

РОССИЙСКАЯ АКАДЕМИЯ НАУК  
ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР

---

**Г.К. Каменев**

**Метод исследования неопределенности,  
возникающей при идентификации парамет-  
ров моделей**



ВЫЧИСЛИТЕЛЬНЫЙ ЦЕНТР РАН  
МОСКВА 2010

УДК 519.6

Ответственный редактор  
доктор физ. матем. наук А.В.Лотов

Статья посвящена методу исследования неопределенности, возникающей при идентификации параметров моделей. Рассматривается метод множеств идентификации – визуальный подход к идентификации параметров моделей, основанный на построении и визуализации многомерного графика функции ошибок, а также множеств квазиоптимальных параметров. Приводятся примеры исследования неопределенности, возникающей в процессе идентификации.

Ключевые слова: неопределенность, идентификация, параметры, квазиоптимальное множество, функция ошибок, визуализация, принятие решений, многокритериальная оптимизация.

Работа выполнена при частичной финансовой поддержке  
– РФФИ, проекты 10-01-00199, 09-01-00599 и 09-01-13534;  
– ПФИ Президиума РАН №14;  
– ПФИ ОМН РАН №3.

Рецензенты: Н.Н.Оленев, М.А.Хохлов

Научное издание

© Вычислительный центр им. А.А.Дородницына  
Учреждение Российской академии наук, 2010  
© Г.К. Каменев, 2010

## 1. Неопределенность в задачах идентификации

Одной из важнейших в математическом моделировании является проблема идентификации, под которой обычно понимается выбор структуры модели (структурная идентификация) и совокупности значений параметров (идентификация параметров), соответствующих выбранной структуре. Соотнесение структуры данных об объекте моделирования со структурой модели является важнейшим этапом процесса идентификации. Наличие неопределенности является фундаментальным свойством процесса математического моделирования [1]. На этом этапе исследователь часто сталкивается с проблемой неопределенности в виде неоднозначности решения задачи идентификации. «Попытка определить параметры данного уравнения на основе лишь эмпирических наблюдений безнадежна, вне зависимости от того, насколько эти последние полны и обширны» [2, стр. 19]. Поэтому требуется привлечь дополнительную содержательную информацию, которой обладает исследователь.

Пусть имеется модель некоторого объекта, заданная в виде уравнений

$$F(z, \lambda) = 0,$$

где  $F$  – вектор-функция  $\{F_i(\cdot), i=1 \dots m\}$ ,  $z$  – вектор наблюдаемых переменных,  $\lambda$  – неизвестный вектор параметров,  $\lambda \in A$ . Пусть требуется идентифицировать вектор параметров  $\lambda$  на множестве наблюдений  $Z := \{z^k, k = 1, \dots, K\}$ .

Пусть задан некоторый вектор параметров  $\lambda$ . Для этого вектора рассмотрим ошибки идентификации на векторе  $z^k$  из множества наблюдений  $Z$ . Эти ошибки будут в рассматриваемом случае иметь вид вектора невязок

$$w(z^k, \lambda) := F(z^k, \lambda).$$

Чтобы вместо совокупности векторов невязок на всех данных наблюдений оценивать качество идентификации одной (скалярной) характеристикой, вводится сначала некоторая *норма невязки* (метрика)  $\|w(z^k, \lambda)\|$  в пространстве векторов ошибок, например:

$$\|w(z^k, \lambda)\| = \max \{ \mu_i |w_i(z^k, \lambda)| : i=1 \dots m \}$$

или

$$\|w(z^k, \lambda)\| = \left( \sum_{i=1}^m \mu_i (w_i(z^k, \lambda))^2 \right)^{1/2},$$

где  $\mu_i \geq 0$  – весовые коэффициенты, характеризующие значимость (масштаб) отклонения по данному структурному уравнению. Потом выбирается некоторая *функция ошибок идентификации*  $\psi(\lambda)$  – свертка норм невязок по всем данным наблюдений. Например, максимальная по норме ошибка

$$\psi_{\max}(\lambda) := \max \{ p_k \|w(z^k, \lambda)\| : z^k \in Z \}$$

или средняя по норме ошибка

$$\psi_{\text{mean}}(\lambda) := \sum_{z^k \in Z} p_k \|w(z^k, \lambda)\|,$$

где  $p_k \geq 0$ ,  $\sum p_k = 1$ , – весовые коэффициенты, отражающие значимость того или иного наблюдения. Задачу идентификации сводят к нахождению значения  $\lambda^* \in A$ , минимизирующего значение функции ошибок  $\psi(\lambda)$ :

$$\psi(\lambda^*) := \min \{ \psi(\lambda) : \lambda \in A \}.$$

Неопределенность возникает, если решение  $\lambda^*$  зависит от вида и параметров нормы невязки  $\|w(z^k, \lambda)\|$  или от вида и параметров функции ошибок идентификации  $\psi(\lambda)$ . Неопределенность возникает и тогда, когда решение задачи идентификации оказывается неоднозначным или неустойчивым, т.е. существуют достаточно отдаленные от решения  $\lambda^*$  допустимые векторы параметров  $\lambda^{**}$  такие, что  $\psi(\lambda^{**})$  близко к  $\psi(\lambda^*)$ . В случае неопределенности проблема идентификации становится частным случаем задачи принятия решения. Действительно, прогноз поведения системы при  $\lambda = \lambda^*$  может существенно отличаться от прогноза ее поведения при  $\lambda = \lambda^{**}$ , в то время как при  $\psi(\lambda^{**}) \approx \psi(\lambda^*)$  правильность предпочтения  $\lambda^*$  вектору  $\lambda^{**}$  неочевидна. Проблема идентификации состоит в том, что структурные уравнения  $F_i(z^k, \lambda) = 0$  понятны специалистам в данной области моделирования, понятны и невязки  $w_i(z^k, \lambda) := F_i(z^k, \lambda)$ , соответствующие этим уравнениям, однако свертка компонент невязок в норму невязки  $\|w(z^k, \lambda)\|$ , как правило, выходит далеко за рамки их интуитивных представлений. В еще большей степени это касается функции ошибок  $\psi(\lambda)$ , являющейся сверткой норм невязок на данных

наблюдений.

Задача идентификации модели в регрессионной (приведенной) форме основана на том, что вектор наблюдаемых переменных  $z$  удастся разделить на векторы независимых (входных)  $x$  и зависимых (выходных) величин  $y$ . В этом случае идентификация есть отождествление модели с объектом в смысле поиска оценки оператора, связывающего выход модели некоторого объекта со входом [3].

Пусть имеется приведенная модель некоторого объекта вида

$$y = f(x, \lambda),$$

где  $f$  – вектор-функция,  $x$  – вектор независимых (входных),  $y$  – вектор зависимых (выходных) переменных,  $\lambda$  – неизвестный вектор параметров,  $\lambda \in A$ . Требуется идентифицировать вектор параметров  $\lambda$  на некотором множестве наблюдений

$$Z := \{(x^k, y^k), k = 1, \dots, K\}.$$

Для вектора параметров  $\lambda$  вектор невязки на паре  $(x^k, y^k)$  из множества наблюдений  $Z$  будет в этом случае иметь вид

$$w(x^k, y^k, \lambda) := y^k - f(x^k, \lambda).$$

Как и при задании модели в структурной форме, в данном случае вводится сначала некоторая норма невязок  $\|y^k - f(x^k, \lambda)\|$  в пространстве векторов ошибок, а затем – некоторая функция ошибок идентификации  $\psi(\lambda)$  – свертка норм невязок по всем данным наблюдений. Как и при задании модели в структурной форме, задача сводится к минимизации функции ошибок  $\psi(\lambda)$  на множестве допустимых значений вектора параметров  $A$ . При идентификации регрессионного описания модели возникают те же источники неоднозначности решения, что и при структурном описании. Однако в то время как точность входных переменных часто можно контролировать, точность наблюдения выходных переменных контролировать не всегда удается.

В этом случае ошибка идентификации может отражать не только неправильный выбор значения параметра или неточность структурного описания модели, но и влияние ошибки наблюдений. Запишем регрессионную модель в виде, учитывающем неточность наблюдений за выходными переменными:

$$y + \varepsilon = f(x, \lambda),$$

где  $\varepsilon$  – неизвестный вектор ошибок наблюдений. Если априори из-

вестно, что  $\varepsilon$  – случайная величина с известным распределением и число наблюдений достаточно велико, то можно перейти к регрессионным уравнениям для математических ожиданий выходных переменных и, таким образом, снять неопределенность, вызываемую ошибками наблюдений.

Часто бывает и так, что исследователь имеет представление только о максимальной величине ошибок наблюдений. Так как пространство векторов невязок  $w$  совпадает с пространством выходных переменных  $y$ , то в этом пространстве можно ввести норму невязки  $\|y^k + \varepsilon^k - f(x^k, \lambda)\|$ , где  $y^k, k=1, \dots, K$ , – неизвестный вектор точных значений выходных переменных, а  $\varepsilon^k, k=1, \dots, K$ , – неизвестный вектор ошибок, соответствующих множеству наблюдений  $Z := \{(x^k, y^k + \varepsilon^k), k=1, \dots, K\}$ . Введем и функцию ошибок  $\psi_\varepsilon(\lambda)$ , основанную на этой норме. Пусть  $\lambda_{\varepsilon^*} \in \Lambda$  такой, что

$$\psi_\varepsilon(\lambda_{\varepsilon^*}) = \min \{ \psi_\varepsilon(\lambda) : \lambda \in \Lambda \}.$$

Спрашивается, какое отношение решение  $\lambda_{\varepsilon^*}$  задачи минимизации «возмущенной» функции ошибок  $\psi_\varepsilon(\lambda)$  имеет к решению  $\lambda^*$  задачи минимизации функции ошибок  $\psi(\lambda)$  при условии, что  $\varepsilon^k=0$ ? Чтобы ответить на этот вопрос, заметим, что для большинства функций ошибок справедливо свойство

$$\psi_\varepsilon(\lambda) \leq \psi(\lambda) + \varepsilon_{\max}, \quad \psi(\lambda) \leq \psi_\varepsilon(\lambda) + \varepsilon_{\max},$$

где  $\varepsilon_{\max} := \max \{ \|\varepsilon^k\| : (x^k, y^k + \varepsilon^k) \in Z \}$ . Поэтому

$$\psi_\varepsilon(\lambda_{\varepsilon^*}) \leq \psi(\lambda_{\varepsilon^*}) + \varepsilon_{\max} \leq \psi(\lambda_{\varepsilon^*}) + \varepsilon_{\max} \leq \psi_\varepsilon(\lambda_{\varepsilon^*}) + 2\varepsilon_{\max}.$$

Следовательно, если известна только максимальная величина  $\varepsilon_{\max}$  ошибок наблюдений, то можно только утверждать, что решение  $\lambda_{\varepsilon^*}$  задачи минимизации функции ошибок  $\psi_\varepsilon(\lambda)$  при условии, что  $\varepsilon^k=0$ , лежит в области значений  $\lambda$  таких, что  $\psi_\varepsilon(\lambda) \leq \psi_\varepsilon(\lambda_{\varepsilon^*}) + 2\varepsilon_{\max}$ .

Аналогично можно учесть ошибки наблюдений входных переменных, если известна, например, константа Липшица оператора  $f$ . Таким образом, неточность в измерении наблюдаемых переменных может быть источником неопределенности при идентификации параметров.

Итак, идентификация параметров, основанная на минимизации функции свертки ошибок, строго обоснована при следующих предположениях: 1) известна структура математической модели; 2) чис-

ло наблюдений асимптотически стремится к бесконечности; 3) наблюдаемые переменные – случайные величины известных типов; 4) найден глобальный минимум функции ошибок, а решение задачи минимизации не вырождено. Широко используемые устойчивые методы обработки данных основываются на этих предположениях. Но во многих задачах идентификации эти предположения не верны. Поэтому в случае неопределенности участие человека в решении задачи идентификации становится неизбежным.

В статье предлагается подход к идентификации математических моделей, основанный на наглядном представлении графика функции ошибок  $\psi(\lambda)$  в окрестности минимума. Исследователь может не ограничиваться единственным сочетанием значений идентифицируемых параметров, а визуально исследовать график функции ошибок, который отражает качественный характер соответствия структуры данных структуре модели. Он получает больше информации для выбора значений параметров модели, на которых будут основаны оценки с помощью данной математической модели.

## 2. Метод достижимых целей

С начала 1980-х г. в Вычислительном центре РАН разрабатывается новый подход к проблеме многокритериальной оценки решений с помощью математических моделей: метод достижимых целей (МДЦ) [4-6].

В рассматриваемых здесь задачах идентификации принимается решение о выборе одного из допустимых значений вектора параметров  $\lambda$ , а критериями качества решения являются значения одной или нескольких функций ошибок  $\psi$ , например, – нескольких компонент невязок  $w_i(z^k, \lambda)$ . По МДЦ с помощью математической модели рассчитывается множество достижимых целей (совокупность достижимых сочетаний значений критериев, оценивающих решения) и представляется пользователю в графическом виде. Важно то, что МДЦ позволяет получить наглядную информацию не только о достижимых значениях критериев, но и о разумных компромиссах между критериями: они задаются паретовской частью границы множества достижимых целей. Это – эффективная граница множества достижимых целей, она показывает, на сколько требуется ухудшить зна-

чение одного критерия для улучшения значения другого. По МДЦ, выбрав точку на эффективной границе (предпочтительную достижимую цель), пользователь получает и решение, реализующее эту цель. Пользователь выбирает решение из множества допустимых вариантов по собственному неформализованному предпочтению одного из разумных компромиссов между критериями. Он может выбрать несколько целевых точек на эффективной границе и подвергнуть сравнительному анализу несколько решений.

Предложенный подход требует заранее построить явное описание множества достижимых в силу модели значений вектора критериев и интерактивно изобразить его средствами современной компьютерной графики. К настоящему времени разработаны численные методы аппроксимации множества достижимых целей и накоплен большой опыт применения визуальной технологии МДЦ для поддержки решений с помощью нелинейных моделей экономических, экологических, технических и управляемых систем [4-9].

В [10] предложено использовать метод, аналогичный МДЦ, для решения задачи идентификации. В этом случае объектом исследования становится не множество достижимых целей, а график и надграфик функции ошибок, а также его поведение в окрестности оптимальных параметров. Визуализация этого множества позволяет исследователю выбирать параметры модели, сочетая результаты наблюдений с собственным неформализованным опытом, дает ему дополнительную информацию об идентифицируемой модели (см. примеры в [10-15]).

Для метода, предложенного в [10], нужна модификация метода достижимых целей. В настоящей статье предлагается визуальный метод множеств идентификации (ММИ). Он основан на явном представлении графика и надграфика функции ошибок и использовании их визуализации для идентификации модели в диалоговом режиме.

### **3. Метод множеств идентификации (ММИ)**

Рассмотрим регрессионную модель некоторого объекта (случай структурной модели рассматривается аналогично). Пусть  $y = f(x, \lambda)$ , где  $x$  – входные,  $y$  – выходные переменные,  $\lambda$  – неизвестный вектор параметров,  $x \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$ ,  $\lambda \in \Lambda \subset \mathbb{R}^p$ . Пусть требуется идентифи-

цировать вектор параметров  $\lambda$  на некотором множестве наблюдений  $Z := \{(x^k, y^k), k = 1, \dots, K\}$ , используя в качестве критерия погрешности идентификации скалярную функцию ошибок идентификации  $\psi: \Lambda \rightarrow \mathbb{R}^1$ , являющуюся некоторой сверткой норм ошибок  $\|y^k - f(x^k, \lambda)\|$  на множестве наблюдений  $Z$ , и пусть  $\psi(\lambda^*) := \min \{ \psi(\lambda) : \lambda \in \Lambda \}$ .

Такой подход, разумный во многих случаях, при наличии неопределенности требует дополнительного исследования. В [17] указывается, что исследователям было бы разумно представить информацию о квазиоптимальных значениях параметров, для которых значение функции ошибок идентификации не превышает некоторую заданную величину, т.е. таких  $\lambda \in \Lambda$ , что  $\psi(\lambda) \leq \delta$ , где  $\delta \geq \psi^* := \psi(\lambda^*)$ .

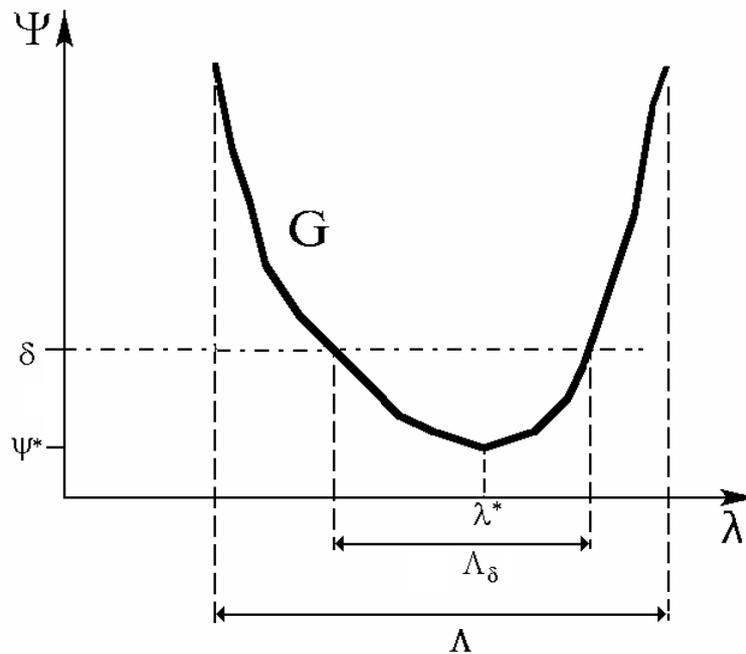


Рис. 1

Рассмотрим *график функции ошибок*  $\psi(\lambda)$ , т.е. множество

$$G := \{(\lambda, \psi(\lambda)) \in \Lambda \times \mathbb{R}^1 : \lambda \in \Lambda\}.$$

График  $G$  представляет собой некоторую, вообще говоря, много-

мерную поверхность в пространстве  $\mathbb{R}^{p+1}$ . Если имеется только один параметр ( $p=1$ ), это множество представляет собой обычный график (рис. 1).

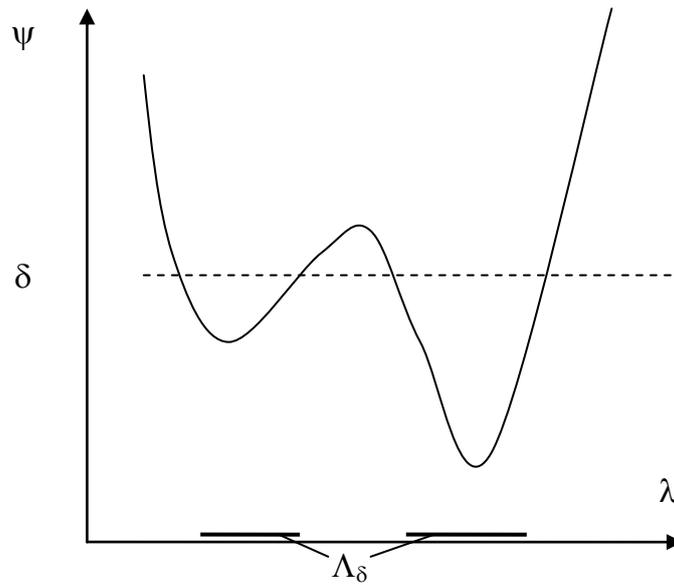


Рис. 2

Рассмотрим также множество

$$A_\delta := \{\lambda \in A: \psi(\lambda) \leq \delta\}.$$

Для  $\delta < \psi^*$  это множество пусто. При  $\delta \geq \psi^*$  оно состоит из значений  $\lambda \in A$ , для которых значение функции ошибок идентификации не превосходит  $\delta$ . Множество  $A_\delta$  представляет собой проекцию части графика функции ошибок, лежащей ниже уровня  $\psi(\lambda) = \delta$ , в пространство параметров  $\lambda$ . Назовем множество  $A_\delta$  *множеством квазиоптимальных параметров уровня  $\delta$*  (квазиоптимальным множеством). Нетрудно видеть, что множество  $A_\delta$  будет являться *сечением* надграфика функции ошибок  $\psi(\lambda)$ , т.е. множества

$$G^+ := \text{epi } G := \{(\lambda, \sigma) \in A \times \mathbb{R}^1: \lambda \in A, \sigma \geq \psi(\lambda)\},$$

гиперплоскостью  $\psi(\lambda) = \delta$ .

Если имеется только один параметр, множество квазиоптималь-

ных параметров  $\Lambda_\delta$  представляет собой подмножество оси абсцисс (рис.1). Если функция ошибок имеет несколько локальных минимумов, то даже в случае одного параметра множество  $\Lambda_\delta$  может быть неодносвязным, т.е. представлять из себя набор отрезков (рис. 2).

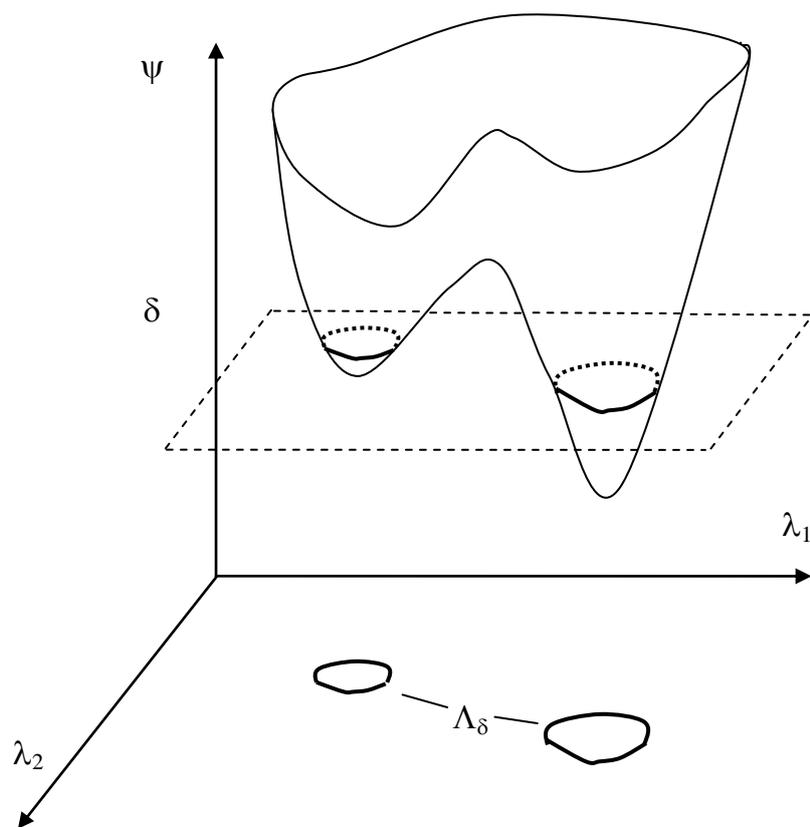


Рис. 3

Если параметров несколько, исследование многомерного множества квазиоптимальных параметров представляет для человека значительную проблему. Она решается рассматриваемым нами ММИ с помощью *диалоговых карт идентификации (ДКИ)*, обеспечивающих интерактивную визуализацию *множеств идентификации*: гра-

фика функции ошибок и множеств квазиоптимальных параметров.

Итак, основными этапами ММИ являются:

- 1) предварительная аппроксимация множеств идентификации простыми фигурами;
- 2) визуализация аппроксимации на основе диалоговых карт идентификации.

#### 4. Диалоговые карты идентификации

Рассмотрим сначала основные методы визуализации с помощью диалоговых карт идентификации. Будем считать, что простое явное описание графика функции ошибок (его аппроксимация) построено, и мы можем достаточно быстро изобразить его различные двумерные сечения, т.е. множества, в которых все, кроме двух параметров, зафиксированы, и срезы, т.е. множества, в которых на все, кроме двух параметров, наложены ограничения типа неравенств. Проиллюстрируем некоторые особенности рассматриваемого нами подхода на примере функции ошибок с различным числом параметров.

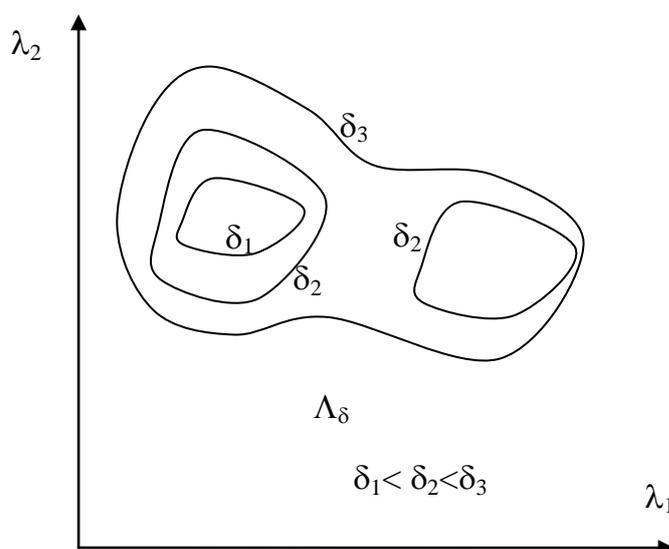


Рис. 4

На рис. 3. представлено изображение графика функции ошибок и квазиоптимальное множество для случая двух параметров.

Квазиоптимальные множества  $A_\delta$  (срезы надграфика функции ошибок) удобно изображать в координатах двух параметров идентификации линиями уровня, соответствующими серии растущих значений уровня функции ошибок  $\delta$  (см. пример на рис. 4). Заметим, что в силу определения квазиоптимального множества  $A_{\delta_1} \subset A_{\delta_2}$  при  $\delta_1 \leq \delta_2$ , поэтому линии с меньшим уровнем ошибки оказываются вложенными в линии с большим уровнем. Это свойство характерно для географических карт, а также для диалоговых карт решений, используемых в ММД [4, 5]. По аналогии изображения, используемые в ММИ, мы будем называть *картами функции ошибок*.

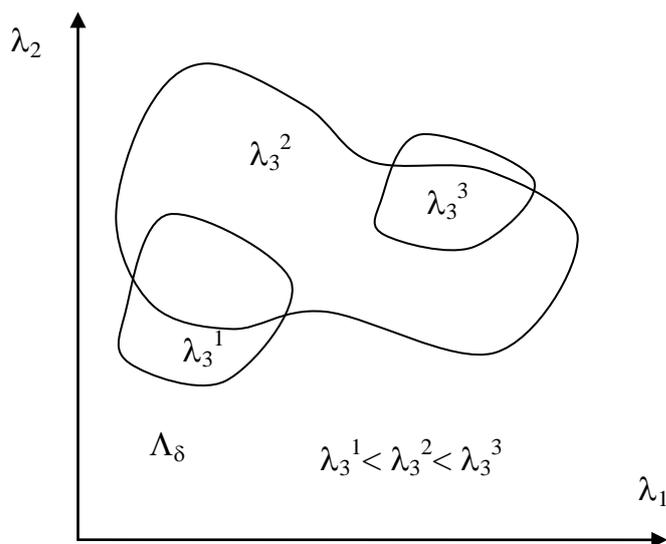


Рис. 5

В случае трех параметров квазиоптимальное множество конкретного уровня  $\delta$  можно изобразить в координатах двух параметров идентификации линиями уровня, соответствующими серии растущих значений третьего параметра (см., например, рис. 5). В этом

случае, однако, нельзя гарантировать вложенности линий уровня, как на картах функции ошибок (рис. 4). Такие изображения мы будем называть *картами параметров*.

В многомерном случае множества идентификации можно представить проекциями их на подпространства, содержащие лишь часть параметров – тех, которые больше интересуют исследователя.

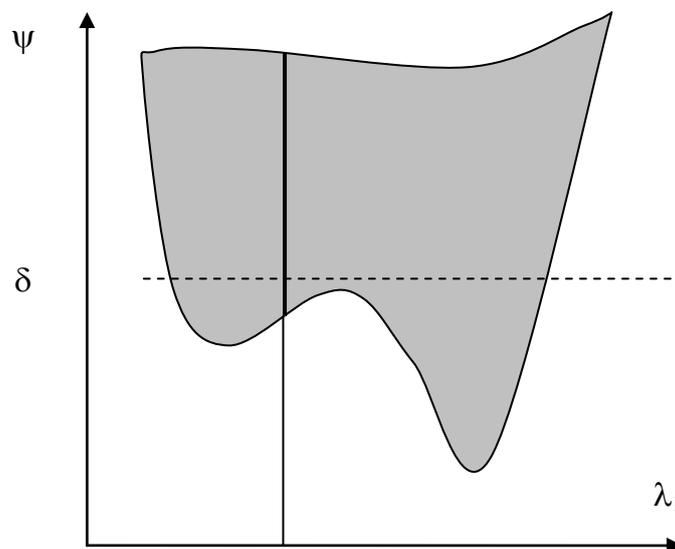


Рис. 6

Например, на рис. 6 изображена проекция графика функции ошибок с рис. 3 в пространство, содержащее лишь первый параметр  $\lambda_1$ , т.е. множество

$$\{(\lambda_1, \psi(\lambda_1, \lambda_2)) \in \mathbb{R}^2: (\lambda_1, \lambda_2) \in A\}.$$

Заметим, что, в отличие от графика на рис. 1, на этом рисунке изображено двумерное множество, так как каждому значению, например, первого параметра соответствуют различные значения второго параметра с соответственно различными значениями функции ошибок, заполняющими собою целый отрезок.

Визуализация, основанная на изображениях проекций, обладает,

однако, тем недостатком, что скрывает неоднозначность, связанную с параметрами, не представленными в данной проекции. Более информативны срезы графика функции ошибок, потому что на значения тех параметров, которые не интересуют исследователя, наложены заданные ограничения.

В рассматриваемом трехмерном случае это будут, например, изображения множеств вида

$$\{(\lambda_1, \psi(\lambda_1, \lambda_2)) \in \mathbb{R}^2: (\lambda_1, \lambda_2) \in A, \lambda_2^* \leq \lambda_2 \leq \lambda_2^{**}\}$$

при различных значениях пороговых ограничений  $\lambda_2^*$  и  $\lambda_2^{**}$ .

Если параметров больше двух, основным средством визуализации множеств идентификации становятся так называемые диалоговые карты идентификации (ДКИ). ДКИ – это карты идентификации, на которых исследователь может интерактивно выбирать состав параметров и ограничения на их значения, а живость диалога обеспечивается быстротой визуализации, основанной на предварительной аппроксимации графика функции ошибок.

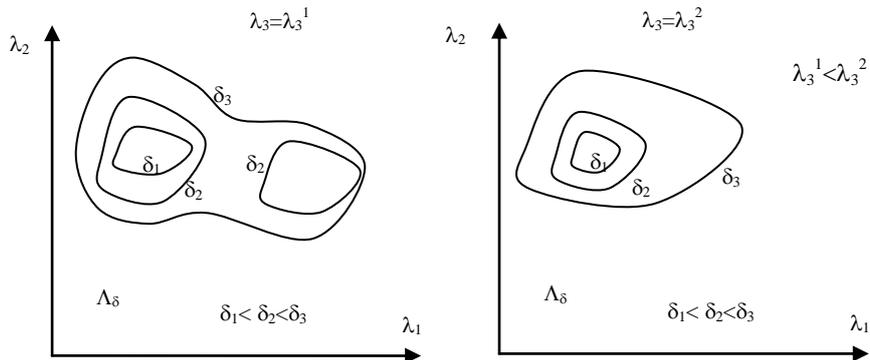


Рис. 7

Например, в случае трех параметров график функции ошибок можно представить совокупностью (например, строкой) карт функции ошибок, на каждой из которых изображено множество  $\Lambda_\delta$  при различных значениях  $\delta$  в пространстве двух выбранных параметров, а значение третьего параметра монотонно меняется от карты к карте (рис. 7). Вместо набора карт можно изобразить только одну: в ней значение третьего параметра можно менять в реальном времени (в виде анимации), либо выбирать на шкале (полосе прокрутки), изоб-

раженной на карте.

В случае четырех параметров можно изобразить совокупности карт функции ошибок, соответствующих различным значениям четвертого параметра, например в виде матриц изображений, аналогичных матрицам карт решений [4, 5].

Случай пяти переменных можно представить в виде атласа (набора листов или вкладок) карт функции ошибок, соответствующих различным значениям пятого параметра.

В типичном случае идентификации модели исследуется множество квазиоптимальных параметров при заданном (или интерактивно изменяемом) уровне ошибки. Для визуализации таких множеств можно применить технологию диалоговых карт, описанную выше. Наглядное представление квазиоптимального множества дает карта (строка, матрица, атлас или динамический объект из карт) параметров, т.е. карта, на которой изображено множество  $\Lambda_\delta$  при фиксированном уровне функции ошибки  $\delta$  в пространстве двух выбранных параметров и различных значениях третьего (рис. 8).

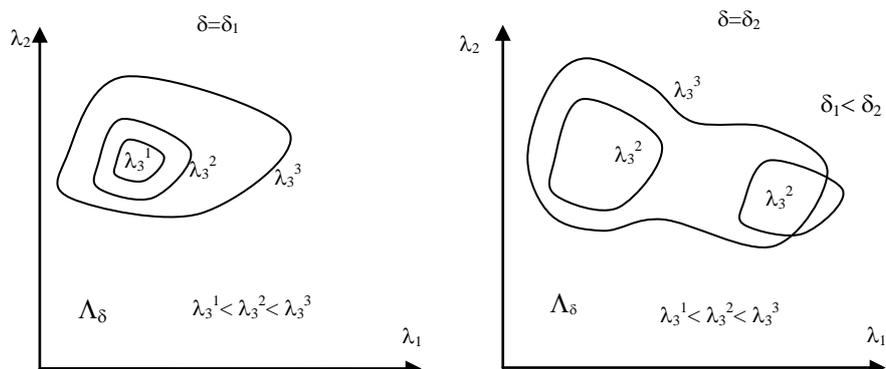


Рис. 8

В общем случае множество идентификации (график функции ошибок или квазиоптимальное множество) визуализируется или в виде *срезов* (когда значения части параметров не интересуют исследователя, но на них наложены дополнительные ограничения), или в виде *сечений* (когда значения части параметров точно заданы). В любом случае визуализация осуществляется исследователем в ин-

тенсивном диалоге с компьютером, что обеспечивается предварительной аппроксимацией визуализируемого множества.

## 5. Исследование функции ошибок

Пусть имеется возможность тем или иным способом изобразить график функции ошибки  $G$  или квазиоптимальные множества параметров  $A_\delta$  при различных уровнях  $\delta$  и представить их для визуального исследования экспертам. Следует рассмотреть две ситуации. В одной исследователь имеет некоторое представление об уровне ошибки, которая может наблюдаться в моделируемой действительности. В другой ситуации, более вероятной, эксперт ничего не может сказать об уровне ошибки, выражаемой функцией  $\psi$ , однако имеет некоторые заранее трудно формализуемые представления о возможных значениях параметров  $\lambda$ . В первом случае исследователя могут заинтересовать множества  $A_\delta$ , соответствующие допустимым, с его точки зрения, значениям функции ошибок  $\delta$ ; во втором – его может заинтересовать устойчивость множества квазиоптимальных решений  $A_\delta$  при значениях  $\delta$ , близких к минимальному, и наличие в

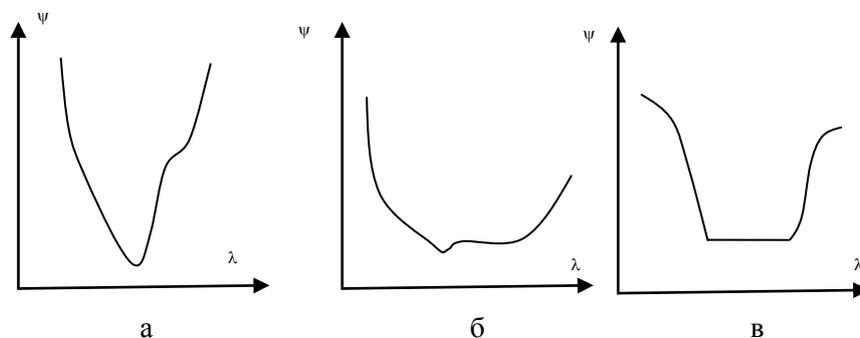


Рис. 9

нем допустимых, с его точки зрения, значений параметров  $\lambda$ .

Поведение функции ошибок вблизи точки оптимума имеет существенное значение для обоснованности решения задачи об идентификации. На рис. 9 представлены типичные случаи поведения графика функции ошибок вблизи глобального минимума:

- *устойчивое* (рис. 9, а), когда при небольшом увеличении уровня функции ошибок расширение квазиоптимального множества не очень значительно;
- *неустойчивое* (рис. 9, б), когда при небольшом увеличении уровня функции ошибок происходит резкое расширение квазиоптимального множества;
- *вырожденное* (рис. 9, в), когда глобальному минимуму функции ошибок отвечает целое множество параметров.

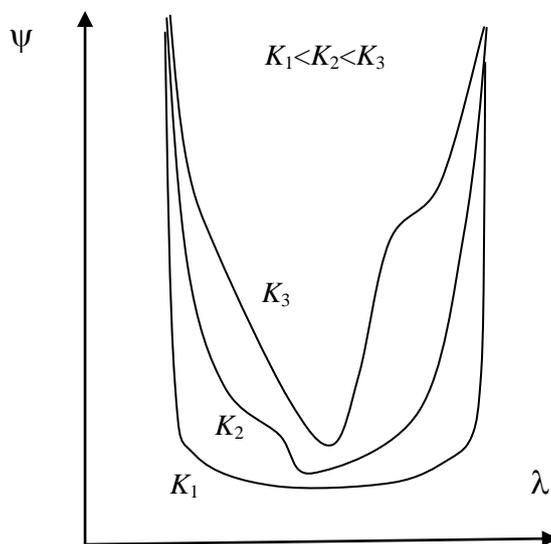


Рис. 10

По мере накопления данных (увеличения числа наблюдений  $K$ ) неустойчивое поведение графика функции ошибок может превращаться в устойчивое. Такой случай изображен на рис. 10. Это может свидетельствовать о соответствии идентифицируемой модели объекту наблюдения и перспективности дальнейших наблюдений для улучшения качества идентификации. Может, разумеется, наблюдаться и обратная картина.

На основе полученной визуальной информации и опыта исследователь должен выбрать некоторое подмножество значений параметров из  $\Lambda$  в качестве решения задачи идентификации. Например,

он может выбрать некоторую пороговую величину уровня ошибки  $\delta^*$  и множество  $A_{\delta^*}$  квазиоптимальных решений уровня  $\delta^*$ . Также можно выбрать один или несколько характерных вариантов сочетаний значений параметров (не обязательно совпадающих с точкой минимума функции ошибок  $\lambda^*$ ). Разумеется, исследователь может выбрать точку  $\lambda^*$  минимума функции ошибок, особенно тогда, *если этот выбор оправдан устойчивостью графика функции ошибок вблизи оптимума, гипотетической адекватностью модели и высокой точностью данных наблюдений.*

Заметим, что не обязательно выбирать в качестве решения задачи идентификации единственное сочетание значений параметров. В общем случае результатом может быть *множество таких сочетаний значений идентифицируемых параметров, которое исследователь считает необходимым учесть при оценке последствий использования данной математической модели.* Результатом исследования может быть и более глубокое проникновение в свойства рассматриваемой модели *на конкретных данных наблюдения*, а также соотнесение структуры данных со структурой модели.

## **6. Аппроксимация множеств идентификации в выпуклом случае**

Теперь рассмотрим проблему аппроксимации множеств идентификации. Выше было отмечено, что визуализация принципиально интерактивна, так как заранее неизвестно, какая область квазиоптимальных решений заинтересует исследователя. Поэтому она должна осуществляться достаточно быстро (если не сказать мгновенно), при этом исследователь должен иметь возможность уточнить построенное описание до необходимой точности. Однако расчет по модели при заданном сочетании значений параметров  $\lambda$  (например, вычисление модельной функции  $f(x, \lambda)$  на множестве наблюдений независимых переменных  $\{x^k\}$ ) может занимать значительное время. Тогда этот предварительный этап идентификации надо проводить на высокопроизводительных, в том числе параллельных, вычислительных машинах. После него визуализация проходит в интенсивном интерактивном (диалоговом) режиме. На предварительном автономном этапе должен быть проведен расчет функции ошибок для всех до-

пустимых сочетаний значений идентифицируемых параметров, или должна быть осуществлена аппроксимация множеств идентификации с точностью, достаточной для содержательного визуального анализа. Ясно, что это очень непростая задача. Аналогичная задача возникала и в МДЦ, поэтому значительные усилия были направлены на разработку методов аппроксимации множеств, заданных неявно с помощью многомерных отображений [18, 4, 5]. Изложим некоторые из этих методов в контексте ММИ.

Сначала рассмотрим случай, когда надграфик функции  $\psi(\lambda)$ , множество  $G^+$ , является выпуклым множеством. В этом случае если  $A$  выпукло, то множества  $A_\delta$ , которые являются сечениями  $G^+$ , также будут выпуклы.

Рассмотрим пример такого подхода. Пусть уравнения модели имеют линейный вид

$$y = Ax + b,$$

где коэффициенты  $\{a_{i,j}\}$  ( $i=1, \dots, m; j=1, \dots, n$ ) и  $\{b_i\}$  ( $i=1, \dots, m$ ) являются искомыми параметрами ( $A$  совпадает с  $\mathbb{R}^{m(n+1)}$ ). Пусть норма на  $\mathbb{R}^m$  имеет вид  $\|y\| = \max\{|y_i|: i=1, \dots, m\}$  и  $\psi(\lambda) := \psi_{\max}(\lambda)$ . Нетрудно видеть, что множество  $G^+$  в этом случае является многогранным множеством в пространстве  $\mathbb{R}^{m(n+1)+1}$  и совпадает с множеством решений системы

$$\begin{cases} \delta \geq y^k - Ax^k - b, & k = 1, \dots, K; \\ \delta \geq -y^k + Ax^k + b, & k = 1, \dots, K. \end{cases}$$

Аналогичная система может быть выписана и для случая  $\psi(\lambda) := \psi_{\text{mean}}(\lambda)$ .

Для аппроксимации надграфика  $G^+$  или его проекций в выпуклом случае можно использовать методы полиэдральной аппроксимации выпуклых тел (см. обзоры в [19, 18, 4, 11, 5, 20, 21]). Практически применяемые оптимальные методы аппроксимации выпуклых тел основаны, как правило, на итерационном уточнении аппроксимирующего многогранника. На каждой итерации многократно вычисляется опорная функция аппроксимируемого множества (решается задача выпуклой оптимизации), поэтому алгоритмы этих методов можно эффективно распараллелить.

## 7. Аппроксимация множеств идентификации в нелинейных моделях

Теперь рассмотрим нелинейный случай. Универсальным способом «кодирования» невыпуклых множеств является построение метрических сетей. Геометрически это означает, что множество аппроксимируется набором простых фигур (таких, как шар или куб) с диаметром тем меньшим, чем выше точность аппроксимации.

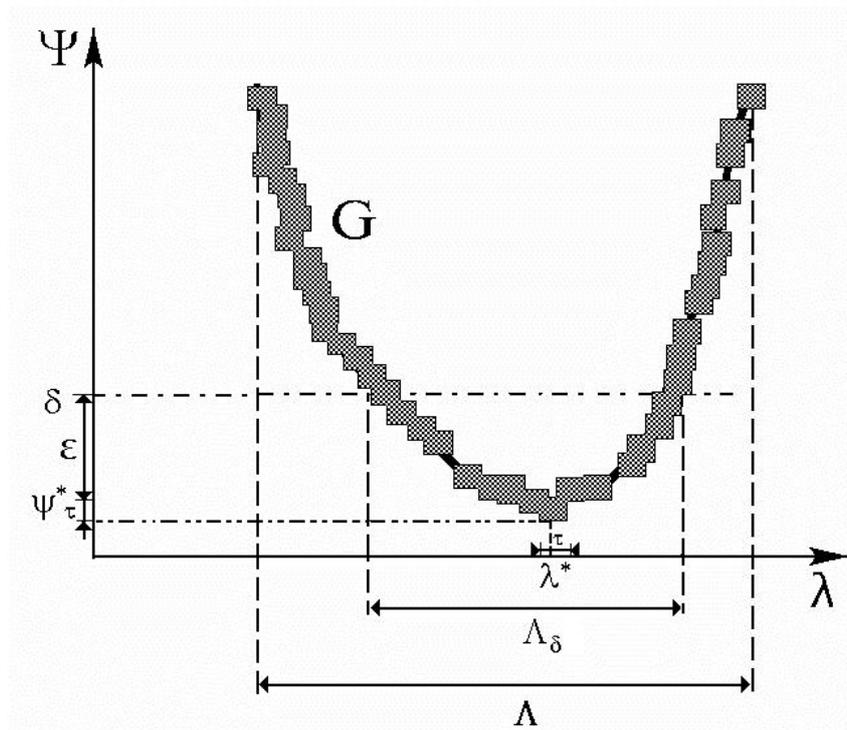


Рис. 11

Под метрической  $\epsilon$ -сетью множества  $A$  понимается такое его подмножество  $T$ , что любая точка  $A$  удалена от  $T$  на расстояние не больше  $\epsilon$ . Если вокруг точек метрической  $\epsilon$ -сети построить «шары» в рассматриваемой метрике пространства, то они покроют все аппроксимируемое множество. В евклидовой метрике с нормой  $\|z\| = (\sum z_j^2)^{1/2}$  это – аппроксимация множества системой обычных ша-

ров; а в метрике Чебышева с нормой  $\|z\| = \max\{|z_j|\}$  – системой кубов. Для каждой точки такой системы можно быстро найти «подлинную» точку аппроксимируемого множества  $A$  (ближайшую точку метрической  $\varepsilon$ -сети  $T$ ), отстоящую от нее не более чем на  $\varepsilon$  в рассматриваемой метрике. При этом чем меньше величина  $\varepsilon$ , тем ближе будут «подлинные» точки из  $T$  к точкам аппроксимируемого множества, тем меньше будет «лишних» точек в аппроксимирующей системе «шаров», но тем больше требуется построить точек в используемой метрической сети  $T$ .

Если известна константа Липшица функции  $\psi(\lambda)$ , для аппроксимации графика ошибки идентификации можно использовать сетки (равномерные или псевдо-случайные) на множестве допустимых параметров  $A$ . В более общем случае могут быть использованы адаптивные стохастические методы аппроксимации ограниченных множеств [11, 10, 22] (см. также [4-6]). С помощью этих методов на аппроксимируемом множестве, неявно заданном с помощью нелинейного отображения, строится метрическая сеть, аппроксимирующая его с заданной точностью и полнотой. Полнота аппроксимации характеризуется тем, что остается лишь небольшая часть аппроксимируемого множества, лежащая вне построенной системы шаров с центрами в точках метрической сети.

На рис. 11 представлен пример достаточно полного покрытия множества  $G$  «шарами» в метрике Чебышева (т.е. кубами) в случае единственного параметра (см. исходное множество  $G$  на рис. 1). «Радиус» «шаров» в системе, т.е. длина ребра куба, соответствует точности аппроксимации и составляет на рисунке величину  $\tau$ . Из рисунка легко понять вид множеств квазиоптимальных параметров  $A_\delta$  при величинах  $\varepsilon = \delta - \psi^*$ , больших  $\tau$ . При малых же  $\varepsilon$  точность аппроксимации должна быть достаточно высокой. Вместе с тем такие качественные характеристики множества квазиоптимальных решений, как устойчивость, становятся ясными и при  $\varepsilon \sim \tau$ .

## 8. Общий случай применения ММИ

Предложенная методика может быть распространена и на случай, когда имеется несколько функций ошибок идентификации, об-

разующих вектор-функцию ошибок  $\Psi: A \rightarrow \mathbb{R}^s$ , а также на случай, когда вместо параметров модели требуется идентифицировать значение вектор-функции от параметров  $\Xi: A \rightarrow \mathbb{R}^t$ .

В общем виде задача сводится к аппроксимации и визуальному анализу множества

$$G^{\psi\xi} := \{(\psi, \xi) \in \mathbb{R}^{s+t}: \lambda \in A, \psi = \Psi(\lambda), \xi = \Xi(\lambda)\}$$

и может быть решена с помощью изложенной выше методологии ММИ.

Например, в линейной регрессии исследователя могут интересовать не коэффициенты матрицы  $\{a_{i,j}\}$ , а ее след, собственные числа или какая-либо норма. Другой пример рассмотрен ниже в задаче об идентификации параметров газовой залежи. В этой задаче параметрами модели являлись объемы газа в небольших частях месторождения, однако исследователя интересовала проблема идентификации суммарного объема газа.

## 9. Простой пример нелинейной регрессии

Проиллюстрируем предложенную визуальную технологию идентификации простым характерным примером нелинейной регрессии. Пусть уравнение модели имеет вид  $y = f(x, \lambda)$ , где

$$f(x, \lambda) := e^{\gamma x} (\alpha + \sin(\omega x + \varphi)),$$

$\lambda = (\gamma, \alpha, \omega, \varphi)$  - неизвестные параметры. Требуется идентифицировать значения, например, параметров  $\omega$  и  $\gamma$  на множестве наблюдений

$$Z = \{(k, e^{-0.05k} (1 + \sin(0.5k + \pi)) + \xi), k=1, \dots, N\},$$

где  $\xi$  - случайная величина, равномерно распределенная на отрезке  $[-0.25, 0.25]$ ,  $N=8$ . Таким образом, «истинное» сочетание параметров есть точка

$$\lambda^* = (-0.05, 1, 0.5, \pi).$$

Предположим, что исследователю неизвестен закон распределения величины  $\xi$ , либо нет возможности воспользоваться средствами усреднения из-за малого числа наблюдений. Пусть для параметров имеется следующая априорная локализация значений:

$$\gamma \in [-0.1, -0.01], \alpha \in [0, 2], \omega \in [0, 1], \varphi \in [0, 2\pi].$$

Пусть функция ошибки имеет вид

$$\psi_{\max}(\lambda) := \max \{|y - f(x^k, \lambda)|: (x^k, y^k) \in Z\}.$$

На рис. 12 изображены проекции в пространство значений параметров  $(\omega, \gamma)$  квазиоптимальных множеств, соответствующих различным значениям  $\delta \leq 0.75$ . Квазиоптимальные множества, соответствующие разным  $\delta$ , различаются штриховкой (на экране – цветом), шкала которой приведена вверху рисунка. Множества, соответствующие меньшей величине  $\delta$  (более темные), наложены на множества, соответствующие большей ее величине. Для иллюстрации методики использована достаточно грубая аппроксимация – на рисунке видны прямоугольники, совокупность которых покрывает график ошибки идентификации с заданной полнотой.

Прежде всего заметим, что минимум «возмущенной» функции ошибок достигается при  $\tilde{\lambda}^{**} = (\omega^{**}, \gamma^{**}) = (0.45, 0.02)$ , тогда как «истинная» точка есть  $\tilde{\lambda}^* = (\omega^*, \gamma^*) = (0.5, 0.05)$ , при этом  $\psi(\lambda^{**}) = 0.18 \leq \psi(\lambda^*) \approx 0.25$ . Проекция  $\tilde{\lambda}^{**} = (\omega^{**}, \gamma^{**})$  точки  $\lambda^{**}$  принадлежит множеству, соответствующему уровню ошибки 0.23, так как существуют сочетания параметров  $(\alpha, \varphi)$ , отличные от  $(\alpha^*, \varphi^*)$ , при которых значение «возмущенной» функции ошибки идентификации меньше, чем  $\psi(\lambda^*)$ .

Из рисунка видно, что при уровне функции ошибки в  $\delta = 0.23$  «истинное» сочетание параметров расположено близко к нижней границе квазиоптимального множества. С увеличением величины  $\delta$  проекция  $\tilde{\lambda}^*$  точки  $\lambda^*$  попадает в середину множества допустимых сочетаний параметров. Заметим, что это множество покрывает почти весь исходный интервал локализации параметра  $\gamma$  и в то же время более узко по  $\omega$ . Это означает, что решение задачи идентификации по параметру  $\omega$  устойчивее относительно ошибки, чем относительно параметра  $\gamma$ . Исследователь может варьировать параметр  $\gamma$  в достаточно широких пределах.

Выбор возможного значения параметра  $\gamma$  влияет на множество допустимых значений  $\omega$ . Так, при уровне функции ошибки  $\delta = 0.23$  если  $\gamma \approx 0.01$ , то  $\omega \in [0.2, 0.55]$ , а если  $\gamma \approx 0.06$ , то  $\omega \in [0.4, 0.65]$ . Пример показывает, что дополнительная информация о значении одного из параметров может существенно повлиять на результат идентификации.

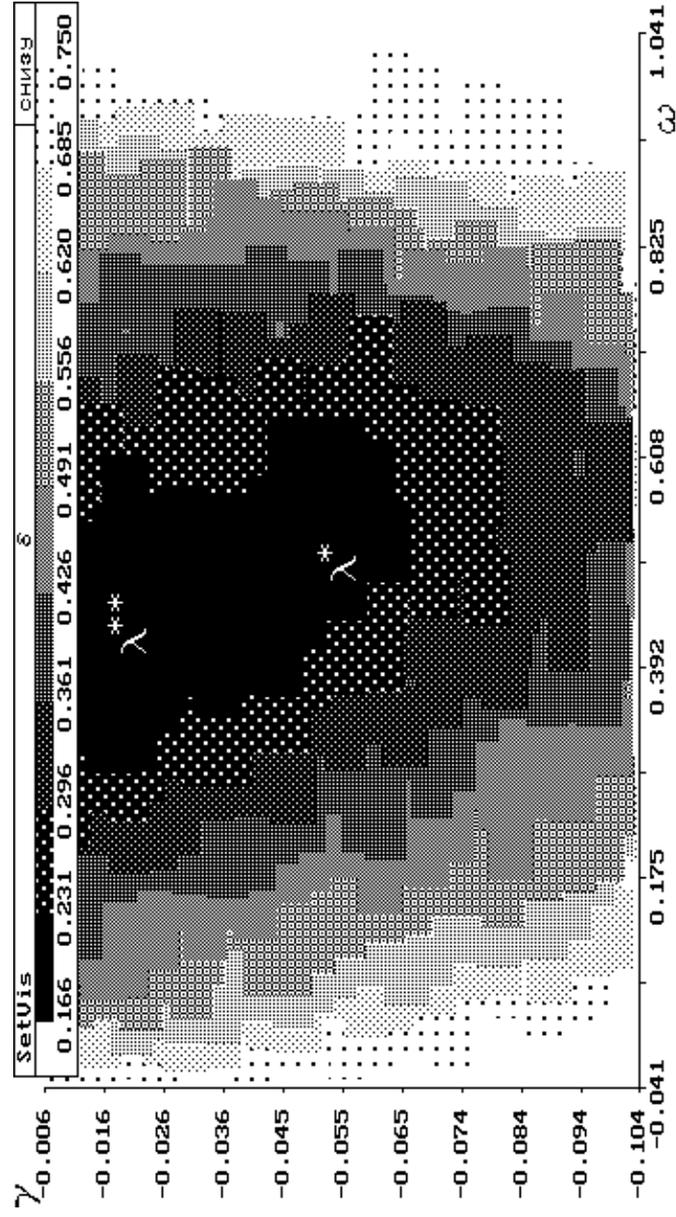


Рис. 12

## 10. Пример структурной идентификации

В следующем примере требуется оценить не только значения параметров, но и решить вопрос об адекватности структуры самой параметрической модели. В [23] приведена модель и данные наблюдений, описывающие относительное содержание  $y_1$  и  $y_2$  радиоактивного трассера (меченых атомов) в двух отделах (частях) человеческого тела в различные моменты времени после инъекции. Модель содержит по три параметра  $\theta_1$ ,  $\theta_2$  и  $\theta_3$  на каждый отдел, соответственно сходных по своему физическому смыслу,  $\theta = (\theta_1, \theta_2, \theta_3)$ . Необходимо:

- 1) решить, можно ли оба отдела описывать одними и теми же тремя параметрами одновременно (согласованная модель);
- 2) идентифицировать эти параметры по данным наблюдений.

Каждый отдел характеризуется своей структурой модели:

$$y_1 = \theta_1 e^{-\theta_2 t} + (1 - \theta_1) e^{-\theta_3 t};$$

$$y_2 = 1 - \varphi(\theta) e^{-\theta_2 t} + (1 - \varphi(\theta)) e^{-\theta_3 t}, \quad \varphi(\theta) := \frac{\theta_1 \theta_3}{\theta_1 (\theta_3 - \theta_2) + \theta_3}.$$

В [23] по методу наименьших квадратов получена оценка параметров сначала на основе только данных по первому отделу  $\theta^{(1)}$ , а затем – по второму отделу  $\theta^{(2)}$ . Оказалось, что эти две оценки слишком далеко расходятся (особенно по параметру  $\theta_1$ ), что ставит под сомнение гипотезу о применимости одних и тех же параметров одновременно (согласованная модель) к описанию обеих частей тела.

Совместная идентификация (т.е. идентификация с одними и теми же параметрами для обоих отделов) дала оценку  $\theta^{(3)}$  [23]. В [17] для той же задачи получена оценка  $\theta^{(4)}$ , близкая к  $\theta^{(2)}$  и  $\theta^{(3)}$ . Наличие различий между оценками  $\theta^{(2)}$ ,  $\theta^{(3)}$  и оценкой  $\theta^{(4)}$  указывает на неоднозначность решения. С другой стороны, так как эти различия невелики, на первый взгляд кажется, что имеет место устойчивость решения рассматриваемой задачи совместной идентификации. Поэтому на основании полученных оценок в [17] делается заключение, что при ошибке идентификации в пределах 6-7% нет оснований отвергать согласованную модель, даже если отдельные модели дают лучшую аппроксимацию.

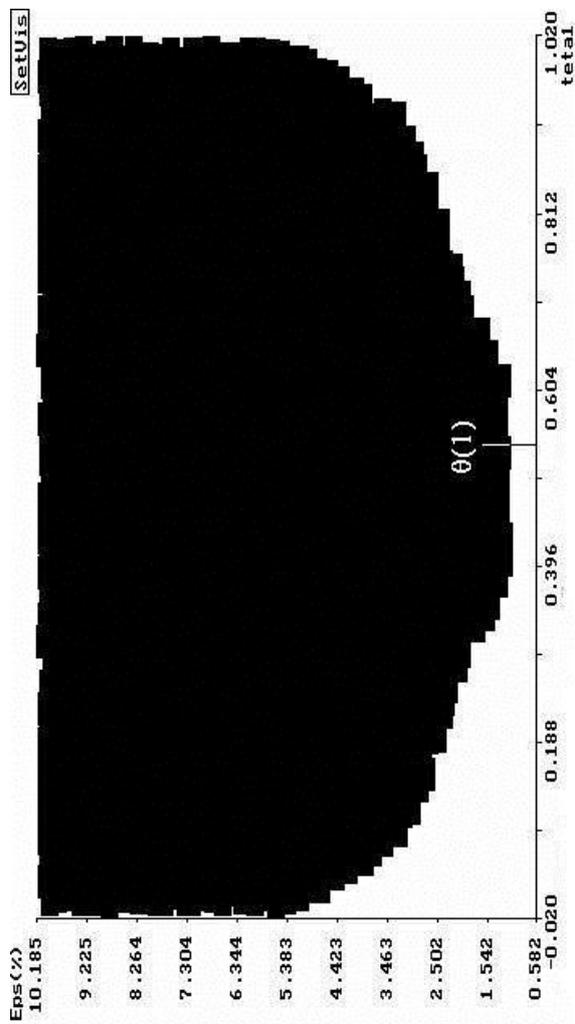


Рис. 13

Таким образом, в [23] согласованная модель поставлена под сомнение, а в [17] допускается при ошибке идентификации в пределах 6-7%.

Для прояснения вопроса об адекватности моделей используем технику визуализации графика функции ошибок.

На рис. 13 изображена проекция графика функции ошибок идентификации (множества  $G$ ) в пространство параметра  $\theta_1$  и критерия среднеквадратичного отклонения  $\varepsilon = \psi_{\text{mean}}(\lambda)$  для первой модели (для данных только по первому отделу  $y_1$ ). Из рисунка видно, что график функции ошибок идентификации для первой модели имеет пологий минимум, приходящийся на значения  $\theta_1$ , близкие к 0.5 ( $\theta_1 \in [0.4, 0.6]$ ) при ошибке не более 1%. Из этого минимума и выбрано решение  $\theta^{(1)}$ .

Рис. 14 позволяет прояснить вопрос об адекватности моделей. На этом рисунке изображены проекции множеств  $G$  в пространство параметра  $\theta_1$  и среднеквадратичного отклонения  $\varepsilon$  для трех моделей:

I - модель для данных только по первому отделу (темная штриховка);

II - модель для данных только по второму отделу (средняя штриховка);

III - согласованная модель, т. е. модель для данных по первому и второму отделам при одних и тех же параметрах (светлая штриховка).

Изображение проекции функции ошибок для I модели (только первый отдел) повторяет изображение на рис. 13. На это изображение наложено изображение проекции функции ошибок для II модели (только второй отдел), а на него наложено изображение проекции функции ошибок для III модели (первый и второй отделы одновременно при одних и тех же параметрах).

Решения задачи идентификации для II модели по параметру  $\theta_1$  гораздо лучше локализованы и при ошибке не более 1% находятся на двух участках, близких к краям априорного диапазона  $[0, 1]$ :  $\theta_1 \in [0, 0.15]$  или  $\theta_1 \in [0.85, 1]$ . Эти участки расположены в противоположных концах априорного диапазона, поэтому об устойчивости решения задачи идентификации даже в этом случае говорить не приходится.

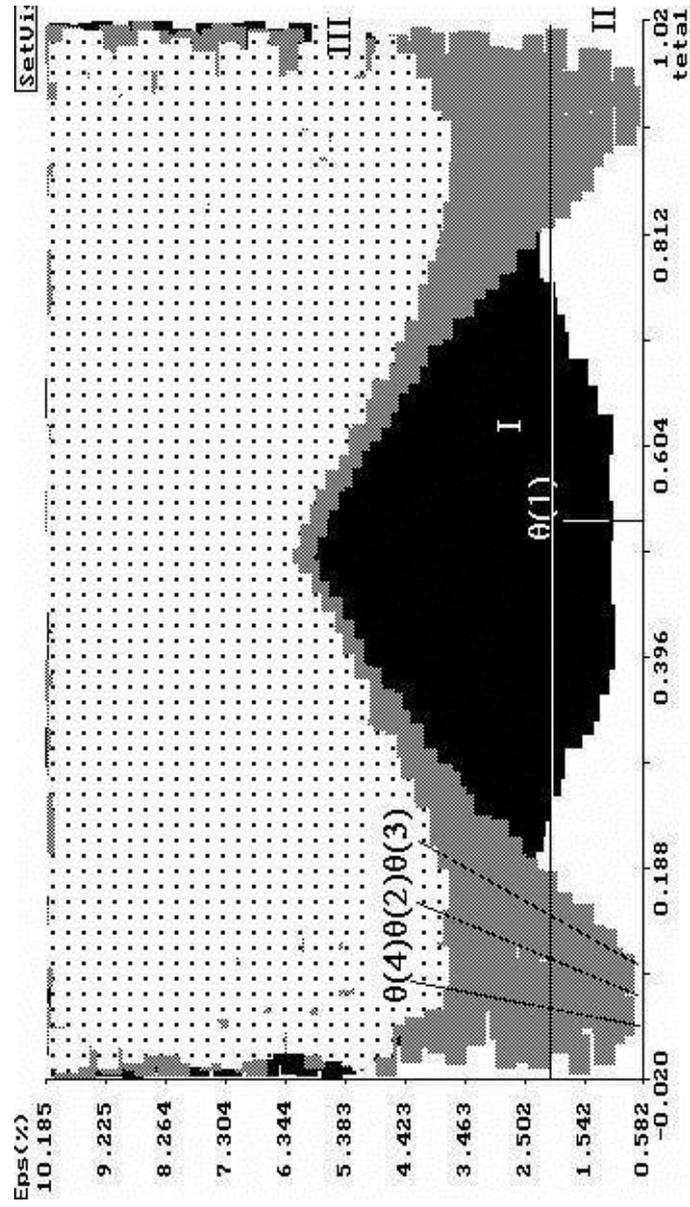


Рис. 14

Решение  $\theta^{(2)}$  выбрано на первом из двух возможных участков ( $\theta_1 \in [0, 0.1]$ ). Там же, на первом участке, выбраны и решения  $\theta^{(3)}$  и  $\theta^{(4)}$  для согласованной модели. Заметим, что в [17] и [23] второй из возможных участков решений задачи идентификации по параметру  $\theta_1$  был проигнорирован.

При экспертной информации об ошибке идентификации, меньшей 2%, множества квазиоптимальных значений параметра  $\theta_1$  для модели I и модели II не пересекаются. Поэтому при максимальном уровне функции ошибок идентификации, не превышающем 2%, согласованную модель следует отвергнуть. (Заметим, что, как было указано выше, ошибка идентификации в 2% может быть вызвана ошибками в наблюдениях величин  $y$ , не превышающими 1%.)

При ошибке идентификации, большей 2%, локализация параметра  $\theta_1$  возможна исключительно за счет данных по второму отделу. Действительно, для первого отдела (рис. 13 и модель I на рис. 14) множество квазиоптимальных значений параметра  $\theta_1$  уровня 2% составляет приблизительно диапазон  $\theta_1 \in [0.2, 0.8]$ . Это означает, что либо для идентификации по модели I недостаточно данных, либо эта модель неадекватно описывает эксперимент.

Предположим, что рассматриваемая модель адекватно описывает эксперимент и итоговый уровень ошибок идентификации в нем выше, чем 2%. В этом случае выбор между II (только второй отдел) и III (два отдела с одними параметрами) моделью по наличным данным практически невозможен, так как приводит практически к одним и тем же значениям параметров (ср. значения  $\theta^{(2)}$  для модели II и значения  $\theta^{(3)}$  и  $\theta^{(4)}$  для модели III).

При ошибках идентификации, больших 6% (это соответствует 3% точности экспериментальных данных по  $y$ ), по всем трем моделям доступен практически весь априорный диапазон параметра  $\theta_1$ , поэтому идентификация параметра и вывод об адекватности невозможны ни для одной из рассмотренных моделей.

Таким образом, если ошибки идентификации больше 6% (в частности, это соответствует 3% точности экспериментальных данных по содержанию трассера), то квазиоптимальные множества значений первого параметра  $y$  всех моделей покрывают практически весь априорный диапазон значений первого параметра.

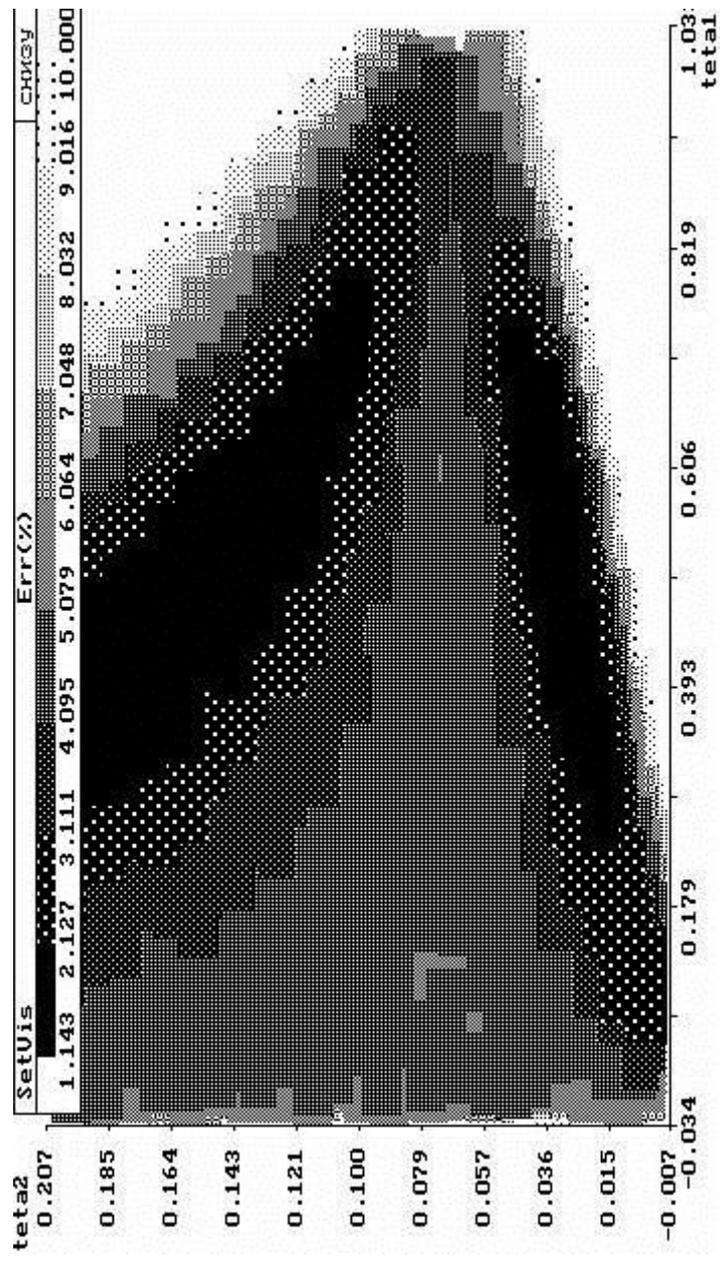


Рис. 15

Следовательно, его идентификация и вывод об адекватности некорректны ни для одной из рассмотренных моделей.

На рис. 15 изображены квазиоптимальные множества  $A_\delta$  в пространстве двух параметров  $\theta_1$  и  $\theta_2$ , ограниченные линиями уровня поверхности ошибок идентификации для первой модели (значения этих уровней представлены на шкале вверху рисунка). Каждому уровню соответствует на рисунке своя штриховка (на экране компьютера – свой цвет). Чем выше уровень допустимых ошибок – тем светлее штриховка и тем шире квазиоптимальное множество параметров. Из рисунка видно, что хотя допустимый при небольших ошибках диапазон значений параметра  $\theta_1$  является единым ( $0.2 \leq \theta_1 \leq 0.8$ ), на самом деле, если учесть влияние  $\theta_2$ , имеет место несвязное множество квазиоптимальных параметров: два кластера, один при  $\theta_2 < 0.05$ , а другой при  $\theta_2 > 0.05$ .

На рис. 16 та же картина представлена для второй модели. Видно, что и в этом случае имеют места две области возможных сочетаний параметров  $\theta_1$  и  $\theta_2$ : два кластера, один при  $\theta_1 < 0.6$ , а другой – при  $\theta_1 > 0.6$ . Однако в отличие от первой модели они и по параметру  $\theta_1$  начинают пересекаться только при значительных допустимых уровнях ошибки идентификации ( $\varepsilon \geq 5.8\%$ ). При большем уровне функции ошибки указанные области сливаются, и даже приближительная идентификация рассматриваемых параметров становится невозможной.

Неодносвязность множества квазиоптимальных параметров создает значительные трудности при использовании методов оптимизации, так как может приводить к получению решений только из одной компоненты связности. Локальные или глобальные оптимальные сочетания параметров из других компонент связности могут остаться незамеченными и неисследованными (см. рис. 14, где решения  $\theta^{(2)}$  и  $\theta^{(3)}$  и  $\theta^{(4)}$  выбраны только из левого минимума функции ошибок).

Предположим теперь, что ошибка идентификации не превышает 2%. Какие сочетания значений параметров в этом случае возможны?

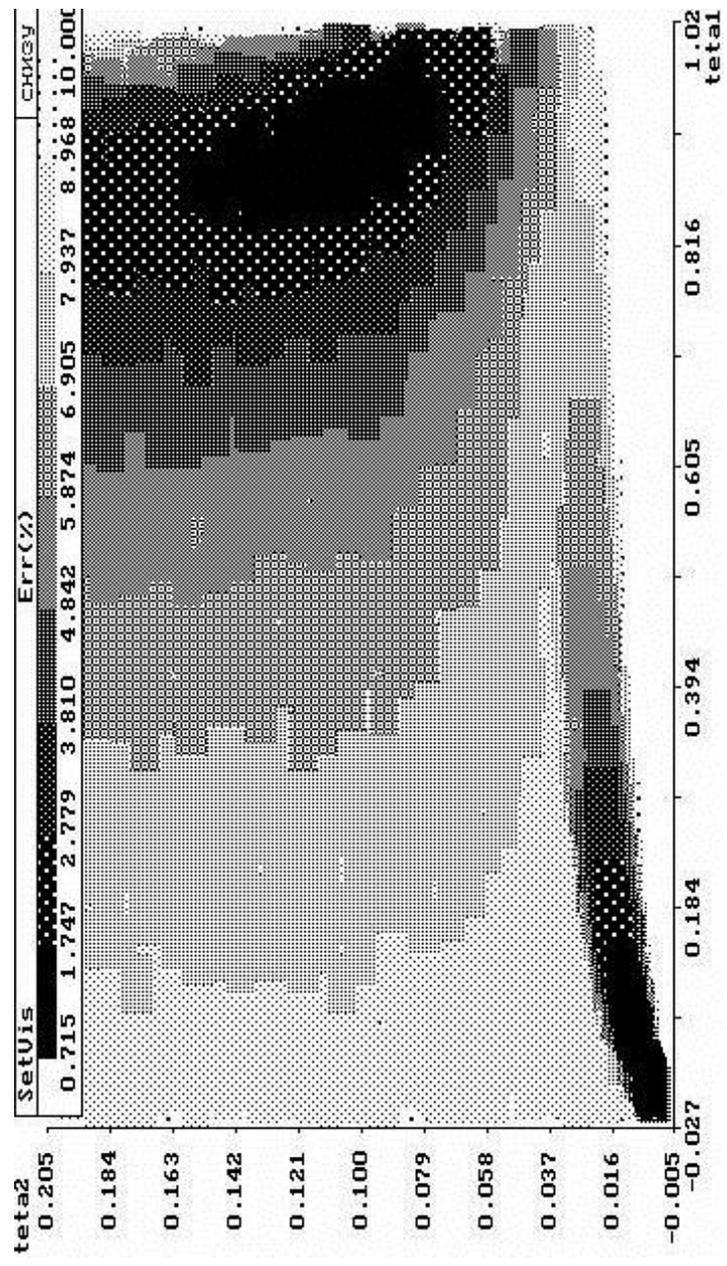


Рис. 16

На рис. 14 видно, что в модели I (первый отдел) множество квазиоптимальных значений параметра  $\theta_1$  составляет приблизительно диапазон  $\theta_1 \in [0.2, 0.8]$ . Для II модели (второй отдел) множество квазиоптимальных значений параметра  $\theta_1$  состоит из двух диапазонов  $\theta_1 \in [0, 0.18] \cup [0.82, 1]$ .

Рассмотрим теперь возможные значения параметров  $\theta_1$  и  $\theta_2$  сразу по двум моделям. На рис. 17 представлены квазиоптимальные множества уровня 2% в пространстве параметров  $\theta_1$  и  $\theta_2$  для I и II моделей. Темная штриховка соответствует первой модели, светлая – второй. Из рисунка видно, что рассматриваемые квазиоптимальные множества состоят каждая из двух кластеров. Для первой модели кластеры расщеплены по параметру  $\theta_2$ . Для второй модели кластеры расщеплены по параметру  $\theta_1$ . Важно отметить, однако, что для двух разных моделей множества квазиоптимальных параметров не пересекаются.

На рис. 18 то же множество представлено расслоенным по параметру  $\theta_3$ . Значения этого параметра представлены на шкале вверху рисунка, сечения множества квазиоптимальных параметров, соответствующие большему уровню  $\theta_3$ , наложены на сечения с меньшим его уровнем. Этот рисунок дает полное представление о структуре трехмерных множеств квазиоптимальных решений 2% уровня ошибки двух рассматриваемых структурных моделей. Например видно, что верхний (для первой модели) и правый (для второй модели) квазиоптимальные кластеры первых двух параметров существуют только при малых значениях третьего параметра.

Мы подробно рассмотрели данный пример, чтобы дать представление об основных возможностях рассматриваемой визуальной технологии в рамках параметрической и структурной идентификации. Безусловно, она требует определенной квалификации от исследователя, однако потраченные усилия окупаются за счет более глубокого проникновения в суть рассматриваемой проблемы.

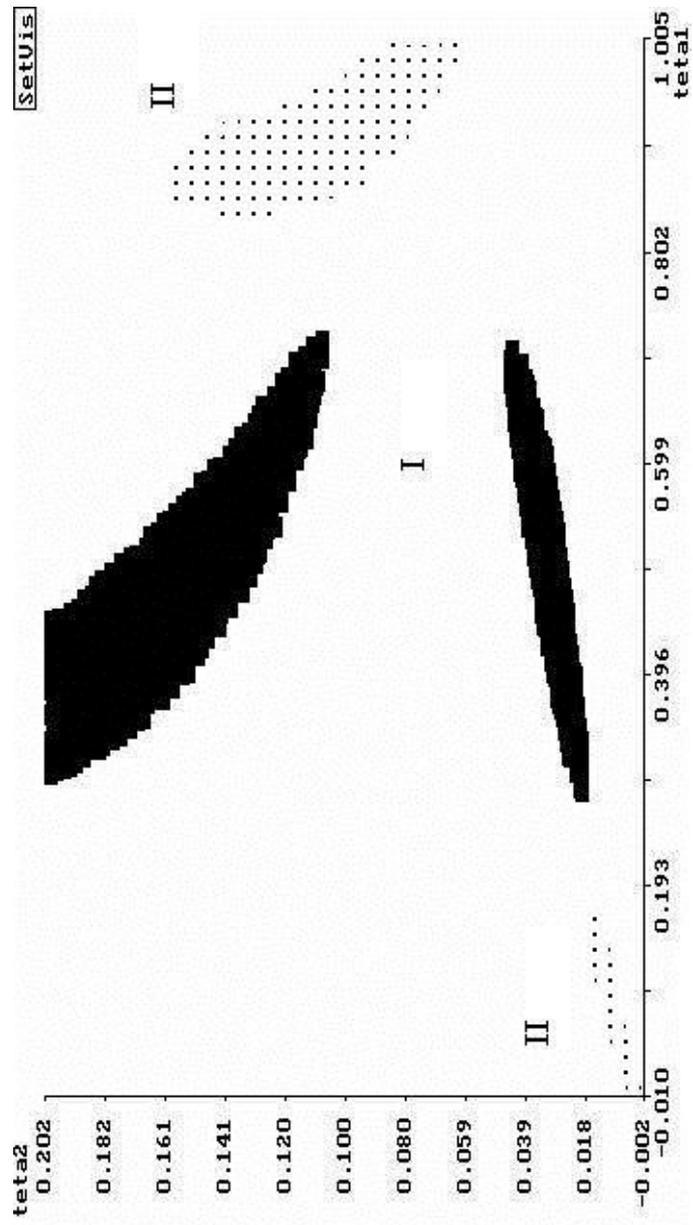


Рис. 17

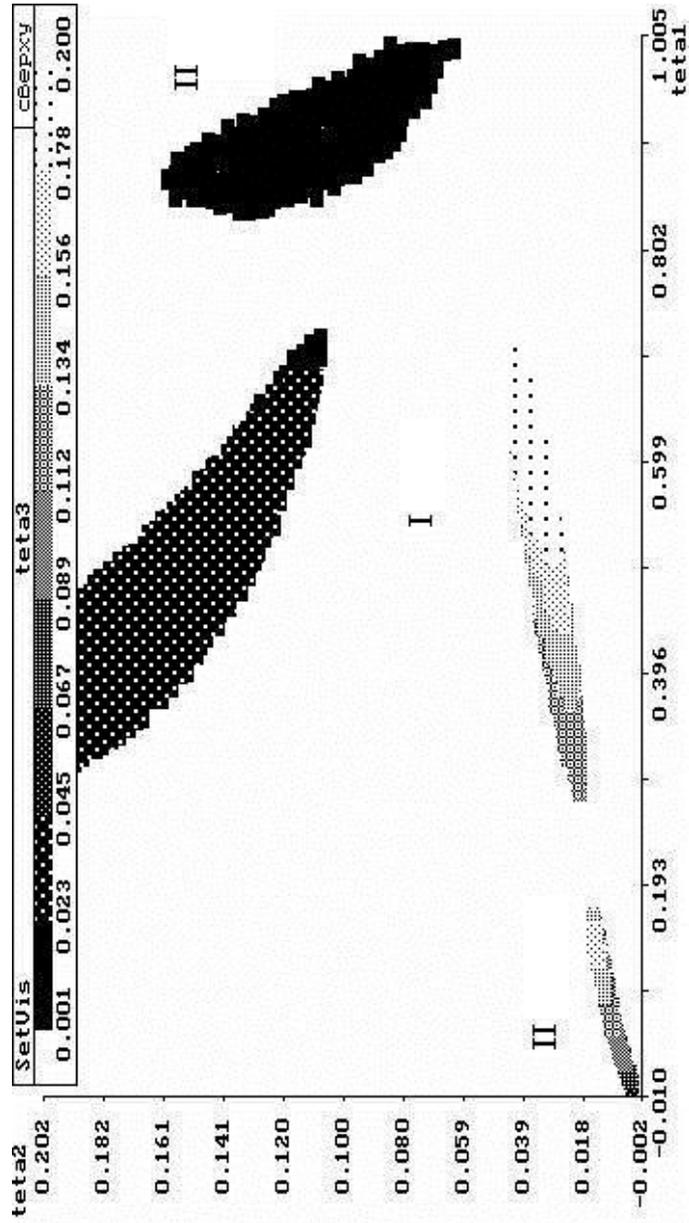


Рис. 18

## 11. Пример идентификации суммарного объема газа в зональной модели газовой залежи

Другой пример применения предлагаемой визуальной технологии содержится в [12], где рассмотрена проблема идентификации параметров грубой динамической модели газовой залежи.

Модель упрощенно описывает движение газа вследствие добычи его из залежи. Весь объем залежи разбит на несколько крупных частей, называемых далее зонами. Предполагается, что давление газа одинаково во всем объеме зоны и падает со временем вследствие уменьшения количества газа из-за добычи. Предполагается также, что давления измеряются с шагом по времени порядка 0.5 - 1 год. Движение газа внутри каждой зоны не описано, рассмотрены лишь перетоки газа через границы из зоны с большим давлением в соседние зоны с меньшим давлением. Величины перетоков вычисляются приближенно по закону Дарси движения газа в пористой среде, а именно: величина перетока из зоны  $i$  в зону  $j$  определяется как произведение  $\gamma_{ij}(P_i^2 - P_j^2)$ ,  $\gamma_{ij} = \gamma_{ji}$ , где  $P_i$  и  $P_j$  – средние давления в этих зонах, а коэффициентом  $\gamma_{ij}$  учитывается геометрия зон, проницаемость пород газоносного слоя и другие факторы, влияющие на движение газа в объеме залежи. Температуру газа в зоне и его объем, предполагая их постоянными во времени на шаге расчета, можно исключить из рассмотрения и выписать уравнение состояния, связывающее давление газа в рассматриваемой зоне с его запасом в ней.

При сделанных предположениях выпишем уравнения балансов для количества газа  $Q_i(t)$  в каждой зоне  $i$  для момента времени  $t$ . Уравнения содержат подлежащие идентификации параметры: коэффициенты  $\gamma_{ij}$  в выражениях для перетока газа из зоны в зону и начальные запасы газа  $Q_i(0)$  в каждой зоне. Наблюдаемыми величинами являются усредненные давления газа и величины добычи газа в каждой зоне на каждый момент времени. С помощью такой модели был построен грубый метод оценки первоначальных запасов газа в зонах залежи и коэффициентов перетока внутри залежи, не требующий ни точного знания характеристик газоносной среды во всех частях месторождения, ни высокой точности измерения давлений в многочисленных точках месторождения.

Исследуем устойчивость решения задачи идентификации. Рас-

смотрим трехзонную модель с типичным распределением начальных запасов  $Q^{0*}(Q_1^{0*}, Q_2^{0*}, Q_3^{0*})$ ,  $Q_i^{0*}=1000$ , коэффициентами перетока  $\gamma^*=(\gamma_{12}^*, \gamma_{13}^*, \gamma_{23}^*)$  и режимом добычи газа, в котором добыча происходит только в первой зоне. На основе этих данных смоделируем экспериментальные данные давлений  $P^*(t)=(P_1^*(t), P_2^*(t), P_3^*(t))$  в зонах (прямой просчет) на период времени  $T$ . Теперь, основываясь на полученных в прямом просчете данных давления, попробуем восстановить (идентифицировать) параметры модели, при которых они были получены.

Зададим диапазоны возможных значений параметров  $Q^0=(Q_1^0, Q_2^0, Q_3^0)$  и  $\gamma=(\gamma_{12}, \gamma_{13}, \gamma_{23})$  и для каждого допустимого сочетания параметров проведем просчет модели. Получим ряды данных по давлению  $P(t)=(P_1(t), P_2(t), P_3(t))$ . Определим ошибку идентификации как максимальное по всем трем зонам отклонение  $PE_{\max}$  этих рядов от рядов  $P^*(t)$ :

$$PE_{\max} := \max \{|P_l^*(t) - P_l(t)|: l=1,2,3\}.$$

На рис. 19 изображена зависимость квазиоптимального множества для начального запаса газа  $Q_1^0$  в первой зоне от ошибки идентификации при различном числе лет наблюдения (число наблюдений представлено на шкале вверху рисунка, причем самая темная штриховка соответствует  $T$ , равному 1 году, а самая светлая – 10 годам). Из рисунка видно, что идентификация запаса газа в отдельной зоне невозможна, так как даже при длительном наблюдении (10 лет) и очень высокой точности идентификации  $PE_{\max} = 3\%$  возможны варианты параметров месторождения с  $Q_1^0$  от 100 до 2000. Аналогичная картина наблюдается и для других зон.

Рассмотрим теперь рис. 20. На нем изображена зависимость квазиоптимального множества для суммарного начального запаса газа  $Q^s=Q_1^0+Q_2^0+Q_3^0$  от максимальной ошибки идентификации  $PE_{\max}$  рядов давления при данных за 5 лет наблюдений.

Из рисунка видно, что для каждого значения  $PE_{\max}$  диапазон возможных значений суммарного запаса  $Q^s$  составляет отрезок, включающий исходное значение суммарного запаса  $Q^{s*}=3000$ . Этот диапазон тем шире, чем больше  $PE_{\max}$ . Так, при  $PE_{\max}=1\%$  имеем  $Q^s \approx 3000$ , а при  $PE_{\max}=8\%$  имеем  $Q^s \in [\approx 2000, \approx 4500]$ . Это означает, что при данной априорной информации о диапазонах возможных параметров месторождения и при 8% точности идентификации

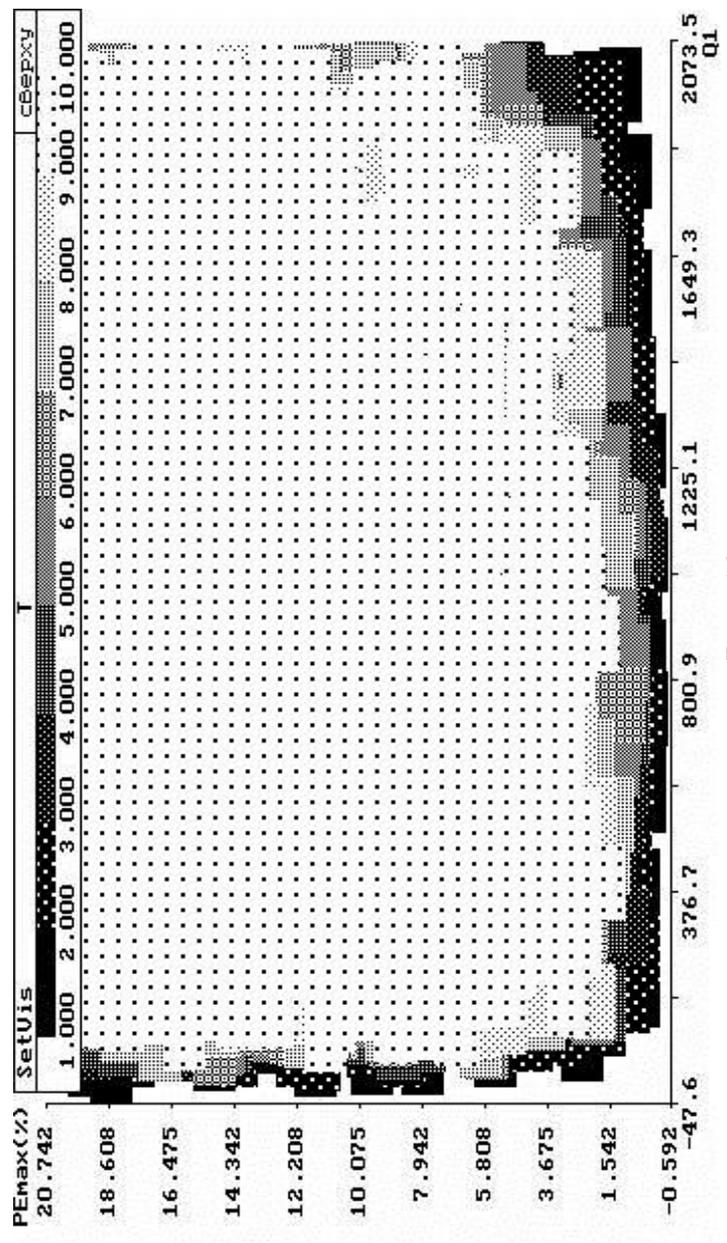


Рис. 19

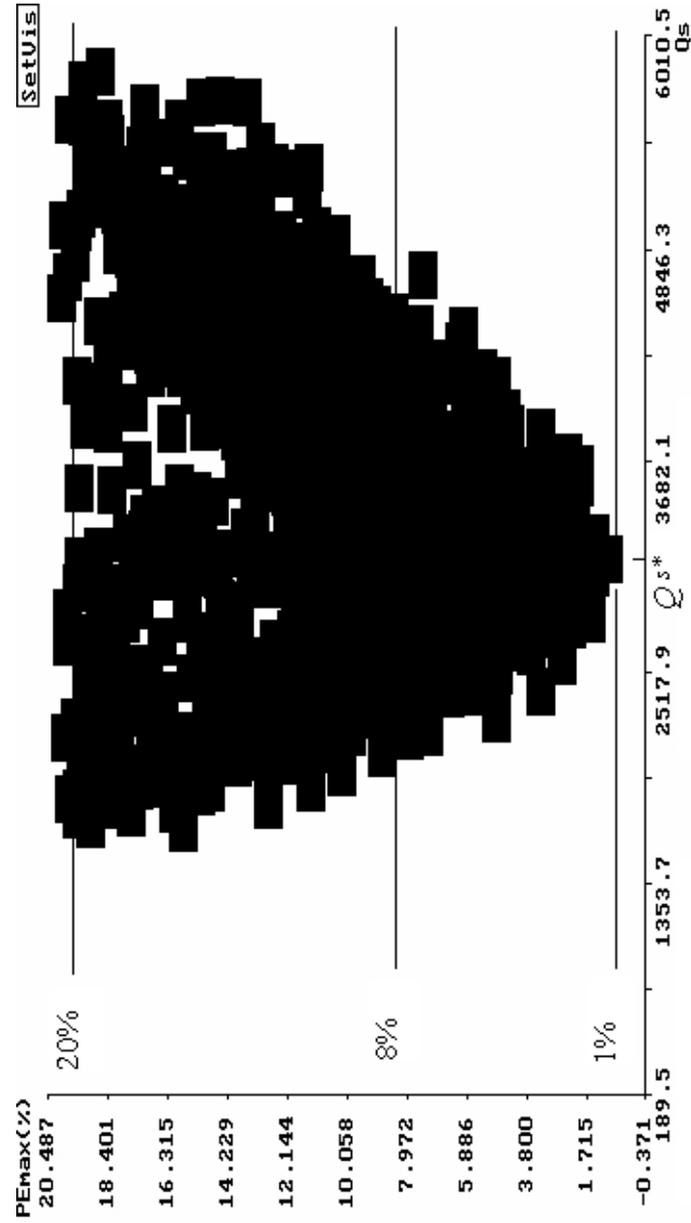


Рис. 20

(что может соответствовать 4% точности агрегированных экспериментальных данных по давлениям в зонах) суммарный запас может быть восстановлен за пять лет с точностью около 35%.

Повышение экспертного уровня точности идентификации приводит к улучшению оценки суммарного запаса, а уменьшение – к ухудшению оценки. Отметим, однако, что даже при 20% точности идентификации нижняя граница суммарного запаса восстанавливается достаточно хорошо:  $Q^s \geq 1800$ .

Рассмотрим теперь, как влияет длительность разработки месторождения на возможности восстановления суммарного запаса месторождения. На рис. 21 изображено множество всевозможных сочетаний величин  $Q^s$  и  $PE_{\max}$  при  $T$  от 1 года до 10 лет. На этом рисунке для каждого  $T$  (значения указаны на шкале штриховок вверху рисунка) множество достижимых сочетаний  $Q^s$  и  $PE_{\max}$  имеет вид, аналогичный рис. 20. Эти множества изображены на рис. 21 различной штриховкой (на экране компьютера – цветом), причем более светлая штриховка соответствует большему  $T$ .

Из рисунка видно, что при  $T=1,2$  сколько-нибудь точное восстановление суммарного запаса по данной модели практически невозможно, так как даже при  $PE_{\max} \approx 0\%$  возможны варианты с  $Q^s$  от  $\approx 2000$  до  $\approx 4500$ , а при  $PE_{\max} = 1\%$  величина  $Q^s$  может изменяться от  $\approx 1300$  до  $\approx 6000$ . С ростом  $T$  возможности восстановления суммарного запаса увеличиваются: множества достижимых сочетаний ( $Q^s$ ,  $PE_{\max}$ ) сужаются к линии  $Q^{s*} = 3000$ . Например, при  $T=10$  (самая светлая штриховка) и оценке точности идентификации в 10% можно утверждать, что величина  $Q^s$  лежит между  $\approx 2600$  и  $\approx 3700$ . т.е. можно оценить суммарный запас с точностью около 20% (отметим, однако, что при заданном темпе добычи за время  $T=10$  вырабатывается уже половина месторождения).

Как и в первом примере, в данной задаче мы наблюдали эффект замещения параметров, когда различные сочетания значений идентифицируемых параметров обеспечивали точность идентификации рядов давлений, близкую к оптимальной. В отличие от примера, рассмотренного в предыдущем пункте, данные рядов давления в зонах не могут, в принципе, быть получены с высокой точностью, так как являются величинами, агрегированными по скважинам зоны и усредненными по значительным отрезкам времени.

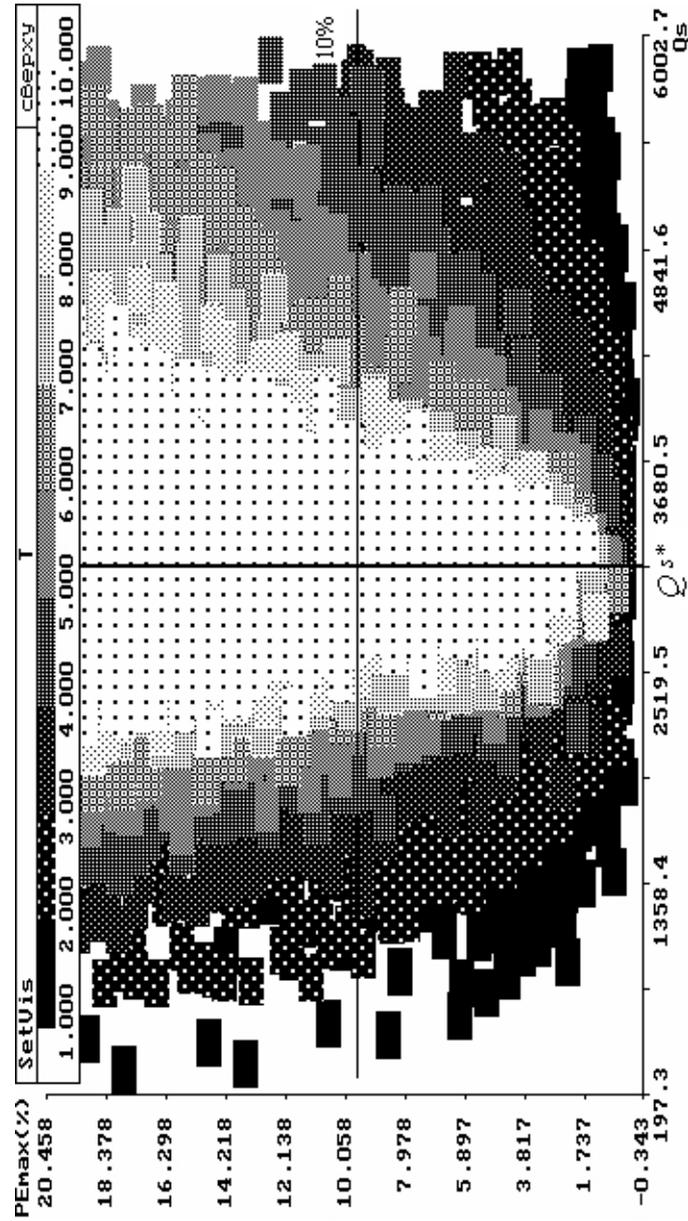


Рис. 21

Поэтому идентификация отдельных параметров была невозможна, однако грубая оценка такого значимого для принятия решений показателя, как суммарный объем газа в месторождении, оказалась, в принципе, выполнимой.

## 12. Заключение

Приведенные примеры показывают, каким образом использование метода множеств идентификации позволяет выявить внутренние ограничения, заложенные в рассматриваемой задаче, и проанализировать те возможности идентификации, которые в ней имеются.

Автор выражает глубокую благодарность А.А.Петрову и А.В.Лотову за многочисленные конструктивные замечания и плодотворное обсуждение.

## Литература

1. *Краснощекоев П.С., Петров А.А.* Принципы построения моделей. М., ФАЗИС, ВЦ РАН: 2000.
2. *Фишер Ф.* Проблема идентификации в эконометрии. М., Статистика, 1978.
3. *Райбман Н.С.* Что такое идентификация. М., Наука, 1970.
4. *Лотов А.В., Бушенков В.А., Каменев Г.К., Черных О.Л.* Компьютер и поиск компромисса. Метод достижимых целей. М.: Наука, 1997.
5. *Lotov A.V., Bushenkov V.A., Kamenev G.K.* Interactive Decision Maps. Approximation and Visualization of Pareto Frontier. Boston: Kluwer Academic Publishers, 2004.
6. *Lotov A.V., Bushenkov V.A., Kamenev G.K., Loucks D.P., Camara A.S.* Water resource conflict resolution based on interactive tradeoffs display // Restoration of Degrad. Rivers: Challenges, Issues and Exper. Kluwer Acad. Publ.: 1998. P. 447-470.
7. *Лотов А.В., Бушенков В.А., Каменев Г.К.* Метод достижимых целей. Математические основы и экологические приложения. The Edwin Mellen Press, Lewiston, NY, USA, 1999.
8. *Lotov A., Bourmistrova L., Bushenkov V.* Efficient strategies: an ap-

- plication in water quality planning. In: G. Kersten, Z. Mikolajuk, M. Rais, and A. Yeh (eds.) *Decision Analysis and Support for Sustainable Development*. Kluwer Academic Publishers: Dordrecht, 1999. Pp. 145-166.
9. Lotov A., Berezkin V., Kamenev G., Miettinen K. Optimal Control of Cooling Process in Continuous Casting of Steel Using a Visualization-Based Multi-Criteria Approach // *Applied Mathematical Modelling*, 2005. 29(7). P. 653-672.
  10. Каменев Г.К. Визуальный метод идентификации параметров // Доклады Академии наук. 1998. Т. 359. № 3. С. 319-322.
  11. Каменев Г.К., Кондратьев Д.Л. Об одном методе исследования незамкнутых нелинейных моделей // *Математическое моделирование*. 1992. №3. С. 105-118.
  12. Гереш П.А., Сушков Б.Г., Фуругян М.Г., Каменев Г.К. и др. Оценка параметров газовой залежи с помощью обобщенной динамической модели. М.: ВЦ РАН, 1994. 40 с.
  13. Петров А.А., Поспелов И.Г., Шананин А.А. Опыт математического моделирования экономики. М.: Энергоатомиздат, 1996.
  14. Поспелов И.Г., Хохлов М.А. Идентификация экономических систем на основе метода множеств достижимости. // II Всероссийская научн. конф. «Матем. моделирование развивающейся экономики ЭКОМОД-2007». Сборник трудов. Киров: ВятГУ, 2007. С. 205-215.
  15. Каменев Г.К. Визуальная технология идентификации параметров моделей технических и биологических систем // III Всероссийская научн. конф. «Матем. моделирование развивающейся экономики, экологии и биотехнологий ЭКОМОД-2008». Киров: ВятГУ, 2008. С. 30.
  16. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. М., Финансы и статистика, 1982.
  17. Бард Й. Нелинейное оценивание параметров. М., Статистика, 1979.
  18. Bushenkov, O.L. Chernykh, G.K. Kamenev, and A.V. Lotov. Multi-dimensional Images Given by Mappings: Construction and Visualization // *Pattern Recognition and Image Analysis*. 1995. V. 5. №1. P. 35-56.
  19. Каменев Г.К. Об одном классе адаптивных алгоритмов аппрок-

- симации выпуклых тел многогранниками // Ж. вычисл. матем. и матем. физ. 1992. Т. 32. № 1. С. 136-152.
20. *Каменев Г.К.* Оптимальные адаптивные методы полиэдральной аппроксимации выпуклых тел. М: Изд. ВЦ РАН, 2007, 230 с.
  21. *Бронштейн Е.М.* Аппроксимация выпуклых множеств многогранниками // Современная математика. Фундаментальные направления. 2007. Т. 22. С. 5-37.
  22. *Каменев Г.К.* Аппроксимация вполне ограниченных множеств методом Глубоких Ям // Ж. вычисл. матем. и матем. физ. 2001. Т. 41. №11. С. 1751-1760.
  23. *Beauchamp J.J. and Cornell R.G.* Simultaneous nonlinear estimation // Technometrics. N8, 1966.

## Оглавление

1. Неопределенность в задачах идентификации .....	1
2. Метод достижимых целей .....	7
3. Метод множеств идентификации (ММИ) .....	8
4. Диалоговые карты идентификации .....	12
5. Исследование функции ошибок .....	17
6. Аппроксимация множеств идентификации в выпуклом случае .....	19
7. Аппроксимация множеств идентификации в нелинейных моделях .....	21
8. Общий случай применения ММИ .....	22
9. Простой пример нелинейной регрессии .....	23
10. Пример структурной идентификации .....	26
11. Пример идентификации суммарного объема газа в зональной модели газовой залежи .....	37
12. Заключение .....	43
Литература .....	43