W)) - (y_W - \hat{y}_W(B))] = \\ = (y_W - \hat{y}_W(A))^T(y_W - \hat{y}_W(W)) + (y_W - \hat{y}_W(B))^T(y_W - \hat{y}_W(W)) - \\ - \|y_W - \hat{y}_W(W)\|^2 - (y_W - \hat{y}_W(A))^T(y_W - \hat{y}_W(B)).$$

Последний член этого выражения уже рассмотрен (см. 5.24). С учетом равенства $\hat{a}_W = (X_W^T X_W)^{-1} X_W^T y_W$ нетрудно установить, что $y_W^T \hat{y}_W(A) = \hat{y}_W^T(W) \hat{y}_W(A)$, $y_W^T \hat{y}_W(B) = \hat{y}_W^T(W) \hat{y}_W(B)$, поэтому

$$(y_W - \hat{y}_W(A))^T(y_W - \hat{y}_W(W)) = (y_W - \hat{y}_W(B)) \times \\ \times (y_W - \hat{y}_W(W)) = \|y_W - \hat{y}_W(W)\|^2 = \varepsilon^2(W). \quad (5.27)$$

Таким образом, получаем формулу, приведенную также в [23]:

$$V^2 + \varepsilon^2(A) + \varepsilon^2(B) = \varepsilon^2(W). \quad (5.28)$$

Учитывая (5.22) или (5.26), можно записать еще две разновидности последней формулы:

$$V^2 + S^2 = d^2 + \varepsilon^2(W), \quad (5.29)$$

$$V^2 + d^2 = n_{cm}^2 + \varepsilon^2(W). \quad (5.30)$$

Интересно, что при рассмотрении (5.28) может возникнуть сомнение: является ли абсолютно помехоустойчивый критерий V^2 квадратичным (неотрицательным)? Однако непосредственно из (5.8) с учетом (5.23) можно установить, что

$$V^2 = (\hat{a}_W - \hat{a}_A)^T X_W^T X_W (\hat{a}_B - \hat{a}_W) = \hat{a}_A^T X_A^T X_A \hat{a}_A + \\ + \hat{a}_B^T X_B^T X_B \hat{a}_B - \hat{a}_W^T X_A^T X_A \hat{a}_W - \hat{a}_W^T X_B^T X_B \hat{a}_W.$$

Это выражение преобразовывается в сумму двух квадратичных форм:

$$V^2 = \|\hat{y}_A(A) - \hat{y}_A(W)\|^2 + \|\hat{y}_B(B) - \hat{y}_B(W)\|^2 = \\ = (\hat{a}_A - \hat{a}_W)^T X_A^T X_A (\hat{a}_A - \hat{a}_W) + (\hat{a}_B - \hat{a}_W)^T X_B^T X_B (\hat{a}_B - \hat{a}_W), \quad (5.31)$$

т. е. $V^2 \geq 0$ всегда. При этом из (5.28), в частности, следует, что ошибка МНК модели, полученной на W , всегда не меньше суммы ошибок МНК модели этой же структуры, оцененной на A и B в отдельности: $\varepsilon^2(W) \geq \varepsilon^2(A) + \varepsilon^2(B)$.

Таким образом, выражения (5.25), (5.26), (5.29) и (5.30) связывают между собой все симметричные квадратичные критерии. Интересно отметить, что сравнение (5.25) и (5.26) с (5.11) — (5.14) позволяет формально отнести критерии S^2 и d^2 к комбинированным. Однако последние принципиально отличаются от комбинированных естественноностью входящих в них компонент — проблема нормировки здесь не возникает.

Аналогично получаются выражения, связывающие несимметричные критерии. Например, критерий регулярности можно представить так [23]:

$$\Delta^2(B) = \varepsilon^2(B) + n_{cm}^2(B), \quad (5.32)$$

где $n_{cm}^2(B)$ — несимметричный критерий смещения вида (5.7а). Формула (5.26) также связывает критерии регулярности с n_{cm}^2 :

$$\Delta^2(A) + \Delta^2(B) = n_{cm}^2 + \varepsilon^2(A) + \varepsilon^2(B). \quad (5.26a)$$

Связь между критерием регулярности и корреляционными критериями $K(B)$ и $J_s(B)$ следует непосредственно из разложения (5.16)

$$\Delta^2(B) = (1 - K^2(B) + J_s^2(B) + J_m^2(B)) N_W z_v^2. \quad (5.33)$$

Критерии (5.20) и (5.21) получены в гл. 2 на основе разложений квадрата ошибки МНК (2.18) и (2.24). С их учетом имеем

$$\Delta^2(B) = \|y_B\|^2 + \Delta_2(B) = \|y_B\|^2 - 2\Delta_1(B), \quad (5.34)$$

$$\Delta^2(B) = h_1(B) + 2h_2(B). \quad (5.35)$$

Приведенные выше представления одних внешних критериев через другие позволяют с помощью вычисления некоторых харак-

теристик моделей, проверяемых в процессе селекции, однозначно определить остальные. Например, вычислив значения квадратов ошибок $\epsilon^2(A)$, $\epsilon^2(B)$, $\Delta^2(A)$ и $\Delta^2(B)$, определяем непосредственно также d^2 , S^2 и n_{cm}^2 . А после оценки $\epsilon^2(W)$ легко найти также V^2 .

4. КАНОНИЧЕСКАЯ ФОРМА КВАДРАТИЧНЫХ КРИТЕРИЕВ

Информация, заключенная в данных наблюдений, используется на всех этапах самоорганизации моделей, вплоть до вычисления значений внешних критериев. Поэтому любой критерий может быть выражен непосредственно через векторы измерений входов и выхода объекта.

В работах [177, 178] предложено представлять квадратичные критерии в так называемой канонической форме:

$$I = \mathbf{y}^T D_I \mathbf{y}, \quad (5.36)$$

где D_I — симметрическая неотрицательно определенная матрица канонической формы критерия I , выражаящаяся через матрицы X , X_A , X_B , X_W . В этих же работах получены канонические представления для основных квадратичных критериев, которые приводятся ниже.

Введем предварительно следующее обозначение:

$$P_{GQ} = X_G (X_Q^T X_Q)^{-1} X_Q^T, \quad (5.37)$$

где G , Q — подмножества (части выборки), которые могут соответствовать A , B или W . При этом на Q вычисляются оценки коэффициентов \hat{b}_Q , а на G — оценки выхода модели $\hat{y}_G(Q)$.

Несложно получить каноническую форму критерия МНК (минимума среднеквадратической ошибки), поскольку

$$\epsilon^2(W) = \| \mathbf{y}_W - X_W \hat{\mathbf{a}}_W \|^2 = \| \mathbf{y}_W - P_{WW} \mathbf{y}_W \|^2 = \mathbf{y}_W^T (E_W - P_{WW}) \mathbf{y}_W.$$

Соответствующая матрица канонической формы имеет вид $D_\epsilon(W) = E_W - P_{WW}$. С учетом $\mathbf{y}_W^T = (\mathbf{y}_A^T | \mathbf{y}_B^T)$, $X_W^T = (X_A^T | X_B^T)$ матрицу P_{WW} можно представить в блочном виде

$$\begin{aligned} P_{WW} &= P_{WW}^2 = \begin{pmatrix} X_A \\ X_B \end{pmatrix} (X_W^T X_W)^{-1} X_W^T X_W (X_W^T X_W)^{-1} (X_A^T | X_B^T) = \\ &= \begin{pmatrix} P_{AW} P_{AW}^T & P_{AW} P_{BW}^T \\ P_{BW} P_{AW}^T & P_{BW} P_{BW}^T \end{pmatrix}. \end{aligned}$$

Поэтому для $D_\epsilon(W)$ эквивалентна следующая запись:

$$D_\epsilon(W) = E_W - P_{WW} = \begin{pmatrix} E_A - P_{AW} P_{AW}^T & -P_{AW} P_{BW}^T \\ -P_{BW} P_{AW}^T & E_B - P_{BW} P_{BW}^T \end{pmatrix}, \quad (5.38)$$

где E_A , E_B — единичные матрицы размерностей $N_A \times N_A$ и $N_B \times N_B$.

Учитывая, что

$$\Delta^2(B) = \| \mathbf{y}_B - X_B \hat{\mathbf{a}}_B \|^2 = \| \mathbf{y}_B - P_{BA} \mathbf{y}_A \|^2,$$

$$n_{cm}^2 = \| X_W \hat{\mathbf{a}}_A - X_B \hat{\mathbf{a}}_B \|^2 = \| P_{WA} \mathbf{y}_A - P_{WB} \mathbf{y}_B \|^2,$$

получаем матрицы канонической формы для критерия регулярности

$$D_\Delta(B) = \begin{pmatrix} P_{BA}^T P_{BA} & -P_{BA}^T \\ -P_{BA} & E_B \end{pmatrix} \quad (5.39)$$

и критерия минимума смещения

$$D_n(W) = \begin{pmatrix} P_{WA}^T P_{WA} & -P_{WA}^T P_{WB} \\ -P_{WB}^T P_{WA} & P_{WB}^T P_{WB} \end{pmatrix}. \quad (5.40)$$

Матрицу D_V абсолютно помехоустойчивого критерия в соответствии с (5.28) запишем в виде

$$D_V(W) = D_\epsilon(W) - D_\epsilon(A) - D_\epsilon(B).$$

Очевидно, что

$$D_\epsilon(A) + D_\epsilon(B) = \begin{pmatrix} E_A - P_{AA} & 0 \\ 0 & E_B - P_{BB} \end{pmatrix}, \quad (5.41)$$

поэтому, с учетом (5.38), получим

$$D_V(W) = \begin{pmatrix} P_{AA} - P_{AW} P_{AW}^T & -P_{AW} P_{BW}^T \\ -P_{BW} P_{AW}^T & P_{BB} - P_{BW} P_{BW}^T \end{pmatrix}. \quad (5.42)$$

Выражения для $D_\epsilon(W)$, $D_\Delta(B)$, $D_n(W)$ и $D_V(W)$ впервые приведены в [177, 178]. С учетом выражений (5.3), (5.22), (5.39) и (5.41) получим матрицы канонических форм также для симметричных критериев регулярности d^2 и стабильности S^2 :

$$D_d(W) = \begin{pmatrix} E_A - P_{BA}^T P_{BA} & -P_{BA}^T - P_{AB} \\ -P_{BA} - P_{AB}^T & E_B - P_{AB}^T P_{AB} \end{pmatrix}, \quad (5.43)$$

$$D_S(W) = \begin{pmatrix} 2E_A - P_{AA} + P_{BA}^T P_{BA} & -P_{BA}^T - P_{AB} \\ -P_{BA} - P_{AB}^T & 2E_B - P_{BB} + P_{AB}^T P_{AB} \end{pmatrix}. \quad (5.44)$$

Представление внешних критериев в канонической форме, нецелесообразное с вычислительной точки зрения (матрица D_I имеет максимально возможную размерность $N_W \times N_W$), оказывается весьма полезным при исследовании свойств и особенностей критериев [177], в задаче о наилучшем разделении исходных данных [178], а также при анализе помехоустойчивости моделирования.

5. АНАЛИЗ ПОМЕХОУСТОЙЧИВОСТИ ИНДИВИДУАЛЬНЫХ КРИТЕРИЕВ СЕЛЕКЦИИ

Под помехоустойчивостью внешнего критерия будем понимать, как и в предыдущей главе, его способность выбирать из всего множества моделей, построенных по зашумленным данным, такую модель, которая достаточно точно восстанавливает незашумленный выход объекта. Характеристикой шума является его уровень $\theta^2 = \sigma^2/c^2$, где σ^2 — дисперсия (мощность) шума, c^2 — «мощность сигнала» (4.4). Тогда характеристикой помехоустойчивости критерия I является тот уровень шума θ_I , при котором модель, выбранная по $I \rightarrow \min$, восстанавливает «точный сигнал» с ошибкой, не превышающей заданное ограничение ρ . Тогда полное решение задачи оценки помехоустойчивости должно сводиться к следующему: получить для каждого критерия I выражение, описывающее линию минимумов $I_{\min}(s)$ (s — сложность модели); построить функцию изменения ошибки восстановления незашумленных данных, соответствующую линии минимумов $R_G^2(s|I_{\min}(s))$, которая выражается через функцию идеального критерия (4.14); определить уровень шума $\theta_I(\rho)$, при котором $R_G^2(s|I_{\min}(s)) = \rho^2$; сравнить полученные значения $\theta_I(\rho)$ для различных внешних критериев и «ранжировать» их по отношению к потенциальной помехоустойчивости $\theta_R(\rho)$ (для идеального критерия).

Однако полностью решить сформулированную задачу чрезвычайно сложно. Поэтому ограничимся здесь некоторыми качественными оценками помехоустойчивости квадратичных внешних критериев в сравнении между собой и по отношению к идеальному критерию. При этом будем рассматривать свойства критериев в «среднем», т. е. по их математическим ожиданиям, поскольку для случая отдельной реализации шума и при учете разделения выборки на две части, как было показано в предыдущей главе, общих выводов относительно помехоустойчивости сделать нельзя (даже для идеального критерия R_B^2). Кроме того, для некоторого упрощения обозначений будем далее считать переменной дисперсию шума σ^2 , имея в виду ее связь с уровнем шума $\sigma^2 = \theta^2 c^2$, $c^2 = \text{const}$.

Для получения ожидаемого значения критерия $M(s, \sigma)$ (далее будем указывать зависимость всех величин от сложности модели s и дисперсии шума σ) предположим, как и ранее, что $u = u + \zeta$, где u — точный (незашумленный) вектор, ζ — некоррелированный нормальный шум с $\zeta \zeta^T = \sigma^2 E_N$. Тогда с учетом канонической формы (5.36) получим

$$\begin{aligned} \bar{I}(s, \sigma) &= M(s, \sigma) = M[(u + \sigma \zeta)^T D_I(s)(u + \sigma \zeta)] = u^T D_I(s) u + \\ &+ \sigma^2 \operatorname{tr}(D_I(s)) \triangleq I_u(s) + I_\sigma(s, \sigma). \end{aligned} \quad (5.45)$$

Таким образом, среднее значение $I(s, \sigma)$ любого внешнего критерия равно сумме двух составляющих: первая определяется точным вектором u и отражает неточность структуры модели, а втор-

рая связана с наличием шума и учитывает его уровень. Составляющая $I_u(s)$ по виду всегда совпадает с канонической формой соответствующего критерия, причем очевидно, что для любого критерия выполняется равенство $I_u(s|s \geq s_0) = 0$ — при точных данных значение критерия для моделей с точной (сложности s_0) и избыточными структурами равно нулю, т. е. функция $I_u(s)$ при $s = \overline{1, s_0}$ всегда убывает от $I_u(1)$ до $I_u(s_0) = 0$ (возможно, немонотонно). Для этой составляющей справедливы все соотношения связи между критериями, описанные в гл. 5, только вместо вектора ошибки $e = y - \hat{y}(s)$ здесь следует рассматривать в соответствии с § 2 гл. 4, вектор $\delta = u - \hat{u}(s)$.

Для выяснения характера зависимости от s второй составляющей $I_\sigma(s, \sigma)$ необходимо вычислить следы $\operatorname{tr}(D_I(s))$ для каждого критерия. С учетом (5.38) — (5.44) получим

$$\Delta_\sigma^2(s, B) = \operatorname{tr}(D_\Delta(s, B)) = \sigma^2 [N_B + \operatorname{tr}(P_{BA}^T(s) P_{BA}(s))], \quad (5.46a)$$

$$d_\sigma^2(s) = \operatorname{tr}(D_d(s)) = \sigma^2 [N_W + \operatorname{tr}(P_{BA}^T(s) P_{BA}(s) + P_{AB}^T(s) P_{AB}(s))], \quad (5.46b)$$

$$S_\sigma^2(s) = \operatorname{tr}(D_S(s)) = \sigma^2 [2N_W - 2s + \operatorname{tr}(P_{BA}^T(s) P_{BA}(s) + P_{AB}^T(s) P_{AB}(s))], \quad (5.46b)$$

$$n_\sigma^2(s) = \operatorname{tr}(D_n(s)) = \sigma^2 [2s + \operatorname{tr}(P_{BA}^T(s) P_{BA}(s) + P_{AB}^T(s) P_{AB}(s))], \quad (5.46g)$$

$$V_\sigma^2(s) = \operatorname{tr}(D_V(s)) = \sigma^2 s. \quad (5.46d)$$

След матрицы $D_n(s)$ имеет вид (5.46g), хотя, согласно (5.40), она выражается через P_{WA} , P_{WB} (5.40). При его получении учтено, что с помощью (5.37) и приложения 1 гл. 4 можно записать

$$\operatorname{tr}(P_{BA}^T P_{BA} + P_{AB}^T P_{AB}) = \operatorname{tr}[(X_A^T X_A)^{-1} X_B^T X_B + (X_B^T X_B)^{-1} X_A^T X_A]. \quad (5.47)$$

Тогда несложно вычислить, что

$$\begin{aligned} \operatorname{tr}(P_{WA}^T P_{WA} + P_{WB}^T P_{WB}) &= \operatorname{tr}[(X_A^T X_A)^{-1} X_W^T X_W + (X_B^T X_B)^{-1} X_W^T X_W] = \\ &= \operatorname{tr}[2E_s + (X_A^T X_A)^{-1} X_B^T X_B + (X_B^T X_B)^{-1} X_A^T X_A] = \\ &= 2s + \operatorname{tr}(P_{BA}^T P_{BA} + P_{AB}^T P_{AB}). \end{aligned}$$

Из (5.46), (5.47) с учетом положительной определенности и невырожденности матриц $H_A(s) = X_A^T(s) X_A(s)$ и $H_B(s) = X_B^T(s) X_B(s)$ следует, что матрица $L_1 = H_A^{-1}(s) H_B(s)$ имеет положительные собственные числа λ_i , $i = \overline{1, s}$ (см. гл. 4). Очевидно далее, что и собственные числа матрицы $L_2 = H_B^{-1}(s) H_A(s)$ также положительны, причем они равны $1/\lambda_i$, $i = \overline{1, s}$, так как $L_2 = L_1^{-1}$. Поэтому все указанные в (5.46) следы являются возрастающими функциями сложности модели s у всех критериев, кроме $S_\sigma^2(s)$, для которого характер зависимости от s сразу выявить трудно.

Рассмотрим подробнее эту функцию. С помощью (5.47) интересующий нас след можно представить через собственные значения $\lambda_i, \lambda_i^{-1}$:

$$\text{tr}[P_{BA}^T(s)P_{BA}(s) + P_{AB}^T(s)P_{AB}(s)] = \sum_{i=1}^s \left(\lambda_i + \frac{1}{\lambda_i} \right). \quad (5.48)$$

Нетрудно установить, что $z_i = \lambda_i + 1/\lambda_i \geq 2$, причем $z_i = 2$ только при $\lambda_i = 1$. Это означает, что справедлива следующая оценка снизу для скорости возрастания следа (5.48):

$$\sum_{i=1}^s \left(\lambda_i + \frac{1}{\lambda_i} \right) \geq 2s, \quad (5.48a)$$

причем равенство здесь практически маловероятно, так как для этого требуется обеспечить $\lambda_i \equiv 1, i = \overline{1, s}$. Поэтому $S_\sigma(s)$ также не убывает при увеличении s , т. е. компоненты (5.46) всех без исключения квадратичных критериев являются монотонно возрастающими функциями сложности модели s , причем скорость возрастания этих функций определяется дисперсией (уровнем) шума σ^2 .

Таким образом, среднее значение каждого из рассматриваемых внешних критериев является суммой двух составляющих: одна при увеличении сложности моделей убывает (возможно, немонотонно), обращаясь в ноль при $s \geq s_0$, а вторая монотонно возрастает и является прямо пропорциональной σ^2 . Это означает, что минимум каждого критерия существует и единствен, причем он может быть только при $s \leq s_0$ (избыточные модели хуже оптимальной при любом шуме). Поскольку при увеличении σ^2 крутизна второй составляющей возрастает, все критерии ведут себя аналогично идеальному — при возрастании уровня шума их минимумы смещаются влево, в область более простых моделей. Интересно отметить, что свойство отсева избыточных моделей при любом шуме, названное в [54] «абсолютной помехоустойчивостью», оказалось общим свойством всех внешних критериев селекции, в связи с чем название критерия V^2 представляется неудачным.

Попытаемся теперь сравнить между собой помехоустойчивость отдельных критериев. Для этого сначала запишем соотношения, связывающие «детерминированные» составляющие критериев (с помощью формул (5.22) — (5.28) и с учетом замены ε на δ):

$$\delta_u^2(s, G) = u_G^T (E_G - P_{GG}) u_G, \quad G = A, B, W, \quad (5.49a)$$

$$\Delta_u^2(s, B) = \|u_B - \hat{u}_B(s, A)\|^2 = \|u_B - P_{BA}(s)u_A\|^2, \quad (5.49b)$$

$$d_u^2(s) = \Delta_u^2(s, B) + \Delta_u^2(s, A), \quad (5.49c)$$

$$S_u^2(s) = d_u^2(s) + \delta_u^2(s, A) + \delta_u^2(s, B), \quad (5.49d)$$

$$n_u^2(s) = d_u^2(s) - [\delta_u^2(s, A) + \delta_u^2(s, B)], \quad (5.49e)$$

$$V_u^2(s) = \delta_u^2(s, W) - [\delta_u^2(s, A) + \delta_u^2(s, B)]. \quad (5.49f)$$

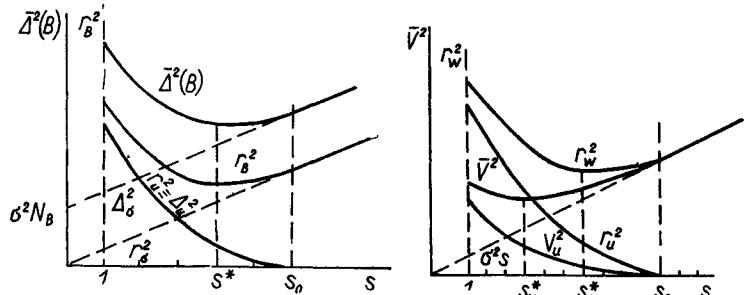


Рис. 16. Сравнение помехоустойчивости идеального критерия r_B^2 и критерия регулярности $\Delta^2(B)$.

Рис. 17. Сравнение помехоустойчивости идеального r_W^2 и абсолютно помехоустойчивого \bar{V}^2 критериев.

Так как все $\delta_u^2(s, G) \geq 0$, то из последних формул следуют такие неравенства:

$$S_u^2(s) > d_u^2(s) > n_u^2(s), \quad (5.50)$$

$$\delta_u^2(s, W) > V_u^2(s). \quad (5.51)$$

Добавим сюда еще выражения для математического ожидания двух вариантов идеального критерия (4.16), выражающих точность моделирования на B и W по оценкам коэффициентов, полученным на A и W (константу c^2 опускаем, имея в виду соотношение $\theta^2 c^2 = \sigma^2$):

$$r_B^2(s) = \Delta_u^2(s, B) + \sigma^2 \text{tr}(P_{BA}^T(s)P_{BA}(s)), \quad (5.52)$$

$$r_W^2(s) = \delta_u^2(s, W) + \sigma^2 s. \quad (5.53)$$

Совокупность выражений (5.46) — (5.53) дает возможность достаточно строго сравнить между собой помехоустойчивость некоторых критериев.

1. Сравнение критерия регулярности $\Delta^2(s, B) = \Delta_u^2(s, B) + \Delta_\sigma^2(s, B)$ и идеального критерия $r_B^2(s)$ (5.52). Из (5.46a), (5.49b) и (5.52) устанавливаем, что «детерминированные» составляющие этих критериев совпадают, а соответствующие шуму отличаются на константу:

$$r_u^2(s, B) = \Delta_u^2(s, B), \quad \Delta_\sigma^2(s, B) = r_\sigma^2(s, B) + \sigma^2 N_B.$$

Это означает, что минимумы $r_B^2(s)$ и $\Delta^2(s, B)$ при любом σ совпадают (рис. 16); т. е. эти критерии в среднем всегда указывают одну и ту же модель оптимальной сложности s^* . Вычисляя для такой модели ошибку воспроизведения (в данном случае прогнозирования на B) незашумленных данных, обнаруживаем, что она совпадает с соответствующим значением идеального критерия $r_B^2(s^*)$. Тем самым доказано, что критерий регулярности «в среднем» дает несмещенную оценку идеального критерия, или обладает потенциальной помехоустойчивостью. К сожалению, столь лест-

ная теоретическая оценка этого критерия не соответствует практическим результатам его применения: он слишком критичен к разделению данных на части в условиях отдельной реализации шума, что будет показано ниже на тестовом примере.

Отметим, что наличие в формулах (5.46), соответствующих критериям точности, членов, пропорциональных числу точек N_B , N_W , связано с записью этих формул в ненормированном виде. На самом деле в качестве оценки квадрата ошибки необходимо брать среднее по числу точек значение или целесообразно нормировать эту сумму квадратов (как это всегда делается на практике), тогда такой пропорциональности не будет. Критерии согласованности, как и идеальный критерий, такой явной зависимости от числа точек не имеют.

2. Сравнение абсолютно помехоустойчивого критерия $\bar{V}^2(s) = V_u^2(s) + V_\sigma^2(s)$ с идеальным $r_W^2(s)$ (5.53). Такое сравнение легко выполнить, так как это единственный критерий, который выражается через алгебраическую сумму ошибок МНК на A , B и W , и вследствие этого след (5.46д) оказывается пропорциональным только сложности модели s , как и у $r_W^2(s)$: $V_\sigma^2(s) = r_\sigma^2(s) = \sigma^2 s$. Детерминированные же части сравниваемых критериев удовлетворяют неравенству (5.51), т. е. при любых s и σ^2 оказывается $r_W^2(s) > \bar{V}^2(s)$, причем это неравенство, как и (5.51), является существенным. Поэтому минимум критерия $\bar{V}^2(s)$ всегда находится левее минимума $r_W^2(s)$, т. е. помехоустойчивость $V^2(s)$ заведомо ниже потенциальной; он всегда указывает на более простую модель, чем наилучшая в смысле идеального критерия (рис. 17).

Отметим, что все симметричные критерии следует сравнивать по помехоустойчивости именно с $r_W^2(s)$, поскольку они оперируют с информацией подвыборок A и B .

3. Сравнение симметричных критериев регулярности $d^2(s)$, стабильности $\bar{S}^2(s)$ и минимума смещения $\bar{n}^2(s)$. Эти критерии не удается связать непосредственно с $r_W^2(s)$, однако их можно сравнить между собой. Детерминированные составляющие этих критериев удовлетворяют неравенству (5.50). Составляющие, отражающие «в среднем» влияние шума, связаны следующим образом (см. (5.46б, в, г)):

$$S_\sigma^2(s) = d_\sigma^2(s) + \sigma^2(N_W - 2s).$$

$$d_\sigma^2(s) = n_\sigma^2(s) + \sigma^2(N_W - 2s).$$

Отсюда видно, что при условии $N_W > 2s$, т. е. при сложностях моделей $s < N_W/2$, будут выполняться неравенства $S_\sigma^2(s) > d_\sigma^2(s) > n_\sigma^2(s)$. Однако при моделировании по комбинаторному алгоритму с применением симметричных критериев необходимо иметь число точек $N_W > 2s_{\max}$ (где $s_{\max} = n$ — максимальная сложность моделей, участвующих в переборе) для обеспечения полного ранга матриц X_A и X_B .

Это означает, что $s < N_W/2$ всегда, поэтому неравенство

$$S_\sigma^2(s) > d_\sigma^2(s) > n_\sigma^2(s),$$

$$s = \overline{1, n},$$

справедливо для рассматриваемых в COMBI моделей любой сложности (равенство возможно лишь при $s = n$, $N_W = 2n$). Отсюда с учетом неравенства $S_u^2(s) > d_u^2(s) > n_u^2(s)$ следует, что при любом σ^2 с ростом сложности моделей минимумы рассматриваемых трех критериев наступают в следующем порядке (рис. 18): сначала \bar{n}_{cm}^2 , затем \bar{d}^2 и, наконец, \bar{S}^2 , т. е. в среднем критерий симметричной стабильности более помехоустойчив, чем критерий симметричной регулярности, а они оба превосходят в этом смысле критерий минимума смещения. В целом ранжировка симметричных критериев в смысле степени помехоустойчивости, полученная в результате сравнения свойств их математических ожиданий соответствует неравенству (5.50).

Из полученных результатов следует, что помехоустойчивость каждого отдельного критерия $I(s)$ связана прежде всего со скоростью возрастания «шумовой» составляющей $I_0(s)$ при увеличении сложности s : чем больше эта скорость, тем ниже помехоустойчивость, т. е. при меньшем σ^2 критерий начинает выбирать более простую модель. Возникает вопрос: можно ли «управлять» этой скоростью? Из (5.46) следует, что для V^2 нельзя, а для остальных критериев можно: необходимо так планировать эксперимент, чтобы уменьшить зависимость следов матриц в (5.46) от сложности s . При этом согласно (5.48) для критерия \bar{d}^2 можно обеспечить в лучшем случае $d_\sigma^2(s) = \theta^2(N_W + 2s)$, т. е. зависимость от s для него весьма существенна. А для критерия \bar{S}^2 теоретически возможно $S_\sigma^2(s) \approx N_W \sigma^2$, т. е. в среднем зависимость данной составляющей от s может быть достаточно слабой. Тем самым усиливается сделанный выше вывод о наибольшей помехоустойчивости критерия симметричной стабильности (и подтверждается целесообразность его названия). Однако при подобном планировании эксперимента возникает вопрос о поведении детерминированной составляющей $S_u^2(s)$ (как и других критериев), требующий более тщательного рассмотрения.

В заключение имеет смысл еще раз обсудить природу «внешнего» характера критериев селекции. Главное проявление «внеш-

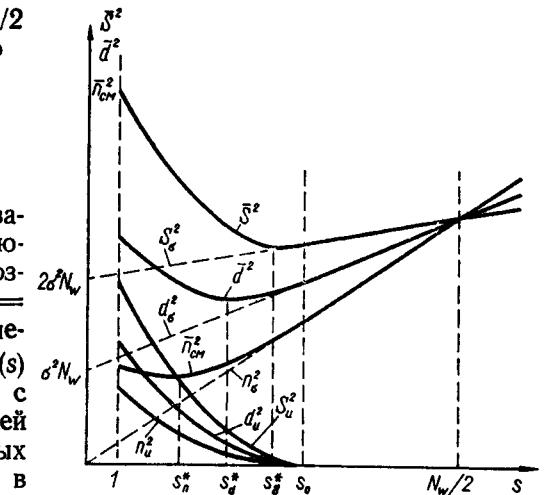


Рис. 18. Сравнение помехоустойчивости симметричных критериев минимума смещения \bar{n}_{cm}^2 , регулярности \bar{d}^2 и стабильности \bar{S}^2 .

ности» — наличие минимума у каждого критерия как функции сложности модели. Из выполненного анализа следует вывод: наличие минимума вызвано тем, что каждый критерий обеспечивает компромисс между степенью уменьшения ошибки восстановления незашумленных (точных) данных и степенью увеличения дополнительной ошибки моделирования из-за наличия шумов в измерениях. И в этом состоит одно из главных достоинств внешних критериев самоорганизации, построенных на разделении имеющейся информации на две и более частей и сопоставлении результатов обработки этой информации: все критерии выполняют указанный компромисс (принципиально необходимый в условиях помех) автоматически, или объективно. Статистические критерии проверки гипотез, применяемые в аналогичной задаче выбора «наилучшей» регрессии [141], также осуществляют компромисс — между уменьшением ошибки аппроксимации и увеличением сложности модели. При этом рост сложности ограничивается с помощью введения в критерии обратной их зависимости от числа степеней свободы, т. е. подразумевается, что уменьшение числа степеней свободы эквивалентно учету возрастания влияния помех измерений при увеличении сложности моделей. Такой априорный, искусственно введенный способ неявного учета влияния помех придает субъективный характер статистическим критериям.

6. ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ПОМЕХОУСТОЙЧИВОСТИ ОДНО- И ДВУХКРИТЕРИАЛЬНОЙ СЕЛЕКЦИИ МОДЕЛЕЙ

Задачи прямого моделирования на ЭВМ реальных объектов или процессов решаются по данным экспериментальных наблюдений, которые всегда в той или иной степени зашумлены помехами различной природы. Статистические характеристики помех нам, как правило, не известны, что затрудняет непосредственную оценку их влияния на результаты моделирования. Поэтому целесообразно исследовать устойчивость выбора моделей на тестовых примерах, когда известны точная модель (или незашумленные данные) и характеристики (закон распределения и интенсивность) помех.

Результаты тестовой проверки алгоритмов самоорганизации на помехоустойчивость селекции моделей приводились в [2, 23, 54], где изучалась работоспособность отдельных критериев селекции при возрастании интенсивности шума. Здесь изложим результаты сравнительной тестовой проверки нескольких наиболее применяемых критериев самоорганизации, причем в условиях как однокритериального, так и двухкритериального выбора моделей.

Методика вычислительного эксперимента по тестовой оценке помехоустойчивости критериев селекции соответствует постановке задачи оценки потенциальной помехоустойчивости (см. § 2 гл. 4). При этом известный незашумленный вектор наблюдений, или точный сигнал u , зашумляют вектором помехи с возрастающей дисперсией: $y = u + \theta c \zeta$, где $\theta^2 c^2 = \sigma^2$ — дисперсия; θ^2 — уровень шума;

c^2 — мощность сигнала (4.4). По данным с шумом ищется с помощью комбинаторного алгоритма МГУА [146] модель оптимальной сложности в полиномиальном базисе, линейная по коэффициентам: $y = a^T x$, где число n аргументов $x[n \times 1]$ больше числа аргументов s_0 в модели истинной структуры. Модель оптимизируется по некоторому критерию селекции или группе критериев при заданном разделении выборки на последовательности A , B и C с числом точек $N_A + N_B + N_C = N$, причем, как и выше, $A \cup B = W$, $N_A + N_B = N_W$, $N_W \geq 2n$. Степень помехоустойчивости одного или нескольких критериев селекции можно установить, постепенно увеличивая уровень шума и определяя такое критическое значение θ^* , выше которого критерий «отказывает».

Понятие «отказ критерия» целесообразно определять по-разному для случаев идентификации, фильтрации и прогнозирования. При идентификации требуется установить точную структуру исходной модели, поэтому θ_i^* находится по моменту потери этой структуры. При этом не требуется точного восстановления значений коэффициентов — оценки могут быть смещены из-за ограниченности выборки данных. Задача фильтрации состоит в восстановлении точного сигнала, т. е. числовых значений вектора u . Точность фильтрации оцениваем величиной идеального критерия вида (4.14):

$$R^2(W) = \sum_{i \in W} (u_i - \hat{y}_i)^2 / \sum_{i \in W} (u_i - \bar{u})^2, \quad (5.54)$$

которая представляет собой приближенный эквивалент дисперсионного отношения и показывает, насколько существенно модель, построенная по зашумленным данным, отличается в смысле точности аппроксимации точного сигнала от «модели» в виде среднего \bar{u} . Примем, что участком фильтрации является объединенная последовательность W , участком прогнозирования — последовательность C , точки которой не участвуют в построении моделей. При оценке помехоустойчивости фильтрации значение θ_F^* будем вычислять по условию $R^2(W) = 1$, т. е. при $R^2(W) < 1$ степень фильтрации будем считать удовлетворительной. При решении задачи прогнозирования коэффициенты моделей определяются на W , а прогнозы (или экстраполяция) проверяются на последовательности C (на экзамене). Так как требуется числовой прогноз, то его точность оценивается аналогично:

$$R^2(C) = \sum_{i \in C} (u_i - \hat{y}_i)^2 / \sum_{i \in C} (u_i - \bar{u})^2 \quad (5.55)$$

и значение θ_p^* устанавливается по условию $R^2(C) = 1$. При $R^2(C) < 1$ степень точности прогнозирования будем считать удовлетворительной.

Итак, для заданного тестового примера степень (уровень) помехоустойчивости каждого критерия селекции можно характеризовать значениями предельных уровней шума θ_i^* , θ_F^* и θ_p^* , при

которых задачи идентификации, фильтрации и прогнозирования еще решаются удовлетворительно. Следует отметить, что эти предельные значения для выбранного критерия, по-видимому, зависят не только от конкретного примера, но и от полученной случайной реализации шума. Поэтому, вообще говоря, при каждом уровне шума целесообразно получить модели для нескольких реализаций шума с последующим усреднением результатов этих статистических испытаний. Однако при этом возникают специфические затруднения, связанные с тем, что рассматриваемая задача моделирования является задачей нелинейного дискретного программирования, при решении которой получаются различные модели для различных реализаций шума. В то же время и при одной реализации шума можно практически рассмотреть следующие вопросы: существенно ли зависит степень помехоустойчивости критерия селекции от числа точек измерений (от объема выборки) и от их разделения на три последовательности; что «теряется» раньше при увеличении шума — структура модели или точность фильтрации (прогнозирования); повышается ли помехоустойчивость выбора модели при применении нескольких критериев селекции? Эти и другие вопросы рассмотрены ниже, при решении тестового примера. Причем следует отметить, что согласно выводам, полученным в гл. 4, результаты оценки решения задачи фильтрации по одной реализации шума являются общими, т. е. справедливыми также и для любой другой реализации, и «в среднем», по математическому ожиданию.

В описываемом вычислительном эксперименте проверялись на помехоустойчивость наиболее часто употребляемые в алгоритмах МГУА критерии селекции: регулярность $\Delta^2(B)$ (5.2), минимум смещения n_{cm}^2 (5.7), симметричные критерии регулярности d^2 (5.3), стабильности S^2 (5.5), а также комбинированные k_1^2 (5.11) и k_2^2 (5.12). На примере сравниваются некоторые варианты одиночно-двухкритериального выбора моделей, оценивается степень помехоустойчивости селекции для этих вариантов.

Отметим, что все рассмотренные выше внешние критерии селекции приведены в виде своих абсолютных значений. На практике же в алгоритмах МГУА эти критерии в таком виде применяются только в процессе перебора и селекции моделей. А при выводе на печать лучших моделей все характеризующие их критерии обязательно нормируются, после чего относительные значения критериев становятся легко интерпретируемыми. Обычно применяется два вида нормировки — запишем их на примере критерия регулярности:

$$\tilde{\Delta}(B) = \sqrt{\|y_B - \hat{y}_B(A)\|^2 / \|y_B\|^2},$$

$$\tilde{\Delta}(B) = \sqrt{\|y_B - \hat{y}_B(A)\|^2 / \|y_B - \bar{y}_B\|^2}.$$

В последнее время чаще применяется второй способ, поскольку при нем полученное выражение близко к обычному дисперси-

Таблица 6. Исходные данные для тестового примера при различных уровнях шума θ

t	Значения зашумленной переменной y при различных θ						
	0	10	20	40	100	200	400
1	9,9	8,43	6,95	4,0	-4,84	-19,58	-49,06
2	9,6	10,41	11,22	12,88	17,68	25,57	41,90
3	9,1	9,67	10,23	11,37	14,76	20,42	31,75
4	8,4	9,93	11,46	14,51	23,68	38,96	69,52
5	7,5	8,40	9,31	11,11	16,56	25,57	43,64
6	6,4	7,58	8,76	11,11	18,19	29,97	53,54
7	5,1	3,39	1,68	-1,75	-12,02	-29,13	-63,36
8	3,6	6,11	8,63	-3,66	28,74	53,87	104,20
9	1,9	1,80	1,70	1,50	0,91	-0,085	-2,07
10	0,0	1,04	1,07	4,14	10,35	20,70	41,40
11	-2,1	-4,32	-6,53	-10,96	-24,25	-46,40	-90,70
12	-4,4	-2,68	-0,95	2,48	12,81	30,01	64,42
13	-6,9	-8,65	-10,41	-13,91	-24,43	-41,19	77,03
14	-9,6	-12,19	-14,78	-19,96	-35,49	-61,39	-113,2
15	-12,5	-11,72	-10,94	-9,37	-4,68	3,14	18,78
16	-15,6	-13,73	-11,85	-8,11	3,33	21,86	59,32
17	-18,9	-19,64	-20,38	-21,86	-26,30	-33,69	-48,49
18	-22,4	-21,19	-21,43	-20,45	-17,52	-12,56	-2,89
19	-26,1	-28,43	-30,75	-35,41	-49,37	-72,63	-119,2
20	-30,0	-28,38	-26,78	-23,56	-13,90	2,90	34,40
21	-34,1	-32,54	-30,97	-27,85	-18,46	-2,82	28,44
22	-38,4	-39,64	-40,87	-43,34	-50,76	-63,11	-87,83

онному отношению. Этот способ удобен тем, что позволяет однозначно оценить качество модели: например, при точности на экзаменационной последовательности $\Delta(C) < 0,5$ модель является хорошей, при $0,5 \leq \Delta(C) \leq 0,8$ — удовлетворительной (гидрометеорологический критерий [49]). Другие критерии селекции нормируются аналогично.

Пример оценки помехоустойчивости селекции моделей. Вычислительный эксперимент по оценке помехоустойчивости различных критериев селекции проводился на следующем тестовом примере:

$$u = 10 - 0,1 t^2.$$

Рассматривался участок этой параболы, соответствующий значениям аргумента $t = 1, 2, 3, \dots, 22$. Были получены 22 значения нормально распределенного случайного шума ξ с нулевым математическим ожиданием и единичной дисперсией. Затем эта реализация шума была просуммирована с исходной параболой при различных значениях уровня шума. Это было выполнено при следующих уровнях θ (в процентах): 3, 5, 10, 20, 40, 60, 80, 100, 130, 160, 200, 230, 260, 300, 330, 360, 400 %. Некоторые из полученных реализаций зашумленного сигнала $y = u + \theta^2 c^2 \xi$, где c^2 — мощность сигнала, вычисленная по (4.4), приведены в табл. 6.

Таблица 7. Значения показателей предельной помехоустойчивости θ_{ii}^* для различных критериев селекции

Критерий	Значения θ_{ii}^* для вариантов разделения данных			
	4+4+3	7+7+8	8+8+6	9+9+4
$\Delta(B)$	20	360	0	40
n_{cm}	20	60	20	80
k_2	0	10	10	0
d	20	80	20	80
S	20	130	20	80
$R(W)$	100	200	160	230

данных: $N_A + N_B + N_C = 4 + 4 + 3$.

В качестве базисного набора функций, в котором с помощью комбинаторного алгоритма МГУА определялась наилучшая по заданному критерию модель, были выбраны члены полинома третьей степени от одного аргумента, из которого можно составить 15 полиномов различной структуры (включая точную структуру). Были выполнены расчеты также при четвертой (полное число структур 31) и пятой (число структур 63) степенях полинома, но это никак не повлияло на полученные результаты — оценки уровня помехоустойчивости оказались одинаковыми для этих достаточно разных вариантов выбора базиса, в котором строились модели.

Отметим, что данный вычислительный эксперимент выполнялся в более общих условиях, чем в гл. 4. Аргументы полинома третьей степени от одной переменной не являются обособленными, причем с выходом тесно коррелированы не только «истинные» аргументы $t^0=1$ и t^2 , но и «избыточные» t и t^3 .

Помехоустойчивость однокритериальной селекции. Рассмотрим результаты испытаний помехоустойчивости всех упомянутых выше критериев, кроме комбинированного критерия k_1 , который оказался неработоспособным практически во всех случаях применения — точнее, он всегда был хуже в смысле уровня помехоустойчивости входящего в него критерия регулярности $\Delta^2(B)$. Очевидно, это свидетельствует о том, что вес внутреннего критерия при нормировке вида (5.9) оказался слишком большим.

Результаты определения предельной помехоустойчивости идентификации, т. е. величины уровня θ_{ii}^* , при котором еще точно восстанавливается структура исходной модели, приведены в табл. 7. Сразу можно отметить, что наименее помехоустойчивым является комбинированный критерий k_2 , что также свидетельствует о нецелесообразности применения нормировки (5.9). Как показано ниже, входящие в данный критерий индивидуальные критерии эффективнее работают при последовательном применении. Наиболее чувствительным к варианту разделения данных оказался критерий

Эксперименты проводились при трех вариантах разделения всей выборки данных на три последовательности общей длиной в 22 точки: 1) $N_A + N_B + N_C = 7 + 7 + 8$; 2) $N_A + N_B + N_C = 8 + 8 + 6$; 3) $N_A + N_B + N_C = 9 + 9 + 4$ (во всех вариантах точки брались подряд). Кроме того, исследовался также случай учета только одиннадцати точек на этом же интервале: из полученных ранее таблиц брались строки, соответствующие нечетным значениям аргумента. При этом рассматривался только один вариант разделения

Таблица 8. Изменение структур моделей, отбираемых по различным критериям при возрастании уровня шума θ (для варианта 8+8+6 точек)

$\theta, \%$	Структурные числа оптимальных моделей для критериев						
	$\Delta(B)$	n_{cm}	k_2	d	S	$R(W)$	$R(C)$
5	1011	1010	1010	1010	1010	1010	1010
10	1011	1010	1010	1010	1010	1010	1010
20	1011	1010	1110	1010	1010	1010	1010
40	1011	1001	1011	1001	1001	1010	1010
60	1011	1001	1011	1001	1001	1010	1010
80	1011	0110	1101	0110	0110	1010	1010
100	1011	0110	1101	0110	0110	1010	1101
130	0110	0110	1010	0110	0110	1010	1101
160	1000	0110	1101	0110	0110	1001	1101
200	0110	0110	1101	0110	0110	1001	1101
230	0110	0110	1101	0110	0110	0011	1101
260	0110	0110	1101	0110	0110	0011	1101
300	0110	1000	1101	1000	0110	0011	1100
330	0110	1000	1100	1000	0110	0011	1100
360	0110	1000	0001	1000	0110	0011	1100
400	1000	1000	0110	1000	1000	0001	0110

рий регулярности: рекордная помехоустойчивость (360 %) при первом варианте разделения 22 точек и нулевая — при втором, что требует осторожности при применении его в задачах с помехами. Во всех остальных вариантах этот критерий работает не лучше других.

Достаточно устойчивыми по отношению к варианту разделения данных оказались критерии симметричного типа n_{cm} , d , S . Они для данной дискретизации уровня шума оказались близкими в смысле помехоустойчивости при индивидуальном применении — только в одном случае критерий S заметно превзошел остальные.

В табл. 7 приведены также значения θ_{ii}^* для идеального критерия $R(W)$ — именно эти значения примерно (с точностью до принятой дискретизации уровня шума) соответствуют предельным оценкам помехоустойчивого выбора модели с точной структурой. Как видно, они не достигаются ни одним из индивидуальных критериев. А «рекордная» помехоустойчивость в варианте разделения 7+7+8 — это грубая ошибка критерия $\Delta(B)$, поскольку при уровне шума выше 200 % модель с точной структурой уже не является оптимальной в смысле минимума идеального критерия.

При увеличении уровня помех выше значений, указанных в таблице, критерии отбирают модели с другой структурой, причем с более простой по числу аргументов или по степени нелинейности. При этом оказывается, что удовлетворительная точность фильтрации и прогнозирования для отобранных моделей сохраняется при существенно более высоких уровнях помех.

Рассмотрим подробнее результаты тестовых испытаний для наименее благоприятного случая разделения данных $N_A + N_B +$

$+N_c=8+8+6$. При увеличении уровня помех структуры отбираемых по различным критериям моделей изменяются в соответствии с табл. 8. Приведенные в этой таблице структурные числа (двоичные векторы) состоят из нулей и единиц, указывающих на отсутствие или наличие в отобранный модели соответствующих членов полного полинома вида $y=a_0+a_1t+a_2t^2+a_3t^3$, т. е. модели правильной структуры соответствует число 1010. Видно, что в данном варианте все практические критерии, кроме критерия регулярности и комбинированного k_2 , практически работают одинаково. Причем отбираемые по различным критериям структуры также отличаются незначительно.

Предельные уровни помехоустойчивости фильтрации и прогнозирования θ_F и θ_i^* оказываются значительно выше уровня помехоустойчивости выбора структуры θ_i^* : селектируемые по внешним критериям модели сохраняют свойство фильтровать и прогнозировать незашумленный сигнал с удовлетворительной точностью при достаточно высоких уровнях шума — порядка 200—300 %.

Из рассмотрения табл. 8 можно сделать следующие выводы: а) симметричные критерии n_{cm} , d и S ведут себя аналогично идеальному $R(W)$ — при увеличении уровня шума лучшей является сначала модель с точной структурой 1010, затем другие модели с двумя аргументами (1001, 0110 и др.) и, наконец, модель в виде среднего зашумленных значений 1000, причем $R(W)$ «держит» соответствующие структуры, дольше практических критериев; б) критерий S переходит на малоинформативную модель со структурой 1000 позже, чем n_{cm} и d , что подтверждает полученные выше теоретические оценки; в) идеальный критерий $R(C)$ при возрастании шума после точной структуры с $s=2$ начинает выбирать смешанную модель 1101 с $s=3$ и только затем переходит на менее сложные структуры, что согласуется с полученным в гл. 4 выводом о помехоустойчивости соответствующего идеального критерия в случае отдельной реализации шума; г) то же самое относится и к критериям $\Delta(B)$ и k_2 , поскольку они отражают оценки точности моделей на одних частях выборки при коэффициентах, полученных на других частях.

Приведенные результаты относятся к моделированию с применением критерии селекции отдельно друг от друга. Однако известно [49], что одновременное или последовательное применение ряда критериев селекции улучшает качество моделирования.

Помехоустойчивость двухкритериальной селекции. Применение «плоскости двух критериев», т. е. двухкритериальная селекция, является распространенным приемом в различных задачах и алгоритмах самоорганизации моделей на ЭВМ. Из опыта моделирования известно, что при этом должны применяться критерии, имеющие различную природу (например, n_{cm} и $\Delta(B)$) или относящиеся к различным частям выборки данных (например, $\Delta(B)$ и $\Delta(C)$ или n_{cm} и $\Delta(C)$).

Два критерия могут использоваться либо в виде «свертки» типа (5.9), либо последовательно. Как показано выше, свертка мо-

жет привести к ошибкам из-за трудности нормировки (взвешивания) критериев. Эта трудность в алгоритмах МГУА устраняется с помощью последовательного применения критериев: сначала по одному критерию (обычно по n_{cm}) отбирается некоторое число F моделей с наименьшими значениями критерия, из которых затем выбирается лучшая по второму критерию.

Изложим основные результаты последовательного применения различных комбинаций двух критериев. В качестве первого критерия используем $\Delta(B)$, n_{cm} , d или S , в качестве второго — точность на экзаменационной последовательности $\Delta(C)$ (после пересчета коэффициентов на W). Проверка эффективности комбинации критериев $n_{cm} \rightarrow \Delta(B)$ дала отрицательные результаты — помехоустойчивость при этом снижалась.

Рассмотрим сначала, как изменяется предельная помехоустойчивость идентификации структуры модели θ_i^* при двухкритериальном выборе на том же примере (22 точки, второй вариант разделения).

1. Комбинация $\Delta(B) \rightarrow \Delta(C)$: из всех моделей отбираются те, которые имеют показатель регулярности не хуже, чем $5\Delta_{min}(B)$, из них выбирается лучшая по второму критерию $\Delta(C)$. Помехоустойчивость при этом возрастает до $\theta_i^* = 60\%$ (для отдельно взятого критерия $\Delta(B)$ она была нулевой). Того же уровня можно достичь и в остальных вариантах текстовых расчетов.

2. Комбинация $n_{cm} \rightarrow \Delta(C)$: из всех моделей отбираются те, которые имеют показатель смещения не хуже $10n_{min}$, из них выбирается лучшая по $\Delta(C)$. Отметим, что здесь порог вдвое выше, чем в предыдущей паре критериев, поскольку разброс значений критерия n_{cm} всегда больше, чем $\Delta(B)$. Очевидно, что в реальных задачах объем перебора составляет десятки тысяч моделей, поэтому моделей, смещение которых не больше, чем на порядок превосходит минимальное значение n_{min} , может быть слишком много. В связи с этим первичную селекцию ограничивают наперед заданным числом F , связанным с возможностями конкретной ЭВМ.

Применение рассматриваемой пары критериев привело к весьма существенному повышению уровня помехоустойчивости: правильная структура модели является лучшей вплоть до значения $\theta_i^* = 200\%$, в то время как для отдельно взятого n_{cm} она равнялась всего 20 %.

Таким образом, вполне целесообразным является применение (последовательное) двух критериев $n_{cm} \rightarrow \Delta(C)$, причем после отбора по n_{cm} можно применять также формулу комбинированного критерия (5.12). Отметим еще, что использование данной комбинации критериев повышает помехоустойчивость идентификации и в остальных вариантах тестового примера: при одиннадцати точках θ_i^* возрастает с 20 до 80 %, при первом варианте разделения 22-х точек — с 60 (при использовании n_{cm}) до 200 %, при третьем — с 80 до 230 %.

3. Комбинации критериев $d \rightarrow \Delta(C)$ и $S \rightarrow \Delta(C)$ при величине порога $5d_{\min}$ или $5S_{\min}$ дают те же результаты, что и предыдущая пара критериев. При этом следует иметь в виду, что критерий n_{cm} вычисляется с меньшим числом элементарных операций, чем d и S (особенно с применением представления через матрицу нормальных уравнений), что существенно в задачах с большим перебором моделей.

Использование двухкритериальной селекции моделей повышает также уровень помехоустойчивости решения задач фильтрации и прогнозирования незашумленной функции (сигнала): в рассматриваемом варианте тестового примера при переходе от индивидуальных критериев к комбинации $n_{cm} \rightarrow \Delta(C)$ помехоустойчивость фильтрации сохраняется: $\theta_{\phi}^* = 360\%$, а помехоустойчивость прогнозирования возрастает с 130 % до значения $\theta_p^* = 400\%$.

Таким образом, численные исследования подтвердили сделанный ранее вывод о том, что в алгоритмах самоорганизации в качестве индивидуальных необходимо применять критерии симметричного типа, учитывающие оценки коэффициентов, полученные на двух частях выборки исходных данных. На примере селекция моделей оказалась помехоустойчивой при уровнях шума примерно до 100 % в задаче идентификации структуры и до 200—300 % — в задачах фильтрации и прогнозирования сигнала. Обычный критерий регулярности при моделировании в условиях помех менее эффективен из-за существенной зависимости от варианта разделения данных.

Помехоустойчивость селекции моделей существенно повышается при применении двухкритериальной поэтапной селекции, когда предварительный отбор выполняется по симметричному критерию, а лучшая модель определяется по точности на экзаменационной последовательности. При этом на тестовом примере достигнута правильная идентификация структуры при уровнях шума порядка 200 % и удовлетворительная фильтрация и прогнозирование сигнала при уровнях шума до 400 %. По-видимому, повышение помехоустойчивости связано с использованием в процессе селекции информации уже не двух, а трех частей выборки A , B и C .

Можно ожидать, что правильный выбор критериев и последовательности их применения позволит осуществлять помехоустойчивое моделирование при отношении помеха/сигнал порядка десяти.

Дальнейшее повышение помехоустойчивости достигается применением многоуровневых схем и специальных алгоритмов, реализующих принципы теории связи. Напомним, что в современных системах связи помеха может превосходить регулярный сигнал в тысячи раз, однако при этом известен класс передаваемых сигналов и применяется высокоеффективное кодирование, направленное на борьбу с известными типами помех. Подобных условий при моделировании, например, природных экологических процессов, как правило, нет — неизвестно, какой незашумленный сигнал может

«передавать» Байкал, измеряем ли мы все существенные переменные и какие помехи присутствуют в данных измерений. Однако помехоустойчивые методы самоорганизации моделей помогают получать оправдывающиеся прогнозы даже в таких неблагоприятных условиях [75].

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ СРАВНИТЕЛЬНОГО ИССЛЕДОВАНИЯ ВНЕШНИХ КРИТЕРИЕВ

Для построения модели оптимальной сложности принципиально необходимо применять внешние критерии селекции. Свойством «внешности» обладают такие количественные критерии выбора моделей, минимум которых на множестве сравниваемых моделей существует и единствен. Критерии селекции, применяемые в алгоритмах самоорганизации, используют только ту информацию, которая заключена в выборке наблюдений за моделируемым объектом. При этом их внешний характер обеспечивается с помощью разделения всей выборки на две и более частей (о способах регуляризации разделения выборки см. [20, 128, 178]).

В настоящей главе приведена классификация основных внешних критериев селекции, применяемых в алгоритмах МГУА. В основу классификации положены следующие характеристики критериев: квадратичный или корреляционный характер; учет определенных требований, предъявляемых к моделям (точность либо согласованность оценок); симметрия использования информации частей выборки данных. Такое представление внешних критериев позволяет глубже понять их сущность, полнее выявить их сходство и различие, а также способствует целенаправленному конструированию новых форм критериев. Приведены также достаточно простые и компактные формулы связи критериев, пригодные как для использования на практике в алгоритмах самоорганизации, так и для теоретического сравнительного анализа различных критериев.

По аналогии с предыдущей главой разработана методика анализа помехоустойчивости квадратичных внешних критериев с помощью исследования их математических ожиданий — т. е. «в среднем» по всем возможным реализациям вектора нормального аддитивного шума. Установлено, что все квадратичные внешние критерии обладают одинаковыми свойствами: в отсутствие шума равны нулю только для истинной модели; при ненулевом шуме имеют единственный минимум, указывающий модель оптимальной сложности; при повышении уровня шума этот минимум смещается и указывает на все более простые модели. Как было установлено в гл. 4, аналогичными свойствами обладает и идеальный критерий R^2 , минимум которого соответствует модели с максимальной (предельно достижимой) помехоустойчивостью, т. е. модели, имеющей минимальную дисперсию ошибки фильтрации или прогноза «истинного сигнала» в условиях помех. Поэтому указанные свойства следует считать принципиально необходимыми для эффективности

любого критерия, применяемого в методах структурного моделирования.

В данной главе получены сравнительные оценки помехоустойчивости некоторых внешних критериев. Например, симметричные критерии «ранжированы» по степени помехоустойчивости следующим образом: стабильность (лучший), регулярность, минимум смещения. Это означает, что в среднем при повышении уровня шума некоторую модель (например, истинную) теряет сначала критерий $n^2_{\text{см}}$, затем d^2 и последним S^2 . Невысокая помехоустойчивость критерия минимума смещения требует осторожности при его индивидуальном применении. Однако сама идея, заложенная в этот критерий — требование согласованности или непротиворечивости оценок коэффициентов, полученных на различных частях выборки — является вполне естественной, и модели, получаемые в результате самоорганизации, с необходимостью должны обладать этим «свойством непротиворечивости». Поэтому критерий $n^2_{\text{см}}$ рекомендуется использовать не индивидуально, а в комбинации с другими — например, в виде свертки или при последовательном применении. В то же время другие симметричные критерии d^2 и S^2 можно применять индивидуально, причем они органично удовлетворяют свойству непротиворечивости, так как непосредственно выражаются через $n^2_{\text{см}}$.

Приведены результаты численного эксперимента по исследованию помехоустойчивости моделирования по COMBI с применением различных внешних критериев. Результаты иллюстрируют теоретические оценки, полученные в четвертой и пятой главах, и демонстрируют высокую помехоустойчивость алгоритмов самоорганизации. В частности, показано, что критерий регулярности весьма критичен к разделению выборки данных, поэтому его не следует применять для структурного моделирования в условиях больших помех. Показана также высокая помехоустойчивость моделирования по комбинаторному алгоритму при последовательном применении критериев минимума смещения (или другого симметричного критерия) и точности на экзаменационной части выборки.

ПОМЕХОУСТОЙЧИВОЕ МОДЕЛИРОВАНИЕ ПРИ НАЛИЧИИ ИНФОРМАЦИИ О КОРРЕЛЯЦИОННЫХ, ЧАСТОТНЫХ И ВРЕМЕННЫХ ХАРАКТЕРИСТИКАХ ПОМЕХ

Разнообразные практические задачи моделирования и прогнозирования сложных экологических, экономических, технических и других процессов и объектов, решенные с помощью индуктивных методов самоорганизации моделей на ЭВМ [49, 79], указывают на высокую эффективность этих методов в условиях помех. Помехоустойчивость алгоритмов МГУА проявляется не только при моделировании по данным, в которых присутствует шум измерений, но и в условиях неизмеряемых возмущений, при неполном информационном базисе и т. д. Изложенные в предыдущих главах результаты теоретических и численных исследований подтверждают и объясняют факт помехоустойчивости алгоритмов МГУА.

Однако в задачах с очень высоким уровнем шума (больших отношениях мощностей шум/сигнал) эта эффективность часто теряется. В таких случаях необходимо применять комбинированный подход — учитывать при решении «плохих» задач как можно больше дополнительной информации о структуре объекта, о взаимосвязи характеристических переменных, свойствах помех и др.

В данной главе рассмотрены особенности и способы решения задач моделирования при наличии информации об одной из следующих характеристик помехи: корреляционная функция, амплитудно-частотная характеристика, временное распределение. При этом анализируются процессы, для которых одним из главных параметров (аргументов) является время, т. е. аналоги сигналов, являющихся предметом общей теории связи. Последовательность измерений значений такого процесса y_k , $k=1, \bar{N}$, будем называть временным рядом, или зашумленным сигналом. Как и раньше, будем исходить из того, что измеренные значения y_k представляют собой сумму точного (регулярного) сигнала u_k и случайной помехи ξ_k , $y_k = u_k + \xi_k$.

1. ПОМЕХОУСТОЙЧИВЫЙ АЛГОРИТМ САМООРГАНИЗАЦИИ МОДЕЛИ ПО ПРИНЦИПУ КОРРЕЛЯЦИОННОГО ПРИЕМНИКА

Одним из наиболее эффективных направлений переноса идей теории информации в теорию самоорганизации моделей является построение алгоритмов, использующих принципы действия приемни-

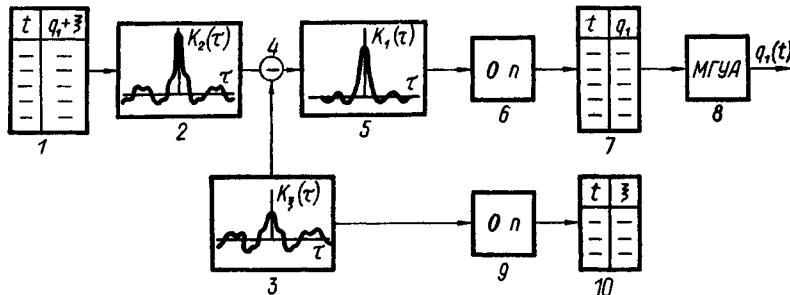


Рис. 19. Схема помехоустойчивого алгоритма самоорганизации модели по принципу корреляционного приемника.

ков с повышенной помехоустойчивостью: корреляционного, взаимокорреляционного, с синхронным детектированием, с накоплением сигнала и др. [100]. Повышение помехоустойчивости алгоритмов моделирования, достигаемое при этом, является крайне необходимым для дальнейшего развития теории самоорганизации [60].

В качестве примера опишем схему корреляционного приемника и основанные на ней алгоритмы моделирования.

Корреляционный помехоустойчивый приемник требует априорного задания общего вида корреляционной функции помех $K_\xi(\tau)$ и наличия тех или иных способов обратного преобразования корреляционных функций в исходную реализацию случайного процесса. Способ преодоления некорректности (регуляризации) обратного преобразования поясняется в § 2. Здесь же полагаем, что мы имеем надежный алгоритм обратного перехода. Тогда схему корреляционного приемника можно описать так. При помощи коррелометра находится корреляционная функция общего зашумленного сигнала $K_2(\tau)$. Так как шум и сигнал не коррелированы между собой, то, вычитая корреляционную функцию шума из корреляционной функции суммарного сигнала, находим корреляционную функцию чистого сигнала без помех: $K_1(\tau) = K_2(\tau) - K_\xi(\tau)$. Обратное преобразование от корреляционной функции к сигналу дает сигнал, освобожденный от помех: $K_1(\tau) \rightarrow q_1(t)$.

В теории связи обратное преобразование еще не применялось: вместо него используются линии задержки, позволяющие отсеять быстро затухающую корреляционную функцию шума от незатухающей корреляционной функции сигнала.

Схема алгоритма моделирования представлена на рис. 19, где 1 — исходная таблица зашумленных данных; 2 — корреляционная функция зашумленных данных; 3 — корреляционная функция помех (задается априори); 4 — элемент вычитания ординат; 5 — корреляционная функция чистого сигнала; 6 — обратное преобразование; 7 — таблица незашумленных данных $q_1(t)$; 8 — алгоритм МГУА для построения гармонической модели; 9 — обратное преобразование; 10 — таблица помехи $\xi(t)$.

Основное затруднение, по-видимому, состоит в выборе корреляционной функции помехи, которая должна быть задана. Ее вы-

бор можно оптимизировать при помощи перебора вариантов по критерию точности прогноза, по конкретным данным, и таким образом расширить действие алгоритма самоорганизации модели.

Алгоритм МГУА с самоорганизацией корреляционной функции помехи. Так как сигнал и помеха независимы друг от друга, независимыми будут и их корреляционные функции:

$$K_2(\tau) = K_1(\tau) + K_\xi(\tau). \quad (6.1)$$

Прямое преобразование сводится к расчету корреляционных функций:

$$q_2(t) \rightarrow K_2(\tau), \quad q_1(t) \rightarrow K_1(\tau), \quad \xi(t) \rightarrow K_\xi(\tau).$$

Обратное преобразование представляет собой следующие операции (выполняемые при помощи решения систем нелинейных уравнений):

$$K_2(\tau) \rightarrow q_2(t), \quad K_1(\tau) \rightarrow q_1(t), \quad K_\xi(\tau) \rightarrow \xi(t).$$

Функции $q_2(t)$ и $K_2(\tau)$ рассчитываются по таблице исходных данных. О функции $K_\xi(\tau)$ известно, что она является монотонной функцией времени вида

$$K_\xi(\tau) = K_0 e^{-\frac{\tau}{\tau_n}}. \quad (6.2)$$

При достаточно большом сдвиге $K_\xi(\tau) = K_2(\tau)$, что дает возможность определить постоянную времени τ_n . Следовательно, для самоорганизации функции $K_\xi(\tau)$ требуется найти только ее максимальное значение K_0 (все функции — ненормированные).

Подбор величины K_0 выполняется по критерию точности краткосрочного прогноза:

$$\Delta^2(C) = \frac{\sum (q_\phi - q_{\text{пр}})^2}{\sum (q_\phi - \bar{q}_\phi)^2}.$$

Величина K_0 постепенно увеличивается, начиная с нуля, пока ошибка $\Delta(C)$ падает. Схема расчета включает в себя такие операции: 1) определение $K_2(\tau)$; 2) задание ряда значений K_0 ; 3) построение соответствующего ряда функций $K_1(\tau)$ и $K_\xi(\tau)$; 4) обратный переход от $K_1(\tau)$ к прогнозу $q_1(t)$; 5) расчет ошибок $\Delta(C)$ и определение минимальной по K_0 ; 6) выбор оптимального значения K_0 и тем самым выбор прогноза $q_{\text{пр}}(t) = q_1(t)$ при $K_0 = K_{\text{опт}}$.

Задачу обратного преобразования от корреляционной функции к исходному сигналу можно свести к решению некоторой системы нелинейных алгебраических уравнений в сочетании с процедурой регуляризации. Как известно, нелинейные системы уравнений в отличие от линейных имеют множество решений. Поэтому в данном случае возникают две задачи: 1) поиск множества решений нелинейной системы уравнений; 2) применение регуляризации для выбора нужного решения из этого множества с помощью критериев,

отражающих дополнительные данные об исследуемом процессе. Решение этих задач рассмотрено в работе [11], которой мы будем следовать.

2. СВЕДЕНИЕ ЗАДАЧИ ОБРАТНОГО ПРЕОБРАЗОВАНИЯ К РЕШЕНИЮ СИСТЕМЫ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотрим случайный процесс, заданный в виде выборки $y^t = (y_1, y_2, \dots, y_N)$. Дискретную корреляционную функцию, соответствующую этому процессу, можно вычислить, пользуясь выражением

$$K(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \overset{\circ}{y_i} \overset{\circ}{y_{i+\tau}}, \quad \tau = \overline{0, N-1}, \quad (6.3)$$

где $\overset{\circ}{y}_i = y_i - \bar{y}$ — отклонение от среднего; τ — сдвиг сравниваемых сигналов по времени. Из выражения (6.3) получим N значений корреляционной функции: $K_\tau = K(\tau)$. При этом предполагаем, что указанные величины не зависят от конкретной выборки сигналов и сохраняются постоянными для всего процесса.

Возьмем теперь отличную от y выборку того же процесса, например, такую: $y'^t = (y'_1, y'_2, \dots, y'_N)$. Так как по предположению значения коэффициентов корреляционной функции должны оставаться постоянными, то можно составить, используя выражение (6.3), N уравнений, которые связывают значения сигналов на новой выборке (на новом временном интервале):

$$\frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \overset{\circ}{y'_i} \overset{\circ}{y'_{i+\tau}} - K_\tau = 0, \quad \tau = \overline{0, N-1}. \quad (6.4)$$

Например, для $N = 5$ и процесса с нулевым средним ($\bar{y} = 0$) эти уравнения имеют вид

$$\begin{aligned} &\frac{1}{5} (y'_1^2 + y'_2^2 + y'_3^2 + y'_4^2 + y'_5^2) - K_1 = 0, \\ &\frac{1}{4} (y'_1 y'_2 + y'_1 y'_3 + y'_1 y'_4 + y'_1 y'_5) - K_2 = 0, \\ &\frac{1}{3} (y'_1 y'_3 + y'_2 y'_4 + y'_3 y'_5) - K_3 = 0, \\ &\frac{1}{2} (y'_1 y'_4 + y'_2 y'_5) - K_4 = 0, \quad y'_1 y'_5 - K_5 = 0, \end{aligned} \quad (6.5)$$

где оценки ординат корреляционной функции

$$K_1 = \frac{1}{5} (y_1^2 + y_2^2 + y_3^2 + y_4^2 + y_5^2),$$

$$K_2 = \frac{1}{4} (y_1 y_2 + y_2 y_3 + y_3 y_4 + y_4 y_5),$$

$$K_3 = \frac{1}{3} (y_1 y_3 + y_2 y_4 + y_3 y_5),$$

$$K_4 = \frac{1}{2} (y_1 y_4 + y_2 y_5), \quad K_5 = y_1 y_5$$

расчитаны по выборке y .

Решение нелинейной системы уравнений (6.4) неединственно, т. е. в пространстве переменных имеется некоторое конечное множество $k \leq 2N$ точек, координаты которых удовлетворяют рассматриваемым уравнениям. Каждая такая точка соответствует некоторой реализации случайного процесса, причем все эти реализации имеют одни и те же значения коэффициентов корреляционной функции. В число указанных входит также искомая реализация y' . Чтобы ее выделить, необходима дополнительная информация о процессе (критерий регуляризации).

Критерий регуляризации — «скользящее среднее». Возможны различные типы таких критериев. В статье [11] в качестве критерия регуляризации принята временная функция скользящего среднего —

$$H_j = \overset{\circ}{y}_{lj} = \frac{1}{l} \sum_{i=j}^{l+j-1} \overset{\circ}{y}_i, \quad j = \overline{1, N-l}, \quad (6.6)$$

где $l < N$ — скользящий интервал усреднения, например $l = N/2$. Преобразуя все k найденных решений системы (6.4) согласно (6.6), получим k новых временных зависимостей. Аналогичная временная зависимость (эталонная) снята также для заданной выборки $H_j = y_{lj}$. То решение системы (6.4), для которого полученная зависимость будет наиболее близка к эталонной, принимаем в качестве искомого.

Наибольшую трудность при изложенном подходе к решению задачи обратного преобразования представляет нахождение множества действительных решений системы уравнений (6.4). Необходимо рассчитывать каждое решение этой системы за достаточно малое время и с заданной точностью. Однако соблюдение этих условий усложняется вследствие того, что для некоторых точек пространства переменных возможна плохая обусловленность решаемой системы уравнений. Последнее связано с линейной зависимостью векторов — градиентов функций невязок отдельных уравнений. В связи с этим в [11] рассмотрен итерационный метод решения нелинейных уравнений, основанный на ортогональном преобразовании соответствующей линеаризованной системы. Такое предварительное преобразование позволяет получить относительно простую расчетную процедуру и, кроме того, повышает точность решения, особенно при плохой обусловленности.

Метод решения системы уравнений типа (6.4), изложенный в [11], основан на результатах [8, 9] и сводится к итерационному решению линеаризованной системы, которое оказывается ближайшим к заданному начальному приближению y^0 . Для нахождения

других решений требуется изменять каким-либо образом начальные приближения. В [11] имеется описание и примеры применения разработанного алгоритма, достаточно простого в реализации.

3. ВЛИЯНИЕ ЧАСТОТНЫХ ПОМЕХ НА РЕЗУЛЬТАТ МОДЕЛИРОВАНИЯ

Рассмотрим особенности решения задачи моделирования в условиях, когда данные измерений y_k представляют собой смесь сигнала с частотной помехой z_k , $k=1, 2, 3, \dots$:

$$y_k = u_k + z_k = u_k + \sum_{i=1}^m (A_i \sin \omega_i k + B_i \cos \omega_i k). \quad (6.7)$$

В общем случае такая помеха представляет собой сумму m гармонических составляющих. Относительно такой помехи может быть известна следующая информация: «полоса частот», т. е. интервал $[\omega_{\min}, \omega_{\max}]$, в котором находятся все частотные составляющие; весь набор m отдельных частот; точная амплитудно-частотная характеристика (АЧХ). По-видимому, на практике возможны только первые два случая, но для выяснения сути дела стоит рассмотреть именно идеальный третий случай.

Очевидно, что при заданной АЧХ помехи z_k сама эта помеха фактически известна и проблемы борьбы с ней нет: достаточно осуществить операцию $u_k = y_k - z_k$, чтобы восстановить точный сигнал. Однако интересно рассмотреть такой вопрос: можно ли с учетом такой информации построить алгоритм моделирования, который, оперируя обычным образом с y_k , мог бы восстановить u_k ? Для этого необходимо попытаться сконструировать некий «частотный критерий», который, «зная» АЧХ помехи, из всех моделей-претендентов, построенных по зашумленным данным, выбирает ту, которая лучше других восстанавливает сигнал, т. е. фильтрует помеху.

Рассмотрим предельно простой пример заданной синусоидальной помехи и линейно зависящего от времени сигнала $u_k = bk$:

$$y_k = u_k + h \sin \omega k = bk + h \sin \omega k, \quad k = \overline{1, N}. \quad (6.8)$$

Отметим сразу же специфику дискретного отсчета всех рассматриваемых функций времени: при заданной дискретизации времени $k=1, 2, 3, \dots$ через точки на заданной синусоиде с частотой ω можно провести сколь угодно много синусоид с большими частотами, но отличить их от исходной никак нельзя. Этот эффект известен в задачах цифровой фильтрации [129, 136] при восстановлении зашумленных сигналов с помощью обработки на ЭВМ данных дискретных измерений непрерывных сообщений. При этом фактически достаточно рассматривать некие «эффективные» гармонические составляющие, удовлетворяющие при заданной дискретизации времени теореме Котельникова о дискретном представлении непрерывной функции.

Итак, полностью зная помеху, попытаемся точно восстановить сигнал, обрабатывая по МНК зашумленные измерения y_k , $k=\overline{1, N}$. Пусть известна также «точная структура» сигнала; тогда

$$\hat{y}_k = \hat{a}k, \quad \hat{a} = \sum_{k=1}^N y_k k / \sum_{k=1}^N k^2,$$

или, с учетом (6.2),

$$\hat{a} = b + h \sum_{k=1}^N k \sin \omega k / \sum_{k=1}^N k^2. \quad (6.9)$$

Зная h и ω , вычисляем b и соответственно точный сигнал $u_k = bk$. Однако то же самое можно было получить прямо из (6.8), т. е. новой информации в результате оценивания \hat{a} мы не получили.

Допустим теперь, что нам не известна амплитуда помехи h , а только частота ω . По-видимому, в этом случае можно, меняя длину выборки N , найти такое N^* , когда $\sum_{k=1}^{N^*} k \sin \omega k = 0$ (при этом $\hat{a} = b$), поскольку $k \sin \omega k$ — это возрастающая синусоида, амплитуда которой ограничена асимптотами $\pm k$. Но при этом, как видно из (6.8) — (6.9), мы должны знать, кроме частоты помехи, также точную структуру модели сигнала (т. е. что $u_k \sim k$), а это уже выходит за рамки наших предположений. И перебор частных моделей вида $y_k = a_1 k$, $y_k = a_2 k^2$, $y_k = a_3 k^3$ и т. д. не даст результата — не зная u_k , нельзя сконструировать внешний критерий, который помог бы выполнить перебор как структур, так и всевозможных длин выборки и восстановить таким образом точную модель u_k . Более того, если точный сигнал и задан, но является более сложной функцией времени, то задача подбора длины выборки, дающей точную оценку коэффициентов, уже неразрешима. Действительно, пусть $u_k = b_1 k + b_2 k^2$, тогда, переходя к векторным обозначениям

$$u^T = (u_1, \dots, u_N), \quad y^T = (y_1, \dots, y_N),$$

$$X^T = \begin{pmatrix} 1, \dots, k, \dots, N \\ 1, \dots, k^2, \dots, N^2 \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix},$$

$$z^T = h z_0^T = h (\sin \omega, \dots, \sin \omega k, \dots, \sin \omega N),$$

получаем оценку МНК для \hat{a} :

$$\hat{a} = (X^T X)^{-1} X^T y = b + h (X^T X)^{-1} X^T z_0.$$

При этом аналогичное рассмотренному выше условие точного оценивания $\hat{a} = b$ имеет вид $X^T z_0 = 0$, т. е. матрица аргументов X должна быть ортогональной вектору z_0 . Однако подбором длины

выборки N этого сделать здесь уже нельзя, поскольку ни при каком N нельзя получить $\sum_{k=1}^N k \sin \omega k = \sum_{k=1}^N k^2 \sin \omega k = 0$.

Итак, даже в случае синусоидальной помехи не удается сконструировать алгоритм моделирования, который с учетом только частоты синусоиды в процессе самоорганизации фильтрует эту помеху и восстанавливает сигнал в виде полиномиальной функции времени. Очевидно, что то же справедливо и для случая, когда задана не одна частота, а полоса частот. Единственный выход — фильтрировать помеху *до моделирования*, используя достаточно хорошо разработанные методы синтеза цифровых фильтров с заданными свойствами — полосовые, заградительные и т. д. [129, 136, 165], для построения которых нет необходимости точно знать АЧХ помехи, достаточно располагать только предполагаемым интервалом частот $[\omega_{\min}, \omega_{\max}]$. В следующем параграфе будет рассмотрено применение простейшего вида фильтрации частотных помех с помощью слаживания, т. е. усреднения соседних точек на «скользящем» интервале. С более сложными и эффективными методами цифровой фильтрации можно познакомиться в указанных монографиях, причем в [127, 136] имеются программы для проектирования фильтров на ЭВМ.

Целесообразно исследовать также задачу моделирования динамических процессов, например, в виде авторегрессионных моделей. Не проявится ли здесь эффект фильтрации частотных помех, поскольку авторегрессионная модель — это частный случай фильтра?

Общая формула для конечного фильтра имеет вид [165]

$$y_k = \sum_{i=-M}^M c_i x_{k-i} + \sum_{i=1}^m a_i y_{k-i}$$

(при $M = \infty, m = \infty$ фильтр называется бесконечным).

Если не все a_i равны нулю, такой фильтр называется рекурсивным, если же $a_i \equiv 0$ — нерекурсивным. Если в фильтре не используются значения x_{k-i} «из будущего», то такой фильтр называется физически реализуемым:

$$y_k = \sum_{i=1}^M c_i x_{k-i} + \sum_{i=1}^m a_i y_{k-i}.$$

Если все $c_i \equiv 0$, приходим к обычному авторегрессионному уравнению, которое, таким образом, является вырожденным рекурсивным фильтром

$$y_k = \sum_{i=1}^m a_i y_{k-i}, \quad (6.10)$$

устанавливающим связь между m прошлыми и одним $(m+1)$ -м будущим значениями процесса.

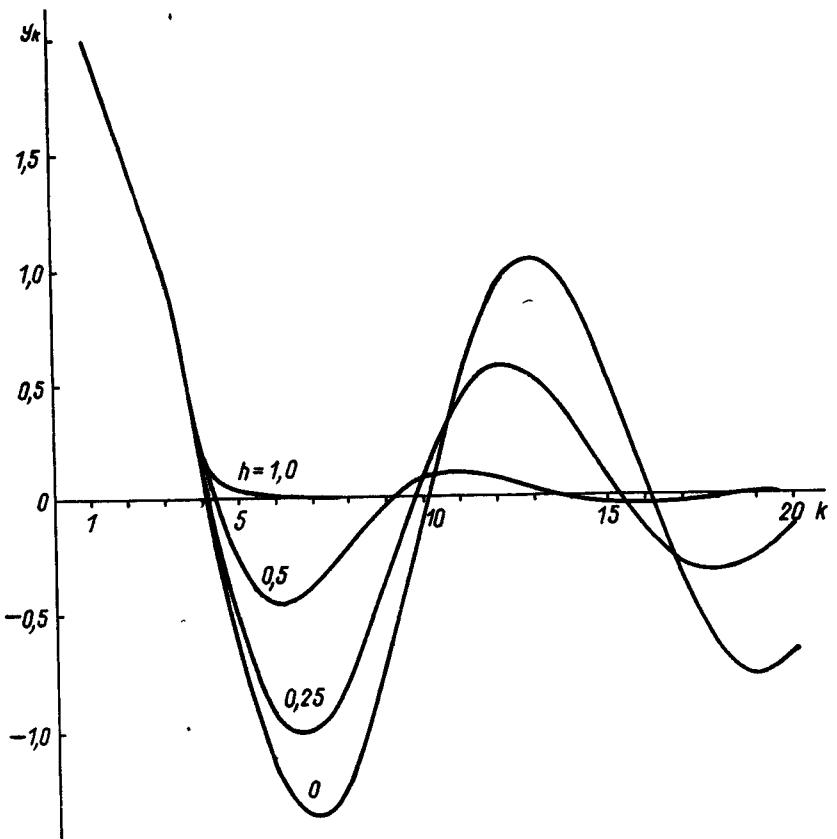


Рис. 20. Интегральные кривые лучших моделей, построенных при различных амплитудах h синусоидальной помехи ($f=0,375$).

Построим АЧХ для частного случая такого фильтра с тремя запаздываниями:

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + a_3 y_{k-3}.$$

Следуя [165], для получения АЧХ подставим в последнее уравнение $y_k = e^{i\omega k}$, причем, поскольку характеристика не зависит от индексации времени, примем $k = 3, \dots, k-3 = 0$:

$$A(\omega) e^{i\omega 3} = a_1 e^{i\omega 2} + a_2 e^{i\omega} + a_3.$$

Разделяя действительную и мнимую части, получаем

$$\operatorname{Re} A(\omega) \cos 3\omega = a_1 \cos 2\omega + a_2 \cos \omega + a_3,$$

$$\operatorname{Im} A(\omega) \sin 3\omega = a_1 \sin 2\omega + a_2 \sin \omega.$$

Отсюда вычисляем АЧХ $A^2(\omega)$ (применяя известные формулы тригонометрии):

$$A^2(\omega) = (\operatorname{Re} A(\omega))^2 + (\operatorname{Im} A(\omega))^2 = a_1^2 + a_2^2 + a_3^2 + 2a_1 a_3 \cos 2\omega + 2a_2(a_1 + a_3) \cos \omega. \quad (6.11)$$

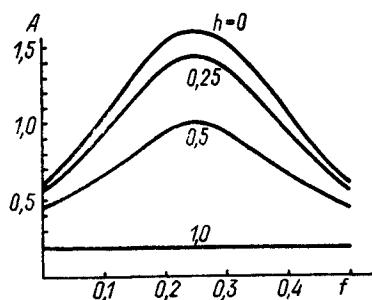


Рис. 21. АЧХ лучших моделей при различных h .

Решение уравнения (6.12) изображено на рис. 20 (при $h=0$), соответствующая ему АЧХ — на рис. 21, причем амплитудно-частотная характеристика, как это принято в работе [165], представлена только для главного интервала частот $f \in [0; 0,5]$ (при больших частотах она зеркально повторяется), где $f = \frac{1}{T} = \frac{\omega}{2\pi}$ — циклическая частота.

Пример. Рассмотрим вычислительный эксперимент по восстановлению сигнала (6.12), зашумленного частотной помехой.

$$y_k = 1,1 u_{k-1} - 0,5 u_{k-3} + h \sin fk,$$

Таблица 9. Данные для моделирования с частотной помехой

k	u_k	Значения $y_k = u_k + h \sin 2\pi fk$			
		$f=0.25$		$f=0.375$	
		$h=0.25$	$h=0.5$	$h=0.25$	$h=0.5$
1	2,0	2,0	2,0	2,0	2,0
2	1,5	1,5	1,5	1,5	1,5
3	1,0	1,0	1,0	1,0	1,0
4	0,1	0,1	0,1	0,1	0,1
5	-0,64	-0,390	-0,140	-0,463	-0,286
6	-1,204	-1,204	-1,204	-1,454	-1,704
7	-1,374	-1,624	-1,874	-1,198	-1,021
8	-1,192	-1,192	-1,192	-1,192	-1,192
9	-0,709	-0,459	-0,209	-0,886	-1,063
10	-0,093	-0,093	-0,093	0,157	0,407
11	0,494	0,244	-0,006	0,317	0,140
12	0,898	0,898	0,898	0,898	0,898
13	1,034	1,284	1,534	1,211	1,388
14	0,890	0,890	0,890	0,640	0,390
15	0,531	0,281	0,031	0,707	0,884
16	0,067	0,067	0,067	0,067	0,067
17	-0,372	-0,122	0,128	-0,549	-0,726
18	-0,674	-0,674	-0,674	-0,424	-0,174
19	-0,775	-1,025	-1,275	-0,952	-1,129
20	-0,667	-0,667	-0,667	-0,667	-0,667

Для примера получим $A^2(\omega)$ для такого конкретного авторегрессионного уравнения:

$$u_k = 1,1 u_{k-1} - 0,5 u_{k-3}, \quad (6.12)$$

решение которого с начальными условиями $u_0=2$, $u_1=1,5$, $u_2=1,0$ является затухающим колебательным процессом. АЧХ этого уравнения (фильтра), согласно (6.11):

$$A^2(\omega) = 1,46 - 1,1 \cos 2\omega.$$

при различных амплитудах h и частотах f помехи, причем на начальные условия помеха не накладывалась. Модель ищется по комбинаторному алгоритму МГУА с интегральным критерием $i^2(N)$ (см. гл. 5) по зашумленным данным y_k , $k=1, N$, среди членов полной модели

$$y_k = a_0 + a_1 y_{k-1} + a_2 y_{k-2} + a_3 y_{k-3}.$$

Исходные данные для самоорганизации модели при различных h и f , $N=20$, представлены в табл. 9. Коэффициенты моделей различной сложности оценивались по МНК на всех двадцати точках данных с помехами. Лучшие по $i^2(W)$ модели при различных h и $f=0,375$ приведены в табл. 10. Частота $f=0,375$ взята для того, чтобы помеха была «высокочастотной» по отношению к собственной частоте сигнала $f_0=0,083$ ($T=12$).

Решения лучших авторегрессионных моделей (из табл. 10) представлены на рис. 20, соответствующие им частотные характеристики — на рис. 21. Видно, что близости интегральных кривых к точному сигналу на рис. 20 однозначно соответствует близость АЧХ к характеристике сигнала без помехи. Аналогичный характер имеют результаты при $f=0,25$. Это означает, что частные модели, проверяемые при моделировании по МГУА, достаточно характеризовать или критерием точности восстановления сигнала, или критерием близости к истинной АЧХ — эти два критерия не дополняют, а повторяют друг друга. Следовательно, эффективный «частотный» критерий сконструировать нельзя.

Из рассмотренного выше следуют два интересных вывода. Во-первых, при $h=0$ (при точных данных) применение МНК к разностной модели дает точные оценки коэффициентов, хотя критерий $e^2(W)$ и неадекватен такой модели; однако при $h>0$ оценки МНК уже не соответствуют специфике разностных моделей (подробнее об этом см. [21]). Во-вторых, оказывается, что и для разностных моделей имеет место «упрощение» структур при повышении уровня h помех.

Итак, авторегрессионная модель, являясь теоретически фильтром, не фильтрует синусоидальную помеху, поскольку АЧХ модели (6.12) является «колоколообразной» и везде положительной, т. е. не имеет интервалов подавления помех (интервалов с нулевой или близкой к нулю амплитудой). По-видимому, это специфика данных вырожденных фильтров.

Следовательно, для самоорганизации алгебраических и разностных моделей необходима предварительная фильтрация частотных помех, имеющихся в данных наблюдений до моделирования по МГУА, т. е. выделение трендапроцесса.

Таблица 10. Коэффициенты лучших моделей ($f=0,375$)

h	a_0	a_1	a_2	a_3
0	0	1,100	0	-0,500
0,25	0	0,991	0	-0,433
0,5	0	0,720	0	-0,265
1,0	0	0,187	0	0

4. ПРЕДВАРИТЕЛЬНАЯ ФИЛЬТРАЦИЯ ЧАСТОТНЫХ ПОМЕХ С ПОМОЩЬЮ СКОЛЬЗЯЩЕГО УСРЕДНЕНИЯ

Наиболее простым способом цифровой фильтрации помех, присутствующих в дискретном временном ряде измерений y_k , $k=1, N$, изучаемого процесса, является усреднение его значений в нескольких соседних точках, т. е. на скользящем интервале усреднения заданной длины l . Рассмотрим усреднение в центральной точке интервала (симметричное усреднение):

$$\bar{y}_k = \frac{1}{l} \sum_{i=k-l_0}^{k+l_0} y_i, \quad l_0 = \frac{1}{2}(l-1). \quad (6.13)$$

При этом полная длина интервала усреднения $l=2l_0+1$ всегда является нечетным числом: суммируются значения в точке k , а также в l_0 предыдущих и l_0 последующих точках, т. е. может быть $l=3, 5, 7, \dots$.

Для фильтра (6.13) можно построить АЧХ [165]. Например, в случае $l=3$ она имеет вид

$$A(\omega) = 1 + 2 \cos \omega. \quad (6.14)$$

Эта характеристика, а также АЧХ для усредняющих фильтров с $l=5$ и $l=7$ изображены на рис. 22. Видно, что приведенные на рисунке характеристики имеют соответственно одну (при $f^* = \frac{\omega^*}{2\pi} = \frac{1}{3}$), две и три точки пересечения с осью ординат, т. е. для некоторых частот амплитуды усредняющего фильтра равны нулю — эти частоты будут отсутствовать в усредненном ряде \bar{y}_k (6.13). В общем случае АЧХ фильтра имеет $l_0 = \frac{1}{2}(l-1)$ «точек нечувствительности», в которых частотная помеха соответствующей частоты ω_j^* , $j=1, l_0$, при любой ее амплитуде полностью отфильтровывается. При этом с ростом порядка l фильтра АЧХ по своему характеру приближается к δ -функции, поскольку амплитуды для более высоких частот уменьшаются относительно амплитуд более низких частот. Таким образом, при низкочастотном сигнале усредняющий дискретный фильтр позволяет бороться с высокочастотными помехами: если задана частота синусоидальной помехи $h \sin \omega k$, то подбором порядка l ее можно в принципе отфильтровать полностью при любой амплитуде h ; если задана полоса частот помехи, то повышением порядка фильтра можно добиться максимального ослабления всех помех в этой полосе частот. Однако следует иметь в виду очевидные недостатки этих про-

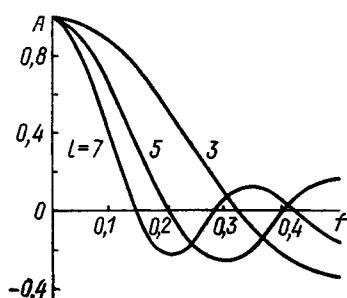


Рис. 22. АЧХ фильтра при различных интервалах l усреднения.

стых усредняющих фильтров: во-первых, при увеличении длины l интервала усреднения сокращается число точек «отфильтрованных» данных $N^* = N - l$ и увеличение l допустимо, пока число N^* остается достаточно большим для дальнейшего моделирования; во-вторых, происходит некоторое искажение характера процесса, прежде всего амплитудное (на рис. 22 амплитудные характеристики приведены в относительных единицах).

Рассмотрим подробнее, как искажается характер процесса (сигнала) при скользящем усреднении для нескольких типичных случаев моделирования (различных видов функций u_k). Предварительно обратим внимание на то, что критические частоты ω_j^* , $j=1, l_0$, усредняющего фильтра не зависят от конкретного вида моделируемой незашумленной функции u_k , т. е. помеха с частотой ω_j^* отфильтровывается в любом случае. Поэтому очень важно представить себе, что происходит с функцией u_k при скользящем усреднении — ведь затем результат усреднения \bar{u}_k (таблица усредненных данных) моделируется с помощью COMBI и полученная модель оптимальной сложности \bar{u}_k^* должна быть максимально близка к незашумленному исходному сигналу u_k .

Линейные функции. Линейная функция времени $u_k = bk$ при скользящем усреднении не искажается совершенно:

$$\bar{u}_k = \frac{1}{l} \sum_{i=k-l_0}^{k+l_0} u_i = \frac{b}{l} \sum_{i=k-l_0}^{k+l_0} i = \frac{b}{l} kl = bk = u_k.$$

Поэтому, если нам заданы зашумленные измерения $y_k = bk + h \sin \omega k$ и известна частота помехи ω , то можно подобрать порядок l фильтра таким, чтобы одна из его критических частот $\omega_j^* = \omega$, и полностью восстановить точный сигнал независимо от амплитуды h помехи. Конечно, при ограниченном числе данных N мы можем таким способом отыскать такое l , при котором помеха подавляется максимально, но не полностью (при этом $\omega_j^* \approx \omega$).

Нелинейные алгебраические функции времени. Для самой простой квадратичной функции $u_k = bk^2$ получаем

$$\begin{aligned} \bar{u}_k &= \frac{b}{l} \sum_{i=k-l_0}^{k+l_0} i^2 = \frac{b}{l} \sum_{i=-l_0}^{l_0} (k^2 + i^2) = \frac{b}{l} k^2 l + \frac{b}{l} 2 \sum_{i=1}^{l_0} i^2 = \\ &= bk^2 + b \frac{2}{l} \frac{1}{6} l_0 (l_0 + 1) (2l_0 + 1) = u_k + \frac{b}{12} (l-1)(l+1) \stackrel{\Delta}{=} u_k + c(l), \end{aligned}$$

т. е. после фильтрации усредненная функция \bar{u}_k отличается от исходной u_k на константу $c(l)$, которая пропорциональна примерно l^2 . Эта константа и определяет искажение нелинейной функции времени в виде смещения на константу, однако при слабой нелинейности (малом b) это искажение незначительно. Например, для рассмотренной в тестовом примере предыдущей главы функции $u_k = 10 - 0,1 k^2$ при $l=5$ получаем $\bar{u}_k = u_k - 0,2 = 9,8 - 0,1 k^2$, т. е. практически ту же функцию. В то же время синусоидальная

помеха любой амплитуды при удачном подборе l может быть полностью отфильтрована или максимально подавлена, и коэффициент b можно восстановить точно.

Тригонометрические функции времени. Рассмотрим, как изменяется при скользящем усреднении обычная синусоида $u_k = b \sin \varphi k$:

$$\begin{aligned} \bar{u}_k &= \frac{b}{l} \sum_{i=-l_0}^{k+l_0} \sin \varphi i = \frac{b}{l} \sum_{i=-l_0}^{l_0} \sin \varphi (k+i) = \frac{b}{l} \left(\sum_{i=-l_0}^{-1} \sin \varphi (k+i) + \right. \\ &\quad \left. + \sin \varphi k + \sum_{i=1}^{l_0} \sin \varphi (k+i) \right) = \frac{b}{l} \left(\sin \varphi k + \sum_{i=1}^{l_0} (\sin \varphi (k-i) + \right. \\ &\quad \left. + \sin \varphi (k+i)) \right) = \frac{b}{l} \sin \varphi k \left(1 + 2 \sum_{i=1}^{l_0} \cos \varphi i \right) = \frac{1}{l} \left(1 + 2 \sum_{i=1}^{l_0} \cos \varphi i \right) u_k. \end{aligned}$$

Мы получили, что рассматриваемый фильтр искажает амплитуду исходного сигнала, не меняя его частоты: $\bar{u}_k = \bar{b}(l) \sin \varphi k$, поскольку преобразованная амплитуда зависит от порядка l фильтра, но не зависит от индекса времени k :

$$\bar{b}(l) = \frac{b}{l} \left(1 + 2 \sum_{i=1}^{(l-1)/2} \cos \varphi i \right) \triangleq bg(l, \varphi). \quad (6.15)$$

Отсюда видно, что при малых частотах амплитуда сигнала искажается незначительно: при $\varphi \rightarrow 0$ получаем $g(l, \varphi) \rightarrow 1$, $\bar{b}(l) \rightarrow b$, т. е. для низкочастотных сигналов такой фильтр вполне применим. Отметим, что в более общем случае сигнала с ненулевой фазой $u_k = b \sin \varphi k + c \cos \varphi k$ результат не изменяется:

$$\bar{u}_k = \bar{b}(l) \sin \varphi k + \bar{c}(l) \cos \varphi k = g(l, \varphi) u_k.$$

Если же имеем сигнал в виде суммы двух синусоидальных колебаний $u_k = b_1 \sin \varphi_1 k + b_2 \sin \varphi_2 k$, то после усреднения получим

$$\bar{u}_k = g_1(l, \varphi_1) b_1 \sin \varphi_1 k + g_2(l, \varphi_2) b_2 \sin \varphi_2 k.$$

При этом частоты сигнала не изменяются, но амплитуды каждого колебания изменяются не одинаково, если $\varphi_1 \neq \varphi_2$. Меняя порядок фильтра l , можно одну из частот ослабить по сравнению с другой, т. е., по существу, мы получили, что $g(l, \varphi)$ с переменной частотой просто совпадает с АЧХ усредняющего фильтра. Таким образом, АЧХ фильтра выводится не только с помощью операций, указанных в предыдущем параграфе, но и непосредственными преобразованиями тригонометрических функций. Возвращаясь к обозначению частоты ω , получаем общий вид АЧХ усредняющего фильтра:

$$A(l, \omega) = 1 + 2 \sum_{i=1}^{(l-1)/2} \cos \omega i. \quad (6.16)$$

В частности, при усреднении по трем точкам $l=3$ получаем уже приведенное выше выражение $A(\omega) = 1 + 2 \cos \omega$. Для этого фильтра критическая частота, при которой $A(\omega^*) = 0$, равна $\omega^* = \frac{2\pi}{3}$. При увеличении l число l_0 таких частот ω_j^* растет и равно числу слагаемых в сумме (6.16), т. е. $l_0 = \frac{1}{2}(l-1)$, причем для каждой из них имеем $A(\omega_j^*) = 0$, $j = \overline{1, l_0}$. Таким образом, если одна из двух рассмотренных выше частот φ_1, φ_2 , например φ_2 , является помехой, то для ее полной фильтрации при любой амплитуде достаточно отыскать l , при котором одна из $\omega_j = \varphi_2$.

Отметим, что вывод о сохранении частот процесса при скользящем усреднении был ранее сделан в [24], где этот факт также использовался для повышения помехоустойчивости моделирования. Однако при этом следует иметь в виду, что в реальных процессах могут присутствовать также высокочастотные колебания, и их можно при увеличении l случайно отфильтровать, попав на критическую частоту. Кроме того, при существенно отличных от нуля частотах искажения амплитуд уже играют заметную роль и возрастают с увеличением l .

Авторегрессионные модели. Поскольку в общем случае решением авторегрессионного уравнения являются нелинейные функции времени, то искажения процесса, описываемого этой моделью, при скользящем усреднении также неизбежны. Рассмотрим простейший случай модели $u_{k+1} = bu_k$, решение которой равно $u_k = b^k u_0$, где u_0 — начальное условие. Тогда результат фильтрации при некотором l следующий:

$$\bar{u}_{k+1} = \frac{1}{l} \sum_{i=-l_0}^{l_0} u_{k+i+1} = \frac{1}{l} \sum_{i=-l_0}^{l_0} b^{k+i+1} u_0 = \left(\frac{1}{l} \sum_{i=-l_0}^{l_0} b^i \right) u_{k+1},$$

так что равенство $\bar{u}_{k+1} = u_{k+1}$ возможно только при $b=1$, т. е. в тривиальном случае $u_k = \text{const}$. Однако здесь очень интересно и важно то, что динамические свойства уравнения сохраняются полностью:

$$u_{k+1} = \left(\frac{1}{l} \sum_{i=-l_0}^{l_0} b^i \right) bu_k \equiv b\bar{u}_k, \quad (6.17)$$

т. е. устойчивость или неустойчивость уравнения $\bar{u}_{k+1} = b\bar{u}_k$ после фильтрации определяется тем же коэффициентом b , что и до фильтрации, поэтому оценка \hat{b} по данным \bar{u}_k должна совпасть с истинной: $\hat{b}=b$. Однако результат пошагового интегрирования уравнения $\bar{u}_{k+1} = b\bar{u}_k$ не будет совпадать с истинным решением $u_k = b^k u_0$, так как после усреднения получаем искаженное начальное условие \bar{u}_0 . Если же допустить, что начальное условие известно точно (если помеха $h \sin \omega k$ начала действовать позже), то только тогда после фильтрации будет восстановлен истинный сигнал.

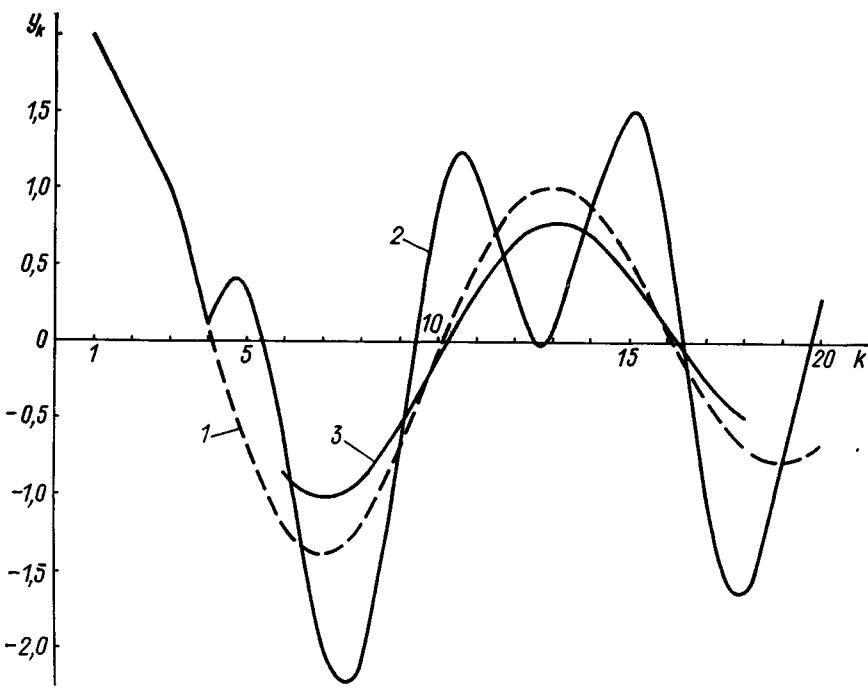


Рис. 23. Истинный процесс (1), наблюдаемый процесс с помехой (2) и результат фильтрации (3) (при $f=0,2$, $h=1$, $l=5$).

Отметим, что в рассматриваемом случае задача моделирования ставится так: имеется уравнение состояния (авторегрессионное) и уравнение измерений

$$u_{k+1} = b_0 u_k + b_1 u_{k-1} + \dots + b_r u_{k-r}, \\ y_k = u_k + h \sin \omega k.$$

По измерениям y_k , $k=1, \overline{N}$, надо восстановить незашумленный сигнал u_k . В предыдущем параграфе фактически показано, что при этом искать модель непосредственно в виде $y_{k+1} = a_0 y_k + \dots + a_r y_{k-r}$ нецелесообразно. Если же теперь с помощью удачно подобранного l и усреднения измерения y_k отфильтровать помеху, то по данным $\hat{y}_k = u_k$, как показано выше, можно успешно решить задачу моделирования.

Действительно, рассмотрим пример (6.12). Приняв, что $l=5$, а частота помехи $f = \frac{\omega}{2\pi} = 0,2$ совпадает с критической частотой фильтра (рис. 22), получим при произвольно большой амплитуде шума h отфильтрованную кривую процесса, изображенную на рис. 23 (кривая 3). Применив к этим новым данным СОМВ1, получим по критерию $i^2(\bar{W})$ наилучшую модель в виде

$$u_{k+1} = 1,100 u_k - 0,500 u_{k-2},$$

которая не отличается от исходной модели (6.12) как по коэффициентам, так и по результату интегрирования (рис. 23, кривая 1) из незашумленных начальных условий $u_0=2$, $u_1=1,5$, $u_2=1,0$.

Полученные результаты показывают, что при известной информации о частотных свойствах помех (в виде отдельных частот или полосы частот, не перекрывающейся с полосой частот сигнала), самоорганизация модели (как и прием сигнала в системе связи) должна происходить в два этапа: на первом фильтруется частотная помеха с помощью скользящего усреднения или более сложных цифровых фильтров [129, 136], на втором отыскивается модель оптимальной сложности по МГУА. Таким образом, применение двухэтапной самоорганизации моделей при заданной полосе частот (спектре) помехи позволяет восстановить незашумленный сигнал теоретически при произвольной амплитуде частотных помех.

5. УЧЕТ ИНФОРМАЦИИ О РАСПРЕДЕЛЕНИИ ПОМЕХ ВО ВРЕМЕНИ

Рассмотрим влияние на моделирование помех другого типа — распределенных во времени. Это может быть отдельный импульс в известной точке времени дискретного интервала наблюдений или импульс конечной длительности, действовавший на части этого интервала, и т. д. При этом результат моделирования будем оценивать с точки зрения точности прогноза, т. е. ошибки восстановления точного сигнала на некотором экзаменационном интервале.

Пусть задана матрица аргументов $X[N \times n]$ и вектор наблюдений выхода процесса $y[N \times 1]$, а вся выборка разделена на части A , B , C длиной N_A , N_B и N_C , причем $N_C=1$, т. е. $N_W=N_A+N_B=N-1$. Допустим, что на N_W точках с применением какого-либо внешнего критерия получена некоторая модель оптимальной сложности по комбинаторному алгоритму МГУА. Имея это в виду, не будем в дальнейшем специально указывать сложность модели. После пересчета коэффициентов такой модели на W прогноз в точке y_N принимает вид

$$\hat{y}_N = \hat{a}^T x_N = x_N^T \hat{a}. \quad (6.18)$$

Очевидно, что с учетом получения оценки \hat{a} по МНК можно это выражение записать и так:

$$\hat{y}_N = x_N^T (X_W^T X_W)^{-1} X_W^T y_W = g^T (x_N, X_W) y_W, \quad (6.19)$$

где, с одной стороны, вектор $g^T (x_N, X_W)$ фактически является частным случаем матрицы проектирования, определявшейся в гл. 5 в общем виде как $P_{GQ} = X_G (X_Q^T X_Q)^{-1} X_Q$, а с другой — его допустимо интерпретировать как некий вектор *весов измерений*: каждая его компонента показывает, с каким весом соответствующая компонента вектора наблюдений y_W участвует в получении прогнозного значения \hat{y}_N .

Полная эквивалентность записи \hat{y}_N в виде (6.18) или (6.19) (обе эти формы приведены, например, в [32, с. 87]) вроде бы не обещает каких-то дополнительных преимуществ от использования того или иного выражения. И тем не менее в [10] из анализа формы (6.19) получены некоторые конструктивные выводы, позволяющие в определенной мере учитывать при прогнозировании специфику распределения помех импульсного вида во времени.

В работе [10] выражение (6.19) трактуется как своеобразный линейный фильтр, аналогичный некоей авторегрессионной модели $\hat{y}_N = g^T y_W$, где g — переходная функция фильтра. Очевидно, что такая трактовка в общем случае неправомерна, поскольку вектор не является функцией времени: $g = g(x_N, X_W)$. Однако в частном случае, когда y является функцией только времени, а не других переменных, такая трактовка допустима: если, например, $y_k = a_0 + a_1 k$, то

$$X^T = \begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & 2 & \dots & k & \dots & N \end{pmatrix}, \quad x_N = \begin{pmatrix} 1 \\ N \end{pmatrix},$$

и вектор g является уже функцией только времени. В этом единственном случае чисто временного ряда y_k , $k = \overline{1, N}$, который, собственно, и исследуется в данной главе, прогнозирующая функция

$$\hat{y}_N = g^T(N) y_W, \quad y_W [(N - 1) \times 1], \quad (6.20)$$

может рассматриваться как линейный фильтр, и то чисто условно — точнее, только для этой выборки данных W и для этой прогнозной точки N . В самом деле, если, сохранив размерность всех параметров, сдвинуться на точку дальше, т. е. перейти от точек $1, 2, \dots, N$ к точкам $2, 3, \dots, N+1$, то можно получить выражение, совпадающее по форме с (6.20): $\hat{y}_{N+1} = g^T(N+1) y_W$, однако две «переходные функции» не совпадут: $g(N) \neq g(N+1)$. Это означает, что и в частном случае функций времени выражение (6.20) не является авторегрессионной зависимостью. Поэтому вектор $g(N) = g(x_N, X_W)$ правильнее называть не «переходной функцией», а «весовым вектором измерений».

Интересно рассмотреть теперь влияние на прогноз \hat{y}_N специфических импульсных помех измерений, когда

$$y_k = u_k + \xi_k, \quad (6.21)$$

а помеха имеет вид

$$\xi_k = \begin{cases} \xi_0 = \text{const}, & k_1 \leq k \leq k_2, \\ 0, & 1 \leq k < k_1, \quad k_2 < k \leq N, \end{cases} \quad (6.22)$$

причем в частном случае дельта-импульса $k_1 = k_2$. Если предположить, что некоторая k_0 -я компонента весового вектора $g(N)$ может оказаться равной нулю, $g_{k_0}(N) = 0$, то получится, что импульс помехи, наложенной на сигнал u_k в момент $k_1 = k_2 = k_0$, никак не

скажется на прогнозе — величина \hat{y}_N будет нечувствительной к такой помехе. И это тем более интересно, что коэффициенты модели при этом, безусловно, будут искажены:

$$\hat{a}(W) = \hat{b}(W) + (X_W^T X_W)^{-1} X_W^T \xi_W \neq \hat{b}(W),$$

где $\hat{b}(W)$ — оценка коэффициентов выбранной структуры по точным данным. Но для того чтобы это было возможно, надо показать, что компоненты вектора $g(N)$ как функция времени могут быть и положительными, и отрицательными. Это сделано в [10] на самом простом примере функции $u_k = 1 + k$: весовой вектор имеет вид (с учетом принятой нами индексации времени)

$$g_k = 2 \frac{3kN - N(N+1)}{(N-1)N(N+1)} = \frac{2(3k-N-1)}{(N-1)(N+1)}.$$

Отсюда видно, что g_k является линейно возрастающей функцией времени k , причем сначала отрицательной, затем положительной, и на интервале $[1, N]$ пересекает ось времени в «точке нечувствительности»

$$k_0 = \frac{1}{3}(N+1).$$

Поэтому, если импульсную помеху любой амплитуды ξ_0 прикладывать поочередно в точках $1, 2, \dots, N-1$, то ошибка прогнозирования точного сигнала

$$\Delta_N^2 = (\hat{y}_N - u_N)^2$$

будет сначала падать, затем в точке k_0 станет нулевой (помеха не влияет), после чего будет уже возрастать; в данном частном случае линейной модели эта точка фиксирована: $k_0 \approx N/3$, если $N \gg 1$.

Можно также рассмотреть случай помехи в виде прямоугольного импульса, т. е. при $k_2 > k_1$. Для этого запишем

$$\hat{y}_N = g^T(N) y_W = \sum_{k=1}^{N-1} g_k u_k + \xi_0 \sum_{k=1}^{N-1} g_k. \quad (6.22)$$

Для точного прогнозирования $\hat{y}_N = \hat{u}_N$ необходимо, чтобы второй член в (6.22) был равен нулю. В рассматриваемом примере это можно обеспечить: если $k_1 = 1$, то по условию $\sum_{k=1}^{N-1} g_k = 0$ находим $k_2 \approx 2N/3$.

Таким образом, если длительность импульса постоянной помехи равна $2/3$ интервала наблюдений, то такая помеха при любой амплитуде ξ_0 опять-таки практически не влияет на точность прогноза — получаем $\hat{y}_N \approx \hat{u}_N$. Интервал $[1, k_2]$ назван в [10] зоной нечувстви-

тельности. Очевидно, что в более общем случае прямоугольного импульса, симметричного относительно $k_0 \approx N/3$ (т. е. при $k_2 - k_0 = -k_0 - k_1$), эффект будет тот же.

По-видимому, аналогичные свойства весового вектора измерений возможны и в других случаях моделируемых функций времени. Поэтому при наличии информации о распределении во времени помехи импульсного типа имеется гипотетическая возможность получать достаточно точный прогноз без корректировки данных и точного оценивания коэффициентов. Для этого требуется ввести в алгоритм самоорганизации дополнительный критерий, отбирающий те частные модели, прогноз которых нечувствителен к помехе такого типа. При этом, правда, остается открытым вопрос о том, насколько структура выбранной модели соответствует модели оптимальной сложности. Однако такой критерий, по-видимому, полезен как дополнительный при выборе одной модели из нескольких, близких по другим критериям.

Отметим, что в этом направлении получены пока только первые результаты. Требуется, например, выяснить, как меняется весовой вектор при изменении длины выборки, при ее сдвиге и т. д. Это поможет полнее оценить целесообразность такого подхода к учету импульсных помех. В нынешнем состоянии результаты, полученные в [10], являются пока лишь средством анализа специфических прогнозирующих свойств моделей как функций времени.

6. ЭФФЕКТИВНОСТЬ МОДЕЛИРОВАНИЯ ПРИ УЧТЕ ХАРАКТЕРИСТИК ПОМЕХ

В предыдущих главах было показано, что помехоустойчивость алгоритмов самоорганизации объясняется прежде всего неявным учетом помех (независимо от их природы) при построении модели оптимальной сложности по зашумленным данным с применением внешних критериев селекции. Такая модель, будучи нефизической и имея коэффициенты, искаженные помехами, до определенного уровня этих помех остается тем не менее эффективной (например, хорошо прогнозирует). Однако после превышения некоторого критического уровня помех эти искажения коэффициентов становятся столь существенными, что и модель оптимальной сложности становится неэффективной. При этом для успешного моделирования необходимо переходить от неявного учета помех к явному использованию имеющейся дополнительной информации о помехах, присутствующих в исходных опытных данных.

В данной главе рассмотрены некоторые возможности учета информации о корреляционных, частотных и временных характеристиках помех. В первом случае показана эффективность использования принципа действия корреляционного приемника сигналов: известная корреляционная функция помехи вычитается из построенной по таблице данных наблюдений корреляционной функции смеси сигнала и помехи, затем решается специфическая задача обратного преобразования полученной корреляционной функции

сигнала в сам сигнал (точнее, дискретную функцию времени), после чего полученный временной ряд незашумленных данных о сигнале моделируется по МГУА. Отметим, что при этом информация о помехах используется до самоорганизации модели, на этапе предварительной обработки данных наблюдений (зашумленного сигнала). Алгоритм обратного преобразования корреляционной функции в функцию времени с помощью решения систем нелинейных уравнений предложен в [11]. Там же имеются примеры успешного решения тестовых и практических задач обратного преобразования.

Для случаев известных частотных помех в главе показано, что такая информация должна использоваться также до моделирования по МГУА, на этапе предварительной обработки зашумленных данных. При этом информация о частотном спектре или полосе частот помех применяется для синтеза эффективного цифрового фильтра, подавляющего эти частотные помехи. В главе исследованы особенности и показана эффективность моделирования с применением простейшего фильтра в виде скользящего усреднения данных наблюдений.

Если же известно, что помехи распределены во времени в виде импульсов заданной длительности, то можно опять выполнить предварительную обработку информации с целью корректировки искаженных данных, или просто исключить их из таблицы наблюдений. Однако характер распределения импульса помехи может быть не задан, а исключение соответствующих зашумленных наблюдений при малом их числе часто нежелательно. В этой ситуации, как показано в [10] на примере прогнозирования на один шаг вперед, имеется возможность использования нечувствительности прогноза к некоторым видам локализации временного импульса помехи в интервале наблюдений. При этом в моделировании используются все исходные данные, как и в специальных алгоритмах с робастным оцениванием коэффициентов [17].

Разработка специальных способов использования различных видов дополнительной информации о характеристиках помех, прежде всего на основе заимствования эффективных схем обработки и приема сигналов, представляет собой наиболее перспективный путь повышения помехоустойчивости алгоритмов самоорганизации моделей на ЭВМ. Отметим, что теоретического предела повышения помехоустойчивости моделирования при учете информации о помехах, по-видимому, не существует.

ПОМЕХОУСТОЙЧИВОСТЬ ДВУХУРОВНЕВЫХ АЛГОРИТМОВ ДОЛГОСРОЧНОГО ПРОГНОЗИРОВАНИЯ

1. МНОГОУРОВНЕВОЕ МОДЕЛИРОВАНИЕ НА НЕСКОЛЬКИХ ЯЗЫКАХ РАЗЛИЧНОЙ СТЕПЕНИ ОБЩНОСТИ

При самоорганизации моделей сложных процессов по МГУА число прогнозируемых переменных определяется разностью

$$m = M - f,$$

где M — число характеристических переменных, f — число связей между ними.

Так как f в сложных объектах моделирования не известно, то величина m определяется по факту успешного прогнозирования. Есть объекты, где $m=1$ (прогноз стока рек [22]) или $m=2$ (прогноз урожая пшеницы [166]) и т. п. Во всех случаях $m \ll M$.

Для прогнозирования одной ($m=1$) или нескольких переменных ($m > 1$), характеризующих изучаемый объект или процесс, применяются различные классы моделей: алгебраические, тригонометрические, экспоненциальные, в виде интегральных или дифференциальных уравнений и их разностных аналогов и др. Неизвестные параметры (коэффициенты) определяются по данным наблюдений. Успех прогнозирования зависит как от соответствия выбранного класса моделей характеру изучаемого процесса, так и от определенных особенностей, присущих этому классу моделей.

Например, полином от одной переменной, которой является время, имеет весьма ограниченное применение — для краткосрочного прогнозирования тренда (основной тенденции) процесса. Причем при усложнении тренда полиномиальные модели нецелесообразны даже для краткосрочного прогнозирования, поскольку поведение полинома за пределом участка интерполяции (как и между узлами интерполяции) становится совершенно непредсказуемым. Значительно больше возможностей для успешного прогнозирования дают модели в виде дифференциальных уравнений, решениями которых могут быть функции времени в виде полиномов, экспонент, синусоид и их комбинаций. При моделировании процессов, представленных дискретными рядами наблюдений, целесообразно пользоваться разностными уравнениями. Дифференциальные или разностные модели в результате идентификации коэффициентов могут оказаться устойчивыми или неустойчивыми, что определяется видом корней соответствующих характеристических уравнений. Неустойчивые уравнения можно считать непримени-

мыми для прогнозирования. Решение же устойчивого уравнения через определенный промежуток времени сходится к постоянному значению, поэтому такие уравнения применимы только для краткосрочного и среднесрочного прогнозирования. А реально наблюдавшиеся в природе процессы, как правило, не являются ни затухающими, ни бесконечно возрастающими. Например, это относится к процессам, связанным с погодой — изменение температуры, влажности воздуха, давления и др. Оказывается, что погодные условия теоретически можно предсказывать на период не более 15 суток с помощью обычно применяемого аппарата дифференциальных уравнений [18, 19], что объясняется устойчивостью этих уравнений.

Различие классов моделей можно назвать «различием языков моделирования» [49]. Тогда можно сказать, что язык дифференциальных уравнений является более общим для задач моделирования, чем язык алгебраических уравнений. С другой стороны, каждый язык можно характеризовать степенью его нечеткости, или «размытия»: язык более высокого уровня, т. е. более общий, является и более размытым. В работе [62] показано, что мерой размытия языка моделирования может быть интервал корреляции корреляционной функции соответствующего класса моделей. В частности, установлено, что интервал корреляции для дифференциальных (разностных) уравнений всегда больше, чем для алгебраических (полиномиальных), что подтверждает вывод о степени общности этих двух языков моделирования. В то же время интервал корреляции является также мерой предела предсказуемости процесса: чем больше интервал корреляции, тем больше заблаговременность прогнозирования этого процесса [62]. Следовательно, более общий (размытый) язык моделирования позволяет увеличить предел предсказуемости.

Однако для каждого отдельно взятого языка моделирования существует свой предел предсказуемости, превзойти который принципиально нельзя. Прогнозы, выполненные на каком-либо одном языке, названы в [49] одноуровневыми. Для увеличения предела предсказуемости необходимо применять несколько языков различной степени общности, или многоуровневое прогнозирование. При этом наиболее четкий язык нижнего уровня называется «детальным» и целью многоуровневого моделирования является увеличение предела предсказуемости детальных прогнозов.

В настоящее время хорошо разработан метод многоуровневого моделирования для случая, когда разным уровням соответствуют различные интервалы усреднения значений моделируемых переменных. Например, «детальными» могут быть среднемесячные значения, переменными второго уровня — среднесезонные, третьего — среднегодовые. Другими примерами переменных различных уровней могут быть такие: часовые и суточные значения; сезонные, годовые и средние за период солнечного цикла (примерно 11 лет), и т. д. Как установлено в работе [62], переход к более усредненным значениям увеличивает интервал корреляции процес-

са, что указывает на возрастание общности (размытости) языка этих уровней и означает увеличение интервала предсказуемости.

Принцип многоуровневого моделирования можно сформулировать так: применяя одновременное моделирование на языках различной степени общности, можно увеличить интервал предсказуемости детального языка нижнего уровня до интервала предсказуемости наиболее размытого языка верхнего уровня. При этом, во-первых, в языке моделирования каждого уровня необходимо использовать элементы языка более высокого уровня и, во-вторых, следует задать связь между элементами соседних уровней. Эти два условия являются необходимыми, другие дополняются в каждом конкретном случае так, чтобы совокупность условий оказалась необходимой и достаточной для получения долгосрочных количественных детальных прогнозов.

Наиболее разработанными являются методы и алгоритмы двухуровневого прогнозирования. Первые результаты по их разработке и применению были опубликованы в [88, 85], затем исследования продолжались в [82, 87], однако полностью сформированным многоуровневый подход представлен в работах [75, 49]. При этом в случае прогнозирования взаимосвязанных переменных перед непосредственным многоуровневым прогнозированием выполняется обязательный этап объективного системного анализа переменных [70, 71, 75] (см. рис. 3).

В данной главе изложены новые теоретические результаты исследования двухуровневых алгоритмов прогнозирования, причем в основном для случая одной прогнозируемой переменной, заданной времененным рядом «детальных» наблюдений. Для определенности интервалами усреднения будем считать сезон и год, так что на первом, нижнем, уровне рассматриваются среднесезонные значения (детальные), на втором, верхнем,— среднегодовые. В этом случае указанные выше два необходимых условия интерпретируются так: в моделях нижнего уровня в качестве аргументов используются среднегодовые значения, а связь между уровнями определяется очевидным условием — среднее значение суммы сезонных величин q_k должно равняться среднегодовой величине Q . Если сезонов четыре, то

$$Q = \frac{1}{4} (q_1 + q_2 + q_3 + q_4)$$

и в случае одновременного моделирования на обоих уровнях появляется возможность выбирать модели по условию минимального разбаланса этого соотношения:

$$B_L^2 = \sum_{T=1}^N \left[\hat{Q}_T - \frac{1}{4} (\hat{q}_{1T} + \hat{q}_{2T} + \hat{q}_{3T} + \hat{q}_{4T}) \right]^2 \rightarrow \min,$$

где T — индекс года; N — число лет наблюдений.

Критерий баланса такого типа впервые был предложен в [85]. Самой существенной особенностью критерия баланса явля-

ется то, что он строго определен как на интервале наблюдений, так и на интервале прогноза: единственное, что нам достоверно известно о будущем,—это то, что выполняется условие баланса [49].

Ниже подробно исследуются свойства критерия баланса как внешнего критерия и анализируется его помехоустойчивость. Кроме того, обсуждается структура двухуровневого алгоритма самоорганизации разностных прогнозирующих моделей и рассматриваются вопросы устойчивости (сходимости) этих моделей, имеющих своеобразный двухмерный отсчет времени — например, в сезонах и в годах.

2. ОСНОВНЫЕ ФОРМЫ КРИТЕРИЯ БАЛАНСА ПРОГНОЗОВ

В различных практических задачах применяются разные критерии балансового типа. Однако все их основные разновидности можно представить в общей форме критерия баланса прогноза. Эта общая форма позволяет установить условия, при которых критерий баланса является типичным внешним критерием.

Критерий баланса при нескольких моделируемых переменных. Для применения в методах самоорганизации моделей критерий баланса впервые был предложен в форме критерия баланса *переменных* [58] для случая одновременного прогнозирования нескольких переменных y_1, y_2, \dots, y_m . При этом предполагалась известной или устанавливалась с помощью моделирования некоторая связь между переменными, существующая в каждый момент времени k :

$$\Phi_k = \varphi(y_k), \quad y_k[m \times 1], \quad k = \overline{1, N}. \quad (7.1)$$

Далее для функций Φ и всех $y^* = (y_1, y_2, \dots, y_m)$ на основе данных наблюдений строились различные прогнозы (например, в виде различных функций времени). Из всех полученных вариантов прогнозов (из «веера прогнозов» [58]) выбиралось то сочетание $\Phi_k^*, y_{1k}^*, \dots, y_{mk}^*$, которое минимизирует критерий баланса

$$B_n^2 = \sum_{k=N+1}^{N+N_y} (\hat{\Phi}_k - \varphi(\hat{y}_k))^2 \rightarrow \min. \quad (7.2)$$

При этом уравнение связи (7.1) являлось ограничением, которому по предположению удовлетворяли все функции y_{ik} , $i = \overline{1, m}$, на интервале наблюдения и на участке уреждения прогноза длиной N_y . В обозначении B_n индекс внизу указывает, что критерий отражает нелинейную связь переменных — условие баланса (7.1). В алгоритмах МГУА «нелинейный» критерий баланса известен, например, в форме «кольца разностей прямых и обратных функций» [49]. Очевидно, что моделируемые переменные могут быть связаны и линейным соотношением

$$\Phi_k = \sum_{i=1}^m \beta_i y_{ik}, \quad (7.3)$$

где «балансовые коэффициенты» $\beta[m \times 1]$ заданы или определены по данным наблюдений. Связь переменных в виде (7.3) порождает «линейный» критерий баланса

$$B_g^2 = \sum_{k \in G} \left(\hat{\Phi}_k - \sum_{i=1}^m \beta_i \hat{y}_{ik} \right)^2, \quad (7.4)$$

где G — произвольная часть множества точек исходных данных или множество будущих (прогнозируемых) точек.

Критерии баланса линейного типа широко применяются в задачах самоорганизации. Часто они основаны на точно известном соотношении типа (7.3): например, изменение числа жителей в городе всегда равно сумме прибывших и убывающих [58]; общая биомасса растения всегда равна сумме биомасс надземной и подземной частей [89], и т. д. В этих примерах балансовые коэффициенты $\beta_i = 1$.

Критерий баланса при скользящем усреднении значений одной переменной. Выше мы рассматривали формы критериев баланса, применяемые при моделировании нескольких различных переменных одновременно и в силу этого называемые критериями баланса переменных. Однако критерии балансового типа успешно используются также и в алгоритмах прогнозирования отдельно взятых функций времени (процессов), заданных своими отсчетами q_k , $k=1, N$. При этом балансовое соотношение имеет место между измеренными и усредненными значениями. Например, на основе формулы усреднения на скользящем интервале длины l (6.13) (см. предыдущую главу) можно построить критерий баланса

$$B_l^2 = \sum_{k \in G} \left(\hat{q}_k - \frac{1}{l} \sum_{j=-\frac{1}{2}(l-1)}^{\frac{1}{2}(l-1)} \hat{q}_{k+j} \right)^2, \quad (7.5)$$

по минимуму которого отбирается наилучший (согласованный) прогноз усредненных \hat{q}_k и неусредненных \hat{q}_k значений процесса, различные варианты моделей для которых строятся отдельно. Отметим, что скользящее усреднение вектора исходных данных $q[N \times 1]$ можно записать с помощью специальной матрицы размера $(N-l) \times N$ [24]:

$$\bar{q}[(N-l) \times 1] = \frac{1}{l} \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 1 & 1 & \dots & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & 1 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{bmatrix}. \quad (7.6)$$

Такая матрица в каждой строке содержит l единиц подряд, причем в соседних строках набор единиц сдвинут на один элемент, поэтому усредненный вектор \bar{q} уже имеет длину $(N-l)$. Очевидно, что критерий (7.5) формально можно привести к виду (7.4), приняв $\Phi_k = q_k$, $\beta_i = 1/l$, $y_{ik} = q_{i-(k-(i-1)/2)}$, хотя такие функции y_{ik} и не являются отдельно моделируемыми переменными.

Критерий баланса для двухуровневого прогнозирования. Наиболее широко критерий баланса применяется в алгоритмах двухуровневого прогнозирования, где он называется критерием баланса прогнозов, поскольку отражает связь между прогнозами искусственных переменных, полученных из исходного временного ряда q_k , $k=1, K$, с помощью усреднения на различных интервалах времени — например, за сезон и за год. В общем случае предположим, что год содержит не четыре, а L произвольных интервалов — сезонов. Тогда на нижнем уровне алгоритма прогнозируются среднесезонные (детальные) значения процесса, на верхнем — среднегодовые. При этом от реального одномерного времени $k=1, 2, 3, \dots$ переходят к двухмерному, или двухмасштабному, отсчету времени в сезонах t и в годах T , и каждому значению k однозначно ставят в соответствие пару значений (t, T) : $t=1, L$, $T=1, N$, так что общее число лет наблюдений равно $N=K/L$, (предполагается, что всегда K кратно L). Среднегодовые Q_T и среднесезонные q_{tT} (численно равные соответствующим q_k) переменные связаны очевидным соотношением, называемым «календарным» (в отличие от скользящего) усреднением:

$$Q_T = \frac{1}{L} \sum_{t=1}^L q_{tT}, \quad (7.7)$$

так что критерий баланса прогнозов имеет вид

$$B_L^2 = \sum_{t \in G} \left(\hat{Q}_T - \frac{1}{L} \sum_{t=1}^L \hat{q}_{tT} \right)^2. \quad (7.8)$$

Критерии этого типа применяются для прогнозирования стока рек [85], температуры воздуха [87], элементов экосистемы озера [75], погодных условий [82] (где он имеет форму баланса усреднений по пространству).

Операцию вычисления среднегодовых значений Q_T через исходные детальные среднесезонные значения q_k можно записать с помощью специальной $N \times K$ -матрицы:

$$Q[N \times 1] = \frac{1}{L} \begin{bmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & 0 \\ \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 1 & \dots & 1 & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_K \end{bmatrix}. \quad (7.9)$$

Здесь в каждой строке имеется L единиц подряд, а в каждом столбце — только одна единица, что отличает календарное усреднение от скользящего (7.6).

В двухуровневых алгоритмах обычно значения для одного и того же сезона t , взятые в различные годы, рассматриваются как отдельные переменные, для которых строятся свои модели с учетом других сезонных значений и годовых. Поэтому, обозначая $\beta_i = 1/L$, $y_{ik} = y_{it} = q_{it}$, получаем формальную эквивалентность выражений (7.4) и (7.8), так что любой критерий баланса можно называть критерием баланса прогнозов, тем более, что в (7.4) также сравниваются прогнозы различных переменных.

Отметим, что в работе [22] рассматривался еще один критерий баланса переменных в виде «баланса ординат колебательного процесса». Однако он применялся не для выбора моделей, а для вычисления балансовых коэффициентов, т. е. являлся чисто *внутренним* критерием, поэтому анализировать здесь его не будем.

Таким образом, критерий вида (7.4) можно считать общей формой записи критерия баланса линейного типа с заданными балансовыми коэффициентами β_i , $i = \overline{1, m}$. Для удобства перейдем к векторной записи (7.3), (7.4), обозначив совокупность N значений переменных Φ_k , y_{ik} , $k = \overline{1, N}$, векторами $\Phi [N \times 1]$, $y_i [N \times 1]$, $i = \overline{1, m}$. Тогда балансовое соотношение (7.3) примет вид

$$\Phi = (y_1; y_2; \dots; y_m) \beta = Y\beta, \quad (7.10)$$

где $\beta^T = (\beta_1, \beta_2, \dots, \beta_m)$ — вектор балансовых коэффициентов: $Y [N \times m]$ — матрица измерений моделируемых переменных. В этих обозначениях критерий баланса (7.4) запишется так:

$$B_n^2 = (\hat{\Phi}_G - \hat{Y}_G \beta)^T (\hat{\Phi}_G - \hat{Y}_G \beta) = \|\hat{\Phi}_G - \hat{Y}_G \beta\|^2. \quad (7.11)$$

Свойства критерия (7.11) и его помехоустойчивость анализируются в следующем параграфе. Отметим одно существенное обстоятельство: соотношения (7.3), (7.10) означают линейную зависимость вектора Φ от векторов y_i , $i = \overline{1, m}$, что должно учитываться при применении критерия баланса в задачах моделирования.

Связь критерия баланса с другими внешними критериями. По своей сущности, в соответствии с принятой в гл. 5 классификацией, B_n^2 относится к критериям согласованности, поскольку отражает определенное требование, предъявляемое к моделям. Непосредственно связать критерий баланса с другими критериями нельзя, так как B_n^2 применяется при моделировании нескольких переменных, характеризующих объект или процесс ($m > 1$), а критерии, описанные в гл. 5, предназначены для оптимизации сложности модели для отдельной переменной (одного уравнения, $m = 1$). Однако при $m = 1$ и различных предположениях о Φ и $Y = y$ можно получить некоторые критерии как частные случаи критерия баланса.

Очевидно, что при $m = 1$ имеем две переменные, отличающиеся согласно (7.10) на константу $\beta_1 : \Phi = \beta_1 y$. Предположив, что $\beta_1 = 1$, получим $\Phi = y$, однако это требование равенства можно интерпретировать по-разному. Например, предположив, что $\hat{\Phi} = \hat{y}_W (A)$, $\hat{y} = \hat{y}_W (B)$, приходим к критерию минимума смещения $n_{cm}^2 = \|\hat{\Phi} - \hat{y}\|^2 = \|\hat{y}_W (A) - \hat{y}_W (B)\|^2$, требующему равенства оценок выхода модели при коэффициентах, полученных на A и B . Если же Φ равно вектору исходных данных $\Phi = y_W$, а оценивается только правая часть требуемого равенства $\Phi = y$, то получаем с учетом $y = \hat{y}_W$ квадратический критерий МНК: $\varepsilon^2 (W) = \|y_W - \hat{y}_W (W)\|^2$. Тогда при $\Phi = y_B$, $y = \hat{y}_B (A)$ частным случаем (7.11) оказывается критерий регулярности $\Delta^2 (B) = \|y_B - \hat{y}_B (A)\|^2$. Если же строится модель не полиномиальная (статическая), а разностная (динамическая), то при $\Phi = y_W$, $y = y_W^* (W)$, где $y_W^* (W)$ — результат пошагового интегрирования разностной модели, приходим к критерию ошибки интегрирования $i^2 (W) = \|y_W - y_W^* (W)\|^2$. Наконец, придавая вектору Φ смысл незашумленного вектора наблюдения u_G (точного сигнала) и предполагая, что $y = \hat{y}_G (A)$, получаем описанный в гл. 4 идеальный критерий $R^2 = \|u_G - \hat{y}_G (A)\|^2$.

Таким образом, при определенных предположениях основные виды внешних критерии селекции можно получить как предельные случаи линейной формы критерия баланса B_n^2 (7.11). С одной стороны, это свидетельствует о важности места, занимаемого критерием баланса среди всех критериев селекции — он является *системным* критерием, а с другой — позволяет предполагать, что он обладает всеми достоинствами внешних критериев, в том числе аналогичен другим критериям в смысле помехоустойчивости.

3. ПОМЕХОУСТОЙЧИВОСТЬ ВЫБОРА МОДЕЛЕЙ ПО КРИТЕРИЮ БАЛАНСА ПРОГНОЗОВ

Для изучения особенностей и результатов применения критерия баланса в условиях помех будем использовать методику аналитического исследования помехоустойчивости внешних критериев, развитую в гл. 4 и 5. При этом переменную Φ , учитываемую в критерии баланса, считаем физически измеряемой (как, например, среднегодовые значения в двухуровневых алгоритмах), а балансовые коэффициенты — заданными (в частном случае — одинаковыми $\beta_i = 1/L$).

Аналогично предыдущему предполагаем, что измерения всех моделируемых совместно переменных Φ , y_1, \dots, y_m зашумлены:

$$\Phi = u_0 + \xi_0, \quad y_i = u_i + \xi_i, \quad i = \overline{1, m}, \quad (7.12)$$

где u_0, u_1, \dots, u_m — векторы незашумленных измерений (точные сигналы), а все векторы шумов независимы между собой и имеют нормально распределенные независимые компоненты с заданными дисперсиями и нулевыми средними:

$$M_{\xi_j} = 0, M_{\xi_j \xi_j} = \sigma_j^2 E_N, j = \overline{0, m}, \quad (7.13)$$

$$M_{\xi_j \xi_v} = 0, M_{\xi_j \xi_v} = 0, j \neq v, j, v = \overline{0, m}.$$

Точные модели переменных u_j имеют вид

$$u_j = \hat{X}_j b_j^0, \hat{X}_j [N \times s_j^0], b_j^0 [s_j^0 \times 1], j = \overline{0, m}, \quad (7.14)$$

где \hat{X}_j — матрицы истинных независимых аргументов, измеряемых без помех; b_j^0 — точные векторы коэффициентов; s_j^0 — сложность точных (истинных) моделей.

Как обычно, предполагаем, что в процессе самоорганизации по критерию баланса ищутся оптимальные приближения к истинным моделям (7.14) по данным зашумленных наблюдений (7.12). При этом модели-претенденты генерируются с помощью перебора базисных наборов аргументов $X_j [N \times n_j]$, в которых по предположению находятся и истинные аргументы \hat{X}_j , т. е. $n_j \geq s_j^0$, $j = \overline{0, m}$. Таким образом, при переборе определяются коэффициенты частных моделей различной сложности для каждой из $(m+1)$ -й переменных из условных уравнений вида

$$X_j(s_j) a_j(s_j) = y_j, X_j(s_j) [N \times s_j], a_j(s_j) [s_j \times 1], \quad (7.15)$$

где s_j — сложность частной модели для j -й переменной, причем все s_j изменяются независимо, а векторы коэффициентов определяются по МНК по зашумленным данным:

$$\hat{a}_j(s_j) = (X_j^T(s_j) X_j(s_j))^{-1} X_j^T(s_j) y_j = X_j^+(s_j) y_j. \quad (7.16)$$

Для существования обратных матриц в (7.16) предполагаем, что $N \geq \max(n_j, j = \overline{0, m})$ и ранги всех $X_j(s_j)$ полные.

Разделение всей выборки длины N на части делать не будем, так как критерий баланса является внешним даже на той последовательности данных, где вычислены коэффициенты моделей (это доказывается ниже). Поэтому индекс выборки везде опускаем, считая, что во всех операциях используются векторы и матрицы полной длины N (учитываются все N точек данных).

Оценки для каждой моделируемой переменной с учетом (7.12), (7.16) имеют вид

$$\hat{\Phi}(s_0) = X_0(s_0) \hat{a}_0(s_0) = X_0(s_0) X_0^+(s_0) \Phi = P_0(s_0) (u_0 + \xi_0), \quad (7.17)$$

$$\hat{y}_i(s_i) = X_i(s_i) \hat{a}_i(s_i) = X_i(s_i) X_i^+(s_i) y_i = P_i(s_i) (u_i + \xi_i), i = \overline{1, m},$$

где $P_j(s_j)$, $j = \overline{0, m}$, — соответствующие матрицы проектирования. Тогда для критерия баланса (7.11) получаем следующее выражение:

$$\begin{aligned} B_n^2 &= \| \hat{\Phi}(s_0) - (\hat{y}_1(s_1); \hat{y}_2(s_2); \dots; \hat{y}_m(s_m)) \beta \|^2 = \| P_0(s_0) \Phi - \\ &- (P_1(s_1) y_1; \dots; P_m(s_m) y_m) \beta \|^2 = \| [P_0(s_0) u_0 - (P_1(s_1) u_1; \dots; \\ &\dots; P_m(s_m) u_m) \beta] + [P_0(s_0) \xi_0 - (P_1(s_1) \xi_1; \dots; P_m(s_m) \xi_m) \beta] \|^2. \end{aligned} \quad (7.18)$$

Целью минимизации критерия баланса является получение оптимально согласованных (сбалансированных) прогнозирующих моделей для $(m+1)$ -й переменной, которые связаны балансовым соотношением (7.10). Для минимизации (7.18) необходимо вычислить B_n^2 для всех вариантов частных моделей различной сложности. Поэтому полный объем перебора определяется выражением

$$p_B = \prod_{j=0}^m p_{n_j} = \prod_{j=0}^m (2^{n_j} - 1), \quad (7.19)$$

а в частном случае при $n_j = n$ получим, что полный перебор пропорционален 2^{nm} . Ясно, что в сложных задачах (при больших n и m) полный перебор становится невозможным. Можно ориентировочно (по аналогии с СОМВИ) принять, что полный перебор эффективен при условии $nm \leq 20$: например, при четырех сезонах ($m=L=4$) в моделях для каждого сезона можно учитывать $n=5$ аргументов. Для более сложных задач необходимо применять целесообразные способы уменьшения перебора. Некоторые из них описаны в следующем параграфе, а здесь будем считать, что полный перебор выполним, и можем точно отыскать минимум критерия баланса:

$$B_{\min}^2 = \min_{v=1, p_{n_j}, j=\overline{0, m}} B_n^2(\hat{a}_v(X_j, y_j)). \quad (7.20)$$

Этим минимальным значением B_{\min}^2 определяется набор моделей оптимальной сложности $s_0^*, s_1^*, \dots, s_m^*$ для каждой из $(m+1)$ -й моделируемых переменных.

Рассмотрим, как изменяется выбор моделей оптимальной сложности при увеличении дисперсий $\sigma_0^2, \sigma_1^2, \dots, \sigma_m^2$ шумов измерений, т. е. исследуем помехоустойчивость выбора по критерию B_n^2 .

Помехоустойчивость критерия баланса, как и других внешних критериев (см. гл. 5), проанализируем «в среднем», т. е. по его математическому ожиданию. Учитывая свойства шумов (7.13), а также то, что в нашем случае $P_j^T P_j = P_j^2 \equiv P_j$, $j = \overline{0, m}$, получаем, применяя оператор математического ожидания к (7.18):

$$\begin{aligned} \bar{B}_n^2 &= MB_n^2 = \| P_0(s_0) u_0 - (P_1(s_1) u_1; \dots; P_m(s_m) u_m) \beta \|^2 + \\ &+ \sigma_0^2 s_0 + \sum_{i=1}^m \sigma_i^2 \beta_i^2 s_i = B_u^2 + B_{\xi}^2 \end{aligned} \quad (7.21)$$

Последнее слагаемое в этом выражении вычисляется так:

$$\begin{aligned} M[\beta^T(P_1(s_1)\xi_1; \dots; P_m(s_m)\xi_m)^T(P_1(s_1)\xi_1; \dots; P_m(s_m)\xi_m)\beta] &= \\ = \beta^T M \left(\begin{array}{cccc} \xi_1^T P_1(s_1) \xi_1 & \xi_1^T P_1(s_1) P_2(s_2) \xi_2 & \dots & \xi_1^T P_1(s_1) P_m(s_m) \xi_m \\ \xi_2^T P_2(s_2) P_1(s_1) \xi_1 & \xi_2^T P_2(s_2) \xi_2 & \dots & \xi_2^T P_2(s_2) P_m(s_m) \xi_m \\ \vdots & \vdots & \ddots & \vdots \\ \xi_m^T P_m(s_m) P_1(s_1) \xi_1 & \xi_m^T P_m(s_m) P_2(s_2) \xi_2 & \dots & \xi_m^T P_m(s_m) \xi_m \end{array} \right) \beta &= \\ = \sum_{i=1}^m \beta_i^2 \sigma_i^2 \text{tr}(P_i(s_i)) &= \sum_{i=1}^m \beta_i^2 \sigma_i^2 s_i \cdot \end{aligned}$$

Таким образом, ожидаемое значение критерия баланса B_u^2 имеет две составляющие: B_u^2 равна разбалансу при моделировании по точным данным, B_ξ^2 отражает действие шумов: $\bar{B}_u^2 = B_u^2 + B_\xi^2$. Рассмотрим подробнее сначала составляющую B_u^2 , а затем в целом \bar{B}_u^2 .

Критерий баланса при точных данных. При отсутствии шумов, как показано в гл. 5, все внешние критерии изменяются при увеличении сложности моделей одинаково: сначала убывают (возможно, немонотонно), а при достижении сложности истинной модели и при дальнейшем росте сложности все они равны нулю. Аналогично ли ведет себя критерий баланса?

Рассмотрим «детерминированную» составляющую критерия \bar{B}_u^2 :

$$\begin{aligned} B_u^2 &= \|P_0(s_0)u_0 - (P_1(s_1)u_1; \dots; P_m(s_m)u_m)\beta\|^2 = \\ &= \|X_0(s_0)\hat{b}_0(s_0) - (X_1(s_1)\hat{b}_1(s_1); \dots; X_m(s_m)\hat{b}_m(s_m))\beta\|^2 = \\ &= \|\hat{u}_0(s_0) - (\hat{u}_1(s_1); \dots; \hat{u}_m(s_m))\beta\|^2 = \|\hat{u}_0(s_0) - \hat{U}\beta\|^2. \quad (7.22) \end{aligned}$$

Поскольку рассматривается случай точных данных, то необходимо установить, когда $B_u^2 = 0$, и проверить, соответствует ли это однозначно получению истинных моделей (7.14), для которых (и только для них) точно выполняется соотношение баланса

$$u_0 = U\beta = (u_1; \dots; u_m)\beta, \quad U[N \times m], \quad (7.23)$$

фактически отраженное в точных исходных данных u_0, \dots, u_m .

Очевидно, что при усложнении моделей для любой из $(m+1)$ -й переменных происходит следующее: при достижении истинной сложности s_j^0 коэффициенты восстанавливаются точно: $\hat{b}_j(s_j^0) = b_j^0$, и при дальнейшем увеличении s_j (добавлении избыточных аргументов к истинным) не изменяются:

$$\hat{b}_j(s_j | s_j \geq s_j^0) = b_j^0,$$

поскольку коэффициенты при избыточных аргументах равны ну-

лю. При этом модели всех переменных являются истинными (7.14) и $B_u^2 = 0$. Но только ли в этом случае можно получить нулевой разбаланс?

Оказывается, что критерий $B_u^2 = 0$ может равняться нулю также в тех случаях, когда при переборе различных сочетаний моделей для всех переменных появляются частные модели с совпадающими структурами: $X^* = X_0(s') = X_1(s') = \dots = X_m(s')$. При этом, очевидно, $P_0(s') = P_1(s') = \dots = P_m(s') = P^*$ и в (7.22) с учетом (7.23) получаем

$$B_u^2 = \|P^*(u_0 - U\beta)\|^2 = 0 \quad (7.24)$$

при любых значениях коэффициентов таких моделей.

Таким образом, при использовании критерия баланса для моделирования взаимосвязанных переменных обязательно должно соблюдаться следующее условие: пересечение всех множеств аргументов должно быть пустым, т. е. во всех матрицах X_0, X_1, \dots, X_m не должно быть *ни одного общего столбца*. Точнее, это относится только к матрицам X_1, \dots, X_m , поскольку в силу линейной зависимости (7.23) матрица X_0 состоит из всех неповторяющихся столбцов этих матриц. Такое условие заведомо соблюдается в случае попарно ортогональных матриц $X_i^T X_j = 0, i, j = \overline{1, m}$, однако столь жесткое условие не является необходимым.

Интересно, что именно о таком вырожденном случае (7.24) было сказано в [99] как об общем свойстве критерия баланса. После этой публикации возник миф о неединственности выбора моделей по критерию баланса, который, как установлено выше, в действительности несправедлив и относится к условиям неправильного применения этого критерия. Случай наличия хотя бы одного общего столбца в матрицах X_1, \dots, X_m , т. е. хотя бы одного общего аргумента в базисах моделирования переменных u_1, \dots, u_m , назовем случаем *вырожденного полного базиса*.

Итак, мы установили следующий результат, согласно которому критерий баланса аналогичен другим внешним критериям.

Утверждение 1. При моделировании по точным (незашумленным) данным получение значения критерия баланса $B_u^2 = 0$ является необходимым и достаточным условием структурной идентификации истинных моделей (7.14), если выполняются условия: 1) точные данные измерений переменных u_0, u_1, \dots, u_m удовлетворяют балансовому соотношению (7.23); 2) базисные матрицы аргументов X_0, X_1, \dots, X_m содержат все истинные аргументы $\hat{X}_0, \hat{X}_1, \dots, \hat{X}_m$; 3) полный базис аргументов невырожден, т. е. при переборе не найдется такая сложность s' , что $X_1(s') = \dots = X_m(s')$.

Все три перечисленных условия не являются ни слишком жесткими, ни идеализированными: первые два дают корректную постановку задачи моделирования нескольких связанных переменных линейным соотношением (7.23), а третье устанавливает условия корректного применения критерия баланса (т. е. единственность

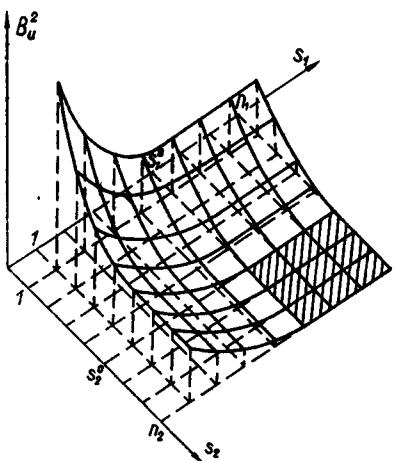


Рис. 24. Критерий баланса как функция сложностей s_1, s_2 в отсутствие помех.
Рис. 25. Линейная зависимость составляющей B_ξ^2 критерия баланса, определяемой наличием нормальных шумов, от сложностей s_1, s_2 .

выбора моделей), причем третье условие на практике несложно обеспечить алгоритмически.

Попытаемся изобразить графически зависимость критерия B_u^2 от сложности частных моделей различных переменных. Как установлено в предыдущем параграфе, критерий баланса строго определен только при $m > 1$: при $m = 1$ он превращается уже в другие критерии. Однако даже при $m = 2$ B_u^2 определен только в четырехмерном пространстве — ведь он зависит от сложности моделей трех переменных — $\hat{u}_0(s_0)$, $\hat{u}_1(s_1)$, $\hat{u}_2(s_2)$. Изобразить графически B_u^2 можно только для случая, когда в критерий вместо модели $\hat{u}_0(s_0)$ подставить точный вектор этой переменной u_0 . Это эквивалентно тому, что перебор моделей $\hat{u}_0(s_0)$ уже выполнен и определена истинная модель $\hat{u}_0(s_0^0) = u_0$. В этих условиях зависимость значений критерия баланса $B_u^2(s_1, s_2)$ от сложности частных моделей переменных $\hat{u}_1(s_1)$, $\hat{u}_2(s_2)$ имеет вид трехмерной фигуры, изображенной на рис. 24, где область значений $B_u^2 = 0$ заштрихована. Видно, что $B_u^2 = 0$ только после достижения сложностей s_1^0, s_2^0 истинных моделей, причем уменьшение критерия до нуля при росте сложности всегда монотонное, что объясняется вычислением B_u^2 на тех же данных, где оцениваются коэффициенты.

Выбор по критерию баланса в условиях помех. Составляющая критерия \bar{B}_n^2 (7.21), отражающая влияние помех, имеет вид

$$B_\xi^2 = \sigma_0^2 s_0 + \sum_{i=1}^m \beta_i^2 \sigma_i^2 s_i. \quad (7.25)$$

Она линейно возрастает при увеличении сложности частных моделей $\hat{u}_0(s_0)$, $\hat{u}_1(s_1)$, ..., $\hat{u}_m(s_m)$, т. е. это плоскость, наклон которой в многомерном пространстве определяется дисперсиями $\sigma_0^2, \sigma_1^2, \dots, \sigma_m^2$ и при их увеличении возрастает. Тогда с учетом свойств составляющей B_u^2 приходим к выводу, что минимум критерия \bar{B}_n^2 как $(m+1)$ -мерной функции переменных (чисел) s_0, s_1, \dots, s_m , во-первых, всегда единственен, во-вторых, он всегда находится в области гиперкуба $1 \leq s_j \leq s_j^0$, $j = \overline{0, m}$ (т. е. переусложенные, избыточные модели всегда отсеиваются) и, в-третьих, при возрастании дисперсий шумов (хотя бы одной из них) происходит его смещение в сторону уменьшения сложности моделей (хотя бы по одной из переменных).

Эти свойства критерия баланса можно отразить графически: на рис. 24 изображены составляющая B_u^2 критерия \bar{B}_n^2 (вогнутая поверхность), на рис. 25 — пересекающая ее возрастающая плоскость — составляющая B_ξ^2 . Их сумма дает трехмерную одноэкстремальную поверхность, сечения (изолинии) которой при различных σ_1^2, σ_2^2 изображены на рис. 26. Видно уменьшение оптимальной сложности моделей, типичное для внешних критериев селекции.

Таким образом, справедливо следующее утверждение.

Утверждение 2. Если при моделировании зашумленных переменных y_0, y_1, \dots, y_m выполнены условия 1—3 из утверждения 1, то справедливы следующие свойства выбора моделей оптимальной сложности по критерию баланса \bar{B}_n^2 : 1) при любых ненулевых $\sigma_0^2, \sigma_1^2, \dots, \sigma_m^2$ минимум критерия как функции s_0, s_1, \dots, s_m существует и единственен; 2) этот минимум находится всегда в ограниченной области $1 \leq s_j \leq s_j^0$, $j = \overline{0, m}$, где s_j^0 — сложность истинных моделей; 3) при возрастании дисперсий шумов $\sigma_0^2, \sigma_1^2, \dots, \sigma_m^2$ минимум смещается в направлении уменьшения сложности оптимальных моделей.

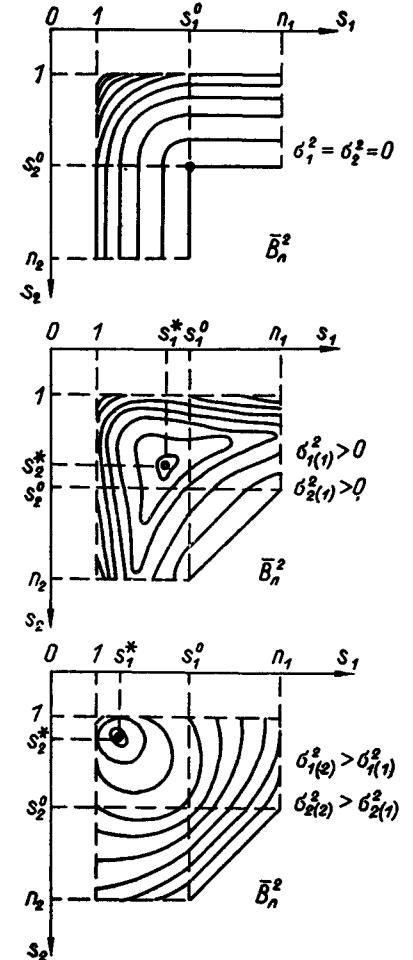


Рис. 26. Изменение выбора моделей оптимальной сложности по минимуму критерия баланса при возрастании дисперсий шумов (изображены линии равных уровней \bar{B}_n^2).

Отметим еще раз замечательную особенность критерия баланса прогнозов: это единственный из практически применяемых критериев селекции, который является внешним даже в случае, когда он вычисляется на той же части выборки данных, где оцениваются коэффициенты. В этом смысле он аналогичен идеальному критерию R^2 , рассмотренному в гл. 4, которому он аналогичен также явной зависимостью составляющей B_L^2 (7.25) от сложности частных моделей. Для критерия B_L^2 можно, по-видимому, определить его идеальный вариант следующим образом:

$$B_{\text{ид}}^2 = \|u_0 - \hat{Y}\beta\|^2 = \|u_0 - (\hat{y}_1 | \dots | \hat{y}_m)\beta\|^2, \quad (7.26)$$

где вместо моделей $\hat{\Phi}$ использован незашумленный вектор u_0 (точный сигнал). На рис. 24 — 26 фактически изображен именно этот критерий для случая $m = 2$. Отметим, что ожидаемое значение $\bar{B}_{\text{ид}}^2 = MB_{\text{ид}}^2$ отличается от (7.21) только отсутствием члена $\sigma_0^2 s_0$.

Интерпретация результатов для случая двухуровневого прогнозирования. В этом случае, как следует из сравнения (7.4), (7.8) и (7.11), число переменных равно числу сезонов $m = L$, вектор связанной переменной является вектором переменной второго уровня $\Phi^T = Q^T = (Q_1, Q_2, \dots, Q_N)$, остальные переменные относятся к первому уровню $y_t^T = q_t^T = (q_{t1}, q_{t2}, \dots, q_{tL})$, $t = \overline{1, L}$, а все балансовые коэффициенты равны: $\beta_i = 1/L$, $\bar{\beta}^T = (1/L, \dots, 1/L)$. Однако для применимости результатов исследования помехоустойчивости критерия B_L^2 к критерию B_L^2 (7.8), или, что то же,

$$B_L^2 = \|\hat{Q} - (\hat{q}_1 | \dots | \hat{q}_L)\bar{\beta}\|^2, \quad (7.27)$$

необходимо доказать справедливость условий (7.13). Векторы Q , q_t , $t = \overline{1, L}$, получены из временного ряда q_k , $k = \overline{1, LN}$, который реально измеряется. Предполагаем, что на эти измерения наложен шум с обычными свойствами:

$$q_k = q_k^0 + \eta_k, \quad M\eta_k = 0, \quad M\eta_k^2 = \sigma^2, \quad M\eta_k\eta_v = 0, \quad k \neq v.$$

Очевидно, что компонентами каждого вектора q_t являются значения временного ряда q_k , взятые через период длины L :

$$q_t^T = (q_k|_{k=t}, q_k|_{k=t+L}, \dots, q_k|_{k=t+(N-1)L}), \quad t = \overline{1, L},$$

поэтому вектор шума ξ_t для каждого q_t удовлетворяет условиям (7.13):

$$q_t = q_t^0 + \xi_t, \quad M\xi_t = 0, \quad M\xi_t\xi_t^T = \sigma^2 E_N, \quad M\xi_t\xi_v = 0, \quad v \neq t,$$

причем для всех t векторов сезонных значений дисперсии шума равны σ^2 (не зависят от t). Далее, с учетом преобразования (7.9)

получаем

$$Q = Q^0 + \xi_0, \quad Q^0 = (q_1^0 | \dots | q_L^0) \bar{\beta},$$

$$\xi_0 = (\xi_1 | \dots | \xi_L) \bar{\beta}, \quad M\xi_0 = 0, \quad M\xi_0\xi_0^T = \frac{\sigma^2}{L} E_N,$$

т. е. дисперсия шума для переменной второго уровня меньше в L раз дисперсии исходного шума: $\sigma_0^2 = \sigma^2/L$ (что повышает помехоустойчивость моделирования на втором уровне), а компоненты вектора шума ξ_0 оказываются также независимыми. И только одно условие из (7.13) не выполняется — условие независимости ξ_0 и всех ξ_t , $t = \overline{1, L}$:

$$M\xi_0\xi_t^T = \bar{\beta}^T M(\xi_1 | \dots | \xi_L)^T \xi_t = \frac{1}{L} M\xi_0^T \xi_t = \frac{\sigma^2}{L} \neq 0.$$

Вследствие этого при вычислении математического ожидания критерия B_L^2 (7.27), записанного в виде (7.18), получим вместо выражения (7.21) несколько другое:

$$\begin{aligned} \bar{B}_L^2 = MB_L^2 &= \|\hat{Q}^0 - (\hat{q}_1^0 | \dots | \hat{q}_L^0)\bar{\beta}\|^2 + \frac{\sigma^2}{L} s_0 + \frac{\sigma^2}{L} \sum_{i=1}^L s_i - \\ &- 2 \frac{\sigma^2}{L} \sum_{i=1}^L \text{tr}(P_0^T(s_0) P_t(s_t)). \end{aligned} \quad (7.28)$$

Вычислить след, фигурирующий в этом выражении, в общем виде не удается:

$$\text{tr}(P_0 P_t) = \text{tr}[X_0(X_0^T X_0)^{-1} X_0^T X_t (X_t^T X_t)^{-1} X_t^T],$$

поэтому трудно установить, как ведет себя \bar{B}_L^2 при усложнении моделей обоих уровней. Если же допустить, что матрица аргументов второго уровня X_0 ортогональна всем X_i , т. е. $X_0^T X_i = 0$, $i = \overline{1, L}$, тогда получим полную аналогию с (7.21):

$$\bar{B}_L^2 = \|\hat{Q}^0 - (\hat{q}_1^0 | \dots | \hat{q}_L^0)\bar{\beta}\|^2 + \frac{\sigma^2}{L} \left(s_0 + \sum_{i=1}^L s_i \right). \quad (7.29)$$

Важно отметить, что сделанное допущение не является искусственным, а полностью согласуется с принципом многоуровневого моделирования, описанным в § 1: из первого необходимого условия применения этого принципа следует, что аргументы моделей первого уровня не должны использоваться на втором уровне. Таким образом, мы пришли к условию, известному из практики двухуровневого прогнозирования и сформулированному в [49] следующим образом: «информационные базисы моделей каждого уровня должны быть различными». В этом случае помехоустойчивость критерия баланса прогнозов B_L^2 характеризуется свойствами, описанными выше.

На практике условие различия информационных базисов обеспечивается, например, тем, что модели первого уровня строятся в виде разностных уравнений, второго — в виде тригонометрических моделей с некратными частотами. При этом на втором уровне самоорганизуются модели \hat{Q} , близкие к точному (незашумленному) вектору Q^0 и тогда критерий B_L^2 оказывается близким к своему идеальному варианту вида (7.26).

4. СТРУКТУРА ДВУХУРОВНЕВОГО АЛГОРИТМА САМООРГАНИЗАЦИИ РАЗНОСТНЫХ ПРОГНОЗИРУЮЩИХ МОДЕЛЕЙ

В настоящее время хорошо разработаны и успешно применяются двухуровневые алгоритмы прогнозирования природных процессов, которые принято называть циклическими [85, 87, 75, 49]. Характерной особенностью таких процессов является то, что график изменения их значений периодически качественно повторяется: например, гидрограф среднемесячного стока равнинной реки (рис. 27) ежегодно имеет пик в период весеннего половодья, хотя высота этого пика год от года изменяется. Аналогичный характер имеют ежегодные сезонные колебания различных метеорологических факторов, показателей состояния экологических систем, суточные колебания температуры и другие природные явления. Циклические изменения характерны и для многих экономических показателей. Именно для таких циклических процессов оказалось эффективным применение критерия баланса прогнозов (7.8), (7.27), хотя сам критерий баланса по своему определению справедлив как для циклических, так и для других колебательных процессов.

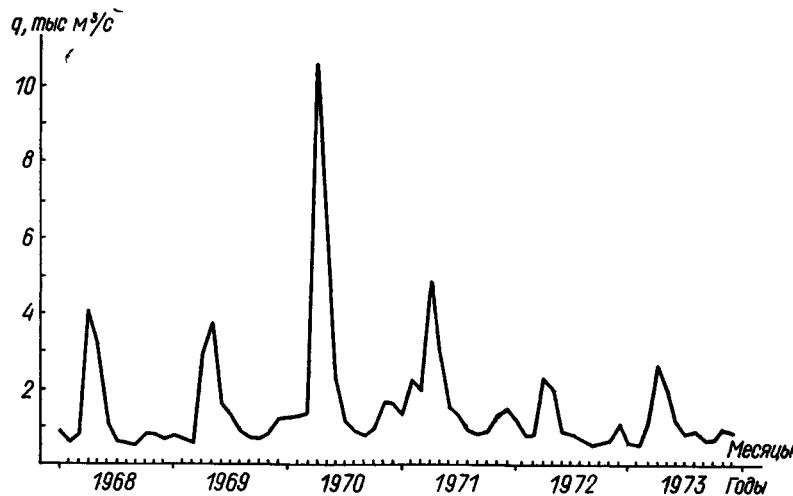


Рис. 27. Гидрограф среднемесячного стока р. Днепр (циклический процесс).

Двухуровневые алгоритмы самоорганизации прогнозирующих моделей отдельных (одномерных) [85, 87] и взаимосвязанных (многомерных) [49, 75] природных процессов с применением критерия баланса подробно описаны в указанных публикациях, поэтому здесь рассмотрим только общую структуру этих алгоритмов и укажем класс используемых моделей для случая одномерного (т. е. не векторного) процесса.

Переход к двухуровневым моделям с помощью введения двухмасштабного отсчета времени. Пусть, например, период циклического процесса равен году, который разделен на L сезонов (или кварталов). Процесс представлен временным рядом среднесезонных значений, которые являются детальными с точки зрения двухуровневого алгоритма, или переменными первого уровня. Заданы наблюдения за N лет, поэтому общее число детальных значений q_k равно NL . С помощью календарного усреднения на годовом интервале вида (7.9) из детальных значений получены среднегодовые, т. е. значения усредненной переменной Q_T , $T=1, \overline{N}$, второго уровня. При этом переменная q_k , $k=1, NL$, имеющая одномерный отсчет времени, переводится в эквивалентную переменную q_{tT} с двухмерным (двурамасштабным) отсчетом времени: $t=1, \overline{L}$, $T=1, \overline{N}$. Соответствие шкал времени k и (t, T) изображено на рис. 28 (для случая $L=4$ сезона). Перевод одной шкалы в другую всегда однозначен: прямой перевод определяется целым числом лет в общем числе среднесезонных значений k :

$$\gamma_k = \text{entier}(k/L), \quad (7.30)$$

тогда t и T равны соответственно:

$$t = k - (T - 1)L, \quad T = \begin{cases} \gamma_k + 1, & k/L \neq \gamma_k, \\ \gamma_k, & k/L = \gamma_k, \end{cases} \quad (7.31)$$

а обратный перевод выполняется непосредственно:

$$k = t + (T - 1)L. \quad (7.32)$$

Как отмечалось в § 2 в настоящей главе, после перехода к

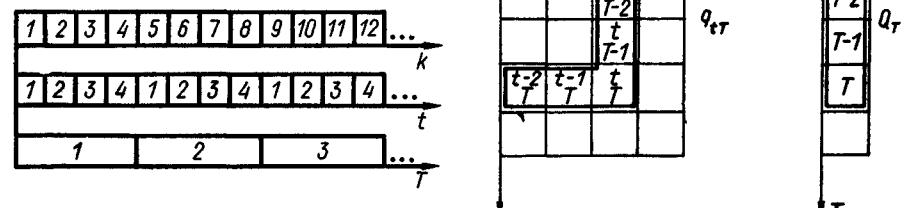


Рис. 28. Соответствие шкал одномерного (реального) и двухмерного дискретного отсчета времени ($L=4$).

Рис. 29. Структура двухмерной (для q_{tT}) и одномерной (для Q_T) таблиц данных и пример шаблонов для построения двухуровневых моделей.

дву层次шабтному отсчету детальной (сезонной) переменной значения $q_{t,T}$, $T=1, \bar{N}$, при каждом $t=1, \bar{L}$ уже рассматриваются как отдельные переменные. При этом «ленточная» таблица данных для q_k , $k=1, \bar{N}L$, представляется в виде таблицы с L столбцами и N строками, где значения каждой сезонной переменной $q_{t,T}$ представлены в столбце с номером t (рис. 29).

Итак, имеются таблицы данных для переменных первого $q_{t,T}$ (размера $N \times L$) и второго Q_T (размера $N \times 1$, рис. 29) уровней, по которым с помощью двухуровневого алгоритма строятся модели для обоих уровней, дающие сбалансированный долгосрочный прогноз усредненных и детальных прогнозов. Общий принцип работы алгоритма можно описать так: генерируются модели различной сложности для Q (независимо от переменных первого уровня) и для всех q_t как самостоятельных переменных (с учетом переменной верхнего уровня Q); затем составляются различные варианты сочетаний моделей для Q и «комплектов» по L моделей для q_t и выбирается сочетание, дающее минимум критерия баланса B_L^2 , который может вычисляться как на интервале обучения $T=1, \bar{N}$, так и на интервале прогноза $T=\bar{N}+1, \bar{N}+N_y$ длиной N_y лет. При этом на обоих уровнях могут применяться различные классы моделей.

Практически установлено, что для успешного долгосрочного прогнозирования следует применять: на верхнем уровне — гармонический алгоритм МГУА [22] (тригонометрический базис с некратными в общем случае частотами), на нижнем уровне — различные алгоритмы самоорганизации разностных моделей (базис — запаздывающие аргументы, т. е. прошлые сезонные значения). Это связано с адекватностью указанных классов моделей прогнозируемым природным процессам: тригонометрические функции с некратными частотами являются ограниченными колебательными функциями времени, аналогичными наблюдаемым в природе изменениям среднегодовых гидрометеорологических, биологических и других экологических показателей; в то же время сезонные циклические изменения этих показателей лучше описываются системой разностных уравнений с двухуровневым отсчетом времени, которые подробно анализируются ниже. Алгебраические модели на обоих уровнях не используются как представляющие наименее «размытый» (общий) язык моделирования.

Основная идея долгосрочного детального прогнозирования по двухуровневому алгоритму в соответствии с принципом многоуровневого моделирования [49] состоит в следующем: с учетом свойства более высокой заблаговременности прогнозов усредненных (среднегодовых) значений получить для них точные долгосрочные прогнозы; затем с помощью минимизации критерия баланса прогнозов получить долгосрочные детальные прогнозы сезонных значений на весь интервал точного прогноза верхнего уровня. При этом происходит существенное превышение предела предсказуемости одноуровневых детальных значений или «подтя-

гивание» [49] заблаговременности детальных прогнозов до уровня предела предсказуемости усредненных прогнозов. Подобного эффекта невозможно достигнуть при одноуровневом прогнозировании независимо от используемых классов моделей.

Структура разностных моделей, применяемых на первом уровне алгоритма. Эта структура соответствует тому или иному выбранному шаблону, по которому таблица данных рис. 29 преобразуется в L таблиц данных для моделирования каждой переменной q_t первого уровня. Например, на таблице рис. 29 изображен «угловковый» шаблон, согласно которому в модель включаются два запаздывания по сезонам (т. е. формально значения других переменных) и два запаздывания по годам для этого же сезона, и этот шаблон применяется для переменной каждого сезона:

$$q_{t,T} = a_0 + a_1 q_{t-1,T} + a_2 q_{t-2,T} + a_3 q_{t,T-1} + a_4 q_{t,T-2}, \quad t = \bar{1}, \bar{4}.$$

Следует учитывать, что, например, при $t = 1$ $q_{0T} = q_{4,T-1}$, т. е. часть шаблона переходит в предыдущий год:

$$q_{1T} = a_{01} + a_{11} q_{4,T-1} + a_{21} q_{3,T-1} + a_{31} q_{1,T-1} + a_{41} q_{1,T-2},$$

$$q_{2T} = a_{02} + a_{12} q_{4,T-1} + a_{22} q_{4,T-1} + a_{32} q_{2,T-1} + a_{42} q_{2,T-2},$$

$$q_{3T} = a_{03} + a_{13} q_{2T} + a_{23} q_{1T} + a_{33} q_{3,T-1} + a_{43} q_{3,T-2},$$

$$q_{4T} = a_{04} + a_{14} q_{3T} + a_{24} q_{2T} + a_{34} q_{4,T-1} + a_{44} q_{4,T-2}.$$

Получена система линейных разностных уравнений, выход каждого из которых зависит от выходов предыдущих — т. е. это как бы модель динамического объекта, состоящего из последовательно или каскадно соединенных блоков. Наличие четырех запаздываний (по два по сезонам и годам) требует задания восьми «начальных условий»; в данном случае это $q_{t,T-4}$, $q_{t,T-2}$, $t = \bar{1}, \bar{4}$. Учет двух запаздываний по годам приводит к уменьшению длины таблицы данных для построения моделей каждого уравнения на две точки: $N_0 = N - 2$.

В случае общего шаблона, включающего τ запаздываний по сезонам и θ по годам, структура полной системы разностных уравнений имеет вид

$$q_{t,T} = a_{0t} + \sum_{i=1}^{\tau} a_{it} q_{t-i,T} + \sum_{j=1}^{\theta} b_{jt} q_{t,T-j}, \quad t = \bar{1}, \bar{L}, \quad (7.33)$$

причем обычно значение τ ограничено: $\tau \leq L$, так как при $\tau = L + 1$ имеем $q_{t-L+1,T} = q_{t,T-1}$, и получается дублирование аргумента (такой уже имеется во второй сумме). Для определения местонахождения в таблице рис. 29 сезонных запаздывающих аргументов при $i \geq t$ справедливо соотношение

$$q_{t-i,T}|_{t \geq t} = q_{L-(i-t),T-1}. \quad (7.34)$$

Однако система, состоящая из L разностных уравнений (7.33), хотя и записана для двухмасштабного отсчета времени, еще не является моделью нижнего уровня двухуровневой системы, поскольку не включает переменной верхнего уровня Q . Поэтому общим видом линейной модели, применяемой в двухуровневом алгоритме, будет модель

$$q_{tT} = a_{0t} + \sum_{i=1}^{\tau} a_{it} q_{t-i,T} + \sum_{j=1}^{\theta} b_{jt} q_{t,T-j} + \sum_{l=0}^{\alpha} c_{lt} Q_{t-l}, \quad t = \overline{1, L}, \quad (7.35)$$

где α — число учитываемых запаздываний для среднегодовых значений (переменной верхнего уровня). Соответствующий линейный шаблон изображен на рис. 29 справа. Отметим, что при всех $t = \overline{1, L}$ (для всех уравнений) учитываются одни и те же значения $Q_t, Q_{t-1}, \dots, Q_{t-\alpha}$, причем в модель входит Q_t — среднее значение для рассматриваемого года T (индексы лет слева q_{tT} и справа Q_t совпадают).

Различие между моделями (7.33) и (7.35) является принципиальным — только последняя удовлетворяет условиям принципа двухуровневого моделирования. Ниже будет показана сущность этого различия. В дальнейшем модели вида (7.33) будем называть *двухмасштабными*, а вида (7.35) — *двухуровневыми*. Число точек, учитываемых при моделировании, для (7.33) равно $N_0 = N - \theta$, для (7.35) — $N_0 = N - \max(\alpha, \theta)$.

Общая схема работы двухуровневого алгоритма. Она состоит из следующих этапов.

1. Генерируются $F_n = 2^n - 1$ частных моделей для каждого из L уравнений вида (7.35) (число аргументов n равно числу членов в (7.35), т. е. $n = 1 + \tau + \theta + \alpha$); при этом учитываются табличные среднегодовые значения Q_t .

2. Независимо строятся F_Q моделей верхнего уровня по данным таблицы для Q_T .

3. Составляются всевозможные «комплекты» систем вида (7.35) из F_n^L моделей различной сложности (по одной частной модели для каждого уравнения).

4. Каждый из полученных комплектов систем нижнего уровня «интегрируется» из соответствующих начальных условий на заданном интервале лет $[1, N], [N+1, N+N_y]$ или вообще $[N_1, N_2]$; при интегрировании в уравнения подставляются либо табличные Q_t , либо модельные \hat{Q}_t (по одному из F_Q) значения переменной второго уровня и для каждого из $F_Q F_n^L$ таких вариантов вычисляется критерий баланса, например, на $[N_1, N_2]$:

$$B_L^2 = \sum_{T=N_1}^{N_2} \left(\hat{Q}_T - \frac{1}{L} \sum_{t=1}^L q_{tT}^* \right)^2, \quad (7.36)$$

где q_{tT}^* — результаты пошагового интегрирования.

5. Минимальное значение B_{\min}^2 указывает согласованные прогнозы годовых и сезонных значений: оптимальный прогноз Q_T^* верхнего уровня и прогнозирующую модель оптимальной сложности нижнего уровня (с оптимальными прогнозами q_{tT}^*).

Очевидно, что в полном объеме описанную схему реализовать не удается из-за большого числа $F_Q (2^n - 1)^L$ вариантов, участвующих в переборе. Поэтому на этапах 1 и 2 применяют различные способы сужения числа частных моделей, участвующих далее в минимизации критерия баланса. В основном используются целесообразные схемы иерархического (постепенного) применения в СОМВИ индивидуальных критериев, с помощью которых получают приемлемые свободы выбора на первом $F_1 \ll (2^n - 1)$ и втором $F_2 \ll F_Q$ уровнях, с тем чтобы объем перебора на четвертом этапе F_2, F_1^L оказался приемлемым. Эти способы уменьшения перебора подробно описаны, например, в [49], поэтому здесь на них останавливаться не будем. Основное внимание уделим анализу свойств двухуровневых моделей вида (7.35).

Краткие сведения о разработке метода двухуровневого прогнозирования. Модели со структурой вида (7.35) применялись в первой публикации [85] по двухуровневым алгоритмам (прогноз среднемесячного стока рек, $L=12$), причем при предельно простом шаблоне $t=a=1, \theta=0$, однако они в общем случае были нелинейными:

$$q_{tT} = f_t(q_{t-1,T}, Q_T).$$

В [87] состав аргументов моделей нижнего уровня был расширен за счет введения скользящего усреднения сезонных значений (в задаче прогнозирования среднемесячной температуры воздуха):

$$q_{tT} = a_{0t} + \sum_{i=1}^{\tau} a_{it} q_{t-i,T} + \sum_{j=1}^{\theta} b_{jt} q_{t,T-j} + \sum_{l=2}^{l_0} d_{lt} \bar{q}_{tT}^{(l)} + \sum_{j=0}^{\alpha} c_{jt} Q_{t-j}, \quad (7.37)$$

где аргумент $\bar{q}_{tT}^{(l)}$ равен среднему от l предыдущих (по отношению к моменту (t, T)) сезонных значений:

$$\bar{q}_{tT}^{(l)} = \frac{1}{l} \sum_{i=1}^l q_{t-i,T}, \quad l = \overline{2, l_0}.$$

Наконец, в [75] модель вида (7.35) была обобщена на случай многомерного процесса, точнее, на случай совместного моделирования нескольких взаимосвязанных циклических процессов, подверженных воздействию независимых процессов (возмущающих воздействий). Это было сделано на примере прогнозирования двух взаимосвязанных показателей экосистемы озера Байкал — биомасс фито- и зоопланктона с учетом воздействия изменений температуры воды и солнечной радиации (при четырех сезонах — весна, лето, осень, зима [75, 49]). Структуру этих моделей (при $t =$

=2) в частном случае $\tau=\theta=a=1$ и при одном возмущении w можно представить в наших обозначениях так:

$$\begin{aligned} q_{it}^{(1)} &= a_{0t}^{(1)} + a_{1t}^{(1)}q_{t-1,t}^{(1)} + a_{2t}^{(1)}q_{t-1,t}^{(2)} + b_{1t}^{(1)}q_{t,t-1}^{(1)} + b_{2t}^{(1)}q_{t,t-1}^{(2)} + \\ &\quad + d_{0t}^{(1)}w_{it} + c_{0t}^{(1)}Q_t + c_{1t}^{(1)}W_t, \quad (7.38) \\ q_{it}^{(2)} &= a_{0t}^{(2)} + a_{1t}^{(2)}q_{t-1,t}^{(1)} + a_{2t}^{(2)}q_{t-1,t}^{(2)} + b_{1t}^{(2)}q_{t,t-1}^{(1)} + b_{2t}^{(2)}q_{t,t-1}^{(2)} + \\ &\quad + d_{0t}^{(2)}w_{it} + c_{0t}^{(2)}Q_t + c_{1t}^{(2)}W_t, \end{aligned}$$

где w_{it}, W_t — переменные первого и второго уровней алгоритма прогнозирования внешнего воздействия w_k (модели этих переменных вычисляются независимо). Отметим, что системные двухуровневые модели типа (7.38) в [75] применялись как в линейной, так и в нелинейной формах. Чем больше различаются между собой сезонные и годовые модели по классу уравнений, виду опорной функции и используемой информации, тем более однозначным является выбор пары прогнозов по критерию баланса.

5. ОСОБЕННОСТИ ДВУХУРОВНЕВЫХ ПРОГНОЗИРУЮЩИХ МОДЕЛЕЙ, ПОСТРОЕННЫХ ПО КРИТЕРИЮ БАЛАНСА

В результате применения двухуровневого алгоритма определяется оптимальная сложность пары моделей как верхнего, так и нижнего уровней, в виде гармонического тренда и системы L уравнений типа (7.35). При этом находятся значения оценок коэффициентов $a_{0t}, a_{1t}, b_{jt}, c_{jt}$, соответствующие минимуму критерия баланса прогнозов B_L^2 (7.36). Модель оптимальной сложности соответствует структуре общего вида (7.35), поэтому рассмотрим свойства моделей данного класса.

Структура двухуровневой системы разностных уравнений. Решение системы (7.35) из заданных начальных условий вычисляется рекуррентно: вычисленное q_{11} подставляется во второе уравнение, затем вычисляется q_{21} и подставляется вместе с q_{11} в третье и т. д., до q_{L1} ; последний результат подставляется снова в первое уравнение системы, вычисляется q_{12} и т. д. Таким образом, система применяется рекуррентно или циклически, причем начальные условия «действуют» ровно θ циклов (рекурсий), где θ — число запаздываний по годам. Все значения $Q_{t-j}, j=0, \theta$, подставляются из таблицы исходных данных или из решений моделей верхнего уровня, т. е. они выступают в модели (7.35) в качестве известных внешних воздействий. Как уже указывалось, структура системы (7.35) подобна структуре объекта, состоящего из каскадно соединенных блоков. Для примера на рис. 30 изображено блочное представление простейшей модели вида $q_{it}=a_{0t}+a_{1t}q_{t-1,t}+c_{1t}Q_t$ при $t=1, 4$, с начальным условием $q_{02}=q_4=q^0$. Видно, что модель можно интерпретировать как дискретный динамический объект с внешним воздействием Q_t и обратной связью (выход блока q_{4t} подается на вход $q_{1,t+1}$). Выходом всего объекта

является совокупность четырех значений $q_{1t}, q_{2t}, q_{3t}, q_{4t}$ — т. е. вектор, состоящий в общем случае из L сезонных значений каждого года $q_t = (q_{1t}, \dots, q_{Lt})^T$. «Каскадное» использование уравнений системы (7.35), соответствующее каскадной структуре блоков объекта на рис. 30, позволило назвать двухуровневые модели такого типа также «каскадно-регрессионными». Наличие обратной связи в объекте рис. 30 требует изучения его устойчивости. С другой стороны, вопрос об устойчивости или неустойчивости системы (7.35) эквивалентен вопросу о сходимости двухуровневого алгоритма долгосрочного прогнозирования, важность которого отмечалась в гл. 1. Если модель оптимальной сложности нижнего уровня алгоритма неустойчива, то это соответствует отсутствию сходимости двухуровневого алгоритма самоорганизации. Если же модель устойчива, то алгоритм сходится, но при этом возникает задача исследования характера сходимости: к чему и за какое время сходится решение — к константе, алгебраическому тренду, периодическому тренду, к какой-то другой функции времени?

Вопрос об устойчивости системы разностных уравнений (7.35) решается непросто, что связано со структурой этой системы: она удобна для ее алгоритмического (рекуррентного) решения (поступативного интегрирования), но не позволяет выразить это решение в общем виде с использованием начальных условий. Поэтому необходимо преобразовать (7.35) к виду, удобному для получения общего решения. Оказывается, данную систему можно свести к двум хорошо известным классам моделей: скалярному разностному уравнению с периодическими коэффициентами или L -мерному разностному уравнению с постоянными коэффициентами.

Представление двухуровневой модели в виде скалярного разностного уравнения. Первый способ рассмотрен в [116], и сводится он к тому, что в (7.35) выполняется возврат к исходному (одномасштабному) времени k (7.32), в котором эта система представляется одним уравнением с переменными коэффициентами:

$$q_k = a_0(k) + \sum_{i=1}^{\theta} a_i(k)q_{k-i} + \sum_{j=1}^L b_j(k)q_{k-L+j} + \sum_{j=0}^{\theta} c_j(k)Q_{T_k-j}. \quad (7.39)$$

В этом скалярном уравнении все коэффициенты периодичны с периодом L :

$$a_i(k+L) = a_i(k), \quad b_j(k+L) = b_j(k), \quad c_j(k+L) = c_j(k),$$

однако эта периодичность задана не функционально, а таблично,

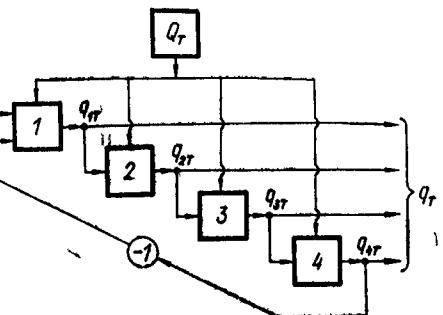


Рис. 30. Пример представления двухуровневой разностной модели в виде многомерного динамического объекта.

что осложняет анализ. Например, $a_i(k)$ на каждом периоде равна $a_i(k) = a_{it}$, $TL + 1 \leq k \leq (T+1)L$.

Очевидно, что «внешнее воздействие» Q_{T-k-i} является кусочно постоянной функцией k , например:

$$Q_{T-k} = Q_T = \text{const}, \quad TL + 1 \leq k \leq (T+1)L.$$

С помощью представления (7.39) в [116] получены некоторые выводы об устойчивости системы (7.35), однако известные трудности анализа уравнений с периодическими коэффициентами [162] (да еще представленными таблично) привели к неполноте и неконструктивности этих результатов: например, не решены вопросы определения типа сходимости и оценки времени переходных процессов.

Представление двухуровневой модели в векторной форме. Рассмотрим второй путь преобразования системы (7.35) — к векторному разностному уравнению с постоянными параметрами. Этот путь более конструктивен, поскольку результатом является система уравнений в нормальной форме (форме Коши), методы анализа которой хорошо разработаны.

Идея предлагаемого преобразования состоит в том, чтобы для объекта типа изображенного на рис. 30 получить модель не в рекуррентной форме (7.35), а в виде явной зависимости вектора выхода $q_T^v = (q_{1T}, q_{2T}, \dots, q_{LT})$ от своих предыдущих (запаздывающих) значений q_{T-1}, q_{T-2} и т. д. Для этого необходимо в правой части каждого уравнения (7.35) исключить переменные с индексом T , что несложно сделать, подставив в каждое t -е уравнение выражения для $q_{1t}, q_{2t}, \dots, q_{t-1,t}$, т. е. предыдущие уравнения; затем привести подобные члены и результат представить в векторно-матричной форме. Например, в простейшем случае $L = 2$, $a_{0t} = 0$, $\tau = 1$, $\vartheta = \alpha = 0$, т. е. для системы $q_{tT} = a_{1t}q_{t-1,T} + c_{0t}Q_T$ получаем (учитывая (7.34))

$$q_T = \begin{pmatrix} q_{1T} \\ q_{2T} \end{pmatrix} = \begin{pmatrix} 0 & a_{11} \\ 0 & a_{12}a_{11} \end{pmatrix} \begin{pmatrix} q_{1,T-1} \\ q_{2,T-2} \end{pmatrix} + \begin{pmatrix} c_{01} \\ c_{02} + a_{12}c_{11} \end{pmatrix} Q_T = Hq_{T-1} + g_T Q_T,$$

где H и g_T — числовые матрица и вектор. Начальным условием для исходной системы является $q_{01} = q_{20} = q^0$, а для полученного уравнения $q_0^v = (0, q^0)$, т. е. начальное условие не изменяется. При $L=2$, $\tau = \vartheta = 2$, $\alpha = 1$ система двух уравнений вида

$$q_{tT} = a_{0t} + a_{1t}q_{t-1,T} + a_{2t}q_{t-2,T} + b_{1t}q_{t,T-1} + b_{2t}q_{t,T-2} + c_{0t}Q_T + c_{1t}Q_{T-1}$$

преобразуется в векторное уравнение

$$q_T = H_1 q_{T-1} + H_2 q_{T-2} + g_0 + g_1 Q_T + g_2 Q_{T-1},$$

где приняты следующие обозначения:

$$H_1 = \begin{pmatrix} a_{21} + b_{11} & a_{11} \\ a_{12}a_{21} & a_{22} + a_{12}a_{11} \end{pmatrix}, \quad H_2 = \begin{pmatrix} b_{21} & 0 \\ 0 & b_{22} \end{pmatrix},$$

$$g_0 = \begin{pmatrix} a_{01} \\ a_{02} + a_{12}a_{01} \end{pmatrix}, \quad g_1 = \begin{pmatrix} c_{01} \\ c_{02} + a_{12}c_{01} \end{pmatrix}, \quad g_2 = \begin{pmatrix} c_{11} \\ c_{12} + a_{12}c_{11} \end{pmatrix}.$$

Очевидно, что в этом случае начальные условия исходной системы для q_{tt} , $t=1, 2$ и для полученного уравнения совпадают:

$$q_0 = \begin{pmatrix} q_{10} \\ q_{20} \end{pmatrix}, \quad q_{-1} = \begin{pmatrix} q_{1,-1} \\ q_{2,-1} \end{pmatrix}.$$

Для системы разностных уравнений общего вида (7.35), т. е. для двухуровневой модели, справедливо следующее эквивалентное представление в векторно-матричной форме (форме Коши):

$$q_T = H_0 q_{T-1} + \sum_{t=1}^{\vartheta} H_t q_{T-t} + g_t + \sum_{j=0}^{\alpha} g_{j+1} Q_{T-j}; \quad (7.40)$$

Здесь q_{T-1} входит, кроме первого члена, также в первую сумму, поэтому фактически

$$H_0 q_{T-1} + \sum_{t=1}^{\vartheta} H_t q_{T-t} = (H_0 + H_1) q_{T-1} + \sum_{t=2}^{\vartheta} H_t q_{T-t}; \quad (7.41)$$

при необходимости это будем учитывать.

Все входящие в (7.40) матрицы и вектора однозначно выражаются через коэффициенты исходной системы (7.35) с помощью рекуррентных соотношений. Элементы матрицы $H_0 = (h_{vt}^0, v, t = \overline{1, L})$ в которой «спрятаны» все коэффициенты a_{it} , $i = \overline{1, \tau}$, из (7.35), имеют вид

$$h_{vt}^0 = a_{L-t+v,v} + \sum_{r=1}^{v-1} h_{rt}^0 a_{v-r,v}, \quad v, t = \overline{1, L}. \quad (7.42)$$

Укажем, что в H_0 всегда ровно τ ненулевых столбцов, и если $\tau < L$, то первые $(L - \tau)$ столбцов (с индексами $t = \overline{1, L - \tau}$) имеют нулевые элементы.

Матрицы H_i , $i = \overline{1, \vartheta}$, выражаются через группу коэффициентов b_{jt} , $j = \overline{1, \vartheta}$, в (7.35) и всегда диагональны, т. е. их элементы h_{vt}^i , $v, t = \overline{1, L}$, имеют вид

$$h_{tt}^i = b_{it}, \quad i = \overline{1, \vartheta}; \quad h_{vt}^i = 0, \quad v \neq t. \quad (7.43)$$

Все векторы $g_0, g_1, \dots, g_{\alpha+1}$ в (7.40) имеют длину L и относятся к членам, имеющим для объекта (7.40) смысл независимых внешних воздействий. Для вычисления их элементов справедливы рекуррентные формулы, аналогичные (7.42):

$$g_{v0} = a_{0v} + \sum_{r=1}^{v-1} g_{rv} a_{v-r,v}, \quad v = \overline{1, L}, \quad (7.44)$$

$$g_{v(j+1)} = c_{jv} + \sum_{r=1}^{v-1} g_{r(j+1)} a_{v-r,v}, \quad j = \overline{0, \alpha}. \quad (7.45)$$

Итак, двухуровневая модель (7.35) может быть представлена в эквивалентной форме L -мерного уравнения (7.40) с постоянными коэффициентами. Очевидно, что при отсутствии последнего члена в (7.40) мы получим эквивалентное представление двухмасштабной модели (7.33). Для интегрирования уравнений (7.35) или (7.40) необходимо задать одни и те же начальные условия: если $\vartheta > 0$, ими являются ϑ первых строк таблицы «двуухмасштабных» сезонных значений — $q_{t1}, \dots, q_{t\vartheta}$; если же $\vartheta = 0$, то это τ последних значений в первой строке этой таблицы: $q_{L-\tau+1,1}, q_{L-\tau+2,1}, \dots, q_{L1}$.

Интересно, что соотношение между (7.35) и (7.40) является взаимно однозначным: если известны числовые значения элементов матриц $H_0, H_i, i = \overline{1, \bar{b}}$, и векторов $g_0, g_j, j = \overline{0, \bar{a}}$, то с учетом структуры уравнений (7.35) выражения (7.42) — (7.45) представляют собой определенные алгебраические уравнения относительно коэффициентов $a_{it}, b_{it}, a_{0t}, c_{jt}, t = \overline{1, \bar{L}}$, решая которые, можно от формы (7.40) однозначно перейти к рекуррентной форме (7.35). Это, во-первых, позволяет применять в двухуровневом алгоритме долгосрочного прогнозирования модель первого уровня в форме (7.40). Критерий баланса для определения оптимальной структуры этой модели принимает вид, полностью эквивалентный (7.36):

$$B_L^2 = \sum_{t=N_1}^{N_2} (\hat{Q}_t - \beta^T q_t^*)^2, \quad (7.46)$$

где q_t^* — результат интегрирования уравнения (7.40); $\beta [L \times 1] = = (\beta_i = 1/L, i = \overline{1, \bar{L}})$. После оптимизации сложности двухуровневой модели в форме (7.40) можно однозначно восстановить ее в рекуррентной форме (7.35).

Во-вторых, эквивалентность моделей (7.35) и (7.40) позволяет подробно исследовать условия устойчивости двухуровневых моделей на основе анализа (7.40). Отметим, что в принципе к виду (7.40) можно привести также двухуровневую модель более общего вида (7.38) для прогнозирования t взаимосвязанных процессов с учетом ряда независимых внешних воздействий.

6. АНАЛИЗ УСТОЙЧИВОСТИ ДВУХУРОВНЕВЫХ РАЗНОСТНЫХ ПРОГНОЗИРУЮЩИХ МОДЕЛЕЙ

Двухуровневая модель в виде (7.40) является L -мерным разностным уравнением с ϑ запаздывающими аргументами (без последнего члена это будет многомерный аналог авторегрессионного уравнения). Известно, что разностное уравнение с заданным числом запаздываний можно свести к уравнению с одним запаздыванием — для этого нужно ввести новый вектор переменных $\vec{q}_t^T = (q_t^T; q_{t-1}^T; \dots; q_{t-\vartheta}^T)$ и выполнить в (7.40) соответствующие преобразования (аналогично тому, как дифференциальное уравнение любого порядка приводится к системе уравнений первого

порядка). Поэтому без ограничения общности в дальнейшем будем анализировать решения этого уравнения для случая $\vartheta = 1$. Кроме того, последняя сумма в (7.40) представляет собой некоторую заданную функцию времени T , поэтому для анализа достаточно учитывать ее в целом или, что проще, принять $a = 0$.

Далее будем исследовать устойчивость решений разностного уравнения

$$q_T = Hq_{T-1} + g_0 + g_1 Q_T, \quad q_T|_{T=0} = q_0, \quad (7.47)$$

где матрица $H [L \times L]$ и векторы $g_0, g_1 [L \times 1]$ определены в результате структурного моделирования по двухуровневому алгоритму самоорганизации, а начальное условие q_0 и внешнее воздействие Q_T (переменная второго уровня) заданы.

Исследовать поведение решений уравнения (7.47) при $T \rightarrow \infty$ важно потому, что тем самым мы выясняем прогнозирующие возможности двухуровневой модели, т. е. устанавливаем, сходится этот прогноз к константе, периодической функции или является незатухающей ограниченной функцией времени.

Чтобы изучить характер решений разностного уравнения с внешним воздействием, следует прежде всего определить собственную устойчивость «автономной» части этого уравнения:

$$q_T = Hq_{T-1} + g_0, \quad q_T|_{T=0} = q_0. \quad (7.48)$$

Очевидно, что это — аналог двухмасштабного уравнения (7.33). Исследуем сначала устойчивость последнего уравнения.

Устойчивость двухмасштабной прогнозирующей модели. В фазовом L -мерном пространстве уравнения (7.48) существует особая точка (точка равновесия), которая единственна ввиду линейности этого уравнения. Она определяется из (7.48) в предположении окончания переходного процесса, т. е. по условию $q_T = q_{T-1}$:

$$q^0 = (E - H)^{-1} g_0. \quad (7.49)$$

Особая точка может быть точкой устойчивого или неустойчивого равновесия, но она существует, если матрица $(E - H)$ невырождена. Однако вырожденность этой матрицы (т. е. случай $H = E$) практически невозможна — это может быть лишь при моделировании строго периодической функции в отсутствие помех. Поэтому дальше считаем, что точка равновесия (7.49) существует и не зависит от начальных условий q_0 .

Поскольку все элементы матрицы H и вектора g_0 , как следует из предыдущего параграфа, различны, то все компоненты вектора q^0 также являются разными. Поэтому, если точка равновесия устойчива, то решение векторного уравнения (7.48) (в виде вектора $q_T = q^0$) в исходном пространстве сезонных переменных, или в реальном времени, будет представлять собой периодическое повторение значений $q_1^0, q_2^0, \dots, q_L^0$. Таким образом, если уравнение (7.48) устойчиво, то его решение может сходиться только к периодической

функции (в реальном «одномерном» времени, т. е. для «детальных» сезонных переменных).

Для решения вопроса об устойчивости точки равновесия q^0 , как обычно, следует перейти к уравнению в отклонениях от этого решения. Введя новую переменную $z_t = q_t - q^0$ и подставив $q_t = z_t + q^0$ в (7.48), получим

$$z_t = Hz_{t-1}, \quad z_t|_{t=0} = q_0 - q^0. \quad (7.50)$$

Теперь устойчивость тривиального решения $z_t = 0$ определяется видом корней характеристического уравнения

$$\det(H - \lambda E) = 0. \quad (7.51)$$

В соответствии с теорией разностных уравнений (см., например, [111]), если все корни (7.51) $|\lambda_i| < 1$, то уравнение (7.50) (и соответствующее решение двухуровневой модели) устойчиво; если же хотя бы один из корней $|\lambda_i^*| > 1$, оно неустойчиво. При этом характер приближения к $z=0$ или ухода от него может быть апериодическим, если все корни действительны и положительны, и колебательным, если имеются корни отрицательные или комплексные. Интересно отметить, что число ненулевых корней (7.51) в соответствии со структурой матрицы H (см. 7.42) при $\vartheta=0$ равно τ , поэтому $i=1, \tau, \tau \leq L$.

Общее время переходного процесса. Допустим теперь, что все $|\lambda_i| < 1, i=1, \tau$, т. е. решение $z_t = 0$ (или $q_t = q^0$) устойчиво, и оценим время окончания переходного процесса. Считаем, что моменту достижения стационарного решения соответствует условие $\|z_T\| \leq \varepsilon$, где ε — малое положительное число.

Решение уравнения (7.50) имеет вид

$$z_t = H^T z_0, \quad (7.52)$$

поэтому ввиду

$$\|z_T\| \leq \|H^T\| \cdot \|z_0\| \leq \|H\|^T \|z_0\|$$

из условия $\|z_T\| = \varepsilon$ получаем

$$T^* = \frac{1}{\ln \|H\|} \ln \frac{\varepsilon}{\|z_0\|}. \quad (7.53)$$

Мы получили оценку сверху времени окончания переходного процесса для (7.50), т. е. T^* — это максимальное время установления периодического решения двухмасштабной прогнозирующей модели, или время сходимости к «жесткому» периодическому тренду. Это время определяется как параметрами модели (нормой $\|H\|$), так и отклонением начальных условий от q^0 (нормой $\|z_0\|$). Очевидно, что $\|H\| = \|\text{diag}(\lambda_1, \dots, \lambda_L)\| < 1$, а ввиду $\|z_0\| \geq \varepsilon$ имеем также $(\varepsilon/\|z_0\|) < 1$, поэтому оба логарифма в (7.53) отрицательны и $T^* \geq 0$. При этом время переходного процесса увеличивается, если уменьшается степень устойчивости уравнения, а также если увеличивается отклонение начальных условий.

Время стабилизации прогноза каждой сезонной переменной. Задача определения момента стабилизации каждой компоненты z_{vt} вектора z_t представляет еще больший интерес, чем оценка общего времени T^* . Для решения этой задачи необходимо получить решение (7.50) не в виде (7.52), а покомпонентно.

Это можно сделать по методу вариации произвольных постоянных [111] (для простоты подразумеваем случай различных корней). Получив корни (7.51) $\lambda_i, i=1, \tau$, подставляем в (7.50) его общее решение в виде

$$z_{vt} = \sum_{i=1}^{\tau} \mu_{vi} \lambda_i^t, \quad v=1, L, \quad (7.54)$$

для $T=1, \tau$ и решаем полученную систему τL уравнений относительно такого же числа неизвестных коэффициентов μ_{vi} . Эти коэффициенты в результате будут выражены через τ начальных условий. Затем строим следующую оценку для модуля решения z_{vt} :

$$|z_{vt}| \leq \sum_{i=1}^{\tau} |\mu_{vi}| |\lambda_i^t| \leq |\lambda|_{\max}^T \sum_{i=1}^{\tau} |\mu_{vi}| \stackrel{\Delta}{=} |\lambda|_{\max}^T \|\mu_v\|,$$

где $|\lambda|_{\max}$ — максимальный корень (7.51); $\mu_v^T = (\mu_{v1}, \dots, \mu_{v\tau})$. Отсюда по условию $|z_{vt}| = \varepsilon$ получаем оценку сверху T_v^* для времени стабилизации каждой v -й компоненты вектора состояния z_t :

$$T_v^* = \frac{1}{\ln |\lambda|_{\max}} \ln \frac{\varepsilon}{\|\mu_v\|}. \quad (7.55)$$

Это и есть время стабилизации прогноза для каждой из $v=1, L$ сезонных переменных q_{vt} , являющихся решениями (7.48), т. е. двухмасштабной модели вида (7.33). Очевидно, что между оценками (7.53) и (7.55) соотношение следующее $T^* \geq \max T_v^*$.

Для примера вычислим все T_v^* для простейшей модели с $\tau=1, \vartheta=0$. При этом матрица H имеет единственный ненулевой столбец (последний), элементы которого равны $h_{iL} = a_{11}a_{12} \dots a_{1i}, i=1, L$. Тогда из (7.51) получаем только один ненулевой корень

$$\lambda = h_{LL} = \prod_{i=1}^L a_{1i}.$$

В этом случае по методу вариации произвольной постоянной с учетом начального условия $z_{0L} = z^0$ получаем $\mu_{v1} = z^0 h_{vL}/h_{LL}$, тогда покомпонентное решение (7.54) принимает вид

$$z_{vt} = \mu_{v1} \lambda^t = z^0 h_{vL} h_{LL}^{t-1}.$$

Следовательно, оценки T_v^* запишутся в виде

$$T_v^* = \frac{1}{\ln |\lambda|} \ln \frac{\varepsilon}{\|\mu_{v1}\|} = 1 + \left(\ln \frac{\varepsilon}{|z^0|} - \ln |h_{vL}| \right) / \ln |h_{LL}|,$$

или через коэффициенты двухмасштабной модели (7.33):

$$T_v^* = 1 + \left(\ln \frac{e}{|q_0 - q^0|} - \sum_{i=1}^v \ln |a_{i1}| \right) / \sum_{i=1}^L \ln |a_{i1}|.$$

Отсюда видно, что если все a_{i1} различны (но $|\lambda| = |h_{LL}| < 1$ — система устойчива), то время стабилизации всех решений q_{vt} разное.

Таким образом, переход от рекуррентной формы (7.35) двухуровневой модели к векторной форме (7.47) позволяет с помощью стандартных методов анализировать характер сходимости решений этой модели. Эти методы можно запрограммировать и применять в алгоритме самоорганизации для оценки качества каждой частной двухуровневой модели, испытываемой в процессе перебора.

Выше мы установили, что в устойчивой двухуровневой модели (7.35), (7.47) собственные колебания, т. е. решения «автономной» двухмасштабной модели (7.33), (7.48), затухают за время, не большее T^* , после чего решения (7.35), (7.47) будут определяться свойствами «внешнего воздействия» Q_t , т. е. переменной второго уровня.

Анализ сходимости решений двухуровневой модели. Общее решение (7.47) имеет вид

$$q_t = H^T q_0 + g_0 \sum_{i=1}^T H^{T-i} + g_1 \sum_{i=1}^T H^{T-i} Q_i. \quad (7.56)$$

Первые два члена определяют «свободное» решение, третий — «вынужденное». За время T^* первая составляющая стабилизируется и дальше остается равной q^0 (7.49), так что

$$q_T^* = q^0 + g_1 \sum_{i=1}^T H^{T-i} Q_i, \quad T > T^*. \quad (7.57)$$

Характер полученного решения q_T^* , $T > T^*$, полностью определяется характером функции Q_t , заданной как решение модели второго уровня алгоритма. Если эта функция постоянна $Q_t = Q_0 = \text{const}$, то уравнение (7.47) имеет просто несколько другую точку равновесия, т. е. его решение сходится не к (7.49), а к

$$q_t^* = q^* = (E - H)^{-1} (g_0 + g_1 Q_0),$$

что в реальном времени дает снова периодическую функцию («жесткий» тренд) с периодом L .

Если функция Q_t периодична (в частном случае синусоидальна) с некоторым периодом T_0 , т. е. $Q_{t+T_0} = Q_t$, то решение (7.57) также периодично с тем же периодом $q_{t+T_0}^* = q_t^*$. В реальном времени такое решение соответствует наложению (биениям) двух периодических колебаний с кратными периодами L и $T_0 L$, что соответствует сходимости прогноза сезонных значений вновь к строго периодическому тренду,

но уже с большим периодом $T_0 L$. Тот же характер будет иметь решение в случае, когда функция Q_t получена в виде суммы гармонических колебаний с *кратными* частотами — при этом Q_t периодична с периодом гармоники минимальной частоты: $T_0 = T_{\max} = 2\pi/\omega_{\min}$.

И только в случае непериодической функции Q_t в виде суммы гармонических колебаний с некратными частотами решение q_t^* (7.57) не будет сходиться к периодическому тренду. Оно будет ограниченной незатухающей функцией времени, имеющей характер типичного циклического процесса — на периодическую функцию q^0 накладывается аддитивная составляющая, которая изменяется от года к году и определяется прогнозом Q_t . Тем самым мы подтвердили известный из практики прогнозирования факт: для долгосрочного детального прогнозирования циклических природных процессов (сезонных колебаний) по двухуровневому алгоритму адекватным является применение на первом уровне (для детальных сезонных переменных) двухуровневых разностных моделей (7.35), (7.45), а на втором (для годовых значений) — гармонического алгоритма МГУА [22], дающего решения в виде суммы гармонических колебаний с некратными частотами.

В заключение заметим, что выражение (7.57) можно несколько упростить. Устойчивость матрицы H означает, что после окончания переходного процесса $H^T \approx 0$, $T > T^*$, поэтому в сумме, фигурирующей в (7.57), существенными оказываются только последние члены:

$$q_T^* = q^0 + g_1 \sum_{i=T-T^*}^T H^{T-i} Q_i, \quad T > T^*. \quad (7.57a)$$

Таким образом, на прогноз среднесезонных значений q_T^* влияют не все значения прогноза среднегодовых значений Q_i , $i = \overline{1, T}$, а только T^* последних.

7. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ АНАЛИЗА АЛГОРИТМОВ ДВУХУРОВНЕВОГО ПРОГНОЗИРОВАНИЯ

Применение принципа многоуровневого моделирования, разработанного в теории индуктивных методов самоорганизации моделей на ЭВМ [49], позволяет повысить предел предсказуемости долгосрочных прогнозов. Главным условием, обеспечивающим эффективность многоуровневых алгоритмов, является использование критерия баланса прогнозов для выбора наиболее согласованных моделей различных уровней. Например, при прогнозировании среднесезонных детальных значений различных показателей природных процессов по двухуровневому алгоритму критерий баланса отражает связь между среднесезонными (на нижнем уровне) и среднегодовыми (на верхнем уровне) значениями. Двухуровневое прогнозирование позволяет увеличить предсказуемость среднесезонных значений процесса до уровня предела предсказуемости среднегодовых значений.

В главе подробно исследованы свойства критерия баланса прогнозов и установлены условия его корректного применения. Показано, что он является типичным внешним критерием: в отсутствие шумов он равен нулю только для истинных моделей (связь между которыми он отражает), а при наличии шума он имеет всегда единственный минимум, указывающий модели оптимальной сложности. Причем при повышении уровня шума в среднем более эффективными (точными) оказываются все более простые модели. По своей природе критерий баланса применим как для алгебраических, так и для разностных моделей, причем он является *системным* — т. е. предназначен для оптимизации сложности моделей нескольких взаимосвязанных переменных. Этот критерий аналогичен критерию минимума ошибки интегрирования в смысле применимости на тех же точках, где вычисляются коэффициенты моделей — оба они не требуют разделения выборки данных на части.

Описана общая структура двухуровневого алгоритма долгосрочного прогнозирования циклических природных процессов с применением критерия баланса прогнозов. Подробно исследован вопрос об устойчивости двухуровневых разностных моделей для определения условий сходимости к тренду решений, получаемых по двухуровневому алгоритму. Показано, что необходимым условием эффективности двухуровневого алгоритма прогнозирования, кроме применения критерия баланса, является использование в моделях нижнего уровня (для сезонных значений) переменной верхнего уровня (среднегодовых значений). Показано также, что алгоритм, в котором на нижнем уровне по комбинаторному алгоритму МГУА самоорганизуются разностные двухуровневые модели, а на верхнем по гармоническому алгоритму МГУА — гармонические тренды с некратными частотами, позволяет получить долгосрочные прогнозы, адекватные моделируемым природным, экономическим и другим процессам. Примеры решения различных практических задач долгосрочного прогнозирования по двухуровневым алгоритмам можно найти в [49], а также в более ранних работах по разработке и применению индуктивных методов самоорганизации моделей на ЭВМ.

ПРИЛОЖЕНИЕ

Список программ МГУА, использующих принцип самоорганизации (на 1 января 1983 г.)

№	Название программы	Автор	Язык, ЭВМ	Публикация
1	Программа на языке АЛГОЛ-БЭСМ-6 для решения интерполяционных задач по алгоритмам МГУА с полиномами первой или второй степени	Ю. В. Коппа	АЛГОЛ-БЭСМ-6	Автоматика, 1971, № 1, а также в [43]
2	Программа на языке транслятора АЛГОЛ-БЭСМ-6 для решения задач самоорганизации математических моделей сложных объектов по алгоритмам МГУА с полиномами первой или второй степени без проекции отдельным переменным	Н. В. Гулян Б. К. Светальский Н. Н. Тодуа	»	Автоматика, 1972, № 6
3	Программа на языке транслятора АЛГОЛ-БЭСМ-6 для выделения тренда оптимальной сложности по алгоритмам МГУА с линейными полиномами	Н. Н. Тодуа	»	Автоматика, 1971, № 5
4	Программа на языке транслятора АЛГОЛ-БЭСМ-6 для алгоритма МГУА с последовательным выделением трендов по наиболее существенным аргументам	Н. Н. Тодуа Ю. В. Чукин	»	Автоматика, 1972, № 2
5	Программа на языке транслятора АЛГОЛ-БЭСМ-6 для выделения гармонического тренда оптимальной сложности	Н. Н. Тодуа Ю. В. Коппа	»	Автоматика, 1972, № 3
6	Программа на языке транслятора АЛГОЛ-БЭСМ-6 для многократного прогнозирования по полиномиальным алгоритмам МГУА	Н. Н. Тодуа Н. В. Гулян	»	Автоматика, 1973, № 2
7	Программа обобщенного алгоритма МГУА на языке транслятора АЛГОЛ-БЭСМ-6	Н. В. Гулян Н. Н. Тодуа	»	Автоматика, 1973, № 1
8	Программа на языке транслятора АЛГОЛ-БЭСМ-6 алгоритма МГУА со случайным выбором партнеров	Г. Петраке Н. Н. Тодуа	»	Автоматика, 1972, № 4

Продолжение табл.

№	Название программы	Автор	Язык, ЭВМ	Публикация
9	Программа синтеза моделей для долгосрочного прогнозирования	Н. А. Ивахненко В. Н. Высоцкий В. П. Боярко	ФОРТРАН, БЭСМ-6	Справочник [144]
10	Программа выделения прогнозирующего тренда оптимальной сложности по многорядному алгоритму МГУА	Н. Н. Тодуа	АЛГОЛ- ГДР, БЭСМ-6	Там же
11	Программа выделения гармонического тренда с некратными частотами для прогнозирования колебательных процессов	В. Н. Высоцкий В. И. Чеберкус	ФОРТРАН, БЭСМ-6	»
12	Программа самоорганизации прогнозирующих моделей циклических процессов	В. С. Степашко	АЛГОЛ- ГДР, БЭСМ-6	Справочник [144], а также [145]
13	Программа идентификации сложных объектов по МГУА	Ю. В. Коппа		Справочник [144]
14	Программа МГУА с ортогональными частными описаниями	О. И. Шелудько	»	Там же
15	Комбинаторная программа для структурной идентификации объектов и процессов управления	В. С. Степашко	»	Справочник [144], а также [146]
16	Программа идентификации сложной системы для оптимального управления (с обязательным включением заданных аргументов-управлений)	Б. К. Светальский	»	Справочник [144]
17	Программа многорядной селекции с применением комбинаторного перебора частных описаний	Н. А. Ивахненко А. А. Марчев	ФОРТРАН, ЕС ЭВМ (ДОС)	Там же
18	Программа МГУА с переобозначением переменных для моделирования динамических процессов	Г. И. Кротов	АЛГОЛ- ГДР, БЭСМ-6	»
19	Программа идентификации динамики полей сложных объектов на основе конечно-разностной аппроксимации	В. С. Степашко П. Ю. Пека	»	»
20	Программа идентификации многомерных полей по нерегулярной сетке	Ю. П. Юрачковский	»	»
21	Программа идентификации линеаризованных уравнений динамики сложных систем и исследование их устойчивости	В. Н. Высоцкий И. К. Тимченко	»	»

Продолжение табл.

№	Название программы	Автор	Язык, ЭВМ	Публикация
22	Программа управления с оптимизацией прогноза линейной дискретной системой с запаздыванием	Б. К. Светальский В. С. Степашко Л. М. Бойчук	АЛГОЛ- ГДР, БЭСМ-6	Справочник [144]
23	Программа оперативного управления каскадом водохранилищ с оптимизацией на скользящем интервале	В. С. Степашко Н. В. Полков	»	Там же
24	Программа выбора ансамбля информативных признаков на основе принципа самоорганизации	Ю. В. Коппа	»	»

СПИСОК ЛИТЕРАТУРЫ

1. Агеев М. И., Алик В. П., Марков Ю. И. Библиотека алгоритмов 516—1006. — М.: Сов. радио, 1976.—136 с.
2. Акишин Б. А., Ивахненко А. Г. О возможности экстраполяции (прогнозирования) по зашумленным данным при монотонном их изменении.— Автоматика, 1975, № 4, с. 22—29.
3. Акишин Б. А., Юнусов Н. Последовательное применение самоорганизации и статистической проверки гипотез для выбора адекватных математических моделей.— Там же, 1978, № 3, с. 64—67.
4. Альберт А. Регрессия, псевдоинверсия и рекуррентное оценивание.— М.: Наука, 1977. - 223 с.
5. Бард Й. Нелинейное оценивание параметров. - М.: Статистика, 1979. - 349 с.
6. Белозерский Е. А., Ивахненко Н. А., Юрачковский Ю. П. Об одном подходе к построению многорядных алгоритмов МГУА с линейными частными описаниями. - Автоматика, 1981, № 5, с. 3-7.
7. Бир С. Кибернетика и управление производством.— М.: Физматгиз, 1963.— 274с.
8. Бойчук Л. М. Метод расчета оптимального управления нелинейными дискретными объектами при наличии ограничений с использованием процедуры ортогонализации.— Кибернетика и вычисл. техника, 1980, вып. 49, с. 51—59.
9. Бойчук Л. М. Синтез многоточечных дискретных моделей непрерывных процессов и их применение для расчета оптимального управления. — Там же, 1981, вып. 53, с. 65—71.
10. Бойчук Л. М., Ивахненко Н. А. Динамические характеристики (переходные функции) линейных регрессионных моделей и их применение. Ч. 1—2. — Автоматика, 1983, № 4, с. 5—12, № 5, с. 15—22.
11. Бойчук Л. М., Козубовский С. Ф., Конна Ю. В. Численное решение нелинейных уравнений для обратного перехода от корреляционных функций к исходным реализациям случайного процесса.— Автоматика, 1983, № 2, с. 22—32.
12. Бойчук Л. М., Кондрашова Н. В. Метод построения многоточечных дискретных моделей непрерывных процессов. — Там же, 1980, № 6, с. 43—53.
13. Бойчук Л. М., Пинчук В. М. Метод синтеза многоточечных дискретных моделей оптимальной сложности непрерывных процессов на основе дробнополиномиальных приближений. - Там же, 1982, № 2, с. 22-31.
14. Бокс Дж., Дженкинс Г. Анализ временных рядов. - М.: Мир, 1974. - Вып. 1-2.
15. Вернадский В. И. Страницы автобиографии. - М.: Наука, 1981, - 300 с.
16. Взаимодействие природы и хозяйства Байкальского региона / Под ред. А. К. Айламазяна.— Новосибирск: Наука, 1981. - 125 с.
17. Винер Н. Кибернетика или управление и связь в животном и машине.— М.: Сов. радио, 1963. - 216 с.
18. Власова И. Л. Предварительные результаты численных экспериментов.— Тр. Гидрометцентра СССР, 1978, вып. 210, с. 41-58.
19. Власова И. Л., Сонечкин Д. М. Исследование предсказуемости в моделях системы гидродинамического типа. - Там же, с. 27—34.
20. Высоцкий В. Н. О наилучшем разделении исходных данных в алгоритмах МГУА - Автоматика, 1976, № 3, с. 71—74.
21. Высоцкий В. Н. Идентификация динамических систем на основе принципов самоорганизации.— В кн.: Математические модели для прогнозирования и управления качеством вод. Киев: ИК АН УССР, 1976, с. 42-58.
22. Высоцкий В. Н., Ивахненко А. Г., Чеберкус В. Я. Долгосрочное прогнозирование колебательных процессов при помощи выделения гармонического тренда оптимальной сложности по критерию баланса переменных.— Автоматика, 1975, № 1, с. 23—31.
23. Высоцкий В. Я., Ихара Д. Повышение помехоустойчивости критериев селекции МГУА при помощи векторного представления и использования минимаксных форм.—Там же, 1978, № 3, с. 3-11.
24. Высоцкий В. Я., Юнусов Я. Повышение помехоустойчивости алгоритма МГУА для выделения гармонического тренда с некратными частотами.— Там же, 1977, № 5, с. 72—76.
25. Гергей И. Обращение матриц и решение систем линейных и нелинейных уравнений методом окаймления. - Журн. вычисл. математики и мат. физики, 1979, 19, № 4, с. 803—810.
26. Глушков В. М. Введение в кибернетику. — Киев: Изд-во АН УССР, 1964.—324 с.
27. Гнеденко Б. В. Курс теории вероятностей.— М.: Наука, 1969. - 406 с.
28. Голованов Я. Я., Пономаренко В. С., Кузьменко Ю. И. Оптимизационный метод сокращения перебора моделей в комбинаторных алгоритмах МГУА.— Автоматика, 1983, № 3, с. 12-17.
29. Грош Д. Методы идентификации систем.— М.: Мир, 1979. - 302 с.
30. Дадаян В. С. Глобальные экономические модели.— М.: Наука, 1981. - 215 с.
31. Демиденко Е. З. Идентификация линейных эконометрических моделей.— Экономика и мат. методы, 1978, 14, вып. 6, с. 17-32.
32. Демиденко Е. З. Линейная и нелинейная регрессия.— М.: Финансы и статистика, 1981. - 302 с.
33. Джонс С. Определение наилучшей формы реакции сервомеханизма при наложении внешней случайной помехи на сигнал ошибки. - В кн.: Автоматическое регулирование: Сб. материалов конф. в Крэнфилде, 1951 г. М.: Изд-во иностр. лит., 1954, с. 141—150.
34. Димитров В. Д. Алгоритмы МГУА на размытых множествах Заде.— Автоматика, 1970, № 4, с. 50—57.
35. Дрейпер Я., Смит Г. Прикладной регрессионный анализ.— М.: Статистика, 1973.—392с.
36. Дубровский С. А. Прикладной многомерный статистический анализ.— М.: Финансы и статистика, 1982.—215 с.
37. Ершов А. А. Стабильные методы оценки параметров. - Автоматика и телемеханика, 1978, № 7, с. 66-101.
38. Ефимов А. Н. Предсказание случайных процессов.— М.: Знание, 1976.—64с.
39. Загускин В. Л. Численные методы решения плохо обусловленных задач.— Ростов: Изд-во Рост. ун-та, 1976.—187 с.
40. Ивахненко А. Г. Электроавтоматика.— Киев: Гостехиздат УССР, 1957.— 439с.
41. Ивахненко А. Г. Метод группового учета аргументов — конкурент метода стохастической аппроксимации.— Автоматика, 1968, № 3, с. 58—72.
42. Ивахненко А. Г. Самообучающиеся системы распознавания и автоматического управления.— Киев: Техніка, 1969.—392 с. „
43. Ивахненко А. Г. Системы эвристической самоорганизации в технической кибернетике.— Киев: Техтка, 1971.—372 с.
44. Ивахненко А. Г. Объективная структурная идентификация дифференциальных уравнений по малому числу наблюдений методами самоорганизации. — Автоматика, 1972, № 2, с. 43—57.
45. Ивахненко А. Г. Долгосрочное прогнозирование и управление сложными системами.— Киев: Техніка, 1975.—311 с.

46. Ивахненко А. Г. Метод группового учета аргументов в задачах прогнозирования.—Автоматика, 1976, № 6, с. 24—33.
47. Ивахненко А. Г. Моделирование водных экологических систем.—Гидробиол. журн., 1980, № 3, с. 12—21.
48. Ивахненко А. Г. Новые акценты в теории самоорганизации моделей.—Автоматика, 1981, № 6, с. 48—60.
49. Ивахненко А. Г. Индуктивный метод самоорганизации моделей сложных систем.—Киев : Наук. думка, 1982.—296 с.
50. Ивахненко А. Г. О «Золотой книге» сбывающихся долгосрочных количественных прогнозов.—Автоматика, 1982, № 2, с. 87—92.
51. Ивахненко А. Г. Кибернетика — наука о моделировании связи и управления в сложных системах.—Там же, 1982, № 1, с. 3—17.
52. Ивахненко А. Г. Особенности МГУА, реализуемые в алгоритме двухуровневого долгосрочного количественного прогноза.—Там же, 1983, № 2, с. 3—11.
53. Ивахненко А. Г., Белозерский Е. А., Козубовский С. Ф. Об открытии законов сложных объектов в случае, когда выходные переменные не указаны.—Там же, 1979, № 4, с. 15—23.
54. Ивахненко А. Г., Высоцкий В. Н., Ивахненко Н. А. Основные разновидности критерия минимума смещения модели и исследование их помехоустойчивости.—Там же, 1978, № 1, с. 32—53.
55. Ивахненко А. Г., Димитров В. Д. Вероятностные алгоритмы метода группового учета аргументов в задаче предсказания случайных событий.—Там же, 1969, № 3, с. 31—49.
56. Ивахненко А. Г., Дылбокова Д. Л., Сирбладзе Д. Д., Дылбоков И. С. Глубина минимума критерия селекции как показатель достоверности модели.—Там же, 1975, № 6, с. 13—18.
57. Ивахненко А. Г., Зайченко Ю. П., Димитров В. Д. Принятие решений на основе самоорганизации.—М. : Сов. радио, 1976.—275 с.
58. Ивахненко А. Г., Ивахненко Н. А. Долгосрочное прогнозирование случайных процессов по алгоритмам МГУА с применением критериев несмещенностии и баланса переменных.—Автоматика, 1974, № 4, с. 52—59.
59. Ивахненко А. Г., Ивахненко Н. А., Высоцкий В. Н. Долгосрочное прогнозирование случайных процессов по алгоритмам МГУА при критериях несмещенностии и баланса переменных. Ч. 3.—Там же, 1976, № 4, с. 21—35.
60. Ивахненко А. Г., Карпинский А. М. Самоорганизация моделей на ЭВМ в терминах общей теории связи (теории информации).—Там же, 1982, № 4, с. 7—26.
61. Ивахненко А. Г., Ковалчук П. И., Высоцкий В. Н. Моделирование динамики случайных полей загрязнения водоемов по алгоритмам МГУА.—Там же, 1974, № 3, с. 30—40.
62. Ивахненко А. Г., Козубовский С. Ф. Интервал корреляции как мера предела предсказуемости случайного процесса и детализации языка моделирования.—Там же, 1981, № 4, с. 3—10.
63. Ивахненко А. Г., Коноваленко В. В., Тулупчук Ю. Н., Тимченко И. К. Метод группового учета аргументов в задаче распознавания образов и принятия решений.—Там же, 1968, № 5, с. 42—53.
64. Ивахненко А. Г., Коппа Ю. В. Алгоритмы метода группового учета аргументов (МГУА) с линейными операторами.—Там же, 1969, № 4, с. 69—78.
65. Ивахненко А. Г., Коппа Ю. В. Выбор ансамбля признаков и синтез многоядного перцептрона по принципу самоорганизации.—Там же, 1974, № 2, с. 41—53.
66. Ивахненко А. Г., Коппа Ю. В., Ву Суан Минь. Полиномиальная и логическая теория динамических систем. Ч. 1.—Там же, 1970, № 3, с. 3—17.
67. Ивахненко А. Г., Коппа Ю. В., Лантайова Д. Н., Ивахненко Н. А. Связь теории самоорганизации моделей на ЭВМ и теории распознавания образов.—Там же, 1980, № 3, с. 3—13.
68. Ивахненко А. Г., Коппа Ю. В., Тимченко И. К., Ивахненко Н. А. Метод группового учета аргументов в задаче идентификации статической характеристики многоэкстремального объекта.—Там же, 1969, № 2, с. 43—53.
69. Ивахненко А. Г., Коппа Ю. В., Тодуя Н. Н., Петраке Г. Метод математического моделирования сложных экологических систем.—Там же, 1971, № 4, с. 20—34.
70. Ивахненко А. Г., Костенко Ю. В. Системный анализ и долгосрочный количественный прогноз квазистатических систем на основе самоорганизации моделей. Ч. 1. Системный анализ на уровне трендов.—Там же, 1982, № 3, с. 11—19.
71. Ивахненко А. Г., Костенко Ю. В., Голеусов И. В. Системный анализ и долгосрочный количественный прогноз квазистатических систем на основе самоорганизации моделей. Ч. 2. Объективный системный анализ без априорного указания внешних воздействий.—Там же, 1983, № 3, с. 3—11.
72. Ивахненко А. Г., Кротов Г. И. Моделирование загрязнения среды при отсутствии информации о выбросах источников загрязнения.—Там же, 1977, № 5, с. 14—30.
73. Ивахненко А. Г., Кротов Г. И. Самоорганизация моделей с переменными коэффициентами для оптимизации проектирования водных объектов.—Там же, 1980, № 6, с. 11—30.
74. Ивахненко А. Г., Кротов Г. И., Степашко В. С. Гармонические и экспоненциально-гармонические алгоритмы МГУА.—Там же, 1983, № 1, с. 3—12.
75. Ивахненко А. Г., Кротов Г. И., Чеберкус В. И. Многорядный алгоритм самоорганизации долгосрочных прогнозов (на примере экологической системы оз. Байкал).—Там же, 1980, № 4, с. 28—47.
76. Ивахненко А. Г., Кротов Г. И., Чеберкус В. И. Гармонические и экспоненциально-гармонические алгоритмы МГУА для долгосрочного прогноза колебательных процессов. Ч. 1. Однорядные алгоритмы.—Там же, 1981, № 3, с. 17—31.
77. Ивахненко А. Г., Кротов Г. И., Юрачковский Ю. П. Экспоненциально-гармонический алгоритм МГУА.—Там же, № 2, с. 23—30.
78. Ивахненко А. Г., Лапа В. Г. Предсказание случайных процессов.—Киев : Наук. думка, 1971.—416 с.
79. Ивахненко А. Г., Мюллер И. А. Самоорганизация прогнозирующих моделей.—Киев : Техника, 1985.—221 с.
80. Ивахненко Ж. Г., Оsipенко В. В. О регуляризации решения задачи обратного развертывания матриц переходов в детальный прогноз.—Там же, 1982, № 5, с. 37—41.
81. Ивахненко А. Г., Оsipенко В. В., Строкова Т. И. Прогнозирование двухмерных физических полей при помощи обратного преобразования матриц вероятности перехода.—Там же, 1983, № 4, с. 12—18.
82. Ивахненко А. Г., Рао М. Прогнозирование и экстраполяция метеорологических полей методами самоорганизации моделей.—Там же, 1979, № 6, с. 18—35.
83. Ивахненко А. Г., Саэтальский Б. К. Самоорганизация модели развития мира и синтез управления по методу перебора вершин гиперкуба возможных управлений.—Там же, 1975, № 1, с. 32—50.
84. Ивахненко А. Г., Спину Е. И., Димитров В. Д. и др. Распознавание продолжительности деструкции и содержания пестицидов в растениях по вероятностному алгоритму МГУА.—Там же, 1970, № 5, с. 42—52.
85. Ивахненко А. Г., Степашко В. С. Самоорганизация моделей и долгосрочное прогнозирование речного стока при помощи балансового критерия.—Там же, 1975, № 5, с. 34—41.
86. Ивахненко А. Г., Степашко В. С. Численное исследование помехоустойчивости многокритериальной селекции моделей.—Там же, 1982, № 4, с. 26—35.
87. Ивахненко А. Г., Степашко В. С., Костенко Ю. В. и др. Самоорганизация комбинированных моделей для прогнозирования циклических процессов с применением критерия баланса прогнозов.—Там же, 1979, № 2, с. 8—21.
88. Ивахненко А. Г., Степашко В. С., Семина Л. П. Долгосрочное прогнозирование среднемесячного стока рек по принципу самоорганизации.—В кн.: Самоорганизующиеся кибернетические системы. Киев : ИК АН УССР, 1974, с. 3—14.
89. Ивахненко А. Г., Степашко В. С., Хомовченко М. Г., Галлямин Е. П. Самоорганизация моделей динамики роста сельскохозяйственных культур для управления орошаемым севооборотом.—Автоматика, 1977, № 5, с. 32—44.

90. Ивахненко А. Г., Тодуя Н. Н., Гулян Н. В., Светальский Б. К. Проблемы статистического прогнозирования случайных процессов по принципу самоорганизации уравнений прогноза. Ч. 3.— Там же, 1973, № 3, с. 15—36.
91. Ивахненко А. Г., Тодуя Н. Н., Чукин Ю. В. Алгоритм МГУА с последовательным выделением полиномиальных трендов по наиболее существенным аргументам.— Там же, 1972, № 2, с. 60—72.
92. Ивахненко А. Г., Шервашидзе В. В., Шелудько О. И. и др. Открытие физических законов методом МГУА по критерию несмещенности.— Там же, 1973, № 6, с. 47—62.
93. Ивахненко А. Г., Юрачковский Ю. П. Выбор оптимального управления, принадлежащего области эффективных решений Парето, на основе долгосрочного прогнозирования.— Там же, 1978, № 2, с. 39—43.
94. Ивахненко Н. А., Кваско М. З. Комбинаторные алгоритмы МГУА, в которых экзамены на регулярность подлежат как «симметричные», так и «несимметричные» полиномы.— Там же, 1972, № 5, с. 42—48.
95. Ивахненко Н. А., Марчев А. А. Самоорганизация математической модели для перспективного планирования строительно-монтажных работ.— Там же, 1978, № 3, с. 12—18.
96. Ивахненко Н. А., Тимченко И. К. Вычислительные программы идентификации сложных объектов, реализующие алгоритмы метода группового учета аргументов (МГУА).— Там же, 1977, № 4, с. 78—81.
97. Игнатов В. А. Теория информации и передачи сигналов.— М.: Сов. радио, 1979.—280 с.
98. Имитационное моделирование экологических систем.— М.: Наука, 1978.—218 с.
99. Ихара Д. О единственности выбора модели по критерию баланса переменных (Письмо в ред.).— Автоматика, 1976, № 1, с. 86—87.
100. Казарин Р. А., Кувшинов Б. И., Назаров М. В. Элементы общей теории связи.— М.: Госэнергоиздат, 1957.—96 с.
101. Каминская В. А. Идентификация динамических систем по дискретным наблюдениям. Ч. 1. Основы статистических методов, оценивание параметров линейных систем.— Вильнюс: Мокслас, 1982.—245 с.
102. Кашиян Р. Л., Рао А. Р. Построение динамических стохастических моделей по экспериментальным данным.— М.: Наука, 1983.—388 с.
103. Кендэл М. Временные ряды.— М.: Финансы и статистика, 1981.—199 с.
104. Кикоть В. С. Планирование эксперимента в задачах самоорганизации математических моделей.— Автоматика, 1984, № 1, с. 32—39.
105. Козубовский С. Ф. Специализированная вычислительная система «Гиперкомп-80».— Зарубеж. радиоэлектрон., 1974, № 10, с. 3—9.
106. Козубовский С. Ф., Юрачковский Ю. П. Информационные критерии селекции моделей.— Автоматика, 1981, № 4, с. 80—89.
107. Колмогоров А. Н. Три подхода к определению понятия количества информации.— Пробл. передачи информ., 1965, 1, вып. 1, с. 18—39.
108. Кондрашова Н. В. Исследование точности и помехоустойчивости многоточечных дискретных преобразований.— Автоматика, 1983, № 1, с. 49—56.
109. Кондрашова Н. В., Бойчук Л. М. Идентификация непрерывных систем при помощи многоточечных дискретных моделей.— Там же, 1981, № 4, с. 14—20.
110. Коппа Ю. В. Программа на языке транслятора АЛГОЛ-БЭСМ-6 для решения интерполяционных задач по алгоритмам МГУА с полиномами первой или второй степени.— Там же, 1971, № 1, с. 88—92.
111. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров.— М.: Наука, 1974.—832 с.
112. Котельников В. А. Теория потенциальной помехоустойчивости.— М.: Госэнергоиздат, 1956.—256 с.
113. Крицкий С. Н., Коренистов Д. В., Раткович Д. Я. Колебания уровня Каспийского моря.— М.: Наука, 1975.—160 с.
114. Крошиц Ф. В., Крехин Е. М., Меншуткин В. В. Сообщество пелагических рыб озера Дальнего (опыт кибернетического моделирования).— Л.: Наука, 1969.—300 с.
115. Кротов Г. И. О соответствии непрерывных и дискретных динамических моделей, имеющих незатухающее колебательное решение.— В кн.: Самоорганизация кибернетических систем.— Киев: ИК АН УССР, 1980, с. 25—29.
116. Кротов Г. И. Исследование устойчивости некоторых двухуровневых алгоритмов долгосрочного прогноза.— Автоматика, 1981, № 5, с. 72—76.
117. Кротов Г. И., Степашко В. С. Способы сокращения необходимой оперативной памяти ЭВМ при использовании алгоритмов МГУА.— Там же, 1978, № 4, с. 89—91.
118. Ланкастер П. Теория матриц.— М.: Наука, 1978.—280 с.
119. Ли Р. Оптимальные оценки, определение характеристик и управление.— М.: Наука, 1966.—176 с.
120. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования.— М.: Статистика, 1979.—254 с.
121. Маригодов В. К. Помехоустойчивая обработка информации.— М.: Наука, 1983.—199 с.
122. Маттес Б., Мюллер И. Прогнозирование временных рядов с помощью динамических моделей.— Автоматика, 1982, № 4, с. 36—48.
123. Моисеев Н. Н. Математика ставит эксперимент.— М.: Наука, 1979.—223 с.
124. Налимов В. В. Теория эксперимента.— М.: Наука, 1971.—207 с.
125. Найдель Э., Ньюмен Д. Теорема Геделя: Пер. с англ.— М.: Знание, 1970.—62 с.
126. Ортега Д., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными.— М.: Мир, 1975.—420 с.
127. Отнес Р., Эноксон Л. Прикладной анализ временных рядов.— М.: Мир, 1982.—428 с.
128. Павлов А. А. Критерий ранжировки для порогового самоотбора переменных в алгоритмах МГУА.— Автоматика, 1969, № 4, с. 89—91.
129. Пелед А., Лиу Б. Цифровая обработка сигналов.— Киев: Вища шк., 1979.—264 с.
130. Перцептрон — система распознавания образов / Под ред. А. Г. Ивахненко.— Киев: Наук. думка, 1975.—430 с.
131. Печчини А. Человеческие качества.— М.: Прогресс, 1980.—300 с.
132. Пирс Дж. Символы, сигналы, шумы.— М.: Мир, 1967.—334 с.
133. Познянский Ю. А. Метод множественной корреляции для свертки нескольких противоречивых критериев выбора оптимального решения.— Автоматика, 1982, № 4, с. 82—86.
134. Программы прямого синтеза моделей (по принципу самоорганизации).— Киев: ИК АН УССР, 1975. (Сб. Респ. фонда алгоритмов и программ; Вып. 1—3).
135. Продуктивная зона экваториальной Атлантики и условия ее формирования / Под ред. К. А. Седых.— Тр. АтлантНИРО, 1971, вып. 37.—470 с.
136. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов.— М.: Мир, 1978.—848 с.
137. Раскин В. Г., Яновский Л. П. О выборе регрессионной модели оптимальной сложности с помощью критерия регулярности.— Автоматика, 1984, № 1, с. 40—43.
138. Светальский Б. К. Методы и алгоритмы идентификации на основе принципов самоорганизации для прогнозирования и оптимального управления в сложных системах: Автореф. дис. ... канд. техн. наук.— Киев, 1980.—24 с.
139. Светальский Б. К., Ковальчук П. И. Многорядный алгоритм МГУА с селекцией первичных аргументов.— Автоматика, 1979, № 4, с. 31—35.
140. Свобода выбора решений в последующие моменты времени как важный критерий оптимальности управления сложными случайными процессами.— Там же, 1972, № 2, с. 82—90.
141. Себер Дж. Линейный регрессионный анализ.— М.: Мир, 1980.—456 с.
142. Сирл С., Госман У. Матричная алгебра в экономике.— М.: Статистика, 1974.—374 с.
143. Смоляк С. А., Титаренко Б. П. Устойчивые методы оценивания.— М.: Статистика, 1980.—206 с.
144. Справочник по типовым программам моделирования / Под ред. А. Г. Ивахненко.— Киев: Техника, 1980.—182 с.

145. Степашко В. С. Алгоритм самоорганизации моделей среднемесячного стока рек.—Киев, 1977.—15 с.—(Респ. фонд. алгоритмов и программ; № 4629).
146. Степашко В. С. Комбинаторный алгоритм МГУА для структурной идентификации объектов управления с применением различных критериев селекции.—Киев, 1978.—18 с.—(Респ. фонд. алгоритмов и программ; № 4951).
147. Степашко В. С. Оптимизация и обобщение схем перебора моделей в алгоритмах МГУА.—Автоматика, 1979, № 4, с. 36—43.
148. Степашко В. С. Комбинаторный алгоритм МГУА с оптимальной схемой перебора моделей.—Там же, 1981, № 3, с. 31—36.
149. Степашко В. С. Потенциальная помехоустойчивость моделирования по комбинаторному алгоритму МГУА без использования информации о помехах.—Там же, 1983, № 3, с. 18—27.
150. Степашко В. С. Конечная селекционная процедура сокращения полного перебора моделей.—Там же, № 4, с. 84—88.
151. Стоун Р. Метод «Затраты—выпуск» и национальные счета.—М.: Статистика, 1966.—205 с.
152. Стоун Р., Ивахненко А. Г., Высоцкий В. Н., Семина Л. П. Открытие законов сложных объектов в случае, когда характеристические переменные не указаны.—Автоматика, 1977, № 6, с. 27—39.
153. Стрэнг Г. Линейная алгебра и ее применения.—М.: Мир, 1980.—454 с.
154. Тарасенко Н. Д., Лукашова Г. И. Что вы знаете о своей наследственности?—Новосибирск: Наука, 1980.—89 с.
155. Типовые линейные модели объектов управления / Под ред. Н. С. Райбмана.—М.: Энергоатомиздат, 1983.—264 с.
156. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач.—М.: Наука, 1979.—285 с.
157. Тукиев А. Д. Метод направляющего вектора и его приложения.—Мат. вопр. кибернетики и вычисл. техники, 1982, вып. 11, с. 116—146.
158. Уилкинсон Дж., Райнш К. Справочник алгоритмов на языке АЛГОЛ: Линейная алгебра.—М.: Машиностроение, 1976.—390 с.
159. Успенский В. А. Теорема Геделя о неполноте.—М.: Наука, 1982.—112 с.—(Попул. лекции по математике).
160. Флейшман Б. С. Основы системологии.—М.: Радио и связь, 1982.—367 с.
161. Форрестер Дж. Мировая динамика.—М.: Наука, 1978.—165 с.
162. Халанай А., Векслер Д. Качественная теория импульсных систем.—М.: Мир, 1971.—309 с.
163. Харкевич А. А. Очерки общей теории связи.—М.: Гостехиздат, 1955.—268 с.
164. Харкевич А. А. Борьба с помехами.—М.: Физматгиз, 1965.—276 с.
165. Хемминг Р. В. Цифровые фильтры.—М.: Сов. радио, 1980.—224 с.
166. Хомовченко М. Г., Коломиц Н. Г. Самоорганизация системы простых моделей, прогнозирующих урожай пшеницы.—Автоматика, 1980, № 1, с. 28—37.
167. Худсон Д. Статистика для физиков.—М.: Мир, 1970.—296 с.
168. Чучадеев Д. Н. Информационный подход в теории распознавания образов.—Автоматика, 1981, № 3, с. 5—16.
169. Шелудько О. И. Алгоритм МГУА с ортогонализированным полным описанием для синтеза моделей по результатам планируемого эксперимента.—Там же, 1974, № 5, с. 32—41.
170. Шелудько О. И., Патерев С. Г. Упрощенный алгоритм идентификации характеристик сложных объектов по МГУА.—В кн.: Программы прямого синтеза моделей (по принципу самоорганизации). Киев: РФАП, 1975, вып. 1.
171. Шелудько О. И., Теряева Г. Н. Некоторые робастные методы оценивания в алгоритмах эвристической самоорганизации.—Автоматика, 1983, № 1, с. 18—26.
172. Шенкок К. Работы по теории информации и кибернетике.—М.: Изд-во иностр. лит., 1963.—828 с.
173. Юрачковский Ю. П. Усовершенствованные алгоритмы МГУА для прогнозирования процессов (обзор).—Автоматика, 1977, № 5, с. 76—87.
174. Юрачковский Ю. П. Сходимость многорядных алгоритмов МГУА.—Там же, 1981, № 3, с. 32—36.
175. Юрачковский Ю. П. Восстановление полиномиальных зависимостей на основе самоорганизации.—Там же, № 4, с. 15—20.
176. Юрачковский Ю. П. Структурное моделирование по выборкам наблюдений.—Там же, 1983, № 1, с. 30—38.
177. Юрачковский Ю. П., Грошков А. Н. Применение канонической формы внешних критериев для исследования их свойств.—Там же, 1979, № 3, с. 85—89.
178. Юрачковский Ю. П., Грошков А. Н. Оптимальное разбиение исходных данных на обучающую и проверочную последовательности на основе анализа функций распределения критерия.—Там же, 1980, № 2, с. 5—9.
179. Юрачковский Ю. П., Маулик М. Ю., Шервашидзе В. В. Моделирование по выборкам точных наблюдений при наличии информации о производной выходной величины.—В кн.: Имитационные и самоорганизующиеся модели сложных систем. Киев: ИК АН УССР, 1982, с. 3—13.
180. Юрачковский Ю. П., Чеберкус В. И. Самоорганизация моделей полигармонических полей.—Автоматика, 1981, № 1, с. 37—45.
181. Aoki J. Optimization of environmental planning under the risk-aversion of non-reparable damage. — Sum. Pap. pen. Fuzzy Probl., 1976, N 2, p. 1-13.
182. Akaike H. A new look at the statistical model identification.—IEEE Trans. Automat. Contr. 1974, 19, N 3, p. 716-723.
183. Beck M. B. Model structure identification from experimental data.—In: Theoretical system ecology. New-York — San-Francisko — London: Acad. press inc., 1979, p. 354—387.
184. Cupp E. E. Seasonal distribution and occurrence of marine diatoms and dinoflagellates at Scotch Cap Alaska. — Bull. Scripps Inst. Oceanogr. Univ. Calif., Techn. Ser., 1937, 4, N 3, p. 71—100.
185. Duffy J., Franklin M. A case study of environmental system modelling with the GMDH. — In: Proc. 1973 JACC Ohio State Univ. Columbus: 1973, p. 101—111.
186. Duffy J., Franklin M. According identification algorithm and its application to an environmental system. — IEEE Trans. Syst. Man and Cybern., 1975, 5, N 2, p. 226—240.
187. Gabor D. Cybernetics and the future of industrial civilization. — J. Cybern., 1971, 2, N 1, p. 1—4.
188. Ihara Dz. The use of protected form at GMDH, for the world population modelling.—Systems and Control, 1973, N 2, p. 21—36.
189. Ikeda S., Ochiai M., Sawaragi Y. Sequential GMDH algorithm and its application to river flow prediction. — IEEE Trans. Syst., Man and Cybern., 1976, 6, N7, p. 473—479.
190. Ivakhnenko A. G. Heuristic self—organization in problems of engineering cybernetics. — Automatica, 1970, N 6, p. 207—219.
191. Ivachnenko A. G. Lösung von Interpolationproblemen der Technische Kybernetik nach der Methode der gruppenweisen Argumentberechnung. — Messen, Steuern, Regeln, 1970, 13, N 4, S. 134—138.
192. Ivakhnenko A. G. Polynomial theory of complex systems.—IEEE Trans Syst., Man and Cybern., 1971, 1, N4, p. 364—378.
193. Ivakhnenko A. G. Problems of complex system modelling and applied mathematical statistics.—Int. J. Comput. and Inform. Sci., 1973, 2, N 1, p. 16—28.
194. Ivakhnenko A. G. The GMDH in long range forecasting.—In: Technological forecasting and social changes.—Elsevier, 1978, p. 223—227.
195. Ivakhnenko A. G. Development models of optimal complexity using self-organization theory. — Int. J. Comput. and Inform. Sci., 1979, 8, N 2, p. 12—21.
196. Ivakhnenko A. G., Ivakhnenko N. A., Visotsky V. N. Longfristige Prognose von Zufallprozessen nach dem Algorithmus der gruppenweisen Argumentberechnung (MGAB). —Meessen, Steuern, Regeln, 1977, 20, N 10, S. 542—549.
197. Ivakhnenko A. G., Krotov G. I., Visolsky V. N. Identification of the mathematical model of complex system by the self-organization method.—In: Theoretical systems ecology. — New-York — San-Francisco — London: Acad. press, inc., 1979, p. 320—353.
198. Ivakhnenko A. G., Muller J.-A. Selbstorganisation von Vorherzagemodellen.—Berlin: VEB Verlag Technik, 1984. - 270 S.

199. Kasperowicz S. K. Modelling complex systems with multinomial networks. — In: IEEE Region 6. Conf. US West. Reg., Honolulu, Hawai, 1973, p. 203—207.
200. Kondo J. Air pollution.— Tokyo: Korono C., 1975. - 300 p.
201. Krepler K. Ein Heuristisches Optimierungverfahren nach der Methode der gruppenweisen Argumentberechnung. - Messen, Steuern, Regeln, 1972, 15, H. 10, S. 369-371.
202. Maciejowsky J. M. The modelling of systems with small observation sets. — Int. J. Comput. and Inform. Sci., 1978, 10, N 6, p. 242—247.
203. Martin W., Muller J.-A. Zur Planung von Modellexperimenten fur komplizierte Modelle. — Messen, Steuern, Regeln, 1983, 26, H. 9, S. 495—497.
204. Mehra R. K. GMDH reviews and experience. — In: Proc. IEEE Conf. on decision and control, New Orleans, 1977, p. 29—34.
205. Muller J. -A., Ivachnenko A. G. Prinzipien der longfristigen Vorhersage auf der Grundlage der Selbst organisation. - Messen, Steuern, Regeln, 1983, 26 H. 7, S. 362-367.
206. Muller J. A., Ivachnenko A. G. Selbstorganization mathematischer Modelle auf der EDV. Anlage und Ihre Anwendung zur Vorhersage.— Ibid., 1980, 23, H. 11, S. 645-648.
207. Pares P. S., Ivachnenko A. G. et al. A self-organization model of the british economy for control with optimal prediction using balans variable criterion. — Int. J. Comput. and Inform. Sci., 1975, 4, N 4, p. 349-379.
208. Sawaragi Y., Soeda T., Tamura H. et al. Statistical prediction of air pollution levels using non-physical models. — Automatica, 1979, 15, N4, p. 453—460.
209. Self-organising methods in modelling: GMDH Type Algorithms / Edited by S. J. Farlow.— Statistics: Textbooks and Monographs, Vol. 54.— New York, Basel: Marcel Decker Inc., 1984. - 350 p.
210. Shankar R. The GMDH: Master thes. univ. of Delaware, 1972. - 250 p.
211. Scott D. S., Hatchison Ch. E. Modelling of economical systems.— Boston: Mass. Techn. Inst., 1975. —115 p.
212. Tamura H., Kondo T. Large spatial pattern identification of air pollution by computer model of source reseptor matrix and revised GMDH.— In: Proc. IFAC Symp. an env. syst. plan, design and control. Kyoto, 1977, p. 167—171.
213. Usher M. B. Developments in the Leslie matrix model.— In: Mathematical model in ecology. — Oxford, 1972, p. 29—69.
214. Widrow B. Generalization and information storage in networks of ADALINE neurons.— In: Self-organizing systems. Washington: Spartan - Books, 1962, p. 21-38.
215. Zyl van J. G. Experimental in socio economic forecasting using Ivakhnenko's approach. - Appl. Modelling, 1978, 2, N3, p. 49-56.

**Алексей Григорьевич Ивахненко
Владимир Семенович Степашко**

**ПОМЕХОУСТОЙЧИВОСТЬ
МОДЕЛИРОВАНИЯ**

Утверждено к печати ученым советом
Института кибернетики им. В. М. Глушкова
АН УССР

Редактор Т. С. Мельник
Художественный редактор И. П. Антонюк
Технический редактор Т. С. Бerezяк
Корректоры Т. Я. Чорная, Л. А. Коваржик

Информ. бланк № 6915.

Сдано в набор 08.08.84. Подп. в печ. 14.01.85.
БФ 01508. Формат 60×90/16. Бум. тип. № 1. Лит.
гари. Выс. печ. Усл. печ. л. 13,5. Усл. кр.-отт.
13,5. Уч.-изд. л. 14,97. Тираж 1500 экз. Заказ
№ 4-561. Цена 2 р. 60 к.

Издательство «Наукова думка», 252001 Киев 4,
ул. Репина, 3.

Киевская книжная типография научной книги.
252004 Киев 4, ул. Репина, 4.