

В.А. ФУРСОВ

**ИДЕНТИФИКАЦИЯ СИСТЕМ
ПО МАЛОМУ ЧИСЛУ
НАБЛЮДЕНИЙ**

2007



САМАРА

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»

В.А. ФУРСОВ

ИДЕНТИФИКАЦИЯ СИСТЕМ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

*Утверждено Редакционно-издательским советом
в качестве учебного пособия*

САМАРА
Издательство СГАУ
2007

УДК 004.9(075)
ББК 39.87
Ф 954



**Инновационная образовательная программа
«Развитие центра компетенции и подготовка
специалистов мирового уровня в области аэро-
космических и геоинформационных технологий»**

Рецензенты: д-р физ.-мат. наук А. И. Жданов,
д-р физ.-мат. наук В. М. Чернов

Фурсов В.А.

Ф 954 Идентификация систем по малому числу наблюдений: учеб. пособие / В.А. Фурсов – Самара: Самар. гос. аэрокосм. ун-т. 2007. – 80 с.: ил.

ISBN 978-5-7883-0626-1

В учебном пособии рассматриваются сравнительно новые подходы к построению оценок параметров систем в условиях неопределенности, связанной с малым числом наблюдений. Развиваемый подход основан на нетрадиционных предположениях, связанных с тем, что при малом числе наблюдений статистическая информация ненадежна либо отсутствует. Изучаются методы контроля информативности малых наборов данных. Приводятся алгоритмы оценивания, работоспособные в условиях статистической неопределенности.

Подготовлено и издано при финансовой поддержке Министерства образования и науки РФ, Администрации Самарской области и Американского фонда гражданских исследований и развития (CRDF).

УДК 004.9(075)
ББК 39.87

ISBN 978-5-7883-0626-1

© В.А. Фурсов, 2007
© Самарский государственный
аэрокосмический университет, 2007

ОГЛАВЛЕНИЕ

| | |
|---|----|
| Введение | 4 |
| 1. Проблема идентификации по малому числу наблюдений..... | 6 |
| 1.1. Понятие и формулировка задачи параметрической идентификации | 6 |
| 1.2. Модели идентифицируемых систем | 8 |
| 1.3. Проблема оценивания по малому числу наблюдений | 12 |
| 1.4. Формы представления и ортогональные разложения ошибок | 19 |
| 1.5. Оценки достижимой точности | 25 |
| 1.6. Прогнозирование относительных ошибок оценивания с учетом погрешностей округления..... | 31 |
| 2. Методы контроля информативности данных..... | 35 |
| 2.1. Меры обусловленности и мультиколлинеарности..... | 35 |
| 2.2. Оценка информативности данных по показателям диагонального преобладания | 40 |
| 2.3. Связь показателей информативности данных..... | 45 |
| 3. Алгоритмы идентификации | 52 |
| 3.1. Итерационное оценивание с корректировкой выходного вектора... | 52 |
| 3.2. Итерационный алгоритм со взвешиванием данных..... | 56 |
| 3.3. Идентификация на основе метода согласованных оценок | 61 |
| 3.4. Алгоритм идентификации с отбором данных по показателям информативности и согласованности..... | 70 |
| Список литературы | 77 |
| Основная | 77 |
| Дополнительная | 77 |

ВВЕДЕНИЕ

Теории и методам идентификации систем посвящена обширная научная и учебная литература. Сложился традиционный взгляд на задачи идентификации как задачи статистического оценивания. При идентификации моделей систем по малому числу наблюдений в рамках статистической теории оценивания возникают следующие трудности.

Информативность исходных данных существенно зависит от выбора интервала наблюдения сигнала. Например, при попадании на участок установившегося процесса задача оказывается плохо обусловленной, что может привести к большим ошибкам в оценках параметров.

Свойство устойчивости статистических характеристик шумов при малом числе наблюдений не проявляется в полной мере, поэтому методы идентификации, основанные на использовании априорных вероятностных распределений помех, оказываются теоретически необоснованными и, как следствие, неработоспособными.

Таким образом, использование теории и методов идентификации, опирающихся на статистическую теорию оценивания и использующих априорные вероятностные распределения, не соответствует существу задачи оценивания по малому числу наблюдений. Для повышения точности идентификации в условиях неопределенности характеристик информативности полезных сигналов обычно применяют регуляризацию, основанную на теории некорректных задач [20]. При статистической регуляризации обычно предполагается, что объем выборки достаточно большой.

Если модель полезного сигнала задана и такова, что приводит к плохой обусловленности задачи [3, 21], эта проблема традиционно считается вычислительной. Это нашло отражение в разработке алгебраической теории возмущений [2]. Проблема обусловленности тесно связана с решением полной или частной проблемы собственных значений. Указанное направление исследований имеет большое самостоятельное значение, возможно поэтому эти результаты, как правило, не учитывают некоторые важные особенности оценивания по малому числу наблюдений.

При решении задачи идентификации на малых фрагментах сигналов их свойства и связанные с ними характеристики обусловленности существенным образом изменяются от фрагмента к фрагменту. Поэтому точностные

свойства оценок по малому числу наблюдений с самого начала должны изучаться с учетом совместного влияния обоих факторов: ошибок в исходных данных и характеристик обусловленности (информативности полезных сигналов). В рамках статистической теории оценивания это не представляется возможным.

В настоящем учебном пособии рассматриваются сравнительно новые подходы к построению оценок параметров систем в условиях неопределенности, связанной с малым числом наблюдений. Развиваемый подход основан на нетрадиционных предположениях. Изучаются методы контроля информативности малых наборов данных, приводятся алгоритмы оценивания, работоспособные в условиях статистической неопределенности.

Учебное пособие в основном ориентировано на подготовку магистров и аспирантов.

1. ПРОБЛЕМА ИДЕНТИФИКАЦИИ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

1.1. Понятие и формулировка задачи параметрической идентификации

Идентификацией называют определение математической модели объекта по измерениям его входа – выхода и априорной информации из определенного класса моделей, которым данный объект эквивалентен. Эквивалентность понимается в смысле какой-либо функции потерь, являющейся функционалом от разности входных сигналов объекта и модели. При этом две модели считаются эквивалентными, если значения функций потерь для этих моделей одинаковы [30].

Если априорная информация об идентифицируемом объекте отсутствует или очень бедная, приходится предварительно осуществлять выбор структуры системы и класса моделей. Обычно эта задача решается на ранних стадиях проектирования системы с использованием идей планирования эксперимента. В ситуациях, с которыми приходится сталкиваться на практике, речь обычно идет об определении параметров модели объекта известного типа (структуры). В отличие от общего случая эту задачу называют параметрической идентификацией.

Задача параметрической идентификации в общем случае формулируется в виде операторного уравнения:

$$y = Fx, \tag{1.1}$$

где F – подлежащий определению и действующий на x математический оператор. Для многомерного объекта y и x – векторы, а в случае одномерного – скаляры. В соответствии с определением понятия параметрической идентификации предполагается, что априори известны класс и структура уравнений, физический смысл и размерности векторов y и x . Требуется на основании совокупности измерений y и x или некоторых других зависящих от них величин определить оператор F .

Поскольку измерения содержат ошибки, истинный оператор найти невозможно и ищется наилучшая в некотором смысле оценка \hat{F} оператора F . Задаваемый критерий качества оценок, как правило, зависит от характера доступной априорной информации. В частности, если известны статистиче-

ские характеристики входных и выходных сигналов объекта, критерии качества оценок целесообразно применять также статистические.

Оператор F может быть известен с точностью до векторного параметра \mathbf{c} т.е. уравнение (1.1) имеет вид

$$\mathbf{y} = F(\hat{\mathbf{n}})\mathbf{x}, \quad (1.2)$$

где $F(\hat{\mathbf{n}})$ - оператор, известным образом зависящий от неизвестного параметра \mathbf{c} . В этом случае ищется

$$\hat{\mathbf{c}}: Q(\hat{\mathbf{c}}) = \min_{\mathbf{c}} Q(\mathbf{c}), \quad (1.3)$$

где $Q(\mathbf{c})$ – некоторый критерий, характеризующий качество восстановления сигнала. Вид критерия качества определяется доступной априорной информацией. В наиболее простом частном случае уравнение (1.2) может быть задано в виде

$$y = \tilde{\mathbf{n}}^T \mathbf{x} + \xi, \quad (1.4)$$

где y – наблюдаемая выходная координата объекта (скаляр), обычно искаженная случайным шумом измерений ξ ; \mathbf{x} – наблюдаемый вектор размерности $M \times 1$, компоненты которого могут представлять собой некоторые функции наблюдаемых величин; \mathbf{c} – подлежащий определению векторный параметр размерности $M \times 1$, T – здесь и далее означает транспонирование.

Если изменение неизвестных параметров объекта за время проведения необходимого для идентификации числа измерений мало, их можно считать постоянными. Тогда после N наблюдений входа и выхода модели вида (1.4) можно записать матричное уравнение:

$$\mathbf{y} = \mathbf{X}\tilde{\mathbf{n}} + \xi, \quad (1.5)$$

где \mathbf{y} – $N \times 1$ -вектор, а \mathbf{X} – $N \times M$ -матрица ($N \geq M$), заданные или полученные в результате измерений, а ξ - неизвестный вектор ошибок. В результате задача идентификации сводится к решению, как правило, переопределенной системы уравнений (1.5). Заметим, что большое число объектов действительно удается описать уравнениями вида (1.4), (1.5). Рассмотрим некоторые, наиболее широко используемые модели.

1.2. Модели идентифицируемых систем

Наиболее простыми, с точки зрения задачи идентификации, являются линейные статические модели [4, 25]. Линейная статическая система, имеющая один выход, может быть описана следующим линейным уравнением:

$$y = \tilde{n}_1 x_1 + \tilde{n}_2 x_2 + \dots + \tilde{n}_M x_M, \quad (1.6)$$

где $x_i, i = \overline{1, M}$ - факторы, определяющие ход некоторого технологического процесса, а $c_i, i = \overline{1, M}$ - подлежащие определению коэффициенты влияния этих факторов на показатель качества y . Проведя N наблюдений факторов и показателя качества системы, получаем совокупность N уравнений вида (1.4), из которых может быть составлена система вида (1.5). Если статический объект имеет несколько показателей качества (выходов), задача идентификации должна решаться для каждой модели, хотя в каждой из этих моделей факторы могут повторяться. Для статических моделей обычно характерно свойство статистической независимости факторов, что существенно упрощает задачу идентификации.

Широкий класс моделей образуют динамические системы. Линейные динамические системы в цифровой системе управления описываются векторными конечно-разностными уравнениями [4, 25]. В частности, для системы порядка n с одним входом - $u(k)$ и одним выходом $-y(k)$ уравнения имеют вид:

$$\mathbf{x}[k+1] = \mathbf{\Phi} \mathbf{x}[k] + \mathbf{d} u[k], \quad (1.7)$$

$$y^*[k] = \mathbf{h}^T \mathbf{x}[k], \sqrt{a^2 + b^2}, \quad (1.8)$$

где \mathbf{d}, \mathbf{h} - $n \times 1$ -векторы, а $\mathbf{\Phi}$ - $n \times n$ матрица (переходных состояний). Для объекта, описываемого системой (1.7),(1.8) можно записать передаточную функцию:

$$W(z) = \frac{Y^*(z)}{U(z)} = \frac{\mathbf{h}^T \cdot \text{Adj}(\mathbf{E}z - \mathbf{\Phi}) \mathbf{d}}{\det(\mathbf{E}z - \mathbf{\Phi})} = \frac{\sum_{i=0}^{n-1} b_{i+1} z^i}{z^n - \sum_{j=0}^{n-1} a_{j+1} z^j},$$

где z - оператор сдвига. Переход от этой передаточной функции во временную область дает следующее выражение:

$$y^*[k+1] = \sum_{i=1}^n a_i y^*[k-i+1] + \sum_{j=1}^n b_j u[k-j+1]. \quad (1.9)$$

Проведя N наблюдений входного и выходного сигналов объекта в соответствии с (1.9), можно записать матричное уравнение вида (1.5), где

$$\mathbf{y} = \mathbf{y}^* + \boldsymbol{\xi},$$

$$\mathbf{y} = \begin{bmatrix} y[k+1] \\ y[k+2] \\ \vdots \\ y[k+N] \end{bmatrix}, \quad \boldsymbol{\xi} = \begin{bmatrix} \xi[k+1] \\ \xi[k+2] \\ \vdots \\ \xi[k+N] \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} y[k], \dots, y[k-n+1], u[k], \dots, u[k-n+1] \\ y[k+1], \dots, y[k-n+2], u[k+1], \dots, u[k-n+2] \\ \vdots \\ y[k+N-1], \dots, y[k+N-n], u[k+N-1], \dots, u[k+N-n] \end{bmatrix},$$

а $\mathbf{c} = [a_1, \dots, a_n, b_1, \dots, b_n]^T$ - вектор искомых параметров объекта. Для определенности будем полагать, что $2n=M$, так что матрица \mathbf{X} по-прежнему имеет размерность $N \times M$, а компоненты вектора ошибок определяются как

$$\xi[k+j] = \varepsilon[k+j] - \sum_{i=1}^n a_i \varepsilon[k-i+j], \quad (1.10)$$

где $\varepsilon[k+j]$ - аддитивные ошибки измерения, порядка модели и др.

Задача идентификации управляемого объекта в данном случае состоит в построении оценки $\hat{\mathbf{c}}$ векторного параметра \mathbf{c} уравнения (1.5) по доступным для непосредственного наблюдения $N \times M$ - матрице \mathbf{X} и $N \times 1$ - вектору \mathbf{y} ($N > M$), при неизвестном $N \times 1$ -векторе ошибок $\boldsymbol{\xi}$. Общая схема формирования данных для идентификации управляемого динамического объекта приведена на рисунке 1.1.

Если ставится задача оценить все элементы матрицы состояния Φ , то уравнения типа (1.5) должны составляться для всех строк матричного уравнения состояний (1.7), т.е. задача идентификации должна решаться n раз.

Для формирования вектора y и матрицы X при этом необходимо, чтобы управляющие сигналы и состояния системы были доступны для непосредственного наблюдения.

Нелинейная динамическая система в общем случае описывается нелинейными уравнениями состояния:

$$\dot{x} = f(x, u). \quad (1.11)$$

Используя обычную процедуру линеаризации относительно некоторого рабочего состояния, уравнение (1.11) можно представить в виде

$$\delta \dot{x} = A \delta x + B \delta u, \quad (1.12)$$

где $A = \nabla_x f(x, u)$, $B = \nabla_u f(x, u)$ – матрицы с элементами

$$a_{i,j} = \frac{\partial^2 f_i}{\partial x_j}, \quad b_{i,j} = \frac{\partial^2 f_i}{\partial u_j} \text{ соответственно.}$$

Соотношение (1.12) по существу является уравнением состояния в (1.7). От него можно перейти к разностному уравнению (1.9), которому *после проведения измерений* соответствует модель вида (1.4). При осуществлении идентификации матриц A и B или коэффициентов соответствующего разностного уравнения в данном случае необходимо помнить, что сигналы, из которых формируется матрица X , представляют собой отклонения от некоторого рабочего режима, в то время как сам этот режим может быть известен не точно. Кроме того, на идентифицируемый объект могут действовать неконтролируемые возмущения.

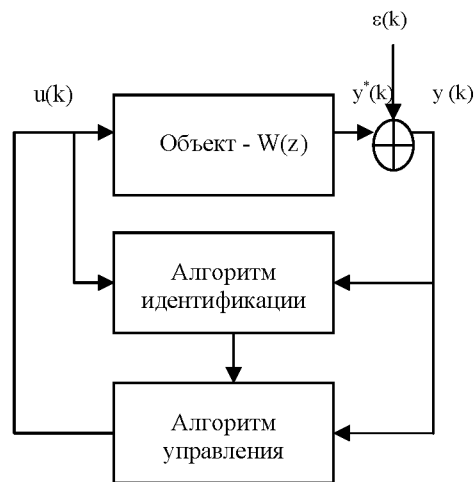


Рис. 1.1. Общая схема системы

Для учета этих факторов наряду с (1.4) для идентификации вводят в рассмотрение расширенную модель

$$y(t) = \tilde{\mathbf{n}}^T \mathbf{x}(t) + f(t) + \xi, \quad (1.13)$$

где $f(t)$ - некоторая, в общем случае неизвестная, составляющая наблюдаемого выходного процесса $y(t)$. По существу ставится задача одновременно-го оценивания параметров системы и неизвестного сигнала. Задача восстановления сигналов в рамках расширенной модели обычно также сводится к оценке некоторого дополнительного числа параметров.

Задача восстановления сигнала может иметь также и самостоятельное значение. Она ставится следующим образом [23]. Пусть сигнал является функцией некоторого аргумента, например, времени t :

$$f(\tilde{n}, t) = y(t) = f(c_1, \dots, c_M, t) = f(\mathbf{c}, t). \quad (1.14)$$

Задача состоит в том, чтобы по принятой последовательности (вектору $\mathbf{Y} = [y_1, y_2, \dots, y_N]^T$) определить вектор параметров $\mathbf{c} = [c_1, \dots, c_M]^T$.

Наиболее широко в задачах восстановления используются линейные зависимости сигнала от искомым параметров. Часто функциональную зависимость общего вида (1.14) специально представляют в виде, допускающем преобразование ее к линейной модели, например, экспоненциальными зависимостями. При этом преобразование к линейной относительно искомым параметров модели осуществляется путем логарифмирования.

В качестве зависимостей (1.14) широко используются также ортогональные представления сигналов [26]:

$$y(t) = \sum_{k=1}^M c_k \varphi_k(t), \quad (1.15)$$

где $\varphi_k(t)$ – заданные ортогональные или ортонормированные базисные функции, а c_k – искомые коэффициенты. Нетрудно заметить, что эти модели также линейные по искомым параметрам. Проведя измерения сигнала $y(t)$ для N фиксированных моментов времени t_i , $i = \overline{1, N}$, можно составить уравнение (1.5). При этом каждый столбец $N \times M$ - матрицы \mathbf{X} будет соответ-

ствовать некоторой базисной функции, а строка – моменту времени, в который проведено наблюдение.

В работе [23] описан алгоритм идентификации, в котором параметры динамической модели объекта и действующих на него возмущений оцениваются попеременно, т.е. на каждом этапе решается отдельно задача идентификации линейной динамической модели и неизвестного сигнала.

1.3. Проблема оценивания по малому числу наблюдений

Проблема: данные \rightarrow система (модель, объясняющая данные) является основной задачей почти для любой отрасли науки. Классические идеи теории систем, касающиеся этой проблемы выражаются следующим образом [7]:

- Принцип единственности. Если данные точные и полные, то существует одна и только одна минимальная система (модель), воспроизводящая эти данные. В теории управления известны строгие доказательства существования и единственности минимальной системы.
- Принцип неопределенности. Неточным (недостовверным) данным соответствует неединственная (недостовверная) система.

Наиболее широко используемый подход к обработке недостовверных данных опирается на следующие предположения:

- Вся неопределенность порождается фиксированным абстрактным вероятностным механизмом (обычно достаточно простого вида).
- Все данные получены путем независимого выбора из фиксированной генеральной совокупности, сформированной посредством этого вероятностного механизма. Другими словами, данные являются конечной, независимой выборкой из генеральной совокупности с фиксированным вероятностным законом [7].

К сожалению, указанная «стандартная статистическая априорная гипотеза» часто оказывается ложной в ситуации, когда она используется. Более того, например, метод наименьших квадратов (МНК) «попадает в эту западню даже без дополнительных (и также априорных) статистических предположений» [7]. В частности, МНК дает единственный ответ в задаче определения модели по зашумленным данным. В то же время это противоречит принципу неопределенности. Поэтому в общем случае метод наименьших квадратов не может рассматриваться как надежный метод идентификации в условиях шумов, поскольку опирается на жесткую априорную гипотезу, связанную с предположением о наличии переменных, свободных от шума [7]. Попытки статистического обоснования МНК не решают

проблему, поскольку при этом априорные гипотезы становятся еще более жесткими.

Еще менее надежным статистическое обоснование метода наименьших квадратов оказывается в случае малого числа наблюдений. Обычно, когда ставится задача построить оценку $\hat{\mathbf{c}}$ вектора параметров \mathbf{c} по $N \times M$ - матрице \mathbf{X} и $N \times 1$ - вектору \mathbf{y} ($N > M$), связанным равенством (1.5), используют следующие предположения [5]: векторы $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_N]^T$ и $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ - случайные; матрица \mathbf{X} детерминирована, то есть ее элементы не являются случайными величинами; $Rank(\mathbf{X})=M$; математическое ожидание вектора $\boldsymbol{\xi}$ равно нулю, то есть

$$M\{\xi_i\}=0, i=1, N \text{ или } M\{\boldsymbol{\xi}\}=0; \text{ для любых}$$

$$i \neq j \quad M\{\xi_i \cdot \xi_j\} = 0, \quad M\{\xi_i^2\} = \sigma^2$$

для всех $i=1, N$. Другими словами,

$$\text{cov}\{\boldsymbol{\xi}\} = \sigma^2 \mathbf{E}_N,$$

где σ^2 - дисперсия отклонений, $\text{cov}\{\cdot\}$ - $N \times N$ -матрица ковариаций отклонений, а \mathbf{E}_N - единичная $N \times N$ -матрица. Если эти предположения выполняются, то оценка метода наименьших квадратов (МНК)

$$\hat{\mathbf{c}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} \quad (1.16)$$

является несмещенной и эффективной [30].

К сожалению, указанные предположения при малом числе наблюдений не отвечают реальному содержанию задачи и оказываются, в лучшем случае, бесполезными. Действительно, даже если они справедливы для шумовой последовательности на всем интервале наблюдения сигнала, оценки математического ожидания и дисперсии на малых фрагментах этого сигнала могут сильно отличаться от декларированных в них свойств. Кроме того, при идентификации моделей систем по малому числу наблюдений всегда сохраняется опасность попадания на неинформативный участок сигнала, характерный для установившихся процессов в системе. При этом условие $Rank(\mathbf{X})=M$ мо-

жет нарушаться, даже если в целом для сигналов, наблюдаемых на большом отрезке, это свойство имеет место.

Таким образом, при оценивании по малому числу наблюдений, во-первых, всегда существует опасность попадания на такой участок сигнала, где задача оценивания оказывается плохо обусловленной или даже вырожденной, во-вторых, основное условие предельных теорем теории вероятностей: существование большого числа наблюдений не выполняются и, как следствие, невозможно обосновать априорную вероятностную модель ошибок в исходных данных.

Даже если существует устойчивое распределение ошибок на одной длинной реализации или множестве реализаций сигнала данного класса, оценки оказываются ненадежными вследствие того, что свойство устойчивости статистических характеристик шумов на малых фрагментах сигнала не проявляется в полной мере. Поэтому мы вынуждены отказаться от предположений классической регрессии. Сформулируем предположения, которые далее будут использоваться при решении задачи оценивания по малому числу наблюдений:

Предположение 1. Матрица X и вектор y фиксированы, то есть

$$x_{i,j}, y_i, \quad i = \overline{1, N}, \quad j = \overline{1, M}$$

известны в результате измерений на одной реализации.

Предположение 2. Число наблюдений мало так, что имеет место неопределенность свойств обусловленности матрицы X и статистических характеристик вектора ξ .

Предположение 3. Относительно вектора ошибок

$$\xi = [\xi_1, \xi_2, \dots, \xi_N]^T$$

известно лишь то, что он ограничен по норме:

$$\|\xi\| \leq R.$$

Предположение 4. В малом наборе данных, несмотря на возможные ошибки, все же содержится достаточное число наблюдений (подсистема, наиболее свободная от шума), по которым оценки \hat{c} могут быть вычислены с требуемой точностью.

Предположение 5. Существует соответствующая уравнению (1.5) точная модель:

$$y^* = Xc, \quad (1.17)$$

где $y^* = y - \xi$.

Существенно, что мы отказываемся от наиболее важных с теоретической точки зрения предположений классической регрессии:

$$M\{\xi\}=0 \text{ и } \text{cov}\{\xi\} = \sigma^2 E_N.$$

Отказ от этих предположений является вынужденным, а предположение 3 – $\|\xi\| \leq R$ более реалистичным. При сделанных предположениях, вся неопределенность связана с произвольной ориентацией вектора ξ в N -мерном пространстве. Заметим, что предположение 3 широко используется в теории возмущений [2] и в теории некорректных задач [20].

Что касается предположения 4, к сожалению, мы не можем построить точную систему (1.17) из (1.5), т.к. не известен вектор ошибок ξ . Однако опираясь на это предположение, можно ставить задачу отыскания подсистемы наиболее свободной от шума [7]. В качестве опознавателя такой подсистемы, в соответствии с предположением 5, может выступать согласованность (взаимная близость) множества оценок, полученных на подсистемах еще более малой размерности. Подробно процедуры идентификации, построенные на основе этого принципа, будут рассмотрены в разделе 3.3.

В дополнение к указанным предположениям могут выдвигаться требования об ограничениях параметров модели в виде неравенств. Эти ограничения задаются на основе априорной информации о допустимых диапазонах оцениваемых характеристик. Такие ограничения часто используются в традиционной постановке задачи оценивания параметров линейной регрессии [4, 30] и в данном случае не являются принципиальными.

С учетом сказанного термин «оценивание» мы будем употреблять, не связывая его во всех случаях с прилагательным «статистическое». Наряду с ним, там, где необходимо подчеркнуть нестатистический характер задачи, мы будем вместо термина «оценивание» использовать также термин «определение». Кроме того, мы будем избегать употребления широко используе-

мого в теории статистического оценивания термина «выборка», применяя вместо него термины «набор данных» или «малое число наблюдений».

В связи с последним замечанием нуждается в уточнении само понятие «малое число наблюдений». Попытки указать конкретное число наблюдений, которое может считаться малым, бесплодны. Например, выборка 500 наблюдений может быть весьма представительной при оценивании одного параметра (сдвига), но это очень мало, если решается задача оценивания параметров разделяющей гиперплоскости в пространстве 400 признаков. Для того, чтобы формально определить малое число наблюдений, нам понадобятся некоторые дополнительные сведения о задаче наименьших квадратов.

Введем в рассмотрение невязку

$$\hat{\xi} = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}} = \mathbf{y} - \hat{\mathbf{y}}. \quad (1.18)$$

Из (1.18) видно, что вектор $\hat{\mathbf{y}}$ принадлежит пространству столбцов матрицы \mathbf{X} :

$$\hat{\mathbf{y}} \in R(\mathbf{X}).$$

Можно показать также [8], что вектор \mathbf{y} декомпозируется на две ортогональные компоненты:

$$\mathbf{y} = \mathbf{X}\hat{\mathbf{c}} + \hat{\xi}, \quad \hat{\xi} \perp \mathbf{X}\hat{\mathbf{c}}.$$

Для более детального обсуждения свойств МНК осуществим стандартное ортогональное разложение.

Известно [8,31], что для $N \times M$ - матрицы \mathbf{X} ранга $k \leq M$ существуют ортогональные $N \times N$ - матрица \mathbf{T} и $k \times k$ - матрица \mathbf{F} такие, что

$$\mathbf{T}^T \mathbf{X} \mathbf{F} = \mathbf{S}, \quad \mathbf{X} = \mathbf{T} \mathbf{S} \mathbf{F}^T, \quad (1.19)$$

где \mathbf{S} - диагональная $N \times M$ - матрица, составленная из неотрицательных чисел, ровно k среди которых строго положительны. Диагональные элементы s_i , $i=1, k$ матрицы \mathbf{S} называются сингулярными числами матрицы \mathbf{X} . Из (1.19), в частности, следует, что

$$\mathbf{X}^T \mathbf{X} = \mathbf{F} \mathbf{S}^T \mathbf{S} \mathbf{F}^T = \mathbf{F} \mathbf{\Lambda} \mathbf{F}^T, \quad (1.20)$$

$$\mathbf{X}\mathbf{X}^T = \mathbf{T}\mathbf{S}\mathbf{S}^T\mathbf{T}^T = \mathbf{T} \begin{bmatrix} \mathbf{\Lambda} & 0 \\ 0 & 0 \end{bmatrix} \mathbf{T}^T, \quad (1.21)$$

где

$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_k)$ - диагональная матрица.

Числа λ_i являются квадратами соответствующих сингулярных чисел s_i , $i=1, k$ и называются характеристическими числами или собственными значениями. Подчеркнем, что собственные значения матрицы $\mathbf{X}^T\mathbf{X}$ и ненулевые собственные значения матрицы $\mathbf{X}\mathbf{X}^T$ совпадают. Векторы-столбцы матриц \mathbf{F} и \mathbf{T} являются соответствующими собственными векторами.

Предположим, что $\text{Rank}\mathbf{X}=M$ и пусть первые M столбцов

$$\mathbf{t}_{\lambda_i}, \quad i = \overline{1, M}$$

матрицы \mathbf{T} соответствуют ненулевым, а последние $N-M$ столбцов

$\mathbf{t}_{0_i}, \quad i = \overline{M+1, N}$ - нулевым собственным значениям матрицы $\mathbf{X}\mathbf{X}^T$, то есть матрица \mathbf{T} представляется в виде блочной:

$$\mathbf{T} = [\mathbf{T}_\lambda; \mathbf{T}_0]. \quad (1.22)$$

Заметим, что с учетом (1.10) – (1.21) и известных свойств

$$\mathbf{F}^{-1} = \mathbf{F}^T, \quad [\mathbf{F}^T]^T = \mathbf{F}, \quad \mathbf{S}^T\mathbf{S} = \mathbf{\Lambda}^2$$

для $N \times M$ - матрицы \mathbf{T}_λ в (1.22) справедливо представление:

$$\mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\mathbf{\Lambda}^{-1/2}. \quad (1.23)$$

Теперь подпространства, связанные с матрицей \mathbf{X} , мы можем ввести следующим образом

$$R(\mathbf{X}) = \text{span}[\mathbf{t}_{\lambda_1}, \dots, \mathbf{t}_{\lambda_M}],$$

$$N(\mathbf{X}^T) = \text{span}[\mathbf{t}_{0,1}, \dots, \mathbf{t}_{0,N-M}].$$

$R(\mathbf{X})$ называют пространством столбцов (или столбцовым пространством) матрицы \mathbf{X} , а $N(\mathbf{X}^T)$ – нуль-пространством матрицы \mathbf{X}^T . Далее, для краткости, мы будем их называть просто пространством и нуль-пространством матрицы \mathbf{X} соответственно.

Нуль-пространство $N(\mathbf{X}^T)$ является ортогональным дополнением для пространства $R(\mathbf{X})$. Поскольку в соответствии с (1.18)

$$\hat{\mathbf{y}} \in R(\mathbf{X}), \text{ а } \hat{\xi} \perp \mathbf{X}\hat{\mathbf{c}} = \hat{\mathbf{y}},$$

ясно, что вектор невязок (1.18) принадлежит нуль-пространству $\hat{\xi} \in N(\mathbf{X}^T)$. На рис. 1.2 приведена геометрическая иллюстрация МНК для случая $M=2, N=3$.

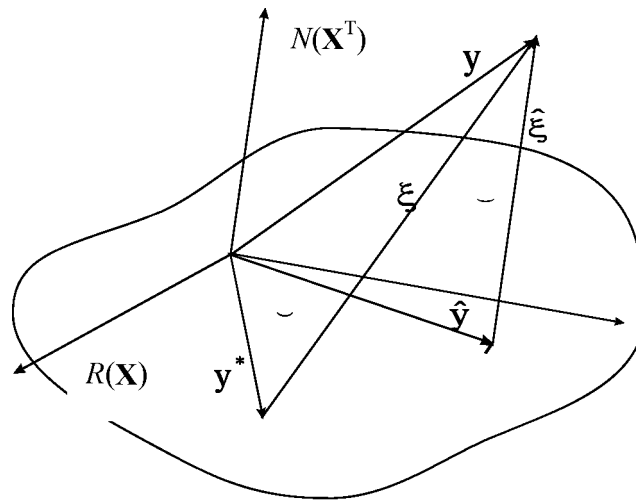


Рис. 1.2. Геометрическая интерпретация МНК

Теперь мы можем сформулировать признак малого числа наблюдений в терминах подпространств: число наблюдений будем считать малым, если размерности подпространств $R(\mathbf{X})$ и $N(\mathbf{X}^T)$ одного порядка.

В заключение еще раз подчеркнем, что результат оценивания существенно зависит от конкретной реализации (фрагмента сигнала). Поэтому при малом числе наблюдений, во-первых, необходимым этапом

является предварительная оценка информативности полученного набора данных. Во-вторых, поскольку используемые предположения отличаются от традиционно принятых в теории статистических решений, процедуры оценивания также должны строиться на основе других (нестатистических) подходов. На эти два основных аспекта и обращено настоящее учебное пособие.

Автор надеется дать читателям новые знания в области решения задач идентификации по малому числу наблюдений. Для рассмотрения возникающих при этом специфических проблем вначале полезно рассмотреть соотношения, показывающие, как формируются ошибки оценивания на конкретном наборе данных.

1.4. Формы представления и ортогональные разложения ошибок

Для установления связи ошибок оценивания с ошибками измерений вначале рассмотрим, как ошибки измерений входят в исходные соотношения, используемые при решении задачи идентификации.

При использовании метода наименьших квадратов (1.16) искомые оценки являются решением системы нормальных уравнений

$$\mathbf{A}\hat{\mathbf{c}} = \mathbf{b} . \quad (1.24)$$

Здесь и далее используются обозначения

$$\mathbf{A} = \mathbf{X}^T \mathbf{X}, \quad \mathbf{b} = \mathbf{X}^T \mathbf{y} ,$$

где \mathbf{X} – $N \times M$ -матрица, а \mathbf{y} – $N \times 1$ -вектор соответственно, фигурирующие в (10.6). Будем использовать также обозначения \mathbf{X}^* и \mathbf{y}^* для незашумленных матрицы и вектора, т.к. в соответствии с предположением 5 соответствующая уравнению (1.5) точная модель (1.17) идентифицируемой системы (15) существует. В действительности матрица \mathbf{X} и вектор \mathbf{y} фиксируются с ошибками. В таблице 1.1 приведены соотношения, показывающие структуру матрицы \mathbf{X} и векторов \mathbf{y} и $\boldsymbol{\xi}$ для статических и динамических моделей общего вида.

Из геометрических представлений (рис. 1.2) ясно, что вектор ошибок $\boldsymbol{\xi}$ оказывает непосредственное влияние на точность МНК-оценок. Искажение шумами матрицы \mathbf{X} и вектора \mathbf{y} напрямую не влияет на точность оценивания.

Это влияние опосредованное (через изменение обусловленности задачи из-за ошибок измерений). Рассмотрим этот вопрос подробнее.

Таблица 1.1. Формирование измерений и ошибок

| Матрицы и векторы | Статический объект | Динамический объект |
|--------------------|------------------------------------|--|
| \mathbf{X} | \mathbf{X}^* | $\mathbf{X}^* + \delta\mathbf{X}$ |
| \mathbf{y} | $\mathbf{y}^* + \delta\mathbf{y}$ | $\mathbf{y}^* + \delta\mathbf{y}$ |
| ξ | $\delta\mathbf{y}$ | $\delta\mathbf{y} - \delta\mathbf{X}\mathbf{c}$ |
| $\delta\mathbf{b}$ | $\mathbf{X}^{*T} \delta\mathbf{y}$ | $\mathbf{X}^{*T} \delta\mathbf{y} + \delta\mathbf{X}^T \mathbf{y}^* + \delta\mathbf{X}^T \delta\mathbf{y}$ |
| $\delta\mathbf{A}$ | 0 | $\mathbf{X}^{*T} \delta\mathbf{X} + \delta\mathbf{X}^T \mathbf{X}^* + \delta\mathbf{X}^T \delta\mathbf{X}$ |

Поскольку в общем случае матрица \mathbf{X} и вектор \mathbf{y} содержат ошибки $\delta\mathbf{X}$ и $\delta\mathbf{y}$, матрица \mathbf{A} и вектор \mathbf{b} также будут содержать возмущения, которые обозначим $\delta\mathbf{A}$ и $\delta\mathbf{b}$ соответственно:

$$\mathbf{A} = \mathbf{A}^* + \delta\mathbf{A}, \quad (1.25)$$

$$\mathbf{b} = \mathbf{b}^* + \delta\mathbf{b}, \quad (1.26)$$

где с учетом (10.14)

$$\mathbf{A}^* = \mathbf{X}^{*T} \mathbf{X}^*,$$

$$\mathbf{b}^* = \mathbf{X}^{*T} \mathbf{y}^*.$$

Выражения, показывающие структуру возмущений $\delta\mathbf{A}$ и $\delta\mathbf{b}$, приведены в последних двух строках таблицы 1.1.

Характеристики обусловленности матриц \mathbf{A} и \mathbf{A}^* могут существенно различаться. Например, если матрица \mathbf{A}^* вырождена, соответствующая ей матрица $\mathbf{A} = \mathbf{A}^* + \delta\mathbf{A}$ может оказаться хорошо обусловленной. Поэтому способ формирования матрицы \mathbf{A} и тип модели (статического или динамического) объекта имеют большое значение.

Построим теперь соотношения для ошибок МНК - оценок. Для этого в соответствии с (10.14) запишем систему уравнений:

$$\mathbf{A}^* \mathbf{c} = \mathbf{b}^* .$$

С учетом (1.25) и (1.26) это равенство можно переписать в виде

$$\mathbf{A} \mathbf{c} = \mathbf{b} - \boldsymbol{\zeta} , \quad (1.27)$$

где

$$\boldsymbol{\zeta} = \delta \mathbf{b} - \delta \mathbf{A} \mathbf{c} . \quad (1.28)$$

Сравнивая (1.27) и (1.24), можно записать выражение для ошибки оценивания $\Delta \mathbf{c} = \hat{\mathbf{c}} - \mathbf{c}$:

$$\Delta \mathbf{c} = \mathbf{A}^{-1} \boldsymbol{\zeta} . \quad (1.29)$$

Можно показать, что векторы ошибок $\boldsymbol{\zeta}$ и $\boldsymbol{\xi}$ связаны соотношением

$$\boldsymbol{\zeta} = \mathbf{X}^T \boldsymbol{\xi} . \quad (1.30)$$

Отсюда, в частности, следует, что

$$\boldsymbol{\zeta}^T \boldsymbol{\zeta} = \|\boldsymbol{\zeta}\|_2^2 = \|\mathbf{X}^T \boldsymbol{\xi}\|_2^2 = \boldsymbol{\xi}^T \mathbf{X} \mathbf{X}^T \boldsymbol{\xi} .$$

Равенство (1.29) можно также получить, произведя в (1.28) замену $\delta \mathbf{b}$, $\delta \mathbf{A}$ их выражениями из таблицы 1.1 или умножив обе части уравнения (1.5) слева на \mathbf{X}^T и затем осуществив подстановку

$$\boldsymbol{\zeta} = \mathbf{b} - \mathbf{A} \mathbf{c}$$

из (1.27).

Если векторы-строки матрицы \mathbf{X}^T нормированы так, что $\|\mathbf{x}_i\|_2 = 1$ для всех $i = \overline{1, M}$, имеет место следующая геометрическая трактовка: компоненты вектора $\boldsymbol{\zeta}$ суть проекции вектора ошибок $\boldsymbol{\xi}$ на направления, задаваемые векторами независимых переменных \mathbf{x}_i , $i = \overline{1, M}$.

С учетом (1.29) и (1.30) ошибку оценивания

$$\Delta \mathbf{c} = \hat{\mathbf{c}} - \mathbf{c}$$

можно выразить как результат преобразования вектора ξ :

$$\Delta \mathbf{c} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \xi = \mathbf{A}^{-1} \mathbf{X}^T \xi. \quad (1.31)$$

Подчеркнем, что матричный коэффициент

$$[\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X} = \mathbf{A}^{-1} \mathbf{X}^T$$

в (1.16), (1.31) при решении задачи оценивания на конкретной реализации всегда известен. Это имеет место как в случае статической модели, где $\mathbf{X} = \mathbf{X}^*$, так и в случае динамической модели, для которой

$$\mathbf{X} = \mathbf{X}^* + \delta \mathbf{X}$$

(см. таблицу 10.2). Поэтому при анализе точности идентификации матрица \mathbf{A} и вектор \mathbf{b} считаются заданными точно, а значение имеет лишь способ формирования ошибок (ζ и/или ξ).

Более ясное геометрическое представление формирования ошибок идентификации дает сингулярное разложение (1.19) – (1.21) задачи наименьших квадратов [8]. Как и ранее здесь, предполагаем, что $Rank \mathbf{X} = M$. Умножим обе части (1.29) слева на \mathbf{F}^T , где \mathbf{F} указанная выше ортогональная матрица:

$$\mathbf{F}^T \mathbf{A} \mathbf{F}^T = \mathbf{A}.$$

Тогда с учетом свойств

$$\mathbf{F} \mathbf{F}^T = \mathbf{E}, \quad \mathbf{F}^T \mathbf{A}^{-1} \mathbf{F} = \mathbf{A}^{-1}$$

можно записать

$$\mathbf{F}^T \Delta \mathbf{c} = \mathbf{A}^{-1} \mathbf{F}^T \xi. \quad (1.32)$$

Равенство (1.32) допускает простую геометрическую трактовку: проекции вектора ошибок оценивания на базис, образованный собственными векторами матрицы \mathbf{A} , образуются из проекций вектора ошибок исходных данных (1.28) на тот же базис с «коэффициентами усиления» обратно пропорциональными соответствующим собственным значениям.

Аналогичное соотношение можно построить для вектора ошибок идентификации, представленного в виде соотношения (1.31). Для этого с учетом

соотношения связи между векторами ошибок ζ и ξ (1.30) перепишем (1.32) в виде

$$\mathbf{F}^T \Delta \mathbf{c} = \Lambda^{-1} \mathbf{F}^T \mathbf{X}^T \zeta = \Lambda^{-1/2} \left[\Lambda^{-1/2} \mathbf{F}^T \mathbf{X}^T \right] \zeta.$$

Подставляя в это равенство вместо матрицы

$$\Lambda^{-1/2} \mathbf{F}^T \mathbf{X}^T$$

в соответствии с (1.23) матрицу \mathbf{T}_λ^T , получаем следующее представление

$$\mathbf{F}^T \Delta \mathbf{c} = \Lambda^{-1/2} \mathbf{T}_\lambda^T \zeta. \quad (1.33)$$

В данном случае имеет место весьма сходная геометрическая трактовка: проекции вектора ошибок идентификации на направления базиса, образованного собственными векторами матрицы

$$\mathbf{A} = \mathbf{X}^T \mathbf{X},$$

образуются из проекций вектора ошибок исходных данных на направления базиса, образованного собственными векторами, соответствующими ненулевым собственным значениям матрицы $\mathbf{B} = \mathbf{X}\mathbf{X}^T$ с «коэффициентами усиления»

$$\lambda_i^{-1/2}, \quad i = \overline{1, M}.$$

Заметим, что из (1.33) может быть получено соотношение (1.32) путем замен

$$\zeta = \mathbf{X}^T \xi \quad \text{и} \quad \mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\Lambda^{-1/2}$$

в соответствии с (1.30) и (1.23) соответственно.

Соотношения (1.32), (1.33) представляются важными для понимания закономерностей формирования ошибок идентификации. Для сравнительного анализа точности удобнее использовать скалярные характеристики векторов ошибок. С использованием разложений (1.32), (1.33) за-

пишем выражения для квадрата евклидовой нормы вектора ошибок оценивания.

Подвергнув операции транспонирования обе части равенства (1.32), перепишем его в виде

$$\Delta \mathbf{c}^T \mathbf{F} = \boldsymbol{\zeta}^T \mathbf{F} \boldsymbol{\Lambda}^{-1}.$$

Далее умножим слева обе части последнего равенства на соответствующие (левую и правую) части (1.32). С учетом свойств матрицы \mathbf{F} получим

$$\Delta \mathbf{c}^T \Delta \mathbf{c} = \|\Delta \mathbf{c}\|_2^2 = \boldsymbol{\zeta}^T \mathbf{F} \boldsymbol{\Lambda}^{-2} \mathbf{F}^T \boldsymbol{\zeta} = \|\boldsymbol{\zeta}\|_2^2 \sum_{i=1}^M \lambda_i^{-2} \cos^2 \varphi_i, \quad (1.34)$$

где $\cos \varphi_i$ – косинус угла $\varphi_i = (\boldsymbol{\zeta}, \hat{\mathbf{f}}_i)$ между вектором ошибок $\boldsymbol{\zeta}$ и направлением собственного вектора $\hat{\mathbf{f}}_i$, соответствующего i -му собственному значению λ_i матрицы $\mathbf{X}^T \mathbf{X}$.

Аналогичное соотношение имеет место для случая представления ошибок идентификации в виде (1.33). Подвергнув операции транспонирования обе части (1.33), запишем равенство

$$\Delta \mathbf{c}^T \mathbf{F} = \boldsymbol{\zeta}^T \mathbf{T}_\lambda \boldsymbol{\Lambda}^{-1/2}.$$

Далее умножим обе части этого равенства на соответствующие части равенства (1.33). Аналогично предыдущему получаем

$$\Delta \mathbf{c}^T \Delta \mathbf{c} = \|\Delta \mathbf{c}\|_2^2 = \boldsymbol{\zeta}^T \mathbf{T}_\lambda \boldsymbol{\Lambda}^{-1} \mathbf{T}_\lambda^T \boldsymbol{\zeta} = \|\boldsymbol{\zeta}\|_2^2 \sum_{i=1}^M \lambda_i^{-1} \cos^2 \psi_i, \quad (1.35)$$

где $\cos \psi_i$ – косинус угла $\psi_i = (\boldsymbol{\zeta}, \hat{\mathbf{t}}_{\lambda_i})$ между вектором $\boldsymbol{\zeta}$ и собственным вектором $\hat{\mathbf{t}}_{\lambda_i}$, (i -м столбцом определяемой соотношением (1.23) матрицы \mathbf{T}_λ).

Соотношения (1.34) и (1.35) имеют ясную геометрическую трактовку. Из правых частей этих равенств видно, что норма вектора ошибок идентификации зависит не только от норм векторов $\boldsymbol{\zeta}$ и $\boldsymbol{\xi}$, но также и от их ориентации относительно соответствующих подпространств. Заметим, что каноническое представление (1.35) во многих случаях оказывается более полезным, т.к. в

нем фигурирует вектор ошибок ξ , которому может быть поставлен в соответствие вектор невязок (1.18).

1.5. Оценки достижимой точности

Получив каким-либо способом оценку, обычно пытаются прогнозировать ее точность и надежность. В теории статистических решений для этого используются интервальные оценки, опирающиеся на априорные данные о распределениях. В рамках принятых предположений эти оценки не могут быть построены. В данном случае подходящим является аппарат вычислительной алгебры, где для прогнозирования точности используются характеристики обусловленности матрицы X и/или A . В настоящем разделе этот подход развивается применительно к представлениям ошибок, приведенным в предыдущем разделе.

Соотношения (1.34), (1.35) вскрывают закономерности формирования ошибок идентификации, однако они не пригодны для непосредственной оценки достижимой точности. Во-первых, самостоятельной проблемой является необходимость вычисления собственных значений матрицы A . Во-вторых, фигурирующие в (1.34), (1.35) косинусы углов неизвестны, т.к. неизвестна ориентация векторов ошибок относительно пространств, в которых они рассматриваются. В настоящем разделе описываются методики, использование которых для прогнозирования точности идентификации, по крайней мере, не требует решения полной проблемы собственных значений.

Для построения количественных оценок точности идентификации по-прежнему будем опираться на предположение 3: $\|\xi\| \leq R$. В этом неравенстве будем использовать обладающую свойством инвариантности к унитарным преобразованиям евклидову норму. При этом указанное ограничение можно переписать в виде

$$\|\xi\|_2 \leq R_\xi. \quad (1.36)$$

Заметим, что в соответствии с выражениями для ошибок, приведенными в таблице 1.1, неравенство (1.36) является следствием того, что в силу задания фиксированной матрицы X заданы также границы для согласованных евклидовых норм возмущений δu и δX . Множество Ξ векторов, удовлетворяющих неравенству (1.36), принадлежит «шару»:

$$\xi \in \Xi, \quad \Xi = \left\{ \xi : (\xi^T \xi)^{1/2} = \|\xi\|_2 \leq R_\xi = Const \right\}. \quad (1.37)$$

Ориентация вектора ошибок ξ относительно пространства столбцов матрицы \mathbf{X} случайна, и не делается никаких предположений относительно вероятности различных направлений.

Как указывалось выше, процедура идентификации часто сводится к решению нормальной системы уравнений, содержащей ошибку ζ (1.28). Из неравенства (1.36) в силу равенства (1.30), устанавливающего связь между различными представлениями ошибок, следует также, что

$$\|\zeta\|_2 \leq R_\zeta, \quad (1.38)$$

где R_ζ - некоторая фиксированная константа. Установим связь величин R_ζ и R_ξ .

Применим к соотношению связи между ошибками (1.30) стандартные преобразования:

$$\mathbf{F}^T \zeta = \mathbf{F}^T \mathbf{X}^T \xi = \mathbf{\Lambda}^{1/2} \left[\mathbf{\Lambda}^{-1/2} \mathbf{F}^T \mathbf{X}^T \right] \xi = \mathbf{\Lambda}^{-1/2} \mathbf{T}_\lambda \xi \quad \text{и} \quad \zeta^T \mathbf{F} = \xi^T \mathbf{T}_\lambda^T \mathbf{\Lambda}^{-1/2}.$$

Далее используя ту же, что и ранее, схему по аналогии с (1.34) и (1.35), получаем

$$\|\zeta\|_2^2 = \|\xi\|_2^2 \sum_{i=1}^M \lambda_i \cos^2 \psi_i. \quad (1.39)$$

Радиус «шара» для вектора ζ должен быть равен минимальной из величин полуосей эллипсоида (1.39) при максимальном значении $\|\xi\|_2^2$. В соответствии с предположением (1.36) максимальное значение квадрата нормы вектора ξ задается равенством

$$\|\xi\|_2^2 = R_\xi^2.$$

Поскольку матрица \mathbf{X} фиксирована, числа $\lambda_i, i = \overline{1, M}$ являются константами. С другой стороны, при экстремальных значениях величины $\|\Delta c\|_2^2$

квадраты косинусов могут принимать значения 0 или 1. Обозначив R_ζ^2 минимально возможное значение суммы в правой части (1.39) при $\|\xi\|_2^2 = R_\xi^2$, можно записать:

$$\|\zeta\|_2^2 \leq R_\xi^2 \lambda_{\min} = R_\zeta^2.$$

Отсюда, в частности, следует, что

$$R_\zeta = \lambda_{\min}^{1/2} R_\xi. \quad (1.40)$$

Построим теперь оценки для скалярной характеристики ошибок оценивания:

$$\Delta \hat{\mathbf{n}}^T \Delta \hat{\mathbf{n}} = \|\Delta \hat{\mathbf{n}}\|_2^2$$

для обоих, указанных выше случаев представления ошибок в исходных данных: в виде векторов ξ и/или ζ . Вначале рассмотрим исходное предположение (1.36) для вектора ξ :

$$\|\xi\|_2 \leq R_\xi.$$

Будем полагать, что норма (длина) вектора ошибок ξ не зависит от ориентации этого вектора относительно пространства столбцов матрицы \mathbf{X} . Для решения задачи воспользуемся каноническими разложениями (1.34), (1.35).

Расщепим задачу на два этапа: вначале решим более простую задачу. Установим экстремальные (максимальную и минимальную) ошибки идентификации, которые могут возникать при фиксированной норме $\|\xi\|_2^2$ вектора ошибок, то есть при условии, что $\xi \in \Xi_R$, а

$$\Xi_R = \left\{ \left(\xi^T \xi \right)^{1/2} = \|\xi\|_2 = R_\xi = \text{Const} \right\}. \quad (1.41)$$

Множество Ξ_R более узкое по сравнению с Ξ в (1.37). Оно представляет собой подмножество векторов ошибок, концы которых принадлежат сфере, радиус которой R_ξ .

По предположению матрица \mathbf{X} фиксирована и задана, а фигурирующие в канонических разложениях (1.34), (1.35) собственные значения

$$\lambda_i(\mathbf{X}^T\mathbf{X}), \quad i = \overline{1, M}$$

не зависят от направления вектора ξ . Таким образом, задача сводится к отысканию ориентации вектора ξ относительно пространства столбцов матрицы \mathbf{X} , при которой квадрат нормы этого вектора будет достигать экстремальных значений (максимума или минимума).

Рассмотрим равенство (1.35) как функцию переменных ψ_i . При соответствующих предположениях непрерывности множества возможных реализаций векторов ξ на заданной сфере, переменные ψ_i , $i = \overline{1, M}$ также непрерывны в интервале $[0, 2\pi]$. Следовательно, квадрат евклидовой нормы вектора ошибок идентификации, заданный в виде (1.35), является непрерывной функцией M непрерывных аргументов ψ_i .

Для фиксированных матрицы \mathbf{X} ($Rank(\mathbf{X})=M$) и собственных векторов, образующих пространство $R(\mathbf{X})$ необходимые условия экстремума функции (1.35) задаются следующей системой M уравнений

$$\frac{\partial}{\partial \psi_i} \|\Delta \mathbf{c}\|^2 = 2R_\xi^2 \lambda_i^{-1} \cos \psi_i \sin \psi_i = 0. \quad (1.42)$$

Одно из возможных решений этой системы соответствует углам ψ_i , при которых все косинусы обращаются в нуль. Это возможно в случае, когда вектор ошибок принадлежит нуль-пространству матрицы \mathbf{X}^T : $\xi \in N(\mathbf{X})$. Если это не так, то (поскольку $Rank(\mathbf{X})=M$ и, следовательно, $\lambda_i \neq 0$) при $\psi_i \in [0, 2\pi]$ углы могут принимать одно из двух значений:

$$\psi_i = 0, \pm \frac{\pi}{2}.$$

Таким образом, необходимые условия экстремума для величины $\|\Delta \mathbf{c}\|_2^2$ выполняются либо при $\xi \in N(\mathbf{X})$, либо на направлениях, задаваемых собственными векторами, образующими пространство $R(\mathbf{X})$. Другими словами, ес-

ли $\xi \in R(\mathbf{X})$, его направление должно совпадать с одним из (ортогональных друг другу) векторов-столбцов $N \times M$ -матрицы \mathbf{T}_λ (Рис. 1.2).

Теперь из геометрических соображений легко установить максимальное и минимальное значения для квадратов норм вектора ошибок идентификации. Ясно, что при $\xi \in N(\mathbf{X})$ имеет место равенство

$$\|\Delta \mathbf{c}\|_2^2 = 0.$$

При $\xi \in R(\mathbf{X})$ зададим в качестве «подозрительного» на экстремум одно, например i -е, направление ($\psi_i = 0$) ортогонального базиса. Но тогда для всех

$$j = \overline{1, M}, \quad j \neq i$$

$$\psi_j = \pm \pi/2, \quad \text{а } \cos \psi_j = 0.$$

Следовательно, сумма в правой части функции (1.35) будет содержать лишь одно слагаемое:

$$\|\Delta \tilde{\mathbf{n}}\|_2^2 = \|\xi\|_2^2 \lambda_i^{-1},$$

соответствующее выбранному (i -му) направлению, а экстремальные значения $\|\Delta \mathbf{c}\|_2^2$ будут достигаться на направлениях, соответствующих максимальному и минимальному собственным значениям. Таким образом, имеет место утверждение 1. Пусть $\text{Rank}(\mathbf{X})=M$, а множество векторов ошибок $\xi \in \Xi_R$:

$$\Xi_R = \left\{ \xi : (\xi^T \xi)^{1/2} = \|\xi\|_2 = R_\xi = \text{Const} \right\}.$$

Тогда

$$\lambda_{\max}^{-1} R_\xi^2 \leq \|\Delta \mathbf{c}\|_2^2 \leq \lambda_{\min}^{-1} R_\xi^2, \quad (1.43)$$

если $\xi \in R(\mathbf{X})$, и

$$\|\Delta \mathbf{c}\|_2^2 \leq \lambda_{\min}^{-1} R_\xi^2 \quad (1.44)$$

в остальных случаях.

Оценка (1.44) широко известна в литературе. Ее, например, можно получить из выражения для вектора ошибок идентификации (1.31), используя неравенство треугольника [3]. Тем не менее, утверждение 1 и использовавшиеся для его обоснования рассуждения дают более полное представление о структуре и границах ошибок оценивания. В частности, приведенная в (1.43) для случая $\xi \in R(X)$ оценка снизу квадрата евклидовой нормы вектора ошибок идентификации может быть весьма полезной.

Аналогичные оценки могут быть построены также в случае, когда величина квадрата нормы вектора ошибок представлена в виде соотношения (1.34). Действуя по той же, что и выше, схеме из (1.34) с учетом ограничения (1.38), получаем неравенство:

$$\|\Delta \mathbf{c}\|_2^2 \leq \lambda_{\min}^{-2} \|\zeta\|_2^2 \leq \lambda_{\min}^{-2} R_\zeta^2, \quad (1.45)$$

где в соответствии с (1.28)

$$\zeta = \delta \mathbf{b} - \delta \mathbf{A} \cdot \mathbf{c}.$$

В отличие от (1.43) здесь, вместо λ_{\min}^{-1} , в качестве «коэффициента усиления» ошибок исходных данных фигурирует λ_{\min}^{-2} . Это совершенно ясно. Ведь последнее равенство можно было получить простой заменой в (1.43) R_ξ на R_ζ в соответствии с (1.40).

Теперь вернемся к исходному предположению (1.36), (1.37). Ясно, что ошибки идентификации в общем случае, когда векторы ξ и/или ζ принадлежат шару, не могут превышать ошибок, которые возникают в случае, когда они принадлежат соответствующим сферам. Поэтому с учетом (1.45), (1.40) справедливо следующее следствие (1.44).

Следствие утверждения 1.

$$\text{Если } \xi \in \Xi, \quad \Xi = \left\{ \xi : (\xi^T \xi)^{1/2} = \|\xi\|_2 \leq R_\xi = \text{Const} \right\},$$

$$\text{то } \|\Delta \mathbf{c}\|_2^2 \leq \lambda_{\min}^{-1} \|\xi\|_2^2 \leq \lambda_{\min}^{-1} R_\xi^2. \quad (1.46)$$

$$\text{Если } \zeta \in \mathbf{Z}, \quad \mathbf{Z} = \left\{ \zeta : (\zeta^T \zeta)^{1/2} = \|\zeta\|_2 \leq R_\zeta = \text{Const} \right\},$$

$$\text{то } \|\Delta c\|_2^2 \leq \lambda_{\min}^{-2} \|\zeta\|_2^2 \leq \lambda_{\min}^{-2} R_{\zeta}^2, \quad (1.47)$$

где $\lambda_{\max}, \lambda_{\min}$ - максимальное и минимальное среди собственных значений λ_i матрицы $\mathbf{X}^T \mathbf{X}$, а R_{ξ} и R_{ζ} удовлетворяют равенству (1.40).

Из (1.46) следует, что оценка сверху для квадрата нормы вектора ошибок оценивания может быть улучшена, если осуществить преобразование данных, направленное на увеличение минимального собственного значения, при котором норма $\|\xi\|_2$ вектора ошибок в исходных данных, по крайней мере, не увеличивается.

Оценки (1.46) и (1.47) получены в предположении, что переменные ψ_i и φ_i могут принимать любые значения в интервале $[0, 2\pi]$. Это предположение можно ослабить. В частности, можно допустить, что векторы ξ и/или ζ принадлежат некоторому конусу. Из геометрических соображений ясно, что указанные неравенства при этом не нарушатся. Более того, если известны априорные вероятностные характеристики, характеризующие преимущественные направления векторов ошибок измерений, оценки достижимой точности могут быть улучшены. В заключение подчеркнем, что в соответствии с (1.46), (1.47) решающее значение при оценке ошибок идентификации имеют собственные значения матрицы \mathbf{A} .

1.6. Прогнозирование относительных ошибок оценивания с учетом погрешностей округления

Полученные выше оценки сверху и снизу определяют границы достижимой точности при наличии ошибок измерений на конкретной реализации без учета погрешностей округления. Если ошибки округления сравнимы с ошибками измерений, прогноз ошибок оценивания будет недостаточно точен. В теории возмущений известен подход к оценке влияния ошибок округления, так называемый обратный анализ ошибок [2]. В соответствии с этим подходом влияние ошибок при вычислениях равносильно дополнительному внесению ошибок в исходные данные. Обозначим эти дополнительные возмущения матрицы \mathbf{A} и вектора \mathbf{b} соответственно $\Delta \mathbf{A}$ и $\Delta \mathbf{b}$. Тогда соответствующая (1.24) система, содержащая эквивалентные возмущения, может быть представлена в виде

$$[\mathbf{A} + \Delta\mathbf{A}] \cdot \hat{\mathbf{c}}_a = [\mathbf{b} - \zeta] + (\zeta + \Delta\mathbf{b}), \quad (1.48)$$

где $\hat{\mathbf{c}}_a$ - оценка, которая кроме ошибок $\Delta\mathbf{c}$, определяемых равенствами (1.29), (1.31), содержит также погрешности округления, возникающие на этапе вычисления оценок. Для наглядности в правой части (1.48) квадратными скобками выделена правая часть точного (по отношению к искомому вектору \mathbf{c}) уравнения (1.27), а круглыми - суммарное возмущение правой части, связанное с вычислениями и неточностью исходных данных.

Для прогнозирования общей ошибки

$$\Delta\mathbf{c}_a = \hat{\mathbf{c}}_a - \mathbf{c},$$

возникающей вследствие ошибок исходных данных и эквивалентного возмущения, учитывающего погрешность округления, воспользуемся результатами теории возмущений. Известна [2,21] следующая оценка для максимальной относительной погрешности

$$\delta_c = \|\Delta\mathbf{c}_a\| / \|\mathbf{c}\|$$

в решениях:

$$\delta_c \leq \frac{K(\mathbf{A})}{1 - K(\mathbf{A})\delta_A} (\delta_b + \delta_A), \quad (1.49)$$

где $K(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}\|^{-1}$ - условное число или так называемое число обусловленности.

Относительные погрешности исходных данных для наиболее общего случая, когда оцениваются параметры динамической модели, в соответствии с (1.48) определяются как

$$\delta_A = \frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}, \quad \delta_b = \frac{\|\zeta + \Delta\mathbf{b}\|}{\|\mathbf{b} - \zeta\|} = \frac{\|\zeta + \Delta\mathbf{b}\|}{\|\mathbf{b}^* + \delta\mathbf{A} \cdot \mathbf{c}\|},$$

где $\delta\mathbf{A}$ - определяется соотношением, приведенным в таблице 1.1.

В исходных предположениях (1.36), (1.38) используется евклидова норма вектора. Поэтому для числа обусловленности будем использовать согласованную с ней спектральную норму матрицы \mathbf{A} .

При этом

$$K(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \quad (1.50)$$

где $\lambda_{\max}(\mathbf{A})$, $\lambda_{\min}(\mathbf{A})$ - максимальное и минимальное собственные значения матрицы \mathbf{A} . Далее для сокращения записей мы будем обозначать их λ_{\max} , λ_{\min} соответственно.

С учетом сказанного неравенство (1.49) можно записать в виде

$$\delta_c \leq \frac{\lambda_{\max}}{\lambda_{\min} - \lambda_{\max} \delta_A} (\delta_b + \delta_A). \quad (1.51)$$

Для характеристики фигурирующих в (1.49) относительных возмущений также должны использоваться евклидовы нормы:

$$\delta_{\tilde{\mathbf{n}}} = \frac{\|\Delta \tilde{\mathbf{n}}_a\|_2}{\|\tilde{\mathbf{n}}\|_2}, \quad \delta_A = \frac{\|\Delta \mathbf{A}\|_2}{\|\mathbf{A}\|_2}, \quad \delta_b = \frac{\|\zeta + \Delta \mathbf{b}\|_2}{\|\mathbf{b}^* + \delta \mathbf{A} \cdot \mathbf{c}\|_2}. \quad (1.52)$$

Если матрица \mathbf{A} хорошо обусловлена, то при соответствующем выборе разрядной сетки вычислителя

$$\lambda_{\min} \gg \lambda_{\max} \delta_A$$

и условие (1.51) принимает более простой вид:

$$\delta_c \leq \frac{\lambda_{\max}}{\lambda_{\min}} (\delta_b + \delta_A) = K(\mathbf{A}) (\delta_b + \delta_A). \quad (1.53)$$

Часто, например, в случае идентификации статистической модели КИХ-фильтра относительной погрешностью δ_A можно пренебречь, по сравнению с δ_b . Тогда справедлива более простая оценка:

$$\delta_c \leq \frac{\lambda_{\max}}{\lambda_{\min}} \delta_b = K(\mathbf{A}) \delta_b. \quad (1.54)$$

Наконец, если составляющей $\delta \mathbf{A} \mathbf{c}$ в выражении (1.28) также можно пренебречь по сравнению с $\delta \mathbf{b}$ (а тем более по сравнению с \mathbf{b}), существенно упрощается и само выражение для относительных возмущений δ_b :

$$\delta_b = \frac{\|\zeta + \Delta \mathbf{b}\|_2}{\|\mathbf{b}^*\|_2}. \quad (1.55)$$

В соответствии с неравенством треугольника наряду с (1.55) можно использовать также ее оценку сверху

$$\delta_b = \frac{\|\zeta\|_2}{\|\mathbf{b}^*\|_2} + \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}^*\|_2}, \quad (1.56)$$

где в соответствии с (1.29) $\zeta = \mathbf{X}^T \xi = \delta \mathbf{b} - \delta \mathbf{A} \mathbf{c}$.

Полезно указать на некоторые частные случаи приведенных выше оценок. Если все собственные значения одинаковы, то есть

$$\lambda_{\max}^{-1} = \lambda_{\min}^{-1} = \lambda^{-1},$$

оценки сверху и снизу в (1.43) обязаны совпадать, т.е.

$$\|\Delta \mathbf{c}\|_2^2 = \lambda^{-1} \|\xi\|_2^2.$$

При этом спектральное число обусловленности $K(\mathbf{A})$ равно единице. Из этого в силу неравенств (1.53), (1.54) следует, что ошибка оценивания не может быть меньше суммарных ошибок в исходных данных. Если вдобавок

$$\lambda_i = 1, \quad i = \overline{1, M},$$

в соответствии с последним равенством норма вектора ошибок идентификации равна норме вектора ошибок в исходных данных.

2. МЕТОДЫ КОНТРОЛЯ ИНФОРМАТИВНОСТИ ДАННЫХ

2.1. Меры обусловленности и мультиколлинеарности

В разделах 1.5. и 1.6. показано, что точность идентификации зависит не только от ошибок в измерениях, но также от спектральных характеристик: минимального собственного значения и числа обусловленности матрицы A . Как следует из соотношений (1.46), (1.47), (1.49) – (1.51), при больших значениях числа обусловленности и/или λ_{\min}^{-1} даже небольшие ошибки в исходных данных могут привести к большим ошибкам в решениях. Задачи, для которых это имеет место, называют плохо обусловленными [21]. Одной из основных причин плохой обусловленности является «почти» линейная зависимость (мультиколлинеарность) [5] векторов - столбцов матрицы X . Можно утверждать, что следствием сильной мультиколлинеарности всегда является плохая обусловленность задачи. Обратное не всегда верно.

Векторы-столбцы матрицы X могут быть почти ортогональными, но сильно различаться параметрами масштаба, что неизбежно ведет к плохой обусловленности. Термин *обусловленность* употребляется для общей характеристики чувствительности решений к ошибкам в исходных данных, а термин *мультиколлинеарность* используется в тех случаях, когда необходимо указать конкретную причину возникновения плохой обусловленности. Поскольку в конечном итоге это всегда приводит к потере точности, мы будем применять также термин *информативность* данных.

В задачах идентификации по малому числу наблюдений характеристики обусловленности могут существенным образом изменяться при переходе от одного фрагмента изображения к другому. Поэтому в данном случае принципиально необходимым этапом является оценка обусловленности и мультиколлинеарности на каждом полученном наборе данных. В настоящем разделе наряду с известными мерами мультиколлинеарности и/или обусловленности описываются меры, которые специально разработаны [24] для оценки информативности данных по малому числу фиксированных наблюдений.

В регрессионном анализе качество оценок обычно характеризуют дисперсией. Оценка МНК становится малоэффективной, т.е. дисперсия оценок существенно возрастает, если имеет место мультиколлинеарность.

Несмотря на существование указанной зависимости, понятие мультиколлинеарности не связано со статистическими характеристиками сигналов, а является мерой сопряженности (но не корреляции) [5] независимых переменных в уравнении (10.6). Поэтому использование этого понятия в контексте сформулированных ранее предположений является вполне уместным. Дадим сравнительную характеристику мерам мультиколлинеарности, рассматривавшимся в работах [5], [24].

1. *Определитель информационной матрицы (Грама) $\mathbf{A}=\mathbf{X}^T\mathbf{X}$:*

$$\det(\mathbf{A}) = \prod_{i=1}^M \lambda_i . \quad (2.1)$$

Определитель может выступать в качестве меры мультиколлинеарности, если матрица Грама определенным образом нормирована. Например, можно вместо исходной матрицы \mathbf{A} рассматривать матрицу $\tilde{\mathbf{A}}$, полученную из нее по правилу

$$\tilde{\mathbf{A}} = \mathbf{A} / \text{tr}\mathbf{A} .$$

Сумма собственных значений нормированной матрицы $\tilde{\mathbf{A}}$ всегда равна единице. Поэтому близость к нулю $\det(\tilde{\mathbf{A}})$ может свидетельствовать либо о близости к нулю некоторых собственных значений, либо о большой размерности задачи. При отсутствии нормировки определитель может быть достаточно большим даже при плохой обусловленности задачи.

2. *Число обусловленности (1.50):*

$$K(\mathbf{A}) = \lambda_{\max} / \lambda_{\min} .$$

Соображения, по которым число обусловленности может выступать в качестве меры мультиколлинеарности, те же, что и для определителя нормированной матрицы. Однако в данном случае не требуется нормировка матрицы \mathbf{A} , т.к. число обусловленности не чувствительно к параметру масштаба. Эта мера широко используется в теории возмущений для анализа ошибок в решениях [2, 21]. Некоторые оценки для относительных ошибок, построенные в рамках этой теории, приведены в разделе 1.5.

3. *Минимальное собственное значение:*

$$\lambda_{\min}(\mathbf{A})$$

информационной матрицы

$$\mathbf{A} = \mathbf{X}^T \mathbf{X}.$$

Это универсальная мера мультиколлинеарности. Она чувствительна как к масштабу (физической размерности) независимых переменных, так и к мультиколлинеарности (сопряженности) соответствующих им векторов. В разделе 1.5 на основе этой меры получены оценки сверху для нормы вектора ошибок идентификации (1.46), (1.47).

4. *Показатель парной сопряженности.* В регрессионном анализе часто используется матрица сопряженности:

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}, \quad (2.2)$$

где

$$r_{ij} = (\mathbf{x}_i^T \mathbf{x}_j) / \left[(\mathbf{x}_i^T \mathbf{x}_i)^{1/2} (\mathbf{x}_j^T \mathbf{x}_j)^{1/2} \right] = \cos(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j),$$

а $\mathbf{x}_i, \mathbf{x}_j$ - i -й и j -й векторы-столбцы матрицы \mathbf{X} . В качестве показателя мультиколлинеарности выступает величина

$$r = \max_{i,j} |r_{ij}|, \quad i \neq j. \quad (2.3)$$

Достоинство этой меры – вычислительная простота. Однако по сравнению с другими она является наиболее слабой. Можно привести пример [5] матрицы

$$\mathbf{A} = \mathbf{X}^T \mathbf{X},$$

составленной с использованием трех *компланарных*, но *попарно линейно независимых* векторов. Матрица \mathbf{A} при этом вырождена, но мера (2.3) «не замечает» этого.

5. *Показатель максимальной сопряженности.* В качестве меры мультиколлинеарности, свободной от указанного недостатка, рассматривают величину [5]

$$R = \max_i |R_i|,$$

где

$$R_i^2 = \frac{\mathbf{x}_i^T \mathbf{X}_{M-1} [\mathbf{X}_{M-1}^T \mathbf{X}_{M-1}]^{-1} \mathbf{X}_{M-1}^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i}, \quad (2.4)$$

\mathbf{x}_i – i -й вектор-столбец матрицы \mathbf{X} , а \mathbf{X}_{M-1} – $N \times (M-1)$ – матрица, полученная из исходной путем «вычеркивания» этого (i -го) столбца. Геометрически R_i – косинус угла между вектором-столбцом \mathbf{x}_i матрицы \mathbf{X} и подпространством, натянутым на множество остальных $M-1$ векторов этой матрицы.

Указанная мера мультиколлинеарности является наиболее сильной. Если $R \neq 1$, гарантируется невырожденность задачи. Недостатком является необходимость вычисления обратной матрицы. Во-первых, это требует значительных вычислительных затрат. Во-вторых, если матрица \mathbf{A} плохо обусловлена, ее обращение вообще становится проблемой.

6. *Показатель минимальной сопряженности с нуль-пространством.* Эта мера мультиколлинеарности [24] позволяет в значительной мере избежать указанных выше трудностей и определяется как

$$S_{\min} = \min_i |S_i|, \quad (2.5)$$

где

$$S_i = \left(\sum \beta_j^2 \right)^{1/2}, \quad j = \overline{M, N}, \quad \beta_j = \langle \mathbf{x}_i, \mathbf{t}_{0j} \rangle$$

– проекция вектора-столбца \mathbf{x}_i матрицы \mathbf{X} на j -й собственный вектор, соответствующий нулевому собственному значению матрицы $\mathbf{X}_{M-1} \mathbf{X}_{M-1}^T$. Здесь \mathbf{X}_{M-1} – $N \times (M-1)$ – матрица, составленная из $(M-1)$ нормированных

$$(\|\mathbf{x}_i\| = 1, \quad i = \overline{1, M})$$

векторов-столбцов матрицы \mathbf{X} после исключения вектора \mathbf{x}_i . Геометрически S_i - косинус угла между вектором \mathbf{x}_i и нуль-пространством матрицы \mathbf{X}_{M-1}^T , составленной из $M-1$ векторов-строк.

Для вычисления меры (2.5) необходимо вначале вычислить все

$$S_i, \quad i = \overline{M, N}.$$

Можно показать [24], что для этого справедливы следующие соотношения.

$$S_i = (\mathbf{x}_i^T \mathbf{T}_{0, M-1} \mathbf{T}_{0, M-1}^T \mathbf{x}_i)^{1/2} / (\mathbf{x}_i^T \mathbf{x}_i)^{1/2}, \quad i = \overline{1, M}. \quad (2.6)$$

Здесь $\mathbf{T}_{0, M-1}$ - матрица, размерности $N \times (N-M+1)$, составленная из $N-M+1$ собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}_{M-1} \mathbf{X}_{M-1}^T$. В данном случае не требуется обращать матрицу. Кроме того, при малом числе наблюдений, когда размерность нуль-пространства не превышает размерность пространства параметров, объем вычислений по сравнению с мерой (2.4) меньше.

7. Показатель диагонального преобладания матрицы Грама. Соображения, по которым показатель диагонального преобладания, определяемый как

$$\phi = \frac{\text{tr}^2 \mathbf{A}}{\text{tr} \mathbf{A}^2} = \left(\sum_{i=1}^M a_{ii} \right)^2 / \sum_{i,j=1}^M a_{ij}^2, \quad (2.7)$$

может использоваться в качестве меры обусловленности и/или мультиколлинеарности связаны с известными свойствами матриц Грама [2,9]:

$$\text{tr}^2 \mathbf{A} = \left(\sum_{i=1}^M \lambda_i \right)^2, \quad \text{tr} \mathbf{A}^2 = \sum_{i=1}^M \lambda_i^2. \quad (2.8)$$

Ограничением на применение показателя (2.7) являются неравенства:

$$M - 1 < \phi \leq M. \quad (2.9)$$

Иногда [24] вместо показателя ϕ удобнее использовать приведенную к интервалу $[0,1]$ величину

$$\bar{\phi} = \phi - M + 1,$$

характеризующую превышение диагонального преобладания ϕ над левой границей допустимых значений в (2.9).

Важным достоинством показателя ϕ (или $\bar{\phi}$) является вычислительная простота. Однако в силу ограничений (2.9) он не всегда дает гарантированные оценки. Это не существенно, если задача заключается в достижении хорошей обусловленности.

Подведем итоги. Использование определителя и показателя парной сопряженности возможно лишь при наличии дополнительной априорной информации. Число обусловленности, минимальное собственное значение и показатель максимальной сопряженности являются достаточно полными характеристиками мультиколлинеарности, но их использование связано со значительными (в т.ч. вычислительными) трудностями. Наиболее подходящими для оценки информативности данных в вычислительном отношении являются показатель диагонального преобладания и показатель сопряженности с нуль-пространством. В следующих разделах мы рассмотрим более детально эти показатели, а также их связь с другими, упоминавшимися выше мерами, с точки зрения эффективности.

2.2. Оценка информативности данных по показателям диагонального преобладания

Как уже отмечалось, решающую роль в формировании ошибок идентификации играют собственные значения матрицы A . Однако в технологиях обработки сигналов и моделирования для контроля информативности данных предпочтительно использовать более простую в вычислительном отношении меру – показатель диагонального преобладания. Основанием для этого является также связь этого показателя с собственными значениями и числом обусловленности. Приведем неравенства, устанавливающие границы для собственных значений и числа обусловленности в зависимости от показателя диагонального преобладания.

Если показатель ϕ диагонального преобладания (2.7) удовлетворяет неравенствам (2.9):

$$M - 1 < \phi \leq M ,$$

имеют место следующие оценки сверху и снизу для собственных значений:

$$\lambda(\mathbf{A}) \leq M^{-1} \text{tr} \mathbf{A} \left(1 + \sqrt{(M/\phi - 1)(M - 1)} \right), \quad (2.10)$$

$$\lambda(\mathbf{A}) \geq M^{-1} \text{tr} \mathbf{A} \left(1 - \sqrt{(M/\phi - 1)(M - 1)} \right). \quad (2.11)$$

Доказательства этого утверждения можно найти в работах [23, 24]. Наглядное геометрическое объяснение неравенств (2.10), (2.11) дает рис. 2.2.

Для практических целей наибольший интерес представляет оценка снизу (2.11), т.к. она может использоваться для оценки верхней границы возможной ошибки идентификации при заданном ограничении на норму вектора ошибок в исходных данных. Действительно, неравенство (1.46) не нарушится, если вместо минимального собственного значения $\lambda_{\min} = \min_i \lambda_i(\mathbf{A})$ подставить его оценку (2.11). При этом получаем следующую легко вычисляемую оценку сверху для погрешности идентификации:

$$\|\Delta \mathbf{c}\|_2^2 \leq M^{-1} \left[\text{tr} \mathbf{A} \left(1 - \sqrt{(M/\phi - 1)(M - 1)} \right) \right]^{-1} R_{\xi}^2,$$

где R_{ξ} – заданная в (1.37) максимально возможная норма вектора ошибок ξ .

Заметим, что если матрица \mathbf{A} нормирована по правилу

$$\tilde{\mathbf{A}} = \mathbf{A} / \text{tr} \mathbf{A},$$

то оценка максимальной погрешности идентификации зависит только от погрешности исходных данных и величины диагонального преобладания информационной матрицы. Если указанной нормировке подвергается набор данных, для которого величина R_{ξ} уже задана, то она также должна быть изменена соответствующим образом. В работе [24] показано, что применение оценок (2.10), (2.11) во многих отношениях является более предпочтительным по сравнению с использованием аналогичных оценок, построенных на основе известных в литературе локализационных теорем [9].

В частности показано, что существует оценка сверху для числа обусловленности:

$$k(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} \leq \hat{k}_1^a, \quad (2.12)$$

где

$$\hat{k}_1^a = \frac{1 + \sqrt{(\phi - M + 2)^{-1}(M - \phi)}}{1 - \sqrt{(\phi - M + 2)^{-1}(M - \phi)}} = \frac{1 + \sqrt{(1 - \bar{\phi})/(1 + \bar{\phi})}}{1 - \sqrt{(1 - \bar{\phi})/(1 + \bar{\phi})}}. \quad (2.13)$$

Здесь $\bar{\phi}$ - приведенная к диапазону [0-1] величина диагонального преобладания:

$$\bar{\phi} = \phi - M + 1. \quad (2.14)$$

Напомним, что для фигурирующего в соотношениях (2.13) показателя ϕ должны выполняться ограничения (2.9):

$$M - 1 < \phi \leq M.$$

С использованием неравенств (2.10) (2.11) при тех же ограничениях на показатель ϕ может быть построена другая оценка. В частности, подставляя в (1.50) вместо λ_{\max} , λ_{\min} оценки (2.10), (2.11) соответственно, получаем следующую оценку сверху для спектрального числа обусловленности [24]

$$k(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} \leq \hat{k}_2^a, \quad (2.15)$$

где

$$\hat{k}_2^a = \frac{1 + \sqrt{(M/\phi - 1)(M - 1)}}{1 - \sqrt{(M/\phi - 1)(M - 1)}} = \frac{1 + \sqrt{(1 - \bar{\phi})(M - 1)(\bar{\phi} + M - 1)^{-1}}}{1 - \sqrt{(1 - \bar{\phi})(M - 1)^{-1}(\bar{\phi} + M - 1)^{-1}}}, \quad (2.16)$$

а показатель $\bar{\phi}$ тот же, что и в (2.13). На рис. 2.1 приведены графики изменения оценок \hat{k}_1^a, \hat{k}_2^a в функции приведенной величины диагонального преобладания $\bar{\phi}$.

Оценки сверху (2.15), (2.12) могут использоваться для построения гарантированных оценок для относительных ошибок идентификации, вычисляемых по соотношениям (1.49), (1.53), (1.54). Поэтому для выработки рекомендаций по их применению представляет интерес провести сравнительный анализ их эффективности.

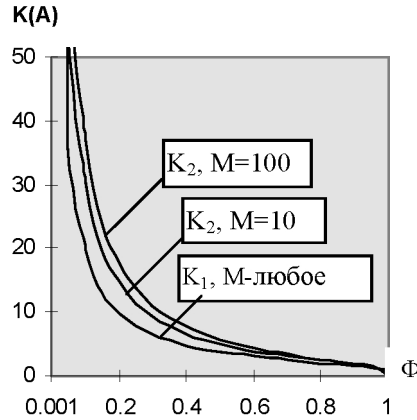


Рис. 2.1. Зависимости оценок k_1, k_2 от показателя $\bar{\phi}$

Из графиков видно, что оценка \hat{k}_1^a более сильная, чем \hat{k}_2^a , но преимущество незначительное и притом быстро убывает с ростом показателя $\bar{\phi}$. Более того, хотя оценка \hat{k}_2^a и зависит от порядка матрицы M , зависимость эта слабая, так что она остается достаточно эффективной даже при $M=100$. Поэтому выбор между этими оценками, по-видимому, должен определяться в первую очередь удобством вычислительной реализации.

Наглядное геометрическое представление приведенных выше неравенств и ограничений для матрицы Грама размерности 3×3 дает рисунок 2.2. Здесь для 3×3 -матрицы с собственными значениями $\lambda_1, \lambda_2, \lambda_3$ в соответствии со свойствами (2.8) матриц Грама показано множество

$$\Theta_2 = \{\lambda : t_1 = \text{Const}, t_2 = \text{Const}\},$$

одновременно удовлетворяющее уравнениям

$$\lambda_1 + \lambda_2 + \lambda_3 = t_1,$$

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = t_2^2.$$

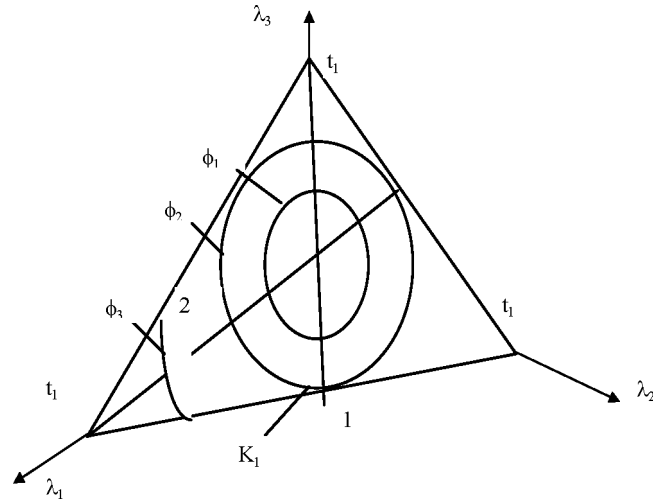


Рис. 2.2. Геометрическая интерпретация множества Θ_2

Множество Θ_2 , образованное пересечением плоскости и сферы, является окружностью. Очевидно, что для различных, но фиксированных пар t_1, t_2 можно построить различные окружности, каждой из которых будет соответствовать фиксированная величина диагонального преобладания ($\phi = t_1^2/t_2$ или $\bar{\phi} = \phi - M + 1$). На рис. 2.2 показаны окружности, соответствующие различным значениям ϕ ($\phi_1 > \phi_2 > \phi_3$). Для окружности, соответствующей величине показателя ϕ_2 цифрами 1 и 2 указаны точки, в которых достигаются оценки (2.10), (2.11) для собственных значений, а в непосредственной близости от точки 1 показана также точка $K_1^a(\mathbf{A})$, в которой имеет место максимально

возможное для соответствующего множества матриц число обусловленности.

Из рисунка 2.2 видно, что построение гарантированных оценок обусловленности по показателю диагонального преобладания при значениях ϕ в диапазоне

$$1 < \phi \leq M - 1$$

возможно только в случае, когда задана граница снизу для минимального собственного значения. Использование такого дополнительного условия неконструктивно, поскольку наличие априорной информации о минимальном собственном значении вообще снимает проблему оценивания обусловленности. Тем не менее, показатель диагонального преобладания ϕ (или $\bar{\phi}$), наряду с минимальным собственным значением и числом обусловленности, является важной характеристикой информативности данных.

Процедуры контроля информативности строятся как проверка достижения текущим значением величины диагонального преобладания порогового значения ϕ_{crit} . Пороговое значение может быть установлено экспериментально или рассчитано с использованием неравенств (2.10) – (2.12), которые определяют связь показателя диагонального преобладания с собственными значениями и числом обусловленности, определяющими достижимую точность.

2.3. Связь показателей информативности данных

Конечно, с точки зрения вычислительных затрат показатель диагонального преобладания является наиболее подходящими для контроля информативности данных. Однако при этом, как мы видели, имеет место ограничение на диапазон существования гарантированных оценок (2.9). Поэтому естественно возникает вопрос: насколько существенно это ограничение, насколько это ограничение влияет на эффективность показателя. Вопрос о степени эффективности возникает также и по отношению к показателю сопряженности с нуль-пространством, который, как отмечалось выше, в вычислительном отношении предпочтительнее показателя максимальной сопряженности. Для ответа на эти вопросы рассмотрим, насколько тесно связаны интересующие

нас показатели информативности с наиболее полными характеристиками обусловленности и мультиколлинеарности.

Связь показателей сопряженности с нуль-пространством - (2.5) и максимальной сопряженности (2.4) дается следующей леммой.

Лемма. Пусть \mathbf{X} - $N \times M$ -матрица: $N > M$ и $\text{Rank}(\mathbf{X}) = M$, а \mathbf{x} - $N \times 1$ - вектор, тогда

$$\mathbf{E}_N - \mathbf{X} \left[\mathbf{X}^T \mathbf{X} \right]^{-1} \mathbf{X}^T = \mathbf{T}_0 \mathbf{T}_0^T, \quad (2.17)$$

где \mathbf{T}_0 - $N \times (N - M)$ -матрица, составленная из собственных векторов, соответствующих нулевым собственным значениям $N \times N$ -матрицы $\mathbf{X}\mathbf{X}^T$.

Равенство (2.17) доказывается с использованием свойств

$$\mathbf{T}_0^T \mathbf{X} = 0 \quad (\mathbf{X}^T \mathbf{T}_0 = 0), \quad \mathbf{T}_0^T \mathbf{T}_0 = \mathbf{E}_{N-M}$$

путем умножения обеих частей (2.17) слева на \mathbf{T}_0^T и справа на \mathbf{T}_0 .

Лемма доказана.

Теперь в соответствии с (2.4), добавив в обеих частях по единице (в правой части она представлена слагаемым $\frac{\mathbf{x}_i^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i}$), запишем:

$$\begin{aligned} 1 - R_i &= \frac{\mathbf{x}_i^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i} - \frac{\mathbf{x}_i^T \mathbf{X}_{M-1} \left[\mathbf{X}_{M-1}^T \mathbf{X}_{M-1} \right]^{-1} \mathbf{X}_{M-1}^T \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i} = \\ &= \frac{\mathbf{x}_i^T \left[\mathbf{E}_N - \mathbf{X}_{M-1} \left[\mathbf{X}_{M-1}^T \mathbf{X}_{M-1} \right]^{-1} \mathbf{X}_{M-1}^T \right] \mathbf{x}_i}{\mathbf{x}_i^T \mathbf{x}_i}. \end{aligned}$$

Откуда, с учетом (2.17), следует, что

$$1 - R_{i,k} = S_{i,k}.$$

Ясно, что

$$R_{i,k} = 1 - S_{i,k}.$$

Наглядное объяснение приведенного результата дает представление величины R_i , фигурирующей в (2.4) в следующем виде

$$R_i = \langle \mathbf{X}_{M-1} \mathbf{z}_i, \mathbf{x}_i \rangle,$$

где

$$\mathbf{z}_i = [\mathbf{X}_{M-1}^T \mathbf{X}_{M-1}]^{-1} \mathbf{X}_{M-1}^T \mathbf{x}_i.$$

Применяя к указанным соотношениям ортогональные преобразования, можно убедиться, что показатель максимальной сопряженности (2.4) можно также трактовать следующим образом

$$R = \max_i |R_i|, \quad R_i = \left\{ \sum_j \alpha_j^2 \right\}^{1/2}, \quad j = \overline{1, M-1}, \quad i = \overline{1, M},$$

где

$$\alpha_j = \langle \mathbf{x}_i, \mathbf{t}_{kj} \rangle$$

– проекция вектора \mathbf{x}_i на j -й вектор базиса, образованного собственными векторами, соответствующими ненулевым собственным значениям матрицы $\mathbf{X}_{M-1} \mathbf{X}_{M-1}^T$. Отсюда, в частности, следует, что если вектор \mathbf{x}_i принадлежит пространству столбцов матрицы \mathbf{X}_{M-1} (задача вырождена, а показатель $R=1$), то он ортогонален нуль-пространству этой матрицы (при этом $S_{\min}=0$). С уменьшением величины R соответствующая ей величина S_{\min} увеличивается. По существу, показатели R и S_{\min} являются мерами одной и той же характеристики, но вычисляются посредством разных подпространств.

Необходимо особо подчеркнуть вычислительные преимущества меры (2.5) при обработке малых наборов данных. Объем вычислительной работы при использовании показателя S_{\min} будет меньше в ситуации, когда размерность нуль-пространства меньше, чем размерность пространства параметров. В действительности эти преимущества проявляются и в случае, когда общее число наблюдений превышает число оцениваемых параметров более чем в два раза. Дополнительный выигрыш имеет место вследствие того, что для определения матрицы $\mathbf{T}_{0, M-1}$, которая используется при вычислении S_i , не

требуется решение полной проблемы собственных значений и (в отличие от R_i) не нужно вычислять обратную матрицу.

На Рис. 2.3, а и б показана связь показателей $\bar{\phi}(A)$ и S_{min} со всеми обсуждавшимися в предыдущем разделе, мерами мультиколлинеарности и обусловленности в двух крайних ситуациях, характеризующихся как хорошо (а) и плохо (б) обусловленная задача. Схема напоминает модель направленного графа.

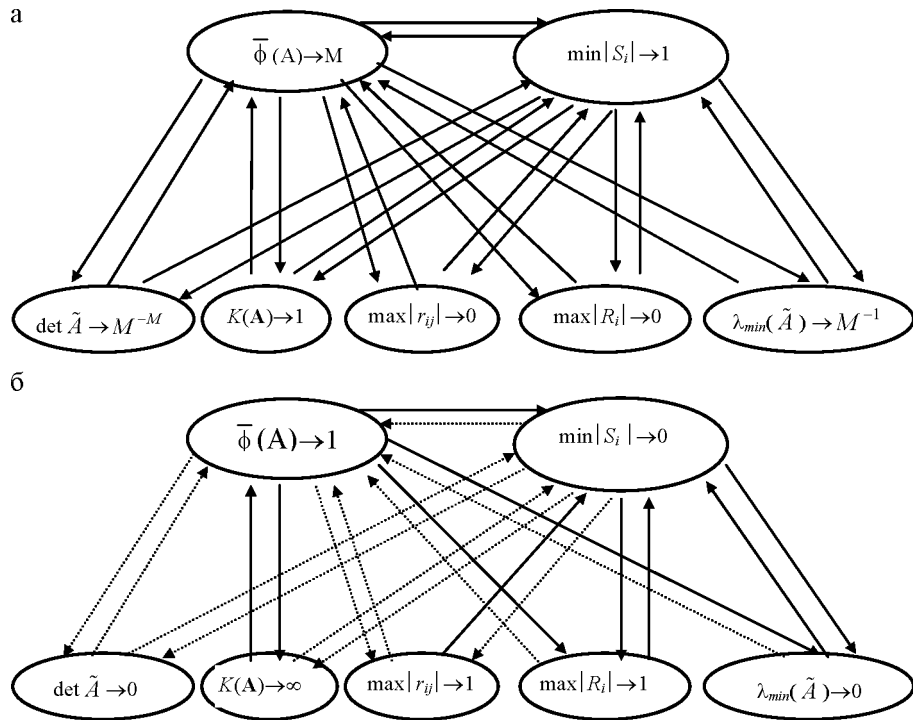


Рис. 2.3. Схема связей показателей мультиколлинеарности и обусловленности: а) в случае хорошо обусловленной задачи; б) в случае плохо обусловленной задачи

Здесь сплошная стрелка означает «из . . . обязательно следует . . .», а пунктирная – «из . . . как правило, но не обязательно, следует . . .». Пока-

занные на схемах связи вытекают из анализа приведенных выше аналитических соотношений мер мультиколлинеарности и обусловленности.

В заключение напомним, что некоторые из мер (S_{min} , $\max R$) учитывают лишь сопряженность (мультиколлинеарность), а другие ($\bar{\phi}(\mathbf{A})$, $K(\mathbf{A})$, λ_{min}) обнаруживают также и различия в масштабах векторов независимых переменных, т.е. являются мерами обусловленности. Ниже приводятся результаты экспериментов, иллюстрирующие связь и сравнительную эффективность мер мультиколлинеарности и обусловленности.

Выбор одной из мер мультиколлинеарности и/или обусловленности в конкретных информационных технологиях, конечно, должен осуществляться в зависимости от требуемой надежности вырабатываемых решений и вычислительных возможностей. Применяя к конкретному набору данных в качестве меры обусловленности показатель диагонального преобладания $\phi(\mathbf{A})$, мы можем всегда получить один из трех вариантов ответа:

1) задача плохо обусловлена и решение с требуемой точностью невозможно;

2) значение меры $\phi(\mathbf{A})$ на заданном фиксированном наборе данных не позволяет сделать уверенное заключение о достижимой точности оценки, т.к. $1 < \phi(\mathbf{A}) \leq M-1$;

3) задача хорошо обусловлена ($\phi(\mathbf{A}) > M-1$) и даются гарантированные оценки разрешимости задачи с требуемой точностью при заданной точности исходных данных.

В первом случае решение, которое следует принять, очевидно. Во втором случае возможность получения удовлетворительных оценок параметров вызывает сомнения и, следовательно, необходимо решить, готовы ли мы пойти на увеличение вычислительных затрат, чтобы провести дополнительный анализ информативности данных по другим показателям.

Иногда удастся увеличить показатель диагонального преобладания путем простого масштабирования исходной матрицы $\mathbf{A} = \mathbf{X}^T \mathbf{X}$ по правилу

$$\tilde{\mathbf{A}} = \mathbf{DAD}, \quad (2.18)$$

где

$$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_M)$$

– диагональная $M \times M$ - матрица масштабных коэффициентов:

$$d_k = 1/\sqrt{a_{k,k}}, \quad k = \overline{1, M}.$$

Заметим, что получаемая в результате такого масштабирования матрица является матрицей сопряженности (2.2):

$$\mathbf{A}(\mathbf{D}) = \mathbf{DAD} = \mathbf{R}.$$

Этот эффект не является неожиданным. Указанное масштабирование исключает влияние на показатель диагонального преобладания различий в масштабах независимых переменных. Поэтому масштабирование (2.18) при слабой корреляции независимых переменных приводит к тому, что вклад внедиагональных элементов оказывается малым по сравнению с диагональными. Матрица \mathbf{R} становится близкой к единичной и для нее

$$\bar{\phi} \cong 1.$$

Если же после масштабирования показатель $\bar{\phi}$ по-прежнему «не срабатывает», это свидетельствует о сильном влиянии мультиколлинеарности.

Это вовсе не означает, что задача оценивания не может быть решена с требуемой точностью. Возможно, на данной конкретной ре-

ализации следует использовать другие более сильные оценки, например, основанные на вычислении показателя сопряженности с нуль-пространством S_{\min} . Объем вычислений при этом, конечно, неизмеримо больше. Но это неизбежная плата за эффективность оценок. Использование показателя сопряженности с нуль-пространством может быть вполне оправдано высокой стоимостью риска, связанного с возможными ошибками оценивания.

В заключение отметим, что применение процедур анализа обусловленности в задачах обработки изображений может преследовать разные цели. Если предъявленный фрагмент изображения является единственным, то результаты сравнения допустимых пороговых значений с фактическими значениями мер обусловленности и/или мультиколлинеарности являются лишь констатацией факта (хорошей или плохой) его информативности.

Иногда преследуется цель подыскать на одной длинной реализации или среди множества различных сигналов участки или сигналы, наиболее

подходящие для решения задачи идентификации. Тогда анализ информативности должен заключаться в сравнении показателей диагонального преобладания на различных фрагментах (сигналах) и отборе тех из них, для которых он оказался выше. Такая задача рассматривалась, например, в работе [24].

3. АЛГОРИТМЫ ИДЕНТИФИКАЦИИ

3.1. Итерационное оценивание с корректировкой выходного вектора

После того как установлено, что сформированный набор данных информативен, можно перейти к решению задачи построения оценок. В этом и следующем разделах рассматриваются два метода оценивания, основанные на итерационной схеме. В настоящем разделе описывается метод с корректировкой выходного наблюдаемого вектора. В следующем разделе в рамках той же итерационной схемы рассматривается метод со взвешиванием наблюдений.

Итерационная схема оценивания широко применяется в задачах идентификации систем. При этом для корректировки данных и параметров алгоритма на промежуточных этапах, как правило, используется априорная статистическая информация. Если такая информация отсутствует, единственным источником дополнительной информации на промежуточных шагах являются сами искомые оценки и соответствующие им векторы невязок выхода объекта и модели. В данном случае итерационная схема строится с использованием векторов невязок следующим образом.

Определяется начальная МНК-оценка $\hat{\mathbf{c}}_0$ и вычисляется вектор невязок

$$\hat{\boldsymbol{\xi}}_0 = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}}_0.$$

По вектору невязок $\hat{\boldsymbol{\xi}}_0$ строится так называемый корректирующий вектор $\check{\boldsymbol{\xi}}$. С использованием этого вектора осуществляется корректировка выходного вектора по правилу

$$\tilde{\mathbf{y}} = \mathbf{y} - \check{\boldsymbol{\xi}}. \quad (3.1)$$

Затем вычисляется новая МНК-оценка:

$$\hat{\mathbf{c}}_1 = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \tilde{\mathbf{y}}, \quad (3.2)$$

и процесс корректировки (теперь уже вектора \tilde{y}) повторяется. Правило останова может заключаться в сравнении меры близости оценок на соседних шагах с заданным порогом.

В рамках описанной итерационной схемы ключевым моментом, определяющим точность оценок, является определение корректирующего вектора $\tilde{\xi}$. Будем строить корректирующий вектор, опираясь на предположение о близости выделяющихся (больших по абсолютной величине) компонентов векторов ошибок и невязок. Справедливость этой гипотезы легко проверить на простейшей задаче оценивания параметра сдвига. Интуитивно ясно, что отличие оценки от истинного параметра меньше всего повлияет на различия больших ошибок и невязок. Небольшие по абсолютной величине ошибки и соответствующие им невязки могут иметь даже разные знаки. Ниже приводится пример, показывающий, что это предположение справедливо и в случае идентификации параметров модели типа регрессии.

Рассмотрим один из методов построения корректирующего вектора $\tilde{\xi}$ или вектора поправок $\Delta\tilde{\xi} = \tilde{\xi} - \hat{\xi}$ по вектору невязок $\hat{\xi}$ для каждого фиксированного набора данных. На рисунке 3.1 видно, что при использовании любого из указанных корректирующих векторов ($\tilde{\xi}$ или $\Delta\tilde{\xi}$) точность оценивания будет одинаковой.

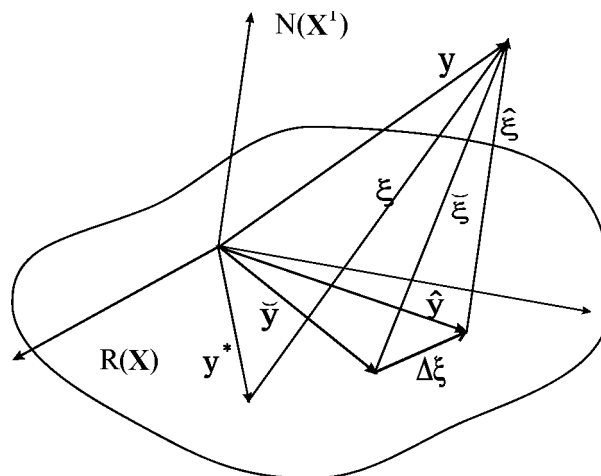


Рис. 3.1. Геометрическая интерпретация метода корректировки выхода модели

Предположим, что с использованием характеристик связи ошибок и невязок или путем дополнительных, более точных измерений нескольких отсчетов удалось задать ровно M (любых) компонент вектора $\check{\xi}$ достаточно близкими к соответствующим компонентам вектора ошибок ξ . Из соображений удобства изложения эти M компонент путем перестановки и перенумерации строк сделаем первыми, а соответствующий им вектор, являющийся составной частью вектора $\check{\xi}$, обозначим $\check{\xi}_M$.

Введем в рассмотрение $(N-M) \times 1$ -вектор $\check{\xi}_{N-M}$, составленный из оставшихся компонент вектора $\check{\xi}$, которые нам пока не известны. Корректирующий $N \times 1$ -вектор

$$\check{\xi} = \left[\check{\xi}_M ; \check{\xi}_{N-M} \right]$$

представляет собой прямую сумму векторов $\check{\xi}_M$ и $\check{\xi}_{N-M}$. Задача заключается в определении недостающих компонент, т.е. вектора $\check{\xi}_{N-M}$ по известным M компонентам вектора $\check{\xi}_M$.

Будем искать вектор $\check{\xi}_{N-M}$ из условия

$$\mathbf{T}_0^T (\check{\xi} - \hat{\xi}) = 0. \quad (3.3)$$

Основанием для этого служит известное равенство:

$$\mathbf{T}_0^T \xi = \mathbf{T}_0^T \hat{\xi} = \mathbf{T}_0^T \mathbf{y}, \quad (3.4)$$

где $\hat{\mathbf{c}}$ – МНК-оценка (10.7),

$\hat{\xi} = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}}$ – вектор невязок (10.8),

а \mathbf{T}_0 – фигурирующая в (10.12) $N \times (N-M)$ матрица, составленная из нормированных собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}\mathbf{X}^T$.

Равенство (3.4) утверждает, что проекции вектора ошибок ξ и вектора невязок $\hat{\xi}$ на собственные векторы, образующие нуль-пространство матрицы \mathbf{X}^T совпадают. Доказательства равенств (3.4) можно найти в работе [8].

Из (3.4) следует, если компоненты вектора $\check{\xi}$ совпадают с истинными значениями компонент вектора ошибок:

$$\check{\xi} = \xi,$$

то корректирующий вектор

$$\Delta\check{\xi} = \check{\xi} - \hat{\xi} = \xi - \hat{\xi},$$

удовлетворяющий этому равенству, будет ортогонален нуль-пространству. Свойство ортогональности вектора $\Delta\check{\xi}$ с нуль-пространством сохранится и в случае, когда это равенство приближительное:

$$\check{\xi} \cong \xi.$$

Таким образом, реализация указанного условия позволит оценить составляющую вектора ошибок, принадлежащую пространству столбцов матрицы \mathbf{X} .

В соответствии с указанным выше условием (3.3) сформируем следующую систему уравнений

$$\mathbf{T}_{0,N-M}^T \check{\xi}_{N-M} = \mathbf{t}_{0,M}. \quad (3.5)$$

Здесь $\mathbf{T}_{0,N-M}^T$ - матрица составленная из последних $N-M$ столбцов матрицы \mathbf{T}_0^T . Вектор $\mathbf{t}_{0,M}$ размерности $(N-M) \times 1$ формируется в соответствии с равенством

$$\mathbf{t}_{0,M} = \mathbf{T}_0^T \hat{\xi} - \mathbf{T}_{0,M}^T \check{\xi}_M,$$

где $\mathbf{T}_{0,M}^T$ - $(N-M) \times M$ - матрица, составленная из первых M столбцов той же матрицы \mathbf{T}_0^T .

Так как

$$\text{Rank}\mathbf{T}_0 = (N-M),$$

существует единственное решение системы (3.5), а искомый вектор $\tilde{\xi}_{N-M}$ определяется как

$$\tilde{\xi}_{N-M} = \mathbf{T}_{0,N-M}^{-1} \mathbf{t}_{0,M}. \quad (3.6)$$

Заметим, что при задании компонент вектора $\tilde{\xi}_M$ равными соответствующим компонентам вектора невязок:

$$\tilde{\xi}_M = \hat{\xi}_M,$$

в (3.6) будут получены оставшиеся компоненты вектора невязок:

$$\tilde{\xi}_{N-M} = \hat{\xi}_{N-M}.$$

Это не удивительно, т.к. задавая компоненты подобным образом, мы не вносим никакого нового знания. Для того, чтобы результат оценивания улучшился, вектор $\tilde{\xi}_M$ должен содержать дополнительное знание. Это могут быть любые априорные сведения, в т.ч. связь между выделяющимися компонентами векторов $\hat{\xi}$ и ξ , о которой говорилось выше. В работе [24] приводится пример, в котором использовался описанный метод. В этом примере показана возможность заметного повышения точности МНК-оценок путем последовательной корректировки одной выделяющейся компоненты наблюдаемого выходного вектора, хотя для этого не использовалась никакая априорная информация. Свойство близости выделяющихся компонент векторов ошибок и невязок не может рассматриваться как априорная информация об ошибках. Это дополнительная информация, характеризующая свойство самой процедуры получения МНК-оценок.

3.2. Итерационный алгоритм со взвешиванием данных

Рассмотрим построение итерационных алгоритмов оценивания, в которых на промежуточных шагах осуществляется линейное преобразование типа взвешивания. Оно заключается в том, что исходные матрица \mathbf{X} и вектор \mathbf{y} умножаются на $N \times N$ -матрицу \mathbf{G} :

$$\tilde{\mathbf{y}} = \mathbf{G}\mathbf{y}, \quad \tilde{\mathbf{X}} = \mathbf{G}\mathbf{X}. \quad (3.7)$$

Здесь

$$\mathbf{G} = \text{diag}(g_1, g_2, \dots, g_N)$$

– диагональная матрица:

$$\mathbf{G}^2 = \mathbf{G}^T \mathbf{G} > 0, \det \mathbf{G} \neq 0.$$

Заметим, что при этом неизвестный вектор ошибок ξ также преобразуется в

$$\tilde{\xi} = \mathbf{G}\xi. \quad (3.8)$$

Итерационная схема по существу такая же, как и в рассмотренном выше методе корректировки выходного вектора. Центральной задачей в данном случае является построение по вектору невязок весовой матрицы \mathbf{G} . Прежде чем рассматривать вопрос построения этой матрицы, полезно выявить условия, при которых с использованием метода взвешивания возможно повышение точности идентификации.

Предположим, что на некотором этапе идентификации получена МНК-оценка (1.16). Для квадрата евклидовой нормы, соответствующей этой оценке ошибки идентификации (10.21), имеет место равенство

$$\|\Delta \mathbf{c}\|_2^2 = \Delta \mathbf{c}^T \Delta \mathbf{c} = \xi^T \mathbf{X} [\mathbf{X}^T \mathbf{X}] \mathbf{X}^T \xi.$$

Пусть также каким-то образом задана диагональная матрица

$$\mathbf{G} = \text{diag}(g_1, g_2, \dots, g_N)$$

и после линейного преобразования данных (1.16)

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

построена новая оценка, для которой мера близости по аналогии с предыдущей имеет вид

$$\|\Delta \tilde{\mathbf{c}}\|_2^2 = \Delta \tilde{\mathbf{c}}^T \Delta \tilde{\mathbf{c}} = \tilde{\xi}^T \tilde{\mathbf{X}} [\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}] \tilde{\mathbf{X}}^T \tilde{\xi}.$$

Зададимся вопросом: всегда ли можно при произвольном, но фиксированном векторе ошибок ξ подобрать диагональную матрицу \mathbf{G} так, чтобы при этом выполнялось неравенство

$$\|\Delta c\|_2^2 \geq \|\Delta \tilde{c}\|_2^2.$$

Нетрудно заметить, что это неравенство эквивалентно требованию неотрицательной определенности квадратичной формы:

$$\xi^T \left[\mathbf{X} [\mathbf{X}^T \mathbf{X}]^{-2} \mathbf{X}^T - \mathbf{G}^T \tilde{\mathbf{X}} [\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}]^{-2} \tilde{\mathbf{X}}^T \mathbf{G} \right] \xi \geq 0. \quad (3.9)$$

Проанализируем, в каких случаях это возможно.

Вновь воспользуемся ортогональной $N \times N$ - матрицей

$$\mathbf{T} = [\mathbf{T}_\lambda : \mathbf{T}_0]$$

(1.23). Учитывая блочную структуру этой матрицы, а также свойства

$$\mathbf{T}_\lambda^T \mathbf{X} [\mathbf{X}^T \mathbf{X}]^{-2} \mathbf{X}^T \mathbf{T}_\lambda = \mathbf{A}^{-1},$$

$$\mathbf{X}^T \mathbf{T}_0 = 0 \quad (\mathbf{T}_0^T \mathbf{X} = 0),$$

$$\mathbf{T}_\lambda^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_\lambda = \mathbf{A}^{-1},$$

где

$$\tilde{\mathbf{V}} = \tilde{\mathbf{X}} [\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}]^{-2} \tilde{\mathbf{X}}^T,$$

неравенство (3.9) можно переписать в виде

$$\xi^T \mathbf{T} \begin{bmatrix} 0 & \mathbf{T}_\lambda^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_0 \\ \mathbf{T}_0^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_\lambda & \mathbf{T}_0^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_0 \end{bmatrix} \mathbf{T}^T \xi \leq 0.$$

Здесь

$$\xi^T \mathbf{T} = \xi^T [\mathbf{T}_\lambda : \mathbf{T}_0] = [\xi^T \mathbf{T}_\lambda, \xi^T \mathbf{T}_0],$$

а

$$\mathbf{T}^T \xi = [\mathbf{T}_\lambda^T \xi, \mathbf{T}_0^T \xi]^T.$$

С учетом блочной структуры матрицы, фигурирующей в последнем неравенстве, квадратичную форму (3.9) можно представить в виде

$$\xi^T \left[\mathbf{T}_0 \mathbf{T}_0^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_\lambda \mathbf{T}_\lambda^T + \mathbf{T}_\lambda \mathbf{T}_\lambda^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_0 \mathbf{T}_0^T + \mathbf{T}_0 \mathbf{T}_0^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_0 \mathbf{T}_0^T \right] \xi \leq 0. \quad (3.10)$$

На основе (3.10) можно сделать следующие выводы. Если вектор ошибок ξ принадлежит нуль-пространству матрицы

$$\mathbf{X}^T (\mathbf{T}_\lambda^T \xi = 0),$$

никаким выбором весовой матрицы нельзя улучшить точность оценок. Действительно, матрица $\mathbf{T}_0 \mathbf{T}_0^T \mathbf{G}^T \tilde{\mathbf{V}} \mathbf{G} \mathbf{T}_0 \mathbf{T}_0^T$ неотрицательно-определенная (по построению) и все ее собственные значения могут обращаться в нуль лишь при

$$\mathbf{G} = \alpha \mathbf{E}_N,$$

где α любое число. При этом квадратичная форма (3.10) принимает вид

$$\xi^T \mathbf{T}_0 \mathbf{T}_0^T \mathbf{X} \left[\mathbf{X}^T \mathbf{X} \right]^{-2} \mathbf{X}^T \mathbf{T}_0 \mathbf{T}_0^T \xi = 0.$$

Для всякой другой весовой матрицы \mathbf{G} собственные значения указанной матрицы неотрицательны и указанное неравенство (3.10) (для отличающегося от нулевого) вектора ошибок невозможно.

Этот вывод не является неожиданным. Действительно, в случае, когда $\xi \in N(\mathbf{X}^T)$ в соответствии с (3.10) ошибка идентификации равна нулю, а всякое изменение ориентации вектора ошибок относительно пространства столбцов матрицы \mathbf{X} может лишь ухудшить качество оценки. Если $\xi \in R(\mathbf{X})$ изменением весовой матрицы \mathbf{G} также ничего сделать нельзя. Этот вывод очевиден, т.к. при

$$\mathbf{T}_0^T \xi = 0$$

все слагаемые в левой части (3.10) обращаются в нуль.

Во всех остальных случаях ($\mathbf{T}_\lambda^T \xi \neq 0$, $\mathbf{T}_0^T \xi \neq 0$) в классе диагональных матриц всегда можно подобрать, и притом не единственную, матрицу \mathbf{G} , для которой в (3.10) будет иметь место строгое неравенство.

Рассмотрим теперь способы построения весовой матрицы. В методе взвешивания, также как и при корректировке выходного вектора в неявном виде, эксплуатируется гипотеза о близости выделяющихся компонентов векторов ошибок и невязок. Имея в виду указанное свойство, ищется матрица

$$\mathbf{G} = \text{diag}(g_1, g_2, \dots, g_N) : Q(\mathbf{G}, \hat{\xi}) \rightarrow \min_{\mathbf{G}} \quad (3.11)$$

при заданных ограничениях.

Общая схема итерационного алгоритма со взвешиванием данных строится следующим образом. Вычисляются начальная МНК-оценка $\hat{\mathbf{c}}_0$ и соответствующий вектор невязок $\hat{\xi}_0 = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}}_0$. Далее в соответствии с (3.11) строится весовая матрица \mathbf{G} . Затем вычисляется новая МНК-оценка:

$$\hat{\mathbf{c}}_1 = [\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}]^{-1} \tilde{\mathbf{X}}^T \mathbf{y}. \quad (3.12)$$

С использованием оценки $\hat{\mathbf{c}}_1$ определяются новый вектор невязок и весовая матрица и т.д. Процесс повторяется до выполнения условия остановки: близости оценок на соседних итерациях.

Нетрудно заметить, что существо метода заключается в попеременном построении оптимальных оценок параметров и оценок весовой матрицы на последовательности шагов. Этот метод в идейном отношении близок к методам построения адаптивных систем, в которых параметры корректируются на промежуточных этапах с учетом дополнительной информации, появляющейся в процессе функционирования.

В таблице 3.1 приведено четыре варианта формулировки критериев и соответствующие им выражения для вычисления весовых коэффициентов.

Нетрудно заметить, что при вычислении весовых коэффициентов по соотношениям, приведенным в п. 1 и 3 указанной таблицы, могут возникнуть проблемы при $\hat{\xi}_k = 0$ ($\check{\xi}_k = 0$). Поэтому при практической реализации в знаменателях соответствующих выражений следует добавлять малые положительные числа. Для других критериев элементы весовой матрицы могут быть вычислены, даже если среди компонент вектора невязок есть нулевые. Это достигнуто соответствующим заданием критерия.

Таблица 3.1 Критерии и соотношения для определения весовых коэффициентов

| № п/п | Вид критерия и ограничений | Выражения для вычисления весовых коэффициентов |
|-------|---|--|
| 1 | $\hat{\xi}^T \mathbf{G}^2 \hat{\xi} \rightarrow \min_{\mathbf{G}},$ $\sum_{i=1}^N g_i = N, \quad g_i > 0.$ | $g_k = \frac{g_0}{\hat{\xi}_k^2}, \quad k = \overline{1, N},$ <p>где $g_0 = N / \sum_{i=1}^N \hat{\xi}_i^{-2}, \quad k = \overline{1, N}.$</p> |
| 2 | $g_q^2 \sum_{i=1}^q \hat{\xi}_i^2 + \sum_{i=q+1}^N g_i^2 \hat{\xi}_i^2 \rightarrow \min_{\mathbf{G}},$ $q \cdot g_q + \sum_{i=q+1}^N g_i = N, \quad g_i > 0,$ $ \hat{\xi}_1 \leq \hat{\xi}_2 \leq \dots \leq \hat{\xi}_q \leq \dots \leq \hat{\xi}_N .$ | $g_q = N / \left[q + q^{-1} \left(\sum_{k=q+1}^N \hat{\xi}_k^{-2} \right) \left(\sum_{i=1}^q \hat{\xi}_i^2 \right) \right]$ $g_k = \left(\frac{g_q}{q} \sum_{i=1}^q \hat{\xi}_i^2 \right) \frac{1}{\hat{\xi}_k^2}, \quad k = \overline{(q+1), N}$ |
| 3 | $\sum_{i=1}^N g_i^2 \hat{\xi}_i \rightarrow \min_{\mathbf{G}},$ $\sum_{i=1}^N g_i = N, \quad g_i > 0.$ | $g_k = g_0 / \hat{\xi}_k , \quad k = \overline{1, N},$ <p>где $g_0 = N / \sum_{i=1}^N \hat{\xi}_i ^{-1}.$</p> |
| 4 | $g_q^2 \sum_{i=1}^q \hat{\xi}_i + \sum_{i=q+1}^N g_i^2 \hat{\xi}_i \rightarrow \min_{\mathbf{G}},$ $q \cdot g_q + \sum_{i=q+1}^N g_i = N, \quad g_i > 0$ $ \hat{\xi}_1 \leq \hat{\xi}_2 \leq \dots \leq \hat{\xi}_q \leq \dots \leq \hat{\xi}_N .$ | $g_q = N / \left[q + q^{-1} \left(\sum_{i=1}^q \hat{\xi}_i \right) \left(\sum_{j=q+1}^N \hat{\xi}_j ^{-1} \right) \right]$ $g_k = \left(\frac{g_q}{q} \sum_{i=1}^q \hat{\xi}_i \right) \frac{1}{ \hat{\xi}_k }, \quad k = \overline{(q+1), N}.$ |

3.3. Идентификация на основе метода согласованных оценок

Метод согласованного оценивания по малому числу наблюдений опирается на предположения 4 и 5. Конечно, по доступным наблюдениям из системы (1.5) невозможно построить точную систему (1.17), т.к. неизвестен вектор ошибок ξ . Однако опираясь на предположение 4, можно ставить задачу отыскания подсистемы наиболее свободной от шума. В качестве опознавателя такой подсистемы в силу предположения 5 предлагается использовать ко-

личественную меру согласованности (взаимной близости) множества оценок, полученных на подсистемах малой размерности. Заметим, что все оценки при этом должны вычисляться по однотипному правилу. В соответствии с изложенным принципом общую схему построения согласованных оценок можно представить следующим образом.

Из N строк исходной системы (1.5) сформируем, например, L подсистем размерности P :

$$\mathbf{y}_l = \mathbf{X}_l \mathbf{c} + \xi_l, \quad l = \overline{1, L}, \quad (3.13)$$

где \mathbf{X}_l - $P \times M$ - матрицы, а \mathbf{y}_l, ξ_l $P \times 1$ - векторы.

Из P строк каждой l -й подсистемы (3.13), в свою очередь, сформируем K_l подсистем размерности S :

$$\mathbf{y}_{l,i} = \mathbf{X}_{l,i} \mathbf{c} + \xi_{l,i}, \quad i = \overline{1, K_l}, \quad l = \overline{1, L}. \quad (3.14)$$

Для различения подсистем (3.13) и (3.14) далее будем их называть подсистемами верхнего и нижнего уровня соответственно. Индекс i здесь и далее служит для обозначения номера подсистемы нижнего уровня и соответствующих ей $S \times M$ - матрицы $\mathbf{X}_{l,i}$ и $S \times 1$ - векторов $\mathbf{y}_{l,i}, \xi_{l,i}$. Предполагается также, что для всех $i = \overline{1, K_l}$ и $l = \overline{1, L}$

$$\text{Rank } \mathbf{X}_{l,i} = M, \quad (3.15)$$

т.е. все подсистемы нижнего уровня имеют полный ранг. Однако обусловленность матриц $\mathbf{X}_{l,i}$ для различных l и i может существенно различаться. Поэтому должна быть предусмотрена проверка обусловленности. Соответствующие процедуры будут рассмотрены в разделе 3.4.

С использованием введенных подсистем принцип согласованности формулируется следующим образом. Пусть для каждой i -й подсистемы нижнего уровня, сформированной из l -й подсистемы верхнего уровня, вычислен вектор оценок $\hat{\mathbf{c}}_{l,i}$. Пусть также задана функция, характеризующая взаимную близость множества решений $\hat{\mathbf{c}}_{l,i}$, на каждой подсистеме верхнего уровня:

$$W[l] = \sum_{i=1, \overline{K_l}} \rho(\hat{\mathbf{c}}_{l,i}, \mathbf{c}_{l,i}^*), \quad l = \overline{1, L}. \quad (3.16)$$

Здесь $\rho(\hat{\mathbf{c}}_{l,i}, \mathbf{c}_{l,i}^*)$ - расстояние между вектором оценок $\hat{\mathbf{c}}_{l,i}$ и вектором $\mathbf{c}_{l,i}^*$, заданным или сформированным некоторым способом на множестве оценок, полученных на l -й подсистеме верхнего уровня. Искомая окончательная оценка $\hat{\mathbf{c}}$ ищется по тому же или иному правилу на подсистеме с номером l^* :

$$W[l^*] = \min_{\forall l} W[l], \quad l = \overline{1, L}.$$

В рамках описанного подхода могут строиться различные процедуры, отличающиеся видом функции $W[l]$ (3.16), способом выбора $\mathbf{c}_{l,i}^*$ и мерой близости $\rho(\hat{\mathbf{c}}_{l,i}, \mathbf{c}_{l,i}^*)$.

В настоящем разделе исследуются свойства оценок и строится алгоритм идентификации для частного случая функции (3.16), так называемой функции парной близости оценок:

$$W_l[l] = \sum_{i,j \in \overline{K_l}} (\hat{\mathbf{c}}_{l,i} - \hat{\mathbf{c}}_{l,j}), \quad l = \overline{1, L}, \quad (3.17)$$

где $\hat{\mathbf{c}}_{l,i}, \hat{\mathbf{c}}_{l,j}, i = \overline{1, K_l}, j = \overline{1, K_l}$ - оценки, получаемые на подсистемах нижнего уровня, сформированных из l -й подсистемы верхнего уровня. Далее мы докажем теорему, устанавливающую количественную оценку величины функции парной близости для случая, когда подсистемы нижнего уровня образованы с использованием квадратных $M \times M$ -матриц, но сначала выясним условия получения согласованных оценок при разбиении исходной системы (1.5) на подсистемы. Для этого проведем анализ ошибок оценивания на подсистемах с использованием ортогональных представлений (1.19) – (1.23).

Лемма.

Для системы (3) с $N \times M$ -матрицей \mathbf{X} , $\text{Rank } \mathbf{X} = M$ МНК-оценка, вычисленная для любой (i -й) подсистемы $\mathbf{y}_i = \mathbf{X}_i \mathbf{c} + \boldsymbol{\xi}_i$ с $S \times M$ -матрицей \mathbf{X}_i , $M \leq S < N$, $\text{Rank } \mathbf{X}_i = M$, может быть представлена в виде

$$\hat{\mathbf{c}}_i = \mathbf{c} + \Delta\mathbf{c} + \Delta\tilde{\mathbf{c}}_i, \quad i = 1, 2, \dots, \quad (3.18)$$

где $\Delta\mathbf{c} = \mathbf{F}\mathbf{A}^{-1/2}\mathbf{T}_\lambda^T\boldsymbol{\xi}$,

$$\Delta\tilde{\mathbf{c}}_i = [\mathbf{X}_i^T\mathbf{X}_i]^{-1}\mathbf{X}_i^T\mathbf{T}_{0,i}\mathbf{T}_0^T\boldsymbol{\xi},$$

а $\mathbf{T}_{0,i}$ - соответствующий i -й подсистеме блок матрицы \mathbf{T}_0 .

Доказательство.

С учетом (12), (13) вектор ошибок $\boldsymbol{\xi} = \mathbf{y} - \mathbf{X}\mathbf{c}$ может быть представлен в виде двух составляющих:

$$\boldsymbol{\xi} = \mathbf{X}\Delta\mathbf{c} + \mathbf{T}_0\Delta\mathbf{d}, \quad (3.19)$$

где

$$\Delta\mathbf{c} = \mathbf{F}\mathbf{A}^{-1/2}\mathbf{T}_\lambda^T\boldsymbol{\xi}, \quad (3.20)$$

$$\Delta\mathbf{d} = \mathbf{T}_0^T\boldsymbol{\xi}. \quad (3.21)$$

Используя представления (3.19)-(3.21), исходную систему (3) можно переписать в виде

$$\mathbf{y} = \mathbf{X}(\mathbf{c} + \Delta\mathbf{c}) + \mathbf{T}_0\Delta\mathbf{d}, \quad (3.22)$$

откуда, в частности, следует известный факт [8], что ошибка $\Delta\mathbf{c}$ МНК-оценки:

$$\hat{\mathbf{c}} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{c} + \Delta\mathbf{c} \quad (3.23)$$

зависит лишь от проекции вектора $\boldsymbol{\xi}$, принадлежащей пространству $\mathcal{R}(\mathbf{X})$.

Для любой (i -й) подсистемы размерности S , $M \leq S < N$ также можно построить разложение типа (3.22):

$$\mathbf{y}_i = \mathbf{X}_i(\mathbf{c} + \Delta\mathbf{c}) + \mathbf{T}_{0,i}\Delta\mathbf{d}, \quad (3.24)$$

где $\mathbf{T}_{0,i}$ - соответствующий i -й подсистеме блок матрицы \mathbf{T}_0 , а $\Delta \mathbf{c}$, $\Delta \mathbf{d}$ те же, что и в (3.20), (3.21). Умножив обе части (3.24) слева сначала на \mathbf{X}_i^T , а затем на $[\mathbf{X}_i^T \mathbf{X}_i]^{-1}$, приходим к (3.18). *Лемма доказана.*

Из леммы вытекают следующие возможные случаи (тривиальный случай $\xi = 0$ мы не рассматриваем).

1. $\mathbf{T}_\lambda^T \xi \neq 0$, $\mathbf{T}_0^T \xi = 0$ ($\xi \in \mathcal{R}(\mathbf{X})$). Тогда оценки $\hat{\mathbf{c}}_i$, $i = 1, 2, \dots$ будут одинаковыми (согласованными) для любых подсистем системы (1.5).
2. $\mathbf{T}_\lambda^T \xi = 0$, $\mathbf{T}_0^T \xi \neq 0$ ($\xi \in \mathcal{M}(\mathbf{X}^T)$). В этом случае оценки $\hat{\mathbf{c}}_i$, $i = 1, 2, \dots$ будут различающимися (не согласованными), т.к. при этом $\Delta \mathbf{c} = 0$, а ошибки $\Delta \tilde{\mathbf{c}}_i$, $i = 1, 2, \dots$ различны.
3. $\mathbf{T}_\lambda^T \xi \neq 0$, $\mathbf{T}_0^T \xi \neq 0$ (вектор имеет проекции как в $\mathcal{R}(\mathbf{X})$, так и в $\mathcal{M}(\mathbf{X})$). В данном случае в соответствии с (3.18) оценки на разных подсистемах также будут различающимися.

Для простоты мы рассмотрели случай одноуровневого разбиения исходной системы (1.5) с $N \times M$ -матрицей \mathbf{X} на подсистемы. Лемма остается справедливой также в случае, когда вместо исходной системы (1.5) рассматривается подсистема (верхнего уровня) (3.13) с $P \times M$ -матрицей \mathbf{X}_l . При этом согласованность множества оценок, соответствующих множеству подсистем нижнего уровня, сформированных на данной подсистеме верхнего уровня, может свидетельствовать о том, что эта подсистема наиболее свободна от шума.

Действительно, при отборе подсистемы верхнего уровня размерности P по показателю $W(l)$ точность оценок на подсистеме в случае 1 будет такой же, как и точность МНК-оценок, вычисленных по всей исходной системе размерности $N > P$. В случаях 2 и 3 отбор оценок по показателю согласованности может существенно улучшить МНК-оценку. Как следует из (3.18), в этих случаях согласованность оценок может свидетельствовать о том, что все ошибки

$$\Delta \tilde{\mathbf{c}}_i, \quad i = 1, 2, \dots$$

малы по сравнению с ошибкой $\Delta \mathbf{c}$. Случай, когда эти ошибки достаточно велики и, тем не менее, одинаковы по величине и знаку (что приводит к согла-

сованности ошибочных оценок), для случайных векторов ошибок ξ по предположению представляется маловероятным.

Получим теперь количественные оценки функции парной близости (3.17) в случае, когда подсистемы нижнего уровня построены с использованием квадратных $M \times M$ - матриц \mathbf{X}_i .

Теорема. Пусть на исходной системе (1.5) или l -й подсистеме верхнего уровня (3.13) с $P \times M$ -матрицей \mathbf{X} , $\text{Rank } \mathbf{X} = M$ задано K подсистем нижнего уровня (3.14) с квадратными $M \times M$ - матрицами \mathbf{X}_i , $\text{Rank } \mathbf{X}_i = M$, $i = \overline{1, K}$.

Тогда величина функции взаимной близости (3.17), вычисленной по соответствующим этим подсистемам МНК-оценкам $\hat{\mathbf{c}}_i$, $i = \overline{1, K}$, определяется равенством:

$$W(l) = 2 \sum_{\substack{i=\overline{1, K-1}, \\ j=i+1, K}} \|\hat{\mathbf{c}}_i - \hat{\mathbf{c}}_j\|^2 = W_1(l) + W_2(l), \quad (3.25)$$

$$\text{где } W_1(l) = (L-1) \sum_{i=1}^K \xi_i^T (\mathbf{X}_i \mathbf{X}_i^T)^{-1} \xi_i, \quad (3.26)$$

$$W_2(l) = 2 \sum_{\substack{i=\overline{1, K-1}, \\ j=i+1, K}} \xi_i^T (\mathbf{X}_j \mathbf{X}_j^T)^{-1} \xi_j. \quad (3.27)$$

Доказательство. Пусть

$$\mathbf{X}_i \mathbf{c} = \mathbf{y}_i - \xi_i \quad (3.28)$$

- подсистема с квадратной $M \times M$ -матрицей \mathbf{X}_i , $\text{Rank } \mathbf{X}_i = M$, соответствующая некоторой i -й, $i = \overline{1, K}$ подсистеме нижнего уровня (3.14). Соответствующая этой подсистеме оценка $\hat{\mathbf{c}}_i$ является решением уравнения

$$\mathbf{X}_i \hat{\mathbf{c}}_i = \mathbf{y}_i.$$

Вычитая из обеих частей этого равенства соответствующие (левую и правую) части равенства (3.28), имеем

$$\mathbf{X}_i \Delta \hat{\mathbf{c}}_i = \xi_i, \quad (3.29)$$

где $\Delta \hat{\mathbf{c}}_i = \hat{\mathbf{c}}_i - \mathbf{c}$ - вектор ошибок оценивания на i -й подсистеме нижнего уровня.

Рассматривая оценки на i -й и j -й подсистемах нижнего уровня можно записать:

$$\hat{\mathbf{c}}_i - \hat{\mathbf{c}}_j = (\hat{\mathbf{c}}_i - \mathbf{c}) - (\hat{\mathbf{c}}_j - \mathbf{c}) = \Delta \hat{\mathbf{c}}_i - \Delta \hat{\mathbf{c}}_j. \quad (3.30)$$

С учетом (3.30) функцию взаимной парной близости оценок (3.17) для всех $i = \overline{1, K}$, $j = \overline{1, K}$ можно представить в виде

$$\begin{aligned} W(l) &= 2 \sum_{\substack{i=\overline{1, K-1}, \\ j=i+1, K}} \|\hat{\mathbf{c}}_i - \hat{\mathbf{c}}_j\|^2 = 2 \sum_{\substack{i=\overline{1, K-1}, \\ j=i+1, K}} \|\Delta \hat{\mathbf{c}}_i - \Delta \hat{\mathbf{c}}_j\|^2 = \\ &= 2 \sum_{\substack{i=\overline{1, K-1}, \\ j=i+1, K}} (\Delta \hat{\mathbf{c}}_i - \Delta \hat{\mathbf{c}}_j)^T (\Delta \hat{\mathbf{c}}_i - \Delta \hat{\mathbf{c}}_j). \end{aligned} \quad (3.31)$$

В соответствии с (3.29) $\Delta \hat{\mathbf{c}}_i = \mathbf{X}_i^{-1} \xi_i$ (напомним, что по предположению $\text{Rank}(\mathbf{X}_i) = M$). Подставив указанное значение $\Delta \hat{\mathbf{c}}_i$, $\Delta \hat{\mathbf{c}}_j$ в (3.31), получим

$$W(l) = \sum_{i,j} \left[\xi_i^T (\mathbf{X}_i \mathbf{X}_i^T)^{-1} \xi_i + \xi_j^T (\mathbf{X}_j \mathbf{X}_j^T)^{-1} \xi_j - \xi_i^T (\mathbf{X}_j \mathbf{X}_i^T)^{-1} \xi_j - \xi_j^T (\mathbf{X}_i \mathbf{X}_j^T)^{-1} \xi_i \right] \quad (3.32)$$

Фигурирующие в (3.32) индексы $i = \overline{1, K}$, $j = \overline{1, K}$ «пробегают» один и тот же набор подсистем. Поэтому первые два слагаемых в правой части можно заменить одной суммой - $W_1(l)$ (3.26). Аналогично последние (два) слагаемых со знаком минус также можно заменить одной суммой - $W_2(Q_j)$ (3.27).

Теорема доказана.

Составляющие (3.26), (3.27) функции взаимной близости (3.25) можно представить в виде более удобном для геометрической интерпретации. Для этого введем замены:

$$\bar{\xi}_i = \mathbf{T}_i^T \xi_i \quad (\bar{\xi}_i^T = \xi_i^T \mathbf{T}_i). \quad (3.33)$$

С учетом (3.33) соотношение (3.26) можно записать в виде

$$W_1(l) = (K-1) \sum_{i=1}^K \bar{\xi}_i^T \Lambda_i^{-1} \bar{\xi}_i = (K-1) \sum_{i=1}^K \sum_{m=1}^M \|\xi_i\|^2 \lambda_{i,m}^{-1} \cos^2 \left(\hat{\mathbf{t}}_{i,k} \xi_i \right), \quad (3.34)$$

где

$$\|\xi_i\|^2 = \xi_i^T \xi_i$$

- квадрат нормы вектора ошибок измерений на i -й подсистеме нижнего уровня, а $\cos(\hat{\mathbf{t}}_{i,m} \xi_i)$ - косинус угла между этим вектором и собственным вектором, соответствующим собственному значению $\lambda_{i,m}$. Здесь $M \times M$ - матрицы \mathbf{T}_i , $i = \overline{1, L}$ такие, что

$$\mathbf{T}_i^T \mathbf{X}_i \mathbf{X}_i^T \mathbf{T}_i = \Lambda_i, \quad \mathbf{X}_i \mathbf{X}_i^T = \mathbf{T}_i \Lambda_i \mathbf{T}_i^T,$$

где

$\Lambda_i = \text{diag}(\lambda_{i,1}, \lambda_{i,2}, \dots, \lambda_{i,M})$ - диагональная матрица собственных значений $M \times M$ -матрицы $\mathbf{X}_i \mathbf{X}_i^T$ (или $\mathbf{X}_i^T \mathbf{X}_i$), $i = \overline{1, K}$.

Аналогично, используя в (3.27) замены (3.33), можно записать

$$W_2(l) = 2 \sum_{\substack{i=1, \overline{L-1}, \\ j=i+1, L}}^L \|\xi_i\| \cdot \|\xi_j\| \sum_{k=1}^M \|\mathbf{x}_{i,k}^{-1}\| \cdot \|\mathbf{x}_{j,k}^{-1}\| \cos \left(\hat{\mathbf{x}}_{i,k}^{-1} \xi_i \right) \cos \left(\hat{\mathbf{x}}_{j,k}^{-1} \xi_j \right), \quad (3.35)$$

где

$\mathbf{x}_{i,m}^{-1}$, $\mathbf{x}_{j,m}^{-1}$ - m -е строки матриц \mathbf{X}_i^{-1} и \mathbf{X}_j^{-1} , а

$\cos \left(\hat{\mathbf{x}}_{i,m}^{-1} \xi_i \right)$, $\cos \left(\hat{\mathbf{x}}_{j,m}^{-1} \xi_j \right)$ - косинусы углов между этими векторами-

строками и соответствующими векторами ошибок измерений.

Свойства слагаемых $W_1(l)$ (3.26) и $W_2(l)$ (3.27) и их вклад в функцию взаимной близости (3.25) существенно различны. Из (3.26), (3.34) видно, что

все слагаемые в $W_1(l)$ неотрицательны и их вклад будет тем значительнее, чем больше нормы векторов ошибок измерений и хуже обусловленность подсистем нижнего уровня. Вклад суммы $W_2(l)$ в значительной степени зависит от взаимной ориентации векторов ошибок измерений ξ_i , ξ_j разных подсистем нижнего уровня, а также ориентации этих векторов относительно образованных матрицами этих подсистем пространств.

Если ошибки случайные и независимые, слагаемые в (3.27), (3.35) могут иметь разные знаки или быть равными нулю. При этом сумма $W_2(l)$ может оказаться весьма малой, а основной вклад в функцию взаимной близости будет вносить сумма $W_1(l)$. Как видно из (3.26), (3.34), эта сумма тем меньше, чем меньше нормы векторов ошибок измерений. Следовательно, есть все основания ожидать, что в ситуации, когда функция взаимной близости минимальна, соответствующая подсистема верхнего уровня наиболее свободна от шума.

Принципу согласованности оценок в случае подсистем нижнего уровня с квадратными $M \times M$ -матрицами можно дать простую геометрическую трактовку. В данном случае векторы оценок $\hat{c}_{i,i}$ являются нормальными векторами соответствующих M -мерных гиперплоскостей. Множеству K подсистем соответствует «облако» гиперплоскостей. В соответствии с принципом согласованности среди подсистем верхнего уровня предпочтение отдается той, для которой имеет место наибольшая взаимная близость гиперплоскостей в этом «облаке». После того как определена подсистема верхнего уровня, необходимо выбрать конкретную гиперплоскость либо построить их некоторую усредненную оценку.

Приведенная геометрическая трактовка использовалась Р. Калманом в работе [8]. В указанной работе он подверг резкой критике методы, опирающиеся на так называемую «стандартную статистическую априорную гипотезу» и показал, что возможность улучшения МНК-оценок связана с наличием переменных, свободных от шума. Используемый в настоящей работе подход является попыткой построить работоспособную процедуру для поиска таких переменных. Успех в данном случае зависит от того, насколько оправданы предположения о существовании наблюдений почти свободных от шума и насколько правильно подобраны размерности подсистем, которые включают эти наблюдения.

Поскольку никогда не известно число наблюдений почти свободных от шума, метод может предусматривать возможность перебора вариантов подсистем с различными размерностями. Такой алгоритм, основанный на критерии относительной согласованности оценок, рассмотрен в работе [9].

3.4. Алгоритм идентификации с отбором данных по показателям информативности и согласованности

Как показано выше составляющая Δc ошибок оценивания в (3.18), обусловленная вкладом составляющей вектора ошибок $\mathbf{T}_\lambda^T \xi \in R(\mathbf{X})$, не может быть обнаружена функцией взаимной близости $W(l)$. Эта составляющая обычно существенно возрастает именно на сигналах, регистрируемых на медленных (установившихся) режимах работы системы. Она может быть уменьшена, если предварительно отбирать лишь те данные, которые соответствуют информативным участкам сигналов. Это приводит к хорошей обусловленности системы (1.5) и в конечном итоге к снижению вычислительных ошибок [2].

Для обоснования процедур отбора данных построим оценку для квадрата евклидовой нормы вектора ошибок оценивания на конкретной реализации. По предположению вектор ξ принадлежит шару (5). Следовательно, вектор ошибок измерений ξ_i подсистемы также удовлетворяет этому условию, т.е.

$$\|\xi_i\|_2 \leq R_\xi.$$

Поэтому для оценки сверху прогнозируемой точности, достижимой на i -й подсистеме, можно воспользоваться разложением (3.34). В частности, заменив во всех слагаемых внутренней суммы в правой части (3.34) $\|\xi_i\|_2^2$ на R_ξ запишем неравенство:

$$\|\Delta c_i\|_2^2 \leq R_\xi \sum_{m=1}^M \lambda_{i,m}^{-1} \cos^2 \left(\mathbf{t}_{i,k} \hat{\xi}_i \right).$$

Из соотношения (3.36) видно, что N -мерный шар (5), которому принадлежит вектор ошибок измерений, преобразуется в M -мерный эллипсоид в пространстве $R(\mathbf{X})$. Этому эллипсоиду принадлежит вектор ошибок оценива-

ния. Его размеры определяются собственными значениями матрицы $\mathbf{X}_i \mathbf{X}_i^T$, которые как раз и характеризуют обусловленность системы.

В связи со сказанным, общую схему согласованной идентификации целесообразно строить в виде двухэтапной процедуры отбора (рисунок 3.2). На первом этапе осуществляется отбор данных, которые приводят к хорошей обусловленности задачи идентификации. Затем из этих данных формируется подсистема, для которой функция взаимной близости минимальна. Для предварительного отбора данных на первом этапе в соответствии с (3.36) можно использовать какую-либо характеристику, связанную с собственными значениями.

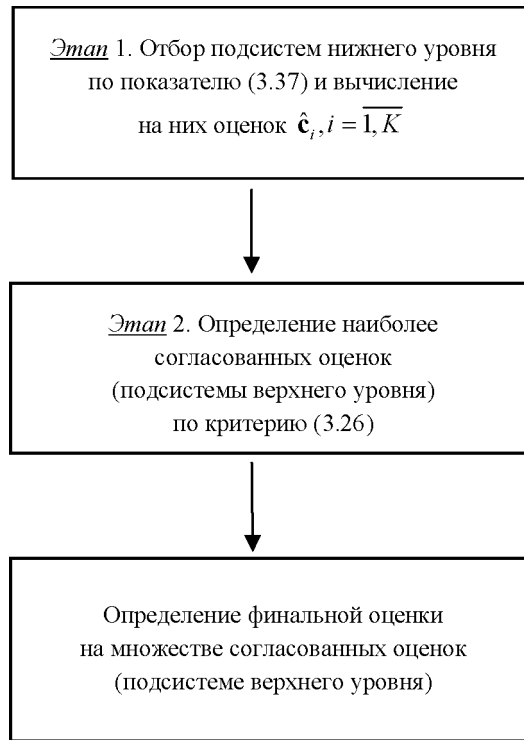


Рисунок 3.2. Схема отбора данных для идентификации

В частности, для построения двухэтапной процедуры отбора данных целесообразно использовать показатель диагонального преобладания:

$$\hat{\mathbf{O}}[\mathbf{X}\mathbf{X}^T] = \left(\sum_{i=1}^M \lambda_i \right)^2 / \sum_{i=1}^M \lambda_i^2 = \left(\sum_{i=1}^M a_{ii} \right)^2 / \sum_{i,j=1}^M a_{i,j}^2, \quad (3.37)$$

где λ_i - собственные значения, а $a_{i,j}^2$ - элементы матрицы $\mathbf{X}\mathbf{X}^T$. В данном случае показатель диагонального преобладания вычисляется по матрице $\mathbf{X}\mathbf{X}^T$, т.к. \mathbf{X} квадратная матрица. В данном случае также сохраняются ограничения (2.9) на допустимый диапазон:

$$M-1 \leq \hat{\mathbf{O}}[\mathbf{X}\mathbf{X}^T] \leq M.$$

Если матрица $\mathbf{X}\mathbf{X}^T$ плохо обусловлена, величина $\hat{\mathbf{O}}[\mathbf{X}\mathbf{X}^T]$ может принимать значения в интервале $[1, M-1]$, для которых гарантированные оценки не существуют. Тем не менее, и в этой ситуации показатель диагонального преобладания, вычисленный по данным, полученным на интервалах «возбуждения» идентифицируемого объекта, всегда выше, чем на установившихся режимах и является простым и надежным индикатором этих сигналов.

Рассмотрим более детально основные этапы отбора данных с использованием показателя диагонального преобладания (3.37) и функции взаимной парной близости (3.17) (рисунок 3.2).

Этап 1. Исходными данными для этого этапа являются измерения входа-выхода объекта (рис. 1.1). На этом этапе осуществляется отбор подсистем нижнего уровня по показателю (3.37) и вычисление на них оценок $\hat{\mathbf{c}}_i, i=1, 2, \dots$. Этот этап реализуется в темпе с процессами, протекающими в системе, в виде следующей последовательности шагов:

Шаг 1. Формирование по текущим измерениям входа - $u(k)$ и выхода - $y(k)$ очередной подсистемы нижнего уровня с квадратной $M \times M$ -матрицей \mathbf{X}_i .

Шаг 2. Вычисление для соответствующей матрицы $\mathbf{X}\mathbf{X}^T$ показателя диагонального преобладания (3.37) и сравнение полученного значения с допустимым значением - $\hat{\mathbf{O}}^*$. Если

$$\hat{\mathbf{O}}[\mathbf{X}_i\mathbf{X}_i^T] < \hat{\mathbf{O}}^*,$$

возвращение к шагу 1. Если

$$\hat{\mathbf{O}}[\mathbf{X}_i \mathbf{X}_i^T] \geq \hat{\mathbf{O}}^*,$$

вычисляется и сохраняется в памяти очередная оценка $\hat{\mathbf{C}}_i, i = 1, \dots$, а к счетчику числа оценок добавляется единица.

Шаг 3. Сравнение текущего значения счетчика числа оценок с заданным значением K . Если K не достигнуто – возвращение к шагу 1, иначе переход ко 2-му этапу.

Этап 2. Исходными данными для второго этапа являются: набор полученных на первом этапе K оценок $\hat{\mathbf{c}}_i, i = \overline{1, K}$ и заданное число k оценок, принадлежащих одной (любой) подсистеме верхнего уровня. На этом этапе для всех возможных сочетаний - \tilde{N}_K^k оценок $\hat{\mathbf{c}}_i, i = \overline{1, K}$ вычисляются функции парной взаимной близости (3.25). Предпочтение отдается тому набору оценок, для которого значение функции $W(I)$ минимально.

Этап 3. На заключительном этапе общей схемы на отобранном множестве согласованных оценок определяется искомая оценка. На этом этапе могут применяться критерии, отличающиеся от метода построения оценок на подсистемах нижнего уровня. В частности, можно в качестве искомой оценки принимать среднее вариационного ряда или среднее всех согласованных оценок, что равносильно вычислению МНК-оценки на подсистеме верхнего уровня.

В работе [36] приводится пример, который иллюстрирует эффективность двухэтапной процедуры идентификации, включающей предварительный отбор данных по показателю диагонального преобладания и последующее согласованное оценивание (рис. 3.2).

ЗАКЛЮЧЕНИЕ

Обычно идентификация управляемого объекта осуществляется с целью адаптации алгоритмов управления к изменяющимся параметрам модели объекта. Если скорость изменения параметров достаточно велика, идентификация должна осуществляться оперативно. Однако не для всякого интервала времени это возможно. Известно, что попытка решать задачу идентификации управляемого объекта по измерениям входа-выхода, полученным в установившемся режиме, приводит к плохой обусловленности или даже вырожденности уравнений, формируемых в процессе идентификации. Пригодными для надежной идентификации являются лишь небольшие участки переходных режимов. Однако при использовании данных, полученных на переходных режимах, возникают следующие проблемы. Во-первых, на этих интервалах возрастает вероятность помех, во-вторых, эти помехи маскируются полезными сигналами большой амплитуды, характерными для этих режимов.

В настоящем учебном пособии рассмотрены методы и алгоритмы, позволяющие эти две проблемы решать последовательно. В частности, рассмотрены методы контроля информативности данных, которые могут использоваться для отбора наблюдений, принадлежащих интервалам переходных процессов. Описаны критерии и процедуры контроля информативности данных, основанные на вычислении показателей обусловленности и мультиколлинеарности информационной матрицы, формируемой в процессе идентификации.

К сожалению, обычно наиболее информативные, с точки зрения свойств информационной матрицы, наблюдения оказываются наиболее зашумленными. Связано это с известным регуляризирующим свойством шумов. Построение надежных оценок по малому числу наблюдений, принадлежащих интервалам переходных процессов, является самостоятельной проблемой. В настоящем учебном пособии описывается нетрадиционный подход к решению этой задачи.

В данном случае, мы отказываемся от традиционных вероятностных моделей для построения алгоритмов идентификации. Предположения, которые мы используем, более реалистичны и менее обременительны, чем те которые обычно используются в теории статистического оценивания. Развиваемый подход не требует выполнения требования статистической устойчивости, которое все равно не выполняется при малом числе наблюдений, а потому статистические оценки оказываются ненадежными. Тот факт, что описанные

методы и алгоритмы показывают принципиальную возможность построения оценок с использованием критерия, не опирающегося на свойство статистической устойчивости, является, безусловно, важным, т.к. содержит не только традиционную критику методов оценивания по малой выборке (которая, обычно, не вызывает возражений), но и вселяет надежду на конструктивное преодоление известных проблем.

В этом отношении наибольший интерес представляют методы, основанные на принципе согласованности оценок, описанные в заключительном разделе. Точность оценок на подсистемах, выделенных с помощью этой функции, как правило, выше, а сама подсистема, как правило, оказывается наиболее свободной от шума. Конечно, для нахождения достаточно точного решения по критерию согласованности оценок может потребоваться перебор большого числа вариантов подсистем. Этот факт не является удивительным. Это неизбежная плата за недостаток априорной информации, связанный с недостаточностью числа наблюдений. Возможность применения простых (а следовательно относительно дешевых) статистических схем обработки, как правило, является следствием значительных затрат на проведение большого числа измерений, необходимого для выполнения закона больших чисел.

Конечно, высокая вычислительная сложность методов, основанных на принципе согласованности оценок, вызывает неудовлетворенность. Тем не менее, применение этого подхода оправдано, когда точное решение требуется получить, несмотря на то, что число информативных наблюдений мало, а проведение дополнительных измерений или ожидание «хороших» режимов для их проведения по каким-либо причинам невозможно. Заметим, что метод является подходящим именно для идентификации управляемого объекта. Связано это с тем, что для объекта, охваченного обратной связью в системе управления, число информативных измерений обычно крайне мало, а требования к точности и надежности оценок высокие.

СПИСОК СПЕЦИАЛЬНЫХ ТЕРМИНОВ

Евклидова норма вектора
Идентификация
Информационная матрица
Матрица Грама
Матрица сопряженности
Метод наименьших квадратов
Мультиколлинеарность
Нуль-пространство матрицы
Обусловленность системы уравнений
Показатель диагонального преобладания
Показатель сопряженности с нуль-пространством
Принцип согласованности
Пространство столбцов матрицы
След матрицы
Собственные числа и собственные векторы
Сопряженность векторов
Спектральное число обусловленности

СПИСОК ЛИТЕРАТУРЫ

Основная

1. **Беллман, Р.** Введение в теорию матриц / Р. Беллман. - М.: Наука, 1976
2. **Воеводин, В.В.** Вычислительные основы линейной алгебры / В.В. Воеводин. - М.: Наука, 1977
3. **Гантмахер, Ф.Р.** Теория матриц / Ф.Р. Гантмахер. - М.: Наука, 1967
4. **Гроп, П.** Методы идентификации систем / П. Гроп; пер. с англ.; под ред. Е.И. Кринецкого. - М.: Мир, 1979

Дополнительная

5. **Демиденко, Е.З.** Линейная и нелинейная регрессии / Е.З. Демиденко. - М.: Финансы и статистика, 1981
6. **Дуда, Р.** Распознавание образов и анализ сцен / Р. Дуда, П. Харт; пер. с англ. - М.: Мир, 1976
7. **Калман, Р.Е.** Идентификация систем с шумами / **Р.Е. Калман** / Успехи математических наук/ - Т. 40. - Вып. 4. - 1985. - № 244. - с. 27-41
8. **Лоусон, Ч.** Численное решение задач метода наименьших квадратов / Ч. Лоусон, Р. Хенсон; пер. с англ. - М.: Наука, Гл. ред. физ.-мат. лит., 1986
9. **Маркус, М.** Обзор по теории матриц и матричных неравенств / М. Маркус, Х. Минк; пер. с англ. - М.: Наука, 1972
10. **Мельканович, А.Ф.** Фотографические средства и их эксплуатация / А.Ф. Мельканович. - М.: Изд-во МО, 1984
11. **Мудров, В.И.** Методы обработки измерений / В.И. Мудров, В.Л. Кушко. - М.: Сов. радио. 1976
12. **Поляк, Б.Т.** Помехоустойчивая идентификация / Б.Т. Поляк, Я.З. Цыпкин // Идентификация и оценка параметров систем. Тр. IV Симпоз. ИФАК/ Ч. I. - Тбилиси, Мецниереба, 1976. - С. 190-213
13. **Прэтт, У.К.** Цифровая обработка изображений / У.К. Прэтт. - М.: Мир, 1982
14. **Сергеев, В.В.** Разработка и применение алгоритма цилиндрической полиномиальной аппроксимации изображения в скользящем окне / Сергеев В.В., Фролова Л.Г. // Автометрия; 1996. - № 1.
15. **Сергеев, В. В.** Оценивание пространственно-зависимых искажений с использованием параметризованных моделей / Сергеев В. В., Фурсов В.А., Максимов

- М.В. // Тр. III конф. "Распознавание образов и анализ изображений: новые информационные технологии" (РОАИ-97). Ч.I; Н.Новгород. 1997. – С. 252-255
16. **Сергеев, В.В.** Информационная технология оценки разрешающей способности видеотракта с использованием модели фильтра с бесконечной импульсной характеристикой / В.В. Сергеев, В.А. Фурсов. С.И. Парфенов // Тезисы докладов IV конференции «Распознавание образов и анализ изображений: новые информационные технологии». Ч.I. – Новосибирск. 1998. - С. 378-381
 17. **Сойфер, В.А.** Построение алгоритмов оперативной коррекции искажений на изображениях в оптико-электронных системах наведения и целеуказания / В.А. Сойфер, В.В. Котляр, В.А. Фурсов // Тр. конф. Волжского регионального центра РАРАН; Саров, 1998. - С. 108-109
 18. **Солодовников, В.В.** Принцип сложности в теории управления / В.В. Солодовников, В.Ф., Бирюков, В.И. Тумаркин. - М.: Наука, 1977)
 19. **Теряев, Е.Д.** Цифровые системы и поэтапное адаптивное управление / ТЛ/ **Теряев, Б.М. Шамриков.**- М.: Наука, 1999
 20. **Тихонов, А.Н.** Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. - М.: Наука, 1974
 21. **Уилкинсон, Дж.Х.** Алгебраическая проблема собственных значений / Дж. Уилкинсон; пер. с англ. - М.: Наука, 1970
 22. **Фурсов, В.А.** Анализ точности и построение алгоритмов идентификации по малому числу наблюдений / В.А. Фурсов // Изв. АН СССР, Техн. кибернетика. – 1991. - № 6. – С. 130-135
 23. **Фурсов, В.А.** Введение в идентификацию по малому числу наблюдений / В.А. Фурсов. – М.: Изд. МАИ, 1991
 24. **Фурсов, В.А.** Идентификация моделей систем формирования изображений по малому числу наблюдений / В.А. Фурсов. – Самара; Изд-во СГАУ, 1998
 25. **Фурсов, В.А.** Определение характеристик объектов в адаптивных системах управления: Учеб. Пособие / В.А. Фурсов. - М.: Изд-во МАИ, 1983
 26. **Фурсов, В.А.** Лекции по теории информации: учеб. пособие / В.А. Фурсов; под ред. Н.А. Кузнецова. Самара; Самар. гос. аэрокосм. ун-т, 2006. – 148 с.
 27. **Фурсов, В.А.** Проекционные методы оценивания параметров моделей прогнозирования по малому числу наблюдений / В.А. Фурсов, А.Г. Чудилин // Искусственный интеллект; Симферополь, 2000. - № 2. – С. 212-218
 28. **Цыпкин, Я.З.** Основы информационной теории идентификации / Я.З. Цыпкин. - М.: Наука, 1983
 29. **Шамриков, Б.М.** Оценивание достоверности параметрической идентификации объекта в адаптивных СУ / Б.М. Шамриков, В.А. Фурсов // Техн. кибернетика, 1979. - №6. – С. 173-180
 30. **Эйкхофф, П.** Основы идентификации систем управления / П. Эйкхофф; пер. с англ.; под ред. Н.С. Райбмана. - М.: Мир, 1975

31. **Bjorck, Ake** Least Squares Methods / Ake Bjorck; Elsevier Science Publishers B.V. North-Holland, 1990
32. **Fursov, Vladimir A.** Constructing unified identification algorithms using a small number of observations for adaptive control and navigation systems / Vladimir A. Fursov // Journal: Proc. SPIE, 1997. - Vol. 3087. - p. 34-44
33. **Fursov, V.A.** Identification of distorting systems with monitoring of data capacity / V.A. Fursov // 5-th International Workshop on Digital Image Processing and Computer Graphics. "Image Processing and Computer Optics"; Samara, Russia. – 1994. - Aug, 22-26. - P.2
34. **Fursov, Vladimir A.** Identification of optical distorting systems by selecting image informative fragments / Vladimir A. Fursov // Journal: Proc. SPIE, 1995. - Vol. 2363. - P. 62-68
35. **Sergeyev, Vladislav V.** Identification of model parameters and correcting filters for space-variant distortions / **Vladislav V. Sergeyev, Vladimir A. Fursov, & M. V. Maksimov** // Journal: Proc. SPIE, 1998. - Vol. 3348. - P. 275-282
36. **Fursov, Vladimir A.** Conforming Identification of the Controlled Object / Vladimir A. Fursov, Andrey V. Gavrilov // Proceeding International Conference on Computing, Communications and Control Technologies. - Austin, Texas, USA, 2004. - August 14-17. - P.326-330.

Учебное издание

Фурсов Владимир Алексеевич

**ИДЕНТИФИКАЦИЯ СИСТЕМ
ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ**

Учебное пособие

Технический редактор С. Н. Х о н и н а
Редакторская обработка А. А. Г н у т о в а
Корректорская обработка Е. С. К о ч е у л о в а
Доверстка Т. Е. П о л о в н е в а

Подписано в печать 07.11.07. Формат 60x84 1/16.
Бумага офсетная. Печать офсетная.
Печ. л. 5,0.
Тираж 120 экз. Заказ . ИП 92/2007

Самарский государственный
аэрокосмический университет.
443086 Самара, Московское шоссе, 34.

Изд-во Самарского государственного
аэрокосмического университета.
443086 Самара, Московское шоссе, 34.