В.А. Фурсов

ИДЕНТИФИКАЦИЯ МОДЕЛЕЙ СИСТЕМ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

Самара 1998

Министерство общего и профессионального образования Российской Федерации
Самарский государственный аэрокосмический университет
имени академика С.П. Королева
Институт систем обработки изображений Российской академии наук

В.А. Фурсов

ИДЕНТИФИКАЦИЯ МОДЕЛЕЙ СИСТЕМ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

Под редакцией
д.т.н., проф. В.А. Сойфера и
д.т.н., проф. Б.М. Шамрикова

Самара 1998

УДК 62.50

Фурсов В.А. Идентификация моделей систем формирования изображений по малому числу наблюдений/ Самар. гос. аэрокосм. ун-т. Самара, 1998, 218 с.

В монографии рассматриваются теория, методы и алгоритмы идентификации моделей систем формирования изображений по их малым фрагментам. Обосновываются меры и строятся процедуры контроля информативности данных. Описываются опирающиеся на алгебраический подход методы анализа точности и алгоритмы идентификации, не требующие задания априорных вероятностных распределений помех. Приводятся примеры созданных на основе этих методов и алгоритмов информационных технологий обработки и анализа изображений.

Основные научные результаты, изложенные в настоящей монографии, получены при выполнении проектов Федеральной целевой программы "Государственная поддержка интеграции высшего образования и фундаментальной науки на 1997-2000 годы" и Федеральной целевой научно-технической программы "Исследования по приоритетным направлениям науки и техники гражданского применения" (подпрограмма "Наукоемкие технологии") при финансовой поддержке Министерства науки и технологий РФ и Центра содействия интеграции высшего образования и фундаментальной науки (Центр "Интеграция"), а также Самарского научно-координационного центра "Перспектива".

Книга предназначена для инженерно-технических и научных работников, специализирующихся в области теории идентификации, цифровой обработки и анализа изображений.

Табл. 34. Ил. 56. Библиогр. 66 наименов.

Рецензенты: д-р физ.-мат. наук, проф. А.И. Жданов
д-р техн. наук, проф. П.К. Кузнецов

ISBN 5-7883-0054-1

© В.А. Фурсов, 1998
© Самарский государственный аэрокосмический университет, 1998

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....5
1. ОБЩАЯ СХЕМА И ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ НА ФРАГМЕНТАХ ИЗОБРАЖЕНИЙ.....9
1.1. Этапы подготовки данных для решения задачи идентификации.....9
1.2. Модели и предположения.....13
1.3. Общая схема решения задачи.....19
1.4. Формы представления ошибок.....23
1.5. Ортогональные разложения ошибок оценивания.....27
1.6. Достаточные оценки точности.....30
2. ОЦЕНКА ИНФОРМАТИВНОСТИ ДАННЫХ.....41
2.1. Понятие информативности данных, связь с обусловленностью задачи.....41
2.2. Оценки для собственных значений матрицы Грама.....46
2.3. Оценки для числа обусловленности и определителя.....53
2.4. Оценка обусловленности и мультиколлинеарности по диагональному преобладанию матрицы Грама.....60
2.5. Оценка мультиколлинеарности по показателям сопряженности с нуль-пространством матрицы измерений.....67
2.6. Связь мер обусловленности и мультиколлинеарности.....71
3. ДОСТИЖИМАЯ ТОЧНОСТЬ ПРИ ПРЕОБРАЗОВАНИЯХ ДАННЫХ.....78
3.1. Качественный анализ ошибок идентификации при преобразованиях взвешивания.....78
3.2. Связь реализаций векторов ошибок и невязок.....87
3.3. Качественный анализ ошибок оценивания при непосредственной корректировке зависимых переменных.....94
3.4. Влияние масштабирования независимых переменных.....100
4. МЕТОДЫ ИДЕНТИФИКАЦИИ С ПРЕОБРАЗОВАНИЕМ ДАННЫХ.....104
4.1. Методы идентификации, основанные на непосредственной корректировке вектора выхода модели.....104
4.2. Омтимизационный метод построения весовой матрицы.....114
4.3. Статистическое обоснование метода взвешивания данных.....118
4.4. Построение преобразований взвешивания путем формального задания параметрических классов распределений.....122
4.5. Различение классов по значениям среднего модулей невязок.....128

4.6. Связь с L_v-оценками 134
5. РЕШАЮЩИЕ ПРАВИЛА И АЛГОРИТМЫ 140
5.1. Общая схема алгоритма идентификации 140
5.2. Процедура масштабирования независимых переменных 144
5.3. Решающие правила и алгоритмы анализа информативности 145
5.4. Решающие правила и алгоритмы отбора данных 151
5.5. Организация итеративного процесса преобразования данных и вычисления оценок 155
5.6. Сравнение с рекуррентным взвешиванием данных 159
6. РЕАЛИЗАЦИЯ МЕТОДОВ И АЛГОРИТМОВ ИДЕНТИФИКАЦИИ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ 163
6.1. Идентификация импульсной характеристики искажающей системы 163
6.2. Построение восстанавливающих фильтров путем непосредственной идентификации инверсной модели видеотракта 175
6.3. Идентификация кусочно-постоянных линейных моделей пространственно-зависимых искажений и восстанавливающих фильтров 183
6.4. Идентификация моделей систем формирования изображений в классе фильтров с бесконечной импульсной характеристикой 187
6.5. Оценки разрешающей способности видеотракта 194
6.6. Информационная технология совмещения изображений по информативным фрагментам 203
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 212

ВВЕДЕНИЕ

Любое научное исследование включает этапы наблюдения, обработки данных и установления основных закономерностей в изучаемой предметной области. Во многих областях знаний (космонавтика, геофизика, астрономия, медицина и др.) для исследования новых явлений широко используются изображения объектов [8]. Оптико-электронные системы формирования изображений неизбежно вносят искажения. Для построения процедур коррекции этих искажений необходимо знать модель искажающей системы.

Задача определения модели искажающей (или восстанавливающей) системы обычно формулируется как задача параметрической идентификации. Теории и методам идентификации систем посвящена обширная литература. Большой вклад в развитие теории идентификации внесли отечественные (Я.З. Цыпкин, Н.С. Райбман, И.И. Перельман, А.А. Ершов, Б.Т. Поляк) и зарубежные (Р.Е. Калман, Р. Ли, Э.П. Сейдж, Дж. Л. Мелса, Д.Гроп, П. Эйкхофф, Л. Люнг) ученые.

При проведении научных исследований и комплексных испытаниях образцов новой техники часто возникает необходимость идентификации модели искажающей системы по весьма малому числу наблюдений на фрагменте изображения [5]. Примеры таких задач: определение локально-линейной инвариантной к сдвигу модели неизопланатичных искажений, оценка разрешающей способности видеотракта по деталям известных объектов на искаженном изображении, оперативная идентификация с целью коррекции изображений в оптико-электронных системах дистанционного наблюдения, навигации и целеуказания.

При решении задач идентификации по малому числу наблюдений на фрагментах изображений возникает, по крайней мере, две проблемы:

1. Информативность исходных данных существенно зависит от выбора фрагмента на изображении, в частности, при попадании на участок фона задача оказывается плохо обусловленной, что может привести к большим ошибкам в оценках параметров.
2. Свойство устойчивости статистических характеристик шумов при малом числе наблюдений на фрагменте изображения не проявляется, поэтому методы идентификации, основанные на использовании априорных вероятностных гипотез относительно помех, оказываются теоретически необоснованными и, как следствие, неработоспособными.

Эти две проблемы возникают во многих приложениях, связанных с построением адаптивных систем, в которых перестройка осуществляется по результатам эпизодической идентификации объекта в режиме его нормального функционирования [3,4,28,60,64]. Тем не менее, анализ работ по теории и прикладным аспектам определения характеристик систем [12,24,56] показывает, что проблема идентификации по малому числу наблюдений остается актуальной.

Анализу свойств оценок с учетом изменения характеристик информативности полезных сигналов в теории оценивания не уделяется достаточного внимания. Обычно предполагается, что модель полезного сигнала задана, либо удовлетворяет определенным требованиям. Если по каким-либо причинам эти требования сильно снижены, констатируется возможность плохой обусловленности задачи. Сама же проблема обусловленности традиционно считается вычислительной, что нашло отражение в разработке алгебраической теории возмущений [9,39] и теории решения некорректных задач [18,38]. Выявление факта плохой обусловленности, обычно, связано с решением полной или частной проблемы собственных значений. Сводку многих известных оценок для собственных значений можно найти в работах [2,10,11,17, 23,25,59].

Причины, по которым эти результаты не всегда включаются в руководства по теории оценивания, заключаются не только в том, что теория возмущений и теория некорректных задач имеют важное самостоятельное значение. Дело в том, что характеристики обусловленности для фиксированного класса полезных сигналов во многих случаях действительно можно считать заданными [6]. Например, при использовании изображений какого-либо определенного типа (портрет, пейзаж и т.д.) корреляционные свойства поля яркости, рассчитанные по всему изображению, слабо изменяются при переходе к другому изображению из этого же класса.

При решении задачи идентификации на малых фрагментах тех же изображений ситуация принципиально иная. Свойства полезных сигналов 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. Однако построение соответствующих методов в рамках статистической теории оценивания не представляется возможным.

Дело в том, что при малом числе наблюдений основное условие предельных теорем теории вероятностей (существование большого числа случайных явлений) не выполняется [7]. Поскольку эти теоремы являются фундаментом теоретической статистики, основанные на ее результатах методы идентификации в теоретическом отношении оказываются также недостаточно обоснованными [31].

По той же причине необоснованным следует считать использование теории робастного оценивания [37,50]. Робастные оценки, методы и алгоритмы являются частным случаем оптимальных на классе оценок. В соответствии с принципом оптимальности на классе [29,51] строится оптимальная функция потерь для наименее благоприятной плотности распределения из заданного класса. Тем самым предполагается существование удовлетворяющих условиям центральных предельных теорем теории вероятностей классов распределений. Вдобавок, при построении оптимального решения, минимизирующего эмпирические средние потери, сходимость к средним потерям имеет место лишь при выполнении условий усиленного закона больших чисел [51].

В качестве альтернативы методам, основанным на фишеровской теории оценивания, пропагандируется [1] мизесовский подход, в котором всерьез принимаются только статистические вероятности. По-видимому, такая точка зрения в некоторых случаях предметно оправдана. Однако выявление статистических вероятностей объективно связано с увеличением числа наблюдений, что для рассматриваемого круга задач неприемлемо.

Резкой критике подходы, связанные с заданием так называемой "стандартной статистической гипотезы", подверг Р. Калман. В работе [19] он рассмотрел несколько известных методов идентификации и выявил предположения, которые основаны на предубеждениях. В частности, показал, что "хорошие" свойства МНК могут быть связаны либо с предположением о наличии переменных свободных от шума, либо с вложением МНК в некоторые вероятностные модели, которые в данном случае, к сожалению, не могут быть обоснованы.

Таким образом, с одной стороны, вероятностные модели не могут служить надежной основой для построения методов и алгоритмов оценивания по малому числу наблюдений [1,19,57,43], с другой стороны, известные в алгебраической теории возмущений методы оценки обусловленности [10,39,52] не учитывают особенности задач оценивания.

В настоящей работе обе указанные проблемы рассматриваются в рамках алгебраической теории возмущений. При этом используется постановка задачи не

требующая задания априорных вероятностных распределений. В частности, относительно вектора ошибок измерений предполагается лишь то, что его норма не превышает некоторой величины. Такая постановка, известная в теории некорректных задач как принцип невязок [38], в какой-то степени, вынужденная. Однако, поскольку вероятностные модели все равно не известны, она более реалистична и, возможно поэтому, широко используется в теории возмущений.

В монографии излагаются опирающиеся на алгебраический подход методы анализа точностных характеристик. Приводятся меры информативности данных и построенные с их использованием алгоритмы отбора фрагментов изображений. Описываются методы идентификации, не требующие задания априорных вероятностных распределений помех. Приводятся результаты исследования проявляющейся и при малом числе наблюдений устойчивой связи выделяющихся компонент векторов ошибок и невязок. Показана возможность использования этой информации для повышения точности оценок.

Описываются построенные на основе разработанных методов новые информационные технологии: оценка параметров моделей искажающих систем и восстанавливающих фильтров пространственно-зависимых искажений на фрагментах изображений, оценка разрешающей способности видеотракта, совмещение изображений.

Основные научные результаты, изложенные в настоящей монографии, получены при выполнении проектов Федеральной целевой программы "Государственная поддержка интеграции высшего образования и фундаментальной науки на 1997-2000 годы" и Федеральной целевой научно-технической программы "Исследования по приоритетным направлениям науки и техники гражданского применения" (подпрограмма "Наукоемкие технологии") при финансовой поддержке Министерства науки и технологий РФ и Центра содействия интеграции высшего образования и фундаментальной науки (Центр "Интеграция"), а также Самарского научно-координационного центра "Перспектива".

Автор выражает признательность д. ф.-м. н., профессору А.И. Жданову и д. т. н., профессору В.В. Сергееву внимательно прочитавшим рукопись и сделавшим ряд весьма ценных замечаний, способствовавших ее улучшению, а также сотрудникам лаборатории "Анализ оптических сигналов" ИСОИ РАН Устинову А.В., Парфенову С.И. и Максимову М.В., подготовившим тестовые изображения и принявшим участие в проведении экспериментов.

1. ОБЩАЯ СХЕМА И ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ НА ФРАГМЕНТАХ ИЗОБРАЖЕНИЙ

В настоящей главе приводятся схемы формирования ошибок исходных данных для двух типичных моделей линейных искажений в системах формирования изображений. Обсуждаются специфические особенности задачи идентификации этих моделей по малым фрагментам изображений и обосновывается общая схема решения задачи в виде последовательности МНК-оценок с промежуточными преобразованиями данных. Рассматриваются формы представления ошибок и их ортогональные разложения. Строятся достаточные оценки сверху и снизу для евклидовой нормы вектора ошибок идентификации. Формулируются частные задачи исследований.

1.1. Этапы подготовки данных для решения задачи идентификации

Для решения многих задач формирования изображений (коррекция, выбор режимов съемки и др.) необходимо знание искажающего оператора системы, преобразующего измерительную видеоинформацию [6,30,58]. Этот оператор рассчитывают на этапе проектирования системы с использованием физических и математических моделей сквозного тракта. Вследствие неизбежных упрощений, связанных как с недостаточной изученностью явлений, так и с вычислительными аспектами, оператор спроектированной и изготовленной системы нуждается в уточнении. Для этого проводят испытания, заключающиеся в подаче на вход системы известного сигнала и его регистрации на выходе. По результатам наблюдений над входными и выходными переменными системы строится оптимальная в некотором смысле модель, т.е. формализованное представление этой системы. Это направление известно как идентификация систем.

В настоящей работе рассматривается задача идентификации в узком смысле [56]. Она состоит в оценивании параметров модели системы формирования изображений из априори заданного класса. Наиболее широко используются линейные модели с постоянными (по пространственным координатам) параметрами. Вычислительные преимущества, связанные с применением таких моделей, очевидны. Если в действительности искажения оказываются пространственно-зависимыми, линейные модели с постоянными параметрами можно применять на малых фрагментах изображений. Это дает возможность использовать унифицированные процедуры обработки данных для линейных искажений, включая неизопланатичные.

Для решения задачи идентификации линейного оператора, в частности импульсной характеристики, искажающей системы необходимо иметь два сигнала: известный входной (неискаженный) и зарегистрированный на выходе системы (искаженный) сигнал. В качестве таких сигналов в нашем случае должны фигурировать изображения. В общем случае связь отсчетов входного (неискаженного) $x(n_1, n_2)$ и выходного (искаженного) $y(n_1, n_2)$ изображений описывается соотношением [14]

$$y(n_1, n_2) = \sum_{\{l_1, l_2\} \in D_x} a(l_1, l_2) x(n_1 - l_1, n_2 - l_2) -$$

$$- \sum_{\{k_1, k_2\} \in D_y} b(k_1, k_2) y(n_1 - k_1, n_2 - k_2) + \xi(n_1, n_2), \qquad (1.1)$$

где $h(k_1, k_2)$ - значения импульсной характеристики $x(n_1 - l_1, n_2 - l_2)$, $y(n_1 - k_1, n_2 - k_2)$ отсчеты поля яркости в точках опорных областей D_x и D_y входного и выходного изображений соответственно, а $\xi(n_1, n_2)$- дискретная последовательность, включающая погрешности измерений и ошибки аппроксимации. Модель, описываемая соотношением (1.1), является фильтром с бесконечной импульсной характеристикой.

Задача идентификации этой модели заключается в оценке неизвестных коэффициентов $a(l_1, l_2)$, $b(k_1, k_2)$ по измерениям $x(n_1, n_2)$ и $y(n_1, n_2)$. В следующих разделах мы рассмотрим также один важный частный случай этой модели. Здесь мы, пока опираясь лишь на соотношение (1.1), обсудим общие вопросы организации подготовки данных и связанные с этим различия в схеме этапов решения задачи идентификации.

В традиционной схеме оценки характеристик модели сквозного видеоинформационного тракта на поверхности Земли создают мишенный комплекс, состоящий из тестовых объектов - мир [26], который затем регистрируется исследуемой системой. При этом на этапе оценивания могут использоваться либо выполненный на некотором носителе рисунок, либо исчерпывающее описание миры (аналитическое или в терминах машинной графики). Если задача заключается в оценке параметров модели линейных искажений, исходное (неискаженное) изображение миры должно быть точно совмещено с изображением, зарегистрированным исследуемой системой, а уровни яркости (контрасты) отдельных фрагментов установлены с учетом требуемых эталонных условий наблюдения.

Оценивание характеристик видеотракта, даже в указанном "классическом" случае, сопряжено со значительными трудностями. Во-первых, достаточно сложной является задача совмещения входного и выходного изображений. Даже если предположить, что регистрация входного изображения осуществляется некоторой "образцовой" системой, имеющей существенно более высокое разрешение, выходное изображение, сформированное идентифицируемой системой, кроме прочего, может отличаться параметрами сдвига, поворота, масштаба. Характер этих различий может изменяться по полю изображения, что является дополнительным источником погрешностей оценивания параметров модели системы. Вклад этой составляющей погрешности, по-видимому, будет тем меньше, чем меньше размеры фрагментов, используемых для идентификации. Если искажения существенно неизопланатичны, задача оценивания должна решаться отдельно для каждого фрагмента по малому числу наблюдений.

Входное изображение не обязательно регистрировать системой с более высоким разрешением. Для его воспроизведения можно использовать описание эталона миры в виде указания размеров элементов и значений отсчетов функции яркости, соответствующих стандартным условиям наблюдения. При этом наряду с погрешностями совмещения вносятся также ошибки, связанные с неточностью поддержания заданных условий освещенности миры в момент регистрации выходного изображения.

Часто задача ставится еще более "жестко". Требуется оценить характеристики модели искажений и/или разрешающей способности видеотракта без использования специальной миры по фрагментам известных объектов на исследуемом изображении. В качестве таких фрагментов должны использоваться элементы объектов, изображения которых слабо зависят от времени суток и сезонных изменений в природе. Желательно также, чтобы эти фрагменты имели участки с заметным контрастом функции яркости. Подходящими для такой цели являются, например, элементы дорожной и аэродромной разметки, фрагменты строительных конструкций, крыши промышленных зданий и сооружений на границе с отбрасываемой ими тенью и др.

Использование этих элементов изображений для идентификации основывается на возможности воспроизведения по априорным описаниям функции яркости, соответствующей их неискаженным изображениям. При этом отпадает необходимость решения задачи совмещения одноименных фрагментов. Используемые фрагменты должны содержать информативные, с точки зрения задачи

оценивания, элементы. Ясно, что количество и размеры фрагментов, для которых имеются априорные описания, и удовлетворяющих всем указанным выше требованиям, невелики. Например, при разрешении на местности около одного метра число отсчетов, фиксирующих фрагменты плоской крыши большинства зданий, не превышает нескольких десятков.

Таким образом, для идентификации модели искажений видеотракта мы, часто, вынуждены использовать фрагменты изображений небольших размеров с числом отсчетов, не превышающим нескольких десятков. При этом, в зависимости от конкретных условий проведения испытаний, возможны, по крайней мере, два варианта схемы этапов подготовки данных для идентификации, показанные на рис. 1.1.

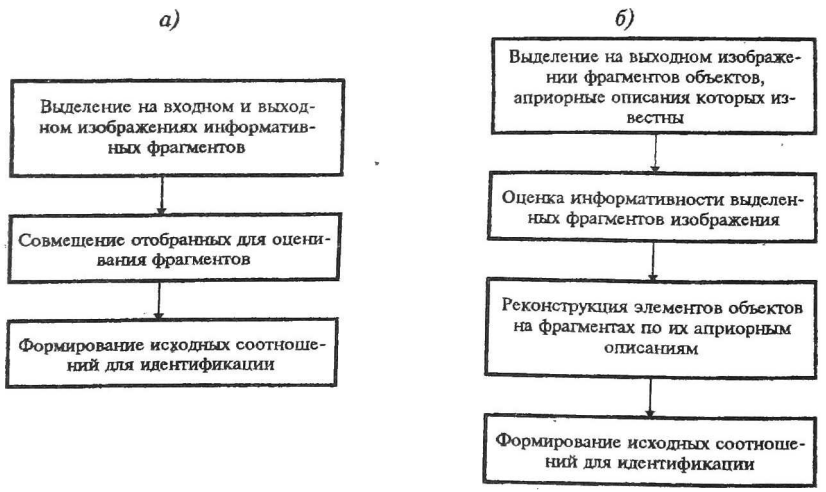


Рис. 1.1. Этапы подготовки данных для идентификации в случаях: а - входное изображение регистрируется; б - входное изображение конструируется по описанию

Эти схемы существенно различны. В схеме (рис. 1.1, а) одним из важнейших этапов является совмещение фрагментов изображений. Задача совмещения фрагментов является в значительной степени самостоятельной задачей. Для ее решения могут использоваться известные методы. Один из вариантов информационной технологии, построенной на основе результатов настоящей работы, будет описан в разделе 6.6.

В схеме, показанной на рис. 1.1, б, отсутствует этап совмещения фрагментов изображений. Однако в данном случае должна решаться сложная задача воспроизведения (реконструкции) какого-либо фрагмента изображения объекта по его

описанию. Реконструкция элементов объектов заключается в устранении искажений на выбранных фрагментах. Для успешной реализации этого этапа необходимо привлекать информацию не только о форме и деталях регистрируемого объекта, но и об условиях его регистрации (направление и относительная скорость движения регистрирующей аппаратуры, разрешение и др.).

Обе схемы завершаются этапом формирования исходных соотношений. Для этого должна быть задана структура модели. В следующем разделе мы рассмотрим два наиболее важных класса моделей, используемых для описания линейных искажений в системах формирования изображений, а также обсудим проблемы, возникающие при оценке параметров этих моделей по малому числу наблюдений.

1.2. Модели и предположения

Рассмотрев варианты подготовки данных для идентификации мы убедились, что определение оценок параметров моделей систем регистрации изображений по малым фрагментам является скорее правилом, чем исключением. Такая необходимость возникает как вследствие того, что размеры элементов объектов, априорные описания которых известны, чаще всего, действительно невелики, так и из соображений уменьшения погрешностей, связанных с неизопланатичностью линейных искажений, относительным сдвигом и поворотом фрагментов.

Идентификация моделей регистрирующих систем по малым фрагментам изображений, в свою очередь, порождает ряд принципиальных проблем. В настоящем разделе мы вначале приведем стандартную постановку задачи оценивания. Затем, обсудив особенности задачи идентификации по малому числу наблюдений, откажемся от традиционных предположений и сформулируем более реалистичные требования.

Широко известна следующая постановка задачи [15]. Построить оценку $\hat{\mathbf{c}}$ вектора параметров $\mathbf{c}$ по доступным для непосредственного наблюдения или сформированным каким-либо способом $N\times M$ матрице $\mathbf{X}$ и $N\times 1$ - вектору $\mathbf{y}$ ($N>M$) уравнения типа линейной регрессии:

$$\mathbf{Xc} = \mathbf{y} + \xi, \qquad (1.2)$$

где ξ - $N\times 1$-вектор ошибок. Источниками ошибок являются погрешности измерений, ограничения на порядок модели и др.

Соотношение вида (1.2) формально легко выписывается для приведенной выше в виде соотношения (1.1) модели фильтра с бесконечной импульсной ха-

рактеристикой (БИХ-фильтра). Предположим, что изображения дискретизованы и представляются в виде совокупности отсчетов значений яркости в узлах заданной сетки (рис. 1.2 и 1.3). Для определенности положим, что размеры опорных областей D_x, D_y (на рис. 1.2 показаны "жирными" точками) соответственно $K_1 \times K_2$ и $L_1 \times L_2$. С учетом принятых обозначений уравнение (1.1) можно переписать в виде [14]:

$$y(n_1, n_2) = \sum_{l_1=0}^{L_1-1} \sum_{l_2=0}^{L_2-1} a(l_1, l_2) x(n_1 - l_1, n_2 - l_2) - \\ - \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} b(k_1, k_2) y(n_1 - k_1, n_2 - k_2) + \xi(n_1, n_2), \quad (k_1, k_2) \neq (0,0) \tag{1.3}$$

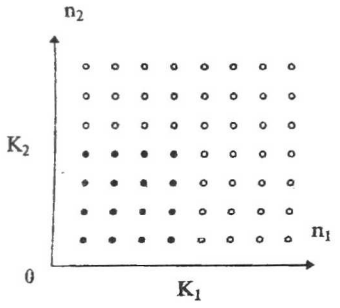


Рис. 1.2. Опорная область D_x на входном изображении

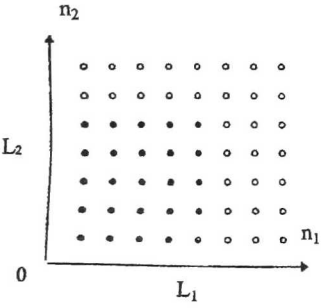


Рис. 1.3. Опорная область D_y на выходном изображении

Введем обозначения $L=L_1 \times L_2$ и $K=(K_1 \times K_2)-1$, а $M=L+K$. Если фрагмент (или изображение) содержит N различных отсчетов $y(n_1, n_2)$, из них можно составить фигурирующий в (1.2) вектор **y** размерности $N \times 1$. Соответствующая матрица **X**, сформированная на том же фрагменте, имеет размерность $N \times M$. Каждая строка этой матрицы состоит из M отсчетов одной опорной области:

$$\mathbf{x}_i = [x_1, x_2, \ldots, x_M] = [x_1, \ldots, x_L, y_1, \ldots, y_K], \quad i = \overline{1, N}.$$

Соответствующий вектор искомых параметров имеет вид

$$\mathbf{c} = [a_1, \ldots, a_L, \ -b_1, \ldots, -b_K]^T = [c_1, c_2, \ \ldots, c_M]^T,$$

а дискретная последовательность "приведенных к выходу" ошибок $\xi(n_1, n_2)$ определяется как

$$\xi(n_1, n_2) = \delta y(n_1, n_2) - \sum_{k_1=0}^{K_1-1} \sum_{k_2=0}^{K_2-1} b(k_1, k_2) \delta y(n_1 - k_1, n_2 - k_2), \tag{1.4}$$

где $\delta y(n_1,n_2)$ ошибки измерений, ограничений на порядок модели и др. Вектор ξ размерности $N\times 1$ составляется из ошибок (1.4) аналогично вектору **y**. Порядок обхода точек фрагмента безразличен. Имеет значение лишь соответствие компонент векторов ξ, **y** и строк x_i^1 матрицы **X**.

Частным случаем модели (1.3) является фильтр с конечной импульсной характеристикой (КИХ-фильтр). Положим, что опорная область D_x на входном изображении прямоугольник размером $K_1\times K_2$ и в ней содержится $M=K_1\times K_2$ различных отсчетов. Тогда связь отсчетов входного $x(n_1,n_2)$ (неискаженного) и выходного $y(n_1,n_2)$ (искаженного) для модели КИХ-фильтра запишется в виде

$$y(n_1,n_2)=\sum_{l_1=0}^{L_1-1}\sum_{l_2=0}^{L_2-1}h(l_1,l_2)x(n_1-l_1,n_2-l_2)+\delta y(n_1,n_2),\quad (k_1,k_2)\neq(0,0),\ (1.5)$$

где $h(l_1,l_2)$ значения импульсной характеристики в точках опорной области D_x, а $\delta(n_1,n_2)$ - ошибки измерений, аппроксимации и др.

Сопоставив (1.5) с (1.2), нетрудно заметить, что вектор искомых параметров **c** в данном случае представляется в виде

$$\mathbf{c}=\left[h_1,h_2,\ldots,h_M\right]^T,$$

а каждая строка матрицы **X** состоит из отсчетов только входного изображения:

$$\mathbf{x}_i=\left[x_{i,1},x_{i,2},\ldots,x_{i,M}\right],\quad i=\overline{1,N}.$$

Еще одно важное отличие модели КИХ-фильтра от (1.3) состоит в следующем. Компоненты $N\times 1$-вектора ошибок ξ не зависят от полезных сигналов (отсчетов поля яркости выходного изображения).

Если искажения являются изопланатичными (пространственно-независимыми), компоненты вектора параметров **c** остаются неизменными при любом сдвиге опорной области на изображении. Если же искажения неизопланатичны, то изображение можно разбить на малые фрагменты, для которых это будет приближенно выполняться. Далее предполагается, что это условие всегда выполнено и искомый вектор **c** в рассматриваемой области изображения, по отсчетам которой формируется система уравнений (1.2), в пределах заданной точности, является постоянным.

Из соотношений (1.3) - (1.5) видно, что задача идентификации модели БИХ (и/или КИХ)-фильтра формально может быть сведена к оценке векторного параметра **c** уравнения типа линейной регрессии (1.2). Классический регрессионный анализ базируется на следующих предположениях [15]:

1. Вектор $\xi = \left[\xi_1, \xi_2, \ldots, \xi_N\right]^T$ случайный. Откуда следует, что $y = \left[y_1, y_2, \ldots, y_N\right]^T$ также случайный вектор.

2. Матрица **X** детерминирована, т. е. x_{ij}, $i = \overline{1,N}$, $j = \overline{1,M}$ не являются случайными величинами.

3. Rank(**X**)=M.

4. Математическое ожидание ξ_i равно нулю, т.е. $M\{\xi_i\}=0$, i=1,N или $M\{\xi\}=0$.

5. Для любых $i \neq j$ $M\left\{\xi_i \cdot \xi_j\right\} = 0$, $M\left\{\xi_i^2\right\} = \sigma^2$ для всех i=1,N. Другими словами, $cov\{\xi\}=\sigma^2\mathbf{E}_N$, где σ^2- дисперсия отклонений, $cov\{\cdot\}$ N×N - матрица ковариаций отклонений, а $\mathbf{E}_N$ - единичная N×N - матрица.

К сожалению, указанные предположения не отвечают реальному содержанию рассматриваемой задачи. При малом числе наблюдений вектор ξ не является случайным в строгом смысле определения случайной величины [7,21]. Поэтому предположения 1, 4 и 5 в данном случае не могут считаться правомерными. Даже если эти предположения справедливы для шумовой последовательности на всем изображении, оценки математического ожидания и дисперсии на малых фрагментах этого изображения могут сильно отличаться от свойств, декларированных в указанных предположениях. При малом числе наблюдений вообще неуместно говорить о математическом ожидании и дисперсии.

Требование *3* является естественным, но его выполнение, часто, вызывает сомнение. Дело в том, что при идентификации моделей искажающих систем по малым фрагментам изображений всегда сохраняется опасность попадания на неинформативный участок (фона). Поэтому условие Rank(**X**)=M должно проверяться перед каждым эпизодом идентификации. Строго говоря, вместо предположения *3* в данном случае более реалистичным является требование Rank(**X**)≤ M.

Таким образом, мы вынуждены отказаться от всех наиболее важных, с теоретической точки зрения, предположений классической регрессии. Этот отказ является вынужденным, но оправданным. Оценки параметров в данном случае должны строиться по малым фрагментам изображений. Поэтому понятно, что свойство устойчивости статистических характеристик шумов и информативных свойств полезных сигналов не проявляется. Это существенным образом отличает рассматриваемый класс задач от задач классической регрессии.

С учетом сказанного будем избегать привычной терминологии, сложившейся в теории статистического оценивания. Во первых, термин "оценивание" мы будем употреблять, как правило, не связывая его с прилагательным "статистическое". Наряду с ним мы будем использовать, на наш взгляд более подходящий в данном случае, термин "определение". Кроме того, мы будем избегать широко используемого термина "выборка", употребляя вместо него термины "набор данных" и "малое число наблюдений".

Необходимо еще раз подчеркнуть, что в настоящей работе изучаются задачи оценивания, в которых число наблюдений превышает число оцениваемых параметров лишь в несколько раз. Конечно, свойство статистической устойчивости может не проявляться и при существенно большем числе наблюдений. Однако при этом многие разрабатываемые ниже процедуры (основанные на том, что при малом числе наблюдений размерность нуль-пространства матрицы измерений сопоставима с размерностью пространства параметров) потеряют свои преимущества. Вместе с тем, некоторые излагаемые ниже теоретические результаты, касающиеся анализа ошибок, могут использоваться и при большом числе наблюдений, если статистические методы исследования оказываются непригодными.

В контексте настоящей работы термин "малое число наблюдений" используется в буквальном смысле. Он служит для обозначения круга задач, в которых число доступных наблюдений не превышает нескольких десятков. Тем не менее, мы будем избегать указаний на конкретное число наблюдений, которое может считаться малым. Это зависит от свойств конкретного сигнала. Если не принимать во внимание вычислительные аспекты реализации, то важнейшим признаком, отличающим задачи оценивания по малому числу наблюдений от задач классической регрессии, является факт отсутствия устойчивости информативных свойств и статистических характеристик. Другими словами, имеют место следующие специфические особенности задачи идентификации по малым фрагментам изображений.

1. Всегда существует опасность того, что из-за попадания фрагмента изображения на участок фона задача идентификации окажется плохо обусловленной или даже вырожденной.
2. Условия предельных теорем теорий вероятностей при малом числе наблюдений не выполняются и, как следствие, невозможно обосновать априорную вероятностную модель ошибок в исходных данных. Свойство устойчивости статистических характеристик шумов на фрагментах изображений может не

проявляться даже в том случае, когда существует устойчивое распределение ошибок на множестве изображений данного класса.

С учетом сказанного сформулируем предположения, которые далее будут использоваться при решении задачи оценивания по малому числу наблюдений на фрагменте изображения.

Предположение 1. Матрица X и вектор y фиксированы, т. е. x_{ij}, y_i, $i = \overline{1,N}$, $j = \overline{1,M}$ известны в результате измерений.

Предположение 2. Априорная информация об информативности этих измерений считается отсутствующей. Спектральные свойства полезных сигналов на фрагменте могут существенно отличаться от априорно заданных для класса или конкретного изображения из заданного класса.

Предположение 3. Статистические характеристики векторов ξ и y на фрагменте изображения считаются неизвестными, даже если они известны для класса или одного изображения, из которого выделен фрагмент.

Предположение 4. Задано ограничение на норму вектора ошибок $\xi = [\xi_1, \xi_2, \ldots, \xi_N]^T : \|\xi\| \le R$, а вся неопределенность связана с произвольной ориентацией этого вектора в N-мерном пространстве.

В дополнение к указанным может использоваться также предположение об ограничениях параметров модели в виде неравенств. Эти ограничения задаются на основе априорной информации о допустимых диапазонах оцениваемых характеристик. Аналогичные ограничения используются в традиционной постановке задачи оценивания параметров линейной регрессии [15] и в данном случае не являются принципиальными.

Существенно то, что мы отказываемся от наиболее важных с теоретической точки зрения предположений классической регрессии: $\mathrm{Rank}(X)=M$, $M\{\xi\}=0$ и $\mathrm{cov}\{\xi\}=\sigma^2 E_N$. Отказ от этих предположений является вынужденным, а дополнительное требование $\|\xi\| \le R$ более реалистичным. Заметим, что предположение 4 широко используется в теории возмущений [9], а в теории некорректных задач формулируется как принцип невязок [38].

Указанные предположения имеют существенное значение для выбора теоретических подходов, общей схемы и методов исследования. Несмотря на то, что приведенные выше выражения (1.2), (1.3), (1.5) внешне совпадают с аналогичными соотношениями в задачах линейной регрессии [15], традиционно используемые для их исследования методы теоретической статистики в данном

случае неприменимы. В настоящей работе соответствующая теория и методы оценивания разрабатываются в рамках алгебраической теории возмущений.

1.3. Общая схема решения задачи

В рамках статистической теории оценивания критерий качества идентификации задается на основе априорной информации о распределении ошибок [24,56]. Если такая информация отсутствует, выдвигается какая-либо правдоподобная априорная гипотеза относительно распределения ошибок, а построенная на ее основе процедура подвергается статистическим испытаниям с целью уточнения вероятностной модели. В предыдущем разделе показано, что такой подход неприемлем для идентификации модели искажающей системы по фиксированному малому фрагменту изображения.

Если о распределении ошибок ничего не известно, опираясь на восходящее к Гауссу мнение, критерий качества оценивания берут в виде суммы квадратов компонент вектора невязок:

$$Q(\hat{\mathbf{c}}) = \hat{\boldsymbol{\xi}}^T\hat{\boldsymbol{\xi}} = \sum_i \hat{\xi}_i^2, \quad i = \overline{1,N}, \tag{1.6}$$

где

$$\hat{\boldsymbol{\xi}} = \hat{\mathbf{y}} - \mathbf{y} = \mathbf{X}\hat{\mathbf{c}} - \mathbf{y} \tag{1.7}$$

- вектор невязок, а $\hat{\mathbf{c}}$ - вектор искомых оценок. В работе [19] показано, что метод наименьших квадратов (МНК), в большинстве случаев, не может обеспечить получение достаточно точных оценок при идентификации систем с шумами, поскольку он опирается на жесткую априорную гипотезу, вытекающую из предположений о переменных, свободных от шума. Улучшить свойства оценок можно путем вложения МНК в некоторые вероятностные модели.

В частности, в случае гауссовых помех с ковариационной матрицей ошибок измерений $\mathbf{K} = M\{\boldsymbol{\xi}\boldsymbol{\xi}^T\}$ оптимальным в среднеквадратическом смысле является обобщенный метод наименьших квадратов (ОМНК):

$$Q(\hat{\mathbf{c}}) = \hat{\boldsymbol{\xi}}^T\mathbf{G}^2\hat{\boldsymbol{\xi}}. \tag{1.8}$$

Здесь $\mathbf{G}^2=\mathbf{K}^{-1}$. Причины, по которым мы ввели такое обозначение для обратной ковариационной матрицы, будут понятны из дальнейшего.

При указанных условиях критерий (1.8) совпадает с методом максимального правдоподобия, поэтому вместо вектора невязок $\hat{\boldsymbol{\xi}}$ здесь может фигурировать вектор ошибок $\boldsymbol{\xi}$. Соответствующая этому критерию оценка

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{y}, \qquad (1.9)$$

а вектор ошибок идентификации (без учета погрешностей, вносимых на этапе вычисления оценок) [53,56]

$$\Delta\mathbf{c} = \hat{\mathbf{c}} - \mathbf{c} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\boldsymbol{\xi}. \qquad (1.10)$$

Использование ОМНК-оценки (1.9) в данном случае вызывает возражения. Если статистические характеристики шумов не известны, выбор матрицы $\mathbf{G}^2$ трудно обосновать. В то же время из (1.9), (1.10) видно, что изменяя, при заданных $\mathbf{X}$, $\mathbf{y}$, матрицу $\mathbf{G}^2$ можно существенным образом влиять на ошибку идентификации. Другой, очевидный, способ изменения ошибки идентификации (1.10) непосредственная корректировка вектора зависимых переменных (выхода модели) $\mathbf{y}$ в уравнении (1.2).

Другими словами, представляется оправданным применение следующего класса модифицированных МНК-оценок:

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^2(\mathbf{y} + \Delta\mathbf{y}), \qquad (1.11)$$

где $\mathbf{G}^2$ - так называемая весовая матрица размерности N×N, а $\Delta\mathbf{y}$ - корректирующий N×1 -вектор. При $\mathbf{G}^2=\mathbf{E}$ и $\Delta\mathbf{y}=0$ (1.11) является обычной МНК-оценкой. Варьируя этими параметрами можно управлять качеством оценки. Однако для их определения, и это следует особо подчеркнуть, мы не можем использовать априорную вероятностную модель ошибок.

Введение весовой матрицы $\mathbf{G}^2$ равносильно тому, что исходная матрица X и вектор $\mathbf{y}$ в (1.2) подверглись линейному преобразованию вида

$$\tilde{\mathbf{y}} = \mathbf{G}\mathbf{y}, \quad \tilde{\mathbf{X}} = \mathbf{G}\mathbf{X}. \qquad (1.12)$$

При этом вектор ошибок ξ преобразуется в $\tilde{\xi} = \mathbf{G}\xi$, а $\Delta\tilde{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}\tilde{\xi}$. В настоящей работе мы ограничимся классом диагональных матриц преобразования: $\mathbf{G} = \mathrm{diag}\left[g_1, g_2, \ldots, g_N\right]$. Таким образом, модифицированная оценка (1.11), по существу, является МНК-оценкой, построенной по предварительно преобразованным исходным данным. Для краткости преобразование (1.12) далее будем называть *взвешиванием*, а преобразование вектора зависимых переменных вида $\mathbf{y} + \Delta\mathbf{y}$ - *корректировкой* данных.

Не вызывает сомнений, что при построении преобразований взвешивания и корректировки должна использоваться вся доступная информация. В частности, если известна, например, (вследствие независимости ошибок) диагональная ко-

вариационная матрица ошибок измерений, весовые коэффициенты следует принять равными обратным значениям элементов этой матрицы. В работе [54] предложено рассматривать три уровня доступной информации: априорная, текущая и дополнительная. Под априорной подразумевается информация, доступная до начала функционирования системы. Под текущей имеется в виду информация, полученная непосредственно в процессе функционирования, например, путем сопоставления измерительной информации от различных источников, а дополнительная появляется непосредственно в ходе вычисления оценок.

Способы получения и использования априорной и текущей информации, конечно, заслуживают внимания и могут быть предметом самостоятельного исследования. В контексте настоящей работы предполагается, что потенциал известных в этой области методов, конечно, используется, а задача заключается в повышении точности идентификации за счет использования, насколько это возможно, дополнительной информации, появляющейся в ходе оценивания.

В качестве источника дополнительной информации рассматриваются реализации вектора невязок. Поэтому схема построения оценки (1.11) итерационная. На первом шаге строится МНК-оценка:

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}. \qquad (1.13)$$

Затем вычисляется вектор невязок (1.7) и анализируется содержащаяся в нем дополнительная информация об ошибках в исходных данных.

С использованием априорной и дополнительной информации о связи реализаций векторов ошибок и невязок определяются диагональная весовая N×N - матрица **G** и/или корректирующий N×1 -вектор $\Delta\mathbf{y}$ [1]:

$$\Delta\mathbf{y} = \mathrm{f}(\hat{\xi}), \quad \mathbf{G} = \mathrm{G}(\hat{\xi}). \qquad (1.14)$$

Далее строится преобразованный набор данных $\tilde{\mathbf{X}}, \tilde{\mathbf{y}}$ по соотношениям

$$\tilde{\mathbf{y}} = \mathbf{G}\mathbf{y}, \quad \tilde{\mathbf{X}} = \mathbf{G}\mathbf{X} \quad (\text{при этом } \tilde{\xi} = \mathbf{G}\xi) \qquad (1.15)$$

и/или

$$\tilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y} \qquad (1.16)$$

и вновь вычисляется МНК-оценка:

$$\hat{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}\tilde{\mathbf{y}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^{2}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{2}(\mathbf{y} + \Delta\mathbf{y}). \qquad (1.17)$$

---

[1] В разделе 4.1 будет рассмотрен метод, в котором по вектору невязок $\hat{\xi}$ вначале строится так называемый корректирующий вектор $\tilde{\xi}$, с использованием которого затем осуществляются преобразования взвешивания и/или корректировки вектора выхода модели - (1.15), (1.16).

При необходимости, в зависимости от способа преобразования, вычисления по соотношениям (1.14)-(1.17) могут повторяться несколько раз. Правило останова заключается в сравнении меры близости оценок на соседних шагах с заданным порогом. Ошибка идентификации на каждом шаге зависит от того, насколько удачно построено очередное преобразование данных и определяется как

$$\Delta\tilde{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}(\tilde{\xi} - \Delta\mathbf{y}). \qquad (1.18)$$

В рамках указанной схемы построения оценок возникают следующие вопросы и, вытекающие из них, задачи исследования.

1. Возможно ли в принципе получение удовлетворительного начального приближения (1.13) на заданном наборе данных **X**, **y**?
2. Если это невозможно, то нельзя ли изменить ситуацию, подвергнув исходные данные какой-либо процедуре отбора?
3. Можно ли, в принципе, построить корректирующий вектор $\Delta\mathbf{y}$ по реализации вектора невязок $\hat{\xi}$?
4. Как должны строиться преобразования **G** и/или $\Delta\mathbf{y}$ по реализациям вектора невязок $\hat{\xi}$, чтобы при этом происходило увеличение точности идентификации?

Если в схеме (1.13) - (1.17) исключить из рассмотрения преобразование корректировки, она формально совпадает с методом взвешенных наименьших квадратов, описанным в работах [15,23,27]. Однако используемые в указанных работах теоретические предпосылки для его построения, основанные на задании априорных вероятностных распределений, в данном случае неприемлемы. В четвертой главе будет показана связь этих методов с методами, построенными на основе используемого в настоящей работе алгебраического подхода.

Еще раз подчеркнем, что принципиальным в описанной схеме построения оценки (1.11) является отсутствие априорных предположений о вероятностном распределении ошибок. Основная задача в данном случае заключается *не в уточнении вероятностных распределений*, а в *непосредственной оценке конкретных реализаций ошибок* на каждом наборе данных. С одной стороны, к этому принуждают обстоятельства, с другой стороны, именно в силу малого числа наблюдений появляется возможность непосредственного анализа каждого отсчета и сопоставления этих отсчетов с априорными данными из других источников, доставляющих более богатую информацию (например, о форме и геометрических размерах регистрируемых объектов), чем статистические характеристики, полученные усреднением по множеству реализаций.

Сказанное, конечно, не исключает использования априорной статистической информации, которая проявляет свойство устойчивости и на малых наборах данных. В данном случае дополнительная информация, имеющая вероятностную природу, как правило, также связана с конкретной ситуацией (видом двумерной функции яркости, условиями регистрации конкретного объекта и др.), вследствие чего сформулировать статистический критерий качества оценивания не удается. Поэтому задача оценивания по малому числу наблюдений, по существу, является *алгебраической*, а для анализа точностных свойств оценок на фиксированном наборе данных должны использоваться подходы, известные в алгебраической теории возмущений и теории некорректных задач. Опираясь на эту точку зрения, ниже развиваются алгебраические методы анализа ошибок.

1.4. Формы представления ошибок

В статистической теории оценивания для анализа качества решений широко используются асимптотические свойства оценок: несмещенность, состоятельность и эффективность. При оценивании по фиксированному малому набору данных асимптотические показатели качества могут ввести в заблуждение. В настоящем разделе описываются формы представления ошибок, пригодные для анализа точности в рамках алгебраического подхода.

В описанной выше схеме оценивания искомые оценки (1.13) и/или (1.17) являются решением системы нормальных уравнений

$$\mathbf{A}\hat{\mathbf{c}} = \mathbf{b}, \tag{1.19}$$

сформированной по $\mathbf{X}$, $\mathbf{y}$ или $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{y}}$. Поскольку мы намерены рассмотреть общую методику анализа точности на основе алгебраического подхода, для нас не имеет значения подвергались ли данные преобразованию. Поэтому здесь мы будем полагать, что $\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$, где $\mathbf{X}$-N×M-матрица, а $\mathbf{y}$- N×1-вектор соответственно.

Предполагается, что соответствующая уравнению (1.2) точная модель идентифицируемой системы:

$$\mathbf{y}^{*} = \mathbf{X}^{*}\mathbf{c} \tag{1.20}$$

существует, где $\mathbf{X}^{*}$ и $\mathbf{y}^{*}$ - не зашумленные матрица и вектор. В действительности матрица $\mathbf{X}$ и вектор $\mathbf{y}$ фиксируются с ошибками. В таблице 1.1 приведены варианты формирования матрицы $\mathbf{X}$ и вектора $\mathbf{y}$ для различных моделей с учетом ошибок, которые реально всегда присутствуют при регистрации изображений.

Выше мы уже указывали на то, что рассматриваемые модели не являются регрессионными в строгом смысле определения, принятого в статистической теории [15]. Поэтому в таблице и далее по тексту, во избежание недоразумений, не используются широко известные понятия: регрессия (Р), авторегрессия (АР) и смешанная регрессия (РАР) [51]. Для обозначения различий в структуре модели здесь и далее используются понятия: КИХ-фильтр и БИХ-фильтр.

Фигурирующие в таблице матрица δX и вектор δy, составлены из погрешностей измерений, аппроксимации и др. Для модели КИХ-фильтра $\delta X=0$. Для модели БИХ-фильтра матрица δX формируется по-разному. В частности, если числитель соответствующей предаточной функции равен единице, то матрица δX не содержит нулевой блок N×L. Если дробно-рациональная передаточная функция содержит полиномы как в числителе, так и в знаменателе, то подматрица размерности N×L содержит нулевые элементы. Эти варианты можно записать в виде

$$\delta X = \begin{bmatrix} \delta y_{1,1}, & \delta y_{1,K} \\ \cdots & \cdots \\ \cdots & \cdots \\ \delta y_{N,1}, & \delta y_{N,K} \end{bmatrix}, \quad \delta X = \begin{bmatrix} 0 & 0 & \delta y_{1,1}, & \delta y_{1,K} \\ & & \cdots\cdots & \cdots\cdots \\ & & \cdots & \cdots \\ 0 & 0 & \delta y_{N,1}. & \delta y_{N,K} \end{bmatrix}, \quad \delta y = \begin{bmatrix} \delta y_1 \\ \vdots \\ \vdots \\ \delta y_N \end{bmatrix}.$$

Таблица 1.1. Варианты формирования исходных данных для различных моделей

Тип модели	Матрица X	Вектор y	Вектор ξ
Фильтр с конечной импульсной характеристикой (КИХ-фильтр)	X^*	$y = y^* + \delta y$	$\xi = \delta y$
Фильтр с бесконечной импульсной характеристикой (БИХ-фильтр)	$X^*+\delta X$	$y = y^* + \delta y$	$\xi = \delta y - \delta Xc$

Если матрица X и вектор y формируются так, что опорные области, соответствующие их различным строкам, (хотя бы частично) пересекаются, среди элементов матрицы δX будут повторяющиеся. Их расположение в матрице зависит от порядка обхода фрагмента изображения. При реализации БИХ-фильтра вид опорной области и порядок обхода фрагмента не могут быть произвольными. Связанные с этим особенности подробно будут обсуждаться в главе 6, посвященной вопросам практической реализации алгоритмов.

Наиболее важными являются различия в соотношениях для формирования вектора ошибок ξ. Дело в том, что хотя матрица X и вектор y формируются по

зашумленным данным, они известны до начала идентификации. Факт вхождения в них ошибок измерений напрямую не влияет на точность оценивания. Это влияние опосредованное (через изменение храктера обусловленности задачи). Поэтому при анализе точности принимают во внимание лишь вектор ошибок ξ.

В случае использования модели БИХ-фильтра вектор ошибок, кроме погрешностей измерений выходного искаженного изображения, содержит также составляющую, зависящую от параметров искажающей системы. Указанный факт имеет принципиальное значение при анализе связи ошибок идентификации с ошибками в исходных данных. Это особенно актуально при использовании модели БИХ-фильтра с единицей в числителе передаточной функции.

Вернемся теперь к уравнению (1.19). 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:

$$\mathbf{A} = \mathbf{A}^* + \delta\mathbf{A}, \tag{1.21}$$

$$\mathbf{b} = \mathbf{b}^* + \delta\mathbf{b}, \tag{1.22}$$

где

$$\mathbf{A}^* = \mathbf{X}^{*T}\mathbf{X}^*, \tag{1.23}$$

$$\mathbf{b}^* = \mathbf{X}^{*T}\mathbf{y}^* \tag{1.24}$$

- точные (невозмущенные) матрица и вектор. Возмущения $\delta\mathbf{A}$ и $\delta\mathbf{b}$ для модели БИХ-фильтра определяются, как

$$\delta\mathbf{A} = \mathbf{X}^{*T}\delta\mathbf{X} + \delta\mathbf{X}^T\mathbf{X}^* + \delta\mathbf{X}^T\delta\mathbf{X}, \tag{1.25}$$

$$\delta\mathbf{b} = \mathbf{X}^{*T}\delta\mathbf{y} + \delta\mathbf{X}^T\mathbf{y}^* + \delta\mathbf{X}^T\delta\mathbf{y}. \tag{1.26}$$

Для модели КИХ-фильтра выражения для возмущений упрощаются:

$$\delta\mathbf{A} = 0, \quad \delta\mathbf{b} = \mathbf{X}^{*T}\delta\mathbf{y}. \tag{1.27}$$

Теперь построим соотношения для ошибок МНК - оценок. Для этого наряду с (1.19) в соответствии с предположением (1.20) запишем точную систему уравнений:

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

С учетом (1.21),(1.22) это равенство можно переписать в виде

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

Сравнивая (1.19) и (1.28) ошибку оценивания можно записать в виде

$$\Delta\mathbf{c} = \hat{\mathbf{c}} - \mathbf{c} = \mathbf{A}^{-1}\zeta, \tag{1.29}$$

где

$$\zeta = (\delta \mathbf{b} - \delta \mathbf{A}\mathbf{c}). \qquad (1.30)$$

Для теоретических исследований представление ошибок исходных данных в виде (1.30) не всегда удобно. Более продуктивным является использование связи ошибок оценивания непосредственно с исходным вектором ошибок ξ, фигурирующим в таблице 1.1. Это представление можно получить произведя в (1.30) замену $\delta \mathbf{b}$, $\delta \mathbf{A}$ их общими выражениями (1.25) и (1.26):

$$\zeta = \mathbf{X}^{*\mathrm{T}} \delta \mathbf{y} + \delta \mathbf{X}^{\mathrm{T}} \mathbf{y}^{*} + \delta \mathbf{X}^{\mathrm{T}} \delta \mathbf{y} - \mathbf{X}^{*\mathrm{T}} \delta \mathbf{X}\mathbf{c} - \delta \mathbf{X}^{\mathrm{T}} \mathbf{X}^{*}\mathbf{c} - \delta \mathbf{X}^{\mathrm{T}} \delta \mathbf{X}\mathbf{c}.$$

Сгруппировав слагаемые в правой части последнего равенства перепишем его в виде

$$\zeta = \mathbf{X}^{*\mathrm{T}} (\delta \mathbf{y} - \delta \mathbf{X}\mathbf{c}) + \delta \mathbf{X}^{\mathrm{T}} (\delta \mathbf{y} - \delta \mathbf{X}\mathbf{c}) = \left(\mathbf{X}^{*\mathrm{T}} + \delta \mathbf{X}^{\mathrm{T}}\right)(\delta \mathbf{y} - \delta \mathbf{X}\mathbf{c}).$$

Далее, поскольку $\mathbf{X}^{*\mathrm{T}} + \delta \mathbf{X}^{\mathrm{T}} = \mathbf{X}^{\mathrm{T}}$ и $\delta \mathbf{y} - \delta \mathbf{X}\mathbf{c} = \xi$, получаем

$$\zeta = \delta \mathbf{b} - \delta \mathbf{A}\mathbf{c} = \mathbf{X}^{\mathrm{T}} \xi. \qquad (1.31)$$

Соотношение (1.31) показывает связь различных представлений ошибок в исходных данных. Компоненты вектора ошибок ζ представляют собой скалярные произведения вектора ξ на соответствующие векторы матрицы $\mathbf{X}$. В частности, если векторы-столбцы матрицы $\mathbf{X}$ нормированы так, что $\|\mathbf{x}_i\|_2 = 1$ для всех i=1,M, имеет место следующая геометрическая трактовка: компоненты вектора ζ - суть проекции вектора ошибок ξ на направления, задаваемые векторами независимых переменных $\mathbf{x}_i$, $i = \overline{1,\mathrm{M}}$.

С учетом (1.29) и (1.31) можно записать следующие равенства:

$$\Delta \mathbf{c} = \hat{\mathbf{c}} - \mathbf{c} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1} \mathbf{X}^{\mathrm{T}} \xi = \mathbf{A}^{-1} \mathbf{X}^{\mathrm{T}} \xi. \qquad (1.32)$$

Это соотношение, показывает связь вектора ошибок оценивания $\Delta \mathbf{c}$ с вектором ошибок ξ. Далее оно будет широко использоваться для установления основных закономерностей формирования ошибок и предельно достижимой точности оценивания. Заметим, что в рамках традиционных методов анализа точности, обе части равенства (1.32) полагается подвергнуть операции математического ожидания, откуда сразу следует несмещенность оценок. Мы не делаем этого, поскольку асимптотические свойства оценок в настоящей работе не используются.

В заключение еще раз подчеркнем, что матричный коэффициент $\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X} = \mathbf{A}^{-1}\mathbf{X}^{\mathrm{T}}$ в (1.32) при решении задачи оценивания всегда известен. Это имеет место как в случае модели КИХ-фильтра, когда матрица X задана точно,

так и в случае модели БИХ-фильтра, для которой $\mathbf{X} = \mathbf{X}^* + \delta\mathbf{X}$ (см. таблицу 1.1). Поэтому при анализе точности идентификации матрица **A** и вектор **b** считаются заданными точно, а значение имеет лишь способ формирования ошибок (ζ и/или ξ).

Тем не менее, выписанные выше возмущения (особенно для матрицы **A**) полезны при анализе информативности данных. Легко заметить, что если матрица $\mathbf{A}^*$ вырождена, соответствующая ей матрица $\mathbf{A} = \mathbf{A}^* + \delta\mathbf{A}$ может оказаться хорошо обусловленной. Поэтому далее в случаях, когда в контексте конкретного исследования способ формирования матрицы **A** имеет принципиальное значение, мы будем указывать тип модели (КИХ-фильтр и/или БИХ-фильтр). При этом общие обозначения, при необходимости, могут быть заменены их конкретными выражениями из соответствующей строки таблицы 1.1.

1.5. Ортогональные разложения ошибок оценивания

Теперь, опираясь на выписанные выше соотношения (1.29), (1.32), вскроем общие закономерности формирования ошибок оценивания. Для этого воспользуемся сингулярным разложением задачи наименьших квадратов [39] и осуществим разложение фигурирующих в указанных равенствах векторов ошибок по соответствующим ортогональным базисам.

Известно [23,59], что для N×M - матрицы **X** ранга k≤M существуют ортогональные N×N - матрица **T** и M×M - матрица **F** такие, что

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

где **S** диагональная N×M - матрица, составленная из неотрицательных чисел, ровно **k** среди которых строго положительны. Диагональные элементы матрицы **S** называются сингулярными числами матрицы **X**.

Известно также, что положительно определенная симметричная невырожденная матрица $\mathbf{A}=\mathbf{X}^T\mathbf{X}$ допускает спектральное разложение $\mathbf{A} = \mathbf{F}\Lambda\mathbf{F}^T$, где **F** - ортогональная M×M матрица такая, что $\mathbf{A} = \mathbf{F}\Lambda\mathbf{F}^T$, $\mathbf{F}^T\mathbf{F} = \mathbf{F}\mathbf{F}^T = \mathbf{E}_M$, $\mathbf{E}_M$ - единичная матрица, а $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_M)$ - диагональная матрица, составленная из собственных значений матрицы **A**. Числа λ_i являются квадратами соответствующих сингулярных чисел и называются характеристическими числами или собственными значениями.

Спектральное разложение можно построить также и для вырожденной N×N - матрицы $\mathbf{B}=\mathbf{X}\mathbf{X}^T$ ранга M. Если X вещественная ограниченная N×M матрица (N>M), RankX=M, а $\mathbf{T}_0$ - N×(N-M) - матрица, составленная из нормированных собственных векторов, соответствующих нулевым собственным значениям той же матрицы ($\mathbf{B}\mathbf{T}_0 = \mathbf{X}\mathbf{X}^T\mathbf{T}_0 = 0$), то существует ортогональная N×N - матрица T такая, что

$$\mathbf{B} = \mathbf{X}\mathbf{X}^T = \mathbf{T}\begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix}\mathbf{T}^T, \quad \mathbf{T}^T\mathbf{T} = \mathbf{T}\mathbf{T}^T = \mathbf{E}_N, \quad \mathbf{T} = [\mathbf{T}_\lambda : \mathbf{T}_0],$$

где

$$\mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\Lambda^{-1/2} \tag{1.33}$$

N×M - матрица, составленная из нормированных собственных векторов, соответствующих ненулевым собственным значениям матрицы $\mathbf{X}\mathbf{X}^T$, а $\Lambda=\mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_M)$ - диагональная матрица, составленная из ненулевых собственных значений матрицы $\mathbf{B}=\mathbf{X}\mathbf{X}^T$. Заметим, что ненулевые собственные значения матрицы $\mathbf{X}\mathbf{X}^T$ совпадают с собственными значениями невырожденной матрицы $\mathbf{A} = \mathbf{X}^T\mathbf{X}$.

Можно показать, что аналогичное утверждение справедливо и в случае, когда матрица $\mathbf{A} = \mathbf{X}^T\mathbf{X}$ вырожденная, т.е. RankX=k<M. В этом случае матрица $\mathbf{T}_0$ будет иметь размерность N×(N-k), $\mathbf{T}_\lambda$ - N×k., а F, Λ, E - k×k.

С использованием приведенных ортогональных представлений построим соотношения для канонических разложений векторов ошибок. Предположим, что RankX=M. Умножим обе части (1.29) слева на $\mathbf{F}^T$. Тогда с учетом свойства $\mathbf{F}\mathbf{F}^T=\mathbf{E}$ можно записать

$$\mathbf{F}^T\Delta\mathbf{c} = \Lambda^{-1}\mathbf{F}^T\zeta, \tag{1.34}$$

где F указанная выше ортогональная матрица: $\mathbf{F}^T\mathbf{A}\mathbf{F}^T=\Lambda$.

Равенство (1.34) допускает простую геометрическую трактовку: проекции вектора ошибок оценивания на базис, образованный собственными векторами матрицы A, образуются из проекций вектора ошибок исходных данных (1.30) на тот же базис с "коэффициентами усиления" обратно пропорциональными соответствующим собственным значениям.

Аналогичное соотношение можно построить для вектора ошибок идентификации, представленного в виде соотношения (1.32). Для этого с учетом соотношения связи между векторами ошибок ζ и ξ (1.31) перепишем (1.34) в виде

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

Подставляя в это равенство вместо матрицы $\Lambda^{-1/2}\mathbf{F}^T\mathbf{X}^T$ равную ей в соответствии с (1.33) матрицу: $\mathbf{T}_\lambda^T$, получаем следующее представление

$$\mathbf{F}^T\Delta\mathbf{c} = \Lambda^{-1/2}\mathbf{T}_\lambda^T\xi . \qquad (1.35)$$

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 ξ на направления базиса, образованного собственными векторами, соответствующими ненулевым собственным значениям матрицы $\mathbf{B} = \mathbf{X}\mathbf{X}^T$ с "коэффициентами усиления" $\lambda_i^{-1/2}$, $i = \overline{1,M}$.

Поскольку ненулевые собственные значения матрицы $\mathbf{B} = \mathbf{X}\mathbf{X}^T$ совпадают с собственными значениями матрицы $\mathbf{A} = \mathbf{X}^T\mathbf{X}$, а также имеет место связь $\zeta = \mathbf{X}^T\xi$ векторов ошибок (1.31), соотношение (1.34) легко может быть получено из (1.35) подстановкой $\mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\Lambda^{-1/2}$ Для теоретического анализа ошибок идентификации соотношение (1.35) представляет больший интерес, т.к. в нем фигурирует исходный вектор ошибок ξ, фигурирующий в уравнении (1.2).

Соотношения (1.34), (1.35) представляются важными для понимания общих закономерностей формирования ошибок идентификации. Для сравнительного анализа точности более предпочтительными являются скалярные характеристики векторов ошибок. Поэтому с использованием разложений (1.34), (1.35) запишем выражения для квадрата евклидовой нормы вектора ошибок оценивания.

Подвергнув операции транспонирования обе части равенства (1.34), перепишем его в виде

$$\Delta\mathbf{c}^T\mathbf{F} = \zeta^T\mathbf{F}\Lambda^{-1} \qquad (1.36)$$

Далее, умножив слева обе части равенства (1.36) на соответствующие (левую и правую) части (1.34), с учетом свойств матрицы $\mathbf{F}$ получим:

$$\Delta\mathbf{c}^T\Delta\mathbf{c} = \|\Delta\mathbf{c}\|_2^2 = \zeta^T\mathbf{F}\Lambda^{-2}\mathbf{F}^T\zeta = \|\zeta\|_2^2\sum_{i=1}^{M}\lambda_i^{-2}\cos^2(\zeta,\hat{\mathbf{f}}_i), \qquad (1.37)$$

где $\cos(\zeta,\hat{\mathbf{f}}_i)$ - косинус угла между вектором ζ и собственным вектором $\mathbf{f}_i$, (i-м столбцом матрицы $\mathbf{F}$).

Аналогичное соотношение имеет место для случая представления ошибок идентификации в виде (1.35). Для этого транспонировав обе части (1.35) запишем равенство

$$\Delta \mathbf{c}^{\mathrm{T}}\mathbf{F} = \xi^{\mathrm{T}}\mathbf{T}_{\lambda}\Lambda^{-1/2} \tag{1.38}$$

Перемножив обе части равенств (1.38) и (1.35), аналогично предыдущему, получаем

$$\Delta \mathbf{c}^{\mathrm{T}}\Delta \mathbf{c} = \|\Delta \mathbf{c}\|_2^2 = \xi^{\mathrm{T}}\mathbf{T}_{\lambda}\Lambda^{-1}\mathbf{T}_{\lambda}^{\mathrm{T}}\xi = \|\xi\|_2^2 \sum_{i=1}^{M}\lambda_i^{-1}\cos^2\left(\hat{\xi}, \mathbf{t}_{\lambda i}\right), \tag{1.39}$$

где $\cos\left(\hat{\xi}, \mathbf{t}_{\lambda i}\right)$ косинус угла между вектором ξ и собственным вектором $\mathbf{t}_{\lambda,i}$, (i-м столбцом определяемой соотношением (1.33) матрицы $\mathbf{T}_{\lambda}$).

Соотношения (1.37) и (1.39) имеют ясную геометрическую трактовку. В частности, из правых частей этих равенств видно, что норма вектора ошибок идентификации зависит не только от нормы вектора ошибок исходных данных, но также и от ориентации этого вектора относительно пространства матрицы X.

Подчеркнем, что в указанных соотношениях $\|\zeta\|_2^2 \neq \|\xi\|_2^2$. В частности, в соответствии с (1.31)

$$\zeta^{\mathrm{T}}\zeta = \|\zeta\|_2^2 = \left\|\mathbf{X}^{\mathrm{T}}\xi\right\|_2^2 = \xi^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\xi. \tag{1.40}$$

Эти различия величин норм в представлениях ошибок идентификации (1.37) и (1.39) "компенсируются" различиями в значениях степеней коэффициентов λ_i, $i = \overline{1,M}$ при $\cos^2(\cdot)$, являющихся собственными значениями матрицы $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Как будет видно из дальнейшего, каноническое представление (1.39) во многих случаях оказывается более полезным, т.к. в нем фигурирует исходный вектор ошибок, которому может быть поставлен в соответствие вектор невязок.

1.6. Достаточные оценки точности

Приведенные выше соотношения (1.37), (1.39) вскрывают закономерности формирования ошибок идентификации, однако они не пригодны для оценки достижимой точности. Вычисление по этим формулам квадрата евклидовой нормы вектора ошибок идентификации представляется проблематичным, во-первых, из-за необходимости вычисления всех собственных значений матрицы A. Во-вторых, точные значения косинусов вообще никогда не известны, т.к. не известна ориентация вектора ошибок относительно пространства матрицы X. В

настоящем разделе мы приведем некоторые известные и получим новые результаты, использование которых для анализа точности идентификации, по крайней мере, не требует решения полной проблемы собственных значений.

Обсуждая в разделе 1.2 содержательную постановку задачи оценивания по малым фрагментам изображений, мы отказались от основных предположений (4,5) классической регрессии. Вместо этого мы ввели в рассмотрение предположение 4 в виде ограничения на норму вектора ошибок: $\|\xi\| \le R$ и констатировали, что вся неопределенность связана с произвольной ориентацией этого вектора в N-мерном пространстве. Для построения количественных оценок точности идентификации здесь и далее используется обладающая свойством инвариантности к унитарным преобразованиям евклидова норма. При этом указанное ограничение имеет вид:

$$\|\xi\|_2 \le R_\xi . \qquad (1.41)$$

В соответствии с выражениями для ошибок, приведенными в таблице 1.1, неравенство (1.41) является следствием того, что при фиксированной матрице **X** могут быть заданы границы для согласованных норм возмущений δy и $\delta \mathbf{X}$.

Множество Ξ векторов, удовлетворяющих неравенству (1.41), принадлежит "шару":

$$\xi \in \Xi, \quad \Xi = \left\{ \xi : \quad (\xi^T \xi)^{1/2} = \|\xi\|_2 \le R_\xi = \mathrm{Const} \right\}. \qquad (1.42)$$

Ориентация вектора ошибок ξ относительно пространства матрицы **X** случайна и не делается никаких предположений относительно вероятности различных направлений. Поэтому в работе отсутствуют какие-либо вероятностные оценки ошибок идентификации. Эта проблема, по-видимому, заслуживает отдельного исследования в рамках традиционной теории статистического оценивания. В рамках настоящей работы эти вопросы не представляют интереса, т.к. статистические характеристики шумов предполагаются неизвестными.

Из неравенства (1.41) в силу равенства (1.31), устанавливающего связь между различными представлениями ошибок, следует также, что

$$\|\zeta\|_2 \le R_\zeta , \qquad (1.43)$$

где R_ζ пока не известная, но фиксированная константа. Позже мы установим связь величин R_ζ и R_ξ. Здесь мы хотим подчеркнуть следующее. Искомый вектор параметров **c** удовлетворяет уравнению $\mathbf{Ac} = \mathbf{b} - \zeta$ (1.28), т.е. $\mathbf{b} - \zeta$ является

"точной" правой частью. Поскольку вектор ζ не известен, вместо уравнения (1.28) мы вынуждены решать уравнение $\mathbf{A}\hat{\mathbf{c}} = \mathbf{b}$ (1.19), относительно которого, учитывая условие (1.43), мы можем утверждать лишь то, что возмущение правой части в смысле евклидовой нормы не превышает величины R_ζ. Кроме того, в ходе последующих вычислений к ним добавятся ошибки округления.

В теории возмущений известен подход к оценке влияния ошибок, так называемый обратный анализ ошибок. В соответствии с этим подходом влияние ошибок округления при вычислениях равносильно дополнительному внесению ошибок в исходные данные. Обозначим эти дополнительные возмущения матрицы $\mathbf{A}$ и вектора $\mathbf{b}$ - $\Delta\mathbf{A}$ и $\Delta\mathbf{b}$ соответственно. Тогда соответствующая (1.19) система, содержащая также и эквивалентные возмущения [9], принимает вид

$$[\mathbf{A} + \Delta\mathbf{A}]\cdot\hat{\mathbf{c}}_в = [\mathbf{b} - \zeta] + (\zeta + \Delta\mathbf{b}), \tag{1.44}$$

где $\hat{\mathbf{c}}_в$ оценка, которая кроме ошибок $\Delta\mathbf{c}$, определяемых равенствами (1.29), (1.32), содержит также погрешности (округления), возникающие на этапе вычисления оценок. В (1.44) для наглядности в правой части квадратными скобками выделена "точная" правая часть уравнения (1.28), а круглыми суммарное возмущение правой части, связанное с вычислениями и неточностью исходных данных.

Из сравнения (1.28) и (1.44) видно, что для анализа общих ошибок $\Delta\mathbf{c}_в = \hat{\mathbf{c}}_в - \mathbf{c}$, возникающих вследствие первоначальных ошибок исходных данных и эквивалентного возмущения при вычислениях, можно воспользоваться известными результатами теории возмущений. В частности, известна [9,40,41] следующая оценка для максимальной относительной погрешности $\delta_c = \|\Delta\mathbf{c}_в\|/\|\mathbf{c}\|$ в решениях:

$$\delta_c \le \frac{K(A)}{1 - K(A)\delta_A}\left(\delta_b + \delta_A\right), \tag{1.45}$$

где

$$K(A) = \|A\|\|A\|^{-1} \tag{1.46}$$

условное число или, так называемое, число обусловленности, а относительные погрешности исходных данных в общем случае идентификации модели БИХ-фильтра определяются как

$$\delta_A = \|\Delta A\|/\|A\|, \tag{1.47}$$

$$\delta_b = \|\zeta + \Delta b\|/\|b - \zeta\| = \|\zeta + \Delta b\|/\|b + \delta Ac\|. \tag{1.48}$$

для согласованной с евклидовой нормой вектора спектральной нормы вещественной матрицы Грама

$$K(\mathbf{A}) = \lambda_{max}(\mathbf{A})/\lambda_{min}(\mathbf{A}), \qquad (1.49)$$

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

Поскольку в исходных предположениях (1.41), (1.43) используется евклидова норма вектора, далее для числа обусловленности мы будем использовать соотношение (1.49). С учетом этого неравенство (1.45) можно записать в виде

$$\delta_c \le \frac{\lambda_{max}}{\lambda_{min} - \lambda_{max}\delta_A}(\delta_b + \delta_A). \qquad (1.50)$$

Подчеркнем, что при этом для характеристики относительных возмущений (1.47), (1.48) также должны использоваться евклидовы нормы:

$$\delta_c = \|\Delta\mathbf{c}_в\|_2/\|\mathbf{c}\|_2, \quad \delta_A = \|\Delta\mathbf{A}\|_2/\|\mathbf{A}\|_2, \quad \delta_b = \|\zeta + \Delta\mathbf{b}\|_2/\|\mathbf{b} + \delta\mathbf{A}\mathbf{c}\|_2.$$

Если данные на фрагменте информативны, т.е. матрица **A** хорошо обусловлена, то при правильной организации вычислений и соответствующем выборе разрядной сетки вычислителя можно обеспечить выполнение неравенства

$$\lambda_{min} >> \lambda_{max}\delta_A.$$

Тогда условие (1.50) принимает более простой вид

$$\delta_c \le \frac{\lambda_{max}}{\lambda_{min}}(\delta_b + \delta_A) = K(\mathbf{A})(\delta_A + \delta_b) = K(\mathbf{A})\left(\frac{\|\Delta\mathbf{A}\|_2}{\|\mathbf{A}\|_2} + \frac{\|\zeta + \Delta\mathbf{b}\|_2}{\|\mathbf{b} + \delta\mathbf{A}\mathbf{c}\|_2}\right) \qquad (1.51),$$

Часто точность вычислений заметно превышает точность исходных данных. Связано это, как правило, с тем, что принятая модель искажений не вполне соответствует реальным физическим процессам в системе. Тогда оценка (1.51) существенно упрощается:

$$\delta_c \le \frac{\lambda_{max}}{\lambda_{min}}\delta_b = K(\mathbf{A})\frac{\|\zeta\|_2}{\|\mathbf{b} + \delta\mathbf{A}\mathbf{c}\|_2}, \qquad (1.52)$$

где в соответствии с (1.28) $\zeta = \mathbf{X}^T\xi = \delta\mathbf{b} - \delta\mathbf{A}\mathbf{c}$. В частном случае идентификации модели КИХ-фильтра имеем

$$\delta_c \le \frac{\lambda_{max}}{\lambda_{min}}\delta_b = K(\mathbf{A})\frac{\|\zeta\|_2}{\|\mathbf{b}\|_2}, \qquad (1.53)$$

а

$$\zeta = \mathbf{X}^T\xi = \delta\mathbf{b}.$$

Таким образом, при заданном ограничении (1.43) на норму $\|\xi\|_2$ для оценки возможной относительной погрешности идентификации в соответствии с (1.50), (1.51) необходимо решать частную проблему собственных значений, т.е. определять максимальное и минимальное характеристические числа матрицы Грама **A**. Если вычислительный процесс организован таким образом, что ошибки округлений малы по сравнению с погрешностями задания исходных данных, можно построить более простые оценки для относительных погрешностей (1.52), (1.53). Заметим, что при этом $\delta_c = \|\Delta \mathbf{c}_в\|_2 / \|\mathbf{c}\|_2 \cong \|\Delta \mathbf{c}\|_2 / \|\mathbf{c}\|_2$. Обычно это требование соблюдается. Поэтому далее мы всюду будем применять для общей ошибки идентификации $\Delta \mathbf{c}_в$ обозначение $\Delta \mathbf{c}$. Если различия между ними в конкретных ситуациях существенны, это будет оговариваться особо.

Построим теперь оценки для скалярной характеристики $\Delta \mathbf{c}^T \Delta \mathbf{c} = \|\Delta \mathbf{c}\|_2^2$ в предположении, что для нормы вектора ошибок исходных данных выполняется предположение $\|\xi\|_2 \le R$ (1.41). Будем использовать допущение, что норма (длина) вектора ошибок ξ не зависит от ориентации этого вектора относительно пространства матрицы **X**. Для решения задачи воспользуемся каноническими разложениями (1.37), (1.39).

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, т.е. при условии, что $\xi \in \Xi_R$, а

$$\Xi_R = \left\{ \left(\xi^T \xi\right)^{1/2} = \|\xi\|_2 = R_\xi = \text{Const} \right\}. \qquad (1.54)$$

Множество Ξ_R более узкое по сравнению с Ξ в (1.42). Оно представляет собой подмножество векторов ошибок, концы которых принадлежат сфере, радиус которой равен радиусу шара (1.42) - R_ξ.

По предположению матрица **X** фиксирована и известна, а фигурирующие в канонических разложениях (1.37), (1.39) собственные значения $\lambda_i(\mathbf{X}^T\mathbf{X})$, i=1,M не зависят от направления вектора ξ, поэтому мы можем считать их константами. Таким образом, задача сводится к отысканию такой ориентации вектора ошибок ξ относительно пространства матрицы **X**, при которой квадрат нормы вектора

ошибок идентификации будет достигать экстремальных значений (максимума или минимума).

Введем в рассмотрение переменную $\psi_i = \xi, \mathbf{t}_{\lambda i}$, обозначающую угол между вектором ошибок ξ и направлением i-го собственного вектора (i=1,M). При соответствующих предположениях непрерывности множества возможных реализаций вектора шума ξ на заданной сфере переменные ψ_i i=1,M также непрерывны в интервале [0,2π]. Следовательно, квадрат евклидовой нормы вектора ошибок идентификации, заданный в виде (1.39), является непрерывной функцией M непрерывных аргументов ψ_i:

$$\|\Delta \mathbf{c}\|_2^2 = \|\xi\|_2^2 \sum_{i=1}^{M} \lambda_i^{-1} \cos^2 \psi_i . \qquad (1.55)$$

Тогда для фиксированной матрицы **X** (Rank(**X**)=M) и при фиксированной процедуре построения базиса пространства $\mathcal{R}(\mathbf{X})$ необходимые условия экстремума функции (1.55) задаются следующей системой M уравнений:

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

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

$$\psi_i = 0, \pm\frac{\pi}{2}. \qquad (1.56)$$

Таким образом, необходимые условия экстремума для величины $\|\Delta \mathbf{c}\|_2^2$ выполняются либо при $\xi \in \mathcal{N}(\mathbf{X})$, либо на направлениях, задаваемых базисными векторами пространства $\mathcal{R}(\mathbf{X})$. Другими словами, если $\xi \in \mathcal{R}(\mathbf{X})$, его направление должно совпадать с одним из (ортогональных друг другу) векторов-столбцов N×M-матрицы $\mathbf{T}_\lambda$.

Теперь из геометрических соображений легко установить максимальное и минимальное значение квадрата нормы вектора ошибок идентификации. Ясно, что при $\xi \in \mathcal{N}(\mathbf{X})$ имеет место равенство $\|\Delta \mathbf{c}\|_2^2 = 0$. При $\xi \in \mathcal{R}(\mathbf{X})$ зададим в качестве "подозрительного" на экстремум одно, например i-е, направление ортогонально-

го базиса. При этом $\psi_i = 0$. Но тогда в соответствии с (1.56) $\psi_j = \pm\pi/2$, $j = \overline{1,M}$, $j \neq i$ и, следовательно, сумма в правой части функции (1.55) будет содержать лишь одно слагаемое $\|\Delta c\|_2^2 = \|\xi\|_2^2 \lambda_i^{-1}$, соответствующее выбранному (i-му) направлению, а экстремальные значения $\|\Delta c\|_2^2$ будут достигаться на направлениях, соответствующих максимальному и минимальному собственным значениям. Таким образом, имеет место

утверждение 1. Пусть Rank(X)=M, а множество векторов ошибок $\xi \in \Xi_R$:

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

Тогда

$$\lambda_{max}^{-1} R_\xi^2 \le \|\Delta c\|_2^2 \le \lambda_{min}^{-1} R_\xi^2, \tag{1.57}$$

если вектор $\xi \in \Xi_R$ принадлежит пространству матрицы X: $\xi \in \mathcal{R}(X)$, и

$$\|\Delta c\|_2^2 \le \lambda_{min}^{-1} R_\xi^2 \tag{1.58}$$

в остальных случаях. ■

Конечно, оценку (1.58) можно получить достаточно просто из выражения для вектора ошибок идентификации

$$\Delta c = \left[X^T X\right]^{-1} X\xi \tag{1.59}$$

используя неравенство треугольника [11]. Однако утверждение 1 и особенно использовавшиеся для его обоснования рассуждения дают более полное представление о возможностях повышения точности идентификации. В частности, видно, что при построении МНК-оценок составляющая вектора ошибок, принадлежащая нуль-пространству матрицы X, не влияет на точность оценок. Проблема заключается в компенсации влияния проекции вектора ошибок на $\mathcal{R}(X)$. В связи с этим, приведенная в (1.57) для случая $\xi \in \mathcal{R}(X)$, оценка снизу квадрата евклидовой нормы вектора ошибок идентификации может быть весьма полезной.

Аналогичные (1.57), (1.58) оценки могут быть построены также в случае, когда величина квадрата нормы вектора ошибок идентификации представляется в виде соотношения (1.37):

$$\|\Delta c\|_2^2 = \|\zeta\|_2^2 \sum_{i=1}^{M} \lambda_i^{-2} \cos^2 \varphi_i, \tag{1.60}$$

где $\varphi_i = (\zeta, \hat{f}_i)$ угол между вектором ошибок ζ и направлением собственного вектора f_i, соответствующего i-му собственному значению λ_i матрицы $X^T X$. Дей-

...ствуя по той же, что и выше, схеме из (1.60) с учетом ограничения (1.43) получаем неравенство аналогичное (1.57):

$$\|\Delta \mathbf{c}\|_2^2 \le \lambda_{\min}^{-2}\|\zeta\|_2^2 \le \lambda_{\min}^{-2} R_\zeta^2, \qquad (1.61)$$

где в соответствии с (1.30) $\zeta = \delta \mathbf{b} - \delta \mathbf{A}\mathbf{c}$. В отличие от (1.57) здесь вместо $\lambda_{\min}^{-1}$ в качестве "коэффициента усиления" ошибок исходных данных фигурирует $\lambda_{\min}^{-2}$. По-видимому, как и в каноническом представлении (1.37), это должно "компенсироваться" соответствующими различиями в величине норм векторов ошибок исходных данных, которые связаны соотношением $\zeta = \mathbf{X}^T\xi$ (1.31).

Теперь установим связь величин R_ζ и R_ξ. Применим к соотношению связи между ошибками (1.31) те же, что и ранее, стандартные преобразования:

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

Далее используя ту же, что и ранее, схему, по аналогии с (1.37), (1.39), получаем

$$\|\zeta\|_2^2 = \|\xi\|_2^2 \sum_{i=1}^{M} \lambda_i \cos^2\left(\hat{\xi}, \mathbf{t}_{\lambda i}\right). \qquad (1.62)$$

Радиус "шара" для вектора ζ, очевидно, должен быть равен минимальной из величин полуосей эллипсоида (1.62) при максимальном значении $\|\xi\|_2^2$.

В соответствии с предположением (1.41) максимальное значение квадрата нормы вектора ξ задается равенством $\|\xi\|_2^2 = R_\xi^2$. Матрица $\mathbf{X}$ фиксирована, поэтому фигурирующие в правой части неравенства собственные значения λ_i, $i = \overline{1,M}$ являются константами. С другой стороны, из схемы доказательства утверждения 1 ясно, что при экстремальных значениях величины $\|\Delta \mathbf{c}\|_2^2$ квадраты косинусов могут принимать значения 0 или 1. Обозначив R_ζ^2 минимально возможное значение суммы в правой части при $\|\xi\|_2^2 = R_\xi^2$ можно записать:

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

Отсюда, в частности, следует, что $R_\zeta = \lambda_{\min}^{1/2} R_\xi$.

Теперь вернемся к исходному предположению (1.41). 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 справедливо следующее следствие (1.58).

Следствие утверждения 1.

Если

$$\xi \in \Xi, \quad \Xi = \left\{\xi: \quad (\xi^T\xi)^{1/2} = \|\xi\|_2 \le R_\xi = \text{Const}\right\}, \tag{1.63}$$

то

$$\|\Delta c\|_2^2 \le \lambda_{\min}^{-1}\|\xi\|_2^2 \le \lambda_{\min}^{-1}R_\xi^2. \tag{1.64}$$

Если

$$\zeta \in Z, \quad Z = \left\{\zeta: \quad (\zeta^T\zeta)^{1/2} = \|\zeta\|_2 \le R_\zeta = \text{Const}\right\}, \tag{1.65}$$

то

$$\|\Delta c\|_2^2 \le \lambda_{\min}^{-2}\|\zeta\|_2^2 \le \lambda_{\min}^{-2}R_\zeta^2, \tag{1.66}$$

где $\lambda_{\max}, \lambda_{\min}$ максимальное и минимальное среди собственных значений λ_i, а $R_\zeta = \lambda_{\min}^{1/2}R_\xi$.

Представляет интерес сопоставить приведенные неравенства с аналогичными оценками, которые используются в теоретической статистике. В фишеровской теории оценивания для проверки адекватности моделей широко используется следующее соотношение для вероятности выполнения неравенства:

$$P\left\{\Delta c^T K_N \Delta c \le 1\right\} = 1 - \alpha, \tag{1.67}$$

где $\Delta c = \hat{c} - c$, $K_N = X^T X / M\hat{s}_N^2 F_a$, $\hat{s}_N^2 = [y - X\hat{c}]^T[y - X\hat{c}]/(N - M)$ - оценка дисперсии, F_a - определяется из условия

$$P\{F \le F_a\} = 1 - \alpha,$$

а F распределение Фишера с M и N-M степенями свободы. Множество тех c, которые удовлетворяют неравенству в фигурных скобках (1.67), образуют замкнутую область - N-мерный эллипсоид с центром в точке $c = \hat{c}$. По определению этот эллипсоид является доверительной областью для параметров c. Считают [20,22], что требуемая точность с уровнем доверия 1-α достигнута, если доверительный эллипсоид оказывается внутри заданного или какая-либо его мера меньше априори заданной величины. Для выполнения условия "вложения" достаточно выполнения неравенства [22]

$$M\hat{s}_N^2 F_a \le \lambda_{\min}\rho^2, \tag{1.68}$$

где

$$\lambda_{\min} = \min_{i=1,M}\lambda_i(B), \quad B = R^{-1}X^TX, \text{ а } \rho^2 = [\hat{c} - c]^T R[\hat{c} - c].$$

Если условие "вложения" (1.68) выполняется, а R единичная матрица, т. е. $\rho^2 = \Delta\mathbf{c}^T\Delta\mathbf{c} = \|\Delta\mathbf{c}\|_2^2$, для квадрата евклидовой нормы вектора ошибок идентификации справедлива оценка:

$$\|\Delta\mathbf{c}\|_2^2 \le \lambda_{min}^{-1} M\hat{s}_N^2 F_a. \qquad (1.69)$$

Несмотря на поразительное сходство (1.69) с (1.64) (если формально задать $R_\xi^2 = M\hat{s}_N^2 F_a$, они совпадают) указанные оценки принципиально различны. Оценка (1.69) имеет вероятностную природу. Фигурирующий в правой части множитель F_a позволяет указать также вероятность, с которой достигается указанная точность. С использованием неравенства (1.64) это сделать невозможно, т.к. при его построении не использовалась априорная информация вероятностного свойства.

В заключение обсудим оценки, фигурирующие в утверждении 1 и его следствии. Из (1.64), в частности, следует, что оценка сверху для квадрата нормы вектора ошибок оценивания может быть улучшена, если осуществить преобразование данных, направленное на увеличение минимального собственного значения, которое будет сопровождаться уменьшением нормы $\|\xi\|_2$ вектора ошибок в исходных данных. Если все собственные значения одинаковы, т.е. $\lambda_{max}^{-1} = \lambda_{min}^{-1} = \lambda^{-1}$, оценки сверху и снизу (1.57) обязаны совпадать. При этом число обусловленности равно единице, а

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

Это равенство отражает известный факт: точность оценивания повышается при возрастании отношения полезный сигнал/шум. В частности, если $\lambda_i = 1$, $\overline{i = 1,M}$ норма вектора ошибок идентификации равна норме вектора ошибок в исходных данных.

Оценки (1.64) и (1.66) 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.

Соотношения (1.64) и (1.66) строились в расчете на наихудший случай, когда направление вектора ошибок совпадает с направлением собственного вектора, соответствующего минимальному ненулевому собственному значению

матрицы XX^T. В действительности вероятность того, что в конкретной идентификации реализуется неблагоприятное направление, будет значительно меньше единицы. Вероятностные характеристики величины и направления вектора ошибок могут иметь различную физическую природу. В частности, свойство статистической устойчивости некоторых направлений вектора ошибок может проявляться в ситуации, когда устойчивость самих распределений ошибок еще не имеет места. Эту информацию, если она имеется, конечно, следует использовать для повышения точности оценок.

В приведенных соотношениях всюду фигурируют векторы ошибок ξ и/или ζ, которые никогда не известны. Для построения практических решающих правил, конечно, могут использоваться лишь их оценки, построенные на основе анализа вектора невязок. Тем не менее, приведенные соотношения представляются полезными, т.к. они вскрывают закономерности формирования ошибок идентификации. В частности, из (1.64), (1.66) видно, что решающее значение при оценке ошибок идентификации имеют собственные значения матрицы A. Следующая, вторая, глава посвящена более глубокому исследованию связи обусловленности и информативности данных со спектром этой матрицы. В частности, с использованием приведенных неравенств будут построены более простые в вычислительном отношении оценки для характеристики информативности данных на фрагментах изображения.

Оценка информативности данных

Выше мы сформулировали общие принципы построения оценок по малому числу наблюдений. В частности, обосновали необходимость предварительного анализа информативности каждого фиксированного набора данных. В настоящей главе строятся учитывающие особенности задачи оценивания по малому числу наблюдений меры мультиколлинеарности и обусловленности и основанные на них методы оценки информативности данных на фрагментах изображений.

2.1. Понятие информативности данных, связь с обусловленностью задачи

В статистической теории оценивания в качестве меры информативности широко используется фишеровская информация. В частности, известна нижняя граница ковариаций оценок, определяемая как обратная информационная матрица Фишера [51]. В рамках сделанных предположений указанная мера информативности неприменима. Здесь мерой точности является евклидова норма вектора ошибок идентификации на конкретной реализации. Тем не менее, по аналогии с вероятностными подходами и в данном случае удобно использовать понятие информативности как характеристику *потенциальных возможностей* фиксированного набора данных, в частности, *достижимой точности оценивания параметров* на заданном фиксированном наборе данных.

В (1.64), (1.66) нижняя граница для погрешности идентификации (аналогичная неравенству Крамера-Рао) отсутствует. Однако это не может служить основанием для оптимизма. Если дополнительная информация об ориентации и величине вектора ошибок отсутствует, гарантированные оценки погрешности идентификации (1.64), (1.66) имеют место лишь в случае, когда указанные неравенства превращаются в равенства. Поэтому, говоря об информативности данных, как о свойстве, определяющем достижимую точность, мы можем (и будем) иметь в виду лишь потенциальную возможность уменьшения верхней границы для ошибок идентификации.

Информативность набора данных в указанном смысле определяется характеристиками обусловленности матрицы Грама. Плохая обусловленность, как правило (но не обязательно), связана с мультиколлинеарностью (сильной сопряженностью) векторов независимых переменных в уравнении (1.2). Поэтому анализ мультиколлинеарности и/или обусловленности задачи должен являться важной составной частью анализа информативности данных. В настоящем разде-

ле приводится обзор известных мер мультиколлинеарности и обусловленности и предлагаются новые меры, позволяющие более эффективно решать задачу оценки информативности данных по малому числу фиксированных наблюдений.

В предыдущей главе показано, что решающее значение при формировании ошибок идентификации имеют собственные значения информационной матрицы. Если собственные значения известны, то при заданной норме вектора ошибок в исходных данных погрешность идентификации может быть определена точно по соотношениям (1.37), (1.39). Однако на практике определение собственных значений представляет серьезные вычислительные трудности. Но даже если пойти на эти вычислительные затраты, дать уверенный прогноз погрешности идентификации все равно вряд ли удастся.

Дело в том, что если задача плохо обусловлена, а среди собственных значений имеются кратные, их вычисление становится серьезной проблемой. Возникает замкнутый круг: для того чтобы дать уверенный прогноз точности неоходимо знать собственные значения, а точность их вычисления вызывает сомнение до тех пор пока не установлены характеристики обусловленности задачи и структура спектра матрицы Грама.

В настоящем разделе мы обоснуем простую в вычислительном отношении и вместе с тем достаточно эффективную меру для оценки степени обусловленности, которая вычисляется непосредственно по элементам матрицы Грама. Эффективность достигается за счет учета специфических свойств этой матрицы. Прежде чем перейти к построению соответствующих оценок, проанализируем известные подходы к этой проблеме.

Если небольшие ошибки в исходных данных приводят к большим ошибкам в решениях, то говорят, что задача плохо обусловлена [38]. Одной из основных причин плохой обусловленности матрицы $\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ является "почти" линейная зависимость (мультиколлинеарность) [15] векторов - столбцов матрицы $\mathbf{X}$. Можно утверждать, что следствием сильной мультиколлинеарности всегда является плохая обусловленность задачи. Обратное не всегда верно. Например, векторы-столбцы матрицы $\mathbf{X}$ могут быть почти ортогональными (nonmulticollinearity), но сильно различаться параметрами масштаба, что неизбежно приведет к плохой обусловленности. Поэтому далее 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.

В традиционном 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 (а не корреляции) независимых переменных в уравнении (1.2). Поэтому использование этого понятия в рамках алгебраического подхода к анализу ошибок идентификации представляется также вполне уместным. Дадим сравнительную характеристику пяти мерам мультиколлинеарности, рассматривавшимся в работе [15].

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

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

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 $\tilde{\mathbf{A}} = \mathbf{A} / \mathrm{tr}\mathbf{A}$. При такой нормировке сумма собственных значений равна единице. Поэтому, если $\det(\tilde{\mathbf{A}}) \to 0$, то минимальное собственное значение также стремится к нулю и, следовательно, задача плохо обусловлена. Заметим, что при анализе обусловленности и достижимой точности по нормированной матрице необходимо учитывать также и связанное с указанной нормировкой матрицы изменение нормы вектора ошибок в исходных данных.

Если нормировка не делается, то указанная мера на множестве матриц может возрастать при стремлении минимального собственного значения к нулю за счет более быстрого роста максимального собственного значения. Следовательно, определитель как мера мультиколлинеарности может "не замечать" ухудшения обусловленности задачи.

2. *Число обусловленности* $K(\mathbf{A}) = \lambda_{max}/\lambda_{min}$. 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 специальная нормировка матрицы A, т.к. число обусловленности не чувствительно к параметру масштаба. Значение этой меры трудно переоценить. Она широко используется в теории возмущений для анализа ошибок в решениях [9]. В работе [36] показана связь между числом обусловленности и определите-

лем матрицы. В частности, приводится следующая оценка сверху числа обусловленности:

$$K(A) \le \left(\mathrm{tr}^{M} A\right)/\det A . \qquad (2.2)$$

3. *Минимальное собственное значение* $\lambda_{\min}(A)$ информационной матрицы $A=X^TX$. Утверждают [15], что эта мера мультиколлинеарности в некотором смысле является универсальной в абсолютном смысле. Такая точка зрения представляется обоснованной в свете приведенных в первой главе оценок сверху для нормы вектора ошибок идентификации (1.64), (1.66). Эта мера отражает как масштаб, зависящий от физической размерности независимых переменных, так и мультиколлинеарность (сопряженность) соответствующих им векторов.

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

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{bmatrix}, \qquad (2.3)$$

где $r_{ij} = \left(x_i^T x_j\right) \Big/ \left[\left(x_i^T x_i\right)^{1/2}\left(x_j^T x_j\right)^{1/2}\right] = \cos\left(\hat{x_i, x_j}\right)$, а x_i, x_j - i-й и j- й векторы-столбцы матрицы **X**. В качестве показателя мультиколлинеарности выступает величина

$$r = \max_{i,j} |r_{ij}|, \quad i \ne j. \qquad (2.4)$$

Эта мера является привлекательной с точки зрения вычислительной простоты. Однако, по сравнению с другими, она является наиболее слабой. В [15] приводится пример матрицы, составленной из трех попарно линейно независимых векторов, которые являются компланарными. Соответствующая 3×3 матрица $A= X^TX$ при этом является вырожденнной, а мера (2.4) "не замечает" этого.

5. *Показатель максимальной сопряженности.* В качестве меры мультиколлинеарности, свободной от указанного недостатка, рассматривают величину [15]

$$R = \max_i |R_i|, \qquad (2.5)$$

где $R_i^2 = x_i^T X_{M-1}\left[X_{M-1}^T X_{M-1}\right]^{-1} X_{M-1}^T x_i$, x_i i-й вектор-столбец матрицы **X**, а X_{M-1} N×(M-1) - матрица, полученная из исходной путем "вычеркивания" этого (i-го) столбца. Предполагается, что $\|x_i\| = 1$, $i = \overline{1,M}$. Геометрически R_i косинус угла

между вектором-столбцом x_i матрицы X и подпространством, натянутым на остальное множество M-1 векторов этой матрицы.

Указанная мера, наряду с минимальным собственным значением и числом обусловленности, является наиболее сильной. Если R≠1, гарантируется невырожденность задачи. Недостатком этой меры является вычислительная сложность, связанная с необходимостью вычисления обратной матрицы. Кроме того, возникает все тот же "замкнутый круг", о котором уже говорилось выше. Для того чтобы сделать заключение о характере обусловленности, необходимо знать точную обратную матрицу. В то же время, если матрица A плохо обусловлена, ее обращение становится серьезной проблемой.

Из приведенной краткой характеристики известных мер мультиколлинеарности следует, что использование определителя и показателя парной сопряженности связано с определенным риском. Их эффективное применение, по-видимому, возможно лишь в случае привлечения дополнительной информации вероятностного характера. Наиболее полно связь с мультиколлинеарностью отражают три меры: число обусловленности, минимальное собственное значение и показатель максимальной сопряженности. Две из них связаны с нахождением экстремальных собственных значений, а третья требует вычисления обратной матрицы. Вычисление собственных значений и/или обратной матрицы в ситуациях, когда задача плохо обусловлена, а среди собственных значений имеются кратные, представляет серьезные трудности. Поэтому даже если пойти на вычислительные затраты, связанные с решением проблемы собственных значений, это не придаст уверенности в надежности результата.

Выход заключается в том, чтобы вырабатывать оценку степени обусловленности задачи непосредственно по элементам матрицы A до начала вычислений, связанных с получением искомых оценок (или собственных значений, если они необходимы). Используемые при этом решающие правила должны обладать устойчивостью к изменению характера обусловленности и структуры спектра матрицы Грама. Для построения таких решающих правил необходимо выявить вычисляемые по элементам матрицы A характеристики, которые, с одной стороны, тесно связаны с ее собственными значениями, с другой стороны, обладают слабой чувствительность к ошибкам округления в ситуациях, когда задача плохо обусловлена.

2.2. Оценки для собственных значений матрицы Грама

Из приведенных в предыдущей главе канонических представлений (1.37), (1.39) видно, что при заданном векторе ошибок исходных данных решающую роль в формировании ошибок идентификации играют собственные значения информационной матрицы Грама $\mathbf{A} = \mathbf{X}^T\mathbf{X}$. Решение полной проблемы собственных значений является сложной вычислительной задачей. Поэтому представляет интерес, с учетом специфических свойств матриц Грама, построить для них простые достаточные оценки сверху и снизу. Для этого нам понадобится результат, который сформулируем в виде следующего утверждения.

Утверждение 2. Пусть $\Omega_k = \left\{\mathbf{A}: \mathrm{tr}\mathbf{A} = t_1, \left(\mathrm{tr}\mathbf{A}^2\right)^{1/2} = t_2, .., \left(\mathrm{tr}\mathbf{A}^k\right)^{1/k} = t_k\right\}$ - множество вещественных ограниченных матриц, образуемых по правилу $\mathbf{A} = \mathbf{X}^T\mathbf{X}$, среди собственных значений которых не менее k различающихся, а $t_j = \left(\mathrm{tr}\mathbf{A}^j\right)^{1/j}$, $j = \overline{1,k}$ фиксированные числа такие, что: $t_i > t_j$, при $i<j$ для всех $i,j=1,2,...,k$, $k<M$, $\mathrm{tr}\mathbf{A}$ - след матрицы A (RankA=M).

Тогда оценки сверху и снизу для собственных значений $\lambda_i = \lambda(\mathbf{A})$, $i = \overline{1,M}$ находятся среди положительных вещественных решений систем из k ($2 \le k < M$) уравнений вида:

$$\sum_{i=1}^{k} m_i \hat{\lambda}_i^j = t_j^j, \quad j = \overline{1,k}, \tag{2.6}$$

где m_i - натуральные числа, удовлетворяющие условию

$$\sum_{i} m_i = M, \quad i = \overline{1,k}. \tag{2.7}$$

Для доказательства утверждения 2 используется свойство собственных значений матрицы Грама:

$$\sum_{i=1}^{M} \lambda_i^j = \mathrm{tr}\mathbf{A}^j,$$

где j - произвольные натуральные числа. Рассматривая эти равенства в качестве ограничений, можно сформулировать задачу отыскания экстремума функции:

$$F(\lambda) = \lambda_M + \sum_{j=1}^{k} \kappa_j \left[\sum_{i=1}^{M} \lambda_i^j - t_j^j\right], \quad j = \overline{1,k} \quad \to \mathrm{extr},$$

где κ_j, $j = \overline{1,k}$ - неопределенные множители Лагранжа. Поскольку все собственные значения равноправны, из соображений удобства записи систем уравнений здесь принято λ_M.

Необходимые условия экстремума (максимума и минимума) выписываются в виде

$$\sum_{i=1}^{k} \kappa_j j\hat{\lambda}_i^{j-1} = 0, \quad i = \overline{1, M-1},$$

$$1 + \sum_{j=1}^{k} \kappa_j j\hat{\lambda}_M^{j-1} = 0,$$

$$\sum_{i=1}^{M} \hat{\lambda}_i^j - t_j^j = 0, \quad j = \overline{1, k}.$$

Последовательно исключая из этой системы неизвестные множители Лагранжа и анализируя возможные решения непосредственно устанавливаем, что среди них может быть любое число (но не более чем M-1) кратных. После исключения (k-1) множителей κ_j, $j = \overline{1, k-1}$ приходим к системе M+1 уравнений простого вида:

$$\left(\hat{\lambda}_i - \hat{\lambda}_{k-1}\right) \cdot k\kappa_k = 0, \quad i = \overline{k, M-1},$$

$$1 + k\kappa_k \prod_{i=1}^{k-1}\left(\hat{\lambda}_M - \hat{\lambda}_i\right) = 0, \tag{2.8}$$

$$\sum_{i=1}^{M} \hat{\lambda}_i^j = t_j^j, \quad j = \overline{1, k}. \tag{2.9}$$

Из уравнения (2.8) видно, что $\kappa_k \neq 0$ и $\hat{\lambda}_M \neq \hat{\lambda}_i$, $i = \overline{1, k-1}$. Это означает, что должны выполняться равенства $\hat{\lambda}_i = \hat{\lambda}_{k-1}$, $i = \overline{k, M-1}$, т.е., по крайней мере, M-k+1 (а возможно и больше) чисел $\hat{\lambda}_i$, $i = \overline{1, M}$ должны быть равны между собой, а кратность любого из них может быть любым целым числом таким, что сумма этих чисел равна точно M (случаи, когда число различающихся чисел $\hat{\lambda}_i$ меньше k мы обсудим особо). Решения указанной кратности должны удовлетворять k уравнениям (2.9). Отсюда сразу следует утверждение 2. ■

Систему уравнений (2.6) можно записать в следующем виде

$$\mathbf{W}\hat{\boldsymbol{\lambda}} = \mathbf{t}, \tag{2.10}$$

где

$$\mathbf{W} = \begin{bmatrix} m_1 & m_2 & \dots & m_k \\ m_1\hat{\lambda}_1 & m_2\hat{\lambda}_2 & \dots & m_k\hat{\lambda}_k \\ \dots & \dots & \dots & \dots \\ m_1\hat{\lambda}_1^{k-1} & m_2\hat{\lambda}_2^{k-1} & \dots & m_k\hat{\lambda}_k^{k-1} \end{bmatrix}$$

- матрица размерности k×k ($2 \le k \le M$). Участвующие в образовании элементов этой матрицы коэффициенты m_i, $i = \overline{1, k}$ удовлетворяют условиям (2.7),

$\hat{\lambda} = \left[\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k\right]^T$ - $k \times 1$ вектор оценок собственных значений матрицы **A**,

$t = \left[t_1, t_2^2, \ldots, t_j^j, \ldots, t_k^k\right]^T$, $t_j^j = trA^j$, $j = \overline{1,k}$ $k \times 1$ вектор, каждая компонента которого является следом различных степеней этой матрицы. Комбинируя различными способами целые положительные числа m_i, $i = \overline{1,k}$ так, чтобы они удовлетворяли условиям (2.7), можно построить различные матрицы **W**, которым будут соответствовать различные решения нелинейной системы алгебраических уравнений (2.10). Среди множества вещественных решений необходимо отобрать те, которые доставляют искомые максимум и минимум.

При k=M можно построить лишь одно уравнение вида (2.10), а матрица **W** является матрицей Вандермонда:

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_M \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_1^{M-1} & \lambda_2^{M-1} & \ldots & \lambda_M^{M-1} \end{bmatrix}.$$

В данном случае вместо оценок $\hat{\lambda}_i$ фигурируют сами собственные значения λ_i, $i = \overline{1,M}$. Если среди собственных значений отсутствуют кратные, эта матрица невырождена и существует решение системы нелинейных алгебраических уравнений (2.10).

Если исходная матрица $\mathbf{A} = \mathbf{X}^T\mathbf{X}$ имеет менее k различающихся собственных значений, то при задании компонент вектора **t** в правой части системы (2.10) с использованием этой матрицы так, как указано в утверждении 2, правая часть системы является линейной комбинацией менеее, чем k векторов. В результате для удовлетворяющего этой системе вектора $\hat{\lambda} = \left[\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k\right]^T = \left[\lambda_1, \lambda_2, \ldots, \lambda_k\right]^T$ (ввиду равенства некоторых компонент) среди столбцов матрицы **W** оказываются линейно зависимые.

Приступая к оценке обусловленности, мы не знаем имеются ли среди собственных значений кратные. Даже если среди собственных значений матрицы A более k различающихся, то вещественные положительные решения могут и не существовать. Кроме того, при увеличении числа уравнений k быстро растет также объем вычислительной работы. Для преодоления указанных трудностей целесообразно снижать размерность системы (2.6), (2.10). Для этого не обязательно

уменьшать старшую степень. Можно, например, взять два уравнения, одно из которых имеет первую степень, а второе j-ю, т.е. "пропустить" все степени со 2-й по (j-1)-ю:

$$L\hat{\lambda}_1 + (M - L)\hat{\lambda}_2 = t_1,$$
$$L\hat{\lambda}_1^j + (M - L)\hat{\lambda}_2^j = t_j^j.$$

Здесь L, j - любые натуральные числа, удовлетворяющие неравенствам: $L \le M-1$, $1<j<M$. Максимальное и минимальное числа в множестве вещественных положительных решений этой системы являются оценками сверху и снизу для собственных значений. При увеличении показателя степени j во втором уравнении объем вычислений также возрастает, однако есть основания ожидать получения более сильных оценок. Для получения вещественных решений необходимо потребовать выполнения неравенства $t_1 > t_j$ для любого $j<M$.

С точки зрения простоты практической реализации наибольший интерес представляют варианты, соответствующие небольшим значениям показателя степени j [45,46]. Рассмотрим наиболее простой частный случай, когда уравнения указанного множества систем содержат лишь первую и вторую степень:

$$\begin{aligned} L\hat{\lambda}_1 + (M - L)\hat{\lambda}_2 &= t_1, \\ L\hat{\lambda}_1^2 + (M - L)\hat{\lambda}_2^2 &= t_2, \end{aligned} \qquad (2.11)$$

где L - натуральные числа: $L \le M-1$. Допустимые решения этой системы (при $(M-1) < \left(t_1^2/t_2^2\right) \le M$) имеют вид

$$\hat{\lambda}_1 = M^{-1}t_1\left(1 \pm \sqrt{(Mt_2^2 - t_1^2)(M - L)L^{-1} \cdot t_1^{-2}}\right),$$

$$\hat{\lambda}_2 = M^{-1}t_1\left(1 \mp \sqrt{(Mt_2^2 - t_1^2)(M - L)^{-1}L \cdot t_1^{-2}}\right).$$

Наибольшее значение для максимального и наименьшее для минимального из чисел $\hat{\lambda}_1, \hat{\lambda}_2$ при фиксированных t_1, t_2, M достигаются при L=1 (или L=M-1). Следовательно, для получения гарантированных оценок сверху и снизу для собственных значений достаточно рассмотреть одну систему при L=1:

$$\hat{\lambda}_1 + (M - 1)\hat{\lambda}_2 = t_1,$$
$$\hat{\lambda}_1^2 + (M - 1)\hat{\lambda}_2^2 = t_2^2.$$

Эта система имеет простое аналитическое решение, из которого вытекает важное следствие.

Следствие утверждения 2. Пусть матрица **A** удовлетворяет условиям утверждения 2 при k=2, т.е. $\mathbf{A} \in \Omega_2$, $\Omega_2 = \left\{\mathbf{A}: \mathrm{trA} = t_1, \left(\mathrm{trA}^2\right)^{1/2} = t_2, t_1 > t_2\right\}$ и кроме того для величины

$$\phi = t_1^2/t_2^2 = \mathrm{tr}^2\mathbf{A}/\mathrm{trA}^2 \tag{2.12}$$

выполняются неравенства

$$M - 1 < \phi \le M. \tag{2.13}$$

Тогда справедливы следующие оценки сверху и снизу для собственных значений:

$$\max_{\mathbf{A}\in\Omega_2} \lambda(\mathbf{A}) = M^{-1}\mathrm{trA}\left(1 + \sqrt{(M/\phi - 1)(M-1)}\right), \tag{2.14}$$

$$\min_{\mathbf{A}\in\Omega_2} \lambda(\mathbf{A}) = M^{-1}\mathrm{trA}\left(1 - \sqrt{(M/\phi - 1)(M-1)}\right). \tag{2.15}$$

Из соотношений (2.14), (2.15) видно, что оценки для собственных значений при фиксированной размерности задачи зависят лишь от характеристики (2.12), которую назовем *показателем диагонального преобладания* матрицы **A**. Обратим внимание, что точность вычисления этого показателя не зависит от характера обусловленности и кратности собственных значений.

Для практических целей наибольший интерес, конечно, представляет оценка снизу (2.15), т.к. она может использоваться для оценки верхней границы возможной ошибки идентификации при заданном в (1.41), (1.43) предельно допустимом значении нормы вектора ошибок исходных данных. Действительно, неравенство (1.64) не нарушится, если вместо минимального собственного значения $\lambda_{\min} = \min_i \lambda_i(\mathbf{A})$ подставить его оценку (2.15). При этом получаем следующую легко вычисляемую оценку сверху для погрешности идентификации:

$$\|\Delta c\|_2^2 \le M^{-1}\left[\mathrm{trA}\left(1 - \sqrt{(M/\phi - 1)(M-1)}\right)\right]^{-1} R_\xi^2, \tag{2.16}$$

где $R_\xi = \|\xi\|_2$ - заданная в (1.41) скалярная характеристика максимальной погрешности исходных данных.

Легко заметить, что если матрица **A** нормирована по правилу $\tilde{\mathbf{A}} = \mathbf{A}/\mathrm{trA}$, то оценка максимальной погрешности идентификации зависит только от погрешности исходных данных и величины диагонального преобладания информационной матрицы. Если указанной нормировке подвергается сформированный набор дан-

ных, для которого величина K_ξ уже задана, то она также должна быть изменена соответствующим образом.

Тот факт, что оценка для минимального собственного значения может быть построена с использованием достаточно просто вычисляемого по элементам матрицы Грама показателя (2.12) представляется крайне важным. Поэтому естественно возникает вопрос: нет ли других более эффективных оценок или таких же по эффективности, но требующих меньших вычислительных затрат? Для ответа на этот вопрос сравним построенные оценки с наиболее сильными известными локализационными результатами Гершгорина и Брауэра [25,52].

В соответствии с теоремой Гершгорина характеристические числа комплексной квадратной матрицы A порядка M лежат в замкнутой области комплексной плоскости, образованной кругами

$$|z - a_{i,i}| \le P_i, \quad i = \overline{1,M},$$

где

$$P_i = \sum_{j=1, j\neq i}^{M} |a_{i,j}|.$$

В частном случае матрицы Грама характеристические числа вещественны, а для минимального собственного значения имеет место оценка

$$\lambda_{min} \ge \min_i (a_{i,i} - P_i). \tag{2.17}$$

Эта оценка существует, если для всех строк матрицы выполняются неравенства

$$|a_{i,i}| > \sum_{j=1, j\neq i}^{M} |a_{i,j}|, \quad i = \overline{1,M}, \tag{2.18}$$

которые по существу и определяют эффективность метода.

Более сильным локализационным методом являются оценки Брауэра [52]:

$$\lambda_{max} = \max_{k,l} \frac{1}{2}\left(|a_{k,k}| + |a_{l,l}| + \sqrt{\left(|a_{k,k}| - |a_{l,l}|\right)^2 + 4P_kP_l}\right), \tag{2.19}$$

$$\lambda_{min} = \min_{k,l} \frac{1}{2}\left(|a_{k,k}| + |a_{l,l}| - \sqrt{\left(|a_{k,k}| - |a_{l,l}|\right)^2 + 4P_kP_l}\right). \tag{2.20}$$

В данном случае условие существования оценки снизу для минимального собственного значения заключается в требовании выполнения для всех возможных комбинаций пар строк (или столбцов) матрицы A следующих неравенств

$$|a_{k,k}| + |a_{l,l}| - \sqrt{\left(|a_{k,k}| - |a_{l,l}|\right)^2 + 4P_kP_l} > 0. \tag{2.21}$$

Для сравнения эффективности построенных выше оценок (2.14), (2.15) с оценками Гершгорина и Брауэра был проведен следующий численный эксперимент. С использованием задаваемых случайным образом базисов формировались 2000 различных матриц размером 4×4. Для половины из них (1000 матриц) спектр задавался таким образом, чтобы существовала оценка снизу (2.15) для минимального собственного значения. Конкретные числовые значения собственных значений в этом эксперименте были следующими λ_1=1, λ_2=3.26, λ_3=5.52, λ_4=7.78. Соответствующий им показатель диагонального преобладания (2.12) равен 3.0031. Для другой половины (1000 матриц) со спектром λ_1=1, λ_2=3.3, λ_3=5.6, λ_4=7.9 оценка снизу (2.15) не существовала, т. к. показатель диагонального преобладания в этом случае равен 2.998675. Для каждой из указанных 2000 матриц проверялись неравенства (2.18) и (2.21), характеризующие возможность построения оценок с использованием локализационных теорем Гершгорина и Брауэра. Результаты испытаний приведены в таблице 2.1.

Из таблицы видно, что число матриц, для которых можно дать оценку (2.17), основанную на теореме Гершгорина, значительно менее числа случаев существования оценок (2.15), вычисляемых по показателю диагонального преобладания (2.12). Число же случаев "срабатывания" оценок Брауэра (2.20) приблизительно равно числу случаев существования оценок (2.15). Для окончательного заключения о преимуществе одного из методов проведем сравнительный анализ их вычислительной сложности.

Таблица 2.1. Сравнение эффективности оценок

Спектр матриц и показатель дигонального преобладания	Число матриц	Число случаев существования оценок:		
		(2.15))	(2.17)	(2.20)
λ_1=1, λ_2=3.26, λ_3=5.52, λ_4=7.78, ϕ=3.0031	1000	1000	177	468
λ_1=1, λ_2=3.3, λ_3=5.6, λ_4=7.9, ϕ=2.998675	1000	-	186	498
Итого:	2000	1000	363	966

Оценка, основанная на теореме Гершгорина, проста в вычислительном отношении. Однако, как видно из таблицы 2.1, она уступает в эффективности как оценке Брауэра (2.20), так и оценке (2.15), построенной по показателю диагонального преобладания. Поэтому сравнение вычислительной сложности имеет смысл лишь для двух, приблизительно равных по эффективности, оценок: (2.15) и (2.20). В таблице 2.2 приведены формулы для определения числа различных

арифметических операций, необходимых при вычисления оценок сверху и снизу для собственных значений (2.14), (2.15) и оценок Брауэра - (2.19), (2.20).

В последней колонке таблицы приведены выражения для разности между числом операций, необходимых для вычисления оценок по методу Брауэра и с помощью соотношений (2.14), (2.15). Там же приведены неравенства, указывающие порядок M матрицы **A**, начиная с которого эти разности положительны.

Таблица 2.2. Сравнение вычислительной сложности оценок

Тип операции	Оценки сверху и снизу для λ_{max} и λ_{min}:		Разница числа вычислительных операций
	по соотношениям (2.14),(2.15);	По методу Брауэра (2.19), (2.20)	
Умножение, деление	$(M^2+M+10)/2$	$(3M^2-3M+2)/2$	M^2-2M-4, $M>3$
Извлечение квадратного корня	1	$(M^2-M)/2$	$(M^2-M)/2-1$, $M>2$
Сложение	M^2+2M	$5M^2-5M$	$4M^2-7M$, $M>2$
Логические		$2M^2-2M$	$2M^2-2M$, $M>2$

Из приведенных соотношений видно, что выигрыш в вычислительной сложности гарантирован при M>3 и этот выигрыш быстро растет с ростом размерности задачи. Таким образом, оценки для собственных значений (2.14), (2.15) более эффективны по сравнению с оценками Гершгорина, а при одинаковой эффективности с оценками Брауэра оказываются более простыми, по сравнению с ними, в вычислительном отношении.

2.3. Оценки для числа обусловленности и определителя

Собственные значения информационной матрицы, конечно, играют большое значение при формировании ошибок идентификации. Однако их числовые значения зависят не только от сопряженности, но и от физической размерности (масштаба) независимых переменных. Если требуется строить оценки степени мультиколлинеарности, свободные от влияния параметра масштаба, то в качестве меры применяют спектральное число обусловленности (1.49):

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

Для определения его точного значения необходимо решать частную проблему собственных значений. С вычислительной точки зрения это не всегда приемлемо, а при плохой обусловленности задачи возможны большие погрешности. Поэтому представляет интерес построение простых в вычислительном отношении и

в то же время достаточно эффективных оценок для числа обусловленности, которые бы вычислялись непосредственно по элементам информационной матрицы.

Оценки сверху и снизу для спектрального числа обусловленности, вообще говоря, могут быть построены с использованием полученных выше оценок сверху и снизу для собственных значений. В частности, подставляя в (1.49) вместо λ_{max}, λ_{min} оценки (2.14), (2.15) соответственно, получаем следующую оценку сверху [45]:

$$k(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \le \hat{k}_2^{в} = \frac{1+\sqrt{(M/\phi - 1)(M-1)}}{1-\sqrt{(M/\phi - 1)(M-1)}}. \qquad (2.22)$$

Соображения, по которым мы обозначили эту оценку $\hat{k}_2^{в}$, будут понятны из дальнейшего.

Использовавшиеся для построения этого неравенства оценки определены на множестве $\Omega_2 = \left\{A: trA = t_1, \left(trA^2\right)^{1/2} = t_2, t_1 > t_2\right\}$. Сама же оценка (2.22) справедлива на более широком, чем Ω_2, классе матриц $A \in \Phi$, $\Phi = \left\{A: tr^2A/trA^2 = \phi = Const\right\}$. В связи с этим возникает вопрос: не существует ли на указанном классе Ф других более сильных оценок? Для ответа на этот вопрос сформулируем и решим задачу отыскания экстремумов непосредственно для условного числа K(A), определяемого отношением λ_i/λ_j, $i,j = \overline{1,M}$, $i \ne j$.

В классе $\Phi = \left\{A: tr^2A/trA^2 = \phi = Const\right\}$ матриц Грама $A=X^TX$ рассмотрим подмножество $\Omega_2 = \left\{A: trA = t_1, \left(trA^2\right)^{1/2} = t_2, t_1 > t_2\right\}$. Каждому элементу из допустимого в Ω_2 множества можно поставить в соответствие точку некоторого множества H евклидова пространства с координатами $(\lambda_1, \lambda_2, ..., \lambda_M)$, где λ_i, $i = \overline{1,M}$ собственные значения матрицы A. Отображение множества Ω_2 на множество H представляет собой некоторую замкнутую связную область. В силу свойства непрерывной зависимости собственных значений матрицы от ее элементов, в каждой точке области H определена и непрерывна функция

$$K(A) = \lambda_1/\lambda_M,$$

где λ_1, λ_M удовлетворяют неравенствам

$$\lambda_i > 0, \quad i = \overline{1,M}, \qquad (2.23)$$

$\lambda_{k+1} - \lambda_k \geq 0, \quad k = \overline{1, M-1}.$ (2.24)

Поскольку характеристические числа невырожденной матрицы Грама положительны, неравенства (2.24) гарантируют выполнение требования $\lambda_1 \geq \lambda_M$, что существенно при определении числа обусловленности.

Множество H компакт, поэтому функция K(A) имеет на нем максимум и минимум. С учетом неравенств-ограничений (2.23), (2.24) задачу отыскания экстремальных значений этой функции на множестве Ω_2 можно сформулировать как общую задачу нахождения седловой точки функции Лагранжа. Однако с методической точки зрения целесообразно решить более простую задачу отыскания максимумов и минимумов функции Лагранжа

$$F(\lambda_1, .., \lambda_M, \chi, \kappa) = \frac{\lambda_1}{\lambda_M} - \chi\left(\mathrm{tr}A - \sum_{i=1}^{M}\lambda_i\right) - \kappa\left(\mathrm{tr}A^2 - \sum_{i=1}^{M}\lambda_i^2\right), \qquad (2.25)$$

учитывающей лишь ограничения-равенства и отобрать среди них допустимые.

Необходимые условия экстремума в задаче (2.25):

$$\hat{\lambda}_M^{-1} - \chi - \kappa 2\hat{\lambda}_1 = 0,$$

$$-\chi - \kappa 2\hat{\lambda}_k = 0, \quad k = \overline{2, M-1},$$

$$-\frac{\hat{\lambda}_1}{\hat{\lambda}_M^2} - \chi - \kappa 2\hat{\lambda}_M = 0,$$

$$t_1 - \sum_{i=1}^{M}\hat{\lambda}_i = 0,$$

$$t_2^2 - \sum_{i=1}^{M}\hat{\lambda}_i^2 = 0,$$

Исключая из первых M уравнений неопределенные множители Лагранжа χ, κ устанавливаем, что M-2 чисел $\hat{\lambda}_k$, удовлетворяющих этой системе, обязаны совпадать. В результате приходим к простой системе из трех уравнений:

$$\lambda_1^2 + \lambda_M^2 = \lambda_k(\lambda_1 + \lambda_M),$$

$$\lambda_1 + \lambda_M = t_1 - (M-2)\lambda_k,$$

$$\lambda_1^2 + \lambda_M^2 = t_2^2 - (M-2)\lambda_k^2,$$

допустимое решение которой доставляет искомый экстремум (максимум). Минимум достигается на границе множеества H при равенстве M-1 (больших по величине) характеристических чисел. Заметим, что хотя функция Лагранжа (2.25) задавалась на множестве H, соответствующем множеству Ω_2, полученные оценки

не зависят от величин t_1, t_2 и являются функцией только величины ϕ, т.е. справедливы для более широкого, по сравнению с Ω_2, класса Φ. Таким образом, справедливо следующее утверждение. ■

Утверждение 3. Пусть A – невырожденная ограниченная матрица Грама порядка M такая, что $A \in \Phi$, $\Phi = \left\{A: \mathrm{tr}^2 A / \mathrm{tr} A^2 = \phi = \mathrm{Const}\right\}$.

Тогда для спектрального числа обусловленности этой матрицы справедливы оценки

$$\max_{A \in \Phi} k(A) = \hat{k}_1^{в} = \frac{1 + \sqrt{(\phi - M + 2)^{-1}(M - \phi)}}{1 - \sqrt{(\phi - M + 2)^{-1}(M - \phi)}}, \tag{2.26}$$

$$\min_{A \in \Phi} k(A) = \hat{k}_1^{н} = \frac{1 + \sqrt{(M\phi^{-1} - 1)(M - 1)}}{1 - \sqrt{(M\phi^{-1} - 1)/(M - 1)}}. \tag{2.27}$$

где ϕ - показатель диагонального преобладания (2.12), удовлетворяющий неравенствам (2.13). Совпадающую с (2.27) оценку снизу для числа обусловленности можно также получить из решений системы (2.11).

Для построения гарантированных оценок для относительных ошибок идентификации по соотношениям (1.50) - (1.53) должны использоваться лишь оценки сверху (2.26). Поэтому сравнительный анализ эффективности имеет смысл провести лишь для оценок (2.22), (2.26).

На рис. 2.1, а приведены графики изменения оценок $\hat{k}_1^{в}$, $\hat{k}_2^{в}$ в функции приведенной к диапазону [0-1] величины диагонального преобладания $\Phi = \phi - M + 1$. Соответствующие соотношениям (2.22) и (2.26) при $\phi = \Phi + M - 1$ выражения для оценок $\hat{k}_1^{в}$, $\hat{k}_2^{в}$, по которым проводились расчеты, имеют вид

$$\hat{k}_1^{в} = \frac{1 + \sqrt{(1 - \Phi)/(1 + \Phi)}}{1 - \sqrt{(1 - \Phi)/(1 + \Phi)}},$$

$$\hat{k}_2^{в} = \frac{1 + \sqrt{(1 - \Phi)(M - 1)(\Phi + M - 1)^{-1}}}{1 - \sqrt{(1 - \Phi)(M - 1)^{-1}(\Phi + M - 1)^{-1}}}.$$

Из графиков видно, что оценка $\hat{k}_1^{в}$ более сильная, чем $\hat{k}_2^{в}$, но преимущество незначительное и притом быстро убывает с ростом показателя Ф. Более того, хотя оценка $\hat{k}_2^{в}$ 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.

В заключение приведем также оценки, устанавливающие границы для определителя матрицы Грама в зависимости от величины ее диагонального преобладаниия. Конечно, определитель, как мера обусловленности, имеет серьезные недостатки, на которые мы уже указывали выше. Тем не менее, эти оценки могут использоваться для оценки обусловленности нормированной матрицы $\tilde{\mathbf{A}} = \mathbf{A} / \mathrm{tr}\mathbf{A}$. Кроме того, они могут представлять самостоятельный интерес в других задачах.

Утверждение 4. Для определителя матрицы $\mathbf{A} \in \Omega_2$, удовлетворяющей условиям следствия утверждения 2, справедливы оценки

$$\max_{\mathbf{A}\in\Omega_2} \det \mathbf{A} = \left(\frac{\mathrm{tr}\mathbf{A}}{M}\right)^M \left[1 + \sqrt{(M\phi^{-1} - 1)(M-1)}\right]\left[1 - \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{M-1}, \quad (2.28)$$

$$\min_{\mathbf{A}\in\Omega_2} \det \mathbf{A} = \left(\frac{\mathrm{tr}\mathbf{A}}{M}\right)^M \left[1 - \sqrt{(M\phi^{-1} - 1)(M-1)}\right]\left[1 + \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{M-1}, \quad (2.29)$$

где ϕ - показатель диагонального преобладания (2.12), удовлетворяющий неравенствам (2.13).

Доказательство утверждения 4 сводится к отысканию в допустимой области экстремумов функции

$$F(\lambda_1,..,\lambda_M,\chi,\kappa) = \prod_{i=1}^{M}\lambda_i - \chi\left(t_1 - \sum_{i=1}^{M}\lambda_i\right) - \kappa\left(t_2^2 - \sum_{i=1}^{M}\lambda_i^2\right), \quad (2.30)$$

где χ, κ - неопределенные множители Лагранжа, $t_1 = \mathrm{tr}A$, $t_2^2 = \mathrm{tr}A^2$ Необходимые условия экстремума задаются системой M+2 уравнений вида

$$\prod_{\substack{i=1,\\ i\neq k}}^{M}\lambda_i - \chi - \kappa 2\hat{\lambda}_k = 0, \quad k = \overline{1,M},$$

$$t_1 - \sum_{i=1}^{M}\hat{\lambda}_i = 0,$$

$$t_2^2 - \sum_{i=1}^{M}\hat{\lambda}_i^2 = 0,$$

Исключая из первых M уравнений неопределенные множители Лагранжа χ, κ, с учетом неравенств (2.23) приходим к системе

$$(\lambda_1 - \lambda_k)(\lambda_1 - \lambda_k) = 0, \quad k = \overline{3,M},$$

$$t_1 - \sum_{i=1}^{M}\hat{\lambda}_i = 0,$$

$$t_2^2 - \sum_{i=1}^{M}\hat{\lambda}_i^2 = 0,$$

из которой видно, что M-1 чисел $\hat{\lambda}_i$, $i = \overline{1,M}$, удовлетворяющих этой системе обязаны совпадать. С учетом этого из последних двух уравнений сразу следует утверждение 4. ■

По свойству определителей для нормированной матрицы $\tilde{\mathbf{A}} = \mathbf{A} / \mathrm{tr}\mathbf{A}$ имеет место равенство $\det \tilde{\mathbf{A}} = \det \mathbf{A} / \mathrm{tr}^M \mathbf{A}$. С учетом того, что величина показателя диагонального преобладания ϕ не зависит от нормировки матрицы, из соотношений (2.28), (2.29) непосредственно вытекает следующее следствие.

Следствие утверждения 4. Пусть $\mathbf{A} \in \Omega_2$- матрица, удовлетворяющая условиям утверждения 4, тогда для определителя нормированной матрицы $\tilde{\mathbf{A}} = \mathbf{A} / \mathrm{tr}\mathbf{A}$ справедливы оценки:

$$\max_{\mathbf{A} \in \Omega_2} \det \tilde{\mathbf{A}} = M^{-1}\left[1 + \sqrt{(M\phi^{-1} - 1)(M-1)}\right]\left[1 - \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{M-1}, \quad (2.31)$$

$$\min_{\mathbf{A} \in \Omega_2} \det \tilde{\mathbf{A}} = M^{-1}\left[1 - \sqrt{(M\phi^{-1} - 1)(M-1)}\right]\left[1 + \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{M-1}, \quad (2.32)$$

С использованием равенств (2.31), (2.32) можно построить оценки для числа обусловленности, отличающиеся от полученных ранее. Обратимся к соотношению (2.2), характеризующему связь числа обусловленности и определителя: $K(\mathbf{A}) \le \left(\mathrm{tr}^M \mathbf{A}\right) / \det \mathbf{A}$. С учетом указанного выше свойства определителей его можно представить в виде $K(\mathbf{A}) \le 1 / \det \tilde{\mathbf{A}}$. Неравенство не нарушится, если вместо $\det \tilde{\mathbf{A}}$ в него подставить оценку снизу (2.32). При этом получаем следующую оценку

$$K(\mathbf{A}) = \hat{K}_3^{\mathrm{B}} \le M^M\left[1 - \sqrt{(M\phi^{-1} - 1)(M-1)}\right]^{-1}\left[1 + \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{1-M}. \quad (2.33)$$

Вместо величины $\mathrm{tr}\mathbf{A}$ в числителе неравенства (2.2) можно использовать оценку для максимального собственного значения, доставляемую формулой (2.14). При этом получается еще одна оценка для числа обусловленности:

$$K(\mathbf{A}) \le \hat{K}_4^{\mathrm{B}}(\mathbf{A}) = \frac{\left[1 + \sqrt{(M\phi^{-1} - 1)(M-1)}\right]^M}{\left[1 - \sqrt{(M\phi^{-1} - 1)(M-1)}\right]\left[1 + \sqrt{(M\phi^{-1} - 1)/(M-1)}\right]^{M-1}}. \quad (2.34)$$

Напомним, что все выписанные выше оценки сверху для числа обусловленности могут быть вычислены лишь в диапазоне $M-1 < \phi \le M$ $(0 < \Phi \le 1)$.

сравнения с оценками k_1^* приведены графики зависимостей оценок $\hat{k}_3^в$, $\hat{k}_4^в$ от приведенного к диапазону [0,1] показателя диагонального преобладания $\Phi = \phi - M + 1$ для матриц порядка M=10. Из графиков видно, что эти оценки слабее оценок $\hat{k}_1^в$, $\hat{k}_2^в$ и их эффективность быстро падает с ростом порядка M. Поэтому в контексте рассматриваемой задачи они не могут рекомендоваться для практического использования. Тем не менее, как оценки для определителя (2.28), (2.29), (2.31), (2.32), так и построенные на их основе оценки для числа обусловленности (2.33), (2.34) могут представлять самостоятельный интерес и применяться там, где это необходимо.

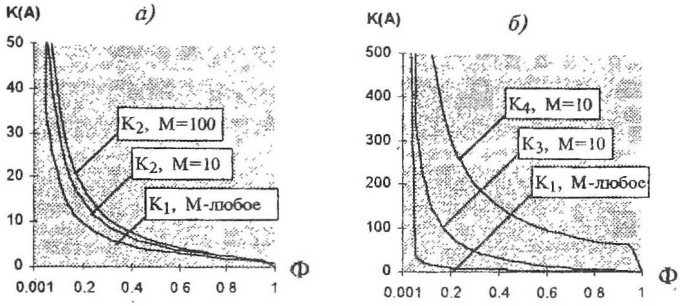


Рис. 2.1. Зависимости оценок числа обусловленности от показателя Ф:
а - оценки $к_1$, $к_2$; б - оценки $к_1$, $к_3$,$к_4$

В предыдущем разделе было показано, что метод построения оценок для собственных значений, основанный на использовании показателя диагонального преобладания, более эффективен, по сравнению с локализационным методом Гершгорина, и обладает вычислительными преимуществами, по сравнению с равным ему по эффективности методом Брауэра (таблицы 2.1 и 2.2). С другой стороны, из выражения (2.26) для наиболее сильной достаточной оценки сверху для числа обусловленности $\hat{k}_1^в = k(M,\Phi)$ видно, что условия ее существования совпадают с условиями (2.13) существования оценок снизу для собственных значений (2.15). На этом основании мы можем утверждать, что сформулированные в разделе 2.2 на основе сравнительных экспериментальных исследований выводы относительно эффективности и вычислительной сложности оценок для собственных значений (2.14), (2.15) в равной мере относятся и к оценке сверху для числа обусловленности (2.26).

Это имеет решающее значение при выборе оценки (2.26) для построении решающих правил контроля информативности данных на фрагментах изображений. С ее использованием на основе неравенств (1.45) и (1.5))-(1.53) можно построить оценки для относительной погрешности идентификации моделей искажающих систем в классе КИХ-фильтров и/или БИХ-фильтров. В частности, справедливые для указанных моделей неравенства (1.52)-(1.53) не нарушатся, если вместо числа обусловленности K(A) подставить его оценку сверху $\hat{k}_1^*$. При этом получаем

$$\delta_c \le K(A)\delta_b = \frac{1+\sqrt{(1-\Phi)/(1+\Phi)}}{1-\sqrt{(1-\Phi)/(1+\Phi)}}\delta_b, \qquad (2.35)$$

где

$$\Phi = \phi - M + 1, \quad \Phi \in [0,1].$$

Обращает на себя внимание вычислительная простота этой оценки. Аналогичные оценки могут быть построены с использованием соотношений (1.45), (1.51). Неравенство (2.35) указывает на существование, возможно более глубокой, чем это может показаться на первый взгляд [55], связи погрешности идентификации с диагональным преобладанием матрицы Грама. Поэтому в следующих разделах мы подвергнем этот показатель более детальному исследованию.

2.4. Оценка обусловленности и мультиколлинеарности по диагональному преобладанию матрицы Грама

В приведенных выше оценках для собственных значений, числа обусловленности и определителя всюду фигурирует характеристика матрицы Грама $A=X^TX$, так называемый показатель диагонального преобладания:

$$\phi = \frac{tr^2 A}{trA^2} = \left(\sum_{i=1}^{M}\lambda_i\right)^2 \Big/ \sum_{i=1}^{M}\lambda_i^2 = \left(\sum_{i=1}^{M}a_{ii}\right)^2 \Big/ \sum_{i,j=1}^{M}a_{ij}^2, \quad M-1 \le \phi \le M. \qquad (2.36)$$

Это не удивительно, т.к. все они строились на множестве Ω_2. В настоящем разделе мы построим неравенства, позволяющие осуществлять оценку возможности решения задачи идентификации с требуемой точностью минуя оценку числа обусловленности и/или минимального собственного значения, т.е. непосредственно по величине диагонального преобладания матрицы Грама ϕ.

Наряду с показателем диагонального преобладания ϕ, определяемым равенством (2.36), будем рассматривать приведенную к интервалу $[0;1]$ величину $\Phi=\phi-M+1$, характеризующую превышение диагонального преобладания ϕ над левой

границы допустимых значений в (2.13). Для установления связи показателя Ф с погрешностью идентификации вначале построим соотношения, связывающие этот показатель с допустимыми значениями минимального собственного значения и числа обусловленности. Для этого будем использовать наиболее сильные из полученных выше оценок: (2.26) для числа обусловленности и (2.15) для собственных значений.

В частности, подставив в неравенство (2.15) вместо показателя ϕ правую часть равенства $\phi=\Phi+M-1$ и осуществив формальную замену λ_{min} на $\lambda_{доп}$, после несложных преобразований получаем следующее неравенство

$$\Phi \geq \frac{\lambda_{доп}(M-1)(2\mathrm{tr}A - \lambda_{доп}M)}{\mathrm{tr}^2A - 2\mathrm{tr}A\lambda_{min} + \lambda^2_{доп}M}, \qquad (2.37)$$

где

$$\lambda_{доп} = R^2_\xi / \|\Delta c\|^2_2 \quad \text{или} \quad \lambda_{доп} = \left(R^2_\xi / \|\Delta c\|^2_2\right)^{1/2}$$

устанавливаются на основе неравенств (1.64), (1.65).

Для нормированной по правилу $\tilde{A} = A / \mathrm{tr}A$ матрицы $\tilde{A}$ как частный случай из неравенства (2.37) получаем более простое условие:

$$\Phi \geq \frac{\lambda_{доп}(M-1)(2-\lambda_{доп})}{M - 2\lambda_{доп} + \lambda^2_{доп}}. \qquad (2.38)$$

Аналогичное неравенство можно также построить с использованием наиболее сильной из полученных выше оценок для числа обусловленности. Для этого произведем в (2.26) формальную замену $\hat{k}^B_1$ на $K_{доп}(A)$. После соответствующих преобразований получаем условие

$$\Phi \geq \frac{2K_{доп}(A)}{K^2_{доп}(A)+1}, \qquad (2.39)$$

где

$$K_{доп}(A) = \frac{\delta_c}{\delta_A + \delta_b + \delta_c}$$

устанавливается непосредственно по соотношению (1.45), а $\delta_A = \|\delta A\|/\|A\|$, $\delta_b = \|\delta b\|/\|b\|$ заданные относительные погрешности исходных данных, $\delta_c = \|\Delta c\|/\|c\|$ - допустимая относительная погрешность оценивания.

На рис. 2.2 показаны области допустимых значений для величины Ф, границы $\Phi=\Phi(\lambda_{доп})$ и $\Phi=\Phi(K_{доп})$ которых построены по соотношениям (2.38) и (2.39). При значениях показателя Ф выше указанных кривых погрешность идентификации не превышает допустимого значения. На рис. 2.2, б показано

несколько кривых для M=2,3,4,5. При фиксированном $\lambda_{доп}$ ординаты кривых растут с ростом M. На основе этих графиков или непосредственно по соотношениям (2.38), (2.39) может быть определено допустимое значение приведенной величины диагонального преобладания для каждого заданного $\lambda_{доп}$ или $K_{доп}(A)$.

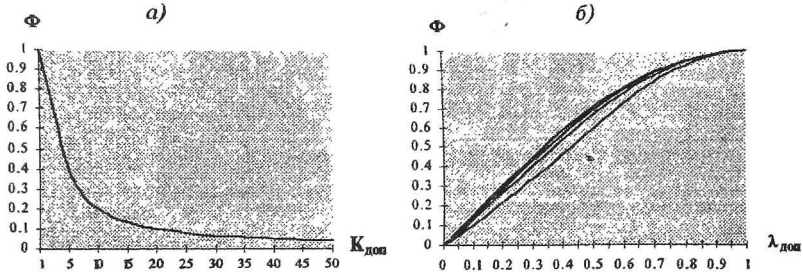


Рис. 2.2. Связь показателя Ф с допустимыми мерами: а - $K_{доп}$; б - $\lambda_{доп}$.

Приведенные соотношения и графики являются важным доказательством тесной связи показателя диагонального преобладания с известными мерами мультиколлинеарности и/или обусловленности. Однако с точки зрения практической реализации алгоритмов контроля информативности данных целесообразно пойти дальше в устранении промежуточных звеньев в общей схеме оценки обусловленности матрицы. В частности, построим зависимости, связывающие допустимые значения величин Ф и ϕ непосредственно с допустимыми погрешностями идентификации.

Для этого исключим в (2.37) $\lambda_{доп}$, в частности, заменим его значением выраженным в соответствии с (1.64) через $\|\Delta c\|_2^2$ и R_ξ^2. В результате получаем следующую, несложную в вычислительном отношении, оценку:

$$\Phi \geq \frac{M-1}{M\delta-1}, \qquad (2.40)$$

где

$$\delta = \|\Delta c\|^4 \Big/ \left(2\|\Delta c\|^2 R_\xi^2 - R_\xi^4\right).$$

Величина δ может быть рассчитана заранее по априорно заданной величине R_ξ^2 и допустимому значению нормы вектора ошибок идентификации. Заметим, что величина δ может изменяться в пределах $M^{-1} < \delta \leq 1$, а требование $\delta = 1$ (при этом Ф=1) влечет за собой требование $\|\Delta c\|^2 = \|\xi\|^2$. Это, как и следовало ожидать, мо-

жет иметь место в случае, когда все собственные значения одинаковы: $\lambda_{max} = \lambda_{min} = \lambda_i = 1$, $i = \overline{1,M}$, а $K(A)=1$.

Аналогично предыдущему, подставляя в неравенство (2.39) фигурирующее там же выражение для $K_{дом}(A)$, получаем, также простую в вычислительном отношении, оценку:

$$\Phi \geq \frac{2\left(\delta_\Sigma\,\delta_c + \delta_c^2\delta_A\right)}{\delta_c^2 + \delta_\Sigma^2 + 2\delta_\Sigma\,\delta_c\delta_A + \delta_c^2\delta_A^2}, \qquad (2.41)$$

где $\delta_\Sigma = \delta_b + \delta_A$, а δ_A, δ_b - указанные в (2.39) относительные погрешности для матрицы **A** и вектора **b**. Напомним, что погрешности δ_A возникают лишь в ходе вычислений, а δ_b включает также и погрешности исходных данных.

Обычно погрешности, связанные с округлениями в ходе вычислений, существенно меньше погрешностей задания исходных данных. Если, вдобавок, задача достаточно хорошо обусловлена, то можно считать $\delta_A = 0$, при этом $\delta_\Sigma = \delta_b$. Тогда из (2.41), как частный случай, вытекает следующая оценка

$$\Phi \geq \frac{2\delta_b\delta_c}{\delta_b^2 + \delta_c^2}. \qquad (2.42)$$

Необходимо иметь в виду, что в данном случае должно выполняться вытекающее из (1.45) неравенство $\delta_c \geq \delta_b$. С учетом этого из (2.42) следует, что уменьшение величины приведенного показателя диагонального преобладания до значений близких к нулю ($\Phi \to 0$) свидетельствует о значительном увеличении погрешности δ_c, т.к. при этом должно выполняться $\delta_c >> \delta_b$. Интересно, что равенство $\delta_c = \delta_b$ возможно, только при $\Phi = 1$.

Наглядное объяснение существа приведенных выше неравенств и ограничений дает геометрическая интерпретация для матрицы Грама размерности 3×3. Предположим, что для такой матрицы с собственными значениями λ_1, λ_2, λ_3 строятся оценки по показателю диагонального преобладания (ϕ или Φ), а множеству $\Omega_2 = \left\{\mathbf{A}: \mathrm{tr}\mathbf{A} = t_1, \left(\mathrm{tr}\mathbf{A}^2\right)^{1/2} = t_2\right\}$ матриц соответствует множество собственных значений одновременно удовлетворяющих уравнениям

$$\lambda_1 + \lambda_2 + \lambda_3 = t_1,$$
$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = t_2^2.$$

В трехмерном пространстве, образованном собственными значениями λ_1, λ_2, λ_3, первое уравнение описывает плоскость, отсекающую на осях величину t_1. Второе уравнение - сфера с центром в начале координат и радиусом t_2. Подмножество Ω_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 $\Phi=\phi-M+1$).

На рис. 2.3 показаны множества Ω_2, соответствующие различным значениям ϕ ($\phi_1>\phi_2$). Для множества Ω_2, соответствующего величине показателя ϕ_2 цифрами 1 и 2 указаны точки этого множества, в которых достигаются оценки (2.14), (2.15) для собственных значений, а в непосредственной близости от точки 1 показана также точка K_1^B, в которой имеет место максимально возможное для соответствующего множества матриц число обусловленности.

Для матриц более высокого порядка такое наглядное представление невозможно. Но даже на приведенном простом примере ясно, что показатель диагонального преобладания ϕ (или Φ) является важной характеристикой матрицы Грама и наряду с минимальным собственным значением, числом обусловленности и определителем может рассматриваться как самостоятельная мера обусловленности. Эта точка зрения подтверждается также результатами экспериментальных исследований на тестовых изображениях, полученных путем моделирования фрагментов с заведомо различной степенью обусловленности. Эти результаты будут приведены в разделе 2.6.

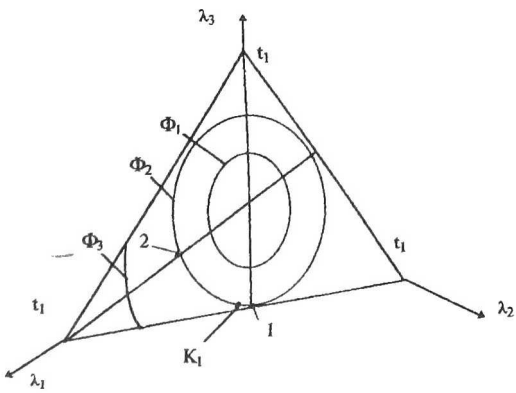


Рис 2.3. Геометрическая интерпретация множества Ω_2

Важным достоинством показателя ϕ (или Φ), по сравнению с указанными известными мерами (минимальное собственное значение, число обусловленности и определитель) является вычислительная простота. Это имеет большое значение в системах обработки изображений с точки зрения практической реализации.

Недостатком этого показателя является то, что гарантированные оценки обусловленности с его использованием могут быть построены лишь для значений, удовлетворяющих неравенствам $M-1<\phi\le M$ (или $0<\Phi\le 1$).

Из рис. 2.4 видно, что построение гарантированных оценок обусловленности по показателю диагонального преобладания при значениях ϕ в диапазоне $1<\phi\le M-1$ возможно только в случае, когда априори 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. Рассмотрим возможность повышения эффективности оценок обусловленности по показателю диагонального преобладания путем преобразования исходной матрицы Грама.

Рассмотрим преобразование матрицы $\mathbf{A}=\mathbf{X}^{\mathrm{T}}\mathbf{X}$, заключающееся в ее масштабировании по правилу

$$\tilde{\mathbf{A}}=\mathbf{DAD}, \qquad (2.43)$$

где $\mathbf{D}=\mathrm{diag}(d_1,d_2,\ldots,d_M)$ - диагональная M×M- матрица масштабных коэффициентов. Известно, что путем уменьшения различий в масштабах независимых переменных можно улучшить обусловленность задачи. Следовательно, изменяя матрицу **D**, можно "управлять" мерами обусловленности. Учитывая выявленную выше связь показателя диагонального преобладания с известными мерами обусловленности и мультиколлинеарности, мы вправе предположить, что путем такого масштабирования можно изменять также (в частности, увеличивать) и диагональное преобладание матрицы Грама.

Будем искать такую $\mathbf{D}_{opt}$, которая доставляет максимум показателю диагонального преобладания матрицы $\mathbf{A}(\mathbf{D})=\mathbf{DAD}$:

$$\mathbf{D}_{opt}=\max_{\mathbf{D}}\left(\frac{\mathrm{tr}^2\mathbf{DAD}}{\mathrm{tr}(\mathbf{DAD})^2}\right)=\max_{\mathbf{D}}\left(\left(\sum_{i=1}^{M}d_i^2a_{i,i}\right)^2\Big/\sum_{i,j=1}^{M}d_i^2d_j^2a_{i,j}^2\right). \qquad (2.44)$$

Необходимые условия экстремума задаются системой M уравнений:

$$\sum_{i,j=1}^{M}d_i^2d_j^2a_{i,j}^2-\left(\sum_{i=1}^{M}d_i^2a_{i,i}\right)\left\{d_k^2a_{k,k}^2+\frac{1}{2}\sum_{\substack{j=1,\\ j\ne k}}^{M}d_j^2a_{k,j}^2+\frac{1}{2}\sum_{\substack{i=1,\\ i\ne k}}^{M}d_i^2a_{i,k}^2\right\}=0,\quad k=\overline{1,M}. \qquad (2.45)$$

Решение достаточно сложной нелинейной системы (2.45) в каждом эпизоде построения оценок обусловленности для фиксированной матрицы **A** вряд ли це-

лесообразно. Если малое диагональное преобладание исходной матрицы A связано лишь с различиями в масштабах независимых переменных, то после масштабирования относительный вес внедиагональных элементов будет значительно меньше диагональных. В указанной ситуации можно пренебречь слагаемыми, содержащими внедиагональные элементы. В результате приходим к более простой системе уравнений

$$\sum_{i=1}^{M} d_i^4 a_{i,i} - d_k^2 a_{k,k} \sum_{i=1}^{M} d_i^2 a_{i,i} \approx 0, \quad k = \overline{1,M}, \tag{2.46}$$

из которой следует, что масштабные коэффициенты можно вычислять по правилу

$$d_k \approx \eta \cdot \left(a_{k,k}\right)^{-\frac{1}{2}}, \quad k = \overline{1,M}, \tag{2.47}$$

где

$$\eta = \left(\sum_{i=1}^{M} d_i^4 a_{i,i} \Big/ \sum_{i=1}^{M} d_i^2 a_{i,i} \right)^{1/2}.$$

Умножение матрицы на скаляр не изменяет ее диагонального преобладания, поэтому величина η может быть произвольной. В частности, можно принять $\eta=1$, а масштабные коэффициенты определять в соответствии с равенством $d_k = 1/\sqrt{a_{k,k}}$, $k = \overline{1,M}$. Полученная в результате такого масштабирования матрица, является матрицей сопряженности: $\mathbf{A(D) = DAD = R}$.

В разделе 2.6 будут приведены результаты экспериментов показывающие, что показатель диагонального преобладания матрицы сопряженности $\Phi(R)$ сформированной по данным, обладающим слабой мультиколлинеарностью, действительно оказывается выше показателя диагонального преобладания $\Phi(A)$ исходной матрицы. Это совершенно ясно, поскольку в результате устранения различий в параметрах масштаба он характеризует лишь сопряженность векторов независимых переменных. Поэтому показатель $\Phi(R)$, в отличие от показателя $\Phi(A)$, характеризующего обусловленность исходной матрицы, является мерой мультиколлинеарности.

При наличии сильной мультиколлинеарности описанная мера повышения диагонального преобладания не приведет к желаемой цели. И хотя нам известно, что при значениях $\phi(A)$ и $\phi(R)$ близких к M-1 задача (как правило, но не обязательно) невырождена, объективно при $1<\phi<M-1$ имеет место неуверенность в оценке ситуации. Если цена ошибки слишком велика, чтобы можно было полагаться на ее малую вероятность, необходимо строить гарантированную оценку

степени обусловленности, возможно не считаясь с вычислительными затратами. В следующем разделе предлагается и исследуется такая мера мультиколлинеарности.

2.5. Оценка мультиколлинеарности по показателям сопряженности с нуль - пространством матрицы измерений

Важнейшее достоинство методов оценки обусловленности по показателю диагонального преобладания матрицы Грама - их вычислительная простота. При сильной мультиколлинеарности векторов независимых переменных ситуация может оказаться такой, что гарантированная оценка по показателю $\Phi(\mathbf{R})$ *уже* не существует, хотя задача *еще* может быть решена с приемлемой точностью. Связано это с ограничениями-неравенствами (2.13). В настоящем разделе мы введем в рассмотрение свободные от указанного недостатка меры, основанные на анализе сопряженности векторов независимых переменных с нуль-пространством.

В качестве меры мультиколлинеарности можно использовать величину

$$S_{min} = \min_i |S_i|, \qquad (2.48)$$

где $S_i = \left(\sum \beta_j^2\right)^{1/2}$, $j = \overline{M,N}$ $\beta_j = \langle x_i, t_j \rangle$ проекция вектора-столбца $\mathbf{x}_i$ матрицы $\mathbf{X}$ на j-й вектор базиса, образованного собственными векторами, соответствующими нулевым собственным значениям матрицы $\mathbf{X}_{M-1}\mathbf{X}_{M-1}^T$. Здесь $\mathbf{X}_{M-1}$ N×(M-1) - матрица, составленная из (M-1) нормированных ($\|\mathbf{x}_i\| = 1$, $i = \overline{1,M}$) векторов-столбцов матрицы $\mathbf{X}$ после исключения вектора $\mathbf{x}_i$. Геометрически S_i косинус угла между вектором $\mathbf{x}_i$ и нуль-пространством подпространства, натянутого на остальные M-1 векторов. Меру S_{min} далее будем назвать *показателем минимальной сопряженности с нуль-пространством.*

Для вычисления меры (2.48) необходимо вначале вычислить все S_i, $i = \overline{M,N}$. Построим для этого явные соотношения, соответствующие введенным выше понятиям. В соответствии с геометрической трактовкой осуществим разложение вектора $\mathbf{x}_i$ по базису, образованному пространством матрицы $\mathbf{X}_{M-1}$:

$$\mathbf{z}_i = \left[\mathbf{X}_{M-1}^T\mathbf{X}_{M-1}\right]^{-1}\mathbf{X}_{M-1}^T\mathbf{x}_i. \qquad (2.49)$$

Проекция этого вектора на подпространство, натянутое на M-1 векторов, из которых составлена матрица $\mathbf{X}_{M-1}$, определяется как

$$\hat{\mathbf{x}}_i = \mathbf{X}_{M-1}\mathbf{z}_i. \qquad (2.50)$$

Вектор "невязок" между вектором $\mathbf{x}_i$ и проекцией $\hat{\mathbf{x}}_i$ на указанное подпространство обозначим

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i.$$

Ясно, что этот вектор ортогонален подпространству, образованному матрицей $\mathbf{X}_{M-1}$ и принадлежит ее нуль-пространству. Следовательно, косинус угла между вектором $\mathbf{x}_i$ и нуль-пространством матрицы $\mathbf{X}_{M-1}$ определяется как

$$\beta_i = \langle \mathbf{x}_i, \tilde{\mathbf{x}}_i \rangle / \left(\|\mathbf{x}_i\| \|\tilde{\mathbf{x}}_i\| \right), \qquad (2.51)$$

где

$$\|\mathbf{x}_i\| = \langle \mathbf{x}_i, \mathbf{x}_i \rangle^{1/2}.$$

Здесь $\langle \cdot,\cdot \rangle$ означает скалярное произведение векторов, а $\|\tilde{\mathbf{x}}_i\|$, с учетом свойств спектрального разложения симметричной вырожденной матрицы и соотношений для векторов $\hat{\mathbf{x}}_i$ и $\mathbf{z}$, определяется как

$$\|\tilde{\mathbf{x}}_i\|_2 = \left\langle (\mathbf{x}_i - \hat{\mathbf{x}}_i), (\mathbf{x}_i - \hat{\mathbf{x}}_i) \right\rangle^{1/2} = \left(\mathbf{x}_i^{\mathrm{T}} \left[\mathbf{E}_{M-1} - \mathbf{T}_{\lambda,M-1}\mathbf{T}_{\lambda,M-1}^{\mathrm{T}} \right] \mathbf{x}_i \right)^{1/2} = \left(\mathbf{x}_i^{\mathrm{T}} \mathbf{T}_{0,M-1}\mathbf{T}_{0,M-1}^{\mathrm{T}} \mathbf{x}_i \right)^{1/2}.$$

Подставляя в равенство (2.51) выписанные выше выражения для векторов $\tilde{\mathbf{x}}_i$ и $\mathbf{x}_i$, после соответствующих преобразований, учитывающих определение величин S_i в (2.48), получаем следующее выражение для их вычисления:

$$S_i = \left(\mathbf{x}_i^{\mathrm{T}} \mathbf{T}_{0,M-1}\mathbf{T}_{0,M-1}^{\mathrm{T}} \mathbf{x}_i \right)^{1/2} \Big/ \left(\mathbf{x}_i^{\mathrm{T}} \mathbf{x}_i \right)^{1/2}, \quad i = \overline{1,M}. \qquad (2.52)$$

Здесь $\mathbf{T}_{0,M-1}$ - матрица, размерности N×(N-M) составленная из N-M+1 собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}_{M-1}\mathbf{X}_{M-1}^{\mathrm{T}}$.

Можно показать тесную связь меры (2.48) с максимальной сопряженностью (2.5) [15]. Для удобства их сопоставления показатель максимальной сопряженности (2.5) представим в следующем виде

$$R = \max_i |R_i|, \quad R_i = \langle \mathbf{X}_{M-1}\mathbf{z}_i, \mathbf{x}_i \rangle, \qquad (2.53)$$

где $\mathbf{z}_i$ определяется в соответствии с (2.49):

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

С использованием обозначений, принятых в (2.48), показатель максимальной сопряженности R, фигурирующий в (2.53), можно представить в виде

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

где $\alpha_j = \langle x_i, t_j \rangle$ проекция вектора x_i на j-й вектор базиса, образованного собственными векторами, соответствующими ненулевым собственным значениям. Ясно, что если вектор x_i принадлежит пространству матрицы 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.

Эти различия, как раз, и являются причиной вычислительных преимуществ меры (2.48) при обработке малых наборов данных. Ясно, что 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 возникает за счет того, что для определения матрицы $T_{0,M-1}$ не требуется решение полной проблемы собственных значений. Кроме того, при вычислении S_i (в отличие от R_i) не нужно вычислять обратную матрицу.

Необходимо подчеркнуть, что в случае, когда показатель S_{min} отличен от нуля, гарантируется невырожденность задачи, а с его возрастанием обусловленность улучшается. В главе 6 будут приведены результаты экспериментов, показывающие, что хотя на отдельных реализациях, характеризующихся одинаковыми значениями показателя S_{min}, точности различаются, имеет место устойчивая тенденция к повышению точности идентификации при $S_{min} \to 1$. Случайный характер погрешности идентификации связан со случайной ориентацией вектора ошибок исходных данных в конкретных экспериментах. Наиболее неблагоприятным является направление, совпадающее с направлением собственного вектора, соответствующего минимальному собственному значению. При этом возможна большая ошибка идентификации, даже если показатель S_{min} принимает не слишком малые значения.

Еще один источник случайности ошибок идентификации связан с определением самой меры мультиколлинеарности. Дело в том, что для оценки степени обусловленности в данном случае используется лишь одно: минимальное среди чисел S_i, $i = \overline{M,N}$, а характер ориентации остальных векторов по отношению

друг к другу не учитывается. Интуитивно ясно, что при одинаковых числовых значениях S_{min} для двух разных протоколов данных, более "плохим" следует считать набор данных, для которого все значения S_i почти равны между собой и близки к нулю.

Для преодоления указанного недостатка необходимо использовать некоторые обобщенные характеристики взаимной ориентации векторов матрицы **X**. В качестве такой меры, в значительной степени свободной от указанного недостатка, можно использовать величину

$$S_{\Sigma} = \left(\sum S_i^2\right)^{1/2}, \tag{2.54}$$

$$S_i^2 = \sum \beta_j^2, \quad j = \overline{M,N}. \tag{2.55}$$

Здесь все обозначения те же, что и в (2.48). Фигурирующая в формуле (2.55) величина S_i^2 в соответствии с (2.52), определяется равенством

$$S_i^2 = \left(\mathbf{x}_i^T \mathbf{T}_0 \mathbf{T}_0^T \mathbf{x}_i\right) / \left(\mathbf{x}_i^T \mathbf{x}_i\right). \tag{2.56}$$

Меру S_{Σ}, определяемую соотношением (2.54), по аналогии с мерой (2.48), назовем - *показателем суммарной сопряженности с нуль-пространством*.

Предложенная мера (2.54) соотносится с мерой (2.48) примерно также как след матрицы (сумма собственных значений) с минимальным собственным значением. Известно, что след матрицы, характеризующий размеры доверительного эллипсоида, не может служить надежной мерой обусловленности. Он может принимать большие значения и в случае, когда задача вырождена.

Точно также и мера (2.54) является менее надежной, чем мера (2.48). Мера S_{Σ} может принимать достаточно большие значения даже в том случае, когда $S_{min}=0$, т. е. задача вырождена. Тем не менее, вычисление показателя S_{Σ} наряду с мерой (2.48) целесообразно, т.к. он несет дополнительную информацию о структуре пространства матрицы **X**. Это тем более оправдано, поскольку не требует, сколько-нибудь значительных, дополнительных вычислительных затрат. Таким образом, практическое использование характеристики S_{Σ} целесообразно лишь в сочетании с показателем S_{min}, либо в ситуациях, когда ранг матрицы не вызывает сомнений (Rank**X**=M) и необходимо дать лишь обобщенную оценку степени ее обусловленности.

В следующем разделе будут приведены результаты экспериментов, показывающие связь предложенных мер мультиколлинеарности с погрешностью

идентификации и дана сравнительная оценка их эффективности. В частности, будет показано, что меры (2.48) и (2.54) тесно связаны с известными мерами мультиколлинеарности (число обусловленности, минимальное собственное значение и определитель).

Необходимо признать, что объем вычислений при определении показателей сопряженности с нуль-пространством, по сравнению с вычислением показателя диагонального преобладания, намного больше. Это, неизбежная, плата за эффективность оценок. Утешением в данном случае является неоспоримый факт меньшей, по сравнению с показателем максимальной сопряженности, вычислительной сложности метода именно на малых наборах данных. Тем не менее, использование обсуждавшихся в настоящем разделе мер может быть оправдано лишь высокой стоимостью риска, связанного с возможными ошибками идентификации.

2.6. Связь мер обусловленности и мультиколлинеарности

В настоящем разделе мы намерены показать связь предложенных мер мультиколлинеарности и обусловленности: показателя диагонального преобладания и показателя минимальной сопряженности с нуль-пространством с известными мерами, обсуждавшимися в разделе 2.1. В частности, там мы отмечали, что наиболее сильными известными мерами являются минимальное собственное значение, число обусловленности и показатель максимальной сопряженности. Все указанные меры, а также определитель нормированной матрицы Грама, могут служить индикаторами обусловленности задачи. Однако некоторые из них учитывают лишь сопряженность (мультиколлинеарность), а другие "чувствуют" также и различия в масштабах векторов независимых переменных.

На рис. 2.4, а и б показана связь предложенных показателей $\phi(\mathbf{A})$ и S_{min} с пречисленными мерами в двух крайних ситуациях, характеризующихся как хорошо и плохо обусловленная задача. Схема напоминает модель направленного графа. Идея этого представления заимствована из [15]. Так же как и там, сплошная стрелка означает "из . обязательно следует . .", а пунктирная - "из . . *как правило*, но не обязательно, следует . . .". Показанные на схемах связи, установлены на основе анализа аналитических соотношений для мер мультиколлинеарности и обусловленности. Поскольку связи, показанные пунктиром, вызывают сомнение, представляет интерес убедиться в их существовании на экспериментальных данных.

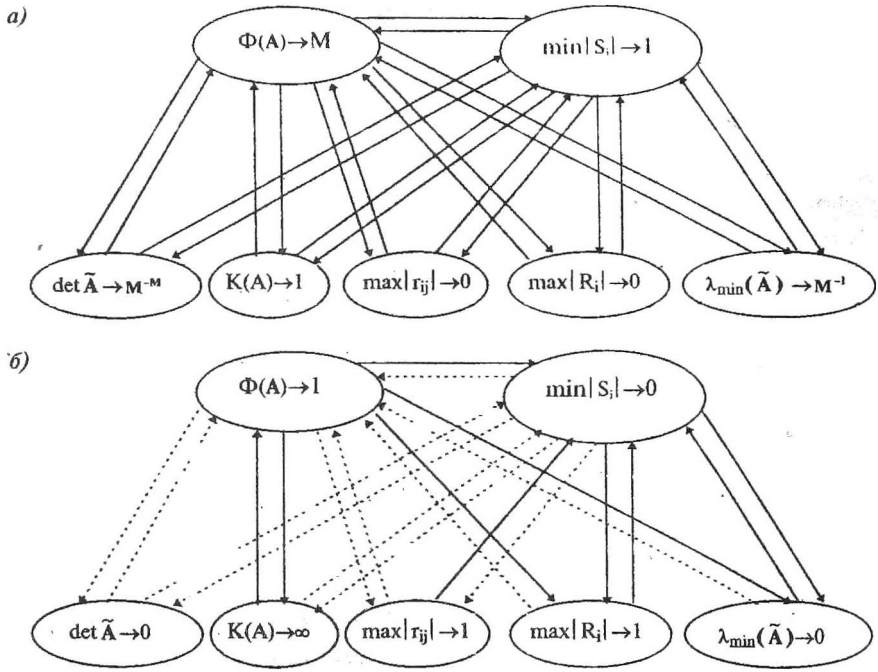


Рис.2.4. Схема связей показателей мультиколлинеарности и обусловленности в случае: а - хорошо обусловленной задачи; б - плохо обусловленной задачи

Для их исследования решалась задача идентификации линейной модели системы формирования изображений с использованием тестовых изображений, обладающих заведомо различной степенью обусловленности в достаточно широком диапазоне. Наряду с проверкой устойчивости связей, показанных пунктирными стрелками, ставилась также задача дать сравнительную оценку эффективности исследуемых мер.

Эксперимент проводился по следующей схеме. Осуществлялась оценка параметров импульсной характеристики искажающей системы по изображениям № 1-5, которые показаны на рисунке 2.5. Интуитивно ясно, что вызываемая мультиколлинеарностью обусловленность задачи идентификации, решаемой по этим изображениям, должна ухудшаться по мере увеличения номера изображения. Наиболее информативным, т.е. приводящим к хорошо обусловленной задаче, является тестовое изображение №1. Его генерация проводилась при помощи датчика псевдослучайных чисел, а каждый отсчет этого изображения является

реализацией случайной величины, равномерно распределенной на интервале от 0 до 255. Тесты №2-5 формировались как фрагменты изображения, использовавшегося в работе [47].

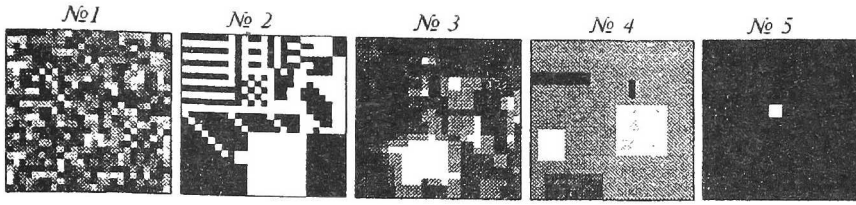


Рис 2.5. Тестовые изображения

Выходные изображения формировались путем моделирования расфокусировки, обладающей круговой симметрией, с использованием разностного уравнения вида [47,62]

$$y(n_1, n_2) = \sum_{k=0}^{4} h(k)x(k), \qquad (2.57)$$

где $x(0)=x(n_1,n_2)$,

$x(1)=[x(n_1+1,0)+x(n_1-1,0)+x(0,n_2+1)+x(0,n_2-1)]/4$,

$x(2)=[x(n_1+1,n_2+1)+ x(n_1+1,n_2-1)+ x(n_1-1, n_2+1)+ x(n_1-1, n_2-1)]/4$,

$x(3)=[x(n_1+2,0)+x(n_1-2,0)+x(0,n_2+2)+x(0,n_2-2)]/4$,

$x(4)=[x(n_1+2,n_2+1)+ x(n_1+2,n_2-1)+ x(n_1-2, n_2+1)+ x(n_1-2, n_2-1)+$

$+x(n_1+1,n_2+2)+ x(n_1+1,n_2-2)+ x(n_1-1, n_2+2)+ x(n_1-1, n_2-2)]/8$.

Уравнение (2.57) соответствует опорной области 5×5 (без угловых отсчетов), показанной на рис. 2.6. Значения импульсной характеристики искажающего фильтра задавались равными: $h(0)=0.102$, $h(1)=0.327$, $h(2)=0.245$, $h(3)=h(4)=0.163$. При указанных значениях параметров импульсная характеристика имеет центральную симметрию. Соответствующая этой опорной области матрица **X** для каждого тестового изображения имела размерность 25×5, а матрица $\mathbf{A}=\mathbf{X}^T\mathbf{X}$ 5×5. Вектор оценок параметров искажающей системы $\hat{\mathbf{h}} = \left[\hat{h}(0), \ldots, \hat{h}(4)\right]$ определялся с помощью метода наименьших квадратов. Было проведено два вида экспериментов: исследование связи показателей $\phi(\mathbf{A})$, $\phi(\mathbf{R})$ и S_{min} с известными мерами обусловленности и/или мультиколлине-

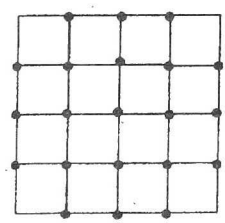


Рис. 2.6. Опорная область

арности и исследование зависимости погрешности идентификации от величины этих показателей.

При исследования связи мер обусловленности для показанных выше пяти тестовых изображений №1-5 (рис. 2.6) размером 25×25 рассчитывались следующие характеристики обусловленности: определитель нормированной матрицы $\det(\tilde{\mathbf{A}})$, спектральное число обусловленности - $K(A)=\lambda_{max}(A)/\lambda_{min}(A)$, минимальное собственное значение $\lambda_{min}(A)$, показатели диагонального преобладания - $\phi(A)$ и $\phi(R)$ (R - матрица сопряженности), а также показатели максимальной сопряженности R и минимальной сопряженности с нуль-пространством S_{min}. Результаты сравнительного исследования мер приведены в таблице 2.3.

Из таблицы видно, что числовые значения всех мер обусловленности указывают на ее ухудшение с ростом номера изображения. Известные меры K(A) и λ_{min} уменьшаются. Показатель S_{min}, как мы и предполагали, принимает значения в диапазоне от 0 до 1 и для изображения № 5 его значение близко к нулю. Порождаемая этим тестовым изображением задача идентификации характеризуется как плохо обусловленная и по другим показателям. Это объясняется тем, что на нем присутствует в основном фон. Для этого изображения

$K(A)\to\infty$, $\lambda_{min}\to 0$, $S_{min}\to 0$.

Таблица 2.3. Сравнение мер мультиколлинеарности и обусловленности

№	det(A/trA)	K(A)	λ_{min}	$\phi(A)$	$\phi(R)$	R	S_{min}
1	3.67×10^{-5}	16.72	12165.9	2.53024	4.26353	0.58013	0.814522
2	6.42×10^{-6}	37.85	2182.3	1.91591	1.88864	0.87135	0.490654
3	7.96×10^{-9}	342.02	672.6	1.16107	1.15292	0.98731	0.158775
4	5.07×10^{-10}	3011.98	35.2	1.21001	1.21259	0.99883	0.057468
5	0.0	∞	0.0	1.00000	1.00000	1.00000	0.0

При хорошей обусловленности $S_{min}\to 1$, а показатели диагонального преобладания $\phi(A)$ и $\phi(R)$ принимают значения существенно более высокие, чем для изображения № 5. Этот случай хорошо иллюстрируется в эксперименте с тестовым изображением №1. Для него $K(A)=16.72$, $S_{min}=0.815452$, $\phi(A)=2.53024$, а $\Phi(R)=4.26353$. Для тестовых изображений 2,3,4, по мере ухудшения обусловленности, все показатели изменяются в соответствии с выводами теории. В

частности, для матрицы, сформированной по тестовому изображению №1, показатель $\phi(\mathbf{R})$ значительно выше показателя $\phi(\mathbf{A})$, что подтверждает возможность повышения диагонального преобладания хорошо обусловленных информационных матриц (Грама) путем масштабирования.

Более наглядное представление о характере связи известных и предложенных в работе мер мультиколлинеарности и обусловленности дают приведенные на рисунках 2.7, а-е графики. Для их построения использовались результаты, приведенные в таблице 2.3, а также дополнительные экспериментальные данные, полученные на том же множестве тестовых изображений (№1-5). На этих графиках цифры по оси абсцисс обозначают номера тестовых изображений. Приведенные кривые показывают изменение различных показателей обусловленности в функции номера изображения. Они убедительно доказывают тесную связь предложенных мер: показателя диагонального преобладания информационной матрицы и показателя минимальной сопряженности с нуль пространством исходной матрицы измерений с известными мерами обусловленности и подтверждают возможность использования указанных мер для анализа информативности исходных данных.

Выбор одной из них, конечно, должен осуществляться в зависимости от требуемой надежности вырабатываемых решений и вычислительных возможностей. Показатель диагонального преобладания является наиболее простым в вычислительном отношении и может использоваться в тех случаях, когда допустим риск принятия неверного решения. Если потери от неверного решения слишком велики, целесообразно использовать показатель сопряженности с нуль-пространством, который можно вычислять на фрагментах изображения, обладающих крайне низкой информативностью (небольшая деталь объекта на сплошном фоне). Это важно с практической точки зрения, т.к. вычисление на таких фрагментах известных мер (λ_{min}, K(A) и R) при плохой обусловленности становится трудноразрешимой проблемой.

В заключительной, шестой, главе будут приведены результаты экспериментальных исследований предложенных мер мультиколлинеарности и обусловленности $\phi(\mathbf{A})$, $\phi(\mathbf{R})$ и S_{min} при решении конкретных задач оценивания параметров моделей систем формирования изображений. В настоящем разделе мы показали их тесную связь между собой и с известными мерами, а также их преимущества. Мы установили также связь этих мер с погрешностями идентификации.

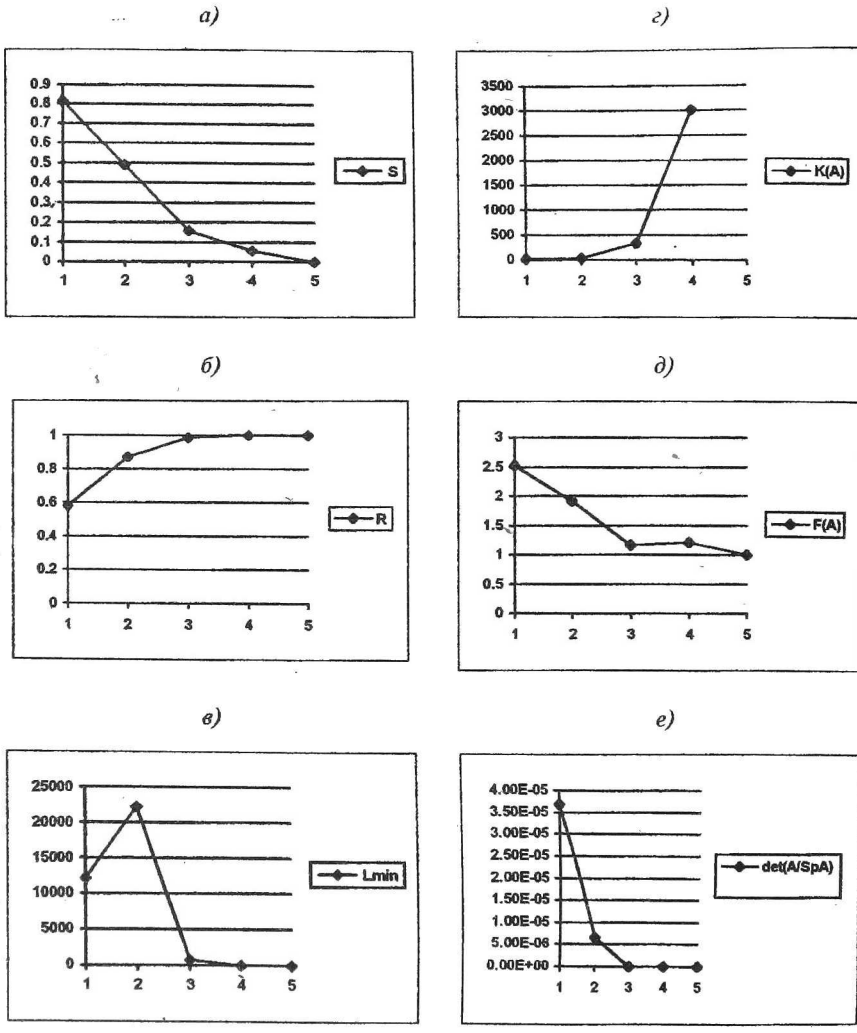


Рис. 2.7. Графики изменения показателей обусловленности:
а - S_{min}; *б* - R; *в* - λ_{min}; *г* - K(A); *д* - ϕ(A); *е* - det $\tilde{A}$

Применив одну из этих мер мультиколлинеарности и обусловленности к конкретному набору данных, мы можем всегда получить, по крайней мере, один из трех вариантов ответа: 1) задача плохо обусловлена и решение с требуемой точностью невозможно; 2) значение меры обусловленности на заданном фиксированном наборе данных не позволяет сделать уверенное заключение о

достижимой точности оценивания; 3) задача хорошо обусловлена так, что существуют и могут быть выработаны (по $\phi(\mathbf{A})$, $\phi(\mathbf{R})$ и/или S_{min}) гарантированные оценки разрешимости задачи с требуемой точностью при заданной точности исходных данных.

В первом и третьем случаях решения, которые следует принять, очевидны. Во втором случае есть надежда на получение удовлетворительных оценок путем преобразований измерительной информации типа взвешивания и непосредственной корректиовки вектора выхода модели. Для того чтобы ответить на вопрос: какие типы и значения параметров преобразований приведут к повышению точности оценок, нам понадобятся некоторые результаты качественного характера. В следующей главе, оставаясь в рамках алгебраической теории, мы проведем качественный анализ ошибок идентификации, установим общие закономерности их формирования при указанных преобразованиях и выявим условия достижения предельной точности.

3. ДОСТИЖИМАЯ ТОЧНОСТЬ ПРИ ПРЕОБРАЗОВАНИЯХ ДАННЫХ

Если в результате анализа информативности данных получен положительный ответ на вопрос о принципиальной возможности построения оценок параметров с требуемой точностью, возникают новые вопросы: какова предельная точность идентификации, которая может достигаться в схеме построения МНК-оценок (1.11) с промежуточными преобразованиями взвешивания и корректировки данных, и как должны выбираться параметры этих преобразований, чтобы достичь наибольшей точности? Настоящая глава посвящена исследованию связанного с этими вопросами круга проблем.

3.1. Качественный анализ ошибок идентификации при преобразованиях взвешивания

В разделе 1.2 из соображений общего характера мы обосновали схему получения модифицированной оценки (1.11) в виде последовательности МНК оценок с перестраиваемой весовой матрицей и корректировкой вектора независимых переменных. В соответствии с этой схемой выбор весовой матрицы G и/или вектора поправок Δy должен осуществляться после каждого эпизода вычисления оценок с учетом информации трех уровней: априорной, текущей и дополнительной (получаемой из невязок).

Для того чтобы научиться эффективно использовать эту информацию, необходимо установить основные закономерности формирования ошибок идентификации. Для этого мы будем использовать методический прием, заключающийся в предположении, что вектор ошибок исходных данных ξ задан. В действительности он никогда не известен. Используя этот прием мы хотим выяснить как различные векторы ошибок "трансформируются" в векторы ошибок идентификации. Цель исследования заключается, во-первых, в выявлении условий, при которых может достигаться предельная точность, во-вторых, в установлении множеств векторов ошибок, для которых возможно улучшение точности оценивания путем преобразований.

В настоящем разделе мы подвергнем исследованию лишь линейное преобразование типа взвешивания (1.12). Оно заключается в том, что исходная матрица **X** и вектор **y** умножаются на диагональную N×N матрицу **G**. При рассмотрении вопросов построения этой матрицы мы будем предполагать, что одним из описанных в предыдущей главе способов установлено, что Rank**X**=M. Важнейший

вопрос, на который нам предстоит ответить, заключается в следующем. Всегда ли можно уменьшить заданную меру вектора ошибок идентификации изменяя матрицу G?

Исследования, в рамках которых получены описываемые ниже результаты, опираются на идеи качественной теории идентификации [48,53]. Следуя традиции, сложившейся в указанных работах, зададим в качестве меры близости оценки к истинному значению скалярную характеристику: квадрат евклидовой нормы вектора ошибок идентификации. Тогда поставленный выше вопрос можно переформулировать следующим образом.

Предположим, что на некотором этапе идентификации получена МНК-оценка (1.13):

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}. \tag{3.1}$$

Соответствующая ей ошибка идентификации $\Delta\mathbf{c} = \hat{\mathbf{c}} - \mathbf{c} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\xi$. Для квадрата евклидовой нормы этого вектора можно записать равенство

$$\|\Delta\mathbf{c}\|_2^2 = \Delta\mathbf{c}^{\mathrm{T}}\Delta\mathbf{c} = \xi^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]\mathbf{X}^{\mathrm{T}}\xi. \tag{3.2}$$

Пусть теперь каким-то образом задана диагональная матрица $\mathbf{G}^2 = \mathrm{diag}\left(g_1^2, g_2^2, \ldots, g_N^2\right)$, $\mathbf{G}^2 = \mathbf{G}^{\mathrm{T}}\mathbf{G} > 0$, $\det\mathbf{G} \neq 0$ и построена новая оценка

$$\tilde{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{G}^2\mathbf{y}. \tag{3.3}$$

Это равносильно повторному оцениванию по соотношению (3.1) с использованием данных, подвергшихся линейным преобразованиям вида (1.12):

$$\tilde{\mathbf{y}} = \mathbf{G}\mathbf{y}, \quad \tilde{\mathbf{X}} = \mathbf{G}\mathbf{X}. \tag{3.4}$$

При этом неизвестный вектор ошибок ξ также преобразуется в $\tilde{\xi} = \mathbf{G}\xi$, а

$$\Delta\tilde{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}\tilde{\xi}. \tag{3.5}$$

Соответствующая преобразованному набору данных (3.4) и ошибке (3.5) мера близости имеет вид

$$\|\Delta\tilde{\mathbf{c}}\|_2^2 = \Delta\tilde{\mathbf{c}}^{\mathrm{T}}\Delta\tilde{\mathbf{c}} = \tilde{\xi}^{\mathrm{T}}\tilde{\mathbf{X}}\left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-2}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\xi}. \tag{3.6}$$

Теперь поставленный выше вопрос формулируется более "узко": всегда ли можно при произвольном, но фиксированном векторе ошибок ξ подобрать диагональную матрицу G так, что при этом возможно достижение меньшего (или

хотя бы сохранение прежнего) значения меры близости (3.2) по сравнению с (3.6):

$$\|\Delta \mathbf{c}\|_2^2 \geq \|\Delta \tilde{\mathbf{c}}\|_2^2 ? \qquad (3.7)$$

Приведенная постановка, являющаяся предметом качественной теории идентификации, по-видимому, впервые была сформулирована в работе [53]. В указанной работе поставленный выше вопрос относился к параметрам (замкнутой и разомкнутой) управляемых динамических систем. Здесь мы хотим решить аналогичную задачу для элементов весовой матрицы **G**.

Подставив в (3.7) меры (3.2), (3.6), нетрудно заметить, что это неравенство будет выполняться, если имеет место неотрицательная определенность квадратичной формы:

$$\xi^{\mathrm{T}}\left[\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-2}\mathbf{X}^{\mathrm{T}} - \mathbf{G}^{\mathrm{T}}\tilde{\mathbf{X}}\left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-2}\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{G}\right]\xi \geq 0 . \qquad (3.8)$$

Проанализируем, когда это возможно.

Введем в рассмотрение уже использовавшуюся в разделе 1.4 ортогональную N×N матрицу $\mathbf{T} = [\mathbf{T}_\lambda \ \vdots \ \mathbf{T}_0]$, где в соответствии с (1.33) N×M - матрица $\mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\Lambda^{-\frac{1}{2}}$, **F** M×M - матрица: $\mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{F}\mathbf{F}^{\mathrm{T}} = \mathbf{E}_{\mathrm{M}}$, $\mathbf{F}^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]\mathbf{F} = \Lambda$, Λ - диагональная матрица собственных значений, а $\mathbf{T}_0$ - N×(N-M) матрица, составленная из собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}\mathbf{X}^{\mathrm{T}}$.

Учитывая блочную структуру ортогональной матрицы **T**, а также свойства

$$\mathbf{T}_\lambda^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{T}_\lambda = \Lambda^{-1}, \qquad \mathbf{X}^{\mathrm{T}}\mathbf{T}_0 = 0 \ (\mathbf{T}_0^{\mathrm{T}}\mathbf{X} = 0),$$

неравенство (3.8) можно переписать в виде

$$\xi^{\mathrm{T}}\mathbf{T}\left\{\begin{bmatrix} \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda & \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0 \\ \mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda & \mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0 \end{bmatrix} - \begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{bmatrix}\right\}\mathbf{T}^{\mathrm{T}}\xi \leq 0 , \qquad (3.9)$$

где $\tilde{\mathbf{V}} = \tilde{\mathbf{X}}\left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-2}\tilde{\mathbf{X}}^{\mathrm{T}}$.

Можно показать, что

$$\mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda = \Lambda^{-1}. \qquad (3.10)$$

Действительно, используя свойство $\mathbf{F} = \left(\mathbf{F}^{\mathrm{T}}\right)^{-1}$ ортогональной матрицы **F** и выражение для матрицы $\mathbf{T}_\lambda = \mathbf{X}\mathbf{F}\Lambda^{-1/2}$ можно записать следующую цепочку равенств:

$$\mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda = \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\mathbf{F}^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right]^{-1}\mathbf{F}\mathbf{F}^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right]^{-1}\mathbf{F}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{T}_\lambda =$$

$$= \Lambda^{-1/2}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\mathbf{F}^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right]^{-1}\mathbf{F}\mathbf{F}^{\mathrm{T}}\left[\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\right]^{-1}\mathbf{F}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\mathbf{L}^{-1/2} =$$

$$= \Lambda^{-1/2}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\left[\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\right]^{-1}\left[\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\right]^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{G}\mathbf{X}\mathbf{F}\mathbf{L}^{-1/2}$$

откуда сразу следует (3.10).

Используя свойство (3.10), неравенство (3.9) можно переписать в виде

$$\xi^{\mathrm{T}}\mathbf{T}\begin{bmatrix} 0 & \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0 \\ \mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda & \mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0 \end{bmatrix}\mathbf{T}^{\mathrm{T}}\xi \le 0. \qquad (3.11)$$

Вектор-строку $\xi^{\mathrm{T}}\mathbf{T}$ размерности $1\times N$ можно представить в следующем виде

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

Ясно, что соответствующий $N\times 1$ - вектор-столбец

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

Теперь с учетом блочной структуры матрицы, фигурирующей в (3.11), квадратичную форму (3.8) можно представить в виде

$$\xi^{\mathrm{T}}\left[\mathbf{T}_0\mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_\lambda\mathbf{T}_\lambda^{\mathrm{T}} + \mathbf{T}_\lambda\mathbf{T}_\lambda^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0\mathbf{T}_0^{\mathrm{T}} + \mathbf{T}_0\mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0\mathbf{T}_0^{\mathrm{T}}\right]\xi \le 0. \qquad (3.12)$$

На основе (3.12) можно сделать следующие выводы. Если вектор ошибок ξ принадлежит нуль-пространству матрицы $\mathbf{X}$ ($\mathbf{T}_\lambda^{\mathrm{T}}\xi = 0$), никаким выбором весовой матрицы нельзя улучшить точность оценок. Действительно, матрица $\mathbf{T}_0\mathbf{T}_0^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\tilde{\mathbf{V}}\mathbf{G}\mathbf{T}_0\mathbf{T}_0^{\mathrm{T}}$ 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 (3.12) принимает вид

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

Для всякой другой весовой матрицы $\mathbf{G}$ собственные значения указанной матрицы неотрицательны и указанное неравенство (3.12) (для отличающегося от нулевого) вектора ошибок невозможно. Этот вывод не является неожиданным. Действительно, в случае, когда вектор ошибок принадлежит нуль-пространству, в соответствии с (1.35) ошибка идентификации равна нулю, а всякое изменение ориентации вектора ошибок относительно пространства матрицы $\mathbf{X}$ может лишь ухудшить ситуацию.

Если вектор ξ принадлежит пространству матрицы $\mathbf{X}$, изменением весовой матрицы $\mathbf{G}$ также ничего сделать нельзя. Этот вывод очевиден, т.к. при $\mathbf{T}_0^{\mathrm{T}}\xi = 0$

все слагаемые в левой части (3.12) обращаются в нуль. Во всех остальных случаях ($T_\lambda^T\xi \neq 0$, $T_0^T\xi \neq 0$) в классе диагональных матриц (включая вещественные и комплексные с "чисто" мнимыми элементами) всегда можно подобрать, и притом не единственную, матрицу G, для которой в (3.12) будет иметь место строгое неравенство.

Для того чтобы убедиться в этом, рассмотрим условие, при котором вектор ошибок идентификации по преобразованным в соответствии с равенствами (3.4) данным, равен нулю:

$$\tilde{\mathbf{X}}^T\tilde{\xi} = \mathbf{X}^T\mathbf{G}^2\xi = 0. \qquad (3.13)$$

Так как матрица $\mathbf{G}^2$ диагональная, в равенстве (3.13), в конечном итоге, ничего не изменится, если мы представим его в виде

$$\mathbf{X}^T\Xi\cdot\mathbf{z} = 0, \qquad (3.14)$$

где Ξ - диагональная матрица, составленная из компонент вектора ошибок так, что цифры индексов матрицы совпадают с цифрами индексов соответствующих компонент вектора ξ: $\Xi = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_N)$, а z - N×1 - вектор-столбец, компоненты которого, наоборот, составлены из элементов диагональной матрицы $\mathbf{G}^2 = \mathrm{diag}(g_1^2, g_2^2, \ldots, g_N^2)$, притом так, что $z_i = g_i^2$, $i = \overline{1,N}$.

По предположению вектор ξ и матрица X фиксированы, следовательно, равенство (3.14) можно интерпретировать как недоопределенную систему уравнений с заданной M×N - матрицей $\mathbf{X}^T\Xi$ и N неизвестными компонентами вектора z. Для получения какого-то решения этой системы N-M компонент вектора z могут быть заданы.

С учетом того, что должно выполняться $z_i = g_i^2$, $i = \overline{1,N}$ зададим их вещественными и положительными. Если RankX=M, задаваемые компоненты всегда могут быть подобраны так, что будет существовать решение системы (3.14) относительно оставшихся M компонент. При этом возможно, по крайней мере, два варианта решения: 1) все M компонент неотрицательны; 2) среди полученных M компонент имеются (или все) отрицательные.

Первый вариант приводит к получению вещественной диагональной весовой матрицы $\mathbf{G} = \mathrm{diag}(g_1, g_2, \ldots, g_N)$. Он возможен, если в N - мерном пространстве существует множество Z точек $z \in Z$:

$$Z = \left\{ z:\ \mathbf{h}_i^T \mathbf{z} = 0,\ i = \overline{1,M},\ \mathbf{h}_i^T = \left[x_{1,i}\xi_1, x_{2,i}\xi_2, \ldots, x_{N,i}\xi_N \right], \right.$$
$$\left. z = \left[z_1, z_2, \ldots, z_N \right]^T,\ z_j > 0,\ j = \overline{1,N} \right\}. \qquad (3.15)$$

Ясно, что это требование может выполняться не всегда.

Во втором случае решение существует всегда. Однако среди элементов матрицы **G**, в силу равенств $z_i = g_i^2$, $i = \overline{1,N}$, мы должны допустить комплексные числа, лежащие на мнимой оси (с нулевой вещественной частью). Следовательно, если $\mathbf{X}^T\xi \neq 0$, с использованием множества чисел, лежащих только на мнимой и/или вещественной оси комплексной плоскости, всегда можно построить, и притом не единственную, диагональную матрицу **G** такую, что $\tilde{\mathbf{X}}^T\tilde{\xi} = \mathbf{X}^T\mathbf{G}^2\xi = 0$.

Таким образом, для любой реализации случайного вектора ошибок ξ в исходных данных можно подобрать, в общем случае комплексную или, при выполнении дополнительных условий (3.15), вещественную, матрицу **G** преобразования взвешивания (1.12), (3.4) так, что при этом повышается (если ошибка оценивания до проведения преобразования взвешивания уже не была равна нулю) точность идентификации. Заметим, что комплексность весовой матрицы **G** в данном случае не создает дополнительных проблем. Дело в том, что для формирования системы нормальных уравнений фактически используется не исходная матрица $\mathbf{G} = \mathrm{diag}\left(g_1, g_2, \ldots, g_N\right)$, а $\mathbf{G}^2$, элементы которой всегда вещественны и принимают отрицательные значения лишь в случае, когда соответствующий элемент исходной матрицы **G** оказывается на мнимой оси комплексной плоскости.

В связи со сформулированными выводами уместно обратить внимание на принципиальное отличие рассматриваемого подхода от традиционного. В теории идентификации, опирающейся на статистическую теорию оценивания, для характеристики точности обычно используют ковариационную матрицу ошибок оценивания. Известно [24,51,56], что при наличии ошибок в исходных данных всегда существует нижняя граница для ковариаций ошибок оценивания, определяемая неравенством Крамера Рао. Поэтому сделанный выше вывод о принципиальной возможности достижения (в конкретном эпизоде идентификации) нулевой ошибки оценивания кажется непривычным.

Однако противоречия здесь нет. В настоящей работе точность характеризуется евклидовой нормой вектора ошибок оценивания на конкретной реализации.

Точность, понимаемая в этом смысле, всегда может быть сделана сколь угодно малой. Успех в данном случае зависит лишь от того, удастся ли в конкретном эпизоде оценивания "узнать" вектор ошибок ξ по вектору невязок $\hat{\xi}$. В следующих разделах будут приведены касающиеся этой проблемы результаты и соответствующие методы преобразования данных.

При построении оптимальных по среднеквадратическому критерию оценок (Гаусса-Маркова) фактически также осуществляется взвешивание данных. В методе Гаусса-Маркова роль весовой матрицы играет обратная ковариационная матрица ошибок измерений. Если ошибки в исходных данных независимы, то эта матрица оказывается диагональной, а "весовые" коэффициенты обратно пропорциональны дисперсиям соответствующих компонент. Однако эти коэффициенты не могут быть отрицательными, поэтому сделать ошибку идентификации равной нулю, даже на фиксированном малом наборе данных, как правило, не удается.

Последнее замечание еще раз указывает на плодотворность развиваемого подхода. Выявленные в настоящем разделе закономерности активизируют усилия на построение "хороших" стратегий определения весовых коэффициентов и дают понимание как это делать. Для построения практических алгоритмов идентификации при малом числе наблюдений это может оказаться более полезным, чем анализ асимптотических свойств оценок.

Ниже приводится пример, иллюстрирующий выводы теории. В частности, получены множества векторов ошибок, для которых существует *вещественная* диагональная матрица **G**, применение которой для проведения преобразования взвешивания обеспечивает нулевую ошибку идентификации. Для построения указанных множеств ставился численный эксперимент. Прежде чем перейти к его описанию, обсудим некоторые необходимые теоретические детали.

Произвольный вектор ошибок ξ представим в виде линейной комбинации векторов, принадлежащих пространству матрицы **X** и ее нуль-пространству:

$$\xi = \alpha\xi_{T\lambda} + (1-\alpha)\xi_{To}, \qquad (3.16)$$

где α - некоторое положительное число ($0 \le \alpha \le 1$). Векторы $\xi_{T\lambda}$, ξ_{To} размерности $N \times 1$ принадлежат пространству матрицы $\tilde{\mathbf{X}}$ и ее нуль-пространству соответственно:

$$\xi_{T\lambda} = \sum_{i=1}^{M} c_{\lambda i} f_{\lambda i}, \qquad \xi_{To} = \sum_{i=M+1}^{N} c_{oi} t_{oi},$$

где $c_{\lambda i}, c_{oi}$ - одновременно не обращающиеся в нуль вещественные числа, а $\mathbf{t}_{\lambda i}, \mathbf{t}_{oi}$ собственные векторы матрицы $\mathbf{XX}^T$ (векторы-столбцы матриц $\mathbf{T}_\lambda, \mathbf{T}_o$ соответственно).

Для вектора ошибок (3.16) введем дополнительно нормировку такую, чтобы он принадлежал сфере (1.54) радиуса $R_\xi=1$. Тогда в соответствии с (1.39) можно записать следующее общее выражение для меры близости $\|\Delta\mathbf{c}\|_2^2$:

$$\|\Delta\mathbf{c}\|_2^2 = \frac{\alpha^2}{\|\sum c_{\lambda i}\mathbf{t}_{\lambda i}\|^2\|\xi\|^2}\sum_{i=1}^{M}\lambda_i^{-1}c_{\lambda i}^2, \qquad (3.17)$$

где $\|\xi\|_2^2$ - квадрат евклидовой нормы вектора (3.16).

Если отказаться от требования дополнительной нормировки вектора (3.16) и задать все коэффициенты $c_{\lambda i}^2$, $i=\overline{1,M}$ одинаковыми, с учетом того, что $\|\sum \mathbf{t}_{\lambda i}\|^2 = M$, $i=\overline{1,M}$, из последнего равенства получаем следующее простое соотношение:

$$\|\Delta\mathbf{c}\|_E^2 = \frac{\alpha^2}{M}\sum_{i=1}^{M}\lambda_i^{-1} \qquad (3.18)$$

Если вдобавок потребовать, чтобы лишь один из коэффициентов, например $c_{\lambda i}^2$, $i\in[1,M]$, был отличен от нуля, последнее соотношение упрощается еще более существенно:

$$\|\Delta\mathbf{c}\|_2^2 = \alpha^2\lambda_i^{-1}$$

Здесь i - индекс собственного значения и/или собственного вектора, для которого соответствующий коэффициент $c_{\lambda i}$ отличен от нуля.

Теперь перейдем к рассмотрению конкретного числового примера. Этот пример иллюстрирует образование множества (3.15) точек $z\in Z$ для случая, когда размерности пространства параметров и пространства исходных наблюдений соответственно равны M=2 и N=3. В частности, построена область значений параметра α, для которой существует вещественная диагональная весовая матрица **G**, обеспечивающая нулевую ошибку идентификации. В примере использовались следующие исходные данные.

Матрица **X** и соответствующая ей матрица $\mathbf{A}=\mathbf{X}^T\mathbf{X}$ формировались случайным образом (но с заданным спектром):

$$\mathbf{X} = \begin{bmatrix} 0.691234 & 0.384033 \\ -0.70629 & 1.072726 \\ -0.302148 & -0.796138 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 1.067942 & -0.251647 \\ -0.251647 & 1.932058 \end{bmatrix}.$$

Собственные значения матрицы **A** (ненулевые собственные значения матрицы $\mathbf{X}\mathbf{X}^{\mathrm{T}}$) соответственно равны

$$\lambda_1 = 1, \quad \lambda_2 = 2.$$

Соответствующие этим собственным значениям ортогональные матрицы определялись в виде

$$\mathbf{F} = \begin{bmatrix} 0.965431 & -0.260658 \\ 0.260658 & 0.965431 \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{T}_\lambda \vdots \ \mathbf{T}_0 \end{bmatrix} = \begin{bmatrix} 0.767440 & 0.134762 & 0.626798 \\ -0.402260 & 0.862489 & 0.307084 \\ -0.499223 & -0.487804 & 0.716117 \end{bmatrix}.$$

Здесь и далее для матриц используются те же обозначения, что и в тексте.

Вектор ошибок формировался в соответствии с (3.16) в виде взвешенной суммы векторов, принадлежащих пространству матрицы **X**:

$$\xi_{T\lambda} = \begin{bmatrix} 0.637953 & 0.325431 & -0.697933 \end{bmatrix}$$

и нуль-пространству этой же матрицы:

$$\xi_{To} = \begin{bmatrix} 0.626798 & 0.307084 & 0.716117 \end{bmatrix}^{\mathrm{T}}.$$

При формировании векторов ошибок все весовые коэффициенты принимались одинаковыми и равными единице.

Далее при указанных исходных данных для значений фигурирующего в (3.16) параметра α в интервале [0,1] с шагом 0.1 рассчитывались: векторы ошибок ξ, весовые матрицы $\mathbf{G} = \mathrm{diag}(g_1, g_2, g_3)$ и разность мер близости оценок: $\|\Delta \mathbf{c}\|_2^2 - \|\Delta \tilde{\mathbf{c}}\|_2^2$. Результаты расчетов в систематизированном виде приведены в таблице 3.1.

Таблица 3.1. Оценки достижимой точности при преобразованиях взвешивания

Параметр	Компоненты вектора ошибок ξ			Компоненты диагональной весовой матрицы G			Разность мер близости
α	ξ_1	ξ_2	ξ_3	g_1	g_2	g_3	
0.1	0.627913	0.308919	0.574712	0.963029	0.961019	1.075951	0.0075
0.2	0.629029	0.310753	0.433307	0.913604	0.909809	1.176586	0.0300
0.3	0.630144	0.310753	0.433307	0.841687	0.836468	1.321845	0.0675
0.4	0.63126	0.314423	0.150497	0.717551	0.711647	1.570802	0.1200
0.5	0.632375	0.316258	0.009092	0.275128	0.272312	2.452560	0.1875

Фигурирующая здесь разность мер близости оценок совпадает с величиной $\|\Delta\mathbf{c}\|_2^2$, т.к. весовая матрица **G** определялась из условия ортогональности пространства матрицы **GX** и преобразованного вектора **G**ξ. В таблице приведены данные, соответствующие лишь тем значениям параметра α, при которых весовая матрица **G** вещественна.

Из таблицы видно, что для половины значений ($0 \le \alpha \le 0.5$) параметра α из диапазона $0 \le \alpha \le 1$ возможно построение *вещественной* весовой матрицы, обеспечивающей нулевую ошибку идентификации. На рис. 3.1 приведена геометрическая интерпретация указанного множества векторов ошибок. Из рисунка видно, что это множество образует внутренность конуса, граница которого соответствует предельному значению параметра $\alpha=0.5$. Ясно, что каждому вектору, принадлежащему внутренности конуса в верхней полусфере, соответствует вектор ξ противоположного знака в нижней полусфере.

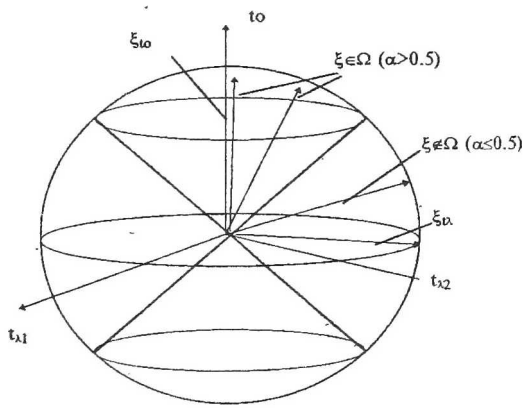


Рис.3.1. Разбиение пространства измерений

Приведенная геометрическая интерпретация, кроме прочего, наглядно иллюстрирует тот факт, что наиболее сложная ситуация при оценивании по методу наименьших квадратов возникает в случае, когда вектор ошибок принадлежит пространству матрицы **X**. При этом, как указывалось выше, никаким выбором весовой матрицы **G** нельзя повысить точность оценок параметров. Возможный путь повышения точности идентификации в этом случае может заключаться в непосредственной корректировке вектора **y** в правой части уравнения (1.2). Для построения процедур такой корректировки необходима дополнительная информация о векторе ошибок измерений. В рамках изучаемой схемы оценивания и сделанных предположений единственным источником такой информации являются реализации вектора невязок. Поэтому представляет интерес установить факт существования и получить обоснованные количественные оценки связи между ошибками и невязками.

3.2. Связь реализаций векторов ошибок и невязок

В настоящем разделе обосновывается принципиальная возможность оценивания реализаций векторов ошибок по реализациям векторов невязок. Эта задача представляется крайне важной, поскольку другие источники информации отсутствуют. Для того чтобы с использованием этих связей можно было строить универсальные алгоритмы идентификации, они не должны быть связаны с конкретными особенностями задачи: схемой измерений, физической размерностью и др. Поэтому изучение этой проблемы мы начнем с выявления и обсуждения общих закономерностей на качественном уровне.

Для этого нам понадобится следующий известный [23] результат. Пусть $\hat{\mathbf{c}}$ МНК-оценка векторного параметра $\mathbf{c}$ уравнения $\mathbf{y} = \mathbf{X}\mathbf{c} + \xi$, а $\hat{\xi} = \hat{\mathbf{y}} - \mathbf{y} = \mathbf{X}\hat{\mathbf{c}} - \mathbf{y}$ - соответствующий этой оценке вектор невязок (1.7). Тогда справедливо равенство

$$\mathbf{T}_0^{\mathrm{T}}\xi = \mathbf{T}_0^{\mathrm{T}}\hat{\xi}, \qquad (3.19)$$

где $\mathbf{T}_0$ - N×(N-M) матрица, составленная из нормированных собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Равенство (3.19) утверждает следующее: проекции вектора ошибок ξ и вектора невязок $\hat{\xi}$ на выбранный базис нуль-пространства матрицы $\mathbf{X}$ совпадают.

Поскольку вектор невязок всегда ортогонален пространству матрицы $\mathbf{X}$, т.е. принадлежит ее нуль-пространству, а также с учетом того, что $\xi = \mathbf{y} - \mathbf{X}\mathbf{c}$ и $\mathbf{T}_0^{\mathrm{T}}\mathbf{X} = 0$ возможны и другие формы записи равенства (3.19):

$$\mathbf{T}_0^{\mathrm{T}}\xi = \mathbf{T}_0^{\mathrm{T}}\mathbf{y} = \mathbf{T}_0^{\mathrm{T}}\hat{\xi} = \hat{\xi}. \qquad (3.20)$$

Из (3.19), (3.20), в частности, следует, что если вектор ошибок принадлежит нуль-пространству, он совпадает с вектором невязок. Ясно, что в некоторой окрестности существует множество (не нулевой меры) векторов ошибок, достаточно близких к соответствующим векторам невязок. Интуитивно кажется, что признаком принадлежности вектора ошибок к этому множеству может служить сравнительно большое значение нормы вектора невязок. Обсудим этот и другие возможные признаки на качественном уровне более подробно.

Для этого обратимся к геометрической интерпретации метода наименьших квадратов, показанной на рис. 3.2. В соответствии с рисунком вектор $\hat{\mathbf{y}}$ является проекцией вектора $\mathbf{y}$ на пространство, натянутое на M векторов независимых переменных, из которых составлена матрица $\mathbf{X}$. Поэтому если в уравнении (1.2) произвести замену вектора (выхода) $\mathbf{y}$ вектором $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{c}}$ (оценок выхода), при по-

вторном решении задачи идентификации будет получена оценка $\hat{c}$, совпадающая с полученной ранее, а вектор невязок также останется прежним.

На основе приведенных геометрических соображений и равенств (3.19), (3.20) можно утверждать следующее:

- малое значение нормы вектора невязок может быть связано с тем, что вектор ошибок принадлежит пространству матрицы **X** (ошибки сильно коррелируют с полезными сигналами), при этом, несмотря на малые невязки, ошибка идентификации может быть очень большой;
- большое значение нормы вектора невязок не может служить достаточным основанием для предположения о большой величине ошибки идентификации, т.к. вектор ошибок может принадлежать нуль-пространству матрицы **X**.

Тем не менее, в N мерном пространстве исходных измерений существует достаточно большое множество векторов ошибок, имеющих промежуточное (по отношению к пространству и нуль-пространству матрицы **X**) направление, для которых значительная составляющая проекции вектора ошибок на нуль-пространство может свидетельствовать также и о большой величине проекции вектора ошибок на пространство матрицы **X**. На основании равенства (3.19) можно также предположить, что выделяющимся компонентам вектора ошибок должны соответствовать выделяющиеся компоненты вектора невязок.

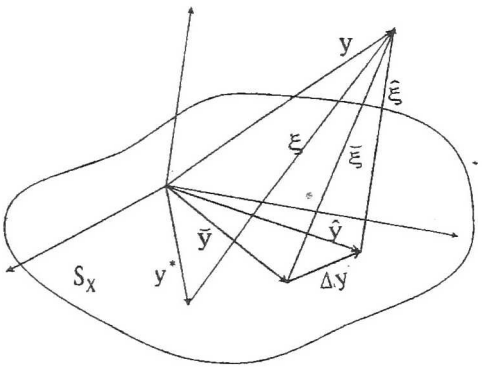


Рис. 3.2. Геометрическая интерпретация МНК

Конечно, из факта равенства проекций векторов на один и тот же базис не следует равенство их компонент. Однако, если какая-либо компонента вектора ошибок значительно превосходит по абсолютной величине остальные, то ее проекция также будет, как правило, превосходить остальные. Следовательно, соответствующая этой проекции компонента вектора невязок также будет выделяющейся. Этот вывод качественного характера был подвергнут экспериментальной проверке и выявлены количественные характеристики связи

Ниже приводятся результаты экспериментов, показывающие, что такая связь действительно существует.

Эксперимент проводился на тестовых изображениях, показанных на рис. 3.3, а,б. Здесь а исходное неискаженное изображение, а б получено из него путем внесения линейных искажений с заданными параметрами модели КИХ-фильтра, который описывается соотношением (1.5), и добавлением гауссовых помех. Размеры опорной области, соответствующие характеристике "размазывающего" фильтра, задавались 5×5, а размеры фрагментов, на которые разбивалось изображение, 25×25 соответственно. В ходе идентификации коэффициентов импульсной характеристики искажающего КИХ-фильтра опорные области на фрагменте размещались "без перекрытия", поэтому размерности векторов ошибок и невязок были 25×1.

а) *б)*

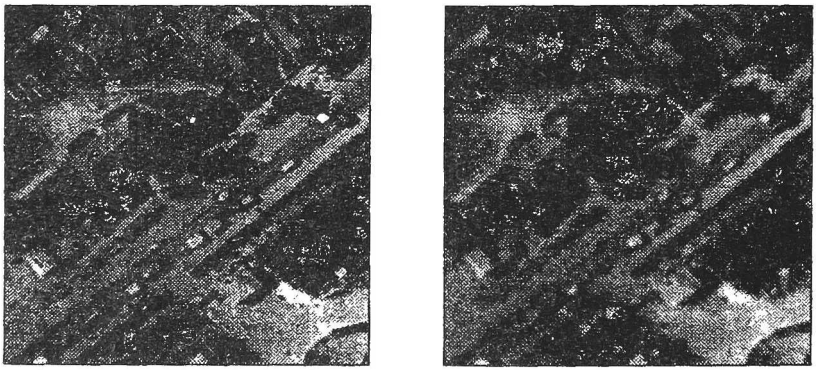


Рис. 3.3. Тестовые изображения: а - входное (неискажённое); б - выходное (искажённое и зашумленное)

После оценки на каждом фрагменте импульсной характеристики вычислялись невязки (1.7). Всего на изображении имелось 100 фрагментов, среди которых 80 информативных [47,62,63]. При моделировании векторов ошибок искусственно вводилась рандомизация, а результаты усреднялись по 100 экспериментам. Цель данного эксперимента для модели КИХ-фильтра заключалась в проверке гипотезы: выделяющимся компонентам вектора ошибок соответствуют выделяющиеся компоненты (с тем же знаком) вектора невязок.

В ходе эксперимента на каждом фрагменте изображения для одноименных компонент векторов ошибок и невязок подсчитывалось число случаев совпадения их знаков, а также вычислялось их отношение. Для фиксированной

реализации шума при разбиении изображения на 100 фрагментов (объем данных на каждом фрагменте - 25) число компонент, для которых знаки совпали в 100% случаев, оказалось равным 14. Как мы и предполагали, ими оказались выделяющиеся по абсолютной величине компоненты вектора ошибок (по 7 на каждом из концов вариационного ряда). Усредненные по всем фрагментам величины отношений компонент вектора ошибок к одноименным компонентам вектора невязок приведены в таблице 3.2.

Таблица 3.2. Отношения компонент

N=25		N=100	
Номер компоненты	Отношение компонент	Отношение компонент	Номер компоненты
1	1.099	0.985	1
2	1.097	1.006	2
3	1.033	1.035	3
4	1.044	1.036	4
5	1.062	1.027	5
6	1.029	1.019	6
7	0.979	1.027	7
⋮	⋮	⋮	⋮
19	0.933	1.043	94
20	0.978	1.038	95
21	1.044	1.036	96
22	1.078	1.032	97
23	1.079	1.038	98
24	1.105	1.032	99
25	1.070	1.025	100

При разбиении изображения на 25 фрагментов (объем выборки на каждом фрагменте 100) знаки совпали у 86 компонент (42 слева и 44 справа). Средние значения отношений компонент вектора невязок к компонентам вектора ошибок для семи первых и семи последних членов соответствующих вариационных рядов приведены в той же таблице. Оказалось, что эти отношения слабо зависят от объема данных, а их числовые значения подтверждают, что выделяющиеся компоненты векторов невязок и ошибок близки по величине и, как правило, компоненты вектора невязок меньше соответствующих компонент вектора ошибок.

Для того чтобы проверить, насколько устойчивы выявленные факты, был проведен также эксперимент, в котором для каждого из 100 фрагментов формировалось 10 различных реализаций нормального шума с нулевым средним и параметром масштаба 0.5 (т.е. эксперимент проводился на 1000 реализациях). Результаты эксперимента приведены в таблице 3.3. Каждая из фигурирующих в таблице цифр подсчитана на 100 фрагментах, т.е. на 100 различных реализациях шума и полезного сигнала. Оказалось, что для выборки объемом 25 наблюдений знаки, по крайней мере, восьми выделяющихся (по 4 на каждом конце вариационного ряда) компонент векторов невязок и ошибок всегда совпадают. Этот эффект устойчиво наблюдался и при любых других малых наборах данных.

Таблица 3.3. Знаки компонент

Номер реализации	Число совпавших знаков "слева"	Число совпавших знаков "справа"
1	7	5
2	7	4
3	4	6
4	8	5
5	4	7
6	7	7
7	5	6
8	7	7
9	6	7
10	7	8

Аналогичные эксперименты проводились с использованием модели БИХ-фильтра. Из соотношения (1.4) видно, что для этого класса моделей компоненты вектора ошибок тесно связаны с характером полезного сигнала (полем яркости на заданном конкретном фрагменте изображения, используемом для идентификации). Поэтому все эксперименты проводились на фиксированном фрагменте.

В эксперименте использовались тестовые изображения, полученные путем моделирования искажений с использованием БИХ-фильтра порядка (2,2), который описывается Z-передаточной функцией вида

$$H(z_1,z_2) = B(z_1,z_2)/A(z_1,z_2), \qquad (3.21)$$

где

$$B(z_1,z_2) = b_{00} + b_{10}z_1 + b_{-10}z_1^{-1} + b_{01}z_2 + b_{0-1}z_2^{-1} +$$
$$+b_{11}z_1z_2 + b_{-1-1}z_1^{-1}z_2^{-1} + b_{1-1}z_1z_2^{-1} + b_{-11}z_1^{-1}z_2$$
$$A(z_1,z_2) = 1 + a_{10}z_1 + a_{-10}z_1^{-1} + a_{01}z_2 + a_{0-1}z_2^{-1} +$$
$$a_{11}z_1z_2 + a_{-1-1}z_1^{-1}z_2^{-1} + a_{1-1}z_1z_2^{-1} + a_{-11}z_1^{-1}z_2 .$$

Переменная z_1 соответствует горизонтальному направлению на изображении, а z_2 - вертикальному. Параметры передаточной функции (3.21) задавались так, чтобы обеспечивалась радиальная симметрия искажений:

$$b_{10} = b_{01} = b_{-10} = b_{0-1}, \qquad b_{11} = b_{1-1} = b_{11} = b_{-1-1},$$

$$a_{10} = a_{01} = a_{-10} = a_{0-1}, \qquad a_{11} = a_{1-1} = a_{11} = a_{-1-1}.$$

При выполнении указанных условий полиномы числителя и знаменателя представляются в следующем компактном виде:

$$B(z_1,z_2) = b_0[1 + b_1\left(z_1 + z_1^{-1} + z_2 + z_2^{-1}\right) + b_2\left(z_1z_2 + z_1^{-1}z_2^{-1} + z_1z_2^{-1} + z_1^{-1}z_2\right)],$$

$$A(z_1,z_2) = 1 + a_1\left(z_1 + z_1^{-1} + z_2 + z_2^{-1}\right) + a_2\left(z_1z_2 + z_1^{-1}z_2^{-1} + z_1z_2^{-1} + z_1^{-1}z_2\right),$$

где $b_0 = 0.02834$, $b_1 = 1.2$, $b_2 = 0.27$, $a_1 = -1.2$, $a_2 = 0.27$. Числовые значения параметров специально подобраны так, чтобы обеспечивалась устойчивость, как искажающего, так и соответствующего ему восстанавливающего (инверсного) фильтра.

исходное тестовое и искаженное изображения приведены на рис. 3.4, а, б соответственно. При моделировании этих изображений исполь-зовался итерационный метод реализации БИХ-фильтров [14]. Шум $\zeta(n_1, n_2)$ моделировался в виде смеси гауссовского с параметрами (M_x=0, D_x=16) и, с вероятностью P=0.03, равномерного в интервале [0-64] импульсного шума.

По сформированным тестовым изображениям осуществлялась идентификация 5-ти параметров фильтра $(b_0, b_1, b_2, a_1, a_2)$ на фрагменте, содержащем 24 наблюдения. Ошибки формировалось в виде векторов одинаковой длины с произвольной ориентацией относительно пространства матрицы **X**. Это достигалось соответствующей нормировкой векторов ошибок, имеющих случайное направление. Для одноименных компонент векторов ошибок и невязок подсчитывались: число совпадений знаков; усредненные по 100 реализациям отношения одноименных компонент и сумма квадратов отклонений этих отношений от найденного среднего значения. Полученные результаты приведены в таблице 3.4.

Номера компонент в первом столбце таблицы упорядочены по убыванию их абсолютных значений. Из таблицы видно, что выделяющимся компонентам вектора ошибок соответствуют выделяющиеся компоненты (с тем же знаком) вектора невязок, а коэффициент связи между ними достаточно устойчив.

а) *б)*

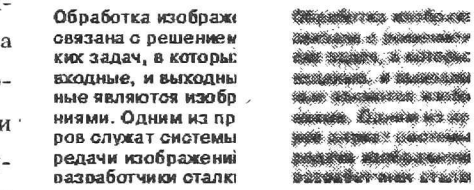


Рис. 3.4. Тестовые изображения: а - исходное; б - искаженное

Таблица 3.4 - Связь реализаций ошибок и невязок для модели БИХ-фильтра

N п./п.	Относитель-ное число совпадений знаков	Отношение $\xi_i/\hat{\xi}_i$	Дисперсия отношений $\xi_i/\hat{\xi}_i$
1	0.992	0.994	0.073
2	0.975	0.987	0.177
3	0.983	1.0480	0.210
4	0.933	1.0167	0.288
5	0.933	1.0689	0.414
6	0.942	1.0962	0.483
7	0.908	1.0250	0.516
8	0.949	1.0065	0.561
9	0.883	1.1810	0.811
10	0.891	1.1005	0.776
11	0.850	1.2369	1.162
12	0.808	1.2023	1.519
13	0.816	1.2514	1.436
14	0.841	1.3291	1.549
15	0.750	1.3395	2.198
16	0.791	1.5177	3.362
17	0.699	1.4881	3.915
18	0.675	1.9899	8.199
19	0.741	2.0329	6.805
20	0.649	2.3555	24.24
21	0.733	3.0389	52.83
22	0.592	3.9789	87.83
23	0.467	5.4769	276.9
24	0.433	12.956	64387

Установленные в экспериментах закономерности могут эффективно использоваться для построения процедур корректировки данных. В частности, приведенные в таблицах 3.2, 3.4 результаты являются достаточным основанием для того, чтобы делать достаточно правдоподобные предположения относительно величины и знака, по крайней мере, пяти выделяющихся компонент вектора ошибок. В связи с этим возникает вопрос: нельзя ли, используя несколько правдоподобно заданных компонент вектора ошибок, восстановить остальные его компоненты. Такая задача и основанный на ее результатах метод идентификации рассматриваются в разделе 4.1.

3.2. Качественный анализ ошибок оценивания при непосредственной корректировке зависимых переменных

Выявленная устойчивая и при малом числе наблюдений связь реализаций векторов ошибок и невязок открывает новые возможности для теоретического обоснования и практической реализации процедур идентификации при отсутствии априорной вероятностной гипотезы относительно распределения ошибок. Действительно, из уравнения (1.2) видно, что если вектор ошибок ξ известен, то исключив его из вектора y в правой части уравнения регрессии можно вычислить точное значение параметра c. Конечно, вектор ошибок никогда не известен, а выявленная связь вектора ξ с вектором невязок $\hat{\xi}$ носит вероятностный характер. Поэтому корректирующий вектор можно задать лишь приблизительно. Далее мы будем обозначать этот вектор $\breve{\xi}$. Пример задания этого вектора показан на рисунке 3.2.

Очевидно, что при задании $\breve{\xi}=\hat{\xi}$ после корректировки сохранится уже полученная МНК- оценка. При задании же этого вектора, даже не слишком отличающимся от вектора невязок, но все же более близким к вектору ошибок хотя бы по некоторым компонентам, после корректировки мы вправе ожидать повышения точности оценок. В настоящем разделе этот вывод, основанный на интуиции, мы подвергнем более детальному анализу.

В первую очередь, нас интересуют локализация и размеры области, к которой принадлежит множество корректирующих векторов $\breve{\xi}$, обеспечивающих повышение точности идентификации. Поскольку на данном этапе будут использоваться элементы качественной теории идентификации [53], как и прежде, мы

будем использовать методический прием, заключающийся в задании фиксированного вектора ошибок ξ.

Обозначим $\Delta\tilde{\mathbf{c}}$ -вектор ошибок оценок, полученных после того как вектор $\mathbf{y}$ заменен на вектор $\tilde{\mathbf{y}} = \mathbf{y} - \tilde{\xi}$. Найдем множество векторов $\tilde{\xi}$, для которых имеет место неравенство

$$\|\Delta\mathbf{c}\|_2^2 > \|\Delta\tilde{\mathbf{c}}\|_2^2. \qquad (3.22)$$

Заметим, что вместо вектора $\tilde{\xi}$ для непосредственной коррекции (вычитания из вектора $\mathbf{y}$) можно использовать показанный на рисунке 3.2 вектор поправок $\Delta\mathbf{y} = \tilde{\xi} - \hat{\xi}$. Это, как и в первом случае, приведет к изменению лишь составляющей вектора ошибок, принадлежащей пространству матрицы $\mathbf{X}$. Компенсация этой составляющей вектора ошибок, по существу, является основной проблемой МНК. Составляющая, принадлежащая нуль-пространству (вектор невязок), как мы уже неоднократно отмечали выше, не влияет на точность МНК-оценок.

Перепишем (3.22) в развернутом виде

$$\xi^T\mathbf{X}\left[\mathbf{X}^T\mathbf{X}\right]^{-2}\mathbf{X}^T\xi > \left(\xi - \tilde{\xi}\right)^T\mathbf{X}\left[\mathbf{X}^T\mathbf{X}\right]^{-2}\mathbf{X}^T\left(\xi - \tilde{\xi}\right).$$

После перемножения и соответствующих преобразований из последнего неравенства получим следующее:

$$\tilde{\xi}^T\mathbf{X}\left[\mathbf{X}^T\mathbf{X}\right]^{-2}\mathbf{X}^T\tilde{\xi} - 2\xi^T\mathbf{X}\left[\mathbf{X}^T\mathbf{X}\right]^{-2}\mathbf{X}^T\tilde{\xi} < 0. \qquad (3.23)$$

Теперь введем замену. Пусть

$$\mathbf{w} = \mathbf{T}^T\tilde{\xi}, \text{ и } \mathbf{v} = \mathbf{T}^T\xi, \qquad (3.24)$$

где $\mathbf{T} = \left[\mathbf{T}_\lambda : \mathbf{T}_0\right]$ ортогональная матрица, которая введена нами в разделе 1.4, $\mathbf{T}_\lambda$ - определяется равенством (1.33), а $\mathbf{T}_0$ та же, что и в (3.19). После подстановки (3.24) в (3.23), с учетом обсуждавшихся в разделе 1.4 свойств матрицы $\mathbf{T}$, получаем

$$\sum_{i=1}^{M}\lambda_i^{-1}w_i^2 - 2\sum_{i=1}^{M}\left(\lambda_i^{-1}v_i\right)w_i < 0. \qquad (3.25)$$

где w_i, v_i - компоненты векторов $\mathbf{w}_\lambda = \mathbf{T}_\lambda^T\tilde{\xi}$, $\mathbf{v}_\lambda = \mathbf{T}_\lambda^T\xi$ соответственно. Таким образом, в $\mathcal{R}(\mathbf{X})$ существует множество векторов $\mathbf{w}_\lambda = \mathbf{T}_\lambda^T\tilde{\xi}$, составленных из проекций векторов $\tilde{\xi}$ на это пространство, удовлетворяющих условию (3.22). Из (3.25) видно, что размеры и локализация этого множества зависят от проекций

векторов ошибок исходных данных $\mathbf{v}_\lambda = \mathbf{T}_\lambda^{\mathrm{T}}\xi$ и собственных значений матрицы Грама, т.е. от полезных сигналов.

Теперь рассмотрим один интересный частный случай, когда корректирующий вектор $\tilde{\xi}$ может быть представлен в виде $\tilde{\xi} = \mathbf{G}\xi$. Вычитание такого поправочного вектора из вектора **y** приводит к изменению вектора ошибок ξ по правилу

$$\tilde{\xi} = (\mathbf{E} - \mathbf{G})\xi . \qquad (3.26)$$

Для преобразованного вектора ошибок (3.26), по аналогии с (3.6), запишем квадрат евклидовой нормы вектора ошибок идентификации:

$$\|\Delta\tilde{\mathbf{c}}\|_2^2 = \Delta\tilde{\mathbf{c}}^{\mathrm{T}}\Delta\tilde{\mathbf{c}} = \xi^{\mathrm{T}}[\mathbf{E} - \mathbf{G}]^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-2}\mathbf{X}^{\mathrm{T}}[\mathbf{E} - \mathbf{G}]\xi . \qquad (3.27)$$

Далее, подставив в неравенство (3.22) соответственно правые части (3.2) и (3.27), после соответствующих преобразований получаем

$$\xi^{\mathrm{T}}\left[\mathbf{G}^{\mathrm{T}}\mathbf{V} + \mathbf{V}\mathbf{G} - \mathbf{G}^{\mathrm{T}}\mathbf{V}\mathbf{G}\right]\xi \geq 0 , \qquad (3.28)$$

где $\mathbf{V} = \mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-2}\mathbf{X}^{\mathrm{T}}$. Можно показать, что всегда можно подобрать вещественную диагональную матрицу **G** так, чтобы матрица в квадратных скобках была неотрицательно-определенной.

Предположим, что **G** строится в классе матриц вида $\mathbf{G}=g\mathbf{E}_N$. Тогда неравенство (3.28) преобразуется к виду

$$(2 - g)\xi^{\mathrm{T}}\mathbf{V}\xi \geq 0 .$$

Применяя ортогональное преобразование $\mathbf{T}^{\mathrm{T}}\mathbf{V}\mathbf{T}$, где $\mathbf{T} = \left[\mathbf{T}_\lambda \ \vdots \ \mathbf{T}_0\right]$, последнее неравенство можно представить в виде

$$(2 - g)\mathbf{w}^{\mathrm{T}}\begin{bmatrix} \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{V}\mathbf{T}_\lambda & \mathbf{T}_\lambda^{\mathrm{T}}\mathbf{V}\mathbf{T}_0 \\ \mathbf{T}_0^{\mathrm{T}}\mathbf{V}\mathbf{T}_\lambda & \mathbf{T}_0^{\mathrm{T}}\mathbf{V}\mathbf{T}_0 \end{bmatrix}\mathbf{w} = (2 - g)\mathbf{w}^{\mathrm{T}}\begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{bmatrix}\mathbf{w} \geq 0 , \qquad (3.29)$$

где $\mathbf{w} = \mathbf{T}^{\mathrm{T}}\xi$. При построении неравенства (3.29) использовались следующие свойства: $\mathbf{X}^{\mathrm{T}}\mathbf{T}_0 = 0$, $\mathbf{T}_\lambda^{\mathrm{T}}\mathbf{V}\mathbf{T}_\lambda = \Lambda^{-1/2}\mathbf{F}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-2}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{F}\Lambda^{-1/2} = \Lambda^{-1}$.

Легко заметить, что в классе вещественных положительных весовых матриц **G** неравенство (3.29) выполняется при $0<g\leq 2$. Для отрицательных значений этого параметра: $g\in(-2,0)$ знак поправки необходимо изменить на противоположный. Очевидно, что максимальный выигрыш имеет место при $g=1$ ($g=-1$) и в соответствии с (3.22) равен $\|\Delta c\|_2^2$. При этом ошибка идентификации равна нулю.

На практике получить выигрыш в точности идентификации не так просто, как это может показаться из приведенных результатов, поскольку вектор ошибок не известен. Тем не менее, установленные неравенства имеют важное значение. Они показывают, как следует использовать информацию о векторе ошибок для построения вектора поправок. Информацию же о векторе ошибок мы можем получить из вектора невязок.

В предыдущем разделе мы показали существование устойчивых коэффициентов связи между выделяющимися компонентами векторов ошибок и невязок. Рассмотрим возможность использования этой информации для формирования компонент корректирующего вектора путем умножения невязки на коэффициент. Другими словами, мы хотим рассмотреть возможность формирования корректирующего вектора непосредственно по вектору невязок: $\tilde{\xi} = G\hat{\xi}$.

Если вектор $\tilde{\xi}$ построен из вектора невязок $\hat{\xi}$ путем умножения каждой его компоненты на некоторый коэффициент g_i, его можно представить в виде

$$\tilde{\xi} = \hat{\Xi} \cdot \mathbf{g}, \qquad (3.30)$$

где $\hat{\Xi}$ диагональная N×N - матрица, составленная из компонент вектора невязок (порядок следования двойных индексов такой же, как и для исходного вектора невязок), а **g** - вектор коэффициентов "связи", компонентами которого, наоборот, являются элементы диагональной матрицы **G**. Если среди компонент вектора невязок отсутствуют нулевые, с учетом (3.24) на основании равенства (3.30) можно записать соотношение для определения вектора **g**:

$$\mathbf{g} = \hat{\Xi}^{-1}\mathbf{T}\mathbf{w}. \qquad (3.31)$$

При указанных условиях для фиксированных **X** (Rank**X**=M) и **y** N×N - матрица $\hat{\Xi}^{-1}\mathbf{T}$ также фиксирована и невырождена, а среди чисел λ_i i=1,M, фигурирующих в (3.25), отсутствуют нулевые. Следовательно, каждой точке множества (3.25) можно поставить в соответствие, по крайней мере, один вектор **g**, для которого выполняется неравенство (3.22). т.е. точность оценивания параметров **c** после корректировки (вычитания вектора $\tilde{\xi}$ из вектора **y**) выше, чем по исходным данным.

Выявленные условия, при которых возможно построение диагональной матрицы коэффициентов для преобразования вектора невязок в корректирующий вектор, вообще говоря, очевидны. Тем не менее, они акцентируют внимание на некоторых важных требованиях, например, отличие всех компонент вектора не-

вязок от нуля. Это требование будет учитываться при построении методов идентификации в следующей главе. В частности, из этого факта следует, что стратегия корректировки должна быть достаточно осторожной и учитывать абсолютные величины компонент вектора невязок на конкретной реализации.

В заключение приведем результаты количественных оценок разности мер, фигурирующих в левой и правой части неравенства (3.22), полученные в численном эксперименте. Воспользуемся той же схемой исследования, что и в предыдущем разделе.

Предположим, что вектор ошибок ξ можно записать в виде (3.16), т.е. любая реализация вектора ошибок может быть представлена в виде линейной комбинации векторов, принадлежащих пространству матрицы **X** и ее нуль-пространству. Зададим матрицу преобразования в (3.26) так, чтобы свести ошибку оценивания по преобразованным данным к нулю: $\mathbf{G}=\mathbf{E}_N$. Ясно, что при этом $\|\Delta\tilde{\mathbf{c}}\|_2^2 = 0$, а выигрыш в точности $\|\Delta\mathbf{c}\|_2^2 - \|\Delta\tilde{\mathbf{c}}\|_2^2 = \|\Delta\mathbf{c}\|_2^2$.

Принципиальное отличие от рассмотренного ранее метода взвешивания данных в данном случае будет заключаться в том, что выигрыш может иметь место для всех допустимых значений параметра α ($0<\alpha<1$). Максимальный выигрыш достигается при $\alpha=1$, т.е. в случае, когда вектор ошибок принадлежит пространству матрицы **X**. Если же вектор ошибок принадлежит нуль-пространству ($\alpha=0$), ошибка идентификации равна нулю и вектор **y** не нуждается в корректировке. Если при этом из него вычесть вектор невязок, то в соответствии с геометрической интерпретацией МНК мы получим вектор **y***, составленный из точных значений зависимых переменных.

Ниже приводятся результаты численных расчетов для примера аналогичного рассмотренному в предыдущем разделе. Здесь специально были сформированы случайным образом другие матрицы **X** и **A**, но такие, что их размеры (N=3, M=2) и собственные значения $\lambda_i(\mathbf{A})$, $i = \overline{1,2}$ ($\lambda_1 = 1$, $\lambda_2 = 2$) совпадали с теми, которые уже использовались в разделе 3.1.

Матрица **X** и соответствующая ей матрица $\mathbf{A}=\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 0.617175 & -1.077193 \\ 0.921965 & -0.151531 \\ -0.748709 & 0.158762 \end{bmatrix}, \qquad \mathbf{A} = \begin{bmatrix} 1.791489 & 0.406244 \\ 0.406244 & 1.208511 \end{bmatrix}.$$

Соответствующие ортогональные матрицы:

$$\mathbf{F} = \begin{bmatrix} -0.45663 & 0.889657 \\ 0.889657 & 0.456630 \end{bmatrix}, \quad \mathbf{T} = [\mathbf{T}_\lambda \vdots \mathbf{T}_0] = \begin{bmatrix} -0.676511 & -0.736064 & 0.023278 \\ -0.555807 & 0.531064 & 0.639570 \\ 0.483127 & -0.419738 & 0.768381 \end{bmatrix}$$

(обозначения те же, что и ранее).

Вектор ошибок формировался в виде линейной комбинации (3.16) из векторов

$$\xi_{T\lambda} = [-0.998842 \quad -0.017496 \quad 0.044823]^T,$$

$$\xi_{T0} = [0.023278 \quad 0.639570 \quad 0.768381]^T$$

для различных значений параметра α ($0 \le \alpha \le 1$) с шагом 0.1. Моделировалось два варианта множеств векторов ошибок: без дополнительной нормировки и с нормировкой. При отсутствии нормировки концы векторов принадлежат поверхности конуса, а после проведения нормировки ($\|\xi\|_2 = R_\xi = 1$) сфере (1.54). Соответствующие указанному набору параметров α разности мер близости оценок $\|\Delta\mathbf{c}\|_2^2 - \|\Delta\tilde{\mathbf{c}}\|_2^2$, рассчитанные по формулам (3.17) и (3.18) для нормированного и ненормированного векторов ошибок приведены в таблице 3.5.

Таблица 3.5. Количественные оценки достижимой точности при непосредственной корректировке вектора выхода модели

Параметр	Компоненты нормированного ($\|\xi\|_e = 1$) вектора ошибок			Компоненты ненормированного ($\|\xi\|_e \neq 1$) вектора ошибок			Разность мер близости для:	
	ξ_1	ξ_2	ξ_3	g_1	g_2	g_3	$\|\xi\|_e = 1$	$\|\xi\|_e \neq 1$
0.0	0.023278	0.639570	0.768381	0.023278	0.639570	0.768381	0.00000	0.0000
0.1	-0.087168	0.633726	0.768631	-0.078934	0.573863	0.696025	0.00915	0.0075
0.2	-0.219672	0.616230	0.756310	-0.181146	0.508156	0.623669	0.04412	0.03
0.3	-0.372067	0.580965	0.723910	-0.283358	0.442450	0.551313	0.11638	0.0675
0.4	-0.534689	0.522449	0.664195	-0.385570	0.376743	0.478958	0.23077	0.12
0.5	-0.689828	0.439873	0.575022	-0.487782	0.311037	0.406602	0.37500	0.1875
0.6	-0.818174	0.340212	0.463516	-0.589994	0.245330	0.334246	0.51923	0.2700
0.7	-0.908911	0.235858	0.343879	-0.692206	0.179624	0.261890	0.63362	0.3675
0.8	-0.963373	0.138145	0.229844	-0.794418	0.113917	0.189534	0.70588	0.4800
0.9	-0.990162	0.053240	0.129402	-0.896630	0.048211	0.117178	0.74085	0.6075
1.0	-0.998842	-0.017496	0.044823	-0.998842	-.017496	0.044823	0.7500	0.7500

Из сравнения данных таблиц 3.1, 3.5 видно, что при значениях параметра $\alpha \le 0.5$ числовые значения разностей мер, вычисленные для обоих методов, совпадают. При $\alpha > 0.5$ выигрыш в точности для метода непосредственной

корректировки вектора y продолжает р[illegible]

ния при этих значениях параметра α весовую матрицу, приводящую к повышению точности оценок, в классе *вещественных* построить не удается. Этот выигрыш при взвешивании данных может быть получен при использовании комплексной весовой матрицы. Из таблицы также видно, что мера близости в случае простой линейной "смеси" векторов (3.16) всегда меньше чем для случая, когда вектор ошибок дополнительно нормируется так, что принадлежит единичной сфере.

Конечно, на практике получить такой заметный, как указано в таблицах, выигрыш в точности идентификации за счет корректировки данных вряд ли удастся, т.к. вектор ошибок не известен. Тем не менее, приведенные примеры полезны, т.к. дают наглядное представление о закономерностях и предельных возможностях повышения точности идентификации. В частности, из геометрических представлений и результатов расчета видно, что путем непосредственной корректировки вектора **y**, в принципе, всегда можно свести ошибку оценивания к нулю. В то же время преобразование взвешивания в случае, когда вектор ошибок исходных данных принадлежит пространству матрицы **X**, ничего не дает.

Во всех остальных случаях ошибка идентификации может быть сделана как угодно малой с использованием любого преобразования (взвешивания и/или непосредственной корректировки данных). Конечно, как уже неоднократно указывалось выше, эти возможности не всегда могут быть реализованы, т.к. для выбора параметров преобразований доступной является лишь информация о реализациях вектора невязок.

3.4. Влияние масштабирования независимых переменных

В разделе 2.4 мы показали целесообразность масштабирования информационной матрицы **A** по правилу $\tilde{\mathbf{A}} = \mathbf{DAD}$ с целью увеличения ее диагонального преобладания. После такого преобразования показатель диагонального преобладания информационной матрицы оказывается мерой мультиколлинеарности векторов независимых переменных уравнения (1.2). Поскольку операция масштабирования улучшает обусловленность матрицы, возникает естественный вопрос: не оказывает ли она также влияние и на точность оценок? Для получения ответа на указанный вопрос в настоящем разделе мы подвергнем этот тип линейного преобразования качественному анализу.

Операция масштабирования матрицы **A** эквивалентна тому, что перед ее формированием по правилу $\mathbf{A} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ каждый из векторов-столбцов $\mathbf{x}_i$, $i=1,M$ матрицы **X** умножается на скалярный множитель d_i. Поэтому масштабирование можно трактовать как операцию изменения физической размерности независимых переменных (отсчетов поля яркости в соответствующих точках опорной области изображения).

Если ввести в рассмотрение диагональную M×M- матрицу

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & \\ \vdots & & & \vdots \\ 0 & & \cdots & d_M \end{bmatrix}, \qquad (3.32)$$

то операция масштабирования информационной матрицы **A**, определяемая формулой (2.43):

$$\tilde{\mathbf{A}} = \mathbf{DAD} = \mathbf{D}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]\mathbf{D}, \qquad (3.33)$$

равносильна тому, что **X** представляется в виде

$$\tilde{\mathbf{X}} = \mathbf{XD}, \qquad (3.34)$$

а затем определяются матрица $\tilde{\mathbf{A}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]$ и вектор $\tilde{\mathbf{b}} = \tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{y}$ системы уравнений вида (1.19). Можно поступить иначе. Указанную систему уравнений (1.19) представить в виде

$$\mathbf{DAD}\tilde{\mathbf{c}} = \mathbf{Db}, \qquad (3.35)$$

где

$$\tilde{\mathbf{c}} = \mathbf{D}^{-1}\hat{\mathbf{c}}. \qquad (3.36)$$

Следуя выводам, сформулированным в разделе 2.4, элементы матрицы **D** будем брать в виде величин обратно пропорциональных евклидовым нормам соответствующих векторов независимых переменных: $d_i = \|\mathbf{x}_i\|_2^{-1}$. Полученная в результате масштабирования матрица будет матрицей сопряженности ($\tilde{\mathbf{A}}$=**R**).

Выясним, изменится ли точность оценивания вектора параметров (в исходной физической размерности) при его вычислении с использованием преобразованных, одним из указанных способов, данных.

Пусть наряду с МНК-оценкой

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

получена оценка

$$\tilde{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{y}$$

для преобразованных по правилу (3.33), (3.34) или (3.35) данных. Подставляя в последнее равенство в соответствии с (3.34) $\tilde{\mathbf{X}} = \mathbf{XD}$, получаем

$$\tilde{\mathbf{c}} = \left[\mathbf{DX}^{\mathrm{T}}\mathbf{XD}\right]^{-1}\mathbf{DX}^{\mathrm{T}}\mathbf{y} = \mathbf{D}^{-1}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{D}^{-1}\hat{\mathbf{c}}. \qquad (3.37)$$

Для того чтобы можно было сравнивать оценки $\hat{\mathbf{c}}$ и $\tilde{\mathbf{c}}$ по точности, необходимо привести их к одинаковой физической размерности. Для этого в соответствии с (3.36) полученную оценку $\tilde{\mathbf{c}}$ необходимо умножить на диагональную матрицу **D**. При этом, поскольку в соответствии с (3.36), (1.32)

$$\mathbf{DAD}\tilde{\mathbf{c}}^{*} = \mathbf{D}(\mathbf{b} - \zeta), \qquad \mathbf{c}^{*} = \mathbf{D}\tilde{\mathbf{c}}, \qquad \tilde{\mathbf{c}}^{*} = \mathbf{D}^{-1}\mathbf{c}^{*},$$

и $$\Delta\tilde{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\xi = \mathbf{D}^{-1}\mathbf{A}^{-1}\mathbf{D}^{-1}\mathbf{DX}^{\mathrm{T}}\xi = \mathbf{D}^{-1}\mathbf{A}^{-1}\mathbf{X}^{\mathrm{T}}\xi = \mathbf{D}^{-1}\Delta\mathbf{c},$$

получаем

$$\mathbf{D}\tilde{\mathbf{c}} = \mathbf{D}\left(\tilde{\mathbf{c}}^{*} + \Delta\tilde{\mathbf{c}}\right) = \mathbf{c}^{*} + \Delta\mathbf{c} = \hat{\mathbf{c}}.$$

Следовательно, масштабирование независимых переменных не изменяет точности оценок.

Этот, вообще говоря, очевидный вывод часто не подтверждается на практике. В частности, решая вместо исходной системы уравнений (1.19) преобразованную (3.35) и переходя затем с учетом (3.36) к исходным физическим размерностям компонент вектора искомых параметров, можно получить заметно отличающийся результат. Связано это с вычислительными аспектами. Во-первых, при масштабировании независимых переменных вносится дополнительная погрешность, связанная с операцией умножения, которая при работе с исходной системой отсутствовала. Во-вторых, и это наиболее существенно, изменяется эквивалентное возмущение обратной матрицы, вследствие изменения ее обусловленности.

Если плохая обусловленность исходной задачи связана с большими различиями в масштабах независимых переменных, то описанное выше масштабирование, как правило, приводит к (возможно незначительному) повышению точности оценок. Следовательно, эта операция может применяться без ущерба для точности оценок с целью повышения диагонального преобладания информационной матрицы. Поскольку при этом исключается фактор плохой обусловленности из-за различий в масштабах независимых переменных, наиболее важным следствием следует считать возможность повышения надежности контроля информативности данных без привлечения дополнительных вычислительных ресурсов.

Анализ операции масштабирования может быть полезен также при выборе физических размерностей полей яркости входного и выходного изображений на стадии планирования экспериментов, связанных с решением задачи оценивания характеристик видеоинформационного тракта. В частности, из приведенных соотношений видно, что для уменьшения ошибки идентификации $\|\Delta\mathbf{c}\|_2^2$ целесообразно выбирать такие масштабы (физические размерности) независимых переменных, при которых обеспечивается приблизительное равенство их норм: $\|\mathbf{x}_i\|_2 = \text{Const}$, а масштаб матрицы **A** должен быть по возможности максимальный. Эти рекомендации аналогичны традиционному требованию увеличения отношения полезный сигнал/шум. Заметим, что для моделей типа БИХ-фильтра матрица **X** "распадается" на два блока, формируемых по входным и выходным сигналам. Обеспечение близости норм для векторов-столбцов в разных блоках матрицы **X** в этом случае сводится к выравниванию физических размерностей (яркостей) входного и выходного изображений.

Методы получения оценок (1.11) в рамках схемы (1.13), (1.17), по существу, различаются лишь способом преобразования данных на промежуточных шагах. В настоящей главе первый раздел посвящен методам идентификации, основанным на непосредственной корректировке вектора выхода модели, а второй — на преобразованиях взвешивания. Затем будет рассмотрена связь метода взвешивания со статистическими методами и основанные на этой связи модификации известных методов идентификации.

4.1. Методы идентификации, основанные на непосредственной корректировке вектора выхода модели

В рамках развиваемого подхода к решению задач идентификации по малому числу наблюдений существенным является предположение об отсутствии априорной информации о вероятностном распределении помех. В данном случае доступными являются лишь реализации векторов невязок. Поэтому основой рассматриваемых методов является выявленная выше связь выделяющихся компонент векторов ошибок и невязок. Хотя эта информация имеет вероятностный характер, она принципиально отличается от традиционно используемых априорных вероятностных распределений. При малом числе наблюдений устойчивость распределений ошибок *еще* не проявляется, а указанные характеристики связи оказываются *уже* устойчивыми.

Информация об отдельных компонентах вектора ошибок может быть получена также путем проведения повторных, возможно более точных, измерений отдельных отсчетов, анализа двумерной функции яркости на конкретном фрагменте, сопоставления данных с измерениями, полученными от других источников информации и др. Подчеркнем, что в любом случае эта информация является дополнительной [54], а не априорной статистической, т.к. она появляется лишь в ходе оценивания.

В настоящем разделе мы рассмотрим метод, основанный на различных процедурах непосредственной корректировки вектора выхода модели. Эти процедуры строятся в рамках алгебраического подхода. Предполагается, что используемая при их реализации дополнительная информация о связях выделяющихся компонент векторов ошибок и невязок может быть получена из различных источников, перечисленных выше.

...тода является построение фигурирующего в (1.11) вектора поправок Δy по вектору невязок $\hat{\xi}$ для каждого фиксированного набора данных. В качестве поправочных будем использовать векторы $\Delta y = \breve{\xi}$ и/или $\Delta y = \breve{\xi} - \hat{\xi}$ и оба вектора $(\Delta y, \breve{\xi})$ далее будем называются корректирующими. Из рис. 3.2 видно, что при использовании любого из указанных корректирующих векторов точность оценивания будет одинаковой. К обсуждению этого факта мы еще вернемся при изложении метода.

Предположим, что каким-то образом (с использованием характеристик связи ошибок и невязок или путем дополнительных более точных измерений нескольких отсчетов [54]) удалось задать ровно M (любых) компонент вектора $\breve{\xi}$ 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 $\breve{\xi}$, обозначим $\breve{\xi}_M$.

Далее введем в рассмотрение (N-M)×1-вектор $\breve{\xi}_{N-M}$, составленный из оставшихся компонент вектора $\breve{\xi}$, которые нам пока не известны. Корректирующий N×1-вектор $\breve{\xi} = \left[\breve{\xi}_M : \breve{\xi}_{N-M}\right]$ представляет собой прямую сумму векторов $\breve{\xi}_M$ и $\breve{\xi}_{N-M}$. Задача заключается в определении недостающих компонент, т.е. вектора $\breve{\xi}_{N-M}$, по известным M компонентам вектора $\breve{\xi}_M$, которые удалось задать с использованием доступной информации.

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

$$T_0^T\left(\breve{\xi} - \hat{\xi}\right) = 0. \qquad (4.1)$$

Основанием для этого служат следующие соображения. В соответствии с (3.19) $T_0^T\xi = T_0^T\hat{\xi}$. Следовательно, если компоненты вектора $\breve{\xi}$ совпадают с истинными значениями компонент вектора ошибок: $\breve{\xi} = \xi$, то корректирующий вектор $\Delta y = \breve{\xi} - \hat{\xi} = \xi - \hat{\xi}$, удовлетворяющий (4.1), будет ортогонален нуль-пространству. Свойство ортогональности вектора Δy с нуль-пространством сохранится и в случае, когда это равенство приблизительное: $\breve{\xi} \cong \xi$. Таким образом, реализация

условия (4.1) позволит оценить составляющую вектора ошибок, принадлежащую пространству матрицы **X**.

С использованием ортогональной матрицы T_0, построенной для матрицы **X**, которая уже подверглась перестановке и перенумерации строк так, что ее первые M строк соответствуют M компонентам вектора $\tilde{\xi}_M$, а последние N-M компонентам вектора $\tilde{\xi}_{N-M}$, в соответствии с условием (4.1) можно сформировать следующую систему уравнений

$$T^T_{0,N-M}\tilde{\xi}_{N-M} = t_{0,M}. \qquad (4.2)$$

Здесь $T^T_{0,N-M}$ - матрица составленная из последних N-M столбцов матрицы T^T_0. Вектор $t_{0,M}$ размерности (N-M)×1 формируется в соответствии с равенством

$$t_{0,M} = T^T_0\hat{\xi} - T^T_{0,M}\tilde{\xi}_M, \qquad (4.3)$$

где $T^T_{0,M}$ - (N-M)×M - матрица, составленная из первых M столбцов той же матрицы T^T_0.

Так как $RankT_0=(N-M)$, существует единственное решение системы (4.2), а искомый вектор $\tilde{\xi}_{N-M}$ определяется как

$$\tilde{\xi}_{N-M} = T^{-1}_{0,N-M}\hat{t}_{0,M}. \qquad (4.4)$$

Это решение удовлетворяет равенству (4.1). Поэтому если M компонент будут заданы равными истинным значениям компонент вектора ошибок: $\tilde{\xi}_M = \xi_M$, остальные, вычисленные в соответствии с (4.4), N-M компонент также будут совпадать с соответствующими истинными ошибками: $\tilde{\xi}_{N-M} = \xi_{N-M}$. По-видимому, при задании близких к истинным значений компонент $\tilde{\xi}_M \approx \xi_M$ есть основания ожидать близкое к ξ_{N-M} решение: $\tilde{\xi}_{N-M} \approx \xi_{N-M}$.

Из (4.1) следует также, что при задании компонент вектора $\tilde{\xi}_M$ равными соответствующим компонентам вектора невязок: $\tilde{\xi}_M = \hat{\xi}_M$, в (4.4) 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 улучшить результат оценивания, мы должны в компоненты вектора $\tilde{\xi}_M$ "вложить" дополнительное знание. Это могут быть любые

априорные сведения, в т.ч. о существующих между реализациями $\hat{\xi}$ и ξ устойчивых связях.

Еще раз подчеркнем, что в соответствии с (4.1) корректирующий вектор $\Delta y = \tilde{\xi} - \hat{\xi}$ всегда ортогонален нуль-пространству. С другой стороны, вектор невязок принадлежит нуль-пространству. Следовательно, найденный описанным выше способом вектор $\tilde{\xi}$ можно использовать для непосредственной корректировки (вычитания его из вектора y). Ясно, что это приведет к изменению составляющей вектора ошибок, принадлежащей пространству матрицы X, которая, по существу, и определяет точность оценивания. Ниже приводятся результаты экспериментальных исследований, подтверждающие возможность построения высокоточных оценок с использованием описанного метода.

В эксперименте использовались показанные на рисунках 3.4, а и б тестовые изображения, которые были получены путем моделирования искажений с использованием БИХ-фильтра порядка (2,2) с передаточной функцией (3.21). Параметры передаточной функции задавались такими же, как и в экспериментах, описанных в разделе 3.2. Опорные области на фрагменте размещались "без перекрытия", поэтому размерности векторов ошибок и невязок были такими же: 25×1.

В соответствии со схемой идентификации (1.13)-(1.17) задача решалась итерационно. Вначале строилась МНК-оценка. С ее использованием вычислялся вектор невязок и задавался вектор $\tilde{\xi}_M$. Затем, с использованием полученного вектора невязок по соотношениям (4.2)-(4.4) вычислялся корректирующий вектор $\tilde{\xi}$, осуществлялась корректировка вектора y и вновь вычислялась МНК-оценка.

Цель эксперимента заключалась в том, чтобы показать принципиальную возможность увеличения точности идентификации по мере более точного задания небольшого числа компонент вектора ошибок. Для этого M (в данном случае 5) компонент вектора $\tilde{\xi}_M$ задавались с возрастающей точностью в соответствии с формулой

$$\tilde{\xi}_M = 1.4\alpha\xi_M + (1 - 1.4)\alpha\hat{\xi}_M . \qquad (4.5)$$

Для сравнения точности идентификации для каждого заданного фиксированного α вычислялась скалярная характеристика относительной погрешности в виде

$\|\Delta\mathbf{c}\|_2/\|\mathbf{c}\|_2$, где $\|\Delta\mathbf{c}\|_2$ и $\|\mathbf{c}\|_2$ евклидовы нормы векторов ошибок оценивания и параметров модели соответственно. Результаты приведены в таблице 4.1.

Таблица 4.1. Связь точности оценивания с точностью задания вектора $\tilde{\xi}_M$

Параметр	Параметры модели БИХ-фильтра					$\frac{\|\Delta\mathbf{c}\|}{\|\mathbf{c}\|}$
α	b0	b1	b2	a1	a2	
Параметры	0.028340	0.034008	0.007652	-1.200000	0.270000	
МНК-оценки	0.022962	0.057481	0.006793	-1.029265	0.114629	0.189
α=0	0.023002	0.057663	0.006808	-1.026394	0.111991	0.191
α=0.14	0.024483	0.050960	0.008005	-1.067358	0.149081	0.147
α=0.28	0.026047	0.044293	0.009344	-1.105419	0.183520	0.104
α=0.42	0.027685	0.037637	0.010813	-1.141070	0.215756	0.065
α=0.56	0.029445	0.031336	0.012390	-1.169678	0.241525	0.034
α=0.71	0.031265	0.025105	0.014061	-1.195811	0.265037	0.011
α=0.86	0.033172	0.019205	0.015788	-1.216209	0.283283	0.022
α=1	0.035153	0.013615	0.017550	-1.231559	0.296894	0.039

Из таблицы 4.1 видно, что при задании компонент вектора $\tilde{\xi}_M$ равными невязкам (α=0) повышения точности, по сравнению с МНК, как и следовало ожидать, не происходит. Однако по мере более точного задания этих компонент происходит увеличение точности и при $\tilde{\xi}_M \cong \xi_M$ (α=0.71) ошибка составляет всего лишь около одного процента. Эта ошибка связана с неизбежными погрешностями округления. Заметим, что погрешность МНК-оценок при этом составляет почти 20%.

В данном случае уточнению подвергались 19 компонент вектора ошибок по заданным 5-ти компонентам. Для получения дополнительной информации об этих компонентах, как указывалось, могут использоваться различные источники. В частности, для этого можно осуществлять предварительное моделирование различных вариантов реализаций векторов ошибок на заданном или "похожем" тестовом изображении.. В шестой главе будут приведены результаты исследования основанных на этой идее информационных технологий, в которых для идентификации используются типовые объекты заданной конфигурации. Нормализованные изображения таких объектов имеют известное поле яркости. В этом случае коэффициенты связи для выделяющихся компонент векторов ошибок и невязок, необходимые для задания вектора $\tilde{\xi}_M$, могут быть определены по результатам моделирования более точно, за счет использования информации о координатах соответствующих отсчетов поля яркости.

Теперь обсудим как может использоваться корректирующий вектор $\tilde{\xi}$ для повышения точности искомых оценок параметров. Наиболее простой способ корректировки данных заключается в вычитании вектора $\tilde{\xi}$ из вектора **y**. Как уже указывалось выше, составляющая вектора $\tilde{\xi}$, принадлежащая нуль-пространству, не влияет на точность оценок. Поэтому для достижения той же точности достаточно вычесть из вектора **y** корректирующий вектор $\Delta\mathbf{y} = \tilde{\xi} - \hat{\xi}$, удовлетворяющий условию ортогональности с нуль-пространством (4.1). Заметим, что это равносильно тому, что сначала из вектора выхода модели вычитается вектор невязок (эта корректировка не влияет на точность оценивания), а затем вычитается корректирующий вектор $\tilde{\xi}$.

Если получен корректирующий вектор $\tilde{\xi}$, то с его использованием можно также реализовать и преобразование взвешивания. В частности, можно построить диагональную весовую матрицу $\mathbf{G}^2$, удовлетворяющую условию (3.13): $\mathbf{X}^T\mathbf{G}^2\tilde{\xi} = 0$. Для этого условие (3.13) представляется в виде (3.14): $\left[\mathbf{X}^T\tilde{\Xi}\right]\cdot\mathbf{z} = 0$, где $\tilde{\Xi} = \mathrm{diag}\left(\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_N\right)$ диагональная матрица, составленная из компонент вычисленного указанным выше способом корректирующего вектора. Для получения какого-то решения N-M компонент $z_i = g_i^2$, $i = \overline{1,N}$ вектора **z** должны быть заданы (произвольными или из каких-либо соображений). Если Rank**X**=M, а заданные компоненты не нулевые, существует решение системы (3.14) относительно оставшихся M компонент.

Напомним, что задание или появление в ходе решения уравнения (3.14) отрицательных компонент не создает принципиальных вычислительных трудностей, поскольку в выражении для МНК-оценок фигурирует матрица $\mathbf{G}^2$. Поэтому задаваемые компоненты формально могут быть любыми вещественными числами. В частности, они могут задаваться из соображений обеспечения хорошей обусловленности задачи, т. е. так, чтобы нормы векторов независимых переменных имели близкие значения (см. раздел 3.4 влияние масштабирования).

Описанный способ взвешивания данных по точности даст тот же результат, что и непосредственная корректировка вектора выхода модели **y**, однако с вычислительной точки зрения он представляется не рациональным. Взвешивание

данных оправдано, когда оно осуществляется непосредственно по невязкам. В следующем разделе мы рассмотрим простой в вычислительном отношении оптимизационный метод построения весовой матрицы, не требующий предварительного построения корректирующих векторов $\tilde{\xi}$ и/или $\Delta y = \tilde{\xi} - \hat{\xi}$.

Заметим, что вычислительная сложность описанного метода определения корректирующего вектора $\tilde{\xi}$ в основном связана с необходимостью вычисления базиса нуль--пространства. Применение метода, конечно, оправдано в тех случаях, когда базис нуль-пространства все равно используется на других этапах сквозной информационной технологии. Важным достоинством метода является возможность достижения высокой точности. Во многих случаях это может служить достаточной компенсацией вычислительной сложности. Если увеличение вычислительной сложности алгоритма недопустимо, можно применять более простые методы корректировки вектора выхода модели.

В частности, можно вообще отказаться от определения всех N компонент корректирующего вектора ошибок и корректировать лишь те компоненты вектора выхода, которые соответствуют выделяющимся компонентам вектора невязок. Для этого могут использоваться как сами невязки, так и уточненные значения компонент с учетом априорно установленных устойчивых коэффициентов связи.

В рамках используемого подхода эффективной является процедура корректировки вектора y, реализуемая в виде следующей последовательности шагов.

Шаг 1. Вычисление начальной МНК-оценки:

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Шаг 2. Вычисление вектора невязок:

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

Шаг 3. Определение среднего

$$\bar{\xi} = \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i.$$

Шаг 4. Упорядочивание невязок по возрастанию:

$$\hat{\xi}_1 < \hat{\xi}_2 < \ldots < \hat{\xi}_i < \ldots < \hat{\xi}_{N-1} < \hat{\xi}_N. \qquad (4.6)$$

Определение среднего члена вариационного ряда:

$$\mathrm{med}\left(\hat{\xi}\right) = \hat{\xi}_{\frac{N-1}{2}}. \qquad (4.7)$$

Шаг 5. Определение величины "сдвига" $\Delta_{фр}$:

$$\Delta = \left[\bar{\xi} - \mathrm{med}\left(\hat{\xi}\right)\right] / \left\|\hat{\xi}\right\|_2 . \qquad (4.8)$$

где $\left\|\hat{\xi}\right\|_2 = \left(\sum \hat{\xi}_i^2\right)^{1/2}$, $i = \overline{1, N}$ - евклидова норма вектора невязок.

Шаг 6. Проверка условия

$$\left|\Delta_{фр}\right| > \Delta_{доп}, \qquad (4.9)$$

где $\Delta_{доп} > 0$ допустимая величина относительного сдвига. Если указанное неравенство имеет место, осуществляется корректировка компонент y_i вектора **y** по правилу:

$$y_i = y_i - \kappa_i \cdot \mathrm{Sign}\left(\Delta_{фр}\right)\left|\hat{\xi}_i\right|, \qquad (4.10)$$

где κ_i - (близкий к единице) коэффициент, характеризующий соотношение величин компонент векторов невязок и ошибок. При выборе этого коэффициента можно ориентироваться на данные таблиц 3.2 - 3.4 или применять более "осторожную" стратегию выбора κ_i. Корректировке подвергаются лишь несколько крайних членов вариационного ряда (4.6). Если $\Delta_{фр} < 0$, то корректируются лишь те компоненты вектора **y**, которые соответствуют наибольшим по абсолютной величине отрицательным (левым) невязкам, если же $\Delta_{фр} > 0$, корректируются компоненты, соответствующие наибольшим положительным (правым) невязкам.

После завершения корректировки осуществляется переход к шагу 1. Если неравенство (4.9) не выполняется, это свидетельствует о том, что невязки достаточно симметрично группируются относительно среднего и медианы. В этом случае, в соответствии с описанной выше последовательностью шагов, можно осуществлять одновременную корректировку компонент вектора **y**, соответствующих как крайним левым, так и крайним правым членам вариационного ряда (4.6). После достижения удовлетворительной симметрии невязок относительно "центра" и удаления сильно выделяющихся невязок обоих знаков целесообразно перейти к преобразованию взвешивания данных. Общая схема алгоритма идентификации с преобразованиями взвешивания и непосредственной корректировки данных будет описана в следующей главе.

Эффективность описанной процедуры корректировки подтверждается результатами экспериментальных исследований. Эксперимент проводился на тестовых изображениях, приведенных на рис. 3.3, а и б. На искаженное изобра-

жение (рис. 3.3, б) дополнительно накладывались гауссовские помехи. Оценивание заданных (при моделировании) параметров линейного искажающего КИХ-фильтра осуществлялось по 25 и 100 фрагментам изображения. В эксперименте ставилась задача выяснить как влияет описанная процедура корректировки вектора выхода модели на точность решения задачи идентификации линейной модели искажений.

Для установления закономерностей, связанных с объемом используемых для оценивания данных, эксперимент ставился для двух случаев разбиения указанных изображений: на 100 и 25 фрагментов. Число наблюдений на каждом фрагменте (поскольку опорные области брались без "перекрытия") составляло 25 и 100 наблюдений соответственно. В ходе эксперимента использовалась одна и та же реализация гауссового шума с различными значениями параметра масштаба, который задавался в диапазоне [0,1] с шагом 0.1. Результаты исследований приведены в таблице 4.2.

Таблица 4.2. Результаты исследования процедуры центрирования невязок

$\Delta_{отн}$	Число наблюдений на фрагменте N=25				Число наблюдений на фрагменте N=100			
	$N_{отн}$	$\Delta_{отн}$	$\|\Delta c_{отн}\|(y)$	$\|\Delta c_{отн}\|(\bar{y})$	$N_{отн}$	$\Delta_{отн}$	$\|\Delta c_{отн}\|(y)$	$\|\Delta c_{отн}\|(\bar{y})$
1	2	3	4	5	6	7	8	9
0.1	0.5250	0.150451	0.047487	0.023925	0.1304	0.078688	0.085386	0.010533
0.2	0.5250	0.150463	0.094967	0.048273	0.1304	0.078544	0.170947	0.020988
0.3	0.5250	0.150485	0.142489	0.072355	0.1304	0.078462	0.255805	0.031255
0.4	0.5250	0.150490	0.189989	0.096519	0.1304	0.078447	0.342631	0.041902
0.5	0.5250	0.150494	0.237450	0.120709	0.1304	0.078418	0.428220	0.052159
0.6	0.5375	0.150381	0.278350	0.142755	0.1304	0.078406	0.513804	0.062677
0.7	0.5250	0.150501	0.332518	0.168966	0.1304	0.078398	0.599668	0.073152
0.8	0.5250	0.150503	0.379958	0.193095	0.1304	0.078386	0.684950	0.083483
0.9	0.5375	0.150387	0.417462	0.214126	0.1304	0.078388	0.769912	0.093935
1.0	0.5250	0.150505	0.474943	0.241379	0.13045	0.078379	0.856072	0.104199

Во второй и шестой колонках таблицы указаны относительные числа фрагментов ($N_{отн}$), для которых среднее ошибок оказалось сдвинутым относительно

нуля на величину, превышающую $\Delta_{доп}=0.1$. Третья и седьмая колонки средние значения сдвига, вычисленные по формуле

$$\Delta_{отн}=\Delta_{фр}/N_{фр},$$

где $N_{фр}$ число использовавшихся для оценивания фрагментов, а $\Delta_{фр}$.- относительный сдвиг среднего и медианы (4.7), вычисленный по формуле (4.8).

Четвертая и восьмая колонки - относительные погрешности МНК - оценок, определяемые как отношение нормы вектора ошибок оценивания к норме самого вектора. Пятая и девятая - относительные погрешности оценок, полученные после корректировки вектора выхода модели **y**, обеспечившей устранение относительного сдвига среднего и медианы с точностью до величины $|\Delta_{фр}|<0.1$. В данном случае корректировке подвергались четыре крайних члена вариационного ряда (4.6).

Из таблицы видно, что относительное число фрагментов, для которых среднее ошибок оказалось сдвинутым относительно нуля, и средние значения сдвига практически не зависят от величины нормы вектора ошибок в исходных данных. Зато наблюдается заметная зависимость указанных параметров от объема данных. В частности, при увеличении числа наблюдений, как и следовало ожидать, указанные характеристики уменьшаются. Это объясняется тем, что на все изображение накладывался гауссовский шум с нулевым средним. Эти результаты подтверждают также, что даже если на изображении в целом распределение помех гауссовское с нулевым средним, среднее ошибок на малых фрагментах может заметно отличаться от нуля.

Относительная погрешность идентификации при возрастании нормы вектора ошибок в исходных данных, как и полагается, растет. При этом относительная погрешность оценок, полученных после устранения относительного сдвига среднего и медианы невязок путем корректировки вектора **y**, всегда меньше. Полученные результаты подтверждают обоснованность приведенного выше правила корректировки (4.10).

Наряду с (4.10) исследовались другие правила корректировки. В частности, хорошие результаты достигаются также с использованием следующего правила:

$$y_i = y_i - \left(\frac{\Delta}{k}\right)\frac{|\hat{\xi}_i|}{\max\limits_{j\in K}|\hat{\xi}_j|},$$

где K множество индексов корректируемых компонент одного знака (на одном из концов вариационного ряда), а k их число. Направление сдвига в данном случае также определяется знаком величины Δ.

В заключение отметим, что в описанном алгоритме корректировки используется критерий (4.8), который в статистической теории оценивания известен как коэффициент асимметрии и определяется как отношение разности среднего и моды к среднеквадратическому отклонению. Подчеркнем, что выполнение этого критерия (сразу или после корректировки данных) при малом числе наблюдений на фрагменте не может служить достаточным основанием для заключения о существовании нормального распределения с нулевым средним. Поэтому, даже если среднее невязок близко к нулю, для вычисления оценок параметров должен применяться метод взвешивания данных.

4.2. Оптимизационный метод построения весовой матрицы

В предыдущем разделе было показано, что с использованием корректирующего вектора $\tilde{\xi}$ можно построить весовую матрицу **G**, обеспечивающую повышение точности искомых оценок параметров. Для оценивания вектора $\tilde{\xi}$, в частности, может использоваться связь выделяющихся компонент реализаций векторов ошибок и невязок. Возникает естественный вопрос: нельзя ли использовать эту связь непосредственно для построения преобразования (взвешивания), минуя сложный в вычислительном отношении этап оценивания корректирующего вектора? В настоящем разделе соответствующий метод строится на основе оптимизационного подхода [42,43].

В своей основе метод близок к оптимизационному подходу в задачах идентификации. Однако, в отличие от традиционных методов, экстремум ищется не в пространстве искомых оценок, а в пространстве параметров весовой матрицы преобразования. Другими словами, идея заключается в построении оптимальных (в заданном смысле) преобразований отдельно для каждого шага процесса последовательного МНК-оценивания. По существу, этот подход наиболее близок к методам построения адаптивных систем, в которых критерий качества корректируется на промежуточных этапах с учетом дополнительной информации, появляющейся в процессе функционирования.

В разделе 3.1 было показано, что в классе комплексных всегда существует весовая матрица, при применении которой ошибка оценивания равна нулю. В

разделе 4.2 установлена связь метода взвешивания с непосредственной корректировкой данных. На основе анализа этой связи можно предположить, что преобразование взвешивания, построенное по более близкому в смысле заданной меры к вектору ошибок корректирующему вектору ξ, обеспечит более высокую точность. С использованием установленного свойства близости выделяющихся компонент векторов ошибок и невязок повышение точности можно обеспечить непосредственно по невязкам. Для этого небходимо непосредственно в функции невязок сформулировать учитывающий указанное свойство критерий и с его использованием строить локально оптимальные (на шаге) процедуры преобразования.

Опираясь на эту точку зрения, рассмотрим методы построения весовой матрицы $\mathbf{G}$ с использованием нескольких критериев качества. Начнем с рассмотрения критерия, который хотя и приводит к алгоритму, имеющему ограниченное практическое применение, но полезен с точки зрения связи с традиционными методами оценивания. Будем искать диагональную весовую матрицу $\mathbf{G} = \mathrm{diag}(g_1, g_2, \ldots, g_N)$, доставляющую минимум квадратичной форме:

$$\hat{\xi}^T \mathbf{G}^2 \hat{\xi} \to \min \tag{4.11}$$

при условии

$$\sum_{i=1}^{N} g_i = N. \tag{4.12}$$

Задача (4.11), (4.12) внешне напоминает задачу отыскания оптимальных средневзвешенных квадратических оценок (1.9). Однако там ищутся оценки параметров $\hat{\mathbf{c}}$ при заданной фиксированной ковариационной матрице ошибок измерений $\mathbf{K} = M\{\xi\xi^T\}$, которую ничто не мешает здесь, как и в формуле (1.9), обозначить $\mathbf{G}^2 = \mathbf{K}^{-1}$. В данном случае весовая матрица $\mathbf{G}$ строится после того, как вектор МНК-оценок $\hat{\mathbf{c}}$ и соответствующий ему вектор невязок $\hat{\xi}$ вычислены.

Такая локальная оптимизационная постановка правомерна для одного шага рассматриваемой схемы последовательных МНК-оценок. В данном случае мы действуем аналогично тому, как это делается при расщеплении задачи совместного оценивания параметров $\mathbf{c}$ и ковариационной матрицы $\mathbf{G}^2$. Заметим, что это расщепление на этапе реализации соответствующей вычислительной схемы является вынужденным (в силу принципиально нелинейного характера задачи совместного оценивания указанных параметров). Тем не менее, оно применяется часто, даже если необходимые условия экстремума выписаны в общем виде.

Решение сформулированной задачи дается следующими соотношениями:

$$g_k = \frac{g_0}{\hat{\xi}_k^2}, \quad k = \overline{1,N}, \tag{4.13}$$

где

$$g_0 = N\left(\sum_{i=1}^{N} \hat{\xi}_k^{-2}\right)^{-1}, \quad k = \overline{1,N}.$$

При определении весовых коэффициентов по соотношению (4.13) вместо невязок $\hat{\xi}_k$ можно использовать компоненты корректирующего вектора $\tilde{\xi}_k$, построенные, например, так, как описано в разделе 4.1. Нетрудно заметить, что при использовании соотношений (4.13) для вычисления параметров весовой матрицы могут возникнуть проблемы при $\hat{\xi}_k = 0$ ($\tilde{\xi}_k = 0$). Конечно, их можно преодолеть путем добавления в знаменателе правой части (4.13) некоторого малого положительного числа. Однако этих трудностей можно избежать изменением начальной постановки задачи.

Поскольку устойчивая связь между реализациями векторов ошибок и невязок имеет место лишь для их выделяющихся компонент, представляется целесообразной постановка задачи, в которой наблюдения, соответствующие малым абсолютным значениям невязок, не подвергаются взвешиванию. Для задания такого критерия построим из невязок $\hat{\xi}_i$, $i = \overline{1,N}$ вариационный ряд:

$$\left|\hat{\xi}_1\right| \le \left|\hat{\xi}_2\right| \le \ldots \le \left|\hat{\xi}_{q-1}\right| \le \left|\hat{\xi}_q\right| \le \left|\hat{\xi}_{q+1}\right| \le \ldots \le \left|\hat{\xi}_N\right|, \tag{4.14}$$

где $\left|\hat{\xi}_q\right| \le \xi_{доп} > 0$. Сформулируем следующую задачу: найти весовую матрицу **G**, доставляющую минимум квадратичной форме:

$$g_q^2 \cdot \sum_{i=1}^{q} \hat{\xi}_i^2 + \sum_{i=q+1}^{N} g_i^2 \hat{\xi}_i^2 \to \min_{\mathbf{G}} \tag{4.15}$$

при условии, что

$$q \cdot g_q + \sum_{i=q+1}^{N} g_i = N. \tag{4.16}$$

где **G**, по-прежнему, диагональная матрица.

В новой постановке первые q весовых коэффициентов, соответствующих первым (небольшим по абсолютному значению) членам вариационного ряда (4.14), будут одинаковыми. Это означает, что мы отказываемся от взвешивания наблюдений, для которых модули невязок не превышают некоторого заданного

значения $\xi_{доп}$. Решение задачи (4.15) при условиях (4.14), (4.16) дается соотношениями:

$$g_q = N \cdot \left[q + q^{-1} \left(\sum_{k=q+1}^{N} \hat{\xi}_k^{-2} \right) \left(\sum_{i=1}^{q} \hat{\xi}_i^2 \right) \right]^{-1}, \tag{4.17}$$

$$g_k = \left(\frac{g_q}{q} \sum_{i=1}^{q} \hat{\xi}_i^2 \right) \frac{1}{\hat{\xi}_k^2}, \quad k = \overline{q+1, N}. \tag{4.18}$$

Аналогичная задача может быть поставлена для критерия качества вида:

$$\sum_{i=1}^{N} g_i^2 \left| \hat{\xi}_i \right| \to \min_{G}, \tag{4.19}$$

при условии

$$\sum_{i=1}^{N} g_i = N, \tag{4.20}$$

где, как и ранее, $G = \mathrm{diag}(g_1, g_2, \ldots, g_N)$.

Решение задачи (4.19) при условии (4.20) дается соотношениями

$$g_k = \frac{g_0}{\left| \hat{\xi}_k \right|}, \quad k = \overline{1, N}, \tag{4.21}$$

где

$$g_0 = N \cdot \left(\sum_{i=1}^{N} \left| \hat{\xi}_i \right|^{-1} \right)^{-1}$$

С учетом того, что близкими являются лишь большие по абсолютной величине компоненты векторов ошибок и невязок, как и в предыдущем случае, изменим постановку задачи:

$$g_q^2 \cdot \sum_{i=1}^{q} \left| \hat{\xi}_i \right| + \sum_{i=q+1}^{N} g_i^2 \left| \hat{\xi}_i \right| \to \min_{G}, \tag{4.22}$$

при условиях (4.14) и нормировки:

$$q \cdot g_q + \sum_{i=q+1}^{N} g_i = N. \tag{4.23}$$

Эта задача подобна (4.15). В результате ее решения получаем следующие соотношения для вычисления весовых коэффициентов:

$$g_q = N \cdot \left[q + q^{-1} \left(\sum_{i=1}^{q} \left| \hat{\xi}_i \right| \right) \left(\sum_{j=q+1}^{N} \left| \hat{\xi}_j \right|^{-1} \right) \right]^{-1}, \tag{4.24}$$

$$g_k = \left(\frac{g_q}{q} \sum_{i=1}^{q} \left| \xi_i \right| \right) \frac{1}{\left| \xi_k \right|}, \quad k = \overline{(q+1), N}, \tag{4.25}$$

где, как и прежде, q означает число первых (малых по абсолютной величине) членов вариационного ряда (4.14). Заметим, что по соотношениям (4.24),(4.25)

параметры весовой матрицы могут быть вычислены, даже если среди компонент вектора невязок есть нулевые.

Вычислительные аспекты, связанные с реализацией описанного в настоящем разделе подхода к построению схемы последовательных МНК-оценок, более подробно будут рассмотрены в следующей главе. Результаты исследования эффективности применения соответствующих процедур оценивания в конкретных информационных технологиях будут приведены в главе 6.

Еще раз подчеркнем, что описанный в настоящем разделе подход близок к схеме идентификация по настраиваемой модели. Однако последовательное МНК-оценивание с перестраиваемой весовой матрицей, по сравнению с идентификацией по настраиваемым моделям, имеет более широкие возможности. Здесь критерий качества может изменяться на каждом шаге с учетом требований к повышению точности оценок на очередном шаге и дополнительной информации, содержащейся в искомых оценках параметров и невязках.

В данном случае задача сходимости (и/или скорости сходимости) в асимптотическом плане не ставится и не решается. В схеме последовательного МНК-оценивания в качестве меры точности оценок используется евклидова норма вектора ошибок идентификации на каждом шаге. При этом анализ сходимости, по существу, сводится к построению оценок точности для последовательности шагов. Это позволяет решать задачи анализа качества оценок по малому числу наблюдений, не поддающиеся решению известными статистическими методами.

Тем не менее, представляет интерес установить связь схемы последовательного МНК-оценивания, использующего преобразование данных типа взвешивания, с традиционными методами идентификации, построенными на основе статистической теории оценивания. Это позволит использовать описываемые здесь схемы идентификации при различном уровне априорной информации (включая вероятностные распределения помех) в ситуациях, когда это возможно и необходимо.

4.3. Статистическое обоснование метода взвешивания данных

В рамках алгебраического подхода выше были установлены основные закономерности формирования ошибок идентификации из ошибок измерений, содержащихся в исходных данных. При изучении этих закономерностей (поскольку все равно вероятностная модель ошибок измерений неизвестна) мы отказались от задания априорных распределений ошибок и предполагали, что

множество векторов ошибок исходных данных принадлежит шару (1.42) или сфере (1.54). В рамках этой постановки нам удалось построить оценки достижимой точности идентификации на конкретной реализации при условии, что выполнены требования к информативности данных на фрагментах изображений.

Тем не менее, если в ходе функционирования системы обработки изображений накапливается статистическая информация, ее целесообразно использовать как для повышения эффективности принимаемых решений в конкретных ситуациях, так и для оценки эффективности процедур оценивания по множеству реализаций. Другими словами, методы и алгоритмы оценивания по малому числу наблюдений должны строиться таким образом, чтобы сохранялась возможность "перехода" к статистическим оценкам при увеличении объема данных.

Для установления связи методов идентификации, основанных на преобразованиях данных типа взвешивания, с известными статистическими методами оценивания построим вычислительную схему последовательных МНК-оценок, опираясь на традиционное задание априорной статистической гипотезы в виде распределения вероятностей ошибок. Введем в рассмотрение функцию потерь так, как в методе максимального правдоподобия [51]:

$$F(\xi(\mathbf{c}),s) = -\ln p(\xi(\mathbf{c}),s), \tag{4.26}$$

где $p(\xi(\mathbf{c}),s)$ - параметрическое семейство плотностей распределений вектора ошибок ξ с неизвестным вектором параметров $\mathbf{c}$ (сдвига) и скалярным параметром s - (масштаба). В соответствии со статистической трактовкой задачи ищутся оценки параметров $\mathbf{c},s$, доставляющие минимум средним потерям

$$Q(\xi(\mathbf{c}),s) = M\{F(\xi(\mathbf{c}),s)\}.$$

Если компоненты вектора ошибок ξ независимы и одинаково распределены, то после замены плотности распределения на эмпирическую плотность распределения из средних потерь получаются эмпирические средние потери $\hat{Q}(\cdot)$ [51] и задачу можно переформулировать следующим образом. Найти $\hat{\mathbf{c}},\hat{s}$:

$$\hat{Q}(\xi(\hat{\mathbf{c}}),\hat{s}) = \min_{\mathbf{c},s} \hat{Q}(\xi(\mathbf{c}),s) = \min_{\mathbf{c},s} \frac{1}{N}\sum_{i=1}^{N} F(\xi_i(\mathbf{c}),s). \tag{4.27}$$

Введем обозначение

$$g(\xi,s) = \left(F_{\xi}(\xi(\mathbf{c}),s)\xi^{-1}\right)^{1/2}, \tag{4.28}$$

где F'_ξ производная функции потерь по ξ (здесь ξ скаляр). Предположим, что плотность распределения ошибок $p(\xi(\mathbf{c}),s)$ такая, что имеют место свойства

$$F'_\xi(\xi(c),s)\xi^{-1} > 0\,,\ \lim_{\xi\to\infty} g(\xi,s) = 0,\quad \lim_{\xi\to 0} g(\xi,s) < \infty. \tag{4.29}$$

В разделе 4.4 будут приведены некоторые параметрические семейства распределений, удовлетворяющих условиям (4.29).

Необходимые условия минимума в задаче оценки неизвестных параметров **c** и s - (4.27) можно записать в виде системы M+1 уравнений:

$$\left[\sum_{i=1}^{N} g^2\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right)\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\right]\hat{\mathbf{c}} - \left[\sum_{i=1}^{N} g^2\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right)\boldsymbol{x}_i y_i\right] = 0\,, \tag{4.30}$$

$$\sum_{i=1}^{N} F'_s\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right) = 0\,, \tag{4.31}$$

где $F'_s\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right)$ частная производная по параметру s, $\hat{\mathbf{c}}$ и $\hat{s}$ - искомые оценки параметров, а $\boldsymbol{x}_i$ - строки матрицы **X**. Если теперь в соответствии с (4.30), (4.31) организовать вычислительный процесс типа релаксации с попеременным решением системы M уравнений (4.30) относительно $\hat{\mathbf{c}}$, и уравнения (4.31) относительно $\hat{s}$, мы приходим к описанной выше схеме последовательных ОМНК - оценок с перестраиваемой весовой матрицей.

Действительно, в соответствии с описанной в разделе 1.2 схемой вычисления оценки (1.11) процедура взвешивания данных строится в виде попеременно реализуемых этапов МНК-оценивания (1.17):

$$\hat{\mathbf{c}} = \left[\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right]^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{y}}\,, \tag{4.32}$$

и линейного преобразования (1.15):

$$\tilde{\mathbf{X}} = \mathbf{GX},\quad \tilde{\mathbf{y}} = \mathbf{Gy}\,, \tag{4.33}$$

где $\mathbf{G} = \mathrm{diag}\left(g_1, g_2, \ldots, g_N\right)$. Для получения начальной оценки матрица преобразования в (4.32) задается единичной: **G=E**. Нетрудно заметить, что соотношение (4.33) представляет собой матричную запись решения уравнения (4.30).

Фигурирующие в уравнении (4.30) векторы $\boldsymbol{x}_i$ являются M×1-столбцами (транспонированной) M×N - матрицы $\mathbf{X}^{\mathrm{T}}$, а весовые коэффициенты $g\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right)$ элементами матрицы преобразования **G**. Процедура формирования весовых коэффициентов определяется соотношением (4.28). Однако для того чтобы ею воспользоваться необходимо сначала решить, общем случае нелинейное, уравнение (4.31) относительно $\hat{s}$.

Если предположить, что компоненты вектора ошибок ξ являются гауссовыми, независимыми и одинаково распределенными, то, как и следовало ожидать, приходим к МНК. Если же, оставаясь в рамках нормальной гипотезы, отказаться от предположения, что ошибки одинаково распределены, т.е. имеют разную дисперсию, формально из (4.30), (4.31) получаем вычислительную схему:

$$\hat{\mathbf{c}}(k) = \left[\sum_{i=1}^{N} \hat{s}_i^{-2}(k-1) x_i x_i^T \right]^{-1} \left[\sum_{i=1}^{N} \hat{s}_i^{-2}(k-1) x_i y_i \right], \tag{4.34}$$

$$\hat{s}_i^2(k) = \hat{\xi}_i^2(\hat{\mathbf{c}}(k)), \quad i = \overline{1,N}. \tag{4.35}$$

Вычислительная схема (4.34), (4.35) была обоснована нами ранее с использованием оптимизационного метода определения весовой матрицы на каждом шаге. В случае, когда решается задача оценивания по *одному* малому набору данных, равенство (4.35), оставаясь в рамках статистической теории, нельзя обосновать. Связано это с тем, что переход от средних потерь к эмпирическим средним потерям в ситуации, когда набор данных содержит N ошибок, имеющих разные распределения, неправомерен.

В данном случае необходимо перейти к схеме многих подвыборок [1]. При этом равенство (4.35) принимает вид

$$\hat{s}_i^2(k) = \frac{1}{M} \sum_{m=1}^{M} \hat{\xi}_{i,m}^2(\hat{\mathbf{c}}(k)), \qquad i = \overline{1,N},$$

параметр s^2 приобретает смысл оценки дисперсии и снимаются проблемы, связанные с возможным равенством нулю невязок на отдельных реализациях: $\hat{\xi}_i^2(\hat{\mathbf{c}}_k) = 0$. Заметим, что для многих негауссовых параметрических семейств $p(\xi(\mathbf{c}),s)$ весовые коэффициенты для одинаково распределенных ошибок оказываются различными и имеют простую и понятную физическую трактовку при отсутствии ансамбля реализаций.

Таким образом, если ошибки независимы и одинаково распределены, а плотность распределения вероятностей известна с точностью до параметра масштаба s и удовлетворяет условиям (4.29), получение оценок максимального правдоподобия можно реализовать в виде вычислительной схемы типа взвешивания данных. И наоборот, по крайней мере, для одного класса процедур формирования весовых коэффициентов [27,15]:

$$g(\xi(\hat{\mathbf{c}}),\hat{s}) = g_0 \cdot \left| \hat{\xi}(\hat{\mathbf{c}},\hat{s}) \right|^{\nu-2}, \tag{4.36}$$

где $0 < \nu < 2$, а g_0 - нормирующая константа, всегда можно поставить в соответствие параметрическое семейство функций, удовлетворяющих обычным требованиям неотрицательности и нормировки, т.е. дать вероятностную трактовку методу взвешивания. Подробно эту возможность мы рассмотрим в заключительном разделе настоящей главы

Следует иметь в виду, что такая трактовка будет носить формальный характер до тех пор, пока не установлена, действительно, вероятностная природа конкретной задачи. При использовании этой связи в ситуации, когда задача решается по *одному* малому набору данных, мы не вправе приписывать параметрическому семейству функций $p(\xi(\mathbf{c}),s)$ вероятностные свойства. Поэтому методы анализа точности оценок, по-прежнему, должны быть алгебраическими [2,11]. Тем не менее, рассмотренная в настоящем разделе связь со статистическими методами может оказаться полезной для более полного учета априорной, текущей и дополнительной (в т. ч. статистической) информации, насколько это возможно.

В следующем разделе на основе выявленной связи будут рассмотрены некоторые классы допустимых параметрических семейств функций $p(\xi(\mathbf{c}),s)$ и соответствующие им весовые коэффициенты $g\left(\xi_i(\hat{\mathbf{c}}),\hat{s}\right)$. В частности, будут приведены результаты экспериментов, подтверждающие сформулированные выше выводы. Еще раз подчеркнем, что статистическое обоснование метода взвешивания при малом числе наблюдений носит формальный характер. Это теоретическое обоснование приобретает реальное содержание при появлении ансамбля реализаций (многих подвыборок). Последнее обстоятельство может иметь исключительно большое практическое значение при обработке наборов данных различного, заранее не фиксированного, объема.

4.4. Построение преобразований взвешивания путем формального задания параметрических классов распределений

В предыдущем разделе мы показали связь метода взвешивания данных со статистическим методом, основанным на задании априорной гипотезы в виде плотности распределения ошибок исходных данных. При оценивании по малому числу наблюдений априорное вероятностное распределение обосновать не удается. В ситуации, когда распределение ошибок на заданном фиксированном наборе данных не известно, задают несколько параметрических семейств функ-

ций распределения $p(\xi(\mathbf{c}),s)$ и решают задачу различения гипотез, т.е. подбора наиболее подходящей функции для данной конкретной реализации [7].

Реализация этой, хорошо известной, идеи при малом числе наблюдений связана с принципиальными трудностями, на которые мы уже указывали в предыдущем разделе. В частности, на практике мы всегда вынуждены вместо средних потерь рассматривать эмпирические средние потери, которые сходятся к средним потерям лишь при выполнении усиленного закона больших чисел [51]. При малом числе наблюдений основное условие предельных теорем теории вероятностей не выполняется. В связи с этим возникает вопрос: можно ли вообще делать заключение о характере распределения *ошибок* (даже не вкладывая в него вероятностный смысл) на *одной* реализации малого фиксированного набора данных по соответствующей *одной* реализации *невязок*?

Положительный ответ относительно выделяющихся компонент векторов ошибок и невязок мы уже дали ранее. В частности, из свойства проекций векторов ошибок и невязок (3.19) следует, что в некоторой окрестности нульпространства существует множество (не нулевой меры) векторов ошибок, достаточно близких к соответствующим векторам невязок. Эти выводы качественного характера в разделе 3.2 подтверждены также количественными оценками. В настоящем разделе исследуются аналогичные количественные характеристики, основанные на связи метода взвешивания со статистическими методами.

Исчерпывающей характеристикой вектора ξ является эмпирическая функция распределения его компонент. Она представляет собой дискретное распределение, приписывающее каждому ξ_k, $k=1,N$ вероятность $1/N$. Это ступенчатая функция со скачками, кратными $1/N$ в точках $\xi_1, \xi_2, ..., \xi_N$. При малом числе наблюдений и статистической неустойчивости неуместно говорить о наличии вероятностных распределений [1]. Невязки и ошибки при данных обстоятельствах вообще не являются случайными величинами [7,21]. Поэтому при использовании этой модели мы будем последовательно исключать из рассмотрения предположения, которые связаны с ее вероятностной трактовкой.

В таблице 4.3 приведены некоторые из известных симметричных относительно центра ξ_0 параметрических семейств функций $p(\xi(\mathbf{c}),s)$, удовлетворяющих условиям (4.29). Для упрощения специально отобраны функции, зависящие лишь от одного (при $\xi_0=0$) параметра (масштаба) s. Во втором столбце приведены выражения для самих функций $p(\xi(\mathbf{c}),s)$, в третьем - соответ-

ствующие им функции потерь (4.26), а в четвертом выражения для квадратов весовых коэффициентов (4.28). Параметрические семейства, фигурирующие в этой таблице под номерами 4 и 5, получаются одно из другого заменой s на 2s. Поэтому при проведении экспериментов используется только семейство 4.

Таблица 4.3. Допустимые параметрические семейства функций $p(\xi(c),s)$

	Функция $p(\xi(c),s)$	Функция потерь $F(\xi(c),s) = -\ln p(\xi(c),s)$	Весовые коэффициенты
1	$p(\xi) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{\xi^2}{2s^2}}$	$\ln(\sqrt{2\pi}s) + \left(-\xi^2/2s^2\right)$	$\frac{1}{s^2}$
2	$p(\xi) = \frac{1}{4s}\left(1+\frac{\lvert\xi\rvert}{s}\right)e^{-\frac{\lvert\xi\rvert}{s}}$	$\ln 4s + \lvert\xi\rvert/s - \ln\left(1+\lvert\xi\rvert/s\right)$	$\frac{1}{s\left(s+\lvert\hat{\xi}\rvert\right)}$
3	$p(\xi) = \frac{s}{\pi} * \frac{1}{s^2+\xi^2}$	$\ln\left(s^2+\xi^2\right) + \ln(\pi/s)$	$\frac{2}{s^2+\hat{\xi}^2}$
4	$p(\xi) = \frac{1}{2s\cdot ch^2\left(\frac{\xi}{s}\right)}$	$2\ln Sech(\xi/s) + \ln 2s$	$\frac{1}{s\hat{\xi}} th\left(\hat{\xi}/s\right)$
5	$p(\xi) = \frac{1}{s} * \frac{e^{-\frac{\xi}{s}}}{\left(1+e^{-\frac{\xi}{s}}\right)^2}$	$\frac{\xi}{s} + 2\ln\left(1+\exp\left(-\frac{\xi}{s}\right)\right) + \ln s$	$\frac{1}{2s\hat{\xi}} th\frac{\hat{\xi}}{2s}$

В качестве меры для различения гипотез о принадлежности эмпирической функции распределения $F'_{эмп}(\xi)$ к одному из указанных однопараметрических семейств $F(s,\xi)$ примем среднеквадратичный критерий вида

$$J_д(s) = \left\{\sum_{i=1}^{N-1}\left[F(s,\eta_i) - p_i\right]^2 * (\xi_{i+1} - \xi_i)\right\}^{1/2} \qquad (4.37)$$

Здесь ξ_i отсчеты вариационного ряда, построенного из компонент вектора ошибок ξ, $\eta_i \in \left[\xi_i;\xi_{i+1}\right]$, p_i, $i = \overline{1,N}$ отсчеты эмпирической функции распределения.

Выбор критерия (4.37) оправдывается тем, что "весовые" множители $\xi_{i+1} - \xi_i$ по смыслу близки к множителям n/p_i при вычислении статистики χ^2 (если N - число разрядов, а n объем выборки). В данном случае, ввиду малого

числа наблюдений, использование свойств χ^2-распределения, конечно, необоснованно. Более того, теряет смысл и разбиение выборки на разряды, поэтому мы полагаем N=n. Однако указанная аналогия может быть полезна при переходе к большим N.

Для проверки эффективности критерия (4.37) проводился эксперимент, заключавшийся в генерации наборов данных с распределениями, соответствующими заданным в таблице 4.3 параметрическими семействами нормированных функций $p(\xi(c),s)$ и последующей их аппроксимацией набором тех же параметрических семейств функций. При построении выборки значения ξ_i вычислялись из условия выполнения равенства

$$\int_{-\infty}^{\xi_i} p(x)dx = \frac{i-0.5}{N}, \quad i = \overline{1,N}. \tag{4.38}$$

Если число наблюдений велико настолько, что функция $p(x)$ может трактоваться как плотность распределения, значения ξ_i, удовлетворяющие уравнению (4.38), являются квантилями [7].

При таком способе построения выборки длина отрезка интегрирования [a;b] растёт с ростом N как $\sim\sqrt{\ln N}$ для функции, фигурирующей в таблице под №1, $\sim\ln N$ для функций №2 и 4 и $\sim N$ для функции №3. Поэтому оценка сверху погрешности традиционных квадратурных формул может не выполняться. Тем не менее, отношение (b-a)/N для функций 1,2 и 4 стремится к нулю, поэтому выражение под корнем в правой части (4.37) является интегральной суммой для интеграла

$$J(s) = \int_{-\infty}^{+\infty} \left[F(s,\xi) - F_{эмп}(\xi)\right]^2 d\xi.$$

При этом, вместо кажущегося естественным выбора η_i по формуле центральных прямоугольников: $\eta_i = (\xi_i + \xi_{i+1})/2$, величины $\eta_i \in \left[\xi_i;\xi_{i+1}\right]$ определялись как решения уравнений

$$\int_{-\infty}^{\eta_i} p(x)dx = \frac{i}{N}, \qquad i = \overline{1,N-1}.$$

Это позволяет исключить методическую ошибку аппроксимации, в частности, обеспечивает равенство нулю критерия (4.37) в случае, когда генерируемое и аппроксимирующее семейства совпадают.

Аппроксимация построенных для каждого набора данных эмпирических функций распределения осуществлялась по принципу: "каждая каждой" Для оценки параметра масштаба s методом дихотомии решалось уравнение $dJ_д(s)/ds = 0$. Результаты исследования приведены в таблице 4.4. Целые числа 1-4 в обозначениях строк и столбцов соответствуют номерам исходных распределений в таблице 4.3. Приведенные результаты получены для N=25. При дальнейшем увеличении N они практически не изменяются.

Таблица 4.4. Результаты аппроксимации эмпирического распределения различными функциями

Тип (№) функции	1		2		3		4	
	s	$J_д(s)$	s	$J_д(s)$	s	$J_д(s)$	s	$J_д(s)$
1 (s=1.000)	1.000	0.00011	0.554	0.02213	0.543	0.07834	1.185	0.01328
2 (s=0.546)	1.071	0.03682	0.546	0.00020	0.510	0.06869	1.172	0.00982
3 (s=0.480)	1.665	0.21448	0.578	0.14051	0.480	0.00005	1.215	0.14803
4 (s=1.176)	1.043	0.02247	0.548	0.00967	0.521	0.07447	1.176	0.00005

Значения s и $J_д(s)$ для распределения 4 оказываются близкими к аналогичным величинам для распределения 2. Поэтому далее мы будем рассматривать лишь хорошо различимые классы 1, 2 и 3. Этот выбор, отчасти, связан с простотой вычисления соответствующих этим классам весовых коэффициентов (таблица 4.3). Кроме того, веса классов 2 и 3 допускают простую физическую интерпретацию: чем тяжелее хвосты функции $p(\xi(c),s)$, тем "жестче" подавляются ошибки. При этом естественным образом обосновывается разумная стратегия формирования весовых коэффициентов, заключающаяся в придании меньшего веса наблюдениям, для которых невязки оказались слишком большими (веса $\sim\left|\hat{\xi}\right|^{-1}$ для функции 2 и $\sim\left|\hat{\xi}\right|^{-2}$ - для функции 3).

При формальном задании параметрических семейств 2 и 3 не возникает принципиальных трудностей, связанных с отсутствием ансамбля наблюдений, которые имели место в (4.35) при использовании нормального распределения. Кроме того, параметрическое семейство 2, часто, действительно оказывается хорошей аппроксимацией ошибок в аналого-цифровых системах. При этом вполне уместна трактовка задачи, как задачи совместного оценивания M-вектора параметров сдвига **c** и скалярного параметра масштаба s на каждом фиксированном наборе данных.

Таким образом, мы подобрали подходящие параметрические семейства функций $p(\xi(\mathbf{c}),s)$ и выбрали меру различения гипотез. В частности, убедились, что критерий (4.37) "работает" при малом числе наблюдений. Теперь приведем результаты исследования возможности оценки распределения ошибок по эмпирическому распределению невязок на малом наборе данных.

Из опирающихся на соотношение (3.19) соображений качественного характера ясно, что близость ошибок и невязок чаще будет иметь место в ситуациях, когда норма вектора невязок (сопряженность с нуль-пространством) выше. Эти соображения подтверждаются экспериментом, который проводился по следующей схеме.

Осуществлялась идентификация коэффициентов импульсной характеристики искажающего КИХ-фильтра по входному (неискаженному) и выходному (искажённому и зашумлённому) изображениям, показанным на рисунках 3.3, а, б. Также как и ранее, размеры опорной области, соответствующие характеристике размазывающего фильтра, задавались 5×5, а размеры фрагментов, на которые разбивалось изображение, 25×25 соответственно. Опорные области на фрагменте размещались "без перекрытия", поэтому размерности векторов ошибок и невязок были 25×1. Реализации шума указанной длины для каждого заданного параметрического семейства $p(\xi(\mathbf{c}),s)$ генерировались как описано выше.

После оценки на каждом фрагменте импульсной характеристики вычислялись невязки (1.7). Далее для каждой реализации вектора невязок строилась эмпирическая функция распределения, которая затем аппроксимировалась каждым из указанных выше трех (1, 2 и 3) параметрических семейств распределений и подсчитывалось значение критерия (4.37). Всего на изображении имелось 100 фрагментов, среди которых 80 информативных [47,62,63]. При моделировании векторов ошибок искусственно вводилась рандомизация, а результаты усреднялись по 100 экспериментам. В таблице 4.5 приведены данные о среднем количестве случаев $K_{ср}$, в которых каждое распределение оказывалось наилучшей, в смысле минимума критерия (4.37), аппроксимацией.

Приведенные результаты хорошо согласуются с соображениями качественного характера, основанными на соотношении (3.19). В частности, несмотря на то, что при малом числе наблюдений отсутствует свойство статистической устойчивости распределений ошибок на фрагментах изображений, существуют устойчивые связи между характеристиками распределений векторов ошибок и

невязок. Эти связи позволяют высказывать достаточно убедительные предположения относительно класса, которому принадлежит вектор ошибок на каждом конкретном наборе данных. В свою очередь, они могут быть достаточным основанием для выбора, по крайней мере, типа преобразования данных.

Таблица 4.5. Характеристика связи распределений ошибок и невязок

Семейство функций (№)	Априорная гипотеза	Количество решений $K_{ср}$ по изображению	Средняя норма вектора невязок
1	1	43.4	4.439
	2	34.7	4.405
	3	1.9	3.578
2	1	22.6	4.614
	2	53.1	4.753
	3	4.3	4.555
3	1	0.3	2.236
	2	20.9	9.843
	3	58.8	10.747

На данном этапе цель исследования была ограничена лишь обоснованием меры для различения гипотез. Решая здесь эту задачу мы, сознательно, использовали критерий наиболее близкий к известному в теоретической статистике критерию χ^2. Это, наряду с соотношениями (4.28), (4.30), (4.31), создает необходимые предпосылки для статистической интерпретации метода взвешивания при поступлении большого объема данных.

При построении практических процедур оценивания, из соображений уменьшения вычислительной сложности и повышения надежности, по-видимому, более плодотворным будет применение более простых критериев, основанных на непосредственном вычислении моментных и других характеристик эмпирического распределения компонент вектора невязок. Ниже приводятся результаты исследований, подтверждающие эту точку зрения.

4.5. Различение классов по значениям среднего модулей невязок

В настоящем разделе предлагается и исследуется критерий различения (обоснованных выше) гипотез 1, 2 и 3, основанный на непосредственном вычислении характеристик эмпирического распределения. Для формулировки

некоторых общих требований к критерию и обоснования методики его сравнения с рассматривавшимся выше критерием (4.37) вначале приведем результаты следующего эксперимента.

Осуществлялись моделирование и идентификация модели искажений на фрагментах изображений, показанных на рисунках 3.3, а, б также, как это было описано в разделе 4.4. Особенность эксперимента заключалась в следующем. Параметрические классы распределений, задаваемые на этапе моделирования, считались известными, а оценивание осуществлялось по трехшаговой схеме с использованием весовых коэффициентов, соответствующих классам №1, 2 и 3 из таблицы 4.3. Погрешность фиксировалась после первого - ε_1 и третьего - ε_3 шагов. Средние (по 100 реализациям) погрешности оценок вектора коэффициентов линейного фильтра на первом и третьем шагах процедуры оценивания в процентах приведены в таблице 4.6.

Таблица 4.6. Результаты исследования точности трехшаговой схемы

Исходное распределение № 1											
s	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
ε_1	0.2374	2.458	4.646	7.736	9.131	12.88	14.66	19.75	22.52	21.54	21.57
ε_3	0.2374	2.458	4.646	7.736	9.131	12.88	14.66	19.75	22.52	21.54	21.57
Исходное распределение № 2											
s	0.01	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
ε_1	0.4982	2.509	4.443	8.065	9.717	14.39	13.64	15.48	18.58	27.58	26.22
ε_3	0.5111	2.568	4.194	7.707	9.645	14.58	13.30	15.05	18.34	25.78	26.05
Исходное распределение № 3											
s	0.01	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
ε_1	1.395	5.480	10.94	21.62	22.62	29.21	38.57	45.82	50.85	56.12	68.07
ε_3	0.723	2.987	6.73	11.04	12.49	15.48	21.74	23.83	27.68	27.44	28.85

Из таблицы видно, что ошибки в определении типа распределения неравнозначны. Ошибки типа 1→2 (стрелка означает, что при моделировании использовался класс 1, а при оценивании класс 2) и 2→1 представляются несущественными, так как при исходных распределениях 1 и 2 погрешности оценивания после первого и третьего шага оказываются близкими. Наиболее значительной является ошибка 3→1.

С учетом полученных результатов сформулируем общие требования к критерию. Во-первых, представляется разумным потребовать, чтобы критерий позволял достаточно уверенно (не менее чем в половине случаев) обнаруживать "свое" параметрическое семейство распределений, реализовавшееся на данном

конкретном фрагменте (наборе данных). Во-вторых, необходимо установить ограничение на число грубых ошибок. В частности, потребуем, чтобы относительное число случаев, когда принимается гипотеза 3, в то время как имело место распределение 1 и наоборот гипотеза 1 при распределении 3, составляло не более 1%. При выполнении этих условий критерии будем сравнивать по суммарному числу правильно принятых решений на выборках, полученных моделированием трёх указанных выше распределений (1,2 и 3) одинаковое число раз.

В качестве критерия для принятия решения о принадлежности реализаций эмпирической функции распределения к одному из классов (№1, 2 или 3) будем использовать нормированную величину:

$$S_{нфм} = S_{эмп}/S_{равн}, \qquad (4.39)$$

где $S_{эмп}$ равна среднему значению модулей величин ξ_i, i=1,N:

$$S_{эмп} = \frac{1}{N}\sum_{i=1}^{N}|\xi_i|, \qquad (4.40)$$

$$S_{равн} = \frac{A^2 + B^2}{2(B - A)}, \qquad (4.41)$$

а [A,B] - отрезок, на котором определены невязки и/или ошибки (при моделировании). Предполагается, что распределение симметризовано так, что величина A всегда отрицательна. Применение критерия (4.39) позволяет избежать построения вариационного ряда и эмпирической функции распределения.

Геометрически критерий (4.39) является отношением площади между эмпирической функцией распределения $F(\xi)$ и осями, соответствующими ординатам 0,1 (на рис. 4.2 эти площади "залиты" серым цветом) к площади равномерного распределения с таким же диапазоном изменения ξ. Легко проверить, что аналитическое выражение (4.41) для величины $S_{равн}$ определяет площадь между функцией равномерного распределения и теми же осями на отрезке [A,B]. Обе указанные площади имеет ту же размерность, что и сама случайная величина. Поэтому после вычисления отношения (4.39) сравниваемые величины $S_{нфм}$ будут безразмерными.

При определении начальных приближений пороговых значений для различения гипотез использовались непрерывные представления соответствующих параметрических семейств распределений, приведенные в таблице 4.3. В частности, для распределений №1, 2 и 3 из указанной таблицы были получены

следующие аналитические выражения для площадей на отрезке интегрирования [A;B]:

$$S_{эмп\,1} = \frac{s}{\sqrt{2\pi}}\left(2 - \exp\left(-\frac{A^2}{2s^2}\right) - \exp\left(-\frac{B^2}{2s^2}\right)\right), \qquad (4.42)$$

$$S_{эмп2} = \frac{3s}{2} - \exp\left(\frac{A}{s}\right)*\left[\frac{3}{8}(s - A) + \frac{A^2}{4s}\right] - \exp\left(-\frac{B}{s}\right)*\left[\frac{3}{8}(B + s) + \frac{B^2}{4s}\right], \qquad (4.43)$$

$$S_{эмп3} = \frac{s}{2\pi}\left(\ln\left(1 + \frac{A^2}{s^2}\right) + \ln\left(1 + \frac{B^2}{s^2}\right)\right) \qquad (4.44)$$

соответственно. Здесь s - фигурирующий в этих функциях параметр масштаба.

Из этих выражений видно, что пороговые значения зависят не только от выбранного параметрического класса распределений, но и от диапазона (интервала [A;B]) ошибок. В свою очередь, этот интервал, для каждого заданного параметрического семейства, зависит от объема данных.

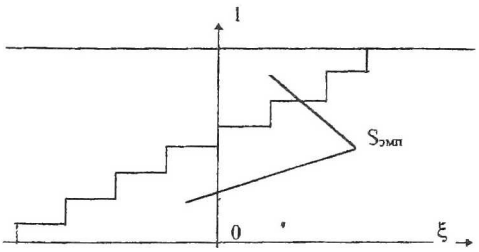


Рис.4.1. Геометрическая иллюстрация критерия (4.38)

Начальные приближения для пороговых значений критерия (4.39) определялись с использованием значений, вычисленных по формулам (4.40)-(4.44). При этом граница A интервала [A,B] являлась решением уравнения (4.38) относительно ξ_i при i=1, а граница B - при i=N. Для упрощения указанные границы интервала можно также вычислять с использованием следующей эмпирической формулы

$$B=-A=k(N)\cdot s,$$

где k(N) фиксированы для каждого класса и объёма выборки N и подобраны так, что указанные границы A и B являются решения уравнения (4.38).

Для уточнения рассчитанных таким образом пороговых значений проводился эксперимент. В ходе эксперимента на каждом (из 80-ти информативных) фрагментов изображений, показанных на рис. 3.3, а и б, осуществлялись следующие этапы.

1. Моделирование реализации вектора ошибок для одного из заданных типов распределений (№1, 2 или 3). Параметры масштаба для них задавались на основе результатов исследований, приведенных в таблице 4.4. В частности,

для распределения №1, $s = \sigma$, для №2, $s = 0.546 \cdot \sigma$ и для №3, $s = 0.48 \cdot \sigma$, где σ - выборочное СКО.

2. Решение задачи оценивания и вычисление соответствующей реализации вектора невязок.
3. Определение текущего значения критерия $S_{норм} = S_{эмп}/S_{равн}$ по невязкам и выбор типа распределения в соответствии с заданным пороговым значением.

Процедура выбора типа распределения на 3-м этапе реализовывалась в виде следующей последовательности шагов.

Шаг 1. Вычисление выборочного с.к.о. $\hat{\sigma}$ и среднего $\bar{\xi}$.

Шаг 2. Центрирование выборочных значений.

Шаг 3. Вычисление:

- среднего значения модулей невязок - (4.40);
- границ интервала $A = -\min \overset{\circ}{\xi}$, $B = \max \overset{\circ}{\xi}$;
- площади $S_{равн}$ по формуле (4.39);
- отношения $S_{норм} = S_{эмп}/S_{равн}$, определяемого формулой (4.39).

Шаг 4. Сравнение $S_{норм}$ с заданным пороговым значением.

В ходе экспериментов уточнялись заданные в качестве начального приближения по формулам (4.39) - (4.44) и зависящие от k(N) пороговые значения. Для этого они варьировались "вручную" таким образом, чтобы максимизировать суммарное число правильно принятых решений при моделировании трёх указанных выше распределений (№1, №2 и №3) одинаковое число раз. В таблице 4.7 приведены пороговые значения, вычисленные на разных стадиях эксперимента.

Таблица 4.7. Обоснование пороговых значений критерия (4.38)

Классы	Пороги, рассчитанные по формулам:		Пороги, подобранные "вручную"
	(4.39)-(4.43)	(4.39), (4.40), (4.41)	
1	0.6828	0.767052	0.896
2	0.6071	0.670155	0.645
3	0.2217	0.323315	0.355

Второй столбец - начальные приближения, полученные по формулам (4.39)-(4.43); третий столбец - точные значения, рассчитанные в ходе эксперимента по формуле (4.39) с использованием только соотношений (4.40) и (4.41); и, наконец, последний столбец - пороговые значения, подобранные "вручную" из условия

рному числу правильных решений при N=25.

Для сравнения критерий (4.39) с рассмотренным ранее критерием (4.37) осуществлялась проверка качества различения классов распределений на тех же тестовых изображениях, приведенных на рис. 3.3, а и б. Для этого на искаженное изображение (рис. 3.3 б) накладывался шум с распределением заданного класса (№1, №2 или №3) и подсчитывалось среднее (по 100 экспериментам) количество правильных и ошибочных решений для каждого распределения на всех фрагментах изображения. Результаты приведены в таблице 4.8. Так как на изображении только 80 фрагментов из 100 являются информативными, сумма чисел в строке равна 80.

Таблица 4.8. Результаты сравнения критериев (4.37) и (4.39)

Фактическое распределение	Количество решений о гипотезе № :			Сумма правильных решений
	1	2	3	
Квадратичный критерий				
1	**43.43**	34.71	1.86	155.33
2	22.60	**53.07**	4.33	
3	0.30	20.87	**58.83**	
Критерий $S_{норм}$				
1	**49.16**	30.82	0.02	170.53
2	23.47	**56.46**	0.07	
3	0.19	14.90	**64.91**	

Осуществлялось также сравнение критериев (4.37) и (4.39) по скорости вычислений. Для этого определялось время работы программы при решении задачи идентификации параметров модели искажающей системы с оценкой класса распределения и преобразованием данных типа взвешивания. Результаты эксперимента, усредненные по данным 50-кратного запуска программы, приведены в таблице 4.9.

В ходе экспериментов для исключения времени работы самого алгоритма идентификации фиксировалось также время работы программы, которая осуществляла вычисления при тех же исходных данных и отличалась лишь тем, что в ней была блокирована процедура оценки класса распределения, т.е. реализовывалось вычисление оценок без распознавания классов. Поэтому в последней строке последней колонки таблицы 4.9 фигурирует цифра 0.

Таблица 4.9. Результаты сравнения критер...

Тип временных затрат	Критерий (4.37)	Критерий (4.39)	Без распознавания классов
Полное время работы программы (с)	727.33	136.22	134.46
В т. ч. на вычисления критерия (с)	592.87	1.76	0

Из таблиц 4.8 и 4.9 видно, что квадратичный критерий (4.37) уступает критерию (4.39) как по эффективности распознавания классов, так и в вычислительном отношении. Поэтому критерий (4.39) (хотя теоретически он и менее обоснован) более предпочтителен для построения практических процедур оценивания с преобразованием данных.

В заключение подчеркнем, что необходимость разработки эмпирических процедур принятия решений о характере распределений невязок и ошибок на конкретных реализациях данных связана с тем, что традиционные статистические методы проверки гипотез в данном случае теоретически необоснованы. Мы стремились, чтобы эти процедуры были "похожими" на аналогичные статистические. Тем не менее, они принципиально отличаются. Здесь эксплуатируется лишь свойство статистической устойчивости связи реализаций векторов ошибок и невязок, которое может проявляться при малом числе наблюдений, в то время как статистическая устойчивость самих распределений ошибок еще не имеет места.

4.6. Связь с L_v - оценками

Выше мы построили методы преобразования данных, основанные на связи метода взвешивания с методом максимального правдоподобия. При выборе классов преобразований мы вынуждены были учитывать соотношения (4.29), которые ограничивают класс допустимых параметрических семейств распределений. Примеры таких распределений приведены в таблице 4.3. В настоящем разделе мы рассмотрим модификации критерия качества идентификации, позволяющие обойти указанные ограничения. В качестве исходного будем использовать класс распределений, порождающих L_v-оценки.

Функция плотности из этого класса [15]

$$p(\xi, v) = Ae^{-\sigma|\xi|^v}, \quad 0 < v < 2, \qquad (4.45)$$

и [illegible] плотности нормального распределения $p(\xi,2) = Ae^{-\alpha\xi^2}$, а при $\nu=1$ распределение (4.45) является распределением Лапласа. Очевидно, что как это, так и другие (при $\nu\neq 2$) распределения из семейства (4.45) не удовлетворяет условиям (4.29). Рассмотрим связь соответствующих классу (4.45) L_ν-оценок регрессий [15] с методом взвешивания и способы преодоления ограничений, связанных с условиями (4.29).

Если ошибки ξ_i независимы и одинаково распределены с функцией плотности (4.45), можно построить оценку максимального правдоподобия (ММП-оценку), минимизирующую сумму

$$Q_\nu(\mathbf{c}) = \sum_{i=1}^{N} F\left(\xi_i(\hat{\mathbf{c}})\right) = \sum_{i=1}^{N} \left|\hat{\xi}_i\right|^\nu , \qquad (4.46)$$

где $\hat{\xi}_i$, $i=1,N$ - компоненты вектора невязок (здесь, как и выше, имеются в виду эмпирические средние потери [51]). Оценки, минимизирующие (4.46) и являющиеся частным случаем ММП-оценок Хюбера [50], называют L_ν-оценками. Для функции (4.46) при $\nu\geq 1$ существует единственный локальный минимум, который совпадает с глобальным. Для нахождения этого минимума предложен "итеративный МНК" [15,27]. Технику построения этого метода мы уже использовали при анализе связи ММП-оценок с методом взвешивания данных в схеме последовательного МНК-оценивания.

Для этого сумму (4.26) переписывают в виде

$$Q_\nu(\mathbf{c}) = \sum_{i=1}^{N} \left|\hat{\xi}_i\right|^\nu = \sum_{i=1}^{N} \left|\hat{\xi}_i\right|^{\nu-2} \hat{\xi}_i^2 = \sum_{i=1}^{N} g_i\left(\hat{\xi}_i\right)\hat{\xi}_i^2 . \qquad (4.47)$$

Из (4.47) видно, что вычисление удовлетворяющей критерию (4.46) оценки может быть реализовано по схеме взвешенного (итеративного) МНК в виде попеременного вычисления оценок

$$\hat{\mathbf{c}} = \left[\sum_{i=1}^{N} g^2\left(\hat{\xi}_i\right)\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}\right]^{-1}\left[\sum_{i=1}^{N} g^2\left(\hat{\xi}_i\right)\boldsymbol{x}_i y_i\right] \qquad (4.48)$$

и весовых коэффициентов

$$g_i\left(\hat{\xi}_i\right) = \left|\xi_i(\hat{\mathbf{c}})\right|^{\nu-2} \qquad (4.49)$$

Поскольку для семейства функций (4.45) не выполняются условия (4.29), при определении весовых коэффициентов (4.49) для $\nu<2$ возникают проблемы, связанные с делением на малое число.

Рассмотрим модификацию критерия (4.46), позволяющую избежать указанные трудности. Построим вариационный ряд

$$\left|\hat{\xi}_1\right| < \ldots < \left|\hat{\xi}_{q-1}\right| < \left|\hat{\xi}_q\right| < \left|\hat{\xi}_{q+1}\right| < \ldots < \left|\hat{\xi}_N\right|. \qquad (4.50)$$

Выделив q первых (малых) членов вариационного ряда (4.50) в отдельную сумму, модифицируем критерий (4.46) следующим образом:

$$Q_v(\mathbf{c}) = \sum_{i=1}^{q} \hat{\xi}_i^{\,2} + \sum_{i=q+1}^{N} \left|\hat{\xi}_i\right|^v \qquad (4.51)$$

Для критерия (4.51) по аналогии с (4.47) будет иметь место соотношение

$$Q_v(\mathbf{c}) = \sum_{i=1}^{q} \hat{\xi}_i^{\,2} + \sum_{i=q+1}^{N} \left|\hat{\xi}_i\right|^{v-2} \hat{\xi}_i^{\,2} = \sum_{i=1}^{q} \hat{\xi}_i^{\,2} + \sum_{i=q+1}^{N} g_i\left(\hat{\xi}_i\right)\hat{\xi}_i^{\,2} \qquad (4.52)$$

Если теперь для (4.52) построить вычислительную схему типа (4.48), (4.49), весовые коэффициенты для q первых членов вариационного ряда окажутся равными единице, т.е. вычисляемые оценки являются МНК-оценками "в центре" и L_v-оценками на "хвостах" вариационного ряда.

Если для соответствующего критерию (4.52) параметрического семейства функций $p(\xi, v)$ при $v=1$ ввести нормировку, то такая модель будет описывать смесь нормального "в центре" и двухстороннего распределения Лапласа на хвостах. Соответствующие этой модели оценки максимального правдоподобия оказываются абсолютно оптимальными на указанном классе распределений. В работе [51] для этого класса используется обозначение P_3.

При малом числе наблюдений такое теоретическое обоснование модификации неприемлемо, т.к. указанные распределения на малых фрагментах просто не существуют. Однако мы можем обосновать степени слагаемых в (4.51) исходя из установленных устойчивых связей компонент векторов ошибок и невязок. Последовательно развивая эту идею в направлении более полного учета различий степени устойчивости связей различных компонент векторов ошибок и невязок, приходим к семейству критериев качества вида

$$Q(\mathbf{c}) = \sum_{i=1}^{N} \left|\hat{\xi}_i\right|^{s_i} \qquad (4.53)$$

Здесь параметры s_i, $i = \overline{1,N}$ должны задаваться для каждого наблюдения (строки матрицы **X**) с использованием априорной, текущей и появляющейся в ходе решения задачи дополнительной информации о невязках. Нетрудно заметить, что при $s_i = \text{Const} = 2$, $i = \overline{1,N}$ критерий (4.53) совпадает с методом наименьших квадратов.

Необходимые условия минимума для задачи отыскания $\hat{\mathbf{c}}$:

$$Q(\hat{c}) = \min_{c} Q(c)$$

для критерия (4.53) имеют вид

$$\nabla_{\hat{c}} Q(\hat{c}) = \sum_{i=1}^{N} s_i \left|\hat{\xi}_i(\hat{c})\right|^{s_i - 1} \mathrm{Sign}\left[\hat{\xi}_i(\hat{c})\right] \nabla_{\hat{c}} \hat{\xi}_i(\hat{c}) = 0. \qquad (4.54)$$

Применив ту же, что и в (4.47) (4.49), технику, систему M уравнений (4.54) можно представить в виде

$$\nabla_{\hat{c}} Q(\hat{c}) = \sum_{i=1}^{N} s_i \left|\hat{\xi}_i(\hat{c})\right|^{s_i - 2} x_i \hat{\xi}_i(\hat{c}) = 0, \qquad (4.55)$$

где x_i, i=1,N - вектор-строка матрицы X. Подставляя в (4.55)

$$\hat{\xi}_i(\hat{c}) = y_i - x_i^T \hat{c},$$

приходим к схеме последовательных МНК-оценок типа (4.48), (4.49):

$$\hat{c} = \left[\sum_{i=1}^{N} g\left(\hat{\xi}_i\right) x_i x_i^T\right]^{-1} \left[\sum_{i=1}^{N} g\left(\hat{\xi}_i\right) x_i y_i\right], \qquad (4.56)$$

$$g\left(\hat{\xi}_i\right) = s_i \left|\hat{\xi}_i(\hat{c})\right|^{s_i - 2}. \qquad (4.57)$$

Из соотношений (4.56), (4.57) видно, что фигурирующие в (4.53) параметры s_i, $i = \overline{1,N}$, в конечном итоге, определяют "коэффициенты доверия" различным наблюдениям. Вычислительная схема (4.56), (4.57), по существу, является одной из форм записи метода взвешивания в задачах наименьших квадратов [23]. Однако в данном случае процедура оказывается более гибкой, чем в методе средневзвешенной квадратической ошибки и/или при вычислении L_v оценок, т.к. параметры s_i, i=1,N могут быть различными для всех строк матрицы X. Конечно, показатели степени s_i должны назначаться так, чтобы избежать деления на нуль или очень малое число. Они должны также учитывать всю имеющуюся априорную и текущую измерительную информацию.

Обычно априорная информация, необходимая для их задания, бедная. В этом случае показатели степени s_i могут быть заданы в виде (положительных) удовлетворяющих условию нормировки:

$$\sum_{i=1}^{N} s_i = N, \qquad s_i > 0, \; i = \overline{1,N} \qquad (4.58)$$

чисел, характеризующих относительную степень доверия различным наблюдениям. Для учета характеристики разброса невязок введем также параметр масштаба s. С учетом сказанного критерию (4.53) можно придать следующий конкретный вид:

$$Q(\mathbf{c}, s) = \frac{1}{N}\sum_{i=1}^{N}\left(\ln\frac{1}{s} - \frac{1}{s_i}\left(\frac{|\hat{\xi}_i|}{s}\right)^{s_i}\right). \qquad (4.59)$$

Критерий (4.59) может быть также формально построен в предположении независимости ошибок по схеме ММП-оценок для параметрического класса функций

$$p(\xi(\mathbf{c}), s) = \frac{A}{s}\exp\left(-\frac{1}{s_i}\left(\frac{|\xi_i|}{s}\right)^{s_i}\right), \qquad (4.60)$$

где A - заданная соответствующим образом нормирующая константа.

Оценки, доставляющие минимум критерию (4.59), могут быть вычислены по соотношениям типа (4.56) и (4.57), в которые теперь войдет также и неизвестный параметр s:

$$\hat{\mathbf{c}} = \left[\sum_{i=1}^{N} g\left(\hat{\xi}_i, \hat{s}\right)\boldsymbol{x}_i\boldsymbol{x}_i^{T}\right]^{-1}\left[\sum_{i=1}^{N} g\left(\hat{\xi}_i, \hat{s}\right)\boldsymbol{x}_i y_i\right], \qquad (4.61)$$

$$g\left(\hat{\xi}_i, \hat{s}\right) = \left|\hat{\xi}_i(\hat{\mathbf{c}}), \hat{s}\right|^{s_i-2}. \qquad (4.62)$$

Вычислению весовых коэффициентов (4.62) должно предшествовать решение (относительно параметра $\hat{s}$) уравнения типа (4.31):

$$\sum_{i=1}^{N} F_s'\left(\xi_i(\hat{\mathbf{c}}), \hat{s}\right) = 0,$$

которое с учетом (4.60) можно представить в следующем виде

$$\hat{s} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\xi_i|^{s_i}}{\hat{s}^{s_i-1}}. \qquad (4.63)$$

Запись нелинейного уравнения типа (4.31) для рассматриваемого конкретного случая в форме (4.63) может непосредственно использоваться для его решения, например, методом простой итерации.

Конечно, с точки зрения практической реализации, более простой и понятной является схема (4.56), (4.57), основанная на критерии (4.53). Рассмотренный частный случай этого критерия в виде (4.59) интересен тем, что перебрасывает "мостик" от оптимизационного метода построения весовой матрицы к вероятностным моделям. Для этого достаточно множеству весовых коэффициентов, полученных в рамках оптимизационного подхода, поставить в соответствие множество параметров s_i, i=1,N. Если окажется, что эти параметры проявляют свойство статистической устойчивости на множестве реализаций, то нет оснований не признавать существование параметрического семейства плотностей (4.60).

Рассмотренная модификация алгоритма является весьма гибкой. В данном случае задаваемые фиксированными параметры s_i могут учитывать априорную и текущую измерительную информацию в любой физической шкале. Неизвестный же параметр $\hat{s}$ оценивается в каждом эпизоде идентификации и учитывает реальный разброс невязок на каждом конкретном наборе данных..

Из (4.63) видно, что если $s_i = \text{Const} = 1, \quad i = \overline{1,N}$ параметр $\hat{s}$ представляет собой среднее модулей, а критерий (4.53) совпадает с методом наименьших модулей. При $s_i = \text{Const} = 2, \quad i = \overline{1,N}$ параметр $\hat{s}$ в (4.63) рассчитывается как среднее квадратов невязок. При этом весовые коэффициенты (4.62) не зависят от величины невязки и равны единице, а соответствующая МНК-оценка получается за один шаг. При $s_i = \text{Const} = 0, \quad i = \overline{1,N}$ критерий (4.53) не зависит от невязок и его непосредственная интерпретация затруднительна. Соответствующая этому случаю весовая матрица, без особых затруднений, была построена в разделе 4.2 на основе оптимизационного подхода. Однако практическая реализация соответствующего преобразования взвешивания невозможна при наличии среди компонент вектора невязок близких к (или равных) нулю.

В заключение еще раз подчеркнем, что при малом числе наблюдений параметрическое семейство плотностей (4.45) может и не существовать. Поэтому при формировании весовых коэффициентов, к сожалению, обычно приходится ориентироваться не на вероятностные модели шумов, а на текущую измерительную и дополнительную информацию и анализ конкретных реализаций невязок. Тем не менее, описанная связь с L_v - оценками показывает, каким образом конкретной вычислительной процедуре преобразования данных может быть поставлена в соответствие вероятностная модель в ситуации, когда статистические характеристики наблюдений проявляют свойство устойчивости.

5. РЕШАЮЩИЕ ПРАВИЛА И АЛГОРИТМЫ

В настоящей главе описываются решающие правила и алгоритмы, реализующие вычисление оценок (1.11). Основное внимание уделяется процедурам анализа информативности и преобразования данных. Приводится общая схема алгоритма идентификации. Обсуждаются вычислительные аспекты, связанные с различными способами организации преобразования данных типа взвешивания.

5.1. Общая схема алгоритма идентификации

Специфические особенности задачи оценивания по малому числу наблюдений не раз обсуждались в предыдущих главах. Вытекающие из них требования к алгоритмам идентификации позволяют сформулировать следующие общие принципы конструирования этих алгоритмов:

- предварительный анализ информативности каждого малого протокола данных, формируемого на отдельном фрагменте изображения, учитывающий как обусловленность задачи, так и ошибки исходных данных;
- отказ от использования априорных вероятностных распределений ошибок;
- построение обеспечивающих повышение точности оценивания преобразований данных на основе дополнительной информации, содержащейся в невязках выхода модели и объекта;
- использование в качестве меры точности оценок нормы вектора ошибок идентификации на каждой реализации.

Указанные принципы, по существу, определяют набор функционально-необходимых процедур обработки и логику их взаимодействия в общей схеме вычисления оценок (1.11) по фиксированному набору данных.

Важнейшими в процессе идентификации по малому числу наблюдений являются этапы анализа информативности и преобразования данных. Каковы бы ни были конкретные методы, реализующие эти этапы, их наличие в общей схеме идентификации является принципиальным. Подчеркнем, что этот вывод касается не какого-либо конкретного "изобретенного" алгоритма, а отражает общие закономерности обработки малых наборов данных, которые могут стать основой для построения автоматизированных систем обработки изображений [13].

В данном случае на основе анализа ситуации на каждом фрагменте изображения должны вырабатываться решения относительно отдельных компонент фиксированного набора данных. Используемая для этого дополнительная ин-

формация содержится в невязках, а вырабатываемые решения должны опираться на объективно существующие устойчивые связи между компонентами векторов ошибок и невязок. За счет многократной циклической обработки (преобразования) одного и того же набора измерений с использованием этой связи достигается наиболее полное извлечение полезной информации из каждого набора данных.

Кроме указанных этапов анализа информативности и преобразования данных циклическая схема обработки должна включать также процедуру останова. Поскольку в данном случае не используются априорные вероятностные модели и критерии, правило останова должно строиться на основе мер близости, характеризующих точность на конкретной реализации. В данном случае целесообразно использовать в качестве меры точности инвариантную к ортогональным преобразованиям евклидову норму вектора ошибок идентификации. Правило останова должно быть также достаточно гибким и учитывать дополнительную информацию, появляющуюся в ходе обработки каждого конкретного набора данных.

Наконец, при решении задач оценки параметров моделей систем формирования изображений искомые параметры, обычно, приблизительно известны. Они рассчитываются на этапе проектирования системы, а целью задачи идентификации является их уточнение в ходе эксплуатации системы. Следовательно, общая схема обработки данных должна также иметь возможность учета этой априорной информации. Использование этой информации часто позволяет снизить размерность исходной задачи идентификации.

Необходимость снижения порядка модели, кроме прочего, может быть вызвана также стремлением улучшить обусловленность задачи. Это возможно, если некоторые компоненты вектора искомых параметров приблизительно известны из измерительных данных других источников или оценок, полученных в предыдущем эпизоде идентификации.

Таким образом, общий алгоритм идентификации по малому числу наблюдений должен включать следующие процедуры:

- анализ обусловленности и/или мультиколлинеарности и отбор наиболее информативных данных;
- снижение (при необходимости учета априорной информации и/или улучшения обусловленности) порядка модели;
- анализ невязок и формирование соответствующего преобразования данных;

- выработка решения об останове процесса попеременного оценивания и преобразования данных.

На рис. 5.1 приведена схема так называемого базового алгоритма идентификации, реализующего указанные выше этапы обработки данных. Процедуры, реализующие отдельные этапы общего алгоритма, могут изменяться в зависимости от конкретной реализации или даже от номера шага преобразования в ходе обработки одной реализации. Ниже приводится краткое описание логики функционирования и назначения отдельных процедур, фигурирующих в этой схеме.



Рис.5.1. Общая схема базового алгоритма идентификации

Как уже отмечалось, важнейшими являются этапы контроля информативности и преобразования данных. В блоке анализа информативности и отбора данных вырабатываются количественные оценки обусловленности. С использованием этих оценок и заданной предельной величины нормы вектора ошибок исходных данных строится прогноз достижимой точности идентификации. Для этого (в зависимости от имеющихся вычислительных ресурсов и требований к качеству принимаемых решений) используются показатель диагонального преобладания и/или показатель минимальной сопряженности с нуль-пространством. Если требуемая степень обусловленности задачи не достигается, включается одна

из процедур отбора наиболее информативных данных. Эти процедуры подробно описываются в разделе 5.2.

Если задача хорошо обусловлена и требуемая точность (при заданных ограничениях на норму вектора ошибок в исходных данных) достижима, осуществляется переход к итерационной схеме обработки информации, заключающейся в попеременном вычислении МНК-оценок и преобразованиях данных типа взвешивания и/или непосредственной корректировки вектора выхода модели.

Если требуемая точность на заданном фиксированном наборе данных недостижима, анализируется возможность улучшения обусловленности задачи путем понижения порядка модели. Для этого используется априорная информация о компонентах вектора параметров c, соответствующих исключаемым переменным. Если такая информация отсутствует, решение задачи идентификации должно быть отложено до поступления нового (более информативного) набора данных.

Процедура снижения порядка строится в виде последовательности шагов, на каждом из которых исключается один из векторов-столбцов матрицы X, при этом порядок модели уменьшается на единицу. После каждого шага проверяется целесообразность дальнейшего снижения порядка. Для этого повторяются анализ обусловленности задачи и оценка достижимой точности идентификации. В первую очередь исключаются те столбцы матрицы X, для которых имеет место сильная мультиколлинеарность. Для их индикации может использоваться коэффициент парной сопряженности и/или показатель сопряженности с нуль-пространством.

Процедуры понижения порядка модели и анализа обусловленности повторяются до тех пор, пока не будут достигнуты требуемые значения мер информативности. Если результаты контроля информативности исходного (или уменьшенной размерности) набора данных удовлетворительные, то осуществляется переход к итерационной обработке сформированного таким образом фиксированного набора данных, которая реализуется в виде следующей последовательности шагов:

1. Вычисление текущей МНК-оценки. На этом этапе, обычно, решается система нормальных уравнений методом исключения Гаусса.
2. Вычисление невязок и выработка признака продолжения процесса уточнения оценок. По этому признаку осуществляется разветвление процесса.

3. Если принято решение о продолжении процесса, вновь анализируются невязки, строится соответствующее преобразование данных и осуществляется переход к шагу 1.

Подчеркнем, что хотя обработка данных в описанной схеме осуществляется итерационно, она не имеет ничего общего с рекуррентными алгоритмами идентификации, в которых происходит уточнение оценок при поступлении каждого нового наблюдения. Описанная схема предназначена для обработки фиксированных наборов данных. Анализ возможности использования для реализации преобразований взвешивания рекуррентной схемы обработки данных с выдачей оценок параметров на промежуточных шагах будет проведен в разделе 5.6. В частности, для двух вариантов взвешивания: рекуррентного и одноразового будут приведены сравнительные оценки вычислительной сложности.

Далее в настоящей главе приводятся подробные описания перечисленных процедур общего алгоритма идентификации. Последовательность этих описаний соответствует последовательности появления этих процедур в общей схеме. Поэтому первой в этом ряду оказалась играющая, вообще говоря, вспомогательную роль процедура масштабирования данных.

5.2. Процедура масштабирования независимых переменных

Выше показано, что масштабированием независимых переменных можно улучшить обусловленность задачи, в частности, увеличить диагональное преобладание матрицы $\mathbf{A}=\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Для этого предложено использовать диагональную M×M-матрицу масштабных коэффициентов $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_M)$, элементы которой определяются по соотношениям:

$$d_k = 1/\sqrt{a_{k,k}}, \quad k = \overline{1,M}, \tag{5.1}$$

где $a_{k,k}$ - диагональные элементы матрицы **A**.

Существуют две возможности для реализации процедуры масштабирования. Выбор какой-либо из них зависит от принятого в общей схеме метода вычисления оценок. Если оценки вычисляются путем решения системы нормальных уравнений $\mathbf{A}\hat{\mathbf{c}} = \mathbf{b}$, то целесообразно масштабировать саму матрицу **A**:

$$\mathbf{A}(\mathbf{D}) = \mathbf{DAD}. \tag{5.2}$$

Если же система нормальных уравнений не формируется, то масштабирование должно заключаться в линейном преобразовании данных по правилу:

$$\mathbf{X(D) = XD}. \qquad (5.3)$$

В любом случае соответствующая матрица $\mathbf{A(D) = DAD = X^T(D)X(D) = R}$ является матрицей сопряженности.

Процедура масштабирования, приводящая к построению матрицы сопряженности, в общей схеме идентификации является подготовительной. Ее задача увеличение диагонального преобладания матрицы **A** в ситуации, когда плохая обусловленность связана с различиями в физических масштабах независимых переменных. Реализация этой процедуры обеспечивает более эффективное решение задачи оценки мультиколлинеарности векторов независимых переменных.

Если масштабирование данных выполнено, то дальнейшее вычисление оценок также целесообразно осуществлять по преобразованным данным. Поэтому для восстановления исходной физической размерности искомых параметров по правилу $\hat{\mathbf{c}} = \mathbf{D} \cdot \hat{\mathbf{c}}(\mathbf{D})$ найденные масштабные коэффициенты должны храниться в памяти до завершения процесса идентификации.

В остальном работа алгоритмов (преобразования данных и вычисления оценок) не зависит от того, какие (масштабированные или не масштабированные) данные используются для решения задачи. Поэтому для матриц **X**, **A** мы будем всегда применять указанные обозначения, даже если при этом, в действительности, имеются в виду матрицы **X(D)**, **A(D)**. Связанные с масштабированием различия будем оговаривать лишь там, где это необходимо.

5.3. Решающие правила и алгоритмы анализа информативности

Во второй главе показана связь информативности данных с обусловленностью задачи и предложены меры обусловленности и мультиколлинеарности: показатель диагонального преобладания матрицы Грама и показатель сопряженности с нуль-пространством. В настоящем разделе описываются основанные на применении указанных мер решающие правила анализа информативности. Приводится несколько вариантов процедур, ориентированных на различные вычислительные возможности и требования к качеству оценок. Последовательность описания вариантов соответствует увеличению вычислительной сложности и эффективности оценок.

Наиболее простым в вычислительном отношении является алгоритм анализа обусловленности и оценки достижимой точности по показателю диагонального преобладания. Основой для построения алгоритма являются неравенства (2.40)

(2.42). Из этих неравенств, в частности, видно, что во всех случаях оценка достижимой точности идентификации сводится к сравнению величины диагонального преобладания матрицы **A** с некоторой заранее вычисленной константой:

$$\Phi = \phi - M + 1 \geq \varepsilon, \quad 0 \leq \varepsilon \leq 1. \tag{5.4}$$

Алгоритм состоит из двух шагов:

Шаг 1. Вычисление диагонального преобладания матриц (**A** и/или **R**) по соотношениям

$$\phi(\mathbf{A}) = \mathrm{tr}^2\mathbf{A}/\mathrm{tr}\mathbf{A}^2 = \left(\sum_{i=1}^{M} a_{ii}\right)^2 \Big/ \sum_{i,j=1}^{M} a_{ij}^2, \quad 1 \leq \phi(\mathbf{A}) \leq M, \tag{5.5}$$

$$\phi(\mathbf{R}) = \mathrm{tr}^2\mathbf{R}/\mathrm{tr}\mathbf{R}^2 = \left(\sum_{i=1}^{M} r_{ii}\right)^2 \Big/ \sum_{i,j=1}^{M} r_{ij}^2, \quad M - 1 \leq \phi(\mathbf{R}) \leq M. \tag{5.6}$$

где tr(·) - след матрицы (**A** или **R**).

Шаг 2. Проверка неравенства (5.4), которое в интересах экономии вычислений здесь представляется в виде:

$$\eta_i = M - 1 + \varepsilon \leq \phi = M - 1 + \Phi. \tag{5.7}$$

Напомним, что если для контроля информативности используется $\phi(\mathbf{R})$, то фактически анализируется лишь мультиколлинеарность векторов независимых переменных. Величина $\phi(\mathbf{A})$ позволяет дать интегральную оценку обусловленности, учитывающую различия в масштабах независимых переменных. Решающее значение при реализации этого алгоритма имеет выбор порогового значение величины η_i. Поэтому проведем более детальное обоснование возможных вариантов его определения.

Ниже приведены построенные с использованием соотношений (2.40)-(2.42) выражения определяющие величину порога η_i для различных типов идентифицируемых моделей:

а) для модели типа КИХ-фильтра ($\delta_A = 0 \quad \delta_\Sigma = \delta_b$)

$$\eta_1 = M - 1 + \frac{2\delta_b\delta_c}{\delta_b^2 + \delta_c^2}, \quad \delta_c \geq \delta_b; \tag{5.8}$$

б) для модели типа БИХ-фильтра

$$\eta_2 = M - 1 + \frac{2\left(\delta_\Sigma\,\delta_c + \delta_c^2\delta_A\right)}{\delta_c^2 + \delta_\Sigma^2 + 2\delta_\Sigma\,\delta_c\delta_A + \delta_c^2\delta_A^2}, \quad \delta_c \geq \delta_b, \tag{5.9}$$

где $\delta_\Sigma = \delta_b + \delta_A$, $\delta_b = \|\delta\mathbf{b}\|/\|\mathbf{b}\|$, $\delta_A = \|\delta\mathbf{A}\|/\|\mathbf{A}\|$, а $\delta_c = \|\Delta\mathbf{c}\|/\|\mathbf{c}\|$;

в) для любых моделей (типа КИХ- и/или БИХ-)фильтра

$$\eta_3 = M - 1 + \frac{k^2 tr^2 A - (M-1)^2 (k \cdot trA - M)^2}{k^2 tr^2 A + (M-1)(k \cdot trA - M)^2}, \qquad (5.10)$$

где $k = \frac{\|\Delta c\|^2}{\|\xi\|^2}$, $\frac{M-1}{trA} \le k \le \frac{M(M-1)}{(M-2)trA}$.

При принятии решения о качестве исходных данных по соотношениям (5.8), (5.9) величины относительных погрешностей могут быть неизвестными. При этом требования к точности оценок могут быть сформулированы в виде допустимого коэффициента $k_{отн}$, показывающего, во сколько раз относительная погрешность оценивания параметров может превышать погрешность исходных данных: $\delta_c = k_{отн}\delta$. Для моделей типа БИХ-фильтра обычно справедливо приближенное равенство $\delta_b = \delta_A = \delta$. При этом фигурирующая в соотношении (5.9) величина $\delta_\Sigma = \delta_b + \delta_A = 2\delta$. С учетом сказанного, наряду с (5.8), (5.9) можно построить более простые выражения для границ η_1, η_2:

а) для моделей типа КИХ-фильтра

$$\eta_1 = M - 1 + \frac{2k_{отн}}{1 + k_{отн}^2}, \quad k_{отн} \ge 1; \qquad (5.11)$$

б) для моделей типа БИХ-фильтра

$$\eta_2 = M - 1 + \frac{2k_{отн}(2 + k_{отн}\delta)}{4(1 + k_{отн}\delta) + k_{отн}^2(1 + \delta^2)}, \quad k_{отн} \ge 1. \qquad (5.12)$$

Вместо выражения (5.10) можно использовать более простое соотношение вида

$$\eta_3 = M - 1 + \frac{k_t^2 - (M-1)^2 (k_t - M)^2}{k_t^2 + (M-1)(k_t - M)^2}, \qquad (5.13)$$

где $k_t = k \cdot trA = \frac{\|\Delta c\|^2}{\|\xi\|^2} \cdot trA$, $M - 1 \le k_t \le \frac{M(M-1)}{M-2}$. Эта оценка справедлива для данных, которые перед обработкой масштабированы по правилу $\tilde{X} = X \cdot (trA)^{1/2}$, $\tilde{y} = y \cdot (trA)^{1/2}$. Поскольку при этом величина $\|\Delta c\|^2$ не изменяется, а $\|\tilde{\xi}\|^2 = \|\xi\|^2 \cdot trA$, величины k_t для масштабированных данных совпадают со значениями параметра $k = \|\Delta c\|^2 / \|\xi\|^2$ в соотношении (5.10) для немасштабированных данных. В таблице 5.1 приведены вычисленные по соотношению (5.8) числовые

значения порога η_1 для M=2. На рис. 5.2 показана зависимость порога η_1 от величины $k_{отн}$ при M=2.

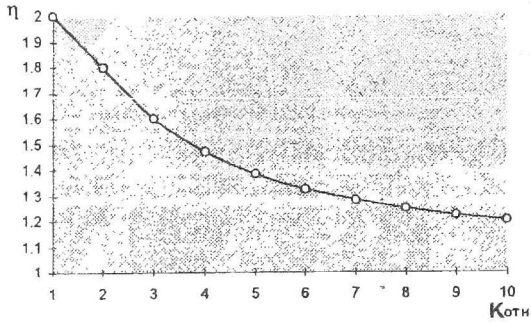


Рис. 5.2. Зависимость порога η_1 от величины $k_{отн}$ для M=2

Таблица 5.1.
Порог η_1

k	η_1
1	2.00
2	1.80
3	1.60
4	1.47
5	1.38
6	1.32
7	1.28
8	1.25
9	1.22
10	1.20

Пороговые значения η_2, вычисленные по соотношению (5.12) для различных $k_{отн}$ при M=2, приведены в таблице 5.2. Соответствующее данным этой таблицы семейство зависимостей показано на рис. 5.3.

Таблица 5.2. Пороговые значения η_2 для различных $k_{отн}$ при M=2

δ	η_2					
	$k_{отн}$=1	$k_{отн}$=2	$k_{отн}$=3	$k_{отн}$=4	$k_{отн}$=5	$k_{отн}$=10
0	1.8	2	1.923	1.8	1.689	1.384
0.1	1.776	1.995	1.965	1.882	1.8	1.550
0.2	1.753	1.983	1.989	1.939	1.882	1.689
0.3	1.731	1.967	1.999	1.976	1.939	1.8
0.4	1.710	1.946	1.998	1.994	1.976	1.882
0.5	1.690	1.923	1.988	2	1.994	1.940
0.6	1.670	1.899	1.973	1.996	2	1.976
0.7	1.651	1.874	1.953	1.984	1.995	1.994
0.8	1.633	1.849	1.931	1.967	1.984	2
0.9	1.616	1.824	1.907	1.946	1.967	1.995
1.0	1.6	1.8	1.882	1.923	1.946	1.984

Наконец, в таблице 5.3 приведены результаты расчета, а на рис. 5.4 соответствующее соотношению (5.13) семейство зависимостей для порогового значения η_3 от величины δ при различных размерностях (M) задачи. Приведенные табли-

цы и графики могут непосредственно использоваться для выбора конкретных числовых значений порогов при построении решающих правил и алгоритмов.

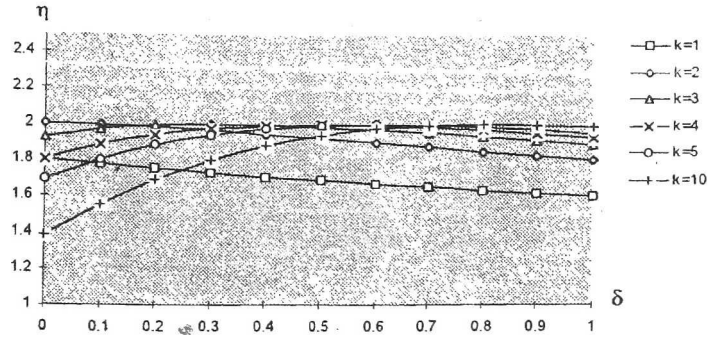


Рис. 5.3. Зависимости порога η_2 от величины δ для различных k при M=2

В задачах обработки изображений процедура анализа обусловленности может иметь различные формы реализации в зависимости от характера конкретной задачи. Если предъявленный набор данных является единственным, то результаты сравнения допустимых пороговых значений с фактическими значениями мер обусловленности и/или мультиколлинеарности являются лишь констатацией факта (хорошей или плохой) информативности данных на данном фрагменте изображения.

Таблица 5.3. Значения порога η_3 для различных k_t и M

k_t	η			
	M=2	M=3	M=4	M=5
1	1.0000			
1.5	1.8000			
2	2.0000	2.0000		
2.5	1.9230	2.7778		
3	1.8000	3.0000	3.0000	
3.5	1.6897	2.8824	3.7692	
4	1.6000	2.6667	4.0000	4.0000
4.5	1.5283	2.4545	3.8571	4.7647
5	1.4706	2.2727	3.5714	5.0000
5.5	1.4235	2.1228	3.2700	4.8400
6	1.3846	2.0000	3.0000	4.5000

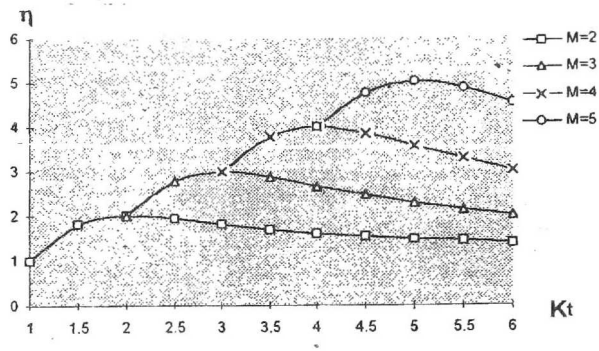


Рис. 5.4. Зависимость порога η_3 от величины k_t для различных M

Если преследуется цель подыскать на изображениях наиболее подходящие для решения задачи идентификации фрагменты, задача сводится к определению набора данных, для которого значения мер обусловленности и/или мультиколлинеарности не только достигают пороговых значений, но и принимают значения, свидетельствующие о том, что чувствительность к ошибкам измерений и вычислений на этом наборе данных наименьшая. Соответствующая процедура фактически осуществляет отбор наиболее информативных данных по всему изображению.

Из приведенных таблиц и графиков видно, что гарантирующие высокую точность оценивания допустимые значения порогов для показателя диагонального преобладания, должны принимать также высокие значения (ϕ>M-1). Это не всегда достигается, особенно на полутоновых, изображениях. В этом случае (если позволяют вычислительные ресурсы) для оценки информативности можно использовать более сильные меры, основанные на вычислении сопряженности векторов независимых переменных с нуль-пространством.

Для построения алгоритма оценки сопряженности используются соотношения (2.48), (2.52) и (2.54). Алгоритм можно представить в виде следующей последовательности шагов.

Шаг 1. Выбор очередного вектора $\mathbf{x}_i$, $i = \overline{1,M}$ и формирование соответствующей N×(M-1)-матрицы $\mathbf{X}_{M-1}$ (из матрицы $\mathbf{X}$ исключается вектор-столбец $\mathbf{x}_i$).

Шаг 2. Определение N×(N-M+1)-матрицы $\mathbf{T}_0$, составленной из собственных векторов, соответствующих нулевым собственным значениям матрицы $\mathbf{X}_{M-1}\mathbf{X}^T_{M-1}$ (базиса нуль-пространства матрицы $\mathbf{X}_{M-1}$).

вычисление очередной величины (2.52):

$$S_i = \left(\mathbf{x}_i^T \mathbf{P}_0 \mathbf{T}_0^T \mathbf{x}_i\right)^{1/2} \Big/ \left(\mathbf{x}_i^T \mathbf{x}_i\right)^{1/2}$$

Если все S_i, $i = \overline{1,M}$ вычислены, переходим к шагу 4, иначе повторяются шаги 1-3 для очередного вектора $\mathbf{x}_i$.

Шаг 4. Вычисление мер (2.48) и (2.54):

$$S_{min} = \min_i |S_i|, \qquad S_\Sigma = \left(\sum S_i^2\right)^{1/2}$$

и принятие решения о возможности решения задачи идентификации. Решающее правило строится аналогично (5.7) и заключается в сравнении указанных мер (обоих одновременно или одной из них) с заданными пороговыми значениями:

$$S_{min} \geq S_{min,доп}, \qquad S_\Sigma \geq S_{\Sigma доп}. \tag{5.14}$$

Описанная процедура в вычислительном отношении, конечно, гораздо сложнее, чем оценка информативности по показателю диагонального преобладания. Такова неизбежная плата за качество оценок. Основные вычислительные затраты связаны с определением базиса нуль-пространства. Поэтому применение этой процедуры целесообразно в тех случаях, когда этот базис используется также для других целей, например, для определения корректирующего вектора при непосредственной корректировке данных (см. раздел 4.1).

Если информативность данных на фрагменте изображения вызывает сомнения, можно попытаться улучшить ее путем отсеивания наблюдений, приводящих к ухудшению обусловленности задачи. Ниже приводятся описания основанных на этой идее решающих правил и алгоритмов формировании наборов данных в задачах идентификации оптических искажающих систем. Результаты исследования эффективность этих процедур можно найти в работах [61,62,63,66].

5.4. Решающие правила и алгоритмы отбора данных

Как это ни парадоксально, но результат оценивания по небольшому числу информативных данных, "изъятых" из заданной выборки наблюдений, иногда может оказаться лучше, чем по всему (возможно большому) набору данных. Этот эффект, в частности, можно наблюдать при решении задачи идентификации с использованием изображения, на котором (информативный) объект занимает относительно небольшую площадь, а фон имеет более высокую, по сравнению с объектом, яркость. При этом, несмотря на усреднение ошибок исходных данных

при большом числе отсчетов, точность оценивания (из-за возрастания чувствительности к ошибкам) может оказаться значительно ниже, чем если бы мы "вырезали" этот информативный объект и решили задачу оценивания по малому числу наблюдений.

В общем случае задачу отбора наиболее информативных данных можно сформулировать следующим образом. Из исходного протокола наблюдений объема N, которому соответствует оценка $\hat{\mathbf{c}}$, необходимо сформировать новый протокол объема N1 (N1<N), так, чтобы построенная для нового набора данных по тому же правилу (критерию) оценка $\hat{\mathbf{c}}1$ была, по крайней мере, не хуже прежней. Заметим, что здесь не ставится задача выбрать наилучший (при заданных N1 и критерии качества) из всех возможных (C_N^{N1}) вариантов. Решение такой задачи, даже при сравнительно небольшом объеме данных, потребовало бы значительных вычислительных затрат. В данном случае формулируются простые правила отбора данных с использованием введенных ранее мер информативности.

Пусть $\hat{\mathbf{c}}$ и $\hat{\mathbf{c}}1$ являются МНК- оценками для наборов данных объема N и N1 соответственно. Из соотношений (1.37), (1.39) видно, что изменение меры качества $\|\Delta\mathbf{c}\|_2^2$ при отбрасывании части наблюдений может происходить вследствие изменения следующих характеристик полезных сигналов и ошибок исходных данных:

- изменение ориентации вектора ошибок измерений относительно пространства матрицы $\mathbf{X}$;
- изменение собственных значений λ_i матрицы $\mathbf{X}^T\mathbf{X}$;
- изменение нормы вектора ошибок $\|\xi\|$.

Общее решение задачи с учетом всех указанных факторов не представляет большого практического интереса. Дело в том, что ориентация вектора ошибок ξ все равно никогда неизвестна, а приближенные значения отдельных его компонент могут быть определены лишь после получения начальных МНК-оценок и вычисления вектора невязок. Поэтому при построении процедур отбора информативных данных имеет смысл учитывать лишь изменения собственных значений и нормы вектора ошибок исходных данных.

Из соотношения (1.39) видно, что при одинаковой ориентации вектора ошибок относительно пространства матрицы $\mathbf{X}$ повышение точности, связанное с

отсеиванием части наблюдений, может произойти, если собственные значения матрицы $\mathbf{X}^T\mathbf{X}$ увеличиваются "быстрее", чем уменьшается величина $\|\xi\|^2$. Следовательно, представляется конструктивным следующее правило: при сортировке целесообразно отсеивать данные, присутствие которых вызывает уменьшение чисел λ_i, $i=\overline{1,M}$.

Связанное с сортировкой относительное увеличение собственных значений λ_i может происходить как из-за увеличения масштаба независимых переменных, так и вследствие уменьшения степени мультиколлинеарности (сопряженности) векторов независимых переменных. Этот эффект будет тем более заметным, чем меньший относительный вес (по норме) имеют отбрасываемые данные. С учетом сказанного, правило отбора информативных данных должно заключаться в отсеивании строк матрицы **X**, которые при сравнительно малой величине нормы вносят значительный вклад в увеличение сопряженности (мультиколлинеарности) векторов-столбцов этой матрицы.

Плохая обусловленность может быть также связана с наличием в исходных данных небольшого числа выделяющихся (по норме) наблюдений. Например, если норма одной из строк матрицы **X** намного превышает сумму норм остальных строк, то M-1 собственных значений будут близки к нулю. Таким образом, "подозрительными" при сортировке являются также строки матрицы **X** сильно выделяющиеся (по норме), по сравнению со средним значением норм остальных строк.

Приведенные рассуждения, конечно, не являются обоснованием процедур отбора информативных данных. В данном случае их цель методическая: объяснить, почему соответствующие алгоритмы оказываются работоспособными. В частности, рассматривалась двухэтапная процедура отбора информативных данных. Первый этап — выявление "подозрительных" данных и второй — анализ степени вносимой этими данными мультиколлинеарности и связанного с этим ухудшения обусловленности.

На первом этапе работа алгоритма описывается в виде следующей последовательности шагов.

Шаг 1. Вычисление норм $\|x_i\|$, $i=\overline{1,N}$ всех строк матрицы **X**.

Шаг 2. Вычисление среднего норм:

$$\bar{\Sigma}_N = \frac{1}{N}\sum_{i=1}^{N}\|x_i\|. \qquad (5.15)$$

Шаг 3. Вычисление весовых коэффициентов

$$k_i = \frac{k_0}{\left|\bar{\Sigma}_N - \|x_i\|\right| + \alpha\bar{\Sigma}_N}, \quad i = \overline{1,N}, \qquad (5.16)$$

где α малое положительное число ($0 \leftarrow \alpha < 1$), исключающее возможность деления на нуль, а k_0 - константа, выбираемая из условий принятой нормировки (например $\sum k_i = N, \quad i = \overline{1,N}$).

В результате реализации указанных шагов будет получен набор N неотрицательных чисел k_i, $i = \overline{1,N}$, каждое из которых будет характеризовать степень доверия отдельным наблюдениям (строкам матрицы **X**). Строки, соответствующие малым и большим значениям коэффициентов k_i, $\sum k_i = N, \quad i = \overline{1,N}$, считаются "подозрительными".

Далее с использованием коэффициентов k_i, $i = \overline{1,N}$ исходный протокол измерений разбивается на три массива, которые из соображений удобства изложения условно назовем: "Низкий", "Средний" и "Высокий". В "Низкий" заносятся данные, для которых $k_i \le k_{доп,min}$, в "Высокий" $k_i \ge k_{доп,max}$ и в "Средний" - все остальные наблюдения. "Низкий" массив отсеивается из исходного набора измерений независимо от оценки степени обусловленности содержащихся в нем данных. "Высокий" массив отсеивается лишь в том случае, если его информативность намного ниже "Среднего"

Для количественной оценки степени обусловленности может использоваться либо показатель диагонального преобладания, либо одна из описанных в разделе 2.5 мер сопряженности векторов с нуль-пространством. Результаты сравнительного исследования эффективности указанных мер содержатся в разделе 2.6. В шестой главе будут приведены информационные технологии, построенные с использованием описанных решающих правил.

Величины порогов $k_{доп,min}$ и $k_i \ge k_{доп,max}$ могут устанавливаться на основе пробных просчетов путем моделирования "похожих" тестовых изображений, либо определяться на каждом фрагменте для заданных допустимых процентных соотношений между объемами "Низкого", "Среднего" и "Высокого" массивов.

Можно существенно сократить вычислительные затраты исключив этап предварительного выявления "подозрительных" измерений. Такая возможность

связана со способом формирования матрицы **X**. Если данные с изображения считываются последовательно, то соседние строки этой матрицы будут соответствовать соседним точкам на изображении. При этом блоки матрицы, соответствующие компактно расположенным участкам изображения с постоянным и притом достаточно заметным уровнем яркости (например светлый фон, на котором отсутствуют информативные объекты) окажутся плохо обусловленными или даже вырожденными и будут существенным образом влиять на обусловленность всей матрицы.

Алгоритм отсеивания таких наблюдений может заключаться в последовательном просмотре (всех подряд) блоков матрицы **X** "скользящим окном", соответствующим предполагаемым размерам участков фона. При этом должно просматриваться N-N1+1 блоков матрицы **X**. Даже при сравнительно небольших размерах фрагмента число вариантов оказывается достаточно большим.

Число вариантов значительно сокращается, если блоки брать "без перекрытия" Для установления факта плохой обусловленности и/или вырожденности блока в данном случае целесообразно использовать наиболее простую характеристику показатель диагонального преобладания матрицы $\mathbf{A}=\mathbf{X}^T\mathbf{X}$. В частности, блок, норма которого превосходит норму матрицы, составленной из остальных данных, и вдобавок для него $\Phi \cong 0$ и/или $\phi \cong 1$, должен быть исключен из общего набора данных.

5.5. Организация итеративного процесса преобразования данных и вычисления оценок

Процесс итеративной обработки каждого фиксированного набора данных начинается с вычисления начальной МНК-оценки. Разрешение на решение исходной системы уравнений (1.19) дается лишь в том случае, если в результате анализа информативности данных установлена принципиальная возможность достижения требуемой точности идентификации. После этого, в соответствии с общей схемой алгоритма, показанного на рисунке 5.1, может быть реализовано итеративное вычисление МНК-оценок с промежуточными преобразованиями взвешивания или непосредственной корректировки данных.

Выбор типа преобразования осуществляется в соответствии со следующими решающими правилами.

1. Если компоненты вектора невязок группируются относительно центра так, что относительный сдвиг среднего и медианы превышает заданное допустимое значение, это свидетельствует о наличии в данных выделяющихся из общей массы (по абсолютной величине) компонент вектора невязок одного знака. В этом случае должна осуществляться непосредственная корректировка вектора **y** выхода модели. Процедура корректировки подробно описана в разделе 4.1, а исследования, послужившие основанием для ее применения, в третьей главе.
2. Если компоненты вектора невязок группируются симметрично относительно "центра" (возможно в результате корректировки вектора **y**), целесообразно осуществлять взвешивание данных. Преобразование заключается в придании меньшего веса наблюдениям, для которых невязки оказались сравнительно большими по абсолютной величине. Различные методы построения весовой матрицы рассматривались в четвертой главе.

Непосредственная корректировка вектора выхода модели **y** осуществляется с целью устранения относительного сдвига среднего и медианы невязок и реализуется в виде следующей последовательности шагов.

Шаг 1. Вычисление очередной (или начальной) МНК-оценки. Для этого методом исключения Гаусса решается система уравнений $\mathbf{A}\hat{\mathbf{c}} = \mathbf{b}$ (1.19).

Шаг 2. Вычисление и анализ невязок $\hat{\xi} = \mathbf{X}\hat{\mathbf{c}} - \mathbf{y}$ (1.7). Анализ невязок начинается с выстраивания вариационного ряда (4.6) и оценки степени его асимметрии по соотношениям (4.8), (4.9). Если в соответствии с неравенством (4.9) величина "сдвига" $\Delta_{\text{фр}} = \left[\bar{\xi} - \text{med}\left(\hat{\xi}\right)\right] / \left\|\hat{\xi}\right\|$ превышает величину 0,1 осуществляется переход к следующему шагу, т.е. к блоку "Преобразование данных".

Шаг 3. Корректировка вектора **y** в соответствии с соотношением (4.10). Корректировка направлена на уменьшение веса выделяющихся ошибок в компонентах вектора **y**. После замены вектора **y** скорректированным вектором $\tilde{\mathbf{y}}$ снова формируется система уравнений вида (1.19) и осуществляется переход к шагу 1.

Операции вычисления невязок, определения степени асимметрии вариационного ряда и симметризации повторяются до тех пор, пока выполняется неравенство (4.9). При отсутствии сильно выделяющихся ошибок одного знака достаточная степень симметрии вариационного ряда, составленного из компонент вектора невязок, может достигаться сразу на исходном наборе данных. В

этом случае корректировка вектора выхода y не производится, а сразу осуществляется переход к следующему этапу обработки информации.

На этом этапе процедура оценки степени асимметрии заменяется процедурой выработки признака целесообразности дальнейшего уточнения оценок (правило останова). При построении этой процедуры используются следующие соображения.

Часто правило останова задают в виде

$$\left\|\Delta\hat{\mathbf{c}}_l\right\| \le \delta_{\text{с.доп}}, \quad \Delta\hat{\mathbf{c}}_l = \hat{\mathbf{c}}_l - \hat{\mathbf{c}}_{l-1}, \tag{5.17}$$

где $\delta_{\text{с.доп}}$ заданное число, а $\hat{\mathbf{c}}_l, \hat{\mathbf{c}}_{l-1}$ - оценки, полученные на соседних (l-м и l-1 м циклах) циклической обработки данных. Малые значения $\left\|\Delta\hat{\mathbf{c}}_l\right\|$ могут быть связаны не только с приближением к точке минимума, но также и с большой крутизной функции потерь. Поэтому наряду с (5.17) необходимо потребовать выполнения следующего условия:

$$\left\|\Delta Q\left(\hat{\mathbf{c}}_l\right)\right\| \le \delta_{\text{Q.доп}}, \quad \Delta Q\left(\hat{\mathbf{c}}_l\right) = Q\left(\hat{\mathbf{c}}_l\right) - Q\left(\hat{\mathbf{c}}_{l-1}\right), \tag{5.18}$$

где $\delta_{\text{Q.доп}}$ заданное число, а $Q\left(\hat{\mathbf{c}}_l\right), Q\left(\hat{\mathbf{c}}_{l-1}\right)$ - значения критерия на тех же соседних (l-м и l-1 м циклах) обработки.

Для того чтобы правило останова было инвариантным к физической размерности задачи, целесообразно также перейти к относительным величинам:

$$\frac{\left\|\Delta\hat{\mathbf{c}}_l\right\|}{\left\|\hat{\mathbf{c}}_l\right\|} \le \delta_{\text{с.отн}}, \quad \Delta\hat{\mathbf{c}}_l = \hat{\mathbf{c}}_l - \hat{\mathbf{c}}_{l-1}, \tag{5.19}$$

$$\frac{\left\|\Delta Q\left(\hat{\mathbf{c}}_l\right)\right\|}{\left\|Q\left(\hat{\mathbf{c}}_l\right)\right\|} \le \delta_{\text{Q.отн}}, \quad \Delta Q\left(\hat{\mathbf{c}}_l\right) = Q\left(\hat{\mathbf{c}}_l\right) - Q\left(\hat{\mathbf{c}}_l\right), \tag{5.20}$$

где $\delta_{\text{с.отн}}, \delta_{\text{Q.отн}}$ - заданные величины, а все остальные обозначения те же, что и в (5.17), (5.18).

С момента включения правила останова и на всех последующих шагах (до его "срабатывания") применяется преобразование данных типа взвешивания. В общей схеме алгоритма и в названиях блоков при этом ничего не меняется и он реализуется в виде следующей последовательности шагов.

Шаг 1. Вычисление очередной (или начальной) МНК-оценки $\hat{\mathbf{c}}$.

Шаг 2. Вычисление и анализ невязок $\hat{\xi} = \mathbf{X}\hat{\mathbf{c}} - \mathbf{y}$ и принятие решения о целесообразности дальнейшего уточнения оценок $\hat{\mathbf{c}}$ (правило останова). Номера

формул, по которым эти правила реализуются для различных способов преобразования данных, перечислены в таблице 5.1.

Шаг 3. Определение весовой матрицы **G** и преобразование данных (взвешивание). Методы построения преобразований взвешивания также перечислены в таблице 5.4. В последней колонке этой таблицы указаны также разделы главы 4, в которых содержится подробное описание соответствующих методов.

Таблица 5.4. Методы преобразования данных

№ п/п	Метод построения преобразования	Правило останова	Раздел
1	Построение корректирующего вектора с использованием базиса нуль-пространства	(5.19)	4.1
2	Корректировка компонент вектора **y**, соответствующих выделяющимся компонентам вектора невязок.	(5.19)	4.1
3	Метод взвешивания. Оптимизационный метод построения весовой матрицы	(5.19), (5.20)	4.2
4	Взвешивание с различением распределений по нормированному среднему модулей невязок.	(5.19), (5.20)	4.5
5	Модификации метода взвешивания, основанные на связи с L_v- оценками	(5.19), (5.20)	4.6

В таблице 5.4 указано пять основных способов построения преобразований и приведены номера формул, по которым могут быть реализованы соответствующие им правила остановки. Правило останова в виде совокупности двух неравенств (5.19) и (5.20), которые должны выполняться одновременно, может применяться в том случае, когда критерий качества идентификации используется для построения преобразований данных в явном виде. Поэтому для двух методов преобразования данных, фигурирующих в таблице 5.4 под номерами 1 и 2, правило останова указано в виде одного соотношения (5.19).

Теперь обсудим перечисленные методы преобразования с точки зрения их эффективности и вычислительной сложности. Наиболее простыми в вычислительном отношении являются методы, фигурирующие в таблице 5.4 под номерами 2 и 3. Более сложным, но эффективным является метод 1. Методы 1 и 2 не требуют задания какого-либо критерия или параметрического семейства функций. В данном случае для анализа ситуации и корректировки используются лишь устойчивые связи между выделяющимися компонентами векторов ошибок и невязок. Это позволяет избежать произвола в задании классов и выборе параметров преобразований при отсутствии другой априорной информации.

Необходимо подчеркнуть, что алгоритм корректировки данных, основанный на методе 2, чрезвычайно прост и обычно дает результаты, по крайней мере, не хуже, чем при взвешивании данных. Используемый здесь критерий согласования среднего и медианы также хорошо известен и обычно не вызывает возражений. В конкретных случаях при выборе параметров и компонент для корректировки может использоваться априорная информация о функции яркости регистрируемого объекта, условиях регистрации и др. Процедура построения корректирующего вектора (метод 1) всегда может рассматриваться как возможное продолжение метода 2. Она, конечно, позволяет восстановить все компоненты корректирующего вектора, но требует проведения большого объема дополнительной вычислительной работы.

5.6. Сравнение с рекуррентным взвешиванием данных

На страницах журналов и монографий часто можно встретить положительные оценки рекуррентных алгоритмов идентификации. Преимущества связывают с тем, что для вычисления оценок $\hat{\mathbf{c}}_M$ при любом сколь угодно большом N данные представляются M×M матрицей и M×1 вектором, а вычисления производятся без обращения матрицы. В связи с этим возникает естественный вопрос: возможна ли экономия вычислений за счет рекуррентной организации взвешивания данных. Для получения ответа на этот вопрос сравним две вычислительные схемы (назовем их условно I и II).

В схеме I, соответствующей базовому алгоритму идентификации (рисунок 5.1), вначале по всем доступным наблюдениям формируется система уравнений вида (1.19). В наиболее общем случае, когда применяется преобразование взвешивания ($\tilde{\mathbf{X}} = \mathbf{GX}$, $\tilde{\mathbf{y}} = \mathbf{Gy}$), для формирования этой системы уравнений можно использовать следующие рекуррентные соотношения:

$$\mathbf{A}_{k+1} = \mathbf{A}_k + g^2_{k+1}\boldsymbol{x}_{k+1}\boldsymbol{x}^T_{k+1}, \quad k = 0, N-1, \quad \mathbf{A}_0 = 0, \tag{5.21}$$

$$\mathbf{b}_{k+1} = \mathbf{b}_k + g^2_{k+1}\boldsymbol{x}_{k+1}y_{k+1}, \quad k = 0, N-1, \quad \mathbf{b}_0 = 0. \tag{5.22}$$

Затем полученная таким образом система M уравнений с M неизвестными (1.19):

$$\mathbf{A}\hat{\mathbf{c}} = \mathbf{b} \tag{5.23}$$

решается методом исключения Гаусса.

В схеме II оценки вычисляются рекуррентно по соотношениям вида [12,24,51,56]

$$\hat{\mathbf{c}}_{k+1} = \hat{\mathbf{c}}_k + g_{k+1}^2 \mathbf{P}_{k+1} \boldsymbol{x}_{k+1} \left(y_{k+1} - \boldsymbol{x}_{k+1}^{\mathrm{T}} \hat{\mathbf{c}}_k \right), \tag{5.24}$$

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \frac{\mathbf{P}_k \boldsymbol{x}_{k+1} \boldsymbol{x}_{k+1}^{\mathrm{T}} \mathbf{P}_k}{g_{k+1}^{-2} - \boldsymbol{x}_{k+1}^{\mathrm{T}} \mathbf{P}_k \boldsymbol{x}_{k+1}}. \tag{5.25}$$

Начальную оценку искомого вектора параметров и матричный коэффициент усиления, как это обычно принято [12,56], зададим в виде

$$\hat{\mathbf{c}} = 0, \quad \mathbf{P}_0 = \alpha \mathbf{E}, \quad \alpha \to \infty. \tag{5.26}$$

Обсудим вначале возможность использования рекуррентной схемы (5.24), (5.25) для получения начального приближения при построении модифицированной оценки (1.11).

Если все $g_k = 1, k = \overline{1,N}$, а начальные значения $\hat{\mathbf{c}}$ и $\mathbf{P}_0$ заданы так, как рекомендуется в (5.26), то, несмотря на быструю (в среднеквадратическом) сходимость матрицы $\mathbf{P}_k$ к $\mathbf{A}_k^{-1}$, какая-то часть полезной информации, содержащейся в "первых" 5-7 наблюдениях, будет безвозвратно утеряна.

Для более полного использования информации можно после первых q ($q \geq M$) шагов задать

$$\mathbf{P}_q = \mathbf{A}_q^{-1}, \text{ и } \hat{\mathbf{c}}_q = \mathbf{A}_q^{-1} \mathbf{b}_q, \tag{5.27}$$

воспользовавшись для этого схемой I. Затем (начав с q+1 шага) можно продолжить оценивание по рекуррентной схеме II. При таком задании начальных значений $\hat{\mathbf{c}}$, $\mathbf{P}_0$ вычислительные схемы (I и II) дадут совпадающие результаты. Однако, поскольку при этом на q-м шаге все равно придется решать систему (5.23) (вычисляя при этом обратную матрицу), то "включение" на оставшихся (N-q) шагах рекуррентной схемы II теряет смысл.

Может быть, получив с использованием вычислительной схемы I начальное МНК-приближение, вычислив невязки и определив тип и параметры преобразования, имеет смысл затем "включить" рекуррентную схему II для уточнения оценок? Поскольку с точки зрения эффективности использования информации в данном случае не возникает противопоказаний, рассмотрим эту возможность с точки зрения вычислительной сложности.

Для наглядности сравнительного анализа общее число операций (умножения и деления), необходимых для проведения вычислений по формулам (5.21), (5.22), (5.24), (5.25), в таблице 5.5 выписано в виде отдельных формул, со-

ставленных для фрагментов указанных выражений, реализуемых на различных этапах вычислительного процесса. Здесь, как и всюду ранее, M - размерность задачи (число оцениваемых параметров модели), а N - объем набора данных (N>M).

Таблица 5.5. Сравнение алгоритмов I и II по числу вычислительных операций

Алгоритм I		Алгоритм II	
Этап	Число длинных операций		Этап
Формирование матрицы **A** по соотношениям (5.25) (сначала умножение $g^2\boldsymbol{x}$)	$N(M^2+M)$	$N(M^2+M)$	Вычисление произведения $Z=\mathbf{P}\boldsymbol{x}$
Формирование вектора **b** по соотношению (5.26)	$N(M+1)$	$\left[\frac{M(M+1)}{2}\right]N$	Вычисление произведения $ZZ^T=\mathbf{P}\boldsymbol{x}\boldsymbol{x}^T\mathbf{P}$ (учитывается симметрия)
Решение (5.27) методом Гаусса (приведение к треугольной форме)	$\sum_{k=1}^{M-1}(M-k)\times$ $\times(M-k+2)$	$n \cdot N$	Вычисление $\boldsymbol{x}^T\mathbf{P}\boldsymbol{x}$ (с учетом того, что $\mathbf{P}\boldsymbol{x}=Z$ уже вычислена)
Решение (5.27) методом Гаусса (обратный ход)	$\frac{M(M-1)}{2}$	$\left[\frac{M(M+1)}{2}\right]N$	Деление всех элементов матрицы $\mathbf{P}\boldsymbol{x}\boldsymbol{x}^T\mathbf{P}$ на скаляр
Итого по всем этапам алгоритма I ⟶	$+\sum_{k=1}^{M-1}(M-k)\times$ $\times(M-k+2)+$ $+N(M^2+2M+$ $+1)+\frac{M(M-1)}{2}$	$N(M^2+2M+$ $+1)+N(M^2+$ $+M-1)$	Итого по всем этапам алгоритма II ⟵

При записи суммарного числа вычислительных операций для каждого алгоритма в последней строке таблицы мы специально сгруппировали слагаемые так, чтобы среди них оказались одинаковые: $N(M^2+2M+1)$. Исключим их из дальнейшего сравнения. Из оставшихся слагаемых видно, что если вместо суммы

$$\sum_{k=1}^{M-1}(M-k)\times\times(M-k+2)$$

взять сильно завышенную ее оценку в виде M^3, то все равно (с учетом того, что всегда выполняется неравенство N>M) алгоритм I оказывается более экономичным. Преимущество возрастает с ростом N. Для точной оценки вычислительной эффективности алгоритма I с использованием фигурирующих в последней строке таблицы 5.5 соотношений запишем разность

$$\Delta(N,M) = N\left(M^2 + M - 1\right) - \sum_{k=1}^{M-1}(M-k)\times(M-k+2) - \frac{M(M-1)}{2} \qquad (5.28)$$

На рис. 5.5 приведены графики зависимостей разности $\Delta(N,M)$ (5.28) от объема данных N при различной размерности задачи (числе оцениваемых параметров M). Вычислительный выигрыш равен ординатам точек пересечения вертикальных прямых, абсциссы которых равны объему данных, с прямой, соответствующей заданной размерности M задачи.

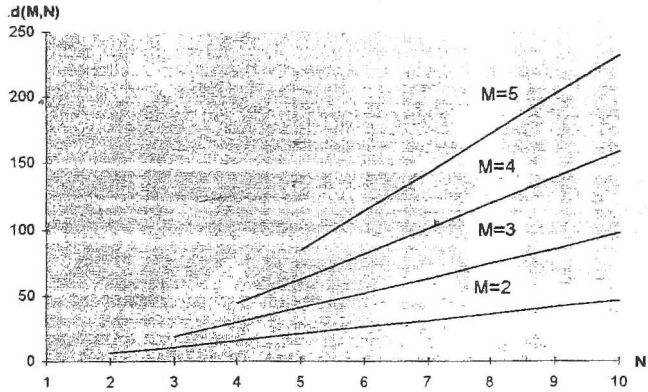


Рис. 5.5. Зависимости вычислительного выигрыша от числа наблюдений (N) *и размерности задачи* (M)

Из графиков видно, что алгоритм I всегда превосходит рекуррентный алгоритм вычисления оценок, если используется лишь одна (последняя) оценка. Он обычно оказывается экономичнее и в случае использования оценок на промежуточных шагах, если допустима их выдача с "прореживанием". В частности, алгоритм I практически не уступает рекуррентному при вычислении оценок через 2 такта при M=5 и через 3 такта при M=7. Для произвольной размерности задачи M (N>M) алгоритм I оказывается всегда более выгодным в вычислительном отношении, если оценку вычислять не чаще, чем через M тактов.

6. РЕАЛИЗАЦИЯ МЕТОДОВ И АЛГОРИТМОВ ИДЕНТИФИКАЦИИ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

В настоящей главе приведены результаты экспериментов по оценке эффективности разработанных методов, алгоритмов и построенных с их использованием новых информационных технологий идентификации моделей систем формирования изображений по малым фрагментам. Рассматриваются следующие задачи: оценивание параметров импульсной характеристики системы; идентификация инверсной модели; оценивание модели искажающей системы и корректирующих фильтров при пространственно-зависимых искажениях; идентификация моделей систем и корректирующих фильтров в классе БИХ-фильтров, оценка разрешающей способности видеотракта; совмещение изображений при сильно различающихся параметрах сдвига и поворота.

6.1. Идентификация импульсной характеристики искажающей системы

Задача идентификации импульсной характеристики искажающей системы является составной частью многих информационных технологий. В системах аэрокосмического мониторинга она решается с целью уточнения характеристик систем формирования изображений в ходе их эксплуатации. Для этого проводят испытания, заключающиеся в регистрации известного (тестового) изображения и последующей оценке параметров (идентификации) модели искажающей системы по сформированным входному (неискаженному) и выходному (искаженному) изображениям. В интересах вычислительной простоты и унификации процедур идентификации предпочтительно использовать линейные модели с постоянными по пространственным координатам параметрами.

В реальных системах искажения чаще всего пространственно-зависимые. Указанное противоречие преодолевается аппроксимацией искажений множеством кусочно-линейных и инвариантных к сдвигу моделей на малых фрагментах изображений. Соответствующая информационная технология будет описана в разделе 6.3. Одним из важнейших этапов этой технологии является оценка параметров импульсной характеристики на малых фрагментах изображения. В настоящем разделе исследуется эффективность решения этой изолированной задачи с использованием разработанных в настоящей работе методов и алгоритмов идентификации.

В разделе 1.2 приведены два основных способа описания импульсной характеристики системы. В настоящем разделе исследуются алгоритмы

идентификации простейшей линейной модели с постоянными параметрами: фильтр с конечной импульсной характеристикой (КИХ-фильтр), для которого связь отсчетов входного (неискаженного) - $x(n_1, n_2)$ и выходного (искаженного) $y(n_1, n_2)$ изображений описывается соотношением (1.5).

С использованием этой модели для совокупности отсчетов на некотором фрагменте изображения можно записать матричное соотношение типа (1.2):

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \xi. \qquad (6.1)$$

Здесь N×1 - вектор $\mathbf{y}$ составлен из N отсчетов $y(n_1, n_2)$, принадлежащих выбранному на выходном изображении фрагменту, N×M-матрица $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T$, где $\mathbf{x}_i$, i=1,N M×1 векторы, составленные из (M) отсчетов опорных областей входного (неискаженного) изображения, $\mathbf{h}$ - соответствующий этим отсчетам M×1-вектор параметров импульсной характеристики, а N×1- вектор ξ составлен из фигурирующих в (1.5) ошибок $\delta y(n_1, n_2)$. Здесь и далее, где речь идет об оценке импульсной характеристики, для вектора искомых параметров вместо ранее использовавшегося в 1.2 обозначения $\mathbf{c}$ применяется (принятое при описании импульсной характеристики) обозначение $\mathbf{h}$.

В первой главе приведена модифицированная МНК-оценка (1.11). С учетом введенного обозначения $\mathbf{h}$ для вектора параметров импульсной характеристики перепишем ее в виде:

$$\hat{\mathbf{h}} = \left[\mathbf{X}^T\mathbf{G}^2\mathbf{X}\right]^{-1}\mathbf{G}^2\mathbf{X}^T(\mathbf{y} + \Delta\mathbf{y}). \qquad (6.2)$$

Выше показано, что после предварительного контроля информативности данных эта оценка может быть вычислена по схеме последовательного преобразования данных с начальными значениями $\mathbf{G}(0) = E$, $\Delta\mathbf{y}(0) = 0$. В настоящем разделе приводятся результаты экспериментального исследования этих алгоритмов на тестовых изображениях.

Эксперименты ставились с целью:

- показать связь предложенных мер информативности (показателя диагонального преобладания Ф и сопряженности с нуль-пространством S_{min}) с характером объектов на регистрируемых изображениях;
- показать, что изменение (увеличение) этих мер информативности, как правило, сопровождается уменьшением погрешности оценивания;

- подтвердить эффективность модифицированной МНК-оценки (6.2) с последовательным преобразованием данных.

Процедуры контроля информативности и отбора данных по показателю диагонального преобладания исследовались в составе общего алгоритма идентификации модели линейных искажений. Для исследования использовались специально сформированные для этой цели тестовые изображения, показанные на рис. 6.1, а исходное, б - искаженное.

Искаженное изображение получено из исходного моделированием расфокусировки. Для этого использовалась модель расфокусировки с круговой (центральной) симметрией. В частности, параметры импульсной характеристики определялись в соответствии с выражением

$$h(i,j) = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(-r_{ij}^2/2\sigma^2\right), \qquad (6.3)$$

где $r_{ij}^2 = i^2 + j^2$, а i,j координаты отсчетов в опорной области. Задавался параметр расфокусировки σ=1.5. Указанное значение параметра σ при допустимой относительной погрешности (около 1%) соответствовало опорной области 9×9. Характер тестовых изображений выбирался из соображений иллюстрации возможности использования для идентификации оптического канала естественных элементов: дорожных разметок, пешеходных переходов и др.

На рис. 6.2 приведены поля значений для показателей диагонального преобладания Ф>0.1 (рис. 6.2, а) и минимальной сопряженности с нуль-пространством S_{min}>0.5 (рис. 6.2, б) по полю тестового изображения. Для удобства визуального восприятия значения Ф и S_{min}, превышающие указанные пороги, заменены максимальным значением в обычной [0-255] шкале яркостей. Эти бинарные изображения в дальнейшем использовались в ходе идентификации для отбора информативных данных.

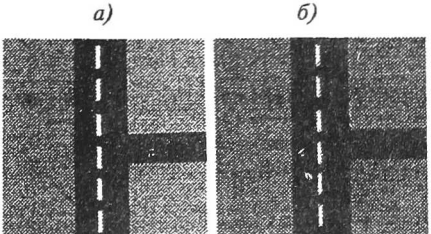


Рис. 6.1. Тестовые изображения: а - исходное; б - расфокусированное

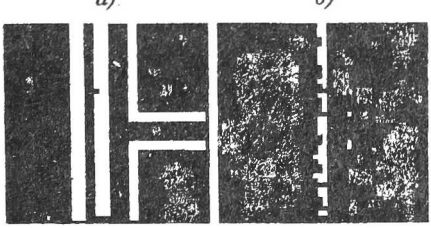


Рис. 6.2. Поля значений мер обусловленности: а - Ф; б - S_{min}

Параметрическая идентификация линейной модели КИХ-фильтра осуществлялась в соответствии с общей схемой алгоритма, показанного на рис. 5.1. В частности, для идентификации отбирались фрагменты изображений, для которых масштабированное значение показателя S_{min} превышало 0,5.

Выражения для расчета весовых коэффициентов зависели от номера итерации. На первой итерации вычислялись оценки МНК с одинаковыми весами равными единице. Весовые коэффициенты на второй и третьей итерациях вычислялись по соотношениям:

$$g_i = g_0\left(\hat{s} + \left|\hat{\xi}\right|\right)^{-1}, \tag{6.4}$$

$$g_i = g_0\left(\hat{s}^2 + \hat{\xi}^2\right)^{-1}. \tag{6.5}$$

Эти классы преобразований подробно рассматривались в разделе 4. Вычисленные в соответствии с соотношениями (6.4), (6.5) весовые коэффициенты подвергались нормировке так, чтобы выполнялось условие: $\sum|g_i| = N, \quad i = \overline{1,N}$.

Для определения параметра расфокусировки σ по оценкам вектора параметров импульсной характеристики решалась задача аппроксимации с помощью этого же алгоритма оценивания. В таблице 6.1 приведены сравнительные данные, показывающие точность оценивания параметра расфокусировки σ при идентификации с использованием построенного алгоритма и обычного МНК. Большое преимущество метода взвешивания в данном случае объясняется тем, что в исходные данные наряду с нормальными помехами, относительная интенсивность которых составляла около 3%, добавлялись (менее 5%) сильно выделяющиеся импульсные помехи.

Таблица 6.1. Сравнение точности МНК-оценок и метода взвешивания

Метод идентификации	$\lvert\Delta\sigma\rvert/\sigma$
Одношаговый МНК	32.33
Трехшаговая процедура с весами (6.4,(6.5)	0.034

В следующей серии экспериментов, кроме прочего, анализировалась также связь показателя диагонального преобладания с погрешностью идентификации. Для этих экспериментов использовалось тестовое изображение, показанное на рис. 6.3. При выборе тестового изображения учитывалась возможность оценивания характеристик реальной системы без использования специально построенной для этой цели миры, например, по крышам зданий и сооружений, элементам аэродромной разметки и др.

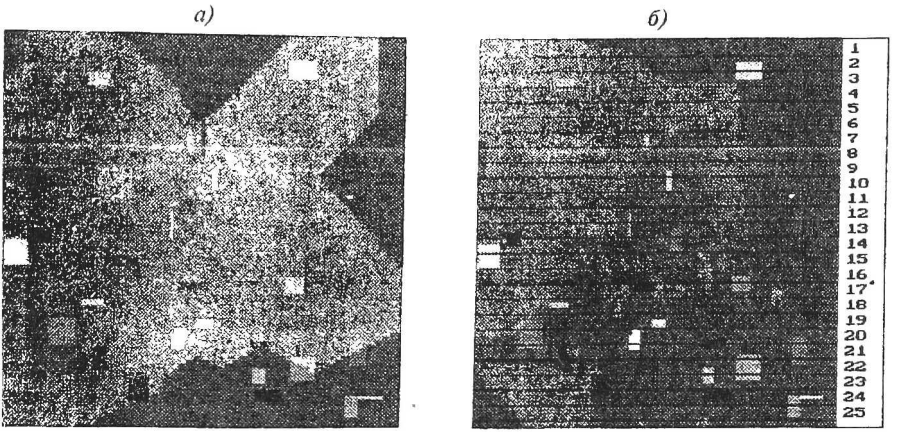


Рис.6.3. Тестовые изображения: а - исходное; б - с разбивкой на фрагменты

Исходя из этих соображений, входное (неискажённое) изображение формировалось в виде прямоугольников на постоянном фоне. Размеры сторон прямоугольников и их яркости задавались в виде случайных равномерно распределенных чисел в диапазонах 1-10 и 0-255 соответственно. Общее количество элементов (прямоугольников) задавалось равным 40. Они располагались на поле изображения случайным образом: равномерно по горизонтали и с возрастающей интенсивностью вниз по вертикали (с ростом порядкового номера фрагмента).

Выходное изображение формировалось "размазыванием" входного с использованием разностного уравнения вида:

$$y(n_1, n_2) = \frac{1}{4} \sum_{k=0}^{2} h(k)x(k), \qquad (6.6)$$

где $x(k) = \left[x(n_1 - k, n_2) + x(n_1 + k, n_2) + x(n_1, n_2 - k) + x(n_1, n_2 + k)\right]$. Уравнение (6.6) соответствует опорной области, показанной на рис. 6.4. Для исследования поведения алгоритма при различной степени искажений моделировалось три искажённых изображения со значениями импульсной характеристики искажающего фильтра соответственно:

вариант 1: $h(0)=0.85$, $h(1)=0.1$, $h(2)=0.05$;

вариант 2:- $h(0)=0.7$, $h(1)=0.2$, $h(2)=0.1$;

вариант 3:- $h(0)=0.5$, $h(1)=0.3$, $h(2)=0.2$.

Во всех трёх случаях для нас имело значение лишь то, что импульсный отклик по осям N_1, N_2 опорной области одинаков, хотя частотная характеристика и не обладает круговой симметрией.

Обрабатывавшиеся в каждом эпизоде идентификации фрагменты изображения имели вид полосок 5x128, показанных на рис. 6.3, б. При этом матрица **X** для каждого фрагмента имела размерность 25x3, а соответствующая ей матрица **A** 3x3. Искомый вектор оценок $\hat{\mathbf{h}} = \left[\hat{h}(0), \hat{h}(1), \hat{h}(2)\right]^T$, так же как и в предыдущем эксперименте, определялся с использованием описанного выше алгоритма (рис. 5.1). Результаты экспериментов сведены в таблицу 6.2. В частности, приведены относительные среднеквадратичные ошибки (СКО) оценивания для каждой из трёх указанных выше моделей искажений.

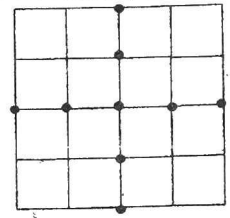


Рис.6.4. Опорная область

Таблица 6.2. Результаты анализа информативности данных и погрешность МНК-оценок

№ фрагмента	$\Phi(\mathbf{R})$	$\Phi(\mathbf{A})$	$\Phi_{min}(\mathbf{A})$	Относительные среднеквадратичные ошибки (СКО) на фрагментах		
				Вариант 1	Вариант 2	Вариант 3
1	1.000	1.000	1.000			
2	1.735	1.005	1.000			
3	1.036	1.027	1.011	0.000005	0.000002	0.000014
4	1.760	1.014	1.000			
5	1.647	1.014	1.007			
6	1.000	1.000	1.000			
7	1.771	1.087	1.000			
8	1.000	1.000	1.000			
9	2.136	1.093	1.066	0.000006	0.000001	0.000002
10	1.013	1.012	1.005	0.000014	0.000011	0.000011
11	2.053	1.220	1.015	0.000001	0.000001	0.000001
12	1.041	1.040	1.011	0.000004	0.000001	0.000003
13	2.742	1.050	1.039	0.000002	0.000003	0.000002
14	1.076	1.069	1.027	0.000001	0.000008	0.000001
15	1.023	1.024	1.003	0.000030	0.000035	0.000002
16	1.780	1.015	1.000			
17	1.091	1.034	1.016	0.000008	0.000002	0.000008
18	1.051	1.022	1.016	0.000004	0.000045	0.000009
19	1.679	1.251	1.033	0.000017	0.000013	0.000018
20	1.263	1.093	1.001	0.000178	0.000016	0.000219
21	1.147	1.134	1.065	0.000001	0.000005	0.000004
22	1.132	1.132	1.095	0.000001	0.000001	0.000001
23	1.043	1.032	1.014	0.000005	0.000003	0.000006
24	1.067	1.064	1.055	0.000004	0.000003	0.000002
25	1.000	1.000	1.000			

В первых двух (после номера фрагмента) столбцах таблиц приведены показатели диагонального преобладания Ф(R) и Ф(A). В следующем столбце приведена еще одна характеристика матрицы Грама Φ_{min}. Вычисляется она следующим образом. Вначале показатели Ф(A), Ф(R) подсчитываются для всех главных миноров матриц A и/или R, а затем из них выбирается минимальное значение. Эта характеристика более надежно обнаруживает факт плохой обусловленности.

Незаполненные строки в таблице соответствуют случаям аварийного завершения программы. Отметим, что это произошло как раз на фрагментах, содержащих небольшое число (либо не содержащих вовсе) информативных элементов - прямоугольников (рис. 6.3, б). На этих фрагментах показатель Ф(A) (а в некоторых случаях и Ф(R)), как и следовало ожидать, оказался близким к единице, что свидетельствует о плохой обусловленности задачи.

Из таблицы видно, что, несмотря на изменение параметров модели искажающего фильтра, показатели диагонального преобладания Ф(A) и Ф(R) на одноимённых фрагментах принимают одинаковые значения. Это подтверждает, что показатель диагонального преобладания характеризует лишь информативные свойства фрагментов изображений, зависящие от характера расположения, размеров и интенсивности объектов на поле изображения.

Для иллюстрации эффективности алгоритма последовательного МНК-оценивания по сравнению с известными процедурами, по той же схеме проводилось оценивание с использованием известного метода регуляризации [16,17]. Результаты приведены в таблице 6.3. Содержание колонок этой таблицы такое же, как и в таблице 6.2.

Из таблиц 6.2, 6.3 видно, что по мере возрастания интенсивности объектов на тестовом изображении (с увеличением номера фрагмента) значения показателей Ф(A), Ф(R) также, в основном, возрастают, а погрешность идентификации уменьшается. Сравнивая их с характером расположения объектов на соответствующих фрагментах тестового изображения (рис. 6.3, б) легко заметить, что значения показателей равны или близки к единице именно на фрагментах, где размеры объектов малы или вовсе отсутствуют.

Из этого правила, конечно, могут быть исключения. Ведь указанные меры обусловленности строились с использованием простых достаточных оценок. В частности, на фрагменте №7 (таблица 6.3), несмотря на сравнительно высокое значение меры мультиколлинеарности Ф(R)=1.771, относительная СКО

составила 0.619, 0.481, 0.229 для вариантов задания параметров моделеи № 1, 2 и 3 соответственно. Этот факт имеет простое объяснение. На тестовом изображении видно, что на этом фрагменте имеется лишь один объект, что и явилось причиной недостаточной информативности. Заметим, что показатель $\Phi_{min}(A)$ зафиксировал факт вырожденности матрицы Грама на этом фрагменте. Показатель Ф(А) не "чувствует" этого, так как определяется для всей матрицы с размерами MxM.

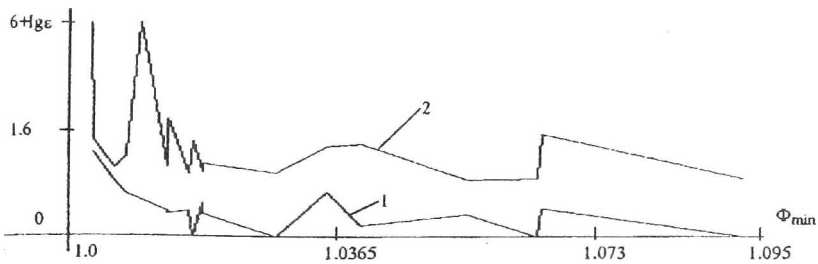
Таблица 6.3. Оценки информативности и погрешность идентификации при использовании алгоритма с регуляризацией

№ фрагмента	Ф(R)	Ф(А)	$\Phi_{min}(A)$	Относительные СКО оценок параметров на фрагментах		
				Вариант 1	Вариант 2	Вариант 3
1	1.000	1.000	1.000			
2	1.735	1.005	1.000	0.041242	0.096226	0.114709
3	1.036	1.027	1.011	0.000075	0.000079	0.000121
4	1.760	1.014	1.000	0.041241	0.096227	0.114710
5	1.647	1.014	1.007	0.659829	0.577350	0.344124
6	1.000	1.000	1.000			
7	1.771	1.087	1.000	0.618590	0.481125	0.229416
8	1.000	1.000	1.000			
9	2.136	1.093	1.066	0.000689	0.000685	0.000728
10	1.013	1.012	1.005	0.000144	0.000162	0.000172
11	2.053	1.220	1.015	0.000389	0.000545	0.000657
12	1.041	1.040	1.011	0.001550	0.001394	0.001152
13	2.742	1.050	1.039	0.000358	0.000402	0.000487
14	1.076	1.069	1.027	0.000056	0.000059	0.000083
15	1.023	1.024	1.003	0.000069	0.000114	0.000167
16	1.780	1.015	1.000	0.618590	0.481125	0.229416
17	1.091	1.034	1.016	0.000056	0.000092	0.000098
18	1.051	1.022	1.016	0.000094	0.000085	0.000150
19	1.679	1.251	1.033	0.000308	0.000357	0.000438
20	1.263	1.093	1.001	0.000396	0.000421	0.000419
21	1.147	1.134	1.065	0.000040	0.000037	0.000024
22	1.132	1.132	1.095	0.000040	0.000043	0.000042
23	1.043	1.032	1.014	0.000050	0.000052	0.000064
24	1.067	1.064	1.055	0.000038	0.000041	0.000059
25	1.000	1.000	1.000			

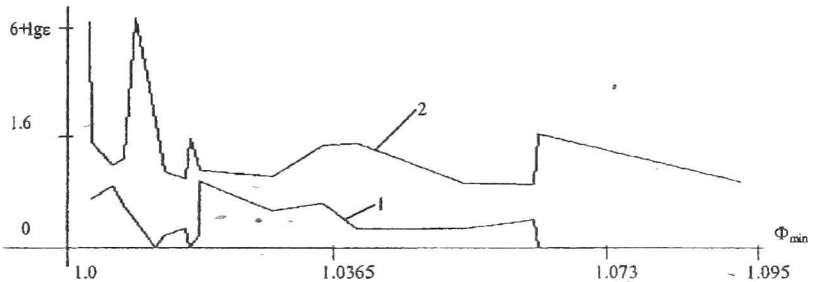
На рис. 6.5, а, б и в по данным таблиц (кривые 1 по таблице 6.2, кривые 2 по таблице 6.3) построены графики, показывающие связь относительной среднеквадратичной ошибки идентификации с показателем диагонального преобладания Φ_{min}. Легко заметить, что с ростом величины Φ_{min} относительная СКО независимо от метода оценивания имеет тенденцию к убыванию для всех

трех указанных выше вариантов задания параметров модели искажений. Следовательно, показатель Φ_{min} можно рекомендовать в качестве надежного индикатора для определения факта вырожденности задачи.

а)



б)



в)

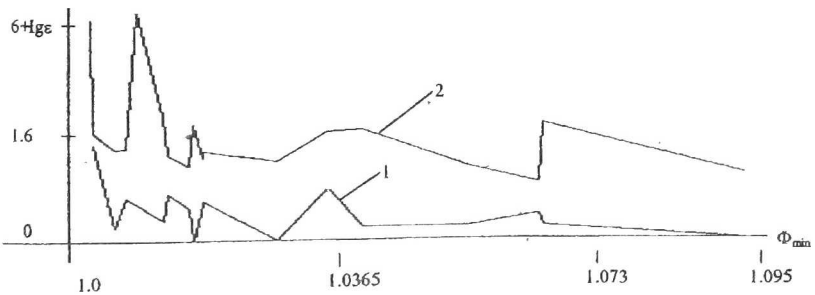


Рис. 6.5. Связь погрешности идентификации с показателем Φ_{min} для различных вариантов параметров модели: а - № 1; б - № 2; в - № 3

Сравнение кривых 1 и 2 на всех трех графиках показывает, что при использовании трехшагового алгоритма с промежуточными преобразованиями взвешивания (6.4), (6.5) при одинаковом значении Φ_{min} всегда достигается более

высокая точность оценивания. Этот факт имеет простое объяснение. Используемый в настоящей работе подход, по существу, является адаптивным и позволяет более полно извлекать информацию об искомых параметрах из фиксированного набора данных. Достигается это за счет использования дополнительной информации, содержащейся в невязках.

Вместе с тем, хотя и имеет место тенденция к возрастанию точности при увеличении диагонального преобладания, зависимости (1 и 2) от Φ_{min} немонотонные. Показатель диагонального преобладания является надежным индикатором для плохо и хорошо обусловленных задач. В диапазоне значений показателя, определяемых неравенствами $1<\phi<M-1$, неопределенность в оценке степени обусловленности существенно выше. В этой ситуации, если позволяют вычислительные ресурсы, должны использоваться показатели сопряженности векторов независимых переменных. Ниже приводятся результаты экспериментов, связанных с исследованием показателя минимальной сопряженности с нуль-пространством - S_{min}, который описывался в разделе 2.5.

В качестве тестового использовался участок изображения типа "текст", показанный на рисунке 6.6. Для наглядной визуальной оценки связи показателя S_{min} с информативностью исходных данных он рассчитывался в каждой точке исходного изображения. Результаты расчетов представлены в виде изображения на рисунке 6.7. Для удобства визуального восприятия диапазон изменения показателя [0,1] приведен к интервалу [0,255]. На рис. 6.7 видно, что показатель S_{min} "безошибочно замечает" строки и слова, т.е. наиболее информативные для идентификации, участки изображения.

Исследовались также количественные оценки связи погрешности идентификации с величиной показателя минимальной сопряженности с нуль-пространством. Для проведения этих исследований наряду с тестовым изображением (рис. 6.6) использовалось изображение, полученное из него "расфокусировкой". При моделировании этого изображения использовалось соотношение (6.3) с параметром расфокусировки $\sigma=1.5$. Соответствующий пространственно-инвариантный КИХ-фильтр имел опорную область 9×9. Расфокусированное изображение показано на рис. 6.8.

Считая изображение, показанное на рис. 6.6 входным, а изображение на рис. 6.8 - выходным, проводилась идентификация на малых фрагментах изображений с оценкой их информативности по показателям S_{min} и S_{Σ}. Для получения количественных оценок связи погрешности идентификации со значениями пока-

зателей S_{min}, S_Σ на искаженное тестовое изображение (рис. 6.8) дополнительно накладывался аддитивный белый гауссовский шум при отношениях шум/сигнал 0.001, 0.01, 0.03 и 0.05. Изображение разбивалось на фрагменты 6x6. На каждом фрагменте определялся вектор оценок параметров импульсной характеристики и вычислялись относительные погрешности идентификации. Полученные на всех фрагментах значения показателей S_{min}, S_Σ были выстроены в вариационный ряд и разбиты на группы по 20 значений в каждой группе. Затем показатели S_{min}, S_Σ и соответствующие им относительные погрешности идентификации усреднялись для каждой группы отдельно. Полученные описанным способом результаты представлены в виде графиков на рис. 6.9, 6.10.

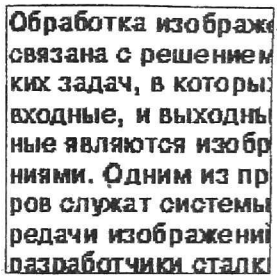


Рис. 6.6. Исходное. изображение

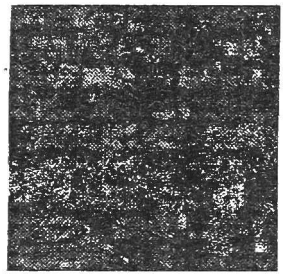


Рис. 6.7. Поле показателя S_{min}

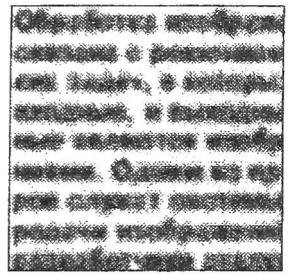


Рис. 6.8. Расфокусированное изображение

На графиках видна устойчивая тенденция к снижению погрешности идентификации по мере увеличения показателей S_{min} и S_Σ. Эти зависимости, конечно, немонотонные, поскольку ориентация вектора ошибок в эксперименте не фиксировалась. Для обоснования пороговых значений этих показателей, при которых гарантируется заданная точность оценивания, в эксперименте необходимо отобрать результаты, полученные при $\xi \in \mathcal{R}(A)$. Подчеркнем, что большие значения показателя S_Σ не могут гарантировать невырожденность задачи, поэтому для отбора информативных фрагментов рекомендуется использовать показатель S_{min} .

Исследовалась также трехшаговая итерационная процедура с преобразованиями взвешивания по соотношениям (6.4) и (6.5) при различных распределениях помех. Подчеркнем, что здесь, как и всюду, где речь идет о распределениях, имеется в виду исходная совокупность отсчетов на всем изображении. Ясно, что на малых фрагментах этого изображения распределения могут существенно отличаться от исходного, использовавшегося на этапе моде-

лирования. Цель эксперимента как раз и заключается в том, что бы убедиться в работоспособности трехшаговой схемы взвешивания по невязкам в условиях статистической неустойчивости распределений помех.

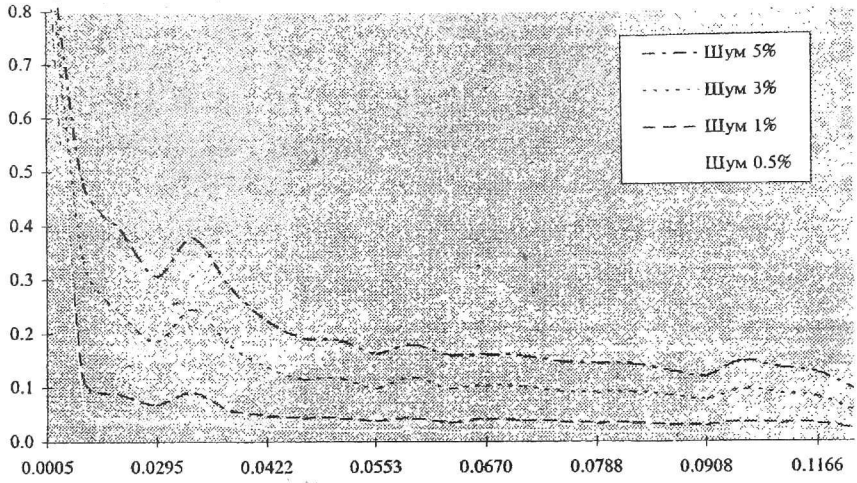


Рис. 6.9. Связь погрешности идентификации с показателем S_{min}

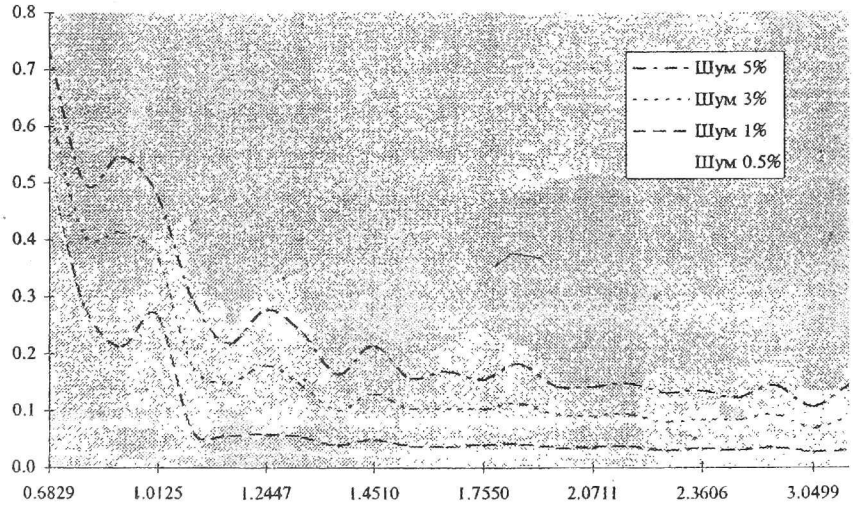


Рис.6.10. Связь погрешности идентификации с показателем S_Σ

Эксперимент проводился на тестовых изображениях, показанных на рисунке 3.3. Использовались два типа допустимых параметрических семейств функций

(№1 и №2 из таблицы 4.3). Поскольку параметры модели искажений при моделировании известны, результаты оценивания сравнивались с истинными значениями параметров.

В таблице 6.4 приведены усредненные по всем использовавшимся фрагментам на изображении относительные погрешности оценивания (в процентах) при указанных распределениях ошибок и различных значениях параметра масштаба -

Таблица 6.4. Точность оценивания с использованием процедуры трехшагового взвешивании по невязкам при различных распределениях помех

Исходное распределение № 1											
s	0.01	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
ε_1	0.0057	0.213	0.451	0.806	0.870	1.118	1.312	1.816	2.015	2.224	2.273
ε_3	0.0057	0.214	0.453	0.817	0.874	1.126	1.320	1.841	2.035	2.247	2.306
Исходное распределение № 2											
s	0.01	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
ε_1	0.0091	0.262	0.601	0.802	1.000	1.275	1.349	1.645	1.706	2.147	2.274
ε_3	0.0093	0.263	0.597	0.793	1.000	1.271	1.344	1.631	1.709	2.173	2.294
Исходное распределение № 3											
s	0.01	0.25	0.5	0.75	1.0	1.25	1.5	1.75	2.0	2.25	2.5
ε_1	0.0282	0.693	1.376	2.026	2.590	3.300	3.804	4.618	5.355	5.767	7.199
ε_3	0.0220	0.541	1.042	1.571	1.970	2.552	2.771	3.573	4.119	4.454	5.492

s. Для каждого распределения приведена относительная погрешность ε_1 начальных МНК-оценок и относительная погрешность ε_3 оценок после третьего шага взвешивания данных весовыми коэффициентами (6.4), (6.5). Из таблицы видно, что выигрыш в точности оценивания в среднем больше, если у исходного распределения более тяжелые хвосты.

6.2. Построение восстанавливающих фильтров путем непосредственной идентификации инверсной модели видеотракта

Проблема восстановления изображений, по существу, сводится к задаче построения обратного оператора искажающей системы. При этом предполагается, что оператор W(·), действующий на входную последовательность $x(n_1, n_2)$, задан. Часто оператор W(.) определяют путем предварительного решения задачи идентификации характеристик искажающей системы по специально подобранным для этой цели тестовым изображениям.

Если задачи идентификации и восстановления решаются изолированно, то усилия, направленные на определение, по возможности, более точной модели искажающей системы на этапе идентификации, могут оказаться, в значительной степени, напрасными. Дело в том, что на этапе восстановления изображений обычно все равно приходится в той или иной степени (в зависимости от свойств оператора) осуществлять регуляризацию решений, по существу, означающую намеренное искажение обратного оператора с целью уменьшения чувствительности к ошибкам в исходных данных. Поэтому представляется целесообразным задачи идентификации моделей искажающих систем и восстановления изображений с самого начала рассматривать с учетом их взаимосвязи.

В качестве модели искажающей системы будем по-прежнему использовать КИХ-фильтр (1.5), которому на фрагменте, содержащем N отсчетов, соответствует уравнение вида (6.1). По аналогии с (1.5) модель инверсного тракта (если она существует) можно представить в виде

$$x(n_1, n_2) = \sum_{\{k_1,k_2\}\in R} h_r(k_1, k_2)y(n_1 - k_1, n_2 - k_2) + \psi(n_1, n_2). \qquad (6.7)$$

Для L опорных областей, каждая из которых включает S отсчетов выходного (искаженного) изображения, по аналогии с (6.1), можно записать матричное соотношение

$$\mathbf{x} = \mathbf{Y}\mathbf{h}_r + \psi, \qquad (6.8)$$

где $\mathbf{x}$, - вектор Lx1, $\mathbf{Y}$ -матрица LxS, а $\mathbf{h}_r$ - вектор Sx1.

Восстанавливающий фильтр будем строить также в классе КИХ - фильтров:

$$\hat{x}(n_1, n_2) = \sum_{\{k_1,k_2\}\in R} \hat{h}_r(k_1, k_2)y(n_1 - k_1, n_2 - k_2). \qquad (6.9)$$

В матричном виде

$$\hat{\mathbf{x}} = \mathbf{Y}\hat{\mathbf{h}}_r, \qquad (6.10)$$

где $\mathbf{Y}$ - фигурирующая в (6.8) матрица размерности LxS, $\hat{\mathbf{h}}_r^T$ - вектор оценок коэффициентов описывающего инверсный фильтр разностного уравнения размерности - Sx1. Компоненты Lx1 вектора $\hat{\mathbf{x}}$ - суть оценки значений поля яркости восстановленного изображения. Они содержат ошибки, связанные с наличием ошибок $\delta y(n_1, n_2)$ в исходных данных, ограничением порядка КИХ-фильтра и погрешностями оценивания его параметров.

При определении вектора $\hat{\mathbf{h}}_r$ есть, по крайней мере, две возможности. Во-первых, можно по вектору $\mathbf{y}$ и матрице $\mathbf{X}$, фигурирующим в (6.1), найти оценку

$\hat{\mathbf{h}}$ вектора $\mathbf{h}$ модели искажающей системы, а затем одним из известных способов построить вектор $\hat{\mathbf{h}}_r$ восстанавливающего фильтра. Во-вторых, можно (поменяв изображения местами) по $\mathbf{x}$ и $\mathbf{Y}$ сразу найти оценку $\hat{\mathbf{h}}_r$ - параметров инверсного фильтра (6.10), решив соответствующую задачу идентификации. В настоящем разделе исследуется второй подход, основанный на непосредственной идентификации характеристик инверсного тракта.

Задача идентификации инверсной модели видеотракта имеет некоторые специфические, по сравнению с оценкой модели прямого тракта, особенности. Если оцениваются непосредственно параметры инверсного фильтра, МНК-оценка $\hat{\mathbf{h}}_r$ вектора параметров $\mathbf{h}_r$ инверсного тракта по аналогии с (6.2) определяется как

$$\hat{\mathbf{h}}_r = \left[\mathbf{Y}^T\mathbf{G}^2\mathbf{Y}\right]^{-1}\mathbf{Y}^T\mathbf{G}^2(\mathbf{x}+\Delta\mathbf{x}). \qquad (6.11)$$

В (6.11) матрица $\mathbf{Y}$, в отличие от фигурирующей в (6.2) матрицы $\mathbf{X}$, составлена из искаженных и зашумленных отсчетов изображения. Поэтому соответствующая (при $\mathbf{G}^2 = \mathbf{E}$) матрица $\left[\mathbf{Y}^T\mathbf{Y}\right]$ может оказаться невырожденной даже в ситуациях, когда исходное неискаженное и незашумленное изображение не содержит никаких информативных элементов, т.е. отсутствуют объекты на фоне с одинаковыми значениями двумерной функции яркости). Результаты оценивания параметров инверсного тракта на таких фрагментах изображения могут ввести в заблуждение.

Вторая важная особенность заключается в следующем. Для задачи (6.2) естественным является требование достижения наивысшей точности определения $\hat{\mathbf{h}}$ в смысле близости к $\mathbf{h}$. В случае идентификации параметров восстанавливающего фильтра требование определения точной инверсной модели не всегда оправдано. Например, начальная (при $\mathbf{G}^2 = \mathbf{E}$, $\Delta\mathbf{x} = 0$) оценка (6.11) является наилучшей в смысле критерия:

$$\hat{\mathbf{h}}_r: \quad Q(\hat{\mathbf{h}}_r) = \min_{\forall h_r} Q(\mathbf{h}_r) = \min_{\forall h_r}\left(\left[\mathbf{x}-\mathbf{Y}\mathbf{h}_r\right]^T\left[\mathbf{x}-\mathbf{Y}\mathbf{h}_r\right]\right). \qquad (6.12)$$

Поскольку $\hat{\mathbf{h}}_r$ ищется в условиях присутствия в элементах матрицы $\mathbf{Y}$ ошибок измерений, то эта оценка фактически оказывается регуляризованной в среднеквадратическом смысле. Конечно, как показывают исследования [30], результаты восстановления наилучшим в среднеквадратическом смысле инверсным фильтром субъективно не всегда воспринимаются как лучшие. Поэтому представляет

интерес построение процедур идентификации, позволяющих достаточно просто реализовывать различную степень близости к истинным значениям параметров модели, в т.ч. и по субъективным оценкам качества восстановления.

Для этого воспользуемся методом взвешивания данных. В данном случае, в отличие от (1.15), взвешиванию должны подвергаться матрица **Y** и вектор **x**:

$$\tilde{\mathbf{Y}} = \mathbf{GY}, \quad \tilde{\mathbf{x}} = \mathbf{Gx}, \tag{6.13}$$

при этом $\tilde{\psi} = \mathbf{G}\psi$. Существо рассматриваемого подхода к построению восстанавливающих фильтров заключается в изменения свойств оценок путем непосредственного изменения матрицы преобразования **G**, которая, в свою очередь, зависти от одного параметра. При таком подходе изменением одного параметра можно добиваться различного регуляризирующего эффекта. Это особенно удобно в случае, когда имеется возможность непосредственного визуального контроля качества восстановления. Проиллюстрируем это на тестовых изображениях.

Для придания указанных свойств использовался метод построения диагональной матрицы **G** преобразования (6.13), основанный на связи с $\mathbf{L}_\nu$-оценками. В частности, использовалось семейство критериев типа (4.53):

$$Q(\mathbf{c},s) = \sum_{i=1}^{N} g_0 \left|\hat{\xi}_i\right|^{s_k} \tag{6.14}$$

где s_k, $k = \overline{1,3}$ параметр, обеспечивающий различную степень близости оценок $\hat{\mathbf{h}}_r$ к истинным параметрам $\mathbf{h}_r$.

Это равносильно построению трех моделей по критерию

$$Q(\mathbf{h}_r) = \left(\left[\tilde{\mathbf{x}} - \tilde{\mathbf{Y}}\mathbf{h}_r\right]^T\left[\tilde{\mathbf{x}} - \tilde{\mathbf{Y}}\mathbf{h}_r\right]\right), \tag{6.15}$$

в котором матрица преобразования **G**, определяющая в (6.13) преобразованные данные $\tilde{\mathbf{Y}}$ и $\tilde{\mathbf{x}}$, изменяется от итерации к итерации так, что при определении моделей на завершающих этапах критерий оказывается "менее регуляризирующим", т.е. обеспечивающим получение более точной инверсной модели.

Исследовалась процедура, предусматривающая построение набора из трех моделей инверсного тракта с помощью последовательности трех оценок с изменяющимися от шага к шагу значениями параметра s_k: $s_1 = 2$, $s_2 = 1$, и $s_3 \to 0$. Подчеркнем, что, несмотря на внешнее сходство, соответствующие оценки не являются L_ν- оценками в строгом значении этого термина [15].

На рис. 6.11 приведены тестовые фрагменты, использовавшиеся для непосредственной идентификации восстанавливающего фильтра, а на рисунке 6.12, в - результаты воеетановления изображения, подвергшегося таким же искажениям, как и тестовые фрагменты. Для идентификации использовалась основанная на критерии (6.19) трехшаговая процедура с весами

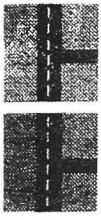


Рис. 6.11. Тестовые фрагменты

а) *б)* *в)*

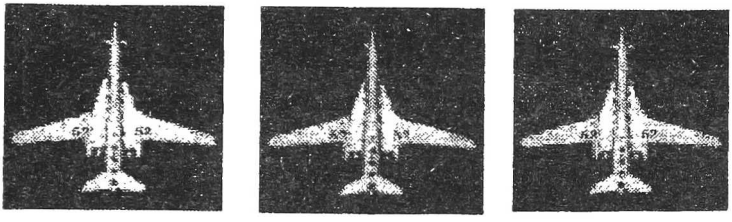


Рис. 6.12. Результаты восстановления фильтрами, полученными путем их непосредственной идентификации: а - исходное изображение; б - искаженное изображение; в - восстановленное изображение

$$g_i(k) = g_0 \left|\hat{\xi}_i\right|^{-s_i(k)},$$

где $s_i(k) = 0$ при всех $k=1,3$ и $i < q$,

q: $\left|\hat{\xi}_q\right| \le \varepsilon_\alpha > 0$, $\left|\hat{\xi}_1\right| \le \left|\hat{\xi}_2\right| \le \ldots \le \left|\hat{\xi}_{q-1}\right| \le \left|\hat{\xi}_q\right| \le \left|\hat{\xi}_{q+1}\right| \le \ldots \le \left|\hat{\xi}_N\right|$,

а для остальных компонент

$$s_i(1) = 0,\ s_i(2) = 1,\ s_i(3) = 2.$$

Параметр g_0 определялся из условия нормировки:

$$g_0 = N \Big/ \sum_{i=1}^{N} \left|\hat{\xi}_i\right|^{-s_i(k)}.$$

При моделировании искажений тестовых фрагментов и изображения "самолет" использовалась параметрическая модель (6.3) с σ=1.5. Соответствующая указанному значению параметра расфокусировки опорная область имеет размеры 9×9. Даже с учетом центральной симметрии искажений число искомых параметров при этом равно 15. Решение задачи идентификации такой размерности объективно требует увеличения размеров используемых для идентификации фрагментов изображений.

Указанная трудность преодолевается применением итерационной процедуры восстановления изображений с использованием набора инверсных КИХ-фильтров невысокого порядка. Общая схема формирования совокупности моде-

лей следующая. По исходным тестовым изображениям идентифицируется модель инверсного тракта невысокого порядка. С помощью полученной модели восстанавливающего фильтра осуществляется обработка искаженного тестового изображения. Затем исходное искаженное тестовое изображение заменяется обработанным и вновь решается задача идентификации. В результате определяется новый КИХ-фильтр невысокого порядка и т.д. Процесс формирования множества моделей останавливается, если очередной шаг не дает существенного улучшения качества восстановления тестового изображения.

Задача восстановления других изображений решается путем последовательной их обработки полученным набором инверсных фильтров в той же последовательности как они были оценены. Если опорная область квадрат со стороной N отсчетов, то каждые два соседних шага обработки формально соответствуют одному шагу обработки фильтром с опорной областью 2N-1. Действительно, подставляя выражение для каждого отсчета обработанного тестового изображения (6.9) в аналогичное соотношение для следующего шага обработки, получаем

$$\hat{x}^{(i+1)}(n_1,n_2) = \sum_{k_1,k_2=\frac{1-N}{2}}^{\frac{N-1}{2}} \hat{h}_r^{(i+1)}(k_1,k_2) \sum_{l_1,l_2=\frac{1-N}{2}}^{\frac{N-1}{2}} \hat{h}_r^{(i)}(l_1,l_2) y(n_1-k_1-l_1, n_2-k_2-l_2) =$$

$$= \sum_{m_1,m_{22}=1-N}^{N-1} \hat{h}_r^{i,i+1}(m_1,m_2) y(n_1-m_{11}, n_2-m_2),$$

где $\quad m_1 = k_1 + l_1, \quad m_2 = k_2 + l_2,$

а $$\hat{h}_r^{i,i+1}(m_1,m_2) = \sum \sum_{\substack{k_1,k_2,l_1,l_2=\frac{1-N}{2}, \\ k_1+l_1=m_1, \\ k_2+l_2=m_2}}^{\frac{N-1}{2}} \hat{h}_r^{(i+1)}(k_1,k_2)\hat{h}_r^{(i)}(l_1,l_2).$$

Для реализации указанной итерационной схемы восстановления изображений даже при сравнительно больших размерах размазывающих масок может применяться набор КИХ-фильтров фиксированного и притом невысокого порядка.

При проведении экспериментов для формирования набора моделей инверсных фильтров использовались тестовые изображения 128×128 с диапазоном яркости 0-256, показанные на рисунках 6.13 (а, б и в). Линейные искажения формировались трехкратным "проходом" КИХ-фильтром, обладающим радиальной симметрией с опорной областью 5×5 (без угловых отсчетов), показанной на рис 6.14.

Компоненты вектора, характеризующего импульсный отклик системы, задавались следующими: h_0=0,359118, h_1=0,39503, h_2=0,19033, h_3=0,04991, h_4=0,00560.

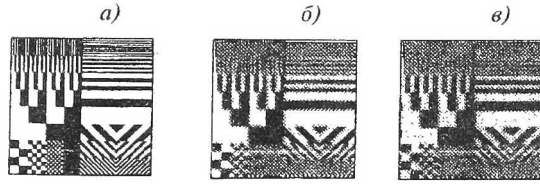


Рис. 6.13. Тестовые изображения: а - неискаженное; б - искаженное; в - искаженное и зашумленное

(корни соответствующего разностного уравнения: -0.4; -0.3; =0.2±j0.3). Аддитивный шум моделировался в виде псевдослучайной последовательности с дисперсией 36. По приведенным тестовым изображениям строился набор из трех инверсных восстанавливающих КИХ-фильтров с такой же опорной областью.

Для анализа обусловленности и отбора информативных данных в данном случае использовалась (неискаженная и незашумленная) матрица **X**. Как уже указывалось выше, показатели информативности матрицы **Y** могут ввести в заблуждение. Операции по отбраковке данных в любом случае должны осуществляться с соответствующими блоками и/или строками матрицы **Y** (искаженного изображения) так, как это описано в разделе 5.4.

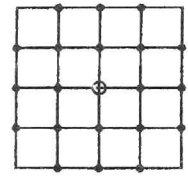


Рис. 6.14. Опорная область

Подлежавшее восстановлению изображение "часы" (размером 256×256) приведено на рисунке 6.15, а. Для его искажения и зашумления использовалась та же, что и для тестового изображения модель (различались лишь генерирующие числа псевдослучайной последовательности шумов). В частности, средний квадрат разности исходного и расфокусированного изображений ($\sigma_\xi^2 = 0$) $\sigma_{\Delta x}^2 = 464$, а при добавлении шумов ($\sigma_\xi^2 = 36$) $\sigma_{\Delta x}^2 = 500$.

На рисунке 6.15, б приведено изображение, полученное путем последовательного применения к показанному на рисунке 6.15, а искаженному и зашумленному изображению набора из трех КИХ-фильтров, построенных по методу наименьших квадратов, а на рисунке 6.15, в - полученное с использованием такого же набора фильтров, оценивавшихся на основе критериев (6.14), (6.15) с изменяющимися от шага к шагу значениями параметра s_k: $s_1 = 2$, $s_2 = 1$, и $s_3 \to 0$. Здесь отчетливо наблюдается эффект подчеркивания шумов инверсным фильтром, хотя контуры и элементы восстановленного изображения воспринимаются более четкими.

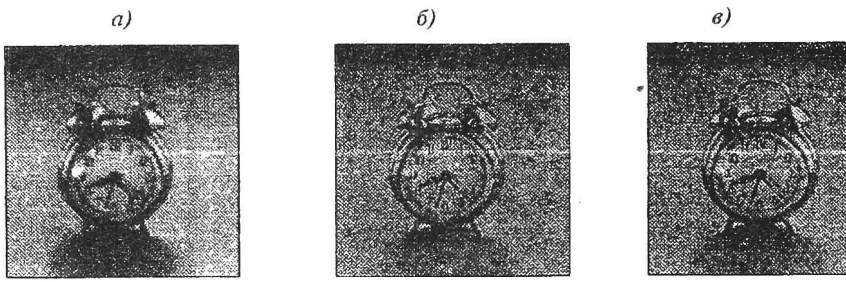


Рис. 6.15. Результаты восстановления изображения "часы" (а) - с использованием фильтра, полученного оцениванием по критерию: б - МНК; в - (6.14), (6.15)

В таблице 6.5 для исследовавшихся методов приведены оценки параметров инверсных фильтров и средние значения квадратов разности исходного искаженного и восстановленных изображений (в последней колонке). Для оценок по критериям (6.14), (6.15) данные, полученные на первом шаге, в таблице опущены, т.к. при $s_i(1) = 2$ они совпадают с оценками МНК.

Таблица 6.5. Оценки параметров восстанавливающих фильтров и средние квадраты разностей восстановленных и исходных изображений.

σ^2_ξ	Критерий	шаг	Параметры инверсных КИХ-фильтров					$\sigma^2_{\Delta x}$
			$\hat{h}_{r0}$	$\hat{h}_{r1}$	$\hat{h}_{r2}$	$\hat{h}_{r3}$	$\hat{h}_{r4}$	
0	МНК	1	8.687516	-4.199416	-4.739308	-2.618809	3.870017	205.50
		2	1.693985	-0.634953	-0.369007	-0.621824	0.931799	149.04
		3	1.033900	-0.67391	0.013760	-0.020369	0.040100	145.36
	(6.14, 6.15)	2	1.830662	-0.630797	-0.749170	-0.669917	1.219222	131.75
		3	1.098201	-0.080034	-0.153213	0.026276	0.108770	114.85
36	МНК	1	4.415751	-0.873153	-2.678178	0.747165	-0.611585	1018.9
		2	0.740701	0.972306	-0.543510	0.110665	-0.280162	1047.19
		3	0.624883	0.956389	-0.635098	-0.088430	0.142256	1100.91
	(6.14, 6.15)	2	0.706808	1.019733	-0.572980	0.177245	-0.330806	1058.3
		3	0.535312	1.017578	-0.616960	0.050619	0.013451	1091.9

Результаты, приведенные в таблице, показывают, что применение взвешенных оценок, обеспечивающих бóльшую близость оценок параметров к истинным значениям, дает выигрыш при отсутствии (или низкой) интенсивности помех в исходных данных. При наличии помех в исходных данных средние значения квадрата разности исходного и восстановленного изображений, полученных с использованием МНК-оценок и фильтра, оцененного по критерию (6.14), при-

близительно одинаковы. В обоих случаях мера близости более чем в два раза выше аналогичной меры для разности исходного и искаженного (и зашумленного) изображений (500). Субъективно изображения, восстановленные с использованием результатов идентификации по критерию (6.14), воспринимаются, по крайней мере, не менее качественными, чем полученные с использованием МНК-оценок.

Полученные результаты позволяют сделать следующие выводы. Восстанавливающий КИХ-фильтр всегда может быть построен путем непосредственной идентификации инверсного тракта. При этом возможно достижение различного регуляризирующего эффекта путем непосредственного "управления" весовой матрицей. Качество восстановления, в конечном итоге, будет зависеть от характеристик модели искажений в реальной системе, а также от того, насколько она в действительности соответствует модели искажений тестовых изображений, использовавшихся на этапе идентификации.

6.3. Идентификация кусочно-постоянных линейных моделей пространственно-зависимых искажений и восстанавливающих фильтров

В системах формирования изображений искажения часто оказываются пространственно-зависимыми (неизопланатичными). Задача коррекции таких искажений может быть решена применением линейных восстанавливающих фильтров с перестраиваемыми по пространственным координатам параметрами. Если модель неизопланатичности может быть представлена в виде зависящего от небольшого числа параметров семейства двумерных функций, параметры восстанавливающего фильтра могут быть легко вычислены в каждой точке обрабатываемого изображения. В настоящем разделе рассматривается информационная технология идентификации параметров такой функции по тестовым изображениям.

Задачу идентификации модели пространственно-зависимых искажений можно свести к совокупности задач оценки пространственно-инвариантных моделей на малых, насколько это возможно, фрагментах изображения. При этом, как уже указывалось, возникает проблема отбора малых фрагментов изображений, на которых измерения являются информативными, т.е. таких на которых задача идентификации может быть решена с требуемой точностью. Разрабатываемая ниже информационная технология учитывает связанный с указанной проблемой круг вопросов.

Наиболее характерным типом искажений, вносимых оптическими системами, является расфокусировка, связанная с аберрациями различных порядков. Предположим, что эти искажения на малом фрагменте изображения могут быть описаны уравнением КИХ-фильтра (1.5). С учетом приведенных (в двух предыдущих разделах) результатов исследования задачу идентификации модели пространственно-зависимых искажений будем решать в виде следующей последовательности подзадач [33,65]:

1. Идентификация импульсной характеристики на малых фрагментах;
2. Оценка на каждом фрагменте параметра расфокусировки σ ;
3. Оценка параметров функции, описывающей изменения параметра расфокусировки по полю изображения (модели неизопланатичности);
4. Оценка параметров восстанавливающих фильтров, соответствующих полученным значениям параметра расфокусировки σ.

Фрагменты для идентификации отбираются по заданному пороговому значению показателя S_{min}. Для каждого выбранного фрагмента задача идентификации параметров импульсной характеристики (этап 1) решается как описано в разделе 6.1.

По полученным на каждом фрагменте оценкам значений импульсной характеристики - $\hat{h}(k_1,k_2)$, $k_1,k_2=\overline{1,M}$ далее строится оценка $\hat{\sigma}$ параметра σ модели (6.3). Для этого решается задача определения

$$\hat{\sigma}_s: \qquad Q(\hat{\sigma}_s)=\min_{\sigma\in\Xi} Q(\sigma), \tag{6.16}$$

где s - номер фрагмента, а Ξ. - область допустимых значений параметра расфокусировки. В частности, для получения МНК - оценки $\hat{\sigma}_s$ параметра σ_s

$$Q(\sigma_s)=\sum_{k_1=-\frac{N}{2}}^{\frac{N}{2}}\sum_{k_2=-\frac{N}{2}}^{\frac{N}{2}}\left(\frac{1}{\sqrt{2\pi}\sigma_s}\exp\left(-\frac{k_1^2+k_2^2}{2\sigma_s^2}\right)-\hat{h}(k_1,k_2)\right)^2 \tag{6.17}$$

Далее (этап 3) по совокупности $\hat{\sigma}_s$ строятся оценки параметров модели неизопланатичности $\sigma=\sigma(n_1,n_2)$. Для этого необходимо задать параметрическое семейство двумерных функций, описывающих характер неизопланатичности. Для выбора этого класса функций можно осуществлять предварительный визуальный просмотр полученного множества оценок $\hat{\sigma}_s$. Например, если расфокусировка увеличивается по мере удаления от центра изображения, удобно изменение параметра σ описать параметрическим семейством функций вида

$$\sigma(n_1, n_2) = \lambda_x n_1^2 + \lambda_y n_2^2 + \lambda_{xy} n_1 n_2 + \beta_x n_1 + \beta_y n_2 + \sigma_0 \ , \qquad (6.18)$$

где n_1, n_2 текущие координаты точки, а $\lambda_x, \lambda_y, \lambda_{xy}, \beta_x \beta_y, \sigma_0$ константы, зависящие от степени неизопланатичности.

Ограничимся рассмотрением частного случая указанной зависимости, когда неравномерность расфокусировки обладает радиальной симметрией. Параметрическое семейство функций, описывающих такую модель неизопланатичности имеет вид

$$\sigma(n_1, n_2) = \lambda_x n_1^2 + \lambda_y n_2^2 + \sigma_0 \ , \qquad (6.19)$$

где $\lambda_x = \lambda_y$ коэффициенты, характеризующие изменения искажений, описываемых моделью (6.3), по мере удаления от центра $n_1 = n_2 = 0$.

Четвертый, заключительный этап идентификации определение параметров восстанавливающих фильтров. С использованием оцененной на третьем этапе модели неизопланатичности, в принципе, в любой точке изображения можно строить соответствующий восстанавливающий фильтр. Однако при этом объем вычислительной работы оказывается весьма значительным. Для сокращения времени обработки изображений целесообразно заранее построить множество фильтров, соответствующих различным значениям параметра σ, в виде некоторой таблицы соответствия, а затем в ходе обработки выбирать "ближайший"

Для составления такой таблицы можно воспользоваться методом непосредственной идентификации характеристик инверсного тракта, описанным в предыдущем разделе 6.2. Для этого необходимо осуществлять идентификацию параметров восстанавливающего фильтра при различных искажениях, полагая входным - искаженное, а выходным - неискаженное изображение.

Достоинство такого подхода мы уже обсуждали в разделе 6.2. Для получения таблицы соответствия параметра σ и параметров инверсных фильтров (4-й этап) в качестве тестового использовалось изображение "текст", показанное на (рис. 6.6). Из этого изображения было получено множество искаженных в соответствии с моделью (1.5), (6.4) изображений при различных значениях параметра σ. Последовательность параметров σ задавалась в диапазоне 0.5-3.5. с шагом 0.1. Полученные таким образом для каждого фиксированного значения σ искаженные изображения (типа показанного на рисунке 6.8) использовались в качестве входных для идентификации параметров восстанавливающего фильтра.

Осуществлялось экспериментальная проверка эффективности описанной выше в виде четырех этапов сквозной информационной технологии построения корректирующих фильтров неизопланатичных искажений. Для этого использовалось изображение (рис. 6.16), полученное из исходного (рис. 6.6) моделированием пространственно-зависимых искажений. Расфокусировка осуществлялась КИХ-фильтром (1.5), значения импульсной характеристики которого определялись по формуле (6.4), а параметр σ в каждой точке изображения рассчитывался по соотношению (6.19) при $\lambda_x = \lambda_y = 0.0002275$, $\sigma_0 = 0.932$.

Для указанных параметров модели неизопланатичности размер опорной области должен увеличиваться от центра изображения к периферии. В качестве критерия для выбора размеров маски использовалась величина отношения значения гауссиана (6.4) на границе опорной области к его значению в центральной точке. Для величины отношения не более 0,01 минимальный размер маски равен 5, а максимальный - 13. На рис. 6.16 видно, что при описанном способе моделирования расфокусировка, как и следовало ожидать, возрастает от центра изображения к периферии.



Рис. 6.16. Тестовое изображение

Идентификация осуществлялась в виде последовательности описанных в предыдущем разделе четырех этапов. В результате реализации первых двух этапов на фрагментах изображения, отобранных по показателю S_{min}, было получено множество оценок параметра расфокусировки $\hat{\sigma}_s$

Далее с использованием полученного множества оценок реализовывался третий этап оценка параметров функции, описывающей изменения параметра σ по полю изображения (модели неизопланатичности). Результаты оценивания приведены в таблице 6.6.

Таблица 6.6. Оценки параметров модели (6.19)

Параметры	Оценки параметров	Абсолютная погрешность	Относительная погрешность
$\lambda_x=\lambda_y$	0.000231	0.00000306	0.01350
σ_0	0.928000	0.00438000	0.00470

Для иллюстрации эффективности описанной схемы идентификации с использованием построенного набора фильтров осуществлялась обработка изображения текста, показанного на рис.6.17, а. Это изображение получено с использованием той же модели искажений.

Восстановленное изображение приведено на рис. 6.18, а. На рис. 6.18, б для сравнительной визуальной оценки приведено изображение этого же текста, полученное обработкой "средним" фильтром, т.е. инверсным фильтром, построенным по той же методике, но без учета пространственной неинвариантности искажений. Как и следовало ожидать, восстановление произошло лишь на отдельных участках изображения.

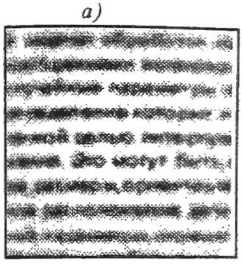


Рис. 6.17. Исходное изображение

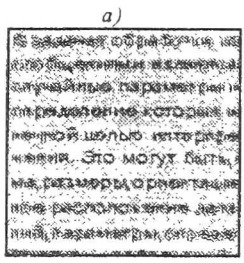


Рис. 6.18. Результаты восстановления: а - пространственно-зависимым фильтром: б - "средним" фильтром

Эффективность предложенной схемы идентификации параметров пространственно-зависимых восстанавливающих фильтров проиллюстрирована на частном примере модели неизопланатичности с радиальной симметрией. Эта схема может использоваться для более широкого класса искажений. Успешность решения задачи в каждом конкретном случае, кроме прочего, будет зависеть от того, насколько удачно подобрано параметрическое семейство двумерных функций, описывающих изменение параметра расфокусировки по пространственным координатам.

6.4. Идентификации моделей систем формирования изображений в классе фильтров с бесконечной импульсной характеристикой

Описанные выше информационные технологии и результаты экспериментальных исследований связаны с использованием КИХ-фильтров. В настоящем разделе мы приведем результаты решения тех же задач с использованием фильтров с бесконечной импульсной характеристикой (БИХ-фильтров).

Модель БИХ-фильтра описывается соотношением (1.1). Будем полагать, что опорные области D_x, D_y на входном (неискаженном) и выходном (зарегистрированном системой) изображениях прямоугольники с размерами $K_1 \times K_2$ и $L_1 \times L_2$ соответственно. При этом соответствующее разностное уравнение записывается в виде (1.3).

При реализации БИХ-фильтра (1.3) возникают принципиальные трудности, связанные с тем, что для вычисления выходных отсчетов $y(n_1, n_2)$ используются отсчеты $y(n_1 - k_1, n_2 - k_2)$, $k_2 = \overline{0,\ (K_1 - 1)}$, $k_2 = \overline{0,\ (K_2 - 1)}$ из той же опорной области. Поэтому для построения физически реализуемых фильтров необходимо согласовывать направления "обхода" изображения и форму опорной области. Сказанное относится как к моделированию искажений, так и их восстановлению.

Это ограничение не имеет значения при решении задачи идентификации. Действительно, матрица **X** и векторы **y** , ξ соответствующей (1.3) матричной записи модели БИХ-фильтра имеют вид

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,L} & y_{1,1}, & y_{1,K} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ x_{N,1} & x_{N,L} & y_{N,1}\cdot & y_{N,K} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{bmatrix}, \quad \xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \vdots \\ \xi_N \end{bmatrix}. \qquad (6.20)$$

Здесь мы для компактности записей изменили систему индексации входных $x(n_1 - l_1, n_2 - l_2)$, $l_1 = \overline{0,\ L_1}$, $l_2 = \overline{0,\ L_2}$ и выходных $y(n_1 - k_1, n_2 - k_2)$, $k_1 = \overline{0,\ (K_1 - 1)}$, $k_2 = \overline{0,\ (K_2 - 1)}$ отсчетов, для чего ввели новые обозначения для размерности описывающих их массивов: $L=L_1 \times L_2$, $K=(K_1 \times K_2)-1$, а $M=L+K$. Компоненты вектора ошибок ξ формируются в соответствии с соотношением (1.4).

Из (1.2) и (6.20) видно, что с точки зрения задачи оценивания безразлично в каком порядке будут взяты строки в матрице **X**. Поэтому задача идентификации модели БИХ-фильтра может решаться точно так же, как и для модели КИХ-фильтра. Отличие заключается в том, что размерность вектора искомых параметров зависит от порядка передаточной функции системы, т.е. от размеров опорных областей как на входном, *так и на выходном* изображении.

Несмотря на это, фактического увеличения размерности задачи (если оценивать вычислительную сложность сквозной технологии оценивания и восстановления) при этом может и не происходить. Дело в том, что моделью БИХ-фильтра невысокого порядка можно аппроксимировать характеристики искажений, для оценивания которых в классе КИХ-фильтров потребуется маска больших размеров. Поэтому в итоге при одинаковом качестве восстановления применение БИХ-фильтра может дать даже экономию вычислений.

Если задача минимизации вычислительных затрат не ставится, то при использовании моделей БИХ-фильтра можно реализовать более высокое качество восстановления. Поэтому, несмотря на некоторые технические трудности реализации процедур восстановления, идентификация моделей систем формирования и восстановления изображений в классе фильтров с бесконечной импульсной характеристикой представляет значительный интерес. В качестве иллюстрации сказанного рассмотрим простейшую задачу моделирования и идентификации линейных искажений.

Для моделирования искаженного изображения будем использовать БИХ-фильтр порядка (1,1), который описывается разностным уравнением вида

$$y(n_1,n_2) = a_{10}y(n_1-1,n_2)+a_{01}y(n_1,n_2-1)+ a_{11}y(n_1-1,n_2-1) + \\ +b_{00}x(n_1,n_2)+b_{10}x(n_1-1,n_2)+b_{01}x(n_1,n_2-1)+b_{11}x(n_1-1,n_2-1)+\xi(n_1,n_2). \quad (6.21)$$

На рис. 6.19 приведено исходное тестовое изображение. Параметры моделирующего фильтра (6.21) для различных вариантов передаточной функции искажающей системы приведены в первых трех строках таблицы 6.7. Соответствующие указанным вариантам искаженные изображения показаны на рис. 6.19, а, б и в. Хотя во всех случаях использовались одинаковые опорные области (3×3), характер искажений существенно различается. По степени искажений указанные варианты искажений (а, б и в) для определенности будем называть "слабым", "средним" и "сильным" соответственно.

Таблица 6.7. Результаты оценивания параметров БИХ-фильтра

Вариант	a_{10}	a_{01}	a_{11}	b_{00}	b_{10}	b_{01}	b_{11}
а	0.125	0.125	0.166	0.250	0.083	0.083	0.166
б	0.165	0.165	0.330	0.110	0.055	0.055	0.110
в	0.200	0.200	0.400	0.066	0.033	0.033	0.066
а	0.125	0.125	0.168	0.249	0.084	0.083	0.164
б	0.163	0.163	0.332	0.110	0.056	0.055	0.110
в	0.200	0.200	0.405	0.066	0.033	0.032	0.067

По полученным тестовым изображениям решалась задача идентификации параметров исходной модели (6.21) для всех указанных вариантов искажений. Оценки $\hat{a}_{ij}$, $\hat{b}_{ij}$ искомых параметров a_{ij}, b_{ij}, $ij = \overline{0,1}$ и относительные погрешности δ_c при отсутствии помех приведены в последних трех строках той же таблицы 6.7.

Исследовалась также возможность непосредственной идентификации инверсной модели в классе БИХ-фильтров. Структура инверсной модели БИХ-фильтра задавалась в виде разностного уравнения того же, что и для модели прямого тракта (6.21), порядка:

$$\hat{x}(n_1, n_2) = \hat{b}^{в}_{10}\hat{x}(n_1 - 1, n_2) + \hat{b}^{в}_{01}\hat{x}(n_1, n_2 - 1) + \hat{b}^{в}_{11}x(n_1 - 1, n_2 - 1) + \\ + \hat{a}^{в}_{00}y(n_1, n_2) + \hat{a}^{в}_{10}y(n_1 - 1, n_2) + \hat{a}^{в}_{01}y(n_1, n_2 - 1) + \hat{a}^{в}_{11}y(n_1 - 1, n_2 - 1). \quad (6.22)$$

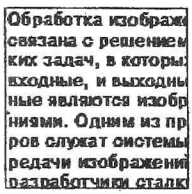


Рис. 6.19. Исходное изображение

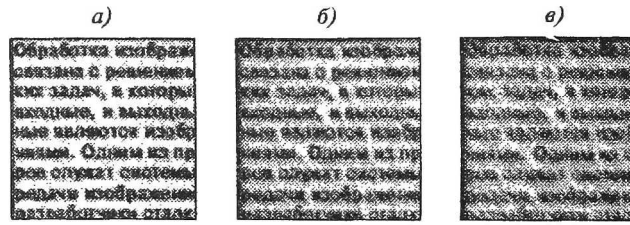


Рис. 6.20. Варианты линейных искажений: а - слабое; б - среднее; в - сильное

Точные значения параметров инверсной модели для трех указанных в таблице 6.7 вариантов искажений (а, б и в), полученные путем вычисления коэффициентов обратной передаточной функции, приведены в первых трех строках таблицы 6.8. Те же коэффициенты, полученные путем непосредственной идентификации инверсной модели, приведены в последних трех строках той же таблицы. Так же, как и для прямого тракта, результаты хорошо согласуются.

Таблица 6.8. Параметры и оценки параметров инверсной модели

Вариант	$\hat{b}^{в}_{10}$	$\hat{b}^{в}_{01}$	$\hat{b}^{в}_{11}$	$\hat{a}^{в}_{00}$	$\hat{a}^{в}_{10}$	$\hat{a}^{в}_{01}$	$\hat{a}^{в}_{11}$
а	-0.332	-0.332	-0.664	4.000	-0.500	-0.500	-0.664
б	-0.500	-0.500	-1.000	9.090	-1.500	-1.500	-3.000
в	-0.500	-0.500	-1.000	15.15	-3.030	-3.030	-6.060
а	-0.335	-0.333	-0.658	4.008	-0.502	-0.499	-0.673
б	-0.510	-0.501	-1.012	9.114	-1.487	-1.484	-3.025
в	-0.492	-0.485	-1.000	15.02	-2.995	-2.941	-6.080

Далее из таблицы 6.8 были взяты коэффициенты восстанавливающего БИХ-фильтра, полученные путем непосредственной идентификации. С использовани-

ем этих коэффициентов фильтром (6.22) осуществлялось восстановление изображений двух типов: "текст" и "город" (рис. 6.21), искаженных с использованием тех же трех вариантов моделей, соответствующих "слабому", "среднему" и "сильному" искажениям. Искаженные изображения в той же последовательности приведены на рис. 6.22.



Рис 6.21. Исходные изображения

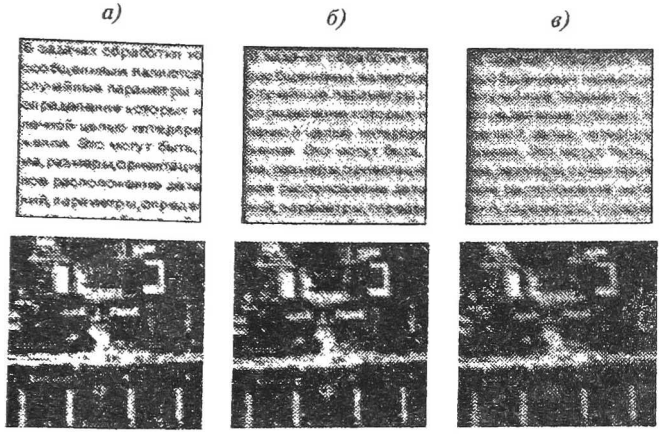


Рис 6.22. Изображения, подвергшиеся линейным искажениям: а - слабое; б - среднее; в - сильное

На рис. 6.23 приведены изображения, полученные путем восстановления наиболее сильно искаженных изображений (по варианту в). Результаты восстановления искажений, соответствующих вариантам моделей а и б, по крайней мере, не хуже, поэтому не приводятся. Подчеркнем, что опорные области восстанавливающих фильтров на входном и выходном изображениях для всех трех вариантов линейных искажений имели размеры 3×3. Ясно, что для "сильных" линейных искажений (по варианту в) размеры опорной области соответствующего КИХ-фильтра должны быть существенно увеличены.

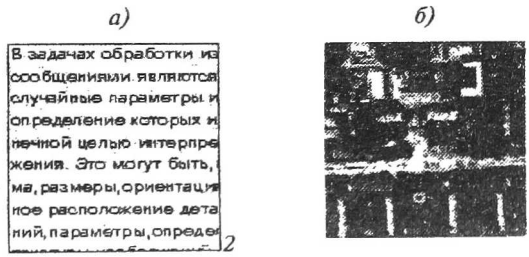


Рис. 6.23. Результаты восстановления изображений: а- "текст"; б "город" БИХ-фильтром, полученным путем идентификации инверсной модели

В проведенных выше экспериментах для простоты опорные области задавались так, чтобы очередное значение выходного отсчета выражалось через значения входных отсчетов, а также уже найденные выходные отсчеты, а коэффициенты "симметричными": $\hat{a}_{10} = \hat{a}_{01}$, $\hat{b}_{10} = \hat{b}_{01}$. При реализации восстанавливающих БИХ-фильтров и коррекции искажений, обладающих радиальной симметрией, возникают принципиальные трудности, связанные с тем, что для вычисления выходных отсчетов используются отсчеты из той же опорной области. В то же время при идентификации таких БИХ-фильтров дополнительных проблем не возникает.

Для иллюстрации возможности непосредственной идентификации восстанавливающих БИХ-фильтров в качестве модели искажающей системы использовался фильтр порядка (2,2), который описывается передаточной функцией (3.21). Параметры передаточной функции (3.21), как и ранее, задавались так, чтобы обеспечивалась радиальная симметрия искажений:

$$b_{10} = b_{01} = b_{-10} = b_{0-1}, \qquad b_{11} = b_{1-1} = b_{11} = b_{-1-1}, \tag{6.23}$$

$$a_{10} = a_{01} = a_{-10} = a_{0-1}, \qquad a_{11} = a_{1-1} = a_{11} = a_{-1-1}. \tag{6.24}$$

Для моделирования и восстановления искаженных изображений использовался итерационный метод [14] реализации БИХ-фильтров. При выполнении условий (6.24), (6.25) полином $A(\omega_1,\omega_2)$, соответствующий знаменателю передаточной функции (3.21), является вещественной и положительной функцией. Поэтому для нахождения частотного отклика можно использовать итерационное выражение вида

$$Y_i(\omega_1,\omega_2) = \lambda \cdot B(\omega_1,\omega_2) \cdot X(\omega_1,\omega_2) + C(\omega_1,\omega_2) \cdot Y_{i-1}(\omega_1,\omega_2), \tag{6.25}$$

где

$$C(\omega_1,\omega_2) = \left|1 - \lambda \cdot A(\omega_1,\omega_2)\right| < 1.$$

Параметр λ выбирался из условия устойчивости итерационной схемы (6.25) в соответствии с неравенствами [14]:

$$0 < \lambda < \frac{2}{\max\limits_{(\omega_1,\omega_2)} A(\omega_1,\omega_2)}.$$

Шум $\zeta(n_1,n_2)$, как и прежде, моделировался в виде смеси гауссовского с параметрами (M_x=0, D_x=16) и, с вероятностью P=0.03, равномерного в интервале [0-64] импульсного шума. Исходное тестовое и полученное при указанных выше условиях итерационным методом искаженное изображения приведены на

рис. 3.4, а и б. Размеры опорных областей на входном и выходном изображениях, соответствующие передаточной функции (3.21), 3×3. Ясно, что для реализации примерно таких же линейных искажений с помощью КИХ-фильтра потребовалась бы маска значительных размеров.

По сформированным тестовым изображениям осуществлялась идентификация восстанавливающего БИХ-фильтра, описываемого передаточной функцией

$$\hat{G}(z_1, z_2) = \frac{\hat{A}(z_1, z_2)}{\hat{B}(z_1, z_2)}, \qquad (6.26)$$

где

$$\hat{A}(z_1, z_2) = \hat{a}_0[1 + \hat{a}_1(z_1 + z_1^{-1} + z_2 + z_2^{-1}) + \hat{a}_2(z_1 z_2 + z_1^{-1} z_2^{-1} + z_1 z_2^{-1} + z_1^{-1} z_2)],$$

$$\hat{B}(z_1, z_2) = 1 + \hat{b}_1(z_1 + z_1^{-1} + z_2 + z_2^{-1}) + \hat{b}_2(z_1 z_2 + z_1^{-1} z_2^{-1} + z_1 z_2^{-1} + z_1^{-1} z_2).$$

Точные параметры фильтра (6.26), соответствующие модели искажений (3.21), соответственно равны: $\hat{a}_0 = 35.2857$, $\hat{a}_1 = -1.2$, $\hat{a}_2 = 0.27$, $\hat{b}_1 = 1.2$, $\hat{b}_2 = 0.27$

Задача идентификации модели БИХ-фильтра [49] решалась на фрагменте, содержащем 24 наблюдения, в соответствии с общей схемой (1.13) (1.17), описанной в разделе 1.3. Вначале строилась МНК-оценка и вычислялся вектор невязок. Затем с использованием полученного вектора невязок строился корректирующий вектор так, как описано в разделе 4.1, осуществлялось преобразование данных и вновь вычислялась МНК-оценка. Восстанавливались 19 компонент вектора ошибок по пяти заданным. Для задания этих пяти компонент использовались полученные ранее в эксперименте коэффициенты связи выделяющихся компонент векторов ошибок и невязок.

В результате идентификации были получены следующие оценки параметров: $\hat{a}_0 = 34.761$, $\hat{a}_0\hat{a}_1 = -41.05279$, $\hat{a}_0\hat{a}_2 = 8.775$, $\hat{b}_1 = 1.216$, $\hat{b}_2 = 0.2669$.

С использованием полученных оценок осуществлялась обработка изображения "часы", подвергшегося тем же искажениям, что и тестовое изображение "текст", показанное на рисунке 3.4. Эти изображения приведены на рисунке 6.24, а - без шума и б - с добавлением указанной выше смеси гауссовых и импульсных помех. Полученные в результате обработки изображения приведены на рис. 6.25, а и б соответственно.

Конечно, такое качество восстановления не всегда достижимо. Если идентифицированный БИХ-фильтр оказывается неустойчивым, мы вынуждены, в интересах реализуемости, изменить его параметры, принеся в жертву качество

восстановления. Проблемы, связанные с обеспечением устойчивости БИХ-фильтров, являются важными и представляют самостоятельный интерес. Наша задача в данном случае состояла лишь в том, чтобы показать принципиальную возможность непосредственной идентификации инверсных БИХ-фильтров по малым фрагментам искаженного и неискаженного изображений.

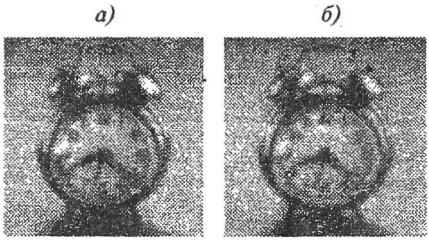


Рис 6.24. Исходные искаженные изображения: а- без шума; б- с шумом

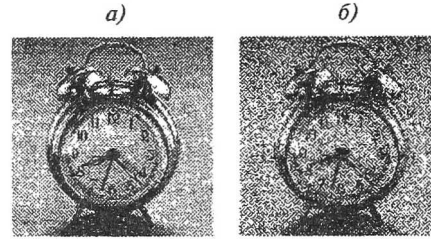


Рис 6.25. Восстановленные изображения: а - без шума; б - с шумом

По существу этой задачи необходимо еще раз подчеркнуть следующее. Несмотря на некоторые трудности, связанные с реализацией БИХ-фильтров, вычислительная сложность сквозной технологии оценивания и восстановления, особенно при интенсивных линейных искажениях, может быть меньшей, по сравнению с использованием КИХ-фильтров. Связано это с возможностью использования опорных областей небольших размеров при любых искажениях. Это представляется особенно выгодным при проведении идентификации моделей искажающих систем и восстанавливающих фильтров по малым фрагментам изображений. Ниже рассматривается информационная технология, в которой это преимущество БИХ-фильтров проявляется в полной мере.

6.5. Оценка разрешающей способности видеотракта

Оценка разрешающей способности видеотракта по одному искаженному изображению, содержащему контрастные объекты, т.е. при отсутствии полномасштабной миры, является наиболее типичной задачей, где возникает проблема идентификации модели искажающей системы по малым фрагментам изображений. В настоящем разделе приводятся результаты сравнительных исследований возможности решения этой задачи с использованием моделей видеотракта в виде фильтров с конечной и бесконечной импульсными характеристиками.

Известная методика [34] измерения разрешающей способности основана на использовании сформированных изображений мир (чередующихся темных и

светлых штрихов), показанных на рис. 6.26. Эта методика сводится к следующему. Предполагается, что видеотракт описывается линейной моделью наблюдения сигнала с шумом с последующей равномерной дискретизацией сигнала по координатам. Частотная характеристика оцениваемой системы

$$H(\Omega_1,\Omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\tau_1,\tau_2)\exp[-j(\Omega_1\tau_1 + \Omega_2\tau_2)]d\tau_1 d\tau_2$$

задает значения коэффициента передачи пространственных гармоник двумерного входного сигнала на частотах Ω_1,Ω_2, а $h(\tau_1,\tau_2)$- импульсная характеристика искажающей двумерной непрерывной линейной системы с постоянными параметрами. В качестве входного сигнала используется сечение поля яркости миры (рис. 6.26) вдоль оси t_1, т.е. в направлении, перпендикулярном штрихам. Соответствующая этому сечению одномерная функция изображена на рис. 6.27.

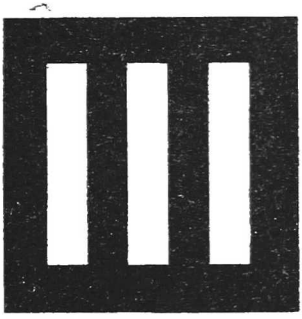


Рис. 6.26. Тестовая мира

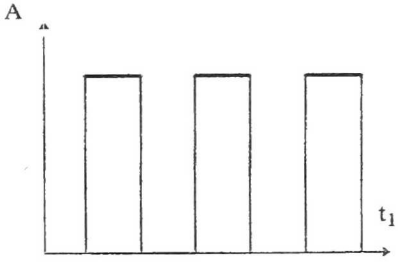


Рис. 6.27. Сечение поля яркости миры

Периодически продолжив данную функцию на всю числовую ось, разложим ее в ряд Фурье на периоде, т.е. на интервале [-1,1]. Несложно показать, что основная, первая гармоника этого разложения записывается в виде

$$\hat{f}_1(t_1) = \frac{2A}{\pi}\mathrm{Sin}(\frac{\pi}{l}t_1),$$

т.е. имеет амплитуду

$$A_1^{(x)} = 2A/\pi \qquad (6.27)$$

и пространственную частоту (измеряемую в радианах на единицу длины)

$$\Omega_1 = \pi/l.$$

Рассматриваемая гармоника после прохождения сквозного тракта будет иметь в выходном сигнале амплитуду, равную уже не (6.27), а

$$A_1^{(y)} = \frac{2A}{\pi}H(\frac{\pi}{l},0). \qquad (6.28)$$

Предполагается, что шум наблюдения $v(t_1,t_2)$ слабо коррелирован и после дискретизации сигнала его можно считать дискретным белым шумом. Пусть σ_v среднеквадратическое отклонение отсчетов шума. Известно, что зрительный анализатор человека обладает способностью усреднять такой шум в пределах наблюдаемых штрихов постоянной яркости. Для рассматриваемого тестового объекта на каждом штрихе выполняется (оценочно)

$$N = \frac{1}{T}\left(\frac{5l}{T}\right) = 5\frac{l^2}{T^2}$$

независимых шумовых отсчетов, поэтому "кажущееся" среднеквадратичное отклонение шума (после его субъективного усреднения) будет равно

$$\tilde{\sigma}_v = \frac{\sigma_v}{\sqrt{N}} = \frac{1}{\sqrt{5}}\frac{T}{l}\sigma_v. \qquad (6.29)$$

Типичная частотная характеристика линейной модели сквозного тракта убывает к нулю с ростом частотных аргументов, т.е. при уменьшении размеров объекта амплитуда (6.28) первой гармоники полезного сигнала на выходе тракта будет также уменьшаться. Одновременно, в соответствии с (6.29) будет наблюдаться рост кажущегося СКО шума. Следовательно, с уменьшением ширины штриха тестового объекта отношение "сигнал/шум"

$$d = \frac{A_1^{(y)}}{\tilde{\sigma}_v} \qquad (6.30)$$

монотонно убывает и при ширине, равной значению искомого показателя линейного разрешения ($l=l_0$), достигает некоторого нижнего порогового значения, определяемого равенством:

$$d_{|l=l_0} = K,$$

где K - безразмерный параметр (порог), зависящий от требуемого уровня вероятности ошибок различения штрихов и обычно выбираемый из диапазона $2<K<5$. Из (6.30) с учетом (6.28), (6.29) получаем:

$$d = \frac{2\sqrt{5}Al_0H\left(\frac{\pi}{l_0},0\right)}{\pi T\sigma_v} = \frac{2\sqrt{5}AH(\Omega,0)}{\Omega T\sigma_v} = K, \qquad (6.31)$$

где в предпоследнем выражении введено обозначение для пространственной частоты, соответствующей значению показателя линейного разрешения:

$$\Omega = \pi/l_0. \qquad (6.32)$$

Из (6.31) следует уравнение относительно определяющей линейное разре-

шение частоты Ω:

$$H(\Omega,0) = Q\Omega, \tag{6.33}$$

где

$$Q = \frac{1}{2\sqrt{5}} KT\left(\frac{\sigma_v}{A}\right) \tag{6.34}$$

- параметр, совокупно характеризующий надежность различения штрихов (через величину K), шаг дискретизации (T) и относительное превышение контраста над шумом (A/σ_v).

Формулы (6.32) (6.34) позволяют определить величину показателя линейного разрешения расчетным путем, без привлечения наблюдателя и без использования тестового изображения. При заданных T и A для расчета необходимо знать уровень шума и соответствующее сечение частотной характеристики оцениваемой системы.

Таким образом, при построения автоматизированной информационной технологии оценки разрешающей способности по описанной методике в условиях, когда тестовое (неискаженное) изображение объекта отсутствует, основные проблемы связаны с решением следующих двух задач: а) формирование фрагментов изображений, на которых искажения устранены с использованием априорной информации о форме объекта; б) оценка частотной характеристики $H(\Omega,0)$ по малым фрагментам изображения.

В рамках указанной методики на искаженном изображении допускается использовать лишь те фрагменты, которые содержат элементы "штрихового" объекта. В частности, ищутся участки, имеющие резкие перепады яркости и в тоже время такие, что изменения яркости в пределах каждого из двух соседних уровней незначительны. Для этого вначале осуществляется отбор двухуровневых (по яркости) фрагментов, имеющих заметный контраст и разделительную линию яркостей точно посередине фрагмента. Затем, среди отобранных площадок "претендентов" отбирается небольшое их число, для которых суммарное СКО, подсчитанное для обоих уровней яркости, минимально.

Заключительный этап формирования "неискаженных" участков заключается в компьютерном "ретушировании" отобранных фрагментов изображения. Предполагается, что наименьшим линейным искажениям подвергаются наиболее удаленные от линии перепада яркостей участки фрагментов. Поэтому процедура ретуширования сводится к замене отсчетов (в пределах ступеньки) средними зна-

чениями яркости, вычисленными по отсчетам, расположенным на наиболее удаленных от центра линиях.

В описанной методике, по существу, реализуется воссоздание отсутствующего эталона (неискаженного) фрагмента изображения по априорной информации о геометрической форме объекта и условиях регистрации изображения. При формировании этих эталонов могут использоваться более сложные (чем перепады яркости вдоль некоторой линии) элементы объектов, если такая априорная информация имеется.

Для оценки частотной характеристики $H(\Omega.0)$ используется ее аппроксимация БИХ-фильтром первого порядка с опорными областями D_x (на входном) и D_y (на выходном) изображениях размерами 3×1:

$$y(n) = \sum_{l=-1}^{1} b(l)x(n-l) - \sum_{k=-1}^{1} a(k)y(n-k) + \xi(n), \quad k \neq 0, \tag{6.35}$$

где $b(l)$, $a(k)$- значения импульсной характеристики, $x(n)$, $x(n)$ - отсчеты поля яркости в точках опорных областей D_x и D_y соответственно, а $\xi(n)$- дискретная последовательность, определяемая как

$$\xi(n) = \delta y(n) + \sum_{k=-1}^{1} a(k)\delta y(n-k), \quad k \neq 0, \tag{6.36}$$

а δy- ошибки измерения отсчетов поля яркости выходного изображения. Для указанных в (6.35) опорных областей передаточная функция, описывающая искажения в направлении оси t_1, имеет вид

$$H(z) = B(z)/A(z), \tag{6.37}$$

где $A(z) = 1 + a_1(z + z^{-1})$, $B(z) = b_0 + b_1(z + z^{-1})$. Соответствующая z-передаточной функции (6.37) частотная характеристика $H(\Omega)$ вещественна:

$$H_1(\Omega) = \frac{b_0 + 2b_1\mathrm{Cos}(\Omega T)}{1 + 2a_1\mathrm{Cos}(\Omega T)}, \tag{6.38}$$

Для N отсчетов на фрагменте изображения для модели (6.35) можно записать уравнение типа (1.2):

$$\mathbf{X}\mathbf{c} = \mathbf{y} + \boldsymbol{\xi},$$

где вектор $\mathbf{c}$ формируется как $\mathbf{c} = [b_0, b_1, -a_1]^T = [c_1, c_2, c_3]^T$, а матрица $\mathbf{X}$ и векторы $\mathbf{y}$, $\boldsymbol{\xi}$ в данном случае соответственно равны

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & y_{1,1} \\ \dots & \dots & \dots \\ x_{N,1} & x_{N,2} & y_{N,1} \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \qquad \xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_N \end{bmatrix}.$$

Компоненты Nx1 векторов y и ξ формируются в соответствии с соотношениями (6.35), (6.36). Для определения параметров БИХ-фильтра строится модифицированная МНК-оценка (1.11) с использованием итерационной схемы получения оценок (1.13)-(1.17). Преобразование данных заключается в корректировке вектора выхода модели. Для корректировки здесь целесообразно использовать описанный в разделе (4.1) метод с восстановлением N-M компонент корректирующего вектора $\tilde{\xi}$. Связано это с тем, что в данном случае вид двумерной функции яркости всегда известен (единичная двумерная ступенька). Поэтому коэффициенты связи между выделяющимися компонентами векторов ошибок и невязок, с учетом этой информации, могут быть определены достаточно точно.

Исследовалась сквозная информационная технология оценки разрешающей способности видеотракта, включающая следующие этапы:

1. Поиск фрагментов на искаженном изображении, для которых перепад яркости превышает заданное пороговое значение;
2. Вычисление для каждого найденного фрагмента среднего значения яркости в каждой из двух зон (темной и светлой) и СКО яркости от соответствующего среднего в каждой зоне;
3. Выбор фрагмента с наименьшим СКО яркости;
4. Формирование "неискаженного" тестового фрагмента изображения [34,35];
5. Оценка параметров передаточной функции БИХ-фильтра (6.37) и вычисление частотной характеристики (6.38);
6. Решение уравнения (6.33) с целью определения пространственной частоты Ω, характеризующей линейное разрешение видеотракта.

На рис. 6.28, а-г приведены использовавшиеся для оценки эффективности технологии изображения (эксперимент №1). Использовалось типичное изображение участка дороги с разметкой (рис. 6.28, а). Исходное изображение было подвергнуто расфокусировке фильтром первого порядка с параметрами:

$b_0 = 0.047$, $b_1 = 0.053$, $a_0 = 1$, $a_1 = -0.9$.

Полученное расфокусированное изображение этого же участка дороги показано на рис. 6.28, б. На рис. 6.28, в приведено то же изображение с двумя выделен-

ными и сформированными "неискаженными" тестовыми фрагментами (в светлых прямоугольниках). В результате решения задачи идентификации были получены следующие оценки параметров БИХ-фильтра первого порядка:

$$\hat{b}_0 = 0.0509, \quad \hat{b}_1 = 0.0534, \quad \hat{a}_0 = 1, \quad \hat{a}_1 = -0.896,$$

с использованием которых в соответствии с (6.38) рассчитывалась частотная характеристика.

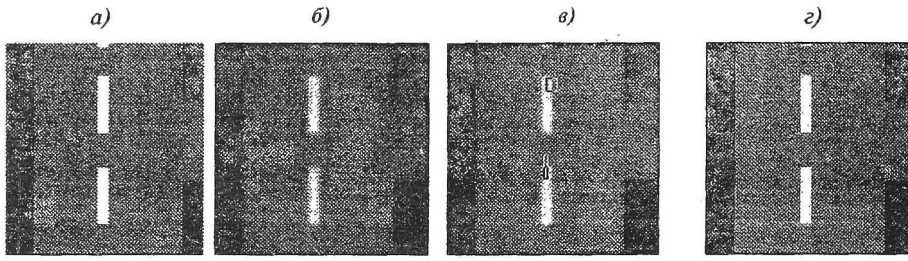


Рис. 6.28. Тестовые изображения: а - исходное; б - расфокусированное; в - с указанием выделенных фрагментов; г - восстановленное

Завершающий (6-й) этап информационной технологии заключался в решении уравнения (6.33) относительно круговой частоты Ω и определении показателя линейного разрешения. При расчетах использовались данные, приведенные в таблице 6.9. Параметры K и T_0 задавались, а контраст и СКО рассчитывались на указанных выделенных фрагментах.

Таблица 6.9. Исходные данные (эксперимент №1)

Характеристики	Фрагмент с горизонтальным перпадом яркости	Фрагмент с вертикальным перпадом яркости
Контраст	71.6	82.2
СКО	12.1	10.5
K	3.5	3.5
T_0	0.01 м	0.01 м

В результате были получены следующие значения, характеризующие разрешение:

- Зачетная частота - 35 1/с.
- Зачетная круговая частота 109.96 рад/с.
- Линейное разрешение 0.0286 м.

Оказалось, что линейное разрешение, определенное по описанной методике, более чем в два раза превысило величину интервала дискретизации (T_0=0.01 м).

Аналогичный эксперимент проводился на тестовом изображении, которое было получено из исходного (рис. 6.28, а) путем расфокусировки в частотной области фильтром с гауссовой частотной характеристикой (6.3):

$$h(i,j) = \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left(-\left(i^2 + j^2\right)/2\sigma^2\right), \qquad (6.39)$$

где i,j - координаты отсчетов в опорной области, а σ - параметр расфокусировки.

В отличие от предыдущего в этом эксперименте использовались два фрагмента, различающиеся видом функции перепада яркости. Использовались переходы "светлый -темный" и "темный-светлый". При расчетах использовались данные, приведенные в таблице 6.10.

Таблица 6.10. Исходные данные (эксперимент №2)

Характеристики	"светлый-темный"	"темный-светлый"
Контраст	66.78	131.81
СКО	10.39	20.04
K	3.5	3.5
T_0	0.01	0.01

В результате были получены следующие характеристики разрешения:

- Зачетная частота - 51.7 1/с.
- Зачетная круговая частота - 162.42 рад/с.
- Линейное разрешение 0.019 м.

Результаты этого эксперимента хорошо согласуются с предыдущим, т.к. параметр расфокусировки подбирался так, чтобы частотные характеристики искажений в обоих экспериментах были достаточно близкими, хотя они, конечно, различны.

Поскольку описанная методика оценки разрешающей способности основана на данных исследований субъективного восприятия, была также проведена и объективная оценка качества определения модели искажений. Для этого осуществлялась идентификация инверсной модели БИХ-фильтра первого порядка и восстановление этим фильтром искаженного изображения. Результат восстановления приведен на рисунке 6.28, г. Достоверность определения частотной характеристики подтверждается также близостью оценок параметров к задававшимся (при моделировании искажений) параметрам БИХ-фильтра.

Производилось также сравнение частотной характеристики (6.38) искажающей системы, полученной путем оценивания передаточной функции БИХ-фильтра (6.37), с гауссовой частотной характеристикой вида (6.39), которая использовалась при моделировании искажений. Результаты сравнения приведены на рис. 6.29. Максимальная и среднеквадратическая погрешности аппроксимации гауссовой частотной характеристики БИХ-фильтром оказались соответственно равными:

$\Delta_{MAX}=0.0455$, $\Delta_{СКО}=0.0007$.

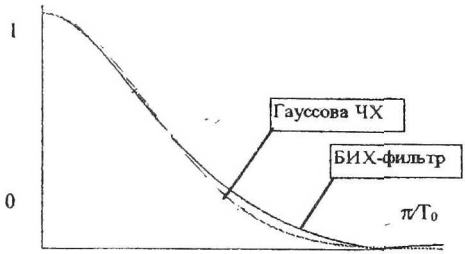


Рис. 6.29. Результаты испытаний

Кроме того, производилось сравнение качества аппроксимации искажений с применением БИХ-фильтра первого порядка и КИХ-фильтров различных порядков для одних и тех же искажений. Результаты исследования приведены в таблице 6.11.

Таблица 6.11. Сравнение с КИХ-фильтром

σ	Тип ф-ра	Маска	Δ_{max}	СКО
1	КИХ	5	0.02398472	0.00038715
	БИХ	3	0.02816998	0.00045672
1.2	КИХ	5	0.04219072	0.00082203
	БИХ	3	0.04545734	0.00075587
1.5	КИХ	5	0.11187530	0.00206668
	БИХ	3	0.09013103	0.00147081
2	КИХ	9	0.10613500	0.00142538
	БИХ	3	0.08086268	0.00129539

Из таблицы видно, что при небольших искажениях изображения КИХ-фильтр третьего порядка (размер маски=5×1) и БИХ-фильтр 1-го порядка сравнимы по точности аппроксимации передаточной функции. При увеличении степени размытия (параметра σ) точность оценивания параметров КИХ-фильтра того же порядка уменьшается гораздо быстрее, чем у БИХ-фильтра.

Повышение порядка КИХ фильтра не позволяет получить погрешность меньшую погрешности БИХ-фильтра первого порядка. Это вызвано тем, что при повышении размерности фильтра (размера маски) в пределы маски попадают уже не только участки, содержащие резкие перепады яркости, но и малоинформативные фрагменты фона. Это приводит к ухудшению обусловленности матрицы **A** и, как следствие, к увеличению погрешности оценки параметров передаточной функции фильтра.

Исследовалась также зависимость погрешности определения зачетной круговой частоты $\Delta\omega$ от параметра Q (6.34) , характеризующего надежность различения штрихов. В данном эксперименте для удобства рассматривался не параметр Q, а параметр α =arctg(Q). На рис. 6.30 представлено семейство зависимостей ошибки оценивания зачетной круговой частоты от параметра α при различных значениях

погрешности аппроксимации ΔH.

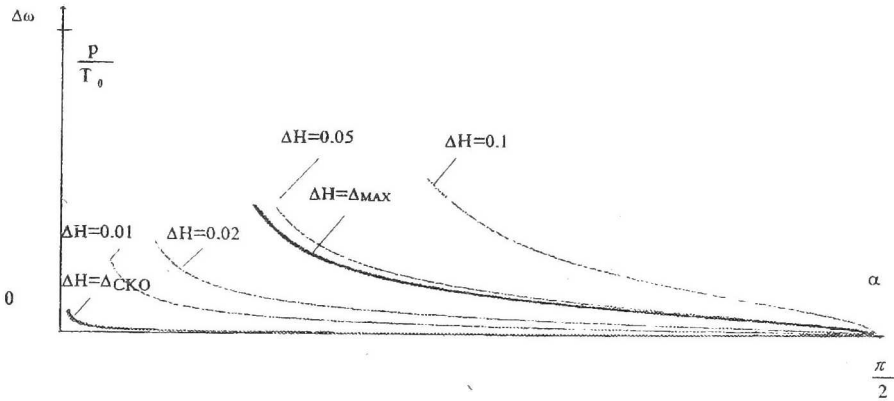


Рис. 6.30. Зависимость ошибки оценивания круговой частоты от параметра α

Приведенные результаты подтверждают принципиальную возможность оценки разрешающей способности видеотракта по одному искаженному изображению, если на изображении имеются малые фрагменты с распределением яркости в виде "ступеньки" Используемая модель тракта в виде БИХ-фильтра позволяет избежать трудностей, связанных с выбором размеров опорной области в условиях априорной неопределенности относительно импульсной характеристики системы. Применяемый для идентификации подход позволяет строить оценки при неизвестных вероятностных распределениях помех. Последнее обстоятельство представляется важным для описанной технологии, т.к. для оценивания используются весьма малые фрагменты изображений.

6.6. Информационная технология совмещения изображений по информативным фрагментам

Во многих алгоритмах и информационных технологиях (калибровка видеокамер, стереозрение, построение рельефа местности, привязка к существующим картам и т.д.), использующих описание одного и того же сюжета, применяют процедуры согласования (совмещения) изображений. При совмещении изображений с целью распознавания (например, идентификации личности по отпечаткам пальцев и др.) параметры взаимных сдвигов и поворота совмещаемых изображений, как правило, сильно различаются. Более того, сюжет совмещаемо-

го изображения может являться лишь частью сюжета, хранящегося в памяти в качестве образца.

Совмещение таких изображений с применением традиционных технологий, основанных на использовании корреляторов и пиковых фильтров, связано с большим объемом вычислительной работы. Для уменьшения вычислительных затрат в работе [32] предложено предварительно осуществлять локальную цилиндрическую аппроксимацию (дактилоскопических изображений) и фиксировать координаты точек, в которых имеют место большие ошибки оценивания. Далее задача распознавания решается с использованием сравнительно небольшого числа точек (названных в указанной выше работе особенностями). Эта технология, конечно, приводит к существенной экономии, однако объем вычислительной работы, связанной с цилиндрической аппроксимацией, все же остается достаточно большим.

В настоящей работе строится информационная технология совмещения изображений, в которой для определения координат точек "претендентов" на соответственные используется показатель диагонального преобладания информационной матрицы, формируемой из отсчетов на каждом заданном фрагменте. Для точного определения соответственных точек предлагается критерий максимума евклидовой нормы матрицы, являющейся результатом перемножения матриц измерений на фрагментах, в окрестности точек "претендентов" на соответственные.

Общая схема решения задачи совмещения изображений включает три этапа:
- выделение информативных фрагментов (точек "претендентов");
- определение среди "претендентов" соответственных точек;
- оценка параметров сдвига и поворота по координатам соответственных точек.

Рассмотрим эти этапы в указанной последовательности.

1. Выделение информативных фрагментов. Список точек "претендентов" на соответственные формируется из наиболее информативных. Анализ информативности фрагментов в окрестности каждой точки с координатами (n_1,n_2) осуществляется с использованием показателя диагонального преобладания [47]:

$$\phi\left[\mathbf{R}(n_1,n_2)\right]=\frac{\mathrm{tr}^2\mathbf{R}(n_1,n_2)}{\mathrm{tr}\mathbf{R}^2(n_1,n_2)}=\frac{M^2}{M+\sum r_{ij}^2},\ i,j=\overline{1,M},\ i\neq j,\ M-1\le\phi\left[\mathbf{R}(n_1,n_2)\right]\le M, \quad (6.40)$$

где

$$\mathbf{R}(n_1,n_2)=\mathbf{C}(n_1,n_2)\mathbf{A}(n_1,n_2)\mathbf{C}(n_1,n_2), \quad (6.41)$$

$$\mathbf{A}(n_1, n_2) = \mathbf{X}^T(n_1, n_2)\mathbf{X}(n_1, n_2) \qquad (6.42)$$

- так называемая информационная матрица,

$$\mathbf{C}(n_1, n_2) = \mathrm{diag}\left(a_{11}^{-1/2}, a_{22}^{-1/2}, \ldots, a_{MM}^{-1/2}\right), \qquad (6.43)$$

а $\mathbf{X}(n_1, n_2)$ N×M матрица, составленная из отсчетов поля яркости на фрагменте в окрестности точки (n_1, n_2).

Для обеспечения инвариантности к углу поворота изображения матрица **X** формируется с использованием фрагмента в виде круга, вписанного в квадрат с нечетной стороной. На рис. 6.31 приведен пример кругового фрагмента, вписанного в квадрат 7×7 (предполагается, что шаг дискретности по осям одинаков). Матрица **X** для этого фрагмента имеет размерность 16×4. Элементы матрицы **X** формируются из отсчетов в узлах прямоугольной сетки, ближайших к точкам пересечений концентрических окружностей и радиальных лучей. Столбцы соответствуют концентрическим окружностям и центральной точке, строки радиальным лучам. Для уменьшения ошибок, связанных с переходом от прямоугольной системы координат к полярной, в некоторых случаях целесообразно осуществлять линейную интерполяцию небольшого числа ближайших к данному пересечению окружности и радиальной прямой отсчетов в узлах прямоугольной сетки.

При описанном способе формирования матрицы **X** соответствующая ей матрица $\mathbf{A}=\mathbf{X}^T\mathbf{X}$ (с точностью до указанных методических погрешностей) не зависит от углового положения фрагмента. При изменении углового положения фрагмента происходит циклическая перестановка строк матрицы **X**, что приводит лишь к изменению прядка суммирования произведений соответствующих элементов. Вследствие умножения слева и справа на матрицу **C**, диагональное преобладание получающейся из **A** матрицы **R** вдобавок не зависит также и от различий среднего уровня яркости регистрируемых изображений.

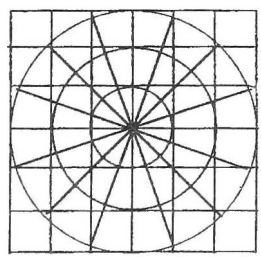


Рис. 6.31. Пример кругового фрагмента, вписанного в квадрат 7×7

Если масштабы зарегистрированных изображений различаются менее более чем в 1±1/2M раза, где M число концентрических окружностей (без центральной точки), то используемый показатель будет также нечувствительным к изменениям масштаба. Исходя из последнего требования, диаметр круговой рамки желательны выбирать, по возможности, небольшим. В то же время из соображений повышения надежности определения соответственных точек диаметр кругового фрагмента целесообразно увеличивать.

2. Определение соответственных точек. Для определения соответственных точек можно использовать обычный коррелятор. Для уменьшения числа сравнений информативные точки предварительно разделяются по величине диагонального преобладания на две три группы. Сравнение точек внутри каждой группы осуществляется по принципу "каждая с каждой". Эта информация является достаточно надежной, т.к. показатель диагонального преобладания информационной матрицы фактически является носителем локальных автокорреляционных свойств сигнала.

Для перебора всех возможных дискретных значений взаимного углового положения в данном случае формируются две матрицы (обозначим их $\mathbf{X}$ и $\ddot{\mathbf{X}}$) для двух точек (претендентов) совмещаемых изображений. Матрица $\ddot{\mathbf{X}}$ путем циклической перестановки строк представляется в одном из 16 видов, соответствующих различным положениям кругового фрагмента (рис. 6.29) на сюжете совмещаемого изображении. Для проверки соответствия фрагментов вычисляется матрица

$$\mathbf{R}_c = \mathbf{C}\mathbf{A}_c\bar{\mathbf{C}}, \tag{6.44}$$

где

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

а

$$\mathbf{C} = \mathrm{diag}\left(\|x_1\|, \|x_2\|, \ldots, \|x_M\|\right), \tag{6.45}$$

$$\bar{\mathbf{C}} = \mathrm{diag}\left(\|\bar{x}_1\|, \|\bar{x}_2\|, \ldots, \|\bar{x}_M\|\right). \tag{6.46}$$

Ясно, что на соответственных фрагментах при их угловом положении, приводящем к одинаковой последовательности строк матриц $\mathbf{X}$ и $\ddot{\mathbf{X}}$ и при отсутствии ошибок в измерениях матрица $\mathbf{R}_c$ совпадает с $\mathbf{R}$, определяемой соотношением (6.41).

След матрицы $\mathbf{R}_c$ представляет собой сумму M коэффициентов корреляции, вычисленных для M концентрических масок:

$$\frac{\mathrm{tr}\mathbf{R}_c}{M} = \frac{1}{M}\sum_{i=1}^{M} k_i = k, \tag{6.47}$$

где $k_i = \langle x_i \cdot \bar{x}_i \rangle / \|x_i\| \cdot \|\bar{x}_i\|$, $i = \overline{1,M}$. Если нормы векторов $\|x_i\|$, $i = \overline{1,M}$ одинаковы, величина k совпадает с обычным коэффициентом корреляции. При наличии шумов или неточном совпадении положений фрагментов будут иметь место различия. Тем не менее, величина k, определяемая соотношением (6.47), является достаточно информативной характеристикой соответствия искомых точек (фрагментов).

Применение описанного способа определения соответственных точек позволяет существенно повысить надежность, по сравнению с применением обычного коррелятора. В данном случае дополнительной информацией являются также знаки элементов матрицы $\mathbf{R}_c$. В частности, если хотя бы один элемент матрицы $\mathbf{R}_c$ оказался отрицательным, соответствующая точка отвергается, а дальнейшие вычисления прекращаются. Если все диагональные элементы положительны, вычисляются внедиагональные элементы. При этом в качестве дополнительной информации также используются знаки элементов матриц $\mathbf{R}$ и $\mathbf{R}_c$. Поскольку в большинстве случаев вычисления прекращаются после обработки отсчетов, соответствующих одной концентрической окружности, достигается существенная экономия вычислений.

Соответственными считаются лишь те точки, для которых оценки $\hat{\varphi}$ угловых положений фрагментов совпадают. При этом используется предположение, что таких точек на изображении больше, чем ложных. Для повышения надежности точки, имеющие похожее, но ориентированное иначе, чем для большинства информативных точек на круговом фрагменте, поле яркости отсеиваются.

3. Оценка параметров сдвига и поворота. Для оценивания параметров сдвига и поворота используются соотношения.

$$\hat{n}_1(i) = n_1(i) \cdot \mathrm{Cos}\alpha - n_2(i) \cdot \mathrm{Sin}\alpha - (x_0 \cdot \mathrm{Cos}\alpha - y_0 \cdot \mathrm{Sin}\alpha), \qquad (6.48)$$

$$\hat{n}_2(i) = n_2(i) \cdot \mathrm{Cos}\alpha + n_1(i) \cdot \mathrm{Sin}\alpha - (x_0 \cdot \mathrm{Cos}\alpha + y_0 \cdot \mathrm{Sin}\alpha), \qquad (6.49)$$

где $\hat{n}_1(i)$, $\hat{n}_2(i)$ оценки координат соответственной (i-й) точки на анализируемом изображении, $n_1(i)$, $n_2(i)$ координаты соответственной (i-й) точки на исходном изображении, x_0, y_0 параметры сдвига, α параметр (угол) взаимного поворота изображений

В соответствии с (6.48), (6.49) по оценкам координат L (i=1,L, L≥3) соответственных точек формируются два соответствующих (6.48), (6.49) уравнения типа (1.2):

$Sc=y+\xi$,

где L×3 матрица S и L×1 вектор y задаются соответственно в виде

$$S = \begin{bmatrix} n_1(1) & -n_2(1) & -1 \\ n_1(2) & -n_2(2) & -1 \\ \vdots & \vdots & \vdots \\ n_1(L) & -n_2(L) & -1 \end{bmatrix}, \qquad y = \begin{bmatrix} \hat{n}_1(1) \\ \hat{n}_1(2) \\ \vdots \\ \hat{n}_1(L) \end{bmatrix},$$

и

$$S = \begin{bmatrix} n_2(1) & n_1(1) & -1 \\ n_2(2) & n_1(2) & -1 \\ \vdots & \vdots & \vdots \\ n_2(L) & n_1(L) & -1 \end{bmatrix}, \qquad y = \begin{bmatrix} \hat{n}_2(1) \\ \hat{n}_2(2) \\ \vdots \\ \hat{n}_2(L) \end{bmatrix},$$

Далее для переопределенных систем (6.48) и (6.49) ищутся соответственно оценки $\hat{c}_1$ и $\hat{c}_2$:

$$\hat{c}_1 = [\hat{c}_{11},\ \hat{c}_{12},\ \hat{c}_{13}] = [\mathrm{Cos}\alpha,\ \mathrm{Sin}\alpha,\ x_0\mathrm{Cos}\alpha - y_0\mathrm{Sin}\alpha]$$

$$\hat{c}_2 = [\hat{c}_{21},\ \hat{c}_{22},\ \hat{c}_{23}] = [\mathrm{Cos}\alpha,\ \mathrm{Sin}\alpha,\ x_0\mathrm{Cos}\alpha + y_0\mathrm{Sin}\alpha]$$

при условии выполнения равенств:

$$\hat{c}_{11} = \hat{c}_{21}, \qquad \hat{c}_{12} = \hat{c}_{22}, \qquad \hat{c}^2_{11} + \hat{c}^2_{12} = 1, \qquad \hat{c}^2_{21} + \hat{c}^2_{22} = 1.$$

Для вычисления параметров сдвига x_0, y_0 с использованием полученных оценок решается система уравнений:

$$\hat{c}_{13}(i) = x_0 \cdot \hat{c}_{11} - y_0 \cdot \hat{c}_{12},$$

$$\hat{c}_{23}(i) = x_0 \cdot \hat{c}_{21} + y_0 \cdot \hat{c}_{22}.$$

Проводилась экспериментальная проверка работоспособности описанной технологии. В качестве тестового использовались два изображения, показанные на рис. 6.32. Изображение на рис. 6.32, б получено поворотом относительно левого нижнего угла на 90^0 и сдвигом на 127 точек (пикселов).

а) *б)*



Рис. 6.32. Тестовые изображения: а - исходное; б - повернутое на 90^0

На рис. 6.33 приведены полученные по этим изображениям так называемые "поля информативности", рассчитанные по соотношению (6.40) для 100 наиболее информативных точек. На этих изображениях видно, что наиболее информативными преимущественно оказались точки в окрестности особенностей (разветвлений папиллярных линий).

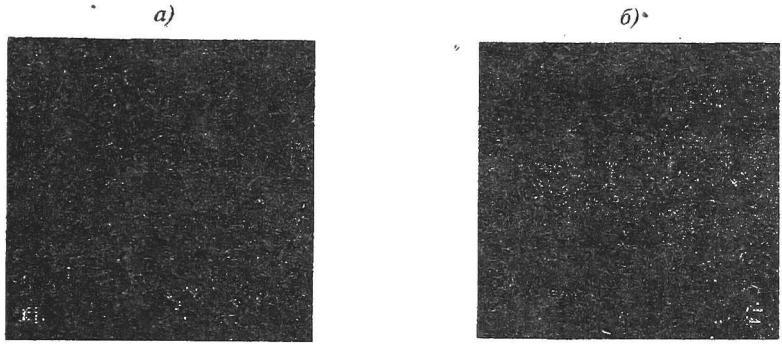


Рис. 6.33. Поля информативности тестовых изображений: а - исходного; б-повернутого на 90^0

Оценки параметров сдвига и поворота в данном эксперименте получились следующими: $x_0=0$, $y_0=127$, $\alpha=-90^0$, а преобразованное изображение в точности совпало с исходным, показанным на рис. 6.32, а. Такое точное совпадение объясняется тем, что при повороте на 90^0 отсутствуют искажения, связанные с преобразованием поворота. Ниже приводится другой пример, в котором такие искажения имеют место.

На рис. 6.34, а и б приведены тестовые изображения участка отпечатка пальца. Координаты точек повернутого изображения, показанного на рис. 6.34, б рассчитывались по формулам

$$\tilde{n}_y(i) = (n_x(i) - x_0 - x_c)\cdot \mathrm{Sin}\alpha - (n_y(i) - y_0 - y_c)\cdot \mathrm{Cos}\alpha + y_c, \qquad (6.50)$$

$$\tilde{n}_x(i) = (n_x(i) - x_0 - x_c)\cdot \mathrm{Cos}\alpha - (n_y(i) - y_0 - y_c)\cdot \mathrm{Sin}\alpha + x_c, \qquad (6.51)$$

где $x_0=25$, $y_0=-5$, параметры сдвига, $x_c=64$ $y_c=64$ - координаты центра, а $\alpha=70^0$ угол поворота изображения. Поля информативности исходного и повернутого изображений (по 100 точек на каждом) показаны на рис. 6.35, а и б соответственно.

В ходе оценивания в соответствии с описанной технологией на каждом изображении были отобраны по 63 точки с одинаковой ориентацией соответственных фрагментов. Затем по этим точкам были оценены параметры

сдвига и поворота, в соответствии с моделью формирования описываемой соотношениями (6.48), (6.49): x_0=87, y_0=-40, α=-69.573^0.

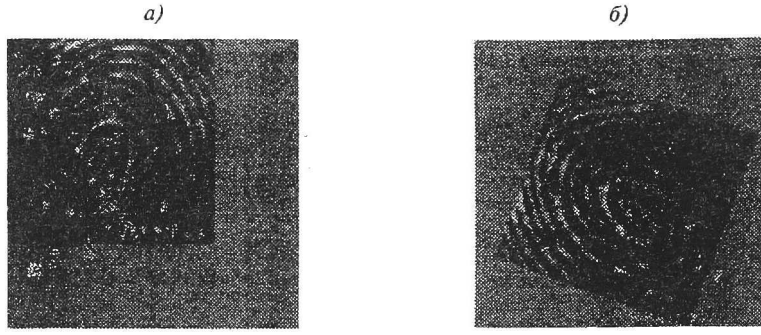


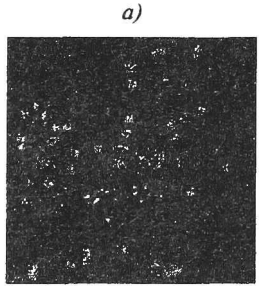
Рис. 6.34. Тестовые изображения: а - исходное; б - сдвинутое и повернутое на 70^0

Изображение, полученное из "повернутого" путем преобразований сдвига и поворота по соотношениям

$$n_x(i) = (\tilde{n}_x(i) - x_0) \cdot \mathrm{Cos}\alpha - (\tilde{n}_y(i) - y_0) \cdot \mathrm{Sin}\alpha\,, \qquad (6.52)$$

$$n_y(i) = (\tilde{n}_y(i) - y_0) \cdot \mathrm{Cos}\alpha - (\tilde{n}_x(i) - x_0) \cdot \mathrm{Sin}\alpha \qquad (6.53)$$

с использованием полученных выше оценок, приведено на рис 6.36. Из рисунка видно, что имеет место неточное совмещение (на 2 пикселя по вертикали).



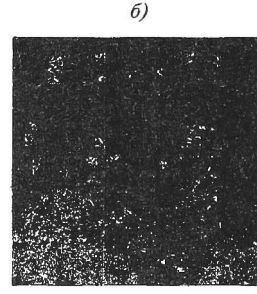


Рис. 6.35. Поле информативных точек на изображении: а - исходном, б - сдвинутом и повернутом

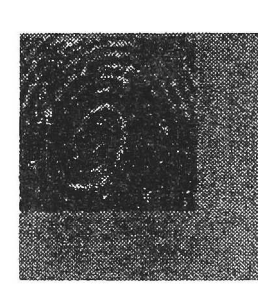


Рис. 6.36. Восстановленное изображение

Мы специально привели этот пример, несмотря на то что в других экспериментах имело место подавляющее число случаев точного совпадения. Ошибки на 1-2 пиксела в описанной технологии возможны вследствие округления на этапе оценивания параметров преобразования. Поэтому заключительный этап технологии всегда должен заключаться в точном совмещении изображений обычными

способами в предположении, что параметры сдвига и взаимного поворота малы. Эта задача может решаться по небольшому числу выделенных ранее информативных фрагментов с использованием обычного коррелятора и прямоугольной маски (без перебора различных угловых положений). Ясно, что этот этап технологии не потребует значительных вычислительных затрат.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Алимов Ю.И. Несовместимость фишеровской теории оценок с требованием многократного воспроизведения экспериментального результата // Сб. Идентификация, прогнозирование и управление в технических системах. Владивосток, Изд-во ДальГУ, 1986.
2. Беллман Р. Введение в теорию матриц. М.: Наука, 1976. - 351 с.
3. Белоногов В.Д., Николаев Ю.А., Теряев Е.Д., Шамриков Б.М. Оценивание параметров динамической модели самолета с целью построения адаптивного цифрового управления // Тез. докл. на IV симпозиуме ИФАК по идентификации и оценке параметров систем. Тбилиси, Изд. Мецниереба, 1976 г.
4. Белоногов В.В., Теряев Е.Д., Фурсов В.А., Шамриков Б.М. Построение алгоритмов идентификации и адаптации в адаптивных цифровых системах управления самолетами // Тр. VIII Всесоюзн. совещ. по пробл. управления. Таллин, 1980. - с. 263-265.
5. Васин Ю.Г. Теоретические и прикладные аспекты создания новых геоинформационных технологий и систем // Тезисы докладов Ш конференции "Распознавание образов и анализ изображений: новые информационные технологии" (РОАИ - 97). Нижний Новгород, 1997, Ч.1. с. 118-119.
6. Василенко Г.И., Тараторин А.М. Восстановление изображений. М.: Радио и связь, 1986. - 304 с.
7. Вентцель Е.С. Теория вероятностей. - Гос. изд-во физ.-мат. литературы. М.: 1958, 464 с.
8. Виттих В.А., Сергеев В.В., Сойфер В.А. Обработка изображений в автоматизированных системах научных исследований. М.: Наука, 1982. 214с.
9. Воеводин В.В. Вычислительные основы линейной алгебры. М.: Наука, 1977. - 303 с.
10. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления.- М.: Наука, 1984. 320 с.
11. Гантмахер Ф.Р. Теория матриц. М.: Наука, 1967. - 575 с.
12. Гроп Д. Методы идентификации систем. Пер. с англ. Под ред. Кринецкого Е.И. М.: Мир, 1979. - 302 с.

13. Гуревич И.Б., Журавлев Ю.И., Сметанин Ю.Г. Алгебры изображений: исследовательские и прикладные задачи // Тезисы докладов IV конференции "Распознавание образов и анализ изображений: новые информационные технологии" (РОАИ - 98). Новосибирск, 1998, Ч.1. - с. 74-78.
14. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. Пер. с англ.- М.: Мир, 1984. - 488 с.
15. Демиденко Е.З. Линейная и нелинейная регрессии.- М.: Финансы и статистика, 1981. - 303 с.
16. Жданов А.И. Вычисление решений некорректных стохастических алгебраических уравнений регуляризованным методом наименьших квадратов // Докл. АН СССР, т. 306, № 2. - с. 324 - 327.
17. Жданов А.И. Рекуррентное оценивание минимальных собственных значений информационных матриц //Автоматика и телемеханика, 1986, № 6.- с. 25-36.
18. Иванов В.К. О некорректно поставленных задачах // Матем. сборник, 1963, т. 61, (103), № 2. - с. 211-223.
19. Калман Р.Е. Идентификация систем с шумами // Успехи математических наук, т. 40, вып. 4 (244), 1985. - с. 27-41.
20. Каминскас В.А.. О правиле останова некоторых последовательных алгоритмов идентификации // Автоматика и вычислительная техника, 1972, № 4. - с 39-44.
21. Колмогоров А.Н. Основные понятия теории вероятностей. - 2 изд., М., 1974.
22. Лецкий Э.К., Вучков И.Н. Об одном методе последовательной идентитфикации // Изв. АН СССР, Техн. кибернетика, 1970, №2. - с. 234-238.
23. Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. Пер. с англ. М.: Наука, Гл. ред. физ.-мат. лит., 1986.
24. Люнг Л. Идентификация систем. Теория для пользователя. Пер. с англ. Под ред. Я.З. Цыпкина. М.: Наука, 1991. - 432 с.
25. Маркус М., Минк Х. Обзор по теории матриц и матричных неравенств. Пер. с англ. М.: Наука, 1972. - 232 с.
26. Мелькановнч А.Ф. Фотографические средства и их эксплуатация. М.: Изд-во МО, 1984. 576 с.
27. Мудров В.И., Кушко В.Л. Методы обработки измерений. М.: Сов. радио. 1976. - 192 с.

28. Петров Б.Н., Теряев В.Д., Шамриков Б.М.. Условия параметрической идентифицируемости управляемых объектов в разомкнутых и замкнутых автоматических системах // Изв. АН СССР, Техн. кибернетика, 1977, №2. с. 160-175.
29. Поляк Б.Т., Цыпкин Я.З.. Помехоустойчивая идентификация // В сб. Идентификация и оценка параметров систем. Тр. IV Симпоз. ИФАК, ч. I, Тбилиси, Мецниереба, 1976. с. 190-213.
30. Прэтт У.К. Цифровая обработка изображений. - М.: Мир, 1982. Кн.2. - 480с.
31. Пытьев Ю.П., Пытьева Ю.Н., Козлов А.С. Проблемы интерпретации измерений: теоретико-вероятностный и теоретико-возможностный подходы // Тезисы докладов Ш конференции "Распознавание образов и анализ изображений" (РОАИ - 97). Нижний Новгород, Ч.1, 1997. - с. 238-242
32. Сергеев В.В., Фролова Л.Г. Разработка и применение алгоритма цилиндрической полиномиальной аппроксимации изображения в скользящем окне // Автометрия, 1996, №1.
33. Сергеев В. В., Фурсов В.А., Максимов М.В. Идентификация параметров моделей и корректирующих фильтров пространственно-зависимых искажений // Международный симпозиум "Информационная оптика. Научные основы и технологии". Москва, Россия, (W3-06P), 1997.
34. Сергеев В.В., Фурсов В.А., Парфенов С.И. Информационная технология оценки разрешающей способности видеотракта с использованием модели фильтра с бесконечной импульсной характеристикой // Тезисы докладов IV конференции "Распознавание образов и анализ изображений: новые информационные технологии. Новосибирскк, Ч. I, 1998. - с. 378-381.
35. Сойфер В.А., Котляр В.В., Фурсов В.А. Построение алгоритмов оперативной коррекции искажений на изображениях в оптико-электронных системах наведения и целеуказания // Тр. конф. Волжского регионального центра РАРАН, Саров, 1998. - с. 108-109.
36. Солодовников В.В., Бирюков В.Ф., Тумаркин В.И. Принцип сложности в теории управления. М.: Наука, 1977. - 342 с.
37. Тарасенко Ф.П. Непараметрическая статистика. Томск, изд-во ТГУ, 1976. 294 с.
38. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М.: Наука, 1974. - 223 с.

39. Уилкинсон Дж.Х. Алгебраическая проблема собственных значений. Пер. с англ. М.: Наука, 1970. - 564 с.
40. Фурсов В.А. Об оценке относительных возмущений систем нормальных уравнений в задачах идентификации // Сб.: Преобразование информации в радиотехнических системах. Тр. МАИ, вып 504,1979. - с 82-86.
41. Фурсов В.А., Шамриков Б.М. Достаточные оценки точности идентификации динамических объектов // Сб.: Оценивание в адаптивных системах управления. Тр. МАИ, 1986.
42. Фурсов В.А. Построение оценок точности идентификации по малому числу дискретных наблюдений // Сб.: Алгоритмическое и техническое обеспечение проектирования систем управления движущимися объектами. Тр. МАИ, 1991. с. 53-61.
43. Фурсов В.А. Анализ точности и построение алгоритмов идентификации по малому числу наблюдений // Изв. АН СССР, Техн. кибернетика, № 6, 1991.
44. Фурсов В.А. Введение в идентификацию по малому числу наблюдений. Изд. МАИ, 1991. 36 с.
45. Фурсов В.А. Построение оценок обусловленности в задачах линейной регрессии // Сб.: Алгоритмическое и техническое обеспечение автоматизированных систем обработки информации и управления. Тр. МАИ, 1993. - с. 38-44.
46. Фурсов В.А. Локализация собственных значений вещественных симметричных матриц // Тез. докл. XI Российского коллоквиума "Современный групповой анализ и задачи математического моделирования" Самара, 1993.
47. Фурсов В.А. Идентификация оптических искажающих систем с отбором информативных фрагментов изображений // Компьютерная оптика. Вып. 14-15, 1995/ - с. 78-79.
48. Фурсов В.А. Восстановление изображений КИХ-фильтрами, построенными путем непосредственной идентификации инверсного тракта // Компьютерная оптика. Вып. 16, 1996, с. 103-108.
49. В.А. Фурсов, С.И. Парфенов. Идентификация моделей систем формирования изображений в классе фильтров с бесконечной импульсной характеристикой. Компьютерная оптика. Вып. 18, 1998.
50. Хьюбер П.Дж. Робастность в статистике. - М.: Мир, 1984.

51. Цыпкин Я.З. Основы информационной теории идентификации. М.: Наука,1983.
52. Чернецкий В.И. и др. Математические методы и алгоритмы исследований автоматических систем. - Л.: Энергия, Ленингр. отделение, 1970, 373 с.
53. Шамриков Б.М. Сравнительный анализ точности параметрической идентификации динамических объектов в разомкнутых и замкнутых автоматических системах // Изв. АН СССР. Техн. кибернетика, №3, 1986. с. 143-150.
54. Шамриков Б.М. Параметрическая идентификация динамических объектов по выборкам ограниченного объема // Изв. РАН. Техн. кибернетика, №2, 1997.
55. Шамриков Б.М., Фурсов В.А. Оценивание достоверности параметрической идентификации объекта в адаптивных СУ // Изд. АН СССР, Техн. кибернетика, №6, 1979. - с.173-180.
56. Эйкхофф П. Основы идентификации систем управления. Пер. с англ. Под ред. Райбмана Н.С. - М.: Мир, 1975. - 683 с.
57. Юсупов Р.М. Элементы теории идентификации технических объектов. - МО СССР, 1974.
58. Ярославский Л.П. Введение в цифровую обработку изображений. - М.: Сов. радио, 1979. - 312 с.
59. Bjorck Ake. Least Squares Methods. Elsevier Science Publishers B.V. (North-Holland). 1990.
60. Belonogov V.V., Teryev E.D., Fursov V.A., Shamrikof B.M. Development of adaptation and identification algorithms in adaptive digital aircraft controlm systems. - VI IFAC Symposium on Identification and System Parameter estimation. Preprints, Kioto, 1981.
61. Fursov, Vladimir A. Constructing unified identification algorithms using a small number of observations for adaptive control and navigation systems. Journal:Proc. SPIE Vol. 3087, p. 34-44, Navigation and Control Technologies for Unmanned Systems II, Scott A. Speigle; Ed., 1997.
62. Fursov V.A. Identification of distorting systems with monitoring of data capacity. 5-th International Workshop on Digital Image Processing and Computer Graphics. "Image Processing and Computer Optics", Samara, Russia, Aug, 22-26, 1994, Proceedings Samara, 1994. p.2.
63. Fursov, Vladimir A. Identification of optical distorting systems by selecting image informative fragments. Journal:Proc. SPIE Vol. 2363, p. 62-68, 5th International

Workshop on Digital Image Processing and Computer Graphics (DIP-94), Nikolai A. Kuznetsov; Viktor A. Soifer; Eds., 1995.

64. Nikolaev J.A., Petrov B.N., Teryaev E.D., Shamrikov B.M. Design a Parameter Identification Algorithm for Adaptive Aircraft Control System. V IFAC Symposium on Identification and System Parameter Estimation. 1979, Preprints, vol. 2, 1053-1057.

65. Sergeyev, Vladislav V., Fursov, Vladimir A., & Maksimov, M. V. Identification of model parameters and correcting filters for space-variant distortions. Journal:Proc. SPIE Vol. 3348, p. 275-282, Optical Information Science and Technology (OIST97): Computer and Holographic Optics and Image Processing, Andrei L. Mikaelian; Ed., 1998.

66. V.A. Fursov. Parameter identification of the models of distorting systems based on a small number of observation. Pattern recognition and image analysis, Vol. 8, No.2, 1998.

Научное издание

Фурсов Владимир Алексеевич

ИДЕНТИФИКАЦИЯ МОДЕЛЕЙ СИСТЕМ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЙ ПО МАЛОМУ ЧИСЛУ НАБЛЮДЕНИЙ

Компьютерный набор и верстка В.А. Фурсов.

Лицензия ЛР № 020301 от 30.12.96 г.

Подписано в печать 01.12.98 г. Формат 60X84$^{1}/_{16}$.
Бумага газетная. Печать офсетная.
Усл. печ. л. 12,79. Усл. кр.-отт. 12,91. Уч.-изд. л. 13,75.
Тираж 1000 экз. Заказ 4.

Самарский государственный аэрокосмический
университет имени академика С.П. Королева.
443086, Самара, Московское шоссе, 34.

ИПО Самарского государственного аэрокосмического
университета имени академика С.П. Королева.
443001, Самара, ул. Молодогвардейская, 151.