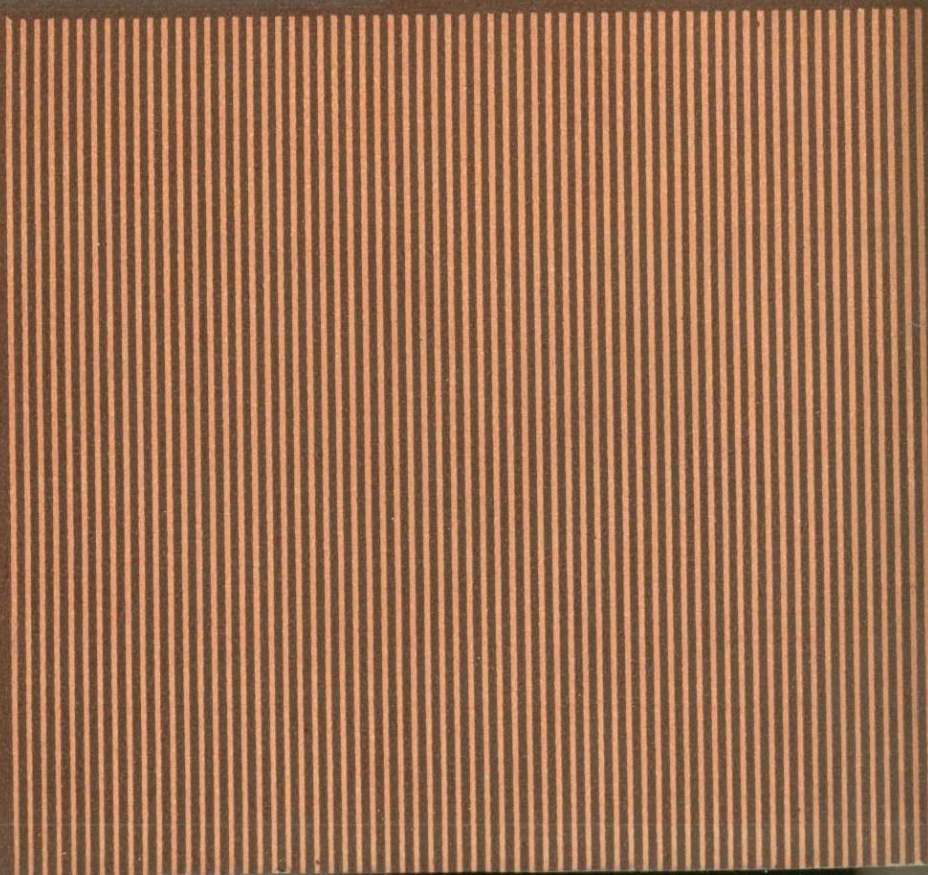


В. И. АРОНОВ

**МЕТОДЫ
МАТЕМАТИЧЕСКОЙ
ОБРАБОТКИ
ГЕОЛОГИЧЕСКИХ
ДАННЫХ
НА ЭВМ**



В. И. АРОНОВ

МЕТОДЫ
МАТЕМАТИЧЕСКОЙ
ОБРАБОТКИ
ГЕОЛОГИЧЕСКИХ
ДАННЫХ
НА ЭВМ

2068



Москва «НЕ Д Р А» 1977



Аронов В. И. Методы математической обработки геологических данных на ЭВМ. М., «Недра», 1977. 168 с.

Рассматриваются наиболее общие и важные задачи обработки геологических данных на ЭВМ: интерполяция геологических характеристик на регулярную сеть и построение карт изолиний, фильтрация случайных ошибок и тренд-анализ, прогнозирование количественных геолого-геофизических характеристик по комплексу косвенных признаков, подсчет запасов и создание автоматизированных систем обработки геологических данных на базе универсальных информационно-поисковых систем.

Основное внимание уделено теории и методической стороне использования методов, уже нашедших применение в практике, вопросам их алгоритмизации и реализации на ЭВМ. Приводятся теоретические и практические примеры.

Книга рассчитана на геологов и специалистов, занимающихся изучением и внедрением математических методов и ЭВМ в геологию.

Табл. 11, ил. 29, список лит. — 96 назв.

А $\frac{20802-190}{043(01)-77}$ 23—77

© Издательство «Недра», 1977

ИБ № 1850

Владимир Ильич Аронов

МЕТОДЫ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ ГЕОЛОГИЧЕСКИХ ДАННЫХ НА ЭВМ

Редактор издательства *Л. С. Цаплина*
Обложка художника *И. А. Дутова*
Художественный редактор *В. В. Евдокимов*
Технический редактор *О. А. Болтунова*
Корректор *А. А. Передерникова*

Сдано в набор 15/XII 1976 г. Подписано в печать 14/III 1977 г. Т-06124. Формат 60×90^{1/16}.
Бумага № 2. Печ. л. 10,5. Уч.-изд. л. 10,84. Тираж 5400 экз. Заказ 2604/5598-14.
Цена 1 р. 08 к.

Издательство «Недра», 103633, Москва, К-12, Третьяковский проезд, 1/19
Московская типография № 6 Союзполиграфпрома при Государственном
комитете Совета Министров СССР по делам издательств полиграфии
и книжной торговли.

109088, Москва, Ж-88, Южнопортовая ул., 24.

В настоящее время совершенствование способов обработки и анализа геологических данных немислимо без привлечения математических методов и современной вычислительной техники. С внедрением математических методов в обработку геолого-геофизических данных в значительной мере связывается прогресс геологии как науки. Не менее актуальным и важным с практической точки зрения является внедрение ЭВМ в практику решения традиционных задач, в частности для сбора, хранения, извлечения и обработки геологических данных. ЭВМ позволяет в широких масштабах производить оперативную обработку данных, решать те задачи, которые традиционными способами обычно не решались из-за большой трудоемкости, причем на более объективной основе, с использованием всего фактического материала.

Современные масштабы использования математических методов и ЭВМ — только начало необратимого процесса. Он вызван не только появившимися возможностями их использования, но и является результатом накопления огромного объема геологических данных в результате совершенствования и широкого внедрения количественных методов геологических исследований. Вследствие этого систематизация данных, их обработка и поиск закономерностей обычными методами оказываются непосильными задачами. Появление новых, более формализованных методов исследований, проводящихся к тому же с использованием ЭВМ, вызывает побочный, но очень важный эффект роста требований к объективности интерпретации, к строгости и однозначности выводов.

Возможности широкого применения математических методов и ЭВМ ограничиваются не только сложностью геологических процессов и невозможностью получения решений на математической основе, но еще и тем, что часто отсутствует четкая геологическая постановка задачи. Несмотря на это от применения ЭВМ в геологии еще недавно ожидали результатов, которые нереальны даже в ближайшем будущем.

При правильной постановке задачи должны быть указаны все исходные данные, которые могут быть использованы, и, фактически, метод ее решения, т. е. даны те четкие соображения, которые в принципе могут быть переложены на математический язык. Сделать это, как правило, очень трудно. Именно поэтому сейчас в геологии наиболее часто применяется математическая статистика, которая отражает лишь внешнюю сторону изучаемого процесса, результаты его проявлений, не за-

трагивая их сути. Получила даже распространение точка зрения, что геологические процессы носят вероятностный характер и что поэтому применение вероятностно-статистических методов является основой их описания. Это априорная точка зрения. Работ, в которых этот вопрос сколько-нибудь серьезно исследовался, по сути дела, нет совсем. Даже если принять эту точку зрения, математическая статистика может быть использована лишь для проверки идей, вытекающих из детерминированной постановки задачи, т. е. из физико-геологических предпосылок. Применение математической статистики всегда подразумевает достаточно глубокое знание изучаемого явления (а также самой статистики, в особенности ее основных принципов). Конечно, выявление статистических закономерностей может натолкнуть исследователя на проверку и обдумывание полученных результатов. Правильное и аккуратное использование статистических методов, несомненно, очень полезно, однако считать их основой изучения геологических процессов, безусловно, неверно. Наоборот, геологические процессы и их проявления являются уникальными, наблюдается лишь одна их реализация, повторение которой в статистическом смысле надо всегда доказывать. Поэтому, несмотря на возможность вероятностно-статистического описания в первом приближении уже известных данных, статистический прогноз (а именно только он и представляет интерес) сравнительно редко бывает удачным. По-настоящему хороших примеров, выдерживающих всестороннюю критику, очень мало. Иными словами, поскольку нам неизвестна генеральная совокупность, надежность статистических оценок не может быть определена. Практика показала, что достоверными оказываются не статистические оценки надежности, а результаты проверки на хорошем контрольном материале. Отсюда следует, что главной задачей является выяснение принципиальной стороны вопроса, правильная постановка задач, а не освоение новых статистических методов или переход к более тонким статистическим критериям, увлечение которыми сейчас наблюдается. Известно, что лучше приближенно решить правильно поставленную задачу, чем точно решить задачу поставленную неверно.

В то же время, несмотря на значительно меньшее число публикаций, все большее значение приобретают детерминированные методы решения задач, основанные на различных методах численного анализа. В принципе большинство геологических задач можно рассматривать как задачи интерполяции и экстраполяции, однако конкретные методы решения получены сейчас лишь для некоторых наиболее подготовленных задач. Это интерполяция геолого-геофизических признаков при автоматизации процесса построения карт; прогнозирование количественных характеристик по комплексу других косвенных признаков; фильтрация случайных ошибок измерений, тренд-ана-

лиз, подсчет запасов, а также целый ряд «специализированных» задач гидрогеологии, геохимии, инженерной геологии и др. При этом, как и в других отраслях науки, выяснилось, что решение на ЭВМ плохо сформулированных геологических задач встречает большие трудности, в то время как эвристическими традиционными методами они решаются относительно хорошо. Это легко прослеживается даже при решении таких, казалось бы, подготовленных формализованными методами задач, как интерполяция, фильтрация и тренд-анализ, поэтому большой интерес представляет изучение эвристических методов решения геологических задач и их комплексирование с формально-математическим аппаратом. Таких примеров еще мало, однако перспективы заманчивые.

В геологии решение задач тренд-анализа и даже классических задач интерполяции и аппроксимации имеет заметную специфику. Кроме того, стандартного математического обеспечения для этих задач на ЭВМ в двухмерном случае практически пока нет. Это привело к необходимости разработки специальных методов и программ на ЭВМ.

Полученные результаты позволили организовать массовое решение на ЭВМ всех перечисленных выше задач и привели к обособлению этапа обработки данных на ЭВМ, причем применяемые методы существенно отличаются от традиционных.

Геологу необходимо понимание технологии вычислительных процессов. В противном случае можно принять математический эффект или эффекты реализованного алгоритма за эффекты геологической модели. «Если ставится цель понять физическое явление, то автор задачи должен понимать и следить за вычислениями. Это не значит, что он должен выполнять всю мелкую работу, но если он не будет в достаточной степени понимать все, что делает машина, то он вряд ли сумеет извлечь из машины максимум пользы, а также понять смысл даже правильно построенных вычислений». Следует всегда помнить, что «цель расчетов — понимание, а не числа». «Прежде чем решать задачу, подумай, что делать с ее решением» [94]. Существует точка зрения, правильная по мнению автора, что научить специалиста в конкретной области программированию и прикладной математике легче, чем математику разобрататься в специфике и постановке геологических задач.

В настоящей книге сделана попытка отразить сложившиеся математические методы решения наиболее общих и важных задач обработки геологических данных, а также алгоритмы их реализации на ЭВМ.

**ИНТЕРПОЛЯЦИЯ ГЕОЛОГИЧЕСКИХ
ХАРАКТЕРИСТИК
НА РЕГУЛЯРНУЮ СЕТЬ.
ПОСТРОЕНИЕ ГРАФИКОВ И КАРТ ИЗОЛИНИИ**

Графическое изображение данных наблюдений играет в геологии огромную роль, являясь в течение многих десятилетий одним из основных инструментов геолога. Поскольку наблюдения практически всегда носят дискретный характер, построение графиков и карт всегда связано с интерполяцией исходных данных. Интерполяция и экстраполяция лежат в основе метода аналогий, который также используется в геологии чрезвычайно широко. Таким образом, процедура интерполяции используется в ряде этапов геологических исследований: начиная с первичной обработки данных и кончая интерпретацией и прогнозом. Это не всегда четко очерченный процесс, наоборот, он часто производится геологом на интуитивном уровне. В целом он плохо поддается формализации, поэтому машинные способы интерпретации пока не получили широкого распространения. В этой главе рассматривается лишь та часть задач интерполяции, которая связана с автоматизацией процесса составления графиков и карт изолиний, а также с подготовкой регулярной сети исходных данных для дальнейших расчетов.

В «домашинный» период интерполяция производилась визуально или графически на основании некоторых эвристических правил и инструкций, не имеющих строгого теоретического обоснования. Естественно, значительная роль при этом отводилась субъективным представлениям и опыту исследователя. Задача приведения наблюдений к регулярной сети даже в тех случаях, когда она возникала, не считалась серьезной самостоятельной задачей, что было обусловлено рядом взаимосвязанных причин, основные из которых приводятся далее.

Во-первых, во многих практических ситуациях, особенно в одномерном варианте, графическая интерполяция позволяет получать удовлетворительные результаты даже при значительной неравномерности сети исходных пунктов. Опытный специалист одновременно может производить не только интерполяцию измерений, но и эффективную фильтрацию случайных ошибок. В некоторых случаях подобная фильтрация является пока идеалом, который не достигнут формальными математическими методами.

Во-вторых, формализованное решение задачи интерполирования сопряжено с выполнением большого объема вычисле-

ний, сопоставимого, а иногда во много раз превышающего объем расчетов, необходимых для решения основных задач. Поскольку общий объем вычислений был сравнительно невелик, предпочтение отдавалось нестандартным подходам, развиваемым с той или иной степенью строгости для каждой конкретной задачи или даже для каждой конкретной ситуации.

Проблема интерполяции результатов экспериментальных наблюдений к узлам регулярной (обычно квадратной) сети стала актуальной, по-видимому, одновременно с внедрением в практику геолого-геофизических исследований электронных вычислительных машин. При разработке «машинных» алгоритмов решения геолого-геофизических задач сразу же выяснилось, что квадратная сеть исходных данных более удобна, так как позволяет упростить и определенным образом стандартизировать последующие вычислительные процедуры. Выяснилось, что при наличии помех и значительной неравномерности наблюдений задача интерполяции далеко не тривиальна и требует самого серьезного внимания, так как точность интерполяции фактически определяет точность решения основных задач, например, подсчета запасов полезных ископаемых, трансформирования потенциальных геофизических полей и т. п. В последние годы эта задача приобрела самостоятельное значение в связи с возникшей возможностью автоматизированного построения графиков и карт изолиний на графопостроителях, современное математическое обеспечение которых обычно предусматривает задание входных данных по квадратной сети.

По теоретическим вопросам интерполяции и приближения дискретно заданных функций написано много трудов. В то же время опыт показывает, что успех решения задачи определяется не только идейными предпосылками, заложенными в алгоритм интерполяции, но и их практическим воплощением на ЭВМ. В настоящей работе мы ограничимся в основном обсуждением лишь тех методов, которые уже реализованы в виде соответствующих программ и применяются для автоматизированной обработки геолого-геофизических данных.

§ 1. Постановка задачи

Пусть имеется множество пунктов наблюдений P с координатами x_i, y_i, z_i

$$P = \{(x_i, y_i, z_i)\},$$

образующих некоторую область D с границей Γ , в которых известны значения некоторого геологического признака $\varphi(x_i, y_i, z_i)$. Необходимо приблизить эти значения некоторой непрерывной функцией $\overline{\varphi}(x, y, z)$, значения которой в точках наблюдений в определенном смысле совпадают со значениями $\varphi(x_i, y_i,$

z_i), $\{(x_i, y_i, z_i)\} = P$ и близки к значениям $\varphi(x, y, z)$ во всей области D , $(x, y, z) \in D$. Эту функцию $\bar{\varphi} \sim \varphi$ в дальнейшем предполагается использовать для вычисления значений φ в любых точках области D , в том числе, например, по координатной сетке с шагом Δx , Δy , Δz , или для прослеживания координат изолиний приближаемого геологического признака, т. е. точек, где $\bar{\varphi}(x, y, z) = c_i$, $c_i = \text{const}$. Как правило, задача рассматривается в более простой постановке, когда влияние высоты измерений z не учитывается*.

Точность решения задачи интерполяции зависит от ряда характеристик, важнейшими из которых являются:

- 1) свойства приближаемой функции $\varphi(x, y)$;
- 2) шаг измерений Δx , Δy и расстановка узлов исходных значений в области интерполяции;
- 3) уровень случайных ошибок наблюдений;
- 4) свойства приближающей функции $\bar{\varphi}(x, y)$ и выбор нормы и величины ε :

$$\|\bar{\varphi}(x_i, y_i) - \varphi(x_i, y_i)\| \leq \varepsilon, \{(x_i, y_i)\} = P;$$

5) качество реализации математического алгоритма. Возможность фильтрации случайных ошибок. Ошибки округления.

Следует подчеркнуть, что, не зная априори свойств функции $\varphi(x, y)$, нельзя однозначно восстановить ее по дискретным значениям $\varphi(x_i, y_i)$. Восстановление функции $\varphi(x, y)$ в этих условиях возможно лишь с некоторой ошибкой, минимальное значение которой δ можно определить, лишь задавшись некоторыми характеристиками функции $\varphi(x, y)$ и сети пунктов наблюдений. Здесь нужно различать две качественно различные ситуации.

1. Данных для построения карты с необходимой детальностью недостаточно; в этом случае при графических построениях особенно сильно сказываются субъективность подхода геолога, его знания и интуиция. Карта на самом деле строится не только по имеющимся замерам, но и с учетом общих соображений о поведении признака, сведений о его корреляции с другими, более детально изученными, и т. п.

2. Исходных данных вполне достаточно для того, чтобы карты, составленные с учетом обычных требований, отличались незначительно. Привлечение косвенных данных при графическом построении не является необходимым.

* В некоторых случаях, в частности при подсчете запасов, проведение интерполяции по вертикальной координате представляет значительный интерес. К настоящему времени учет влияния высоты пунктов наблюдений разработан лишь для тех случаев, когда наблюдаемый признак является гармонической функцией, например Δg или Δz , аномалии гравитационного и магнитного поля [8, 16].

В первом случае минимальные ошибки интерполяции математическими методами составляют десятки процентов, во втором — несколько процентов. Положение в настоящее время таково, что математические методы приближения функций, применяемые в геологии и геофизике, дают хорошие (сравнительно с графической интерполяцией) результаты лишь во второй ситуации. Первая из них — это в значительной мере область экстраполяции, прогноза, требующая разработки специальных, не обоснованных строго с математической точки зрения, методов. В этих условиях большой интерес представляет моделирование тех эвристических приемов, которые использует геолог при построении карт вручную, моделирование процесса, который можно назвать геологической интерполяцией.

С учетом неоднозначности получаемых результатов и трудностей их объективного сравнения на фактическом материале возникает серьезная задача оценки применяемых методов, задача разработки критериев для их сравнения. Такими критериями могут быть, в частности, следующие.

1. Точность на определенном классе функций. Возможность проведения теоретических оценок точности при некоторых необходимых предположениях о свойствах функции $\varphi(x, y)$.

2. Устойчивость к ошибкам измерений, возможность эффективной фильтрации этих ошибок.

3. Устойчивость относительно структуры сети исходных данных.

4. Устойчивость параметров вычислительной конструкции по отношению к изменению уровня сигнал/помеха, параметров, характеризующих «сложность» функции $\varphi(x, y)$ относительно значений Δx , Δy , и непосредственно связанная с этим простота определения параметров вычислительной схемы, минимизация их числа.

5. Технологичность алгоритма.

Таким образом, желательно, чтобы метод был достаточно универсальным, годился к использованию при очень малом и при существенном уровне помех, слабо зависел от структуры сети исходных данных, был экономичным в расходовании машинного времени и давал высокую точность на сравнительно широком классе функций. В то же время известно, что чем уже класс функций $\varphi(x, y)$, т. е. чем больше данных о его свойствах, тем лучший по точности метод приближения можно построить. В геологии, однако, в большинстве случаев можно установить априори лишь некоторые очень общие свойства наблюдаемых признаков, поэтому относительно большую, чем обычно, роль должны играть вопросы, перечисленные в пунктах 2—5. Это не оправдывает, естественно, то положение, когда применяемый метод не имеет вообще никаких оценок точности, кроме качественных, и, к тому же, полученных из рассмотрения результатов на отрывочных теоретических или практических

примерах. Рассмотрение таких примеров может быть эффективным в тех случаях, когда они иллюстрируют полученные теоретически результаты, пусть даже качественные. В противном случае требуется слишком детальное и всестороннее изучение всех возможных ситуаций.

Остановимся очень кратко на некоторых исходных предпосылках и определениях теории полиномиальных приближений [20, 22, 42, 43, 94].

Функции вида

$$Q_m(x) = c_0\psi_0(x) + c_1\psi_1(x) + \dots + c_m\psi_m(x), \quad (I.1)$$

где c_0, c_1, \dots, c_m — постоянные коэффициенты, называются обобщенными полиномами порядка m . Чаще всего функцию $Q_m(x)$ отыскивают в виде алгебраического полинома (многочлена), где функции ψ представляют собой целые положительные степени переменной x :

$$Q_m(x) = c_0 + c_1x + c_2x^2 + \dots + c_mx^m. \quad (I.2)$$

Если $\psi_0=1$, $\psi_1(x) = \cos x$, $\psi_2(x) = \sin x$, $\psi_3(x) = \cos 2x$, ..., то

$$Q_m(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + \dots + \\ + a_m \cos mx + b_m \sin mx \quad (I.3)$$

называется тригонометрическим полиномом порядка m . Такие полиномы обычно применяют в тех случаях, когда приближаемая (интерполируемая) функция $f(x)$ является периодической.

Если значения приближаемой функции $f(x)$ и полинома $Q_m(x)$ в исходных узлах x_i , $i=0, 1, 2, \dots, n$ (узлах интерполяции) совпадают лишь приближенно $f(x_i) \approx Q_m(x_i)$, полином $Q_m(x)$ называется аппроксимационным. В случае когда $f(x_i) = Q_m(x_i)$, он называется интерполяционным. Теория и методы построения интерполяционных и аппроксимационных полиномов существенно различны. Соответственно этому обычно рассматриваются две задачи теории приближений — интерполяция и аппроксимация функций, хотя ясно, что по сути решаемых задач первая из них является частным случаем второй. В обоих случаях задача заключается в том, чтобы, используя некоторые известные значения функции $f(x_i)$, построить некоторый многочлен, а затем уже выполнять дальнейшие расчеты с этим многочленом.

При построении интерполирующей функции ставится задача нахождения такого полинома $Q_m(x)$ возможно низшей степени m , который в узлах интерполяции принимает те же значения, что и исходная функция $f(x)$, т. е. такой, что

$$Q_m(x_i) = f(x_i) \quad (i = 0, 1, 2, \dots, n; \quad x_i \neq x_j \text{ при } i \neq j). \quad (I.4)$$

Коэффициенты полинома (I.2) в этом случае можно опреде-

лить из следующей системы линейных алгебраических уравнений при $n=m$:

$$\begin{cases} c_0 + c_1 x_0 + c_2 x_0^2 + \dots + c_m x_0^m = y_0 \\ c_0 + c_1 x_1 + c_2 x_1^2 + \dots + c_m x_1^m = y_1 \\ \dots \\ c_0 + c_1 x_m + c_2 x_m^2 + \dots + c_m x_m^m = y_m, \end{cases} \quad (I.5)$$

где $y_i = f(x_i)$ ($i=0, 1, 2, \dots, n=m$), или более коротко:

$$\sum_{j=0}^m c_j x_i^{(j)} = f(x_i) \quad (i=0, 1, 2, \dots, m).$$

Известно, что определитель этой системы (определитель Вандермонда) не равен нулю:

$$\Delta = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^m \\ 1 & x_1 & x_1^2 & \dots & x_1^m \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_m & x_m^2 & \dots & x_m^m \end{vmatrix} \neq 0. \quad (I.6)$$

Отсюда следует, что система (I.5) является линейно-независимой и имеет единственное решение, т. е. коэффициенты $c_0, c_1, c_2, \dots, c_m$ всегда определяются однозначно. Такой полином называется полиномом Лагранжа. Его коэффициенты могут быть определены в явном виде из следующего выражения:

$$L_m(x) = \sum_{i=0}^m y_i \frac{(x-x_0)(x-x_1)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_m)}{(x_i-x_0)(x_i-x_1)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_m)}. \quad (I.7)$$

Имеется ряд других формул интерполирования многочленами, однако все они являются лишь иной формой записи формулы Лагранжа и при одинаковых узлах интерполирования дают одинаковые результаты.

Если необходимо последовательно вычислять значения $f(x)$ в точках $x=x_i+t\Delta x$, $t=\text{const} < 1$, $i=0, 1, 2, \dots$ при $\Delta x = \text{const}$, коэффициенты c_i можно вычислить заранее и затем многократно использовать. Расчеты в этом случае значительно упрощаются и сводятся к вычислению сумм вида

$$f(x_i + t\Delta x) = \sum_{i=0}^m f(x_i) c_i. \quad (I.8)$$

Ошибка интерполяции по формуле (I.7) равна

$$R_m(x) = f(x) - L_m(x) = \frac{f^{(m+1)}(\xi)}{(m+1)!} \omega_m(x), \quad (I.9)$$

где ξ — внутренняя точка отрезка $[x_{\min}, x_{\max}]$, $\omega_m = (x-x_0) \times \dots \times (x-x_1) \dots (x-x_m)$, $x_{\min} = \min(x_0, x_1, x_2, \dots, x_m, x)$, $x_{\max} =$

$= \max(x_0, x_1, \dots, x_m, x)$, R_m — остаточный член полинома Лагранжа.

Полагая

$$\|f^{(m+1)}(\xi)\| = \sup_{x \in [x_{\min}, x_{\max}]} |f^{(m+1)}(x)|,$$

получим

$$|R_m(x)| \leq \frac{\|f^{(m+1)}(\xi)\|}{(m+1)!} |\omega_m(x)|. \quad (I.9')$$

Таким образом, величина $R_m(x)$ оценки (I.9) определяется величинами $\|f^{(m+1)}(\xi)\|$ и $|\omega_m(x)|$. Первая из них определяется свойствами $f(x)$ и при получении $f(x_i)$ в результате измерений редко бывает известна, вторую в принципе можно изменять по своему усмотрению в порядке планирования эксперимента. Задача в этом случае ставится следующим образом: необходимо выбрать такие узлы x_i , чтобы величина

$$\sup_{[x_{\min}, x_{\max}]} |\omega_m(x)| = \min.$$

Доказано, что величина $|\omega_m(x)|$ достигает минимума, если в качестве узлов интерполирования брать корни многочленов Чебышева [22]:

$$x_i = \frac{1}{2} \left[(x_{\max} - x_{\min}) \cos \frac{2i+1}{2m} \pi + (x_{\max} + x_{\min}) \right]. \quad (I.10)$$

В этом случае

$$R_m(x) \leq \frac{\|f^{(m+1)}(x)\|}{(m+1)!} \frac{(x_{\max} - x_{\min})^{m+1}}{2^{2m+1}}. \quad (I.11)$$

По ряду причин, однако, эта возможность при обработке дискретных измерений практически не используется.

Легко видеть, что ошибка приближения полиномом Лагранжа степени m будет равна нулю для функций $f(x)$, представимых полиномами степени не выше m (поскольку для них $\|f^{(m+1)}(x)\| = 0$).

Если функция $f(x)$ принадлежит к классу функций, для которых $\|f^{(m+1)}(x)\| \neq 0$, интерполяция на отрезке $[x_{\min}, x_{\max}]$ при конечных значениях m будет производиться с некоторой ошибкой, причем оценка (I.9') дает обычно значительно завышенные значения ошибок интерполяции. Обычно без строгого обоснования считается, что чем выше степень полинома, тем меньше эта ошибка (значения $f(x_i)$ предполагаются точными). В то же время известно, что в некоторых случаях величина $R_m(x)$ не стремится к нулю даже при $m \rightarrow \infty$, а для случая равномерного расположения узлов интерполяции доказано только, что величина $R_m(x) \rightarrow 0$ при $m \rightarrow \infty$ лишь в том случае, когда

функция $f(x)$ целая¹. Так, например, при равноотстоящих узлах ($\Delta x = \text{const}$) интерполяционный процесс не сходится даже для такой простой функции, как

$$\hat{f}(x) = \frac{1}{1+x^2},$$

совпадающей, как известно, с выражением для аномалий силы тяжести, создаваемой горизонтальной бесконечной вещественной линией с глубиной залегания $H=1$ и значением коэффициента $2f\lambda=1$.

Рассмотрим случай, когда значения $f(x_i)$ заданы с ошибкой $f(x_i) = \varphi(x_i) - \xi(x_i)$, где $\xi(x_i)$ — ошибки наблюдений. Ясно, что сам принцип построения интерполяционного полинома $Q_m(x_i) = \varphi(x_i)$ становится нецелесообразным, поскольку равенство этих значений теперь не свидетельствует о близости функций $f(x)$ и $Q_m(x)$ даже в узлах интерполяции, точнее, оно справедливо в них с точностью до величины $\xi(x_i)$. Полином $Q_m(x)$ теперь одновременно интерполирует и функцию $\xi(x)$, которая при $x \neq x_i$ неизвестна и полностью определяется самим полиномом. Поскольку значения ошибок наблюдений $\xi(x_i)$ изменяются обычно более прихотливо, чем $f(x)$, проведение полинома высокой степени, имеющего большое число коэффициентов и, соответственно, степеней свободы, может стать неустойчивым: чем выше степень полинома, тем более сложную функцию он может описать и тем сильнее сказываются значения $\xi(x_i)$. Начиная с некоторого m точность приближения начинает падать, а амплитуда осцилляций функции $Q_m(x)$ между узлами увеличиваться, поэтому обычно ограничиваются полиномами низких степеней. Если число измерений значительно больше выбранной степени полинома, интерполирование можно производить методом «скользящего окна» — путем последовательного перемещения отрезка $[x_{\min}, x_{\max}]$ таким образом, чтобы в него всегда попадало число точек, равное m . Если значение $\Delta x = \text{const}$, расчеты удобно производить по формуле (I.8).

Более естественным и эффективным выходом из положения, когда необходимо, чтобы выполнялось условие $m < n$, является переход от интерполяции к аппроксимации. В этом случае рассматриваются следующие меры сравнения функций $f(x)$ и $Q_m(x)$:

$$I = \int_{x_{\min}}^{x_{\max}} p(x) [Q_m(x) - f(x)]^2 dx \quad (\text{I.12})$$

и

$$\delta = \max_{x_{\min} < x < x_{\max}} |Q_m(x) - f(x)|, \quad (\text{I.13})$$

* Функция $f(x)$ называется целой, если ее можно представить в виде сходящегося при всех значениях x ряда: $f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \dots + a_n(x - x_0)^n + \dots$

где $p(x)$ — заданная неотрицательная функция, называемая весом, обычно $p(x) = 1$.

В первом случае процесс построения полинома называется интегральной квадратичной аппроксимацией, во втором — равномерной аппроксимацией. Среднее квадратичное отклонение определяется по формуле

$$s = \sqrt{\frac{1}{x_{\max} - x_{\min}} \int_{x_{\min}}^{x_{\max}} p(x) [Q_m(x) - f(x)]^2 dx}. \quad (I.14)$$

Легко видеть, что в первом случае достигается аппроксимация в «среднем», для некоторых интервалов значений x отклонения могут быть значительными, не отражаясь существенно на значении s .

В случае равномерной аппроксимации из условия $\delta < \varepsilon$ следует, что $|Q_m(x) - f(x)| \leq \varepsilon$ во всех точках отрезка x_{\min}, x_{\max} .

При обработке данных измерений вид функции $f(x)$ часто неизвестен (в геологии это типичное явление), а сами измерения обычно производятся в дискретных точках x_i . Ясно, что значения I_m и δ в (I.12 и I.13) не могут быть исходными параметрами при построении полинома $Q_m(x)$. Наоборот, как правило, стоит задача определения этих величин, т. е. определения точности интерполяции на всем отрезке и интегрирования, поэтому интегральная квадратичная аппроксимация заменяется точечной квадратичной, а за меру отклонения полинома принимается величина

$$I_m = \sum_{i=0}^n p(x_i) [Q_m(x_i) - f(x_i)]^2. \quad (I.15)$$

Среднее квадратическое отклонение равно

$$s_m = \sqrt{\frac{1}{n} \sum_{i=0}^n p(x_i) [Q_m(x_i) - f(x_i)]^2}. \quad (I.16)$$

Аналогично можно ввести меру равномерной точечной аппроксимации

$$\delta_m = \max_{\{(x_i)\}=P} |Q_m(x_i) - f(x_i)|. \quad (I.17)$$

Однако теперь условие $\delta_m \leq \varepsilon$ однозначно свидетельствует о близости функций лишь в узлах интерполяции $|Q_m(x_i) - f(x_i)| \leq \varepsilon$, в то время как между узлами значения δ могут быть сколь угодно велики. Это же замечание полностью относится и к точечной квадратичной аппроксимации, где, к тому же, даже малость значений $|Q_m(x_i) - f(x_i)|$ гарантируется лишь в «среднем». Таким образом, если величина n значительна, в некоторых точках могут быть расхождения, значительно превышающие значения s_m .

задача квадратичной аппроксимации функции $f(x)$, непрерывной на отрезке $[a, b]$, упрощается. Необходимые коэффициенты вычисляются по формуле

$$c_j = \frac{\int_a^b f(x) \psi_j(x) dx}{\int_a^b \psi_j^2(x) dx} \quad (j = 0, 1, 2, \dots, m) \quad (I.22)$$

и называются коэффициентами Фурье функции $f(x)$ относительно заданной ортогональной системы функций $\{\psi_i(x)\}$. Таким образом, например, если приближение функции с периодом 2π производить с помощью тригонометрического полинома

$$Q_m(x) = \frac{a_0}{2} + \sum_{k=1}^m (a_k \cos kx + b_k \sin kx), \quad (I.23)$$

его коэффициенты необходимо определять по формулам

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx,$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx.$$

Полином (I.23) называется тригонометрическим полиномом Фурье, а его коэффициенты — тригонометрическими коэффициентами Фурье функции $f(x)$. Известно, что обобщенный полином с коэффициентами Фурье данной функции обладает наименьшим квадратичным отклонением от этой функции по сравнению со всеми другими обобщенными полиномами того же порядка m . Ясно, что если функция $f(x)$ известна лишь в дискретных точках, задача вычисления c_j фактически требует интерполяции $f(x)$ на весь отрезок $[a, b]$ для вычисления интегралов в (I.22). Эта задача особенно трудна при нерегулярной сети измерений в двухмерном случае.

Остановимся теперь на некоторых особенностях интерполяции двух переменных. Эта задача является несравненно более сложной по объему вычислений, кроме того, имеются различия принципиального порядка [22]. Запишем полином в следующем виде:

$$Q_m(x, y) = c_{00} + c_{10}x + c_{01}y + c_{20}x^2 + c_{11}xy + c_{02}y^2 + \dots + c_{m0}x^m + c_{m-1,1}x^{m-1}y + \dots + c_{0m}y^m. \quad (I.24)$$

Отсюда следует, что для определения коэффициентов c_j необходимо, чтобы число точек было равно $(m+1)(m+2)/2$. Таким образом, количество узлов здесь фиксировано при $m=1, n=3; m=2, n=6; m=3, n=10$ и т. д.

Если рассмотреть определители решаемой системы уравнений, то можно заметить, что при $m=1$ он будет обращаться в нуль, если точки лежат на одной прямой, а при $m=2$ — если шесть точек лежат на одной кривой второго порядка. Таким образом, узлы не могут быть расположены произвольно. Большие трудности возникают и при оценке точности интерполяции.

§ 2. О точности интерполяции геологических характеристик

Вопросы определения точности интерполяции геологических характеристик — это вопросы первостепенной важности. Чем точнее удастся определить точность восстановления функций $\varphi(x)$, тем более правильно можно рассчитать необходимый шаг наблюдений, т. е. именно этим в значительной части определяется уровень затрат на производство работ. Для того чтобы иметь гарантированную точность на всем отрезке интерполирования, величину ошибки будем определять максимальным значением разности функций $\varphi(x)$ и $\bar{\varphi}(x)$, т. е. нормой функции ошибок в метрике пространства C непрерывных на отрезке функций:

$$\|\delta\|_C = \max_{x_{\min} < x < x_{\max}} |\varphi(x) - \bar{\varphi}(x)|, \quad (I.25)$$

$$\|\delta\|_C = \max_{(x, y) \in D} |\varphi(x, y) - \bar{\varphi}(x, y)|.$$

В первом параграфе уже отмечалось, что если функция $\varphi(x)$ определена в дискретных точках наблюдений $\varphi(x_i)$, то ее поведение между узлами можно определить различными способами, причем значения могут сильно различаться друг от друга. Для того чтобы получить оценки точности интерполяции, совершенно необходимо сделать некоторые предположения о свойствах интерполируемых функций — об их принадлежности к некоторому классу и о значениях параметров, характеризующих этот класс. Чем удачнее мы определим этот класс и особенно параметры, необходимые для расчетов, тем лучше будут оценки точности. Наиболее естественным и эффективным было бы определение соответствующего класса функций на основе априорной информации, т. е. данных о физической или физико-геологической природе приближаемых характеристик. Так, например, потенциальные функции, измеряемые в геофизике (аномалии силы тяжести), удобно отнести к классу гармонических функций $U(H)$, введенному В. Н. Страховым. Это позволило ему получить предельные оценки точности восстановления функций класса $U(H)$, заданных в узлах регулярной сети [85]. В боль-

шинстве случаев физическая природа количественных характеристик геологических процессов весьма сложна, а математической теории их практически не существует. В результате пока нет никаких конструктивных критериев для определения классов функций. В настоящее время имеется лишь возможность замены истинных свойств геологических признаков (за исключением, конечно, тех случаев, когда они все-таки могут быть определены) такими классами функций, которые, с одной стороны, достаточно хорошо приближают их, т. е. близки к ним по некоторому набору характеристик, а с другой стороны, позволяют определять значения параметров, необходимых для оценок. Не менее важным обстоятельством является и возможность получения хороших оценок точности в выбранных классах.

Рассмотрим с этих позиций возможность отнесения функций $f(x)$ к классу алгебраических полиномов. В этом случае наиболее целесообразно и в качестве приближающей функции $\bar{f}(x)$ выбрать интерполяционный полином. Тогда для получения оценок по формуле (I.9) необходимо знать производные $m+1$ его порядка. В то же время по эмпирическим данным можно более или менее надежно определить только значение первой производной функции $f'(x)$ и иногда $f''(x)$. Отсюда следует, что оценки могут быть сделаны лишь для полинома первой степени. Этого явно недостаточно, тем более, что величины, получаемые по формуле (I.9), являются, как правило, сильно завышенными. Выбирая различную степень полинома и изменяя расстановку узлов интерполяции, мы можем получить существенно разные результаты. Значение $R_m(x)$ может сильно зависеть от m . В то же время нет критериев для определения такого значения m , которое соответствует действительности наилучшим образом. Все это в данном случае делает использование классических полиномиальных приближений при оценке точности нецелесообразным. Другим выводом является необходимость получения расчетных параметров по самой функции $f(x)$, а не по ее производным высоких степеней.

Использование многочленов является конструктивным лишь при изучении непрерывных функций, заданных на конечном отрезке. В то же время, несмотря на то, что при решении практических задач в большинстве случаев рассматриваются функции, заданные на конечном интервале, часто оказывается очень полезным и удобным задание функции на всей оси. В частности, это позволяет естественным образом перейти к изучению спектральных характеристик изучаемых функций.

Учитывая высказанные соображения, приходится, к сожалению, констатировать, что выбор класса функций S , $f \in S$ в значительной мере произволен. Таким образом, при его определении целесообразно учитывать возможность хорошей практической аппроксимации функций f и соображения, связанные с

возможностью создания необходимой теории расчетов. Несомненно лишь, что функции φ можно считать непрерывными (за редкими исключениями, такими как, например, отметки структурных поверхностей при разрывных дислокациях). Подобным образом поступают обычно в теории случайных процессов, аппроксимируя реальные автокорреляционные функции одной из трех «удобных» моделей нормированной функции автокорреляции:

$$B_H(\tau) = \exp\left(-\frac{\tau}{r}\right), \quad (1.26)$$

$$B_H(\tau) = \exp\left(-\frac{\tau^2}{r^2}\right), \quad (1.27)$$

$$B_H(\tau) = \frac{a^2}{\tau^2 + r^2}. \quad (1.28)$$

Различие этих моделей в первую очередь определяется не видом функциональной зависимости, а значением параметра r . Его можно подобрать таким образом, что функции $B_H(\tau)$ будут отличаться незначительно, если учесть приближенность их эмпирического определения. В нашем случае, очевидно, наиболее естественно воспользоваться последней из них. Этот вид $B_H(\tau)$ (1.28) непосредственно связан с потенциальными функциями (определяет их нормированную функцию автокорреляции $B_H(\tau)$):

$$U(x, H) = \frac{cH}{H^2 + (x - \xi)^2}, \quad H > 0, \quad (1.29)$$

с помощью которых, как показывает опыт [11, 16, 84], можно хорошо аппроксимировать геолого-геофизические признаки*. При замене $\varphi(x)$ функциями $U(x, H)$ для получения оценок точности интерполяции необходимое наиболее подходящее значение H получается путем приближенной минимизации величин

$$\|U(x, H) - \varphi(x)\|_C = \max_{x_{\min} \leq x \leq x_{\max}} |U(x, H) - \varphi(x)| \quad (1.30)$$

или

$$\|U(x, H) - \varphi(x)\|_{L^2} = \left(\int_{x_{\min}}^{x_{\max}} [U(x, H) - \varphi(x)]^2 dx \right)^{1/2}.$$

* Для аномалий силы тяжести (Δg) функции $U(x, H)$ полностью соответствуют физике процесса, являясь экстремальными, в смысле величины ошибок интерполяции, функциями класса $U(H)$.

Величину H легко определить, используя известный в гравиметрии способ характерных точек: H равно абсциссе x_1 точки, где $U(x_1, H) = 0,5 U(0, H) = 0,5 \max U(x, H)$ в одномерном случае или $1,3 x_1$ в двухмерном. Другим способом является определение H через значения радиуса автокорреляции: $H = \pi \tau_0$ для одномерного случая и $H = 2,8 \tau_0$ для двухмерного [34].

Количество необходимых функций $U(x, H)$ с различными значениями s и H зависит от сложности аппроксимируемой функции; поскольку в данном случае аппроксимация производится лишь для получения оценок точности, их количество будет небольшим. При этом последовательно выбираются аномальные особенности различной протяженности и амплитуды — сначала более протяженные, затем более мелкие [24]. После этого необходимые величины оценок определяются для каждой аппроксимирующей компоненты $U(x, H)$. Обычно можно вообще ограничиться одним экстремальным или средним значением H , определив соответственно максимальную или среднюю ошибку интерполяции.

В дальнейшем будут существенно использоваться спектральные характеристики функций, поэтому важным обстоятельством является принадлежность функций $U(x, H)$ пространству $L^1(-\infty, \infty)$ абсолютно интегрируемых функций $U(x, H) \in L^1(-\infty, \infty)$ при $H > 0$ и, следовательно, существование преобразования Фурье и его обращения:

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} U(x, H) e^{-i\omega x} dx, \quad (I.31)$$

$$U(x, H) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(\omega) e^{i\omega x} d\omega.$$

Учитывая (I.29), (I.31), можно определить класс функций $\Phi \in U'(H)$ дополнительным условием, аналогично тому, как это сделано в работе [85]:

$$|F(\omega)| \leq K_1 H e^{-|\omega| H}, \quad (I.32)$$

$$K_1 = \frac{1}{H} \max_{\omega} |F(\omega)|. \quad (I.33)$$

Таким образом, к классу $U'(H)$ относятся функции, принадлежащие пространству $L^1(-\infty, \infty)$, спектр которых ограничен оценкой (I.32).

Прежде чем перейти к методике расчета ошибок, отметим, что имеет смысл оценивать точность интерполяции только тех составляющих геологического признака, которые в достаточной мере обеспечены шагом наблюдений. Если размеры некоторых проявлений признака меньше шага измерений, величина оши-

бок интерполяции их достигает 100% и более. Будем называть эти составляющие геологических характеристик непредставительными. Действительно, на региональной карте, составляемой по редкой сети наблюдений, не выделяют локальные детали, не подтвержденные достоверно значениями в ряде соседних узлов. Говоря на языке спектральных представлений, отбрасывается высокочастотная часть геологического признака, проявляющаяся лишь в отдельных точках наблюдений. С учетом изложенного можно ввести понятие о некоторой граничной частоте, являющейся функцией шага измерений Δx , таким образом, чтобы ошибки интерполяции без учета непредставительной части признаков были сравнительно небольшими.

Сформулируем теперь понятие оптимальной точности интерполяции. Пусть для простоты дискретная сеть измерений является равномерной, $y_i - y_{i-1} = \Delta x = \text{const}$. Как известно, согласно теореме Котельникова, функция $\tilde{\varphi}(x) \in W_\sigma^*$ может быть точно восстановлена на всей оси $-\infty < x < \infty$ по своим значениям в регулярной сети дискретных точек $\tilde{\varphi}(k\Delta x)$, $k=0, \pm 1, \pm 2, \pm \dots, \pm \infty$ с помощью ряда Котельникова [93, 94, 95]:

$$\tilde{\varphi} = \sum_{k=-\infty}^{\infty} \tilde{\varphi}(k\Delta x) \frac{\sin \frac{\pi}{\Delta x} (x - k\Delta x)}{\frac{\pi}{\Delta x} (x - k\Delta x)}, \quad \frac{\pi}{\Delta x} \leq \sigma, \tilde{\varphi} \in W_\sigma. \quad (I.34)$$

Высокие частоты ($|\omega| < \sigma = \frac{\pi}{\Delta x}$), которые могут содержаться в спектре функции $\varphi(x)$, не могут быть выделены и сливаются с низкочастотными составляющими ($|\omega| \leq \sigma$). Таким образом, если в узлах задана функция $\varphi \in W_\sigma$, с помощью ряда (I.34) можно восстановить лишь целую функцию $\tilde{\varphi}$ степени $\sigma = \frac{\pi}{\Delta x}$, такую, что $\varphi(k\Delta x) = \tilde{\varphi}(k\Delta x)$, $k=0, \pm 1, \pm \dots, \pm \infty$, т. е. $\varphi(x_i) = \tilde{\varphi}(x_i)$. Такое восстановление функции φ будем называть оптимальным, поскольку без дополнительной информации более точное восстановление функции φ , $-\infty < x < \infty$, невозможно. При $\Delta x \rightarrow 0$ $\tilde{\varphi}(x) \rightarrow \varphi(x)$ [8, 16].

На самом деле функция φ задана на конечном интервале $x_{\min} \leq x \leq x_{\max}$. Казалось бы, в этом случае оптимальным целесообразно считать восстановление функции полиномом Лагранжа $L_m(x)$ степени $m = 2k$. Однако мы не можем выбрать от-

* Функция $f(x)$ с интегрируемым на оси $0x$ квадратом $f(x) \in L^2(-\infty, \infty)$ называется функцией с ограниченным спектром (класса W_σ), если ее спектр $F(\omega)$ удовлетворяет условию $F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx = 0$, $|\omega| > \sigma$. Наименьшее из чисел σ , для которого выполняется это условие, называется степенью функции с ограниченным спектром.

счетные значения произвольно (например, в узлах Чебышева, чтобы ошибки были минимальными), а в случае равномерного шага Δx последовательность интерполяционных многочленов Лагранжа $L_m(x)$ является сходящейся лишь для целых функций, в частности $f(x) \in W_\sigma$. В то же время, ввиду равномерной сходимости приближения рядом (I.34) и $L_m(x)$ целых функций*, начиная с некоторого значения $k \geq n$ результаты разложений будут отличаться сколь угодно мало, т. е.

$$|\tilde{\varphi}_n(x) - L_m(x)| < \varepsilon, \quad (\text{I.35})$$

где $\tilde{\varphi}_n(x)$ — значения функции $\tilde{\varphi}(x)$, определенные рядом (I.34) с числом членов, равным n . В большинстве случаев результаты будут практически совпадать, если n равно нескольким десяткам точек. Таким образом, и в этом случае целесообразно оставить прежнее определение оптимальности, тем более, что, как было показано, использование теории полиномиальных приближений не приводит к нужному результату.

Восстановление функции φ с помощью конечного ряда (I.34) будем называть асимптотически оптимальным. Такой подход позволяет получить конструктивные оценки величины δ в формуле (I.25) [8, 16].

Ошибки интерполяции δ можно разделить теперь на два слагаемых. Первое (δ_1) определяется той точностью, которую теоретически можно получить с помощью оптимального метода, второе (δ_2) — ошибками конкретного метода интерполяции, т. е. $\|\delta\|_C \leq \|\delta_1\|_C + \|\delta_2\|_C$, где

$$\|\delta_1\|_C = \max_{-\infty < x < \infty} |\varphi(x) - \tilde{\varphi}(x)|, \quad (\text{I.36})$$

$$\|\delta_2\|_C = \max_{-\infty < x < \infty} |\bar{\varphi}(x) - \tilde{\varphi}(x)|. \quad (\text{I.37})$$

Оценки δ_1 и δ_2 имеют принципиально различный смысл. Первая из них характеризует практически неустранимую погрешность и зависит только от свойств восстанавливаемой функции $\varphi(x)$. Вторая определяет расхождение между конкретной аппроксимирующей функцией $\bar{\varphi}(x)$ и той целой функцией, которая может быть восстановлена при помощи ряда Котельникова. Ясно, что при соответствующем выборе $\varphi(x)$ величину $\|\delta_2\|_C$ можно сделать сколь угодно малой. В соответствии с введенным выше определением метод интерполяции будет оптимальным, если $\|\delta_2\|_C = 0$. Соответственно для асимптотически оптимальных оценок δ_n будем иметь

$$\|\delta_{1,n}\|_C = \max_{x_{\min} < x < x_{\max}} |\varphi(x) - \tilde{\varphi}_n(x)|, \quad (\text{I.36}')$$

$$\|\delta_{2,n}\|_C = \max_{x_{\min} < x < x_{\max}} |\bar{\varphi}(x) - \tilde{\varphi}_n(x)|. \quad (\text{I.37}')$$

* В данном случае можно считать, что приближается целая функция $\tilde{\varphi}(k\Delta x) = \varphi(k\Delta x)$, $\sigma = \pi/\Delta x$.

В большинстве случаев ограниченность числа пунктов измерений n лишь незначительно ухудшает полученные оценки и может не учитываться вследствие приближенности определяемых величин.

Поскольку цель данного параграфа — найти оптимальные оценки точности интерполяции, которые могли бы служить эталоном для сравнения практически используемых методов, определим величину $\|\delta_1\|_C$.

Учитывая, что для $\varphi(x) \in U'(H)$ существует прямое и обратное преобразование Фурье, вместо (I.36) запишем

$$\delta_1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |F(\omega) - \tilde{F}(\omega)| e^{i\omega x} d\omega, \quad (\text{I.38})$$

где $F(\omega)$ — преобразование Фурье функции $\varphi(x)$ или, короче, спектр функции $\varphi(x)$, $\tilde{F}(\omega)$ — спектр функции $\varphi(x)$.

Воспользовавшись суммационной формулой Пуассона [18], можно показать, что

$$\tilde{F}(\omega) = \begin{cases} \sum_{n=-\infty}^{\infty} F(\omega + 2n\sigma), & |\omega| \leq \sigma \\ 0, & |\omega| > \sigma; \end{cases} \quad (\text{I.39})$$

$\sigma = \frac{\pi}{\Delta x}$ — граничная частота.

Отсюда следует

$$\begin{aligned} \delta_1 &= \frac{1}{\sqrt{2\pi}} \left\{ \int_{\sigma}^{\infty} F(\omega) e^{i\omega x} d\omega + \int_{-\infty}^{-\sigma} F(\omega) e^{i\omega x} d\omega - \right. \\ &\quad \left. - \sum_{n=1}^{\infty} \int_{-\sigma}^{\sigma} [F(\omega + 2n\sigma) + F(\omega - 2n\sigma)] e^{i\omega x} d\omega \right\} \leq \\ &\leq \frac{1}{\sqrt{2\pi}} \left\{ \int_{\sigma}^{\infty} |F(\omega)| d\omega + \int_{-\infty}^{-\sigma} |F(\omega)| d\omega + \right. \\ &\quad \left. + \int_{-\sigma}^{\sigma} \left| \sum_{n=1}^{\infty} [F(\omega + 2n\sigma) + F(\omega - 2n\sigma)] \right| d\omega \right\}. \quad (\text{I.40}) \end{aligned}$$

В силу неравенства (I.32) для $\varphi(x) \in U'(H)$

$$\begin{aligned} \|\delta_1\|_C &\leq \frac{K_1 H}{\sqrt{2\pi}} \left\{ 2 \int_{\sigma}^{\infty} e^{-|\omega|H} d\omega + 2 \sum_{n=1}^{\infty} e^{-2n\sigma H} \int_0^{\sigma} (e^{|\omega|H} + \right. \\ &\quad \left. + e^{-|\omega|H}) d\omega \right\} = \frac{4K_1 e^{-\sigma H}}{\sqrt{2\pi}}; \quad \sigma = \frac{\pi}{\Delta x}. \quad (\text{I.41}) \end{aligned}$$

Для функции $U(x, H)$

$$F(\omega) = \frac{c}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-i\omega x}}{(x-\xi)^2 + H^2} = c \sqrt{\frac{\pi}{2}} e^{-|\omega|H - i\omega\xi}. \quad (I.42)$$

Учитывая это, на основании формул (I.32), (I.33) и (I.41) получим

$$\|\delta_1\|_c \leq \frac{2e^{-\sigma H}}{H}, \quad \varphi(x) \in U'(H), \quad c = 1. \quad (I.43)$$

Аналогичным образом получим величину $\|\delta_1\|_c$ для двухмерного случая

$$\begin{aligned} \|\delta_1\|_c &= \max_{\substack{-\infty < x < \infty \\ -\infty < y < \infty}} |\varphi(x, y) - \tilde{\varphi}(x, y)| \leq \\ &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(U, V) - \tilde{F}(U, V)| dU dV, \end{aligned} \quad (I.44)$$

где $F(U, V)$ и $\tilde{F}(U, V)$ — двукратные преобразования Фурье функций $\varphi(x, y)$ и $\tilde{\varphi}(x, y)$ соответственно, $\varphi(x, y)$ — целая функция с ограниченным спектром, определяемая разложением типа (I.34):

$$\tilde{\varphi}(x, y) = \frac{\pi}{(\Delta x)^2} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \varphi(p\Delta x, q\Delta x) \frac{\sin \frac{\pi}{\Delta x} (x - p\Delta x) \sin \frac{\pi}{\Delta x} (y - q\Delta x)}{(x - p\Delta x)(y - q\Delta x)}. \quad (I.45)$$

Преобразование Фурье функции $\tilde{U}(x, y)$ определяется по формуле

$$\tilde{F}(U, V) = \begin{cases} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} F(U + 2p\sigma, V + 2q\sigma), & |U| \leq \sigma, \quad |V| \leq \sigma, \\ 0 & |U| > \sigma, \quad |V| \leq \infty, \\ & |U| < \sigma, \quad |V| > \sigma. \end{cases} \quad (I.46)$$

Таким образом,

$$\begin{aligned} \|\delta_1\|_c &\leq \frac{1}{2\pi} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(U, V)| dU dV - \int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} F(U, V) dU dV + \right. \\ &+ \left. \frac{1}{2\pi} \int_{-\sigma}^{\sigma} \int_{-\sigma}^{\sigma} \left| \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} F(U + 2p\sigma, V + 2q\sigma) - F(U, V) \right| dU dV. \right] \end{aligned} \quad (I.47)$$

Функция $U(x, y, H)$ равна

$$U(x, y, H) = \frac{cH}{(x^2 + y^2 + H^2)^{3/2}} \quad (I.48)$$

и ее спектр при $c=1$

$$F(U, V) = e^{-H\sqrt{U^2+V^2}}. \quad (I.49)$$

Соответственно изменяются и условия (I.32) и (I.33):

$$|F(U, V)| \leq K_2 e^{-H\sqrt{U^2+V^2}}, \quad (I.50)$$

$$K_2 = \frac{1}{H^2} \max_{U, V} |F(U, V)|. \quad (I.51)$$

После некоторых преобразований приходим к следующему результату [11, 16]:

$$\|\delta_1(x, y)\|_C < \frac{2}{H^2} e^{-\sigma H} (\sigma H + 1), \quad (I.52)$$

$$\mu = \|\delta_1\|_C / \|U\|_C.$$

Таблица 1

Значения относительных ошибок для одномерного и двухмерного случаев

$H/\Delta x$	0,5	1,0	1,5	2,0	2,5	3,0
$\tau_0/\Delta x$	1,57	3,14	4,71	6,28	7,85	9,42
$\mu_1(x) \%$	42	8,6	1,8	0,36	0,08	0,02
$\tau_0/\Delta x$	—	2,8	4,2	5,6	7,0	8,4
$\mu_1(x, y) \%$	—	36	10	2,7	0,7	0,2

Из табл. 1 видно, что уже при $\tau_0/\Delta x > 3$ в одномерном случае и $\tau_0/\Delta x > 5$ в двухмерном теоретические ошибки интерполяции невелики и во многих случаях не превышают уровня ошибок наблюдений. Значения ошибок получены для равномерной сети исходных данных $\Delta x = \Delta y = \text{const}$, тогда как обычно наблюдения производятся неравномерно.

Если в ряде Котельникова значения композиционной функции одни и те же для всех x_i :

$$Q_i(x) = \frac{\sin \sigma(x - i\Delta x)}{\sigma(x - i\Delta x)}$$

и достигают максимума, равного 1 в точках x_i , то при неравномерных отсчетах наибольшее значение $Q_i(x)$ может значительно возрасти, причем максимум достигается в интервале между отсчетами, следовательно, возрастают требования к точности исходных данных, кроме того, композиционные функции становятся слишком сложными. Это обстоятельство в совокупности с медленной сходимостью подобных рядов и отсутствием разработанных способов одновременной фильтрации

случайных ошибок измерений приводит к тому, что они используются, как правило, лишь для решения теоретических вопросов. В данном случае важно, что все изложенное остается верным и для случая неравномерной сети, если вместо Δx использовать среднее значение $\overline{\Delta x}$, такое что

$$\lim_{i \rightarrow \infty} \frac{x_i - x_{-i}}{2i} = \overline{\Delta x}, \quad x_{i+1} - x_i > 0 \text{ при } i = 0, \pm 1, \pm 2, \pm \dots, \quad (1.53)$$

т. е. функция $\tilde{\varphi} \in W\sigma$, $\sigma = \pi/\Delta x$ может быть восстановлена точно [49, 95]. Конечно, такие оценки могут быть эффективными лишь при не слишком больших колебаниях Δx_i . Их конструктивность в этом случае подтверждается опытом проведения расчетов при изучении ошибок интерполяции на теоретических примерах [8, 11, 16]. Таким образом, используя полученные результаты — значения μ_1 в табл. 1 и определив приближенно значения H или τ_0 по уже проведенным наблюдениям или по аналогии с ранее изучавшимися объектами, мы можем оценить теоретически возможную точность интерполяции по имеющимся данным или, наоборот, определить необходимый для достижения нужной точности оптимальный (максимальный) шаг измерений.

В последнее время в геологии все чаще используются понятия теории случайных функций, обычно без достаточного на то обоснования. Интересно отметить, однако, что изложенный детерминированный подход, основанный на использовании теории функций с финитным спектром, может быть перенесен на случайные процессы с финитным спектром [95]: если $\varphi(x)$ ($-\infty < x < \infty$) — стационарный случайный процесс с финитным спектром на $(-\sigma, \sigma)$, то почти для всех выборочных функций (реализации) $\tilde{\varphi}(x)$ справедливо представление (1.34). Следовательно, полученные результаты можно трактовать и в терминах теории случайных функций, аппроксимируя исходный сигнал случайными функциями с радиусом автокорреляции τ_0 (связь которого с величинами H функций $\varphi \in U'(x, H)$ указана выше) и нормированной функцией автокорреляции типа (1.28).

§ 3. Интерполяция алгебраическими полиномами

Методы интерполяции алгебраическими полиномами очень детально освещены в математической литературе. Реализация их сравнительно несложна, поэтому они получили довольно широкое распространение.

Наиболее простым является построение алгебраического полинома степени m по методу наименьших квадратов одновре-

менно для всей площади исследований:

$$\begin{aligned} f(c_0, c_1, \dots, c_m) &= \sum_{i=0}^n [Q_m(x_i, y_i) - \varphi(x_i, y_i)]^2 = \\ &= \min, \{(x_i, y_i)\} = P, \end{aligned} \quad (I.54)$$

коэффициенты которого c_j определяются из системы нормальных уравнений (I.20). Степень полинома подбирается таким образом, чтобы средняя квадратическая ошибка точечного приближения s_m , определяемая по формуле (I.16), была достаточно мала. Ясно, что этот алгоритм можно использовать лишь при сравнительно простом характере приближающей функции, так как в противном случае приходится конструировать полиномы слишком больших степеней. Помимо общеизвестных вычислительных трудностей, связанных с плохой обусловленностью систем нормальных уравнений высоких порядков, это приводит к неравномерному приближению на различных участках. Поскольку производится приближение «в среднем», в отдельных точках, которые могут быть очень информативными, разность $Q_m(x_i, y_i) - \varphi(x_i, y_i)$ может достигать недопустимо больших значений. Повышение степени полинома приводит к сокращению этих разностей, но в этом случае становится слишком неустойчивым поведение полинома между узлами, поэтому на практике полиномиальная интерполяция на всей площади применяется обычно лишь для приближения региональных компонент исследуемых признаков, т. е. для построения поверхностей тренда, или для описания сравнительно простых явлений, когда на участке имеется небольшое число экстремумов (чаще всего один).

Важной положительной особенностью приближения полиномами является возможность простого и удобного использования дополнительной информации. Например, если картируемая величина корреляционно связана с какой-либо другой известной на всей площади характеристикой, то включение последней в аппроксимирующий полином в качестве дополнительной независимой переменной приводит к повышению точности интерполяции [13]. Однако это относится к тому случаю, когда сеть исходных значений слишком редка, т. е. к случаю, в котором, как было определено выше, производится геологическая интерполяция, иными словами, происходит совмещение процедуры интерполяции и прогнозирования. Этот вопрос будет рассмотрен в главе III.

Другой вариант полиномиальной интерполяции заключается в следующем: среднеквадратический полином строится в каждой расчетной точке по исходным узлам, расположенным в некоторой ее локальной окрестности — окне, причем для ослабления влияния удаленных точек вводится функция веса измерений, убывающая от центра (расчетная точка) к периферии.

Большого сглаживания можно достигнуть, используя вес вида

$$p_i = \left(\frac{R^2 - d_i^2}{d_i^2 + b^2} \right)^N, \quad (I.59)$$

где величина b определяется опытным путем. На основании опытных данных получено «наилучшее» значение N , $N=2$. Значение R в расчетах является постоянным. Оно выбирается экспериментально для каждого участка, но обязательно должно быть таким, чтобы внутри круга попадало не менее 6 точек — по числу коэффициентов c_j в уравнении (I.57).

Программа осуществляет поиск сомнительных значений $\varphi(x_i, y_i)$. Этот тест основан на анализе отклонения значений $\varphi(x_i, y_i)$ от значений, получаемых путем построения по методу наименьших квадратов полинома первой степени,

$$Q_1(x, y) = c_0 + c_1x + c_2y \quad (I.60)$$

в окне, рекомендуемый радиус которого равен $r=0,4R$. Поскольку такая отбраковка ненадежна, окончательное решение остается за пользователем, которому выдается ведомость сомнительных точек.

Точность интерполяции исходной функции и возможности фильтрации случайных ошибок регулируются в данном случае параметрами — степенью полинома и размерами скользящего окна. Повышение точности требует соответствующего повышения степени полинома или сокращения размеров окна, а усиление эффекта фильтрации — выполнения обратных процедур. Фиксация хотя бы одного из этих параметров приводит к принципиальным затруднениям. Например, при заданной степени полинома m возможность сокращения размеров окна существенным образом ограничивается неравномерностью исходной сети, так как для построения однозначно разрешимой системы нормальных уравнений необходимо, чтобы в окно попадало не меньше чем $(m+1)(m+2)/2$ узлов. В то же время повышение степени полинома при фиксированных размерах окна прямо противоречит требованию фильтрации случайных ошибок.

Серьезным недостатком рассматриваемого метода является его чувствительность к расположению исходных точек в скользящем окне — наличием сгущений узлов и купюр, слабо охарактеризованных наблюдениями. Возникающие при этом эффекты частично подавляются введением функции веса, однако полностью исключить их не удастся. Например, если все узлы, попавшие в окно, располагаются в одном или двух его квадрантах, то, независимо от выбора веса, вместо интерполяции исходной функции будет производиться ее экстраполяция, что, несомненно, скажется на точности результатов. При резко неравномерном распределении узлов по квадрантам окна пре-

обладающим может оказаться суммарный вес периферийных сгущений узлов, тогда как вес действительно информативных центральных точек будет сравнительно мал. Задавая различные функции веса, можно, по-видимому, воспроизвести такие вполне реальные и в то же время парадоксальные ситуации, когда увеличение числа определенным образом расположенных исходных нунктов будет приводить не к повышению, а к понижению точности интерполяции. Таким образом, необходимы анализ характера расположения узлов в окне и введение соответствующих «весов», зависящих, например, от количества точек в каждом квадранте.

Теоретический подход к вопросам выбора степени полинома, размеров окна и веса измерений в зависимости от характера приближаемой функции, уровня помех и структуры исходной сети пока не разработан, поэтому все эти параметры определяются чисто эмпирически по результатам модельных расчетов. Ясно, однако, что параметры «оптимальные» на модельных примерах или в одном из районов могут оказаться непригодными в другом. В решающей роли правильного выбора параметров нетрудно убедиться, рассматривая примеры, приведенные в работе [61].

Некоторые из отмеченных выше недостатков устранены в работе [23]. Аппроксимация производится алгебраическим полиномом степени m

$$Q_m(x, y) = \sum_{k=0}^m \sum_{j=0}^{m-k} c_{kj} x_i^k y_i^j, \quad (I.61)$$

коэффициенты которого определяются по методу наименьших квадратов. Степень полинома выбирается автоматически, в зависимости от количества узлов, попадающих в окно, так чтобы она была максимальной при выполнении неравенства

$$(m+1)(m+2)/2 \leq N, \quad (I.62)$$

где N — число точек в окне.

Предполагается, что съемка проведена с учетом сложности конфигурации картируемого признака на отдельных участках, т. е. сгущения замеров соответствуют участкам усложнения поведения признака. Если это не так, то может оказаться, что для одних участков выбранный размер окна обеспечивает необходимую точность интерполяции при разумной степени полинома, а в других (участках детализации) — степень полинома, определяемая по формуле (I.62), может стать слишком высокой. Поэтому предусмотрена возможность сужения окна до тех пор, пока в нем не останется такое количество точек, которое необходимо для построения полинома заданной предельной степени. Максимальное значение этой предельной степени может достигать шести. При переходе к следующей точке вос-

становливается заданный размер окна и подобный анализ производится снова.

Как уже отмечалось, решение системы нормальных уравнений может быть неустойчивым. В связи с этим в работе [23] уделено большое внимание проблеме их обусловленности. С целью ее повышения координаты узлов вычисляются в подвижной системе координат с началом в центре окна, а если точки группируются в одной половине окна, то в центре рассеяния точек. Затем координаты нормируются по половине окна (R), в результате чего элементы матрицы получаются приблизительно одного порядка.

Другой причиной плохой обусловленности является неблагоприятное расположение замеров, в результате чего определитель системы может обращаться в нуль. Подобные ситуации будут встречаться тем чаще, чем разнообразнее кривые, на которых не должны располагаться узлы, т. е. чем больше вероятность того, что узлы окажутся вблизи от них. Эта вероятность в общем случае растет с повышением степени полинома. Однако уже при $m=2$ практические ситуации, когда система нормальных уравнений не может быть решена, встречаются не так уже редко, поэтому предлагается следующая процедура. Система нормальных уравнений наращивается постепенно, в соответствии с последовательным добавлением членов полинома (I.61) при возрастании m . Одновременно изучается обусловленность системы уравнений. Коэффициенты и, соответственно, члены полинома, приводящие к существенному ухудшению устойчивости системы уравнений, отбраковываются. Пример интерполяции с помощью этой программы изображен на рис. 1.

Разработанный алгоритм допускает возможность учета неравноточности замеров. Составленная программа для ЭВМ позволяет умножать вес типа (I.58) на некоторую оценку достоверности наблюдений, в результате аппроксимирующая функция будет «подтягиваться» к значениям узлов с большими весами. Интересной особенностью является возможность точного проведения функции $\varphi(x, y)$ через некоторые «опорные» точки, т. е. таким образом, что $\varphi(x_i, y_i) = \bar{\varphi}(x_i, y_i)$. Алгоритм основан на решении задачи Лагранжа об отыскании условного минимума:

$$\Phi = \frac{1}{2} \sum_1^n [\varphi(x_i, y_i) - \bar{\varphi}(x_i, y_i)]^2 p_i + \sum_{j=1}^k \lambda_j \theta_j \rightarrow \min, \quad (\text{I.63})$$

где λ_j — формальный параметр, k — количество фиксируемых замеров, $\theta_j = \varphi(x_j, y_j) - \bar{\varphi}(x_j, y_j) = 0$ — условие равенства значений φ и $\bar{\varphi}$ в точке (x_j, y_j) . Минимизация (I.63) приводит к расширенной за счет коэффициентов λ_j системе нормальных

уравнений. Программа предусматривает возможность фиксации одновременно до 5 точек, $k \leq 5$.

Для обработки профильных съемок разработан вариант программы, использующий прямоугольное окно, ориентирован-

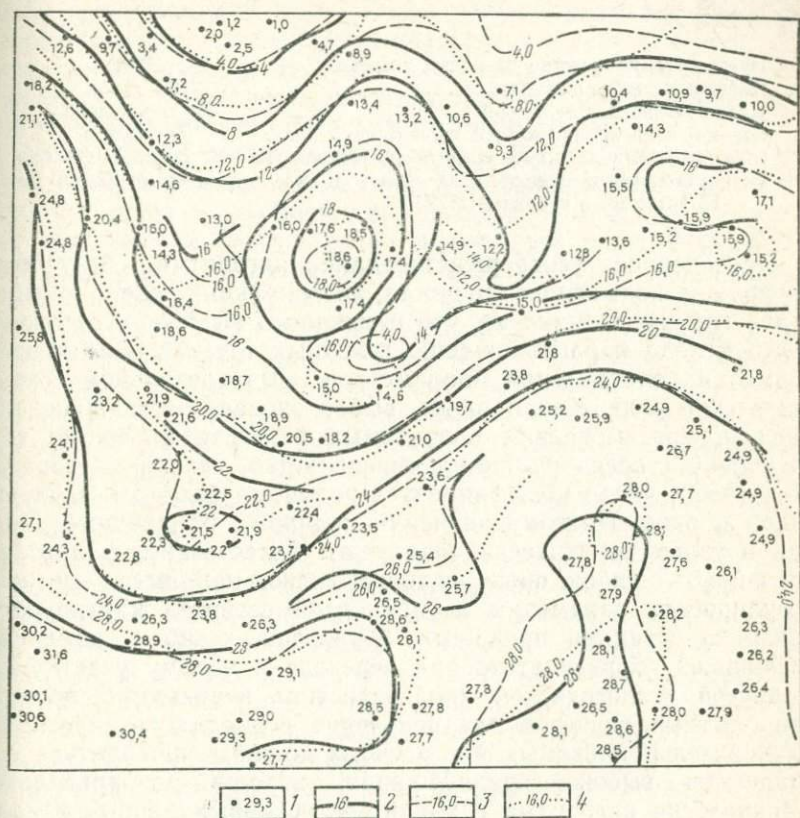


Рис. 1. Сопоставление результатов графической и машинной интерполяции: 1 — значения исходных данных (по разреженной сети); 2 — изолинии, проведенные вручную по неразрезанной сети пунктов наблюдений; 3 — изолинии, проведенные по значениям в узлах квадратной сети, рассчитанным на ЭВМ по разреженной сети исходных данных; 4 — изолинии, проведенные по разреженной сети пунктов опытным специалистом

ное перпендикулярно к профилям, с соотношением сторон, равным соотношению шагов вдоль и вкрест простирания аномалий. Этот вариант учитывает то обстоятельство, что профили наблюдений обычно закладываются вкрест ожидаемого простирания, так чтобы различие в сложности поведения признака компенсировалось изменением шага наблюдений.

В отличие от других программ построения карт имеется возможность приведения данных в единую систему наблюдений при расположении участка съемки на стыке двух шестиградус-

ных зон прямоугольной координатной сетки Гаусса—Крюгера.

Программа, реализующая рассмотренный алгоритм, составлена на языке Алгол-60 и транслирована на ЭВМ БЭСМ-4 с помощью транслятора ТА-2М. Основные характеристики программы:

Предельное количество исходных значений	9600
Ширина матрицы результатов	100 строк
Длина матрицы результатов	Не ограничена
Степень аппроксимирующего полинома	До шестой
Скорость счета для карты средней сложности	1000 точек/30—60 мин
Результаты счета выдаются на печать в виде полос сетки 22×22 по 12 значений в строке АЦПУ	

Проведенное усовершенствование алгоритма, безусловно, привело к повышению точности интерполяции, сделало метод более гибким. Однако то, что отмечалось выше о чувствительности метода к расположению исходных точек в окне и эмпиризме в определении параметров вычислительной схемы, остается в силе. Как и ранее, выбор значений R и m основан лишь на рассмотрении получаемых результатов, кроме того, поскольку степень полинома определяется, исходя из неравенства (1.62), число коэффициентов полинома близко к значению узлов в окне. Поэтому значения полинома и исходные значения в узлах практически совпадают. Естественно, фильтрация случайных ошибок очень слаба. Неясно, например, как надо регулировать параметры m и R , если свойства картируемого признака остаются прежними, а уровень ошибок измерений увеличился. В результате при переходе к новому участку или к другой геологической характеристике необходимо проведение опытных расчетов для получения подходящих параметров вычислительной схемы. Эти расчеты должны проводиться специалистом высокой квалификации, хорошо разбирающимся в технологии алгоритма и характере исходных данных и дальнейших задачах.

Подводя итоги, следует еще раз подчеркнуть, что метод наименьших квадратов обеспечивает минимум погрешностей лишь в самих пунктах наблюдения, а не во всей области приближения. Более того, вывод авторов работы [63] о том, что при использовании алгоритмов полиномиальной аппроксимации «квадрат разности между исходной функцией и аппроксимирующей сведен к минимуму», на самом деле верен лишь для конкретной степени полинома, заранее выбранных размеров окна и принятого принципа определения весов. Что касается практических приложений, то при сравнительно равномерной исходной сети и небольшом уровне помех эти алгоритмы позволяют производить интерполяцию с приемлемой точностью. Наиболее предпочтительным из них, по-видимому, является алгоритм, изложенный в работе [23].

Метод непересекающихся треугольников. Из общих соображений ясно, что аппроксимация в скользящем окне полиномом степени $m > 1$ практически равноценна линейной аппроксимации в окне меньших размеров*. Поскольку размеры каждого окна (треугольника) значительно меньше размеров окна при полиномиальных среднеквадратических приближениях, то алгоритм, основанный на линейной интерполяции внутри непересекающихся треугольников, заполняющих всю область приближения, может не уступать последним в точности.

Алгоритм, предложенный в работе [2], условно можно назвать эвристическим, так как здесь в определенной мере моделируется процесс графической интерполяции. В то же время при графической интерполяции выбор сети непересекающихся треугольников приводится обычно многократно — сеть меняется в зависимости от хода изолиний, кроме того, при этом учитывается и поведение изолиний в соседних треугольниках и проводится сглаживание влияния случайных ошибок измерений. Как получить «оптимальную» сеть треугольников с помощью ЭВМ, остается неясным. Ясно только, что результаты могут быть различными, а какой вариант сети лучше — формальным способом определить пока не удалось. Не решены и вопросы фильтрации случайных ошибок, не говоря уже о том, что линейная интерполяция на ЭВМ даже с так называемых инженерных позиций не может считаться сколько-нибудь близкой к оптимальной. Поэтому метод рекомендуется для получения материалов в наглядной графической форме при предварительном анализе. Авторы считают, что такие операции, как сглаживание, не относятся к этапу первоначальной обработки данных и должны применяться впоследствии. Обоснованием этого может служить то положение, что параметры вычислительных схем зависят от характера поля, расположения точек на участке и других факторов, которые до предварительной обработки изучить крайне трудно.

Интересно отметить, что сам алгоритм построения сети непересекающихся треугольников на ЭВМ далеко не прост, как это может показаться с первого взгляда. Здесь мы встречаемся с большими трудностями формализации процесса, который при построении карт вручную почти незаметен. Вот как выглядит, например, первый шаг [2]. Из множества точек $P = \{(x_i, y_i)\}$ выберем точку с координатами x_0, y_0 , о которой заранее известно, что она лежит внутри выпуклого многоугольника. Затем находятся точки, которые вместе с уже выбранной образуют треугольник минимального периметра и ненулевой площади. Для этого точка x_1, y_1 выбирается как ближайшая к x_0, y_0 , а поиск точки x_2, y_2 производится последовательным просмотром

* Для некоторых полиномов это показано в гл. II на основании рассмотрения их частотных характеристик.

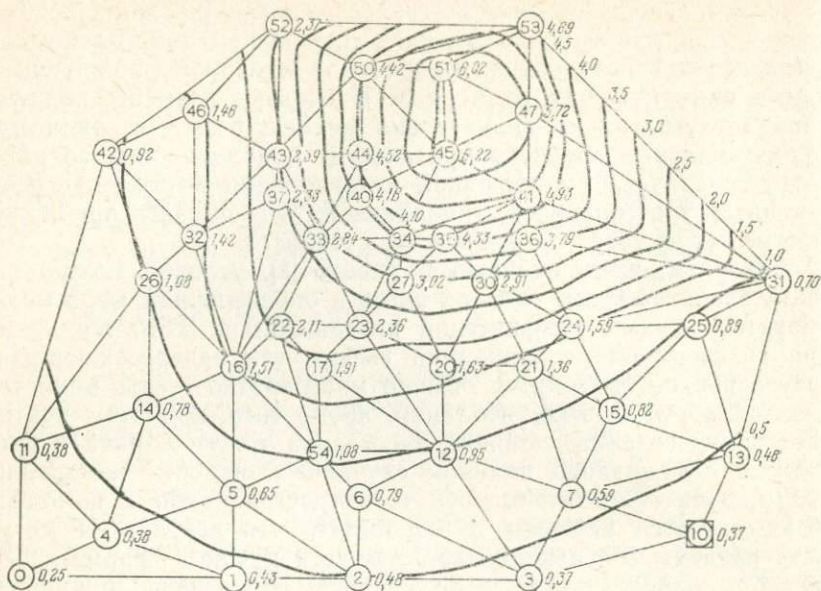


Рис. 2. Фрагмент построения карты при начальном пункте 10

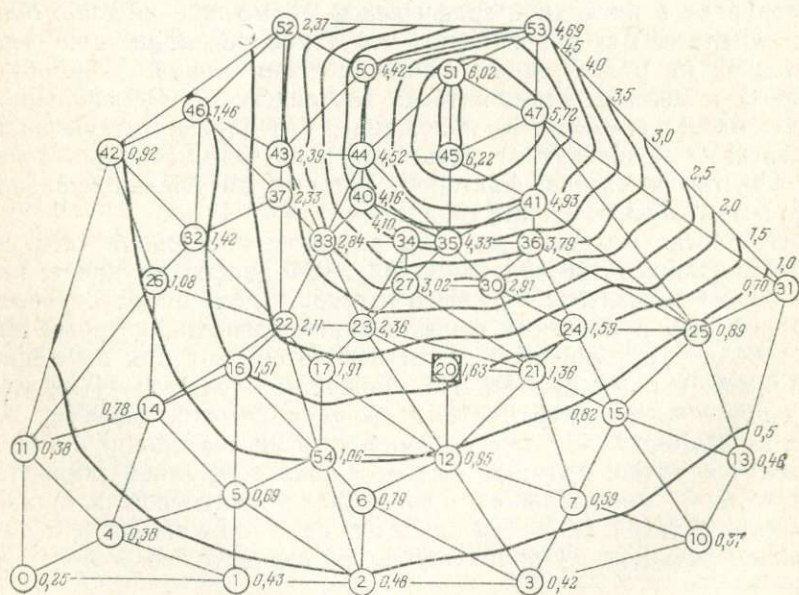


Рис. 3. Фрагмент построения карты при начальном пункте 20

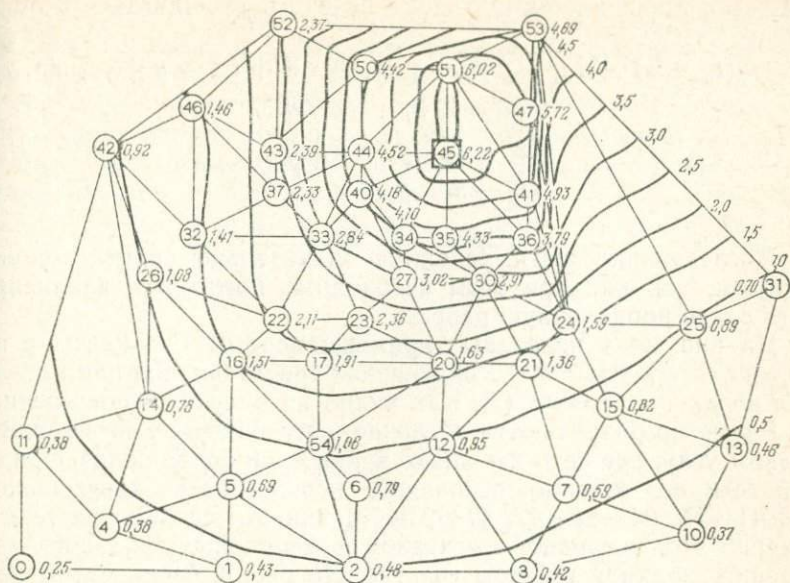


Рис. 4. Фрагмент построения карты при начальном пункте 45

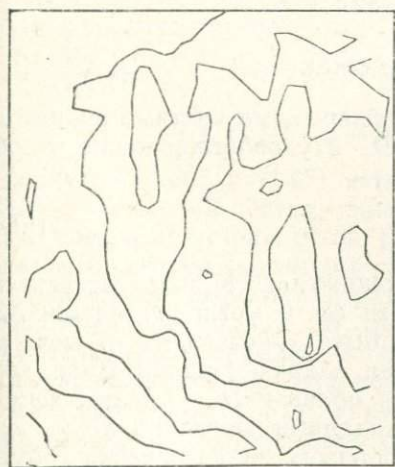


Рис. 5. Автоматически построенная карта изолиний (практический пример)

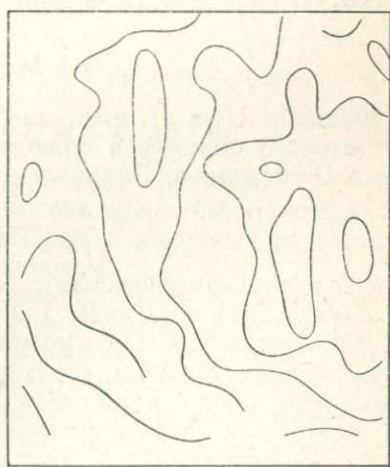


Рис. 6. Карта, построенная вручную по тем же исходным данным, что и карта на рис. 5

ром всех точек до тех пор, пока не будут выполняться условия

$$\sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2} + \sqrt{(x_1 - x_i)^2 + (y_1 - y_i)^2} = \min,$$

$$i = 1, 2, \dots, n.$$

$$\begin{vmatrix} x_1 - x_0 & y_1 - y_0 \\ x_0 - x_i & y_0 - x_i \end{vmatrix} \neq 0.$$

Последующие шаги делаются значительно более сложным образом, так как при этом необходимо помнить и анализировать ранее проделанную работу.

На рис. 2—4 показаны примеры построения карты с помощью алгоритма непересекающихся треугольников при различных начальных точках [2]. Как видно из рисунков, построенные сети заметно различаются. Отличие карт существенно в верхнем правом углу, где сеть исходных данных наиболее редка. При такой сети нет никаких оснований использовать треугольники 36—41—24, 24—41—47, 47—53—24, как это сделано на рис. 4. Именно это, очевидно, — основной фактор неоправданного проведения изолиний на этом участке. На рис. 5, 6 приведены примеры автоматического и графического построений карт по одному из участков Днепровско-Донецкой впадины. Карты построены по почти равномерной и густой сети точек, общее число которых около 200. Как отмечают авторы, различия имеются лишь в тех местах, где в исходной информации есть бракованные значения, которые при ручной обработке, естественно, исключались из рассмотрения.

§ 4. Метод сеток

Функции $U(x, y)$, имеющие непрерывные частные производные второго порядка в области D и удовлетворяющие в области D уравнению Лапласа

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0, \quad (I.64)$$

называются гармоническими функциями. К ним относятся, например, аномалии силы тяжести Δg и магнитного поля ΔZ .

Для таких функций может быть поставлена и решена единственным образом следующая задача (задача Дирихле): найти функцию, гармоническую в области D и принимающую на ее границе S непрерывные заданные значения, непрерывную в области $\bar{D} = D + S$. Таким образом, зная значения гармонической функции $U(x, y)$ на границе области S , можно определить значения этой функции во всей области D . При этом гармоническая функция в области D , где справедливо равенство (I.64), не может принимать значений больших, чем мак-

симум ее значений на границе области S , и меньших, чем ее минимум на границе S . Это так называемая теорема о максимуме и минимуме гармонической функции.

Решение задачи Дирихле для ограниченной области D (внутренняя задача Дирихле) может быть проведено на основе замены дифференциального уравнения Лапласа (I.64) уравнением Лапласа в конечных разностях [20, 22, 42]. Поскольку

$$\frac{\partial^2 U}{\partial x^2} \approx \frac{U(x+h, y) - 2U(x, y) + U(x-h, y)}{h^2}, \quad (I.65)$$

$$\frac{\partial^2 U}{\partial y^2} \approx \frac{U(x, y+h) - 2U(x, y) + U(x, y-h)}{h^2},$$

вместо (I.64) получим

$$\begin{aligned} & \frac{U(x+h, y) - 2U(x, y) + U(x-h, y)}{h^2} + \\ & + \frac{U(x, y+h) - 2U(x, y) + U(x, y-h)}{h^2} = 0, \end{aligned}$$

где шаг $h > 0$.

Отсюда

$$U(x, y) = \frac{1}{4} [U(x+h, y) + U(x-h, y) + U(x, y+h) + U(x, y-h)]. \quad (I.66)$$

Аналогичным образом можно получить формулу, использующую несколько другие узлы:

$$U(x, y) = \frac{1}{4} [U(x+h, y-h) + U(x-h, y-h) + U(x-h, y+h) + U(x+h, y+h)]. \quad (I.67)$$

Уравнения (I.66) и (I.67) есть уравнения Лапласа в конечных разностях. Теперь поставленная задача может быть легко решена с помощью метода сеток (или метода конечных разностей), идея которого заключается в следующем.

Область D покрывается равномерной сетью узлов, аппроксимирующей эту область. В граничных узлах, наиболее близко прилегающих к границе S , на основании граничных условий устанавливаются значения искомого решения. Используя уравнение (I.66) или (I.67), получим систему конечно-разностных уравнений, т. е. фактически систему линейных алгебраических уравнений. Порядок системы зависит от размеров области D и величины h : чем меньше значение h , тем точнее результаты. Однако порядок системы уравнений обратно пропорционален квадрату величины h и это быстро приводит к нежелательности излишнего дробления даже при использовании ЭВМ. Итак,

задавая в граничных точках значения $\varphi_h(x_i, y_i) = \varphi_S(x_i, y_i)$, где $\varphi_S(x_i, y_i)$ — значение функции φ в точке границы S , ближайшей к соответствующему граничному узлу, вместо уравнения (I.64) решаем систему конечно-разностных уравнений:

$$\varphi_{i,j} = \frac{1}{4} (\varphi_{i-1,j} + \varphi_{i+1,j} + \varphi_{i,j-1} + \varphi_{i,j+1}). \quad (\text{I.68})$$

В результате получаем значения $\varphi_{i,j}$ во всей области D , которые можно считать результатом интерполяции значений $\varphi_S(x, y)$, заданных на границе S . Система (I.68) обычно решается итерационным способом по методу Либмана. Значения $\bar{\varphi}_{i,j}^{(k)}$ для внутренних узлов сетки D_h определяются по формуле

$$\bar{\varphi}_{i,j}^{(k)} = \frac{1}{4} [\bar{\varphi}_{i-1,j}^{(k-1)} + \bar{\varphi}_{i+1,j}^{(k-1)} + \bar{\varphi}_{i,j-1}^{(k-1)} + \bar{\varphi}_{i,j+1}^{(k-1)}] \quad (\text{I.69})$$

$$(k = 1, 2, 3, \dots).$$

За начальное приближение $\bar{\varphi}$ можно взять любую систему чисел, однако, учитывая теорему о максимуме и минимуме гармонических функций, целесообразно ограничить их значения неравенством

$$\min_S \varphi(x, y) \leq \bar{\varphi}^{(0)} \leq \max_S \varphi(x, y). \quad (\text{I.70})$$

Во многих случаях может оказаться, что простой перенос в граничные узлы регулярной сетки ближайших значений наблюдения $\varphi(x_i, y_i)$ определяет значения в них слишком грубо. Их можно подправить путем линейной интерполяции в каждом цикле интерполяционного процесса.

$$\bar{\varphi}^{(k)}(A_h) = \varphi(A) + \frac{\bar{\varphi}^{(k-1)}(B) - \varphi(A)}{h + \delta} \delta \quad (\text{I.71})$$

$$(k = 1, 2, 3, \dots),$$

где A — ближайшая к A_h точка наблюдений, B — ближайший к A_h внутренний узел сетки D_h , δ — удаление узла A_h от точки A (рис. 7).

Такой подход к решению задачи интерполяции геологических признаков в 1965 г. предложил использовать А. Р. Зайдель, с тем чтобы использовать хорошо разработанный вычислительный аппарат решения внутренней задачи Дирихле. Ясно, что в данном случае нельзя использовать только значения $\varphi_S(x, y)$, заданные по границе области, так как при этом

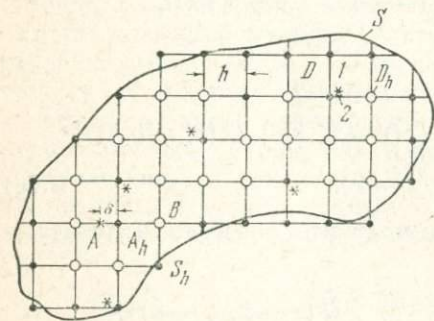


Рис. 7. Метод сеток:

1 — S_h — граничные точки сетки; 2 — пункты наблюдений внутри области D

теряется большая часть информации, заданной внутри ее. Не говоря даже об устойчивости и точности решения методом сеток, это следует уже из того, что на основании теоремы о максимуме и минимуме гармонических функций в области D не может быть наибольших или наименьших значений, т. е. экстремумы могут находиться лишь на границе S . Это, конечно, совершенно неприемлемое положение. Поэтому было предложено считать граничными и внутренние точки области D_h , ближайшие к точкам наблюдений или совпадающие с ними. При решении системы уравнений итерационным способом значения в этих точках, как и в точках, действительно расположенных на границе, не изменяются. Теорема о максимуме и минимуме теперь действует в том смысле, что все экстремальные значения функции $\bar{\varphi}(x, y)$ могут быть приурочены только к пунктам наблюдений $\varphi(x_i, y_i)$. Кроме того, это означает, что значения функции $\bar{\varphi}(x, y)$ в промежутках между узлами исходных значений $\varphi(x_i, y_i)$ всегда меньше значений в этих узлах, т. е. в последних всегда должны быть экстремумы. Это обычно хорошо заметно на составляемых картах, если они не подвергнуты сглаживанию. Впрочем, в некоторой степени сглаживание всегда достигается уже тем, что получаемые экстремумы не всегда превышают величину сечения изолиний.

В работах [19, 68] в целях интерполяции используется и решение уравнения вида

$$\frac{\partial}{\partial x} \left(K \frac{\partial U}{\partial x} \right) + \frac{\partial}{\partial y} \left(K \frac{\partial U}{\partial y} \right) = 0, \quad (1.72)$$

где K также является функцией координат x, y и удовлетворяет уравнению Лапласа $\Delta K = 0$.

Рассмотрим примеры из работы [68], в которой проведен анализ влияния величины h и количества итераций на качество построенной структурной карты. Участок характеризуется спокойным поведением изолиний, число исходных точек (24) достаточно для их уверенного проведения (рис. 8). На рис. 9 представлен вариант карты, рассчитанный когда число узлов квадратной сетки N составляло 13431 (!), т. е. примерно в 500 раз больше исходного. В этом случае можно считать, что граничные значения $\bar{\varphi}_h$ очень хорошо совпадают с ближайшими значениями наблюдений. Однако итерационный процесс оказался неустойчивым, а полученная карта совсем не похожа на действительную. Почти каждому значению $\varphi(x_i, y_i)$ соответствует локальный экстремум или соответствующий изгиб изолиний. На рис. 10 изображена карта, составленная при $N=527$, напоминающая рис. 8. Промежуточные варианты $N=2133$ и $N=1492$ практически непригодны. Таким образом, эти примеры показывают, что выбор шага h очень существенно влияет на получаемые результаты. Несмотря на это, никаких рекоменда-

даций о выборе значения h в работах [19, 68] не содержится. Кроме того, расчеты при решении уравнения (1.72) требуют задания значений K , от которых, естественно, зависимость по-

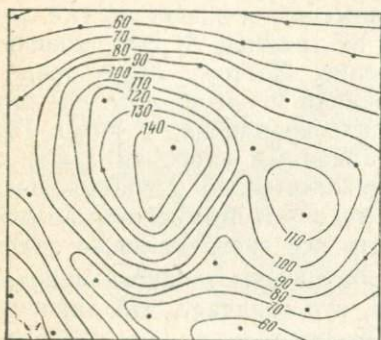


Рис. 8. Структурная карта, составленная графическим способом

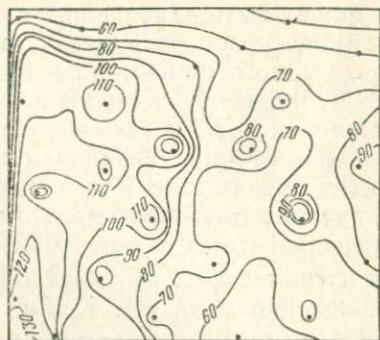


Рис. 9. Вариант структурной карты, рассчитанной при $N=13431$

лучаемых результатов также очень сильна. Для того чтобы уменьшить влияние значений $\bar{\varphi}$, расположенных по границе сеточной области S_h , которые получают-ся предварительным «ручным» интерполированием,

в работе [68] предлагается для точек, находящихся внутри контура S_h , полагать $K=1$, а на контуре $0 < K(x, y) < 1$. Определить эту величину, очевидно, можно в результате получения большого опыта работы по анализу используемого алгоритма. В работе [19] по этому поводу отмечено: «Конкретных рекомендаций для выбора коэффициентов неоднородности или анизотропии при построении числовой модели того или иного геологического параметра дать в настоящем описании невозможно. В каждом случае исследователь

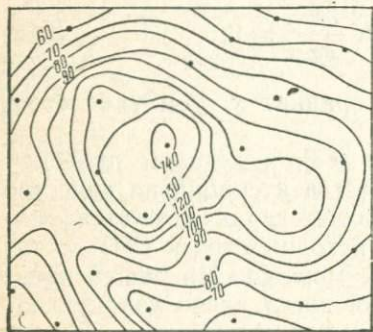


Рис. 10. Вариант структурной карты, рассчитанной при $N=527$

должен исходить из специфики решаемой задачи». Несомненно, это говорит о существенных недостатках в реализации алгоритма. Чем меньше величина h , тем точнее происходит перенос исходных значений в узлы квадратной сетки, но тем менее устойчивой становится решаемая система уравнений. Даже если считать возможным перенос на $0,2h$, т. е. с учетом переноса по x и y , до $0,3h$, количество узлов сетки D_h увеличивается в 25 раз по сравнению с исходным. Существенным

недостатком является отсутствие фильтрации случайных ошибок, без чего интерполяция большинства геологических признаков может привести к существенным искажениям. Не может спасти положение и последующая фильтрация (например, с помощью осреднения) по уже полученной регулярной сети данных, хотя полученные карты будут иметь вполне удовлетворительный вид. Дело в том, что в случае неравномерной сети наблюдений случайные ошибки могут при интерполяции проявить себя весьма сложным образом, значительно, иногда во много раз увеличивая радиус своего влияния. Радиус автокорреляции этих помех может стать близким к аналогичной характеристике полезной составляющей, и эффективность осреднения резко падает.

Возможно, некоторые из этих недостатков можно было бы устранить, используя для решения задачи Дирихле более совершенные алгоритмы, однако, учитывая наличие принципиального недостатка, связанного с сильным влиянием теоремы о максимуме и минимуме, это вряд ли целесообразно. Следует отметить, что если функция φ будет гармонической, этот недостаток проявится в еще более четкой форме. Здесь обнаруживается то, что на самом деле решение внутренней задачи Дирихле для уравнения Лапласа рассчитано на существенно иной класс задач, в результате чего вопросы о принадлежности функций φ к классу гармонических функций приобретают второстепенный характер.

Ниже дается характеристика программ на ЭВМ БЭСМ-4, АУЛ-2, основанных на алгоритмах, разработанных и используемых в работе [19].

Вместо формулы (I.69) используется несколько другое выражение:

$$\bar{\varphi}_{i,j}^{(k)} = \bar{\varphi}_{i,j}^{(k-1)} + \frac{1}{N} [\bar{\varphi}_{i-1,j}^{(k)} + \bar{\varphi}_{i+1,j}^{(k-1)} + \bar{\varphi}_{i,j-1}^{(k)} + \bar{\varphi}_{i,j+1}^{(k-1)} - N\bar{\varphi}_{i,j}^{(k-1)}], \quad (I.73)$$

где

$$N = p_{i-1,j} + p_{i+1,j} + p_{i,j-1} + p_{i,j+1},$$

$$p_{i\pm 1,0; j\pm 1,0} = \begin{cases} 1 & \varphi_{i\pm 1,0; j\pm 1,0} \neq 0 \\ 0 & \varphi_{i\pm 1,0; j\pm 1,0} = 0. \end{cases}$$

Имеется возможность построения сетки узлов с разным шагом по осям x и y , $\Delta x \neq \Delta y$. В этом случае

$$\bar{\varphi}_{i,j}^{(k)} = \bar{\varphi}_{i,j}^{(k-1)} + \frac{1}{N} [\bar{\varphi}_{i-1,j}^{(k-1)} + \bar{\varphi}_{i+1,j}^{(k)} + \alpha (\bar{\varphi}_{i,j+1}^{(k-1)} + \bar{\varphi}_{i,j-1}^{(k)}) - N \left(\frac{\alpha + 1}{2} \right) \bar{\varphi}_{i,j}^{(k-1)}], \quad (I.74)$$

где

$$\alpha = \left(\frac{\Delta x}{\Delta y} \right)^2.$$

При решении уравнения (I.72) конечно-разностная формула имеет вид

$$\begin{aligned} \bar{\Phi}_{i,j}^{(k)} = & \bar{\Phi}_{i,j}^{(k-1)} + \frac{1}{N} [(K_{i+1,j} + K_{i,j}) \bar{\Phi}_{i+1,j}^{(k-1)} + (K_{i,j+1} + K_{i,j}) \bar{\Phi}_{i,j+1}^{(k-1)} + \\ & + (K_{i-1,j} + K_{i,j}) \bar{\Phi}_{i-1,j}^{(k-1)} + (K_{i,j-1} + K_{i,j}) \bar{\Phi}_{i,j-1}^{(k-1)} - (K_{i+1,j} + \\ & + K_{i,j+1} + K_{i-1,j} + K_{i,j-1} + NK_{i,j}) \bar{\Phi}_{i,j}^{(k-1)}], \end{aligned} \quad (I.75)$$

где

$$N = \sum_1^4 p_{i \pm 1, 0; j \pm 1, 0}.$$

Итерационный процесс продолжается до выполнения условия

$$\bar{\Phi}_{i+1,j}^{(k-1)} + \bar{\Phi}_{i,j+1}^{(k-1)} + \bar{\Phi}_{i-1,j}^{(k)} + \bar{\Phi}_{i,j-1}^{(k)} - N \bar{\Phi}_{i,j}^{(k)} \leq \varepsilon. \quad (I.76)$$

Следует отметить, что величина ε служит лишь для контроля итерационного процесса и не может быть непосредственно использована для оценки точности интерполяции.

Область интерполяции может быть многосвязной, т. е. внутри внешней границы могут быть участки, где наблюдения отсутствуют и интерполяция производиться не должна. Исходными данными для программы служат значения $\varphi(x_i, y_i)$ внутри области и на ее границе, информация о контурах интерполируемой области, которые представляют собой замкнутую ломаную линию, шаги Δx и Δy , величина ε , масштаб и сечение изолиний карты результатов. Предусмотрена возможность выдачи числовой модели карты на алфавитно-цифровое печатающее устройство (АЦПУ). Расчеты могут производиться одновременно для 20 признаков.

Программа АУЛ-2 состоит из двух самостоятельных блоков — АУЛ-2П и АУЛ-2С. Первый из них осуществляет подготовительные операции: контролирует данные, формирует интерполируемую область, засылает значения наблюдений в соответствующие узлы сетки, записывает исходные данные на магнитную ленту. Второй осуществляет расчеты по формулам (I.73) или (I.75) и записывает результаты на магнитную ленту, а при необходимости выдает их на АЦПУ. Время счета по программе АУЛ-2С примерно равно

$$t_{\text{мин}} = 0,006pq, \quad (I.77)$$

а для программы АУЛ-2П

$$t_{\text{мин}} = 0,0002pq + 0,006nr, \quad (I.78)$$

где n — число точек наблюдения, r — число параметров,

$$q = \frac{y_{\max} - y_{\min}}{\Delta y},$$

$$p = \frac{x_{\max} - x_{\min}}{\Delta x},$$

x_{\max} , x_{\min} , y_{\max} , y_{\min} — максимальное и минимальное значения координат на контуре — по осям абсцисс и ординат, т. е. pq — это фактически число узлов прямоугольной сетки. Оно всегда в десятки раз больше числа проведенных измерений, поэтому время, определяемое по формулам (I.77) и (I.78), сравнительно велико.

§ 5. Приближение гармоническими функциями

В § 2 было дано определение оптимального и асимптотически оптимального восстановления функций φ как такого, при котором восстанавливается целая функция с ограниченным спектром $\sigma = \frac{\pi}{\Delta x}$, значения которой в узлах наблюдений совпадают с исходными, т. е. $\varphi(x_i) = \bar{\varphi}(x_i)$ при $i=1, 2, 3, \dots, \infty$ и $i=1, 2, 3, \dots, n$. Это может быть сделано с помощью ряда Котельникова, если $\Delta x = \text{const}$. Однако получение коэффициентов соответствующих рядов при неравномерной сети измерений слишком сложно [95]. Кроме того, подобные ряды сходятся медленно, что сказывается не только в необходимости удерживать большое число членов ряда, но и требует наличия достаточного числа наблюдений. Эти соображения, а также отсутствие разработанного алгоритма фильтрации не позволяют пока использовать их в практических расчетах. Далее будет рассмотрен другой подход к решению этой задачи, позволяющий получать асимптотически оптимальные решения. Он основан на предложенном автором в 1963 г. [6] методе решения некоторых задач гравиметрии и теории фигуры Земли, развитом затем в ряде работ для решения задач интерполяции и фильтрации случайных ошибок [7, 8, 11, 16]. Наиболее полно метод рассмотрен в работе [16].

Метод предполагает аппроксимацию значений $\varphi(x)$ функцией вида

$$\bar{\varphi}(x, z) = \sum_k \alpha_k G_k(x - k\Delta x, z_1), \quad (\text{I.79})$$

где G_k — гармонические функции,

$$G_k = \frac{1}{\pi} \frac{z_1}{(x - k\Delta x)^2 + z_1^2} \quad (\text{I.80})$$

или

$$G_k = \frac{1}{\pi} \int_{\Delta x \left(k - \frac{1}{2}\right)}^{\Delta x \left(k + \frac{1}{2}\right)} \frac{z_1 d\xi}{(x - \xi)^2 + z_1^2} = \frac{1}{\pi} \operatorname{arctg} \frac{(x - \xi)}{z_1} \Big|_{\Delta x \left(k - \frac{1}{2}\right)}^{\Delta x \left(k + \frac{1}{2}\right)}, \quad (I.81)$$

$$k = 0, \pm 1, \pm 2, \dots$$

α_k — некоторые коэффициенты, определяемые таким образом, чтобы выполнялось неравенство

$$\max_{\{x_i\}} |\varphi(x_i) - \bar{\varphi}(x_i)| \leq \varepsilon_0, \quad \{(x_i)\} = P, \quad (I.82)$$

где ε_0 — достаточно малая величина.

Если элементы α_k определены, дальнейшие расчеты, связанные с различными линейными преобразованиями исходных данных (в том числе и интерполяции), не представляют никаких затруднений. В общем случае для трансформации, описываемой уравнением

$$V = L\varphi, \quad (I.83)$$

где L — линейный ограниченный оператор, имеем

$$V \approx \bar{V} = \sum_k b_k \alpha_k, \quad (I.84)$$

где

$$b_k = LG_k. \quad (I.85)$$

При решении задачи интерполяции на регулярную сеть оператор L равен 1. При этом используются функции $G_k(x - k\Delta x)$ из (I.79), где значениями x в этом случае являются координаты узлов регулярной сетки.

Легко видеть, что точность этих расчетов полностью определяется близостью функций φ и $\bar{\varphi}$, так как для последней из них вычисление трансформант производится точно. Имеется в виду, конечно, близость не в точках наблюдений, как в формуле (I.82), а на всем участке приближения, т. е.

$$\max_{x_{\min} < x < x_{\max}} |\varphi(x) - \bar{\varphi}(x)| \leq \delta. \quad (I.86)$$

Если наблюдаемый признак — аномалии силы тяжести, метод получает простую физическую трактовку. Конструкции (I.79), (I.81) соответствуют аппроксимации возмущающих масс простым слоем с кусочно-постоянной (на отрезках длиной Δx) плотностью, распределенным на внутренней плоскости S' , отстоящей от плоскости наблюдений на расстоянии z_1 . Таким образом, построение конструкции (I.79) сводится к определе-

нию таких значений α_h , чтобы гравитационный эффект подби-
раемого простого слоя с заданной точностью ϵ_0 совпадал с из-
меренными значениями на множестве пунктов наблюдения
(см. формулу (I.82)). Эти значения плотности простого слоя
(α_h) затем легко использовать для вычисления различных
линейных трансформаций, решая прямую задачу для простого
слоя, т. е. фактически по формулам (I.83)—(I.85).

Для функций, не являющихся гармоническими, такой физи-
ческой трактовки, к сожалению, дать нельзя. Впрочем, по ана-
логии с изложением аппроксимации алгебраическими полино-
мами, можно ограничиться следующим: производится аппрок-
симация системой функций определенного вида (G_h) с коэф-
фициентами α_h , такая, что выполняется неравенство (I.82).
Таким образом на множестве пунктов измерений, в отличие
от среднеквадратического приближения полиномами дости-
гается равномерная аппроксимация, в частности, величина ϵ_0
в (I.82) может быть равна нулю.

Следует отметить, что между методом сеток, изложенным
в предыдущем параграфе, в котором функции φ считаются гар-
моническими, и рассматриваемым методом есть принципиаль-
ное отличие. Если в методе сеток принадлежность интерполи-
руемых функций к классу гармонических является существен-
ной, то в данном случае это вовсе необязательно, хотя аппрок-
симирующая функция $\bar{\varphi}$ при использовании функций G_h из
(I.80) и (I.81) является гармонической. Главное заключается
в том, что в методе сеток решается граничная задача, в резуль-
тате чего в полной мере сказывается влияние теоремы о мак-
симуме и минимуме. В данном случае при решении задачи
интерполяции такого рода ограничений нет, наоборот, метод
в определенной степени позволяет восстанавливать амплитуду
и положение экстремумов, если они находятся между пунктами
наблюдений¹.

В работах [8, 11, 16] обычно рассматривались функции G_h
вида (I.81). Это, однако, связано лишь со спецификой обра-
ботки аномалий силы тяжести в горных районах. Принципиаль-
ная эквивалентность функций вида (I.80) и (I.81) с точки зре-
ния точности получаемых результатов следует из ряда сооб-
ражений, на которых здесь не имеет смысла останавливаться.
Подробное теоретическое доказательство этого приведено в [16].
Более того, можно показать, что аналогичные результаты могут
быть получены и для некоторых других видов функций G_h ,
не обязательно являющихся гармоническими.

* При решении некоторых других задач, например при пересчете анома-
лий силы тяжести в верхнее полупространство, теорема о максимуме и мини-
муме и здесь сказывается в полной мере, однако это полностью отвечает
сути этих задач.

Таким образом, основную роль здесь играет принцип построения аппроксимирующей конструкции, применение же различных, но обладающих необходимыми свойствами функций G_k приводит фактически к одинаковым результатам (значения z_1 при этом различаются) интерполяции. В целях экономии машинного времени для решения задач целесообразно использовать G_k из (I.80).

При практической реализации метода значения Δx принимаются примерно равными расстоянию между исходными пунктами, а значения a_k определяются из системы линейных уравнений

$$F = AX, \quad (\text{I.87})$$

где $x = [a_k]$, $F = [\varphi(x_i)]$ или $F = [\varphi(x_i, y_i)]$ — свободные члены уравнения (I.87) — значения признака в пунктах наблюдений ($i=0, 1, 2, \dots, n$), $A = [a_{i,k}]$ — матрица, элементы которой определяются по формулам

$$a_{i,k} = \frac{1}{\pi} \frac{z_1}{(x_i - \xi_k)^2 + z_1^2} \quad (\text{I.88})$$

или

$$a_{i,k} = \frac{1}{\pi} \operatorname{arctg} \frac{x_i - \xi_k}{z_1} \left| \begin{array}{c} \xi_k + \frac{\Delta x}{2} \\ \xi_k - \frac{\Delta x}{2} \end{array} \right. , \quad (\text{I.89})$$

$\xi_k = (0, \pm 1, \pm 2, \dots, \pm n) \Delta x$, для функций G_k вида (I.80) и (I.81) соответственно в одномерном случае и

$$a_{i,k} = \frac{1}{2\pi} \frac{z_1}{[(x_i - \xi_k)^2 + (y_i - \eta_k)^2 + z_1^2]^{3/2}}, \quad (\text{I.90})$$

$$a_{i,k} = \frac{1}{2\pi} \operatorname{arctg} \frac{(x_i - \xi_k)(y_i - \eta_k)}{z_1 \sqrt{(x_i - \xi_k)^2 + (y_i - \eta_k)^2 + z_1^2}} \left| \begin{array}{c} \xi_k + \frac{\Delta x}{2} \quad \eta_k + \frac{\Delta x}{2} \\ \xi_k - \frac{\Delta x}{2} \quad \eta_k - \frac{\Delta x}{2} \end{array} \right. \quad (\text{I.91})$$

в двухмерном.

Поскольку функция $\bar{\varphi}$ является гармонической, из условия $\sum a_k G_k(x - k\Delta x, z) = 0$ следует, что $a_k \equiv 0$, т. е. совокупность векторов G_k ($k=0, \pm 1, \pm 2, \dots$) линейно независима и, следовательно, система уравнений (I.87) всегда имеет решение.

Уравнение (I.87) удобно решать по методу последовательных приближений:

$$X^{(k+1)} = (E - A) X^{(k)} + F, \quad (\text{I.92})$$

где

$$X^{(k+1)} = X^{(k)} + \delta^{(k)}, \quad (\text{I.93})$$

$$\delta^{(k)} = F - AX^{(k)}, \quad (I.94)$$

E — единичная матрица.

Сходимость последовательных приближений легко доказывается теоретически (доказательство будет приведено в гл. II при рассмотрении алгоритма фильтрации случайных ошибок), а также с помощью численного вычисления наибольших собственных значений λ_1 матрицы $(E-A)$ для различных соотношений $z_1/\Delta x$ [16]. Из графика наибольших собственных значений матрицы $(E-A)$ видно, что при $z_1/\Delta x > 1$ значения λ_1 быстро приближаются к 1, поэтому при $z_1/\Delta x > 1$ скорость сходимости заметно снижается. Это, однако, не имеет значения, поскольку, как это будет показано ниже, всегда можно ограничиться значениями $z_1/\Delta x \approx 1$.

Уравнение (I.87) может быть решено и точными, не итерационными способами решения системы линейных уравнений, как, например, это сделано в работе [82], где рассмотрена несколько иная реализация метода. Это, однако, нецелесообразно, потому что приходится ограничиваться решением систем сравнительно небольшого порядка, к тому же плохо обусловленных, для того, чтобы машинное время не превышало разумных пределов. Решение систем уравнений при этом необходимо получать в каждой расчетной точке. Принципиальным преимуществом итерационного способа в данном случае является организация одновременной эффективной фильтрации случайных ошибок. Действительно, как и при решении задачи интерполяции другими методами, нет смысла полагать значение ϵ_0 в формуле (I.82) равным нулю или вообще делать значение ϵ_0 меньше уровня случайных ошибок, поэтому вместо метода простой итерации (I.92) — (I.94) на самом деле реализуется следующий процесс:

$$\alpha_i^{(k+1)} = \begin{cases} \alpha_i^{(k)} + \delta_i^{(k)}, & |\delta_i^{(k)}| > \epsilon_0 \\ \alpha_i^{(k)}, & |\delta_i^{(k)}| < \epsilon_0, \end{cases} \quad (I.95)$$

где ϵ_0 — некоторая константа, равная, например, величине среднеквадратической ошибки исходных данных (с учетом величины непредставительных характеристик интерполируемых признаков) или несколько большей этой величины, если требуется большая гладкость функции φ . Подробно этот алгоритм будет рассмотрен в гл. II. Отметим еще лишь то, что его использование не увеличивает, а, наоборот, сокращает время решения системы (I.87).

Исследование точности метода. В § 2 ошибка интерполяции δ в (I.25) была разложена на составляющие δ_1 и δ_2 , причем для первой из них на классе функций $U'(H)$ были получены соответствующие оценки. Получим теперь оценки второй

составляющей — δ_2 в (I.37) для рассматриваемого здесь метода [8, 16]. Оценки, как и для δ_1 , будем проводить для класса функций $U'(H)$ (см. § 2), в частности, на примере экстремальной функции этого класса $U(x, H)$.

Рассмотрим интерполирующую конструкцию

$$\bar{\varphi}(x) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \alpha_k \frac{z_1}{(x - k\Delta x)^2 + z_1^2}, \quad (\text{I.96})$$

где вид G_k определяется формулой (I.80), $k=0, \pm 1, \pm 2, \dots$

При получении оценок рассматривается случай, когда $\Delta x \leq 2H$, $\Delta x \leq 2z_1$. Это позволяет упростить вычисления, не ограничивая существенно их общности, так как использование сетей при $\Delta x > 2H$ заведомо приводит к слишком большим ошибкам. То же самое, как это будет показано ниже, имеет место и для соотношений Δx и z_1 . Спектр конструкции (I.96) равен

$$\bar{F}(\omega) = \frac{1}{\sqrt{2\pi}} e^{-|\omega|z_1} \sum_{k=-\infty}^{\infty} \alpha_k e^{-i\omega k\Delta x}. \quad (\text{I.97})$$

Учитывая, что $\bar{\varphi}(k\Delta x) = \tilde{\varphi}(k\Delta x)$, на основании (I.39) получим

$$\begin{aligned} \tilde{F}(\omega) &= \begin{cases} \sum_{-\infty}^{\infty} e^{-|\omega+2n\sigma|z_1} \sum_{-\infty}^{\infty} \alpha_k e^{-i(\omega+2n\sigma)k\Delta x} = \\ 0 \end{cases} = \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} e^{-|\omega+2n\sigma|z_1} \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k\Delta x} = \\ 0 \end{cases} = \\ &= \begin{cases} \sqrt{\frac{\pi}{2}} \sum_{-\infty}^{\infty} e^{-|\omega+2n\sigma|H}, & |\omega| \leq \sigma, \\ 0, & |\omega| > \sigma \end{cases}. \end{aligned} \quad (\text{I.98})$$

Отсюда

$$\frac{1}{\pi} \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k\Delta x} = \frac{\sum_{-\infty}^{\infty} e^{-|\omega+2n\sigma|H}}{\sum_{-\infty}^{\infty} e^{-|\omega+2n\sigma|z_1}}, \quad |\omega| \leq \sigma. \quad (\text{I.99})$$

Левая часть последней формулы $\frac{2\pi}{\Delta x}$ — периодическая функция*, поэтому

* Отметим, что ввиду этого равенство (I.99) справедливо и при $|\omega| > \sigma$. Таким образом, значения α_k являются коэффициентами ряда Фурье функции в правой части (I.99).

$$\frac{1}{\pi} \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k \Delta x} \leq \sup_{|\omega| \leq \sigma} \frac{\sum_{-\infty}^{\infty} e^{-|\omega + 2n\sigma| H}}{\sum_{-\infty}^{\infty} e^{-|\omega + 2n\sigma| z_1}}. \quad (I.100)$$

В то же время при $\Delta x < 2H$, $\Delta x < 2z_1$,

$$\sup_{|\omega| \leq \sigma} \frac{\sum_{-\infty}^{\infty} e^{-|\omega + 2n\sigma| H}}{\sum_{-\infty}^{\infty} e^{-|\omega + 2n\sigma| z_1}} =$$

$$= \sup_{|\omega| \leq \sigma} \frac{e^{-|\omega| H} \left[1 + \frac{e^{-2\sigma H}}{1 - e^{-2\sigma H}} (e^{2|\omega| H} + 1) \right]}{e^{-|\omega| z_1} \left[1 + \frac{e^{-2\sigma z_1}}{1 - e^{-2\sigma z_1}} (e^{2|\omega| z_1} + 1) \right]} =$$

$$= \left\{ \begin{array}{l} \frac{1 + \frac{2e^{-2\sigma H}}{1 - e^{-2\sigma H}}}{1 + \frac{2e^{-2\sigma z_1}}{1 - e^{-2\sigma z_1}}} \\ e^{-\sigma H} \left[1 + \frac{1 + e^{-2\sigma H}}{1 - e^{-2\sigma H}} \right] \\ e^{-\sigma z_1} \left[1 + \frac{1 + e^{-2\sigma z_1}}{1 - e^{-2\sigma z_1}} \right] \end{array} \right\} \leq \left\{ \begin{array}{l} 1, \quad z_1 < H \\ 1,02 \frac{e^{-\sigma H}}{e^{-\sigma z_1}} \approx \frac{e^{-\sigma H}}{e^{-\sigma z_1}}, \quad z_1 > H. \end{array} \right.$$

(I.101)

На основании соотношений (I.96)–(I.101) и формулы (I.38) находим окончательный результат:

$$\begin{aligned} \|\delta_2\|_C &\leq \frac{2}{\sqrt{2\pi}} \int_{\sigma}^{\infty} |\bar{F}(\omega)| d\omega + \frac{2}{\sqrt{2\pi}} \int_0^{\sigma} |\bar{F}(\omega) - \tilde{F}(\omega)| d\omega = \\ &= \frac{1}{\pi} \int_{\sigma}^{\infty} \left| e^{-|\omega| z_1} \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k \Delta x} \right| d\omega + \int_0^{\sigma} \left| e^{-|\omega| z_1} - \right. \\ &\quad \left. - \frac{1}{\pi} \sum_{-\infty}^{\infty} e^{-|\omega + 2n\sigma| z_1} \sum_{-\infty}^{\infty} \alpha_k e^{-i\omega k \Delta x} \right| d\omega \leq \end{aligned}$$

$$\begin{aligned} & \left\langle \frac{e^{-\sigma z_1}}{z_1} + \frac{e^{-2\sigma z_1}}{1 - e^{-2\sigma z_1}} \int_0^\sigma (e^{|\omega| z_1} + e^{-|\omega| z_1}) d\omega \right. \\ & \left. \leq \frac{e^{-\sigma H}}{e^{-\sigma z_1}} \int_\sigma^\infty e^{-|\omega| z_1} d\omega + \frac{e^{2\sigma z_1}}{1 - e^{-2\sigma z_1}} \frac{e^{-\sigma H}}{e^{-\sigma z_1}} \int_0^\sigma (e^{|\omega| z_1} + \right. \\ & \left. + e^{-|\omega| z_1}) d\omega \right. \\ & = \begin{cases} \frac{2e^{-\sigma z_1}}{z_1}, & z_1 < H, \\ \frac{2e^{-\sigma H}}{z_1}, & z_1 > H. \end{cases} \end{aligned} \quad (I.102)$$

В случае двухмерной задачи, повторяя ход получения величины $\|\delta_1\|_C$ (§ 2), можно найти

$$\|\delta_2\|_C \leq \frac{2\bar{K}_2}{\pi} e^{-\sigma z_1} (\sigma z_1 + 1), \quad (I.103)$$

где \bar{K}_2 — некоторая константа, которая, судя по аналогии с одномерным вариантом, по порядку величин совпадает с величиной K_2 , определяемой (I.51). Для случая $z_1 < H$ в этом можно убедиться, проделав выкладки, совершенно аналогичные выкладкам одномерного варианта.

Итак, интерполяционный процесс является сходящимся: при $\Delta x \rightarrow 0$, $\|\delta\|_C$ также равномерно стремится к нулю. Из (I.102) и (I.103) легко видеть, что при $z_1/\Delta x \rightarrow \infty$, $\|\delta_2\| = 0$, т. е. при $z_1/\Delta x \rightarrow \infty$, $\tilde{\varphi}(x) = \bar{\varphi}(x)$, $\varphi(x, y) = \bar{\varphi}(x, y)$. При достаточно больших z_1 значение

$$\|\delta\|_C \leq \|\delta_1\|_C \left[1 + O\left(\frac{1}{z_1}\right) \right], \quad (I.104)$$

т. е. с точностью до величины порядка $O\left(\frac{1}{z_1}\right)$ совпадает с оценками, получаемыми на основании теоремы Котельникова. Таким образом, в соответствии с определением, данным в § 2, при $z_1/\Delta x \rightarrow \infty$ метод является оптимальным. На самом деле, как уже упоминалось, производить расчеты при очень больших значениях $z_1/\Delta x$ нельзя из-за плохой обусловленности систем уравнений. Однако практически это вовсе не обязательно. Учитывая малость величин $\|\delta_2\|_C$ в формулах (I.102) и (I.103), вполне достаточно того, чтобы выполнялось условие $z_1 \geq (1,0 \div 2,0) H$. С учетом же реального уровня ошибок измерений, независимо от величины H , в большинстве случаев можно ограничиться значениями $z_1 \geq \Delta x$, а в некоторых случаях $z_1 \geq 0,5 \Delta x$. Полученные оценки позволяют конструктивно оценить точность метода в зависимости от значений Δx , z_1 и H ,

а в совокупности с оценками величины $\|\delta_1\|$ (§ 2) — получить полную величину ошибок интерполяции. Их эффективность полностью подтверждается результатами исследований на теоретических примерах.

Исследования на теоретических примерах [11, 16]. Расчеты производились для функций $U(x, H)$ (1.29). Для удобства сравнения моделей с реальными условиями все линейные расстояния выражены в безразмерных единицах x/H , z_1/h и т. д., где h — шаг вспомогательной квадратной или равномерной сетки, принятый за условную единицу длины.

Значения наложенной случайной помехи $\xi(x_i)$ получены по таблицам случайных чисел и характеризуются следующими величинами:

$$\sigma_{\xi} = \sqrt{\frac{\sum \xi_i^2}{n-1}} \quad (I.105)$$

— среднеквадратическая погрешность исходных данных;

$$K^{(\text{исх})} = \frac{\sigma_{\xi}}{\max |U|}; \quad K_1^{(\text{исх})} = \frac{\max |\delta U|}{\max |U|}; \quad (I.106)$$

$$K_2^{(\text{исх})} = \sqrt{\frac{\sum \xi_i^2}{\sum U_i^2}}.$$

Аналогично характеризуются и погрешности результатов вычислений

$$\sigma_u = \sqrt{\frac{\sum \delta U_i^2}{n-1}}; \quad (I.107)$$

$$K(U) = \frac{\sigma_u}{\max |U|}; \quad K_1(U) = \frac{\max |\delta U|}{\max |U|}; \quad K_2(U) = \sqrt{\frac{\sum \delta U_i^2}{\sum U_i^2}}. \quad (I.108)$$

В таблицах и тексте приняты следующие обозначения: x_i — абсциссы исходных пунктов (в единицах h); $\max \Delta x_i$ — максимальное расстояние между пунктами; $\mu_1(U) = \delta_1(U)/\max |U|$ — характеристика теоретически возможной точности восстановления функции $U(x, H)$ (см. табл. 1); r_s — радиус автокорреляции полезного сигнала; r_{μ} — радиус автокорреляции случайной помехи. С целью сокращения текста в таблицах представлены лишь результаты для центральных частей моделей, где ошибки интерполяции максимальны.

1. Наиболее простой вариант: значения U заданы в узлах равномерной сети с шагом $\Delta x = h$, $H/h = 2,0$.

Интерполяция на середину интервала между узлами производилась по формулам весового скользящего среднего, полу-

ченным по предлагаемой методике (I.84), $z_1/h=1,5$ и на основе интерполяционной формулы Лагранжа при $n=2$ (линейная интерполяция) и $n=10$ (табл. 2).

Таблица 2

Результаты интерполяции на модели 1

x	U теоретические	U фактические	δU	По формуле Лагранжа $n=10$		Линейная интерпретация	
				U	δU	U	δU
0,0	1,0000	0,9949	-0,0051	0,9946	-0,0054	0,9412	-0,0588
0,5	0,9412	0,9415	-0,0003	0,9416	+0,0004	0,9000	-0,0412
1,0	0,8000	0,8041	+0,0041	0,8047	+0,0047	0,7906	-0,0094
1,5	0,6400	0,6393	-0,0007	0,6400	0,0000	0,6500	+0,0100
2,0	0,5000	0,4974	-0,0026	0,4974	-0,0026	0,5151	+0,0151
2,5	0,3902	0,3904	+0,0002	0,3901	-0,0001	0,4038	+0,0136
3,0	0,3077	0,3089	-0,0012	0,3090	+0,0013	0,3182	+0,0105
3,5	0,2462	0,2459	-0,0003	0,2465	+0,0003	0,2538	+0,0076
4,0	0,2000	0,1993	-0,0007	0,1996	-0,0004	0,2056	+0,0056

Как видно из табл. 2, результаты интерполяции по Лагранжу при $n=10$ и с использованием предлагаемой методики примерно равноценны, но на порядок лучше, чем при линейной

Таблица 3

Результаты интерполяции на модели 2

x	U исходные	U теоретические	$z_1/h = 1,0$		$z_1/h = 3,0$	
			U фактические	δU	U фактические	δU
-6,0	0,200	0,200	0,200	0,000	0,200	0,000
-5,5	0,229	—	—	—	—	—
-5,0	—	0,265	0,263	0,002	0,265	0,000
-4,0	0,360	0,360	0,360	0,000	0,361	-0,001
-3,0	—	0,500	0,509	-0,009	0,498	+0,002
-2,0	—	0,692	0,695	-0,003	0,690	+0,002
-1,5	0,800	—	—	—	—	—
-1,0	—	0,900	0,900	0,000	0,901	-0,001
-0,8	0,934	—	—	—	—	—
0	—	1,000	1,005	-0,005	1,000	0,000
+0,5	0,973	—	—	—	—	—
+1,0	—	0,900	0,898	+0,002	0,900	0,000
+1,5	0,800	—	—	—	—	—
+2,0	0,692	0,692	-0,693	-0,001	0,692	0,000
+3,0	—	0,500	0,500	0,000	0,500	0,000
+3,2	0,468	—	—	—	—	—
+4,0	—	0,360	0,360	0,000	0,360	0,000
+5,0	0,265	0,265	0,262	0,003	0,264	0,001
+6,0	0,200	0,200	0,200	0,000	0,200	0,000

интерполяции. Обращает внимание резкое увеличение ошибок в точках $x_i=0,0; 1,0; 2,0$ и т. д., которое объясняется смещением исходных узлов относительно эпицентров аномалии. При отсутствии смещения точность повышается примерно на порядок. Полученная величина $K_1=0,5\%$ всего в 1,5 раза больше соответствующего значения $\mu=0,36\%$.

2. Исходные данные заданы в неравноотстоящих узлах y_i ; максимальная купюра $\max \Delta x_i=2,5 h$ смещена относительно максимума функций U ; $H/h=3,0$; $\min (H/\Delta x_i)=1,2$.

Расчеты проводились при значениях параметров $\varepsilon_0=0,001$; $z_1/h=1,0$ и $z_1/h=3,0$. Результаты представлены в табл. 3.

Рассматриваемый пример хорошо иллюстрирует эффект повышения точности при увеличении значения z_1/h : при $z_1/h=1,0$ величина $K_1(U)=0,9\%$, тогда как при $z_1/h=3,0$ $K_1(U)=0,2\%$. Если $z_1/h \geq 2,0$, то этот эффект уловить очень трудно, так как подбор интерполирующей конструкции нужно производить с очень высокой точностью. Следует отметить, что полученное значение K_1 значительно меньше величины $\mu_1(U)$ при $H/h=1,2$. Это объясняется быстрым убыванием ошибок интерполяции при убывании от эпицентра аномалии.

3. В табл. 4 ($H/h=3,0$) и 5 ($H/h=5,0$) приведены результаты расчетов в более сложных условиях, когда максимальные купюры $\Delta x_{\max}=3,0 h$ и $\Delta x_{\max}=4,8 h$ приурочены к эпицентру аномалий $U(x, 0)$. Исходные значения заданы точно $K_2^{\text{исх}}=0,001$ и со случайной ошибкой $K_2^{\text{исх}}=0,2$. Расчеты проводились при значениях $z_1/h=2,0$; $\varepsilon_0 \approx \sigma_\xi$.

Как видно из табл. 4, $K_2^{\text{исх}}=0,001$ величина $K_1(U)$ равна 2,5% при $\max \Delta x_i=3,0 h$ и 12% при $\max \Delta x_i=4,8 h$, что значительно меньше соответствующих значений $\mu_1(U)$. Это же соотношение между теоретическими и фактическими ошибками интерполяции наблюдается при увеличении H . При $H/h=5,0$ и тех же параметрах вычислительной схемы $K_1(U)=1,1\%$ ($\max \Delta x_i=3,0 h$) и $K_1(U)=5\%$ ($\max \Delta x_i=4,8 h$).

Анализ получаемых результатов (табл. 4 и 5) показывает, что значения U в эпицентре аномалии уверенно восстанавливаются и в том весьма сложном случае, когда $K_2^{\text{исх}}=0,2$. Сравнительно большие значения $K_1=19\%$ ($H/h=3,0$; $\max \Delta x_i=3,0 h$) и $K_1=17\%$ ($H/h=5,0$; $\max \Delta x_i=3,0 h$) обусловлены тем, что в ближайших к эпицентру узлах $x_i=-2,8; -1,5; +1,5$ значения ξ_i оказались систематически завышенными. При более благоприятном распределении случайных ошибок даже при максимальной купюре $\max \Delta x_i=4,8 h$, $K_1=17\%$ при $H/h=3,0$ и $K_1=7\%$ при $H/h=5,0$.

Вследствие близости значений радиусов автокорреляции полезного сигнала и помехи при $H/h=3,0$ ($r_s/r_\xi \approx 2,0$) эффек-

Результаты интерполяции на модели 3 ($H/h = 3,0$)

x/h	U теоретические	$\max \Delta x_i = 3,0 h$				$\max \Delta x_i = 4,8 h$			
		$K^{(исх)} = 0,001$		$K^{(исх)} = 0,2$		$K^{(исх)} = 0,001$		$K^{(исх)} = 0,2$	
		U исходные	U фактические	U исходные	U фактические	U исходные	U фактические	U исходные	U фактические
-7,0	0,155	0,155	0,156	0,168	0,189	0,155	0,155	0,165	0,176
-6,5	0,176	0,176	—	0,207	—	0,176	—	0,200	—
-6,0	0,200	—	0,200	—	0,162	—	0,200	—	0,167
-5,5	0,229	0,229	—	0,167	—	0,229	—	0,181	—
-5,0	0,265	—	0,267	—	0,266	—	0,263	—	0,266
-4,0	0,360	—	0,362	—	0,378	—	0,358	—	0,376
-3,5	0,423	0,424	—	0,476	—	0,424	—	0,465	—
-3,0	0,500	—	0,499	—	0,524	—	0,501	—	0,510
-2,8	0,534	0,534	—	0,481	—	0,534	—	0,493	—
-2,0	0,692	—	0,696	—	0,671	—	0,670	—	0,649
-1,5	0,800	0,800	—	0,794	—	—	—	—	—
-1,0	0,900	—	0,891	—	0,776	—	0,813	—	0,767
0,0	1,000	—	0,976	—	0,812	—	0,883	—	0,828
+1,0	0,900	—	0,892	—	0,774	—	0,874	—	0,796
+1,5	0,800	0,799	—	0,682	—	—	—	—	—
+2,0	0,692	0,692	0,693	0,730	0,659	0,692	0,692	0,722	0,667
+3,0	0,500	—	0,499	—	0,502	—	0,501	—	0,501
+3,2	0,468	0,467	—	0,393	—	0,467	—	0,409	—
+4,0	0,360	—	0,360	—	0,336	—	0,358	—	0,340
+4,5	0,308	0,308	—	0,329	—	0,308	—	0,325	—
+5,0	0,265	0,264	0,264	0,116	0,188	0,264	0,264	0,148	0,206
+6,0	0,200	0,200	0,200	0,189	0,169	0,200	0,200	0,191	0,179
+7,0	0,155	0,156	0,156	0,261	0,210	0,156	0,155	0,238	0,200
$\max \delta U$		0,001	0,024	0,149	0,188	0,001	0,116	0,117	0,171
σ_{ξ}		0,000	0,007	0,074	0,074	0,000	0,040	0,056	0,070

тивной фильтрации случайных ошибок не происходит $-\sigma U \approx \sigma_{\xi}$. В то же время при $H/h = -5,0$ ($r_s/r_{\xi} \approx 3,0$) $\sigma_U = 0,6 \sigma_{\xi}$. Эти результаты находятся в полном согласии с возможностями оптимальной фильтрации (гл. II). Весьма важной является уже и констатация того факта, что при интерполяции в таких сложных условиях не происходит усиления случайных ошибок.

Следует отметить, что оставшиеся погрешности носят случайный характер. Это хорошо иллюстрируется рис. 11, где приведены результаты расчетов при $H/h = 5,0$ и различном положении максимальных купюр. Заметим, что при точном задании исходных данных купюры ($\max \Delta x_i = 5,0 h$) на периферии ано-

Результаты интерполяции на модели 3 ($H/h = 5,0$)

X/h	U теоретические	$\max \Delta x_i = 3,0 h$				$\max \Delta x_i = 4,8 h$			
		$K^{(исх)} = 0,001$		$K_2^{(исх)} = 0,2$		$K^{(исх)} = 0,001$		$K_2^{(исх)} = 0,2$	
		U исходные	U фактические	U исходные	U фактические	U исходные	U фактические	U исходные	U фактические
-7,0	0,338	0,338	0,338	0,355	0,372	0,338	0,338	0,353	0,366
-6,5	0,372	0,372	—	0,415	—	0,372	—	0,410	—
-6,0	0,410	—	0,409	—	0,421	—	0,410	—	0,419
-5,5	0,453	0,452	—	0,365	—	0,452	—	0,375	—
-5,0	0,500	—	0,500	—	0,507	—	0,498	—	0,516
-4,0	0,610	—	0,612	—	0,606	—	0,608	—	0,639
-3,5	0,671	0,672	—	0,745	—	0,672	—	0,736	—
-3,0	0,735	—	0,735	—	0,703	—	0,736	—	0,747
-2,8	0,761	0,761	—	0,687	—	0,761	—	0,695	—
-2,0	0,862	—	0,864	—	0,781	—	0,845	—	0,829
-1,5	0,917	0,917	—	0,908	—	—	—	—	—
-1,0	0,962	—	0,957	—	0,821	—	0,917	—	0,892
0,0	1,000	—	0,989	—	0,829	—	0,950	—	0,929
+1,0	0,962	—	0,956	—	0,836	—	0,937	—	0,922
+1,5	0,917	0,917	—	0,752	—	—	—	—	—
+2,0	0,862	0,862	0,863	0,914	0,817	0,862	0,862	0,908	0,851
+3,0	0,735	—	0,736	—	0,734	—	0,736	—	0,734
+3,2	0,709	0,710	—	0,605	—	0,709	—	0,676	—
+4,0	0,610	—	0,609	—	0,589	—	0,609	—	0,583
+4,5	0,553	0,553	—	0,583	—	0,553	—	0,579	—
+5,0	0,500	0,499	0,500	0,291	0,397	0,499	0,500	0,315	0,407
+6,0	0,410	0,410	0,410	0,394	0,374	0,410	0,410	0,396	0,379
+7,0	0,338	0,339	0,338	0,486	0,416	0,338	0,338	0,469	0,409
$\max \delta U$		0,001	0,011	0,209	0,170	0,001	0,059	0,185	0,071
σ_{ϵ}		0,000	0,004	0,104	0,082	0,000	0,019	0,089	0,046

мали значения U восстанавливаются практически без погрешности.

Реализация метода на ЭВМ. Изложенный алгоритм был реализован в системе программ, подробно рассмотренной в книге [81], которая в течение ряда лет использовалась при практических расчетах. В настоящее время А. И. Ширгиновой составлена новая система программ «Карта», позволяющая решать задачи интерполяции данных на квадратную сеть, фильтрации случайных ошибок, получения поверхностей тренда и задачи обработки гравиметрических и магнитных аномалий при произвольной поверхности наблюдений. Следует отметить, что различие наблюдений по высоте иногда целесообразно

(или даже необходимо) учитывать и для геологических признаков, не являющихся гармоническими. Для этого можно воспользоваться рассматриваемым здесь методом, с учетом, однако, того обстоятельства, что интерполяция по высоте производится на основе решения граничной задачи теории потенциала, и в данном случае будут проявляться (по оси Oz)

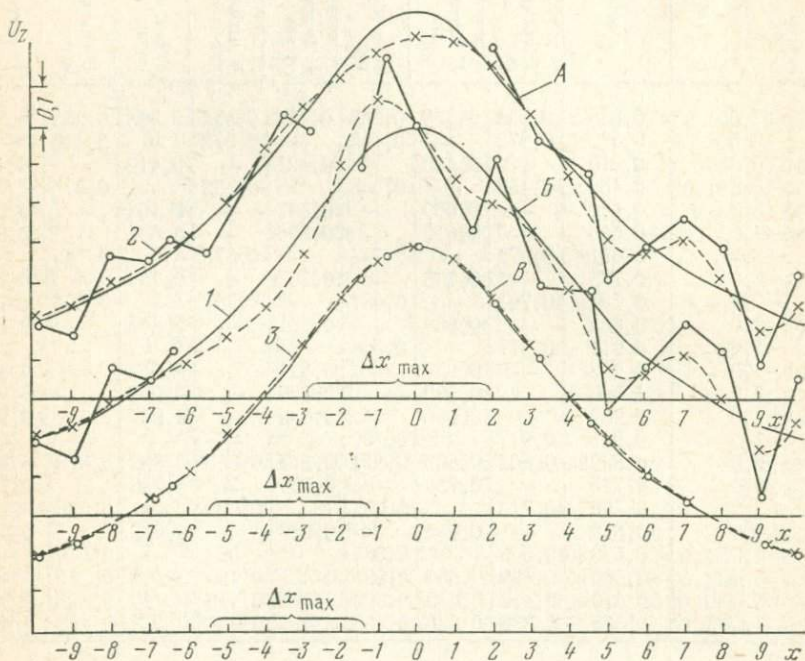


Рис. 11. Результаты расчетов на теоретическом примере:
 А — $K_2^{\text{исх}} = 0,2$; $\max \Delta x_j = 4,8$ в эпицентре аномалии; Б — $K_2^{\text{исх}} = 0,2$; $\max \Delta x_j = 5,0$ на периферии аномалии; В — $K_2^{\text{исх}} = 0$; $\max \Delta x_j = 5,0$ на периферии аномалии; 1 — теоретическая кривая; 2 — исходные значения; 3 — значения, полученные в результате интерполяции к узлам равномерной сетки

те свойства гармонических функций, на которые указывалось при рассмотрении метода сеток.

Основной особенностью вновь составленной для машины БЭСМ-4 системы программ* является то, что она соответствует некоторым стандартам, принятым во ВНИГНИ при составлении программ обработки данных. Это позволяет при подготовке исходных данных использовать автоматизированную информационно-поисковую систему (ИПС) (гл. V).

* Возможность решения геофизических задач с помощью программы здесь не рассматривается.

Все входящие в систему массивы исходных и промежуточных данных хранятся на магнитной ленте (МЛ) с соблюдением стандартов программ «Лента» и «Поиск». Можно, например, легко организовать выборку по любому из признаков геологических объектов, хранящихся в массиве, в прямоугольнике $-a < x < a$, $-b < y < b$ или на площади более сложной структуры, выбрать измерения определенной точности (если она указана) и т. п. При решении небольших по объему задач предусмотрена возможность работы без ИПС.

Управление работой системы осуществляется с помощью управляющей программы, функции которой заключаются в анализе задания, вызове с МЛ в оперативную память той или иной рабочей программы и ее настройке. Порядок проработки рабочих программ задается в информационных картах. Предусмотрена возможность подключения к системе дополнительных программ.

Подготовка исходных данных. При использовании ИПС информация готовится согласно требованиям программы «Лента», записывается на МЛ и затем выбирается на барабан по необходимым форматам с помощью программы «Поиск» (гл. V). Значения признаков, координаты пунктов и данные, характеризующие точность наблюдений, записываются в прямом виде (последовательно по каждому объекту (пункту) наблюдений). Порядок последовательности объектов произвольный. Если использование ИПС не предусматривается, то исходные данные выписываются по отдельным прямоугольным площадям со сторонами ph и qh , где h — среднее расстояние между пунктами наблюдений, $p \cdot q \leq 32\,000$.

Программа «Формат». Предназначена для упорядочения исходной информации и формирования массива исходных данных. Если $p \cdot q > 4000$, то обрабатываемый прямоугольник автоматически разбивается на ряд прямоугольных форматов $p_f \cdot q_f \leq 4000$.

Формируется поисковое задание и с помощью программы «Поиск» выбираются необходимые исходные данные для каждого формата, осуществляется их расстановка по ближайшим узлам квадратной сети. Если в узел попадает несколько пунктов, предпочтение отдается ближайшему к центру. Часть избыточных пунктов записывается на МЛ и может быть использована для независимого контроля точности интерполяции.

Программа «Граница». Исходной информацией служат координаты вершины контура исследуемой области. Программа находит все граничные узлы и определяет в них значения f с помощью линейной интерполяции. Если в некотором квадрате вспомогательной квадратной сетки наблюдения отсутствуют (имеет купюра), то вычисление соответствующего значения α_k не производится. Однако с целью повышения точности интерполяции значения α_k в этих квадратах определяются и впослед-

ствии исправляются в каждом итерационном цикле с помощью линейной интерполяции. Последняя выполняется следующим образом: в каждом квадранте отыскивается ближайший квадрат, в котором имеются наблюдения, производится интерполяция по двум значениям a_k на ось абсцисс, а затем по оси абсцисс — в соответствующий узел. Размеры купюр не должны превышать $10h \cdot 10h$. Это ограничение не является недостатком алгоритма и сделано лишь для упрощения программы.

Программа «Итерации». Решение системы уравнений (I.87) производится по формуле (I.95). Предусмотрена возможность работы с различными функциями $G_k(x - k\Delta x)$, т. е. различными коэффициентами $A = [a_{i,k}]$. Порядок матрицы может изменяться, максимально его значение равно $2N+1=21$. Если высоты пунктов наблюдений считаются постоянными, имеется возможность учитывать неравноточность исходных данных — в формуле (I.95) значения принимаются равными ϵ_i .

Подбор аппроксимирующей конструкции производится одновременно для $2N$ столбцов вспомогательной квадратной сетки. При достижении заданной точности $|\delta_i^{(k)}| < \epsilon_{0i}$ полученный массив смещается на N столбцов, вводится очередной массив исходной информации и вычисления продолжаются для следующих $2N$ столбцов (с перекрытием в N столбцов). По окончании подбора программа вычисляет значения φ в узлах квадратной сетки (с шагом h) и также формируется массив $\{a_k\}$ на МЛ. По нему с помощью специальных программ легко получить необходимые линейные трансформанты, а также более густую равномерную сеть значений интерполируемого признака — с шагом $h/2$, $h/4$ и т. д. Наиболее просто вычисляются значения поверхности тренда: увеличением (после определения a_k) значения $z_1 B G_k$ в 2, 3, 4 и т. д. раз, которые можно рассматривать как результаты весового осреднения исходной функции. Чем больше увеличение z_1 , тем более плавными получаются поверхности тренда.

Программа «Контроль» предназначена для проведения независимого контроля качества интерполяции. Исходные значения «избыточных» наблюдений и значения, вычисляемые в них, сводятся в таблицу и выдаются на печать.

Выдача результатов. Результаты выдаются на печать по форматам в виде таблиц или по прямоугольной сети на АЦПУ или изображаются в виде изолиний на графопостроителе.

Отличительной особенностью системы программ, отражающей свойства самого метода, является простота определения параметров вычислительной схемы. Работа программ определяется всего тремя величинами (без учета сведений о границах участка и самих исходных данных): ϵ_0 , h , z_1 . E_0 — это величина среднеквадратической ошибки наблюдений (или ошибки измерений в каждом пункте) плюс амплитуда непредставительных

в данном масштабе, точнее при данных расстояниях между наблюдениями, характеристик признаков, которые необходимо отфильтровать. Величина h обычно принимается равной среднему расстоянию между значениями исходных данных с учетом следующих соображений: уменьшение h приводит к увеличению времени счета за счет увеличения общего числа узлов вспомогательной регулярной сетки, а увеличение — к появлению «избыточных» значений φ_i . Значение z_1 обычно принимается равным h^* . Таким образом, параметры вычислительной схемы на каждом участке исследований легко определяются на основе априорных сведений, практически без какого-либо проведения пробных расчетов.

Важным обстоятельством является и возможность определения точности интерполяции по величине ошибок $\|\delta_1\|_C$ и $\|\delta_2\|_C$, а также наличие эффективной фильтрации случайных ошибок (см. § 3, гл. II).

Время работы программы может довольно сильно колебаться в зависимости от соотношения z_1/h , сложности поведения интерполируемой функции, точности подбора интерполируемой конструкции и др. В большинстве случаев производительность не ниже 500 пунктов/ч, составляя в среднем примерно 1000 пунктов/ч и выше.

На рис. 12 и 13 представлены результаты «ручной» интерполяции по нерегулярной сети данных и получаемой по изложенному методу квадратной сети. Сравнение исходных и интерполированных значений показывает, что различия лежат в пределах субъективных особенностей в подходе к проведению изолиний. Следует отметить, однако, более закономерный и, очевидно, более обоснованный ход изолиний на «машинных» картах. Это особенно четко видно по профилю 1—1 на рис. 13, в.

В работе [82] приведена иная реализация метода. При небольшом числе точек ($n \leq 60$) значения a_h находят путем решения системы линейных уравнений (I.87), вид функций G_h определяется формулой (I.90). Если число $n > 60$, то значения a_h определяются специальным приемом на основе ортонормирования функций G_h . Чтобы ограничить порядок решаемых систем уравнений, учитывают лишь значения, лежащие в некоторой окрестности. Радиус этой окрестности выбран таким же, как и в работах [11, 84] и равен $5\Delta x$. Аналогично выбрано и значение z_1 : $0,5\Delta x \leq z_1 \leq 2\Delta x$. Фильтрация случайных ошибок может осуществляться двумя способами. Однако критерием сглаживания служит точечное среднеквадратическое отклонение.

* Получение значений z_1 и h легко можно автоматизировать. Однако их определение пользователем не представляет никаких затруднений, позволяя в ряде случаев «нестандартно» приспособляться к конкретным условиям и в результате сокращать время расчетов.

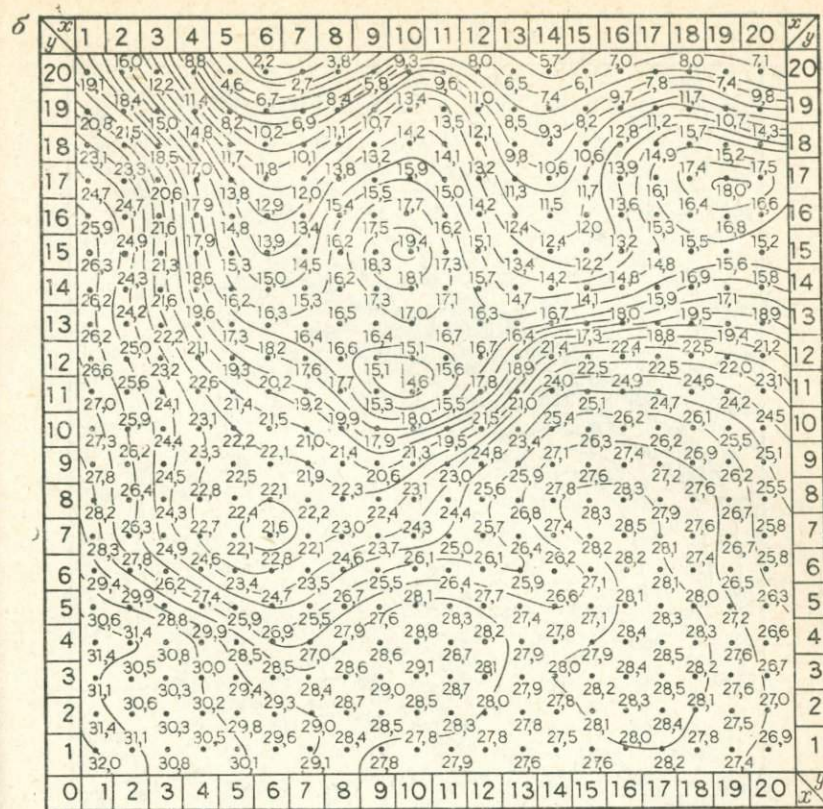
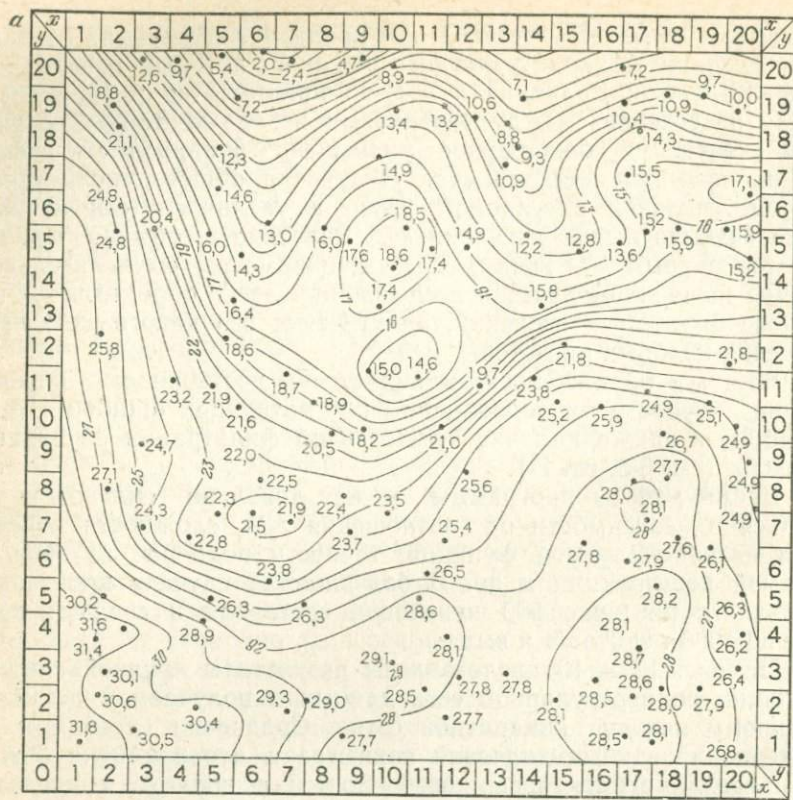


Рис. 12. Участок карты, построенной графическим способом:

а — по неравномерной исходной сети; б — результаты машинной интерполяции, $\epsilon_0 = \pm 0,3$, $z' = \Delta x$

Еще два варианта реализации метода предложены в работе [2]. В первом варианте вместо решения системы уравнений (1.87) предлагается решать аналогичную, но переопределенную систему уравнений по методу наименьших квадратов. Этим достигается уменьшение количества определяемых значений a_k и, очевидно, времени счета, если сокращение их числа будет значительным. Однако выполнение неравенства (1.82) теперь уже не гарантируется и, таким образом, точность аппроксимации даже в исходных узлах может быть неравномерной, а в некоторых случаях значения $|\varphi_i - \bar{\varphi}_i|$ могут стать недопустимо большими. Второй алгоритм предусматривает использование функций $G_k(x - k\Delta x)$ весьма сложного вида, в ре-

зультате чего вычисления станут слишком громоздкими. Однако на основании оценки точности подобного рода аппроксимаций, проведенной выше, следует ожидать, что точность расчетов при этом не изменится.

В заключение следует подчеркнуть, что с практической точки зрения все рассмотренные алгоритмы пригодны лишь в тех случаях, когда плотность исходной сети обеспечивает проведение интерполяции с погрешностью, не превышающей нескольких процентов. Если же сеть наблюдений редка и предельные ошибки интерполяции, определяемые на основании теоремы Котельникова, составляют десятки процентов и более, применение формальных методов приближения функций становится малоэффективным или просто неудовлетворительным. При построении карт вручную геолог в таких ситуациях широко использует эвристические приемы, позволяющие привлекать косвенные данные. Моделирование этого процесса — геологиче-

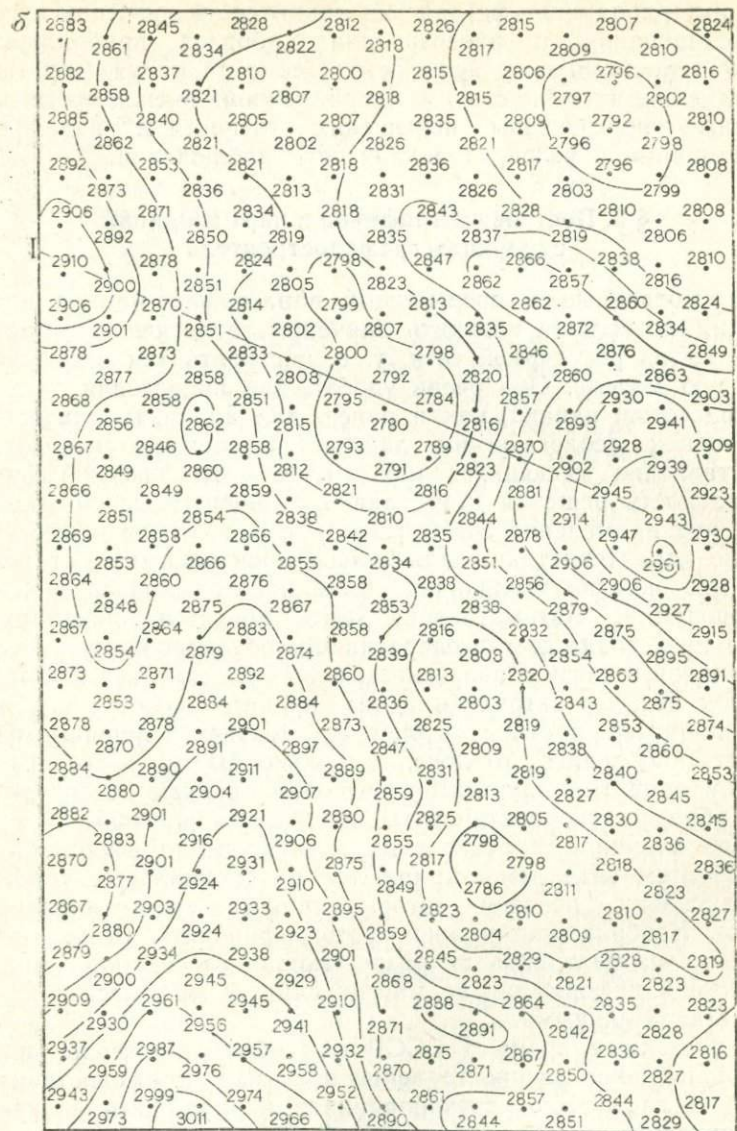
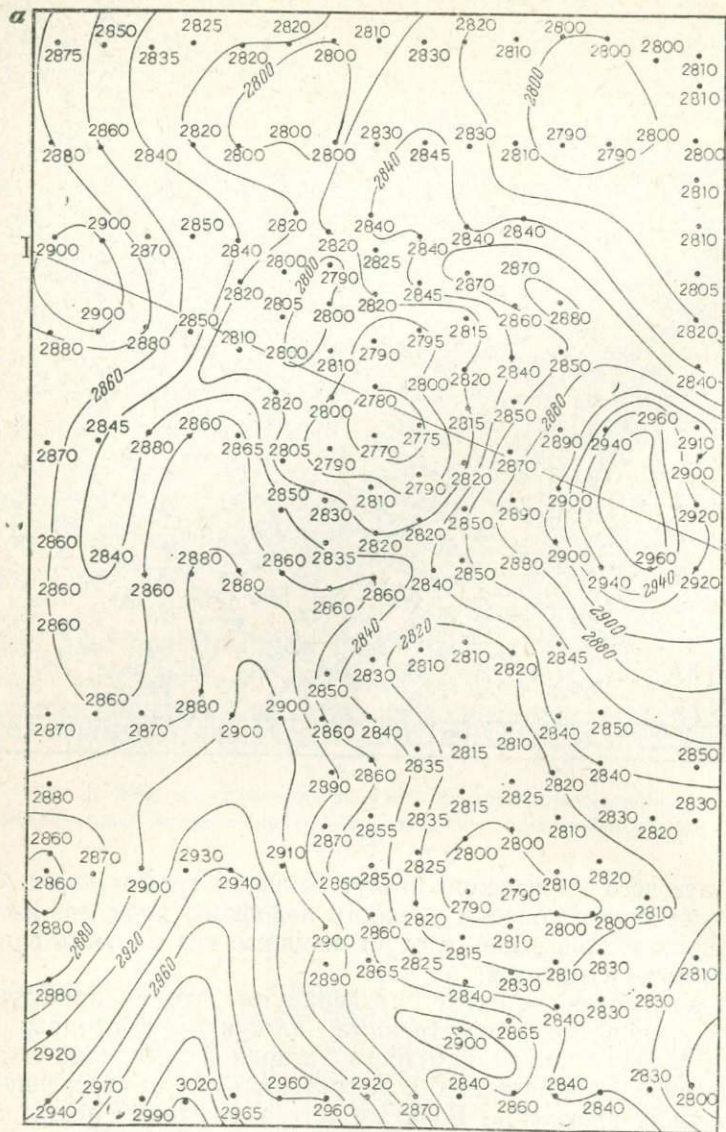
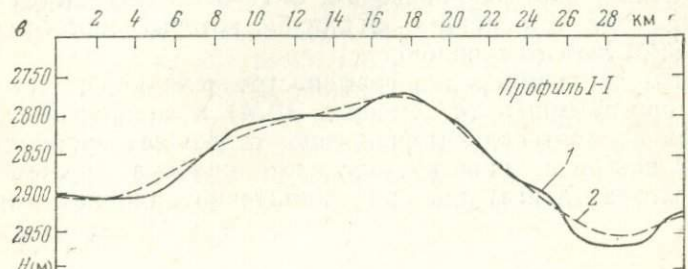


Рис. 13. Построение графическим способом структурной карты по сейсмическим данным:

а — по неравномерной исходной сети; б — по результатам машинной интерполяции, $\epsilon_0 = 10$ м, $z' = -2$ км; в — кривые (1, 2), построенные по профилю I—I по исходной и равномерной сети соответственно



ской интерполяции — фактически сводится к решению задач прогнозирования и распознавания образов приближаемых функций по комплексу признаков. Несмотря на ряд серьезных затруднений, именно с этим направлением, очевидно, связаны основные перспективы повышения точности и надежности интерполяции в условиях редкой сети наблюдений.

§ 6. Построение графиков и карт изолиний с помощью графопостроителя

Рассмотренные в предыдущих параграфах методы интерполяции позволяют получить значения геологических признаков в узлах регулярной сети, т. е. построить так называемую числовую модель. Она очень удобна для проведения дальнейших расчетов, однако, как и ранее, для анализа и качественных оценок геологических данных совершенно необходимо представление имеющихся материалов в графической форме. Актуальность решения этой задачи становится ясной уже потому, что применение ЭВМ позволяет рассчитывать многочисленные характеристики, являющиеся производными от исходных, многократно увеличивая объем необходимой графической интерполяции. Автоматизация этого процесса обеспечивает, кроме того, возможность оперативного обновления графических материалов в связи с получением новых данных, а в некоторых случаях без нее затруднен и сам процесс расчетов на ЭВМ. Так, например, расчеты в системе человек — машина очень существенно зависят от того, насколько быстро исследователь сможет оценить полученные расчетные данные и определенным образом скорректировать процесс вычислений. В таких случаях получаемые результаты в алфавитно-цифровом или графическом виде демонстрируются на специальных устройствах, дисплеях, основой которых служит электроннолучевая трубка. Изображения можно сфотографировать. Однако качество этих изображений, их разрешающая способность, емкость вмещаемой информации во многих случаях являются пока неудовлетворительными.

В настоящее время в СССР при обработке геологических данных широкое распространение получили графопостроители типа «Атлас». Это двухкоординатный планшетный графопостроитель с шаговым приводом. С 1973 г. выпускается модификация этого устройства — «Атлас-3», который способен работать с ЭВМ любого типа [58].

Подготовка команд для графопостроителя может осуществляться программным (с помощью ЭВМ) и аппаратным путем. Наличие аппаратного интерполятора сокращает время работы ЭВМ и повышает производительность прибора за счет плавного разгона двигателей на достаточно длинных прямых линиях.

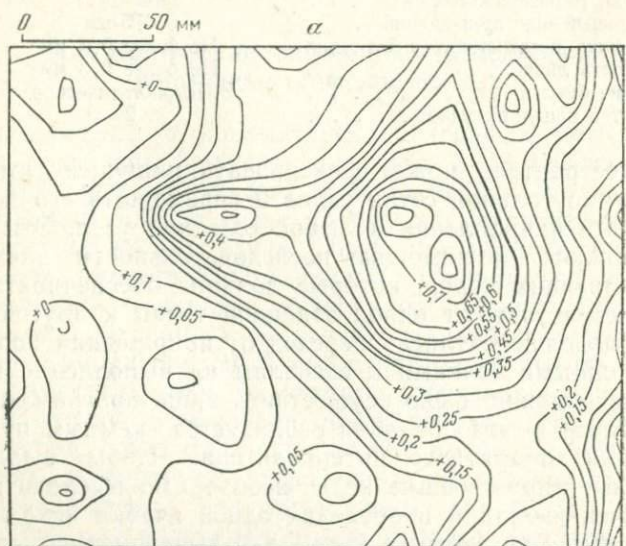
Технические характеристики устройства «Атлас-3»

Рабочая поверхность планшета	60×800 мм
Режим работы	Шаговый и режим приращений
Скорость в шаговом режиме	до 800 мм/с
Формат записи	Восьмивекторный
Максимальная скорость в режиме приращений (с интерполятором)	1300 мм/с
Минимальный шаг приращений	0,075/мм
Абсолютная погрешность построения	Не более 0,4 мм
Волнистость линии	Не более 0,3 мм
Крепление бумаги	Электростатическое
Количество пишущих узлов	2

Графопостроитель может фиксировать пишущее устройство в узлах прямоугольной сетки узлов и перемещать его по одному из восьми направлений в любой соседний узел. Эти операции выполняются с помощью последовательности элементарных управляющих кодов, которые должны обеспечивать перемещение через узлы от одной заданной точки к другой точке. В старых моделях «Атласа» переход от приращений координат между исходными точками к сигналам на выполнение элементарного шага можно было осуществить лишь программным путем, в «Атласе-3» эта операция реализуется схемным путем, за счет наличия аппаратного интерполятора. Чтобы выработать элементарные управляющие коды, необходимо вычислить координаты точек изолиний в пределах одной ячейки исходной регулярной сети. Для этого, а также для вычерчивания надписей координат произвольных осей составлен ряд программ, входящих в математическое обеспечение «Атласа» [75, 76].

Программы, входящие в математическое обеспечение «Атласа», осуществляя поиск координат изолиний с помощью линейной интерполяции. Исходными данными для них служат числовые модели, которые определяют качество получаемых графических материалов. Чем мельче сеть числовой модели, тем точнее могут быть определены координаты ее изолиний, тем менее заметны изломы вычерчиваемых кривых. Однако сгущение сети связано со значительным ростом времени работы ЭВМ (именно поэтому поиск координат изолиний проводится с помощью линейной интерполяции) и быстрым, пропорционально квадрату величины уменьшения шага, ростом объема данных. Считается, что достаточно, когда шаг сети в масштабе строящейся карты не превышает 5 мм. В этом случае угловатые очертания изолиний, связанные с этой причиной, становятся незаметными. По данным работы [76], при сети исходных узлов 10×10 мм и относительно спокойном рельефе признака максимальная ошибка из-за использования линейной интерполяции равна примерно 0,5 мм, в наиболее трудных ситуациях она может достигать величины диагонали квадрата сетки—1,4 см. Ошибки, зависящие от конструкции графопостроителя «Атлас», равны примерно 0,5—0,6 мм.

В математическое обеспечение «Атласа» входит целый комплекс программ, составленный Л. Г. Перфильевым и группой соавторов. Это программа вычерчивания изолиний, прямых линий и надписей, штриховых линий и др. В этих программах основ-



б

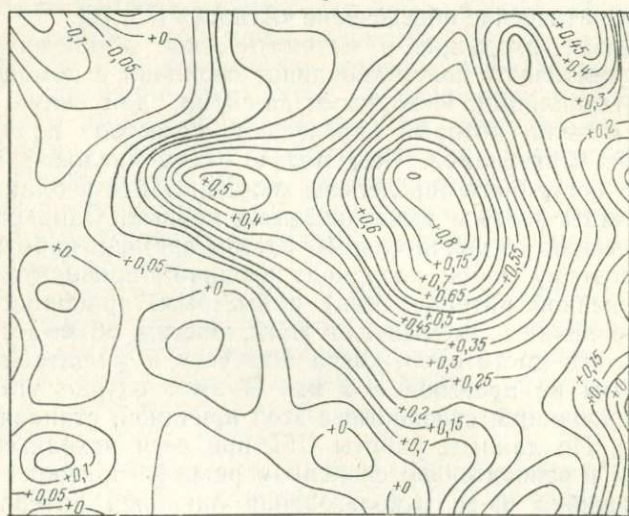


Рис. 14. Карты, построенные с помощью «Атласа»: а — по сети точек $16,25 \times 16,25$ мм; б — по сети точек 5×5 мм

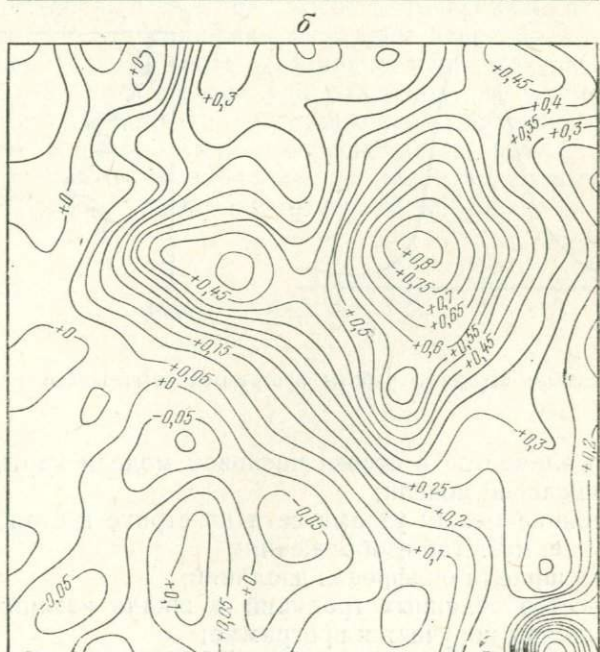
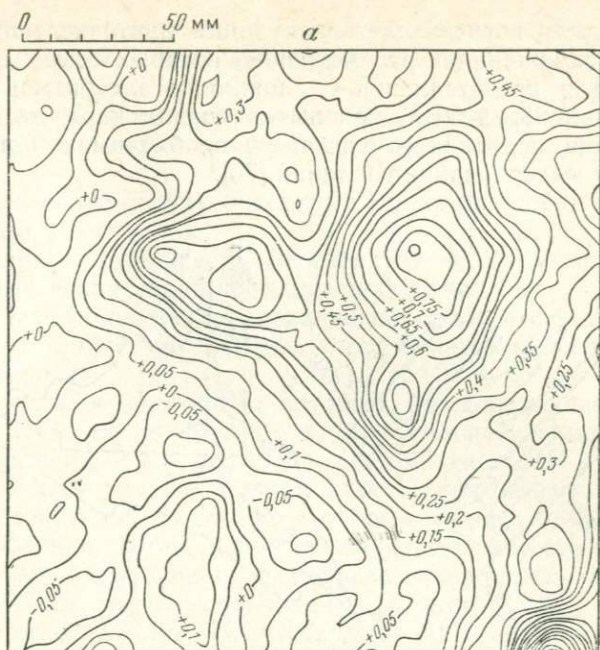


Рис. 15. Сглаживание карт изолиний:
 α — карта, искаженная нормально распределенной ошибкой с дисперсией 0,05 единиц;
 β — та же карта после осреднения в окне радиусом 1,42 шага сети

ную сложность представляет реализация многочисленных алгоритмов обслуживания, рассмотрение которых здесь нецелесообразно. Для представления о сложности алгоритмов и задачах, стоящих перед ЭВМ при вычерчивании карт, рассмотрим, какая нужна исходная информация, необходимая для работы программы вычерчивания изолиний [76]:

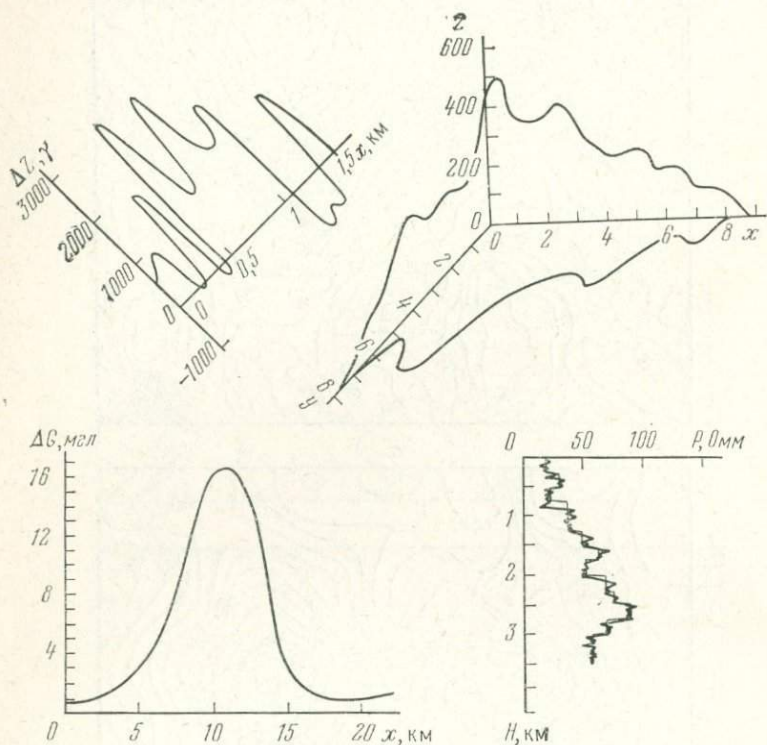


Рис. 16. Различные варианты работы программы вычерчивания графиков

- число элементов в строке числовой модели карты; число строк всей числовой модели;
- расстояние между узлами сети по строке и столбцу матрицы в мм, т. е. масштаб изображения;
- высота шрифта оцифровки изолиний;
- количество заданных градаций в шкале изолиний, если построение шкалы поручается программе;
- шаг изолиний;
- градации шкалы изолиний, которые задаются, когда шкалу нельзя построить программным путем;

— число, обычно кратное десяти, которое определяет рядность представления значений функций при оцифровке чертежей;

— условное число, определяющее режим вывода помеченных точек: не выводить, чертить первым пером, чертить вторым пером;

— расстояние между соседними линиями координатной сетки ($\Delta\xi$, $\Delta\eta$) в мм;

— размер плеча креста (в долях от $\Delta\xi$, $\Delta\eta$), наносимого в точке пересечения линий координатной сетки;

— условное число, определяющее режим вывода сетки координатных линий: не чертить, чертить первым пером, чертить вторым пером;

— уровень, относительно которого вычерчиваемые изолинии могут отличаться по фактуре и (или) по цвету, данные о координатах в исходной системе и системе числовой модели карты и некоторые другие. Время работы программы вычерчивания изолиний зависит от шага сетки исходных значений и густоты изолиний. В большинстве случаев расстояние принимается равным 1 см. При среднем расстоянии между изолиниями примерно 1 см время вычерчивания масштаба 400×400 мм составляет около 30 мин. На каждую подпись значения признака в узле расходуется 5—7 с. Скорость вычерчивания линий программами «Ось» и «График» в среднем равна 3—4 см/с.

На рис. 14 представлены сравнительные результаты вычерчивания карт на «Атласе» при различных шагах исходной сетки. На рис. 15 показано сглаживающее действие фильтрации случайных ошибок методом осреднения на результаты графического изображения. Образцы работы стандартных программ «Ось» и «График» представлены на рис. 16 [75].

**ФИЛЬТРАЦИЯ СЛУЧАЙНЫХ ОШИБОК.
ТРЕНД-АНАЛИЗ**

Фильтрация случайных ошибок и тренд-анализ относятся к важнейшим этапам обработки данных даже при традиционных методах обработки, хотя обычно в этом случае они не доставляют особых хлопот. Особое значение эти методы приобретают в связи с переходом к массовой обработке геологических данных математическими методами на ЭВМ. Их использование необходимо при решении задач интерполяции и построения карт, подсчета запасов и во всех тех случаях, когда наблюдения содержат ошибки и (или) когда необходимо выделить региональную (локальную) компоненту геологического признака, характеризующуюся относительно большими (меньшими) размерами, чем остальные составляющие, и, соответственно, меньшей (большей) изменчивостью, т. е. различающиеся по своей частотной характеристике. Наиболее высокочастотными являются, естественно, случайные ошибки измерений, наиболее низкочастотными — региональные составляющие.

Области применения тренд-анализа весьма разнообразны. Он применяется при структурно-тектонических построениях, в литологии, палеонтологии, геохимии и др. Очень разнообразны и методы тренд-анализа — от сложного современного математического аппарата до простейших эвристических построений, поэтому далее рассматривается формально-методическая основа лишь основных тенденций этого направления.

Как известно, измерения всегда производятся с той или иной ошибкой. Это особенно справедливо в отношении измерений геолого-геофизических характеристик, где неудовлетворительная во многих случаях точность аппаратуры усугубляется наложением компонент, отражающих влияние побочных факторов, имеющих аналогичную геолого-геофизическую природу, но не интересующих исследователя в данный момент или интересующих лишь в «чистом» виде, без смешения с другими составляющими. Еще недавно основой при обработке данных в геологии была лишь глазомерная оценка явлений, которая во многих случаях при фильтрации случайных ошибок и тренд-анализе является не только достаточной, но и вполне эффективной. Этому способствовало широкое применение графических методов обработки и интерпретации, со своей стороны в немалой степени также обязанные эффективности глазомерных оценок. В настоящее время с появлением огромных массивов

данных такая индивидуальная обработка стала очень обременительной. К тому же появилась возможность использовать при их обработке электронные вычислительные машины. Это привело к быстрому возрастанию интереса к математическим, формализованным методам фильтрации случайных ошибок и тренд-анализа. Особенно глубокое и широкое применение они получили в геофизических методах разведки.

§ 1. Постановка задачи

Обозначим через $f(x)$ и $\xi(x)$ соответственно точное значение измеряемой функции и аддитивной функции ошибок или, как принято в теории связи, будем называть $f(x)$ полезным сигналом, $\xi(x)$ — помехой. Тогда наблюдаемые значения

$$\varphi(x) = f(x) + \xi(x). \quad (II.1)$$

Еще раз подчеркнем, что помеха $\xi(x)$ может быть обусловлена как ошибками измерений и тогда она является дискретной функцией координат, так и наложением на $f(x)$ аналогичной по природе характеристики геологического процесса, которые обычно являются непрерывными функциями координат.

Если $\xi(x)$ — систематическая ошибка, то она может быть обнаружена и устранена лишь в результате контрольных наблюдений или при вычислении аппаратурных ошибок. Если же значения ошибок можно считать случайными, то их влияние может быть уменьшено в результате соответствующей математической обработки. Это и есть задача фильтрации случайных ошибок. В геологии к ошибкам измерений часто добавляется высокочастотная составляющая полезного сигнала, которая при данном шаге наблюдений Δx неотличима от них по своему частотному составу. Такие компоненты являются непредставительными и не только не могут быть удовлетворительно восстановлены, но, так же как и ошибки измерений, во многих случаях существенно ухудшают решение задач, поэтому, за некоторыми исключениями, между этими непредставительными компонентами, радиус автокорреляции которых $r \approx \Delta x$, и ошибками измерений не делается различий и ставится задача их подавления. Далее, там где в этом нет особой необходимости, под ошибками измерений будет всегда подразумеваться сумма этих компонент полезного сигнала и собственно ошибок измерений*.

В тех случаях, когда поведение $\xi(x)$ связано с положением и расстоянием между пунктами наблюдений очень слабо и определяется физико-геологическими причинами, выделение полезного сигнала является задачей тренд-анализа.

* Наличие непредставительных компонент геологических признаков легко установить в тех случаях, когда амплитуды их существенно превышают аппаратурные ошибки измерений.

Эти задачи могут решаться одновременно. Между фильтрацией случайных ошибок и тренд-анализом нет принципиальных различий. В обоих случаях фактически решается одна и та же задача — разделение сигналов, отличающихся друг от друга по своей частотной характеристике. Однако частотная характеристика случайных ошибок в основном определяется шагом измерений Δx ($r_{\xi} \approx \Delta x$) и, в принципе, при необходимости может изменяться по желанию экспериментатора. Например, если нельзя уменьшить ошибки измерений, то точность получаемых в результате фильтрации значений $\hat{f}(x)$ можно повысить, уменьшая величину Δx . Если необходимо, например, избавиться от региональных проявлений геологического признака, такой прием почти ничего не изменит, так как масштабы этих региональных проявлений являются объективной характеристикой и от методики эксперимента не зависят (при уменьшении Δx может лишь добавиться влияние еще более мелких неоднородностей). То же самое будет и в случае, когда необходимо, наоборот, избавиться от локальной составляющей. По сути это фактически одно и то же, так как локальная составляющая приближенно может быть определена как разность $\varphi(x) - \hat{f}(x)$, где $\hat{f}(x)$ — выделенная региональная компонента.

Таким образом, при фильтрации случайных ошибок определение искоемых характеристик помехи, необходимых для расчета фильтра, является сравнительно простым и формальным этапом, а при тренд-анализе представляет основные трудности, причем и характеристики полезного сигнала в этом случае определяются несравненно труднее. Действительно, обычно определять значения Mf^2 и $M\xi^2$ и их частотный состав приходится по суммарной функции $\varphi(x)$, в то же время, чтобы сделать это достаточно хорошо, необходимо их сначала разделить, т. е. сначала решить поставленную задачу. Лишь в некоторых случаях частотные и даже амплитудные характеристики можно получить по косвенным данным, например, если известны необходимые сведения о геологических объектах, создающих аномалии геофизических полей — помех или полезного сигнала. Как правило, необходимые для проведения тренд-анализа характеристики определяются очень приближенно, а в ряде случаев до проведения фильтрации они практически остаются количественно неизвестными, точнее неоднозначными: исследователь, например, делает целый набор карт тренда, находя подходящий вариант путем их анализа на основе какой-либо одной или даже группы гипотез, т. е. по степени соответствия предполагаемых и получаемых результатов*.

* Конечно, такой прием — только вынужденная необходимость, и к результатам следует относиться с соответствующей долей скептицизма, учитывая степень соответствия результатов и надежность самих гипотез.

Таким образом, фильтрация случайных ошибок и тренд-анализ предназначены для решения одной и той же задачи — выделения полезного сигнала на фоне помех* на основе имеющих различия в их $f(x)$ и $\xi(x)$ частотном составе. Однако для определенности, из-за существенных различий в самих исходных условиях решения задач и традиций, сложившихся в теории связи, где задачи борьбы с помехами получили особенно большое распространение и где впервые были заложены теоретические основы их решения, в тех случаях, когда помеха — случайная ошибка измерений или непредставительная «случайная» при данном значении Δx составляющая геологического признака, будем говорить о фильтрации случайных ошибок, а в тех случаях, когда помеха создается физико-геологическими причинами, причем размеры ее проявлений достаточно велики и не зависят сколько-нибудь существенно от величины Δx , — о тренд-анализе. При этом составляющая, характеризующая большей закономерностью в поведении — $f(x)$, называется трендовой или просто трендом, а $\hat{\xi}(x) = \varphi(x) - \hat{f}(x)$ — остаточной или локальной составляющей.

Результат, получаемый при фильтрации ошибок измерений и непредставительных характеристик признаков, всегда имеет четкий физический смысл: достигается уменьшение дисперсии ошибок значений наблюдаемого признака $f(x)$. Применение тренд-анализа возможно только в случае наличия четких физико-геологических предпосылок или тогда, когда различия в частотном составе составляющих очевидны, хотя формально выделение различных составляющих может быть проведено неограниченным количеством вариантов**. И, наконец, несмотря на физико-математическую общность рассматриваемых процессов, их реализация в ряде случаев проводится таким образом, что фильтры, рассчитанные на проведение тренд-анализа, не пригодны для фильтрации случайных ошибок и наоборот. Следует отметить, что из-за несовершенства технологии методов фильтрации при определении характеристик фильтра безразлично, что считать полезной составляющей, т. е. какую составляющую определять непосредственно с их помощью или с помощью вычисления разности между исходными значениями $\varphi(x)$ и этой составляющей.

В настоящей книге сделана попытка дать достаточно общее определение тренд-анализа, это важно потому, что очень часто понятие тренда изменяется в зависимости от конкретной задачи и от выбранного метода ее решения. Так, например, в рабо-

* В тренд-анализе понятие «помехи» может быть чисто условным, так как во многих случаях «помеха» представляет такой же интерес, как и полезный сигнал.

** Такой необоснованный тренд-анализ проводится, пожалуй, значительно чаще, чем при правильной постановке задачи.

те [21] тренд определяется как «любая составляющая процесса, период которой превышает длину реализации». Там это важно потому, что иначе при вычислении по данной реализации корреляционных функций и спектральных плотностей могут быть получены существенные ошибки. В геологии такое определение во многих случаях будет абсурдным, поскольку по своему смыслу решаемая тренд-анализом задача не зависит от длины реализации (в пределах величин, определяемых технологией расчетов). В работе [57] под трендом подразумеваются систематические составляющие изменчивости признака по площади в задачах картирования характеристик совокупности. Такое определение вряд ли является конструктивным, кроме того, вообще непонятно, почему остальная часть составляющих не должна быть систематической. На самом деле и здесь, естественно, подразумевается разделение составляющих по частотному составу. Еще более неудачным является понятие о полном тренде, содержащем компоненты первого, второго и более высоких порядков [57]. Вводится, в частности, понятие полного тренда второго порядка, являющегося суммой «линейных и квадратичной составляющих».

По сути дела это определение предусматривает наличие тренда во всех случаях, поскольку линейная и квадратичная составляющие (имеется в виду приближение полиномами первого и второго порядка) могут быть выделены всегда*, даже когда в этом нет никакого физического смысла. Дело не только в том, что такое определение непосредственно подталкивает к опасному формальному подходу, но и в том, что в зависимости от размеров участка исследований будут получаться сколь угодно существенно различные результаты.

В работе [70] предлагается в качестве тренда рассматривать любую упорядоченную последовательность чисел или элементов, отличающуюся от случайной. Такое определение направлено на выделение тренда с помощью статистических критериев, однако его эффективность определяется методикой выбора элементов, так как получаемые результаты полностью зависят от их размеров, т. е. фактически от их частотного состава.

В проблеме борьбы с помехами наиболее часто рассматриваются следующие задачи: 1) обнаружение сигнала; 2) восстановление сигнала.

Первая задача возникает тогда, когда уровень помех очень велик по сравнению с полезным сигналом (амплитуда помехи достигает и даже превышает амплитуду полезного сигнала) и можно лишь надеяться на обнаружение полезного сигнала, на установление того факта, присутствует ли он вообще. Такого рода задачи часто встречаются при определении наличия сла-

* За исключением того тривиального случая, когда процесс полностью описывается полиномом первого порядка.

бых геофизических аномалий малоамплитудных локальных структур в осадочной толще, выделенных по сейсмическим данным и др., когда определяются участки для дальнейшей, более детальной разведки. Иногда рассматривается задача различения сигналов. В этом случае необходимо определить, имеется ли только один полезный сигнал $f_1(x)$ или присутствует второй полезный сигнал $f_2(x)$, но и здесь ставится задача обнаружения наличия этих сигналов, т. е. она решается качественно.

Во второй задаче необходимо так восстановить полезный сигнал, чтобы он по возможности меньше отличался от истинного. Оптимальные методы восстановления — это методы, минимизирующие расстояние между $f(x)$ и $\hat{f}(x)$ в каком-либо определенном заранее смысле. Если различие полезного сигнала $f(x)$ и помехи $\xi(x)$ по частотному составу и амплитуде невелико, то, несмотря на использование методов выделения сигналов, значения $\hat{f}(x)$ могут характеризоваться большими ошибками и использоваться лишь при качественных оценках. Результаты, полученные с помощью решения первой и второй задач могут существенно различаться, поскольку в них используются разные критерии эффективности или верности результатов.

При обнаружении обычно рассматривается изменение величины отношения сигнал-помеха:

$$\rho_0 = \frac{\bar{f}^2}{\bar{\xi}^2} = \frac{Df}{D\xi}, \quad \rho_\Phi = \frac{Df}{D\varepsilon}, \quad (\text{II.2})$$

где ε — ошибка определения $f(x)$ после фильтрации, причем считается, что применяемый метод тем лучше, чем больше отношение ρ_Φ/ρ_0 .

При восстановлении сигналов наиболее часто употребляемой мерой сравнения функций f и \hat{f} , т. е. мерой расстояния между ними, является норма пространства L^2

$$\|\delta\|_{L^2} = \left(\int_a^b [f(x) - \hat{f}(x)]^2 dx \right)^{1/2}, \quad (\text{II.3})$$

т. е. квадратичное отклонение. Применение этой метрики, как и в методе наименьших квадратов, обуславливается главным образом тем, что при этом более просто получить необходимые результаты.

Можно использовать и расстояние в пространстве L^1 .

$$\|\delta\|_{L^1} = \int_a^b |f(x) - \hat{f}(x)| dx. \quad (\text{II.4})$$

Наиболее сильной нормой является норма в пространстве непрерывных функций C

$$\|\delta\|_C = \max_{a < x < b} |f(x) - \hat{f}(x)|, \quad (\text{II.5})$$

гарантирующая, в отличие от двух предыдущих, что расстояние между $f(x)$ и $\hat{f}(x)$ не превышает $|\delta|$ на всем $[a, b]$.

Необходимо отметить, что «нет решительно никаких общих оснований для предпочтения одного критерия другому» [93]. В идеале использование того или иного критерия должно зависеть от характера решаемых после фильтрации геологических задач. Однако, как правило, критерии верности для их решения неизвестны.

Ясно, что работа «частотного» фильтра основана на использовании различия в спектрах разделяемых сигналов. Если их спектры не перекрываются ($0 < \omega_f \leq \omega_0$, $\omega_0 < \omega_\xi < \infty$), то разделение, в принципе, можно произвести точно. На самом деле при решении задач фильтрации случайных помех и особенно тренд-анализа в геологии спектры полезного сигнала и помехи не только перекрываются, но и могут быть даже довольно близкими между собой. Разделение сигналов в этом случае может быть основано на том, что функция $\varphi(x)$ подвергается такому преобразованию, при котором по возможности сохраняются частоты, формирующие полезный сигнал, и, наоборот, «обрезаются» частоты, наиболее свойственные помехе.

Как и в ряде других отраслей науки, методы решения задачи тренд-анализа в геологии очень широко используют вероятностный подход. «Вероятностное описание является моделью, обусловленной необходимостью установления закономерностей в тех случаях, когда априори известны весьма скудные феноменологические данные, но имеется возможность сопоставить много наблюдений над явлениями, причем условия, в которых происходит опыт, остаются практически неизменными» [95]. На самом деле, как уже отмечалось, в геологии, как правило, нет ансамбля реализаций, рассматривается и обрабатывается фактически лишь единственная реализация функции $\varphi(x)$. Мало того, геолога обычно интересуют не среднестатистические результаты по всей реализации, а конкретные данные на ее участках — отдельных аномалиях, структурах, площадях и т. п. В таких условиях функции $f(x)$ и $\xi(x)$ (за исключением того случая, когда $\xi(x)$ — ошибки измерений) следует считать детерминированными.

Несмотря на это, вероятностная модель может быть полезной и даже эффективной, причем не только из-за наличия хорошо развитого аппарата, что, конечно, очень важно, но и ее приемлемого во многих случаях соответствия реальным условиям, особенно в случае фильтрации случайных ошибок и непредставительных компонент признаков. Здесь можно рассматривать две постановки: 1) случайной считается только помеха $\xi(x)$ и 2) случайными считаются обе составляющие — $f(x)$ и $\xi(x)$. Первая постановка в геологии обычно используется при решении задач обнаружения по методу накопления при боль-

шом уровне помех. При этом полезный сигнал $f(x)$ обычно считается постоянным. Это обусловлено в первую очередь тем, что необходимые сведения о полезном сигнале обычно неизвестны, а также слабым проникновением в геологию методов оптимальной фильтрации в такой постановке. Как правило, она рассматривается при решении геофизических задач, иногда в несколько смягченной форме, когда полезный сигнал известен с точностью до некоторых параметров, причем ставится не только задача обнаружения наличия полезного сигнала, но и определения этих параметров.

В условиях недостатка сведений о функциях $f(x)$ и $\xi(x)$ при проведении тренд-анализа в геологии возможна лишь вторая постановка. В связи с этим ниже будет рассмотрен линейный фильтр, получивший название фильтра Колмогорова-Винера, оптимальный (минимально-квадратический) для стационарных эргодических случайных процессов. Вместе с тем, не следует забывать, что его действительная оптимальность определяется соответствием реального и модели случайного процессов, а также качеством определения параметров последней.

Случайные процессы характеризуются своими вероятностными характеристиками, определяемыми по множеству возможных реализаций. Исключение составляют так называемые эргодические случайные процессы.

Случайный процесс называется эргодическим, если его вероятностные характеристики могут быть получены не только по ансамблю реализаций случайного процесса, но и по одной реализации за достаточно большой период времени. Важнейшими его характеристиками являются:

1. Математическое ожидание (среднее значение процесса $\xi(t)$)

$$M\xi(t) = \overline{\xi(t)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t) dt. \quad (\text{II.6})$$

Далее будем считать, что рассматриваемые случайные процессы являются центрированными, $\overline{\xi} = 0$.

2. Средний квадрат или дисперсия центрированного процесса

$$D\xi(t) = M[\xi(t)]^2 = \overline{\xi^2(t)} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi^2(t) dt. \quad (\text{II.7})$$

3. Корреляционная (автокорреляционная) функция

$$B_\xi(\tau) = M\xi(t)\xi(t+\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t)\xi(t+\tau) dt \quad (\text{II.8})$$

и взаимная корреляционная функция процессов $\xi(t)$ и $\eta(t)$

$$B_{\xi\eta}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \xi(t) \eta(t + \tau) dt. \quad (\text{II.9})$$

Если случайный процесс является стационарным, его среднее значение и средний квадрат не зависят от t , а функция $B_{\xi}(\tau)$ зависит только от τ *. Корреляционная функция $B(\tau)$ является четной $B(\tau) = B(-\tau)$, причем $B(0) \geq B(\tau)$. При $\tau \rightarrow \infty$ для централизованного процесса $\lim_{\tau \rightarrow \infty} B(\tau) = B(\infty) = 0$, при $\tau \rightarrow 0$ $\lim_{\tau \rightarrow 0} B(\tau) = \overline{\xi^2}$.

Таким образом, средний квадрат (и дисперсия, если $\bar{\xi} = 0$) стационарного случайного процесса равен $B(0)$. Значение $B(\tau)/\overline{\xi^2}$ есть коэффициент корреляции R между $\xi(x)$ и $\xi(x + \tau)$ или нормированная корреляционная функция:

$$B_n(\tau) = \frac{B(\tau)}{B(0)}. \quad (\text{II.10})$$

Максимальное значение $B_n(\tau) = B_n(0) = 1$; при $|\tau| > 0$ $|B_n(\tau)| \leq 1$, причем при $\tau \rightarrow \infty$ $B_n(\tau) \rightarrow 0$. Для стационарных случайных процессов можно указать такое τ_0 , что при $\tau > \tau_0$ величины $\xi(\tau)$ и $\xi(t + \tau)$ можно считать некоррелированными для любого t , т. е. можно указать такое τ_0 , что при $\tau > \tau_0$ величина $B_n(\tau)$ не превысит заданной. Величину $\tau_0 = r$ называют радиусом или интервалом корреляции. Обычно радиус корреляции r принимают равным тому значению τ , при котором $B_n(\tau) = (0,3 - 0,4) B(0)$. Иногда величину $r = \tau_0$ определяют из выражения

$$r = \frac{1}{2} \int_{-\infty}^{\infty} |B_n(\tau)| d\tau. \quad (\text{II.11})$$

По корреляционной функции стационарного случайного процесса можно определить его энергетический спектр $G(\omega)$. Они связаны друг с другом парой преобразований Фурье (теорема Хинчина — Винера):

$$G(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} B(\tau) e^{-i\omega\tau} d\tau = \frac{2}{\pi} \int_0^{\infty} B(\tau) \cos \omega\tau d\tau \quad (\text{II.12})$$

и

$$B(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} G(\omega) e^{i\omega\tau} d\omega = \int_0^{\infty} G(\omega) \cos \omega\tau d\omega. \quad (\text{II.13})$$

* Обратное не справедливо. Однако если случайный процесс является гауссовым (имеет нормальное распределение), то он полностью определяется своим смешанным моментом второго порядка — корреляционной функцией.

Поскольку энергетический спектр является Фурье-преобразованием корреляционной функции, он не учитывает фазовую характеристику сигналов и дает лишь усредненное распределение энергии процесса по частоте.

Формулы (II.12) и (II.13) справедливы только при условии, если

$$\int_{-\infty}^{\infty} |B(\tau)| d\tau \leq M, \quad \int_{-\infty}^{\infty} |G(\omega)| d\omega \leq N, \quad (\text{II.14})$$

где M и N — константы. Поэтому здесь рассматривается центрированный случайный процесс $\bar{\xi}(t) = 0$. При $\tau = 0$ из (II.13) находим

$$\int_0^{\infty} G(\omega) d\omega = B(0), \quad (\text{II.15})$$

т. е. площадь между кривой энергетического спектра и осью абсцисс равна среднему квадрату $\bar{\xi}^2(t)$ или средней мощности стационарного случайного процесса.

Аналогичным образом связаны между собой взаимный энергетический спектр $G_{\xi\eta}(\omega)$ и взаимная корреляционная функция $B_{\xi\eta}(\tau)$.

При линейных преобразованиях случайных реализаций энергетический спектр изменяется следующим образом:

$$G_{\text{пр}}(\omega) = |K(\omega)|^2 G(\omega), \quad (\text{II.16})$$

где $K(\omega)$ — частотная характеристика линейного преобразования. В результате, в зависимости от вида $K(\omega)$, усиливается или ослабляется влияние соответствующих частот в энергетическом спектре исходного сигнала.

В общем виде решение задачи тренд-анализа может быть записано следующим образом:

$$\hat{f}(x) = L[\varphi(x)], \quad (\text{II.17})$$

где L — некоторый линейный оператор свертки, удовлетворяющий условию $\|L\| = A < \infty$. Поскольку при расчетах на ЭВМ всегда используются дискретные значения функции $\varphi(x)$, оператор L является матрицей, форма и распределение коэффициентов которой зависят от методики вычислений.

§ 2. Оптимальный фильтр Колмогорова-Винера

Пусть, как в формуле (II.1),

$$\varphi(x) = f(x) + \xi(x),$$

где $f(x)$ и $\xi(x)$ — реализации стационарного эргодического случайного процесса, не коррелированные между собой, т. е.

$B_{f\xi}(\tau) = 0$. Необходимо построить такой фильтр (линейное преобразование) [28, 62, 93], чтобы

$$M\varepsilon_{\Phi}^2 = M | f(x) - \hat{f}(x) |^2 = \min, \quad (\text{II.18})$$

где $\hat{f}(x)$ — оценка полезного сигнала после фильтрации. Таким образом, в качестве критерия верности используется минимум среднеквадратической погрешности. Учитывая формулу (II.15), можем записать

$$M\varepsilon_{\Phi}^2 = \int_0^{\infty} G_e(\omega) d\omega. \quad (\text{II.19})$$

Обозначим через $\psi(\omega)$ вещественную (поскольку сдвиг по фазе полагается равным нулю) частотную характеристику линейного оптимального фильтра L , энергетические спектры полезного сигнала и помехи — через $G_f(\omega)$ и $G_{\xi}(\omega)$.

Тогда, после проведения линейного преобразования L , величина $G_e(\omega)$ будет равна

$$G_e(\omega) = G_f(\omega) [1 - \psi(\omega)]^2 + G_{\xi}(\omega) \psi^2(\omega) \quad (\text{II.20})$$

или после соответствующих преобразований

$$G_e(\omega) = \left[\psi(\omega) \sqrt{G_f(\omega) + G_{\xi}(\omega)} - \frac{G_{\xi}(\omega)}{\sqrt{G_f(\omega) + G_{\xi}(\omega)}} \right]^2 + \frac{G_f(\omega) G_{\xi}(\omega)}{G_f(\omega) + G_{\xi}(\omega)}.$$

Легко видеть, что $G_e(\omega) = \min$ тогда, когда выражение в скобках равно нулю. Отсюда получим частотную характеристику оптимального фильтра

$$\psi(\omega) = \frac{G_f(\omega)}{G_f(\omega) + G_{\xi}(\omega)}. \quad (\text{II.21})$$

Из формулы (II.20) с учетом (II.21) получим

$$G_e(\omega) = \frac{G_f(\omega) G_{\xi}(\omega)}{G_f(\omega) + G_{\xi}(\omega)} \quad (\text{II.22})$$

и соответственно

$$M\varepsilon_{\Phi}^2 = \int_0^{\infty} \frac{G_f(\omega) G_{\xi}(\omega)}{G_f(\omega) + G_{\xi}(\omega)} d\omega. \quad (\text{II.23})$$

Из формулы (II.23) видно, что $M\varepsilon_{\Phi}^2$ может быть равно нулю, если $G_{\xi}(\omega) = 0$ и если спектры $G_f(\omega)$ и $G_{\xi}(\omega)$ не перекрываются, т. е. когда $G_f(\omega) G_{\xi}(\omega) = 0$.

Итак, поставленная задача получения оптимальной оценки полезного сигнала с помощью преобразования L , частотная

характеристика которого $\psi(\omega)$ определена по (II.21), сводится к следующему:

$$\widehat{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} L(x-t) \varphi(t) dt, \quad (\text{II.24})$$

где

$$L(x-t) = 2 \int_0^{\infty} \psi(\omega) \cos \omega(x-t) d\omega. \quad (\text{II.25})$$

При расчетах на ЭВМ интегралы типа (II.24) обычно вычисляются по методу весового скользящего среднего:

$$\widehat{f}(x) = \sum_{j=-N}^N c_j \varphi(x - j\Delta x), \quad (\text{II.26})$$

где $\Delta x = \text{const}$ — расстояние между значениями $\varphi(k\Delta x)$, $k=0, 1, 2, \dots, n$, а значения c_j в силу (II.25) определяются следующим образом:

$$c_j = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(\omega) \cos \omega j\Delta x d\omega. \quad (\text{II.27})$$

Пусть теперь необходимо выполнить не только оптимальное выделение сигнала, но и его линейное преобразование $\widehat{f}_{\text{тр}} = K[f(x)]$. Тогда вместо формулы (II.18) будем иметь

$$M \varepsilon_{\text{тр}}^2 = M |K[f(x)] - \widehat{f}_{\text{тр}}(x)|^2 = \min, \quad (\text{II.28})$$

а вместо $\psi(\omega)$ в (II.21)

$$\Psi_{\text{тр}}(\omega) = \frac{K(\omega) G_f(\omega)}{G_f(\omega) + G_{\xi}(\omega)}, \quad (\text{II.29})$$

где $K(\omega)$ — частотная характеристика линейного преобразования K , причем $\|K\| < \infty$.

Такой подход к вычислению линейных трансформаций целесообразен в тех случаях, когда результаты расчетов сильно зависят от ошибок входных данных, т. е. когда решаемая задача является некорректной (например, при вычислении производных $f^n(x)$).

Таким образом, частотная характеристика $\psi(\omega)$ оптимального в смысле среднеквадратического критерия преобразования получена. Для нормальных случайных процессов $f(x)$ и $\xi(x)$ этот оптимум сохраняется и в классе нелинейных преобразований. Это, однако, не означает того, что возможности повышения эффективности решения задач восстановления функций этим полностью исчерпываются. Вывод справедлив только для данного подхода к решению задачи: минимально-квадратический критерий и модель случайного стационарного эргодического про-

цесса без учета конкретных свойств подвергаемых разделению функций и специфики решаемых задач.

Дискретный оптимальный фильтр. При практической реализации оптимального фильтра необходимо учитывать дискретность измерений или, во всяком случае, необходимость дискретизации функций при расчетах на ЭВМ. Естественно считать, что при шаге Δx нет смысла пытаться восстанавливать частоты выше граничной: $|\omega| \geq \sigma = \pi/\Delta x$.

Действительно, если заменить исходный критерий (II.18), поставив задачу минимизации величины

$$M\epsilon_{\Phi}^2 = M |f(k\Delta x) - \hat{f}(k\Delta x)|^2, \quad (\text{II.30})$$

частотная характеристика оптимального дискретного фильтра $\tilde{\psi}(\omega)$ будет выглядеть аналогично $\psi(\omega)$ в (II.21) [28, 37]:

$$\tilde{\psi}(\omega) = \frac{\tilde{G}_f(\omega)}{\tilde{G}_f(\omega) + \tilde{G}_{\xi}(\omega)}, \quad (\text{II.31})$$

однако $\tilde{G}_f(\omega)$ и $\tilde{G}_{\xi}(\omega)$ — дискретные энергетические спектры полезного сигнала и помехи, определяются следующим образом:

$$\tilde{G}(\omega) = \sum_{n=-\infty}^{\infty} G\left(\omega + \frac{2\pi n}{\Delta x}\right). \quad (\text{II.32})$$

Его ошибка ϵ_{Φ}^2 равна

$$M\epsilon_{\Phi}^2 = \int_{-\pi/\Delta x}^{\pi/\Delta x} \frac{\tilde{G}_f(\omega) \tilde{G}_{\xi}(\omega)}{\tilde{G}_f(\omega) + \tilde{G}_{\xi}(\omega)} d\omega, \quad (\text{II.33})$$

а коэффициенты дискретного линейного преобразования, минимизирующего (II.30), определяются из выражения

$$c_j = \frac{\Delta x}{2\pi} \int_{-\pi/\Delta x}^{\pi/\Delta x} \frac{\tilde{G}_f(\omega)}{\tilde{G}_f(\omega) + \tilde{G}_{\xi}(\omega)} \cos \omega j \Delta x d\omega. \quad (\text{II.34})$$

Интересно отметить аналогию с введенным в § 2 гл. I на основе теоремы Котельникова определением оптимального восстановления функций: восстанавливается функция $\tilde{f}(x)$, спектр которой

$$\tilde{F}(\omega) = \sum_{-\infty}^{\infty} F\left(\omega + \frac{2\pi n}{\Delta x}\right), \quad |\omega| \leq \frac{\pi}{\Delta x}.$$

Реализация оптимального фильтра связана с решением следующих основных вопросов:

1) расчетом вычислительных конструкций;

2) определением исходных параметров \bar{f}^2 , $\bar{\xi}^2$, r_f , r_ξ — средних квадратов амплитуд и радиусов корреляции полезного сигнала и помехи;

3) определением степени эффективности фильтра, т. е. целесообразности его применения.

Вычислительные схемы. Отметим сразу, что численная реализация оптимальной фильтрации к настоящему времени разработана недостаточно. В основном способы рассчитаны на использование ручного счета, для ЭВМ пока нет даже реализации в двумерном варианте и, кроме того, нет вообще реализаций, рассчитанных на нерегулярную сеть.

Одна из первых вычислительных схем рассмотрена в работе [5], где дан пример расчета коэффициентов c_j при $\bar{f}^2 = \bar{\xi}^2$, $d_\xi = 1$,

$$r_f = 5r_\xi, \quad B_n(\tau) = \exp(-\tau^2/r^2).$$

Согласно (II.10), (II.12), (II.21) и (II.27), получим

$$G_f(\omega) = \frac{2}{\pi} \int_0^\infty \bar{f}^2 B_n(\tau) \cos \omega \tau d\tau = \bar{f}^2 r_f \frac{1}{\sqrt{\pi}} \exp\left(-\frac{\omega^2 r_f^2}{4}\right), \quad (\text{II.35})$$

$$G_\xi(\omega) = \bar{\xi}^2 r_\xi \frac{1}{\sqrt{\pi}} \exp\left(-\frac{\omega^2 r_\xi^2}{4}\right), \quad (\text{II.36})$$

$$\Psi(\omega) = \frac{1}{1 + \frac{G_\xi(\omega)}{G_f(\omega)}} = \frac{1}{1 + \frac{1}{5} e^{6\omega^2}}, \quad (\text{II.37})$$

$$L(x-t) = 2 \int_0^\infty \frac{\cos \omega(x-t)}{1 + \frac{1}{5} e^{6\omega^2}} d\omega. \quad (\text{II.38})$$

С помощью численного интегрирования получены следующие значения коэффициентов c_j : $c_0 = 0,148$; $c_1 = 0,136$; $c_2 = 0,107$; $c_3 = 0,069$; $c_4 = 0,033$.

Результаты обработки случайной функции [5] с указанными выше характеристиками показаны на рис. 17.

Использование дискретного оптимального фильтра позволяет ограничить пределы интегрирования величиной $\sigma = \pi/\Delta x$. Такие формулы для двух типов корреляционных функций получены в работе [38].

1.

$$B_n(\tau) = \left[1 + \left(\frac{\tau\pi}{2r}\right)\right]^{-1}. \quad (\text{II.39})$$

Тогда

$$\begin{aligned} \tilde{G}_f(\omega) &= \frac{2\bar{f}^2 r_f}{\pi} \sum_{-\infty}^{\infty} \exp\left(-\frac{2r_f}{\pi} \left|\omega + \frac{2\pi n}{\Delta x}\right|\right) = \\ &= \frac{2\bar{f}^2 r_f}{\pi} \sum_{-\infty}^{\infty} \exp(-2\alpha |t + 2n|) = \frac{2\bar{f}^2 r_f}{\pi} \cdot \frac{\text{ch}[2\alpha(1-|t|)]}{\text{sh } 2\alpha} \end{aligned} \quad (\text{II.40})$$

и

$$\tilde{G}_{\xi}(\omega) = \frac{2\bar{\xi}^2 r_{\xi}}{\pi} \cdot \frac{\text{ch}[2\beta(1-|t|)]}{\text{sh } 2\beta}, \quad (\text{II.41})$$

где $t = \omega \Delta x / \pi$, $\alpha = r_f / \Delta x$, $\beta = r_{\xi} / \Delta x$.

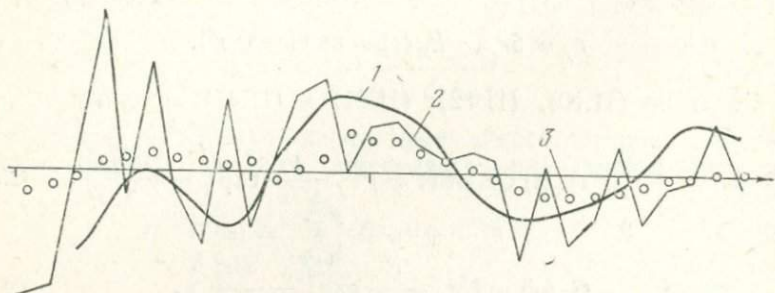


Рис. 17. Фильтрация помех с помощью оптимального фильтра:
1 — функция $f(x)$; 2 — функция $\varphi(x)$; 3 — функция $\hat{f}(x)$

Используя теперь формулу (II.34) и полагая $\bar{f}^2 / \bar{\xi}^2 = \rho_0$, получаем формулы для определения c_j и Me_{Φ}^2 :

$$c_j = \rho_0 \frac{\alpha}{\beta} \int_0^1 \frac{\cos j\pi t}{\rho_0 \frac{\alpha}{\beta} + \frac{\text{sh } 2\alpha}{\text{sh } 2\beta} \cdot \frac{\text{ch}[2\beta(1-t)]}{\text{ch}[2\alpha(1-t)]}} dt, \quad (\text{II.42})$$

$$Me_{\Phi}^2 = \frac{2\bar{f}^2 \alpha}{\text{sh } 2\alpha} \int_0^1 \frac{\text{ch}[2\alpha(1-t)]}{1 + \rho_0 \frac{\alpha}{\beta} \cdot \frac{\text{sh } 2\beta}{\text{sh } 2\alpha} \cdot \frac{\text{ch}[2\alpha(1-t)]}{\text{ch}[2\beta(1-t)]}} dt. \quad (\text{II.43})$$

2.

$$B_H(\tau) = \left[1 + \left(\frac{\tau\pi}{2r}\right)^2\right]^{-1}. \quad (\text{II.44})$$

В этом случае

$$c_j = \rho_0 \frac{\alpha}{\beta} \int_0^1 \frac{\cos j\pi t}{\rho_0 \frac{\alpha}{\beta} + \frac{\varphi(t, \beta)}{\varphi(t, \alpha)}} dt, \quad (\text{II.45})$$

$$M\varepsilon_{\Phi}^2 = 2\bar{f}^2\alpha \int_0^1 \frac{\varphi(t, \alpha)}{1 + \rho_0 \frac{\alpha}{\beta} \cdot \frac{\varphi(t, \alpha)}{\varphi(t, \beta)}} dt, \quad (\text{II.46})$$

где $\varphi(t, x) = \sum_{-\infty}^{\infty} \exp[-\pi x^2(t + 2n)]$.

Одновременно в работе [38] получены формулы для оценки влияния величины N , т. е. определено значение $\varepsilon^2(N)$.

Учитывая (II.33),

$$M\varepsilon_{\Phi}^2 = \frac{1}{2} \int_{-\pi/\Delta x}^{\pi/\Delta x} [|1 - \tilde{\Psi}(\omega)|^2 \tilde{G}(\omega) + | \tilde{\Psi}(\omega) |^2 \tilde{G}_\varepsilon(\omega)] d\omega, \quad (\text{II.47})$$

а $\tilde{\Psi}(\omega)$ при конечном значении N равно

$$\tilde{\Psi}^{(N)}(\omega) = \sum_{-N}^N c_j e^{-i\omega j \Delta x} \quad (\text{II.48})$$

можно получить значение $\bar{\varepsilon}^2(N)$. Для моделей $B_H(\tau)$, определяемых из формул (II.39) и (II.44), они соответственно равны

$$M\varepsilon_{\Phi}^2(N) = \bar{f}^2 \left\{ 1 - 2 \sum_{-N}^N c_j \frac{4\alpha^2}{4\alpha^2 + j^2\pi^2} + \sum_{-N}^N \sum_{-N}^N c_j c_k \left[\frac{4\alpha^2}{4\alpha^2 + (k+j)^2\pi^2} + \frac{1}{\rho_0} \frac{4\beta^2}{4\beta^2 + (k+j)^2\pi^2} \right] \right\}, \quad (\text{II.49})$$

$$M\varepsilon_{\Phi}^2(N) = \bar{f}^2 \left\{ 1 - 2 \sum_{-N}^N c_j \exp\left(-\frac{j^2\pi}{4\alpha^2}\right) + \sum_{-N}^N \sum_{-N}^N c_j c_k \left[\exp\left(-\frac{(k+j)^2\pi}{4\alpha^2}\right) + \frac{1}{\rho_0} \exp\left(-\frac{(k+j)^2\pi}{4\beta^2}\right) \right] \right\}. \quad (\text{II.50})$$

Сопоставляя величины $\bar{\varepsilon}^2(N)$ и $\bar{\varepsilon}^2$, можно определить необходимое значение N . Проведенные расчеты показали, что для достижения точности

$$\Delta = [\bar{\varepsilon}^2(N) - \bar{\varepsilon}^2] / \bar{\varepsilon}^2 = 0,01$$

при $\bar{f}^2 = \bar{\xi}^2 = 1$, $r_{\xi} = 1$, $r_f = 5r_{\xi}$ необходимое значение N в первом случае равно 3, а во втором $N=4$.

Составленная программа для ЭВМ предусматривает автоматическое определение N на основе полученных формул по значениям исходных параметров \bar{f}^2 , $\bar{\xi}^2$, r_f , r_{ξ} и вида $B_H(\tau)$.

В двухмерном варианте расчет коэффициентов резко усложняется. К настоящему времени разработаны лишь способы приближенного их определения для осесимметричного случая при расчетах вручную с помощью палеток [33, 83]. Если сеть ис-

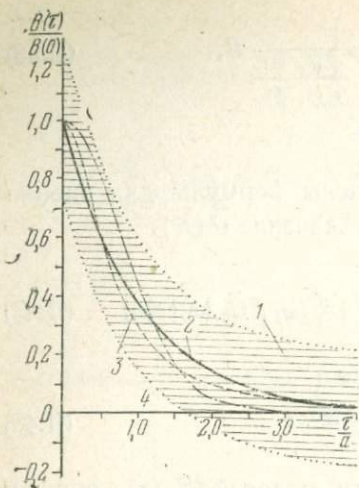


Рис. 18. Точность вычисления автокорреляционных функций:

1 — область значений $[B(\tau) \pm \sigma(\tau)]/B(0)$ для автокорреляционной функции $B_H(\tau) = \exp(-\tau/a)$; 2 — кривая $B_H(\tau) = \exp(-\tau/a)$; 3 — кривая $B_H(\tau) = (1 + \tau^2 \pi^2 / 4a^2)^{-1}$; 4 — кривая $B_H(\tau) = \exp(-\tau^2 \pi^2 / 4a^2)$

ний и более. При небольших значениях $|\tau|$ вычисление корреляционных функций не имеет смысла, так как среднеквадратическая ошибка определения может превышать величину $B(\tau)$.

Пусть оценка корреляционной функции равна

$$\widehat{B}(\tau) = \frac{1}{2T - |\tau|} \int_{-\tau + \frac{|\tau|}{2}}^{\tau - \frac{|\tau|}{2}} \varphi\left(x - \frac{\tau}{2}\right) \varphi\left(x + \frac{\tau}{2}\right) dx. \quad (\text{II.51})$$

Тогда, если $\varphi(x)$ — реализация нормального случайного центрированного процесса, дисперсия $\widehat{B}(\tau)$ может быть найдена по формуле [62]

$$\sigma^2(\tau) = \frac{2}{2T - |\tau|} \int_0^{2T - |\tau|} \left(1 - \frac{z}{2T - |\tau|}\right) [B^2(z) + B(\tau + z)B(\tau - z)] dz. \quad (\text{II.52})$$

Необходимые значения $\sigma^2(\tau)$ можно получить, подставив в формулу (II.52) соответствующие значения $B(\tau)$ и выполнив интегрирование. Анализ получаемых значений $\sigma^2(\tau)$ был проведен в работе [40] и показал следующее. Для того чтобы среднеквадратическая погрешность не превышала 10% $B(0)$, отноше-

ходных данных $\varphi(x_i)$ нерегулярна, трудности реализации оптимального фильтра приобретают принципиальный характер. Во всяком случае этот вопрос нигде даже не обсуждается.

Определение исходных параметров — значений f^2 , ξ^2 , r_f , r_ξ , необходимых для расчета оптимального фильтра, — является весьма сложной задачей. Если помеха — случайные ошибки наблюдений, то величину ξ^2 можно определить на основе контрольных измерений, а $r_\xi = (1 \div 2) \Delta x$. В остальных случаях необходимо знать все эти параметры априори или определить их на основе анализа суммарных значений $\varphi(x)$. В то же время известно [21, 62, 91], что для надежной оценки автокорреляционной функции даже если иметь отдельные, а не суммарные реализации $f(x)$ и $\xi(x)$, необходимы тысячи наблюдений

ние $2T/r$ должно быть свыше 100. Особенно показательным является рис. 18, на котором изображены ошибки вычислений в случае $2T/r=20$. Несмотря на то что доверительный интервал на графике корреляционной функции (1.26) равен всего $\pm\sigma$, в нем полностью умещаются графики корреляционных функций (1.27) и (1.28) и, кроме того, в области значений $B(\tau)/B(0) = 0,3-0,4$, т. е. в области, в которой определяется значение r , величина $\sigma(\tau)$ сравнима с величиной $B(\tau)/B(0)$. Из этого следует, во-первых, что анализ эмпирических функций автокорреляции при тех ограниченных значениях $2T/r$, которые обычно имеются в геологии, не может служить основой для выбора модели автокорреляционной функции* и, во-вторых, что значение r может быть определено лишь с очень большой погрешностью. Еще более плохие результаты будут получены при исследовании суммарного сигнала $\varphi(x)$. В этом случае весьма трудной становится и задача определения значений \bar{f}^2 и $\bar{\xi}^2$, даже без учета влияния недостаточной длины реализации. Учитывая все эти трудности, можно даже предположить, что ошибки определения значений параметров «на глаз», особенно при проведении предварительного, не формального разделения полезной составляющей и помехи могут быть не больше ошибок, получаемых на основе анализа корреляционных функций.

Изучение ошибок, возникающих в практике работ, насколько это известно, никем не проводилось. Можно отметить к тому же, что в опубликованных работах по использованию оптимального фильтра рассматриваются лишь теоретические примеры. Во всяком случае, проблема определения параметров оптимального фильтра чрезвычайно сложна и, очевидно, в большинстве ситуаций, которые встречаются при тренд-анализе геологических процессов, вряд ли может быть удовлетворительно решена.

Эффективность оптимального фильтра. Рассмотрим вопросы эффективности фильтра в условиях, когда исходные параметры определяются точно, а реальная модель соответствует теоретической [9]. Для этого воспользуемся формулой (II.23) и вычислим значение $\bar{\epsilon}_\Phi^2$ для случая $B_n(\tau) = \exp\left(-\frac{\tau^2}{r^2}\right)$:

$$G_f(\omega) = \frac{2\bar{f}^2}{\pi} \int_0^\infty B_n(\tau) \cos \omega\tau d\tau = \frac{\bar{f}^2 r_f}{\sqrt{\pi}} e^{-\omega^2 r_f^2/4}, \quad (\text{II.53})$$

$$G_\xi(\omega) = \frac{\bar{\xi}^2 r_\xi}{\sqrt{\pi}} e^{-\omega^2 r_\xi^2/4}, \quad (\text{II.54})$$

$$\bar{\epsilon}_\Phi^2 = \frac{\bar{f}^2 \bar{\xi}^2 r_f r_\xi}{\sqrt{\pi}} \int_0^\infty \frac{e^{-\omega^2 (r_f^2 + r_\xi^2)/4}}{\bar{f}^2 r_f e^{-\omega^2 r_f^2/4} + \bar{\xi}^2 r_\xi e^{-\omega^2 r_\xi^2/4}} d\omega. \quad (\text{II.55})$$

* В то же время модель автокорреляционной функции играет вполне заметную роль при получении частотной характеристики $\psi(\omega)$ и, соответственно, преобразования L .

Результаты вычисления интеграла (II.46) [9, 16] представлены на рис. 19, из которого видно, что средний квадрат погрешности после фильтрации $\bar{\epsilon}_{\Phi}^2$ в условиях, которые являются типичными в геологических ситуациях, уменьшается вовсе не так уже сильно. Например, даже при $r_{\xi}/r_f=0,2$; $f^2/\xi^2=1$ $\bar{\epsilon}_{\Phi}^2=0,25$, т. е. среднеквадратическая ошибка равна половине амплитуды полезной аномалии. При $r_{\xi}/r_f=0,5$ и том же соотношении амплитуд функций $\xi(x)$ и $f(x)$ величина $\bar{\epsilon}_{\Phi}^2=0,42$, т. е. близка к своему предельному значению $\bar{\epsilon}_{\Phi}^2=0,5$, достигаемому при $r_f/r_{\xi}=1,0$ *. Можно даже, очевидно, считать, что полученные результаты не противоречат интуитивным представлениям о возможности графического сглаживания, во всяком случае не отличаются от них принципиально. Более того, далее на основе изучения характеристик оптимальной фильтрации и полиномиальных методов будет показано, что примерно те же результаты можно получить с помощью обычного осреднения, но при правильном выборе его радиуса.

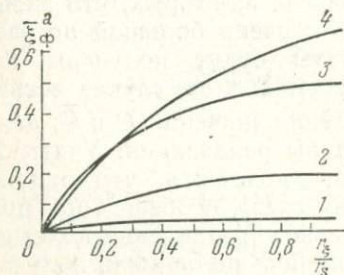


Рис. 19. Графики ошибок оптимального фильтра (1—3) и пересчета на высоту (4).

1 — $\bar{\xi}^2/f^2=0,04$; 2 — $\bar{\xi}^2/f^2=0,25$; 3 — $\bar{\xi}^2/f^2=1$;
4 — $z_0=0,2 r_f$, $\bar{\xi}^2/f^2=1$.

Итак, рассмотрена очень красивая теория метода, которая может быть с большим успехом использована при изучении сравнительной эффективности других методов фильтрации и тренд-анализа, для определения возможностей тренд-анализа, основанного на частотном подходе, а также при решении неустойчивых, некорректных задач, когда необходима одновременная фильтрация случайных ошибок. В то же время реализация его в геологических приложениях встречается с существенными трудностями**. Рассмотрим основные из них.

1. Определение исходных параметров, необходимых для расчета оптимального фильтра, представляет весьма сложную

* В последнем случае, когда аномалии фактически разделить невозможно, уменьшение амплитуды помехи достигается просто соответствующим уменьшением амплитуды суммарного поля. Это легко видеть из формул

$$(II.21) \text{ и } (II.23): \text{ при } G_f(\omega) = G_{\xi}(\omega), \psi(\omega) = 0,5 \text{ и } \bar{\epsilon}_{\Phi}^2 = \frac{1}{2} \int_0^{\infty} G_{\xi}(\omega) d\omega = \frac{1}{2} \bar{\xi}_2^2.$$

** Примеров практического использования фильтра Колмогорова — Винера при проведении тренд-анализа в геологии пока, очевидно, не имеется.

задачу, а их определение «на глаз» не позволяет говорить об оптимальности фильтра.

2. Предполагается статистически однородное распределение полезного сигнала и помехи в пределах обрабатываемой территории, тогда как на самом деле их радиусы автокорреляции и средние квадраты амплитуд на отдельных участках колеблются в широких пределах.

3. Получаемые результаты оптимальны (если все предположения верны) лишь в смысле минимизации величины $M[\hat{f}(x) - \hat{f}(x)]^2$ на всей реализации $(-T, T)$, т. е. в среднем по всей площади (профиле) исследований. Геолога в первую очередь интересуют именно отдельные участки съемки, отдельные аномалии, структуры и т. п. Поэтому принятый критерий оптимальности нельзя считать удачным.

Очевидно, неправильно в данном случае говорить об оптимальной фильтрации, поскольку соответствие модели и действительного геологического процесса сколько-нибудь серьезно даже не обсуждается. Поэтому вполне целесообразно применение способов, не являющихся оптимальными, но практически обеспечивающих хорошие результаты. Эти способы, не отличаясь существенно по точности от теоретической точности оптимального фильтра, могут иметь и серьезные преимущества, такие, например, как реальная возможность фильтрации случайных ошибок непосредственно в точках произвольной сети наблюдений и учет различия в точности определения значений $f(x)$ на каждом пункте, что в некоторых случаях очень важно.

Оптимальные фильтры получили большое распространение в радиотехнике. И это не случайно, поскольку там имеются огромные реализации процессов, проходящих в мало изменяющихся условиях, причем параметры процессов часто известны априори. Геология имеет дело с несравненно меньшими реализациями, значительно отличающимися друг от друга не только своими статистическими параметрами, но и физикой процесса. Учет этой специфики, очевидно, может приводить к весьма положительным результатам. К сожалению, пока можно, насколько это известно автору, привести лишь один такой пример [9, 16]. Он касается восстановления потенциальных функций, к которым относятся аномалии силы тяжести Δg и магнитного поля ΔZ (или другие составляющие гравитационного и магнитного потенциалов). Для таких функций можно по наблюдениям значений на плоскости $z=0$ вычислить их значения на высоте $z_0 > 0$, причем известно, что аномалии различного размера существенно по-разному убывают с высотой. Помеха, характеризующаяся меньшими значениями радиуса автокорреляции, убывает быстрее полезного сигнала, так как частотная характеристика этого преобразования равна $\psi(\omega) = e^{-|\omega|z_0}$.

Скорость убывания помехи с высотой можно оценить следую-

щим образом:

$$\begin{aligned}
 M\varepsilon_n^z &= \int_0^{\infty} G_{\varepsilon}(\omega) \Psi^2(\omega) d\omega = \frac{\bar{\varepsilon}_0^2 r_{\varepsilon}^2}{\sqrt{\pi}} \int_0^{\infty} e^{-\omega^2 r_{\varepsilon}^2/4 - 2\omega z_0} d\omega = \\
 &= \bar{\varepsilon}_0^2 e^{4z_0^2/r_{\varepsilon}^2} \left[1 - \Phi\left(\frac{2z_0}{r_{\varepsilon}}\right) \right], \quad (\text{II.55}')
 \end{aligned}$$

где $\Phi\left(\frac{2z_0}{r_{\varepsilon}}\right)$ — интеграл вероятности,

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt,$$

$G(\omega)$ определяется формулой (II.35).

Вычисленные по (II.55') значения $\bar{\varepsilon}_n^2$ представлены на рис. 19. Уже при $z_0 = 0,2r_f$ полученные результаты не хуже, чем при оптимальной фильтрации. В то же время ослабление (но не искажение!) полезного сигнала при этом незначительно.

Интересно отметить, что при определенных параметрах коэффициенты расчетных формул оптимальной фильтрации пересчета на высоту практически совпадают. При $z_0/\Delta x = 2,0$ $c_0 = 0,155$; $c_1 = 0,126$; $c_2 = 0,080$; $c_3 = 0,050$; $c_4 = 0,032$, что очень близко к приведенным выше коэффициентам оптимальной фильтрации. Во всяком случае применение их дает практически идентичные результаты. Однако значения, полученные с помощью оптимального фильтра, относящиеся к исходному условию $z=0$, тогда как на самом деле (исходя из близости значений c_j) они должны относиться к высоте $z_0 = 2\Delta x$. При опробовании на теоретической модели с соответствующими параметрами $\bar{\varepsilon}_n^2 = \pm 0,2$, а $\bar{\varepsilon}_n^2 = \pm 0,07$, если полученные результаты сравнивать с теоретическими значениями функции $f(x)$ на высоте z_0 (см. табл. 6). Мало того, если не учитывать изменение плоскости относимости аномалий, даже при идеальном проведении интерпретации ошибка определения глубины до источника составит $2\Delta x$, т. е. в данном случае 80% от истинного значения глубины. С этой точки зрения критерий верности оптимального фильтра в данном случае вообще непригоден.

Можно надеяться, что и в некоторых других случаях на основе использования специфических свойств геологических процессов будут найдены методы, обладающие значительными преимуществами перед рассмотренным здесь универсальным подходом, основанном на аппроксимации их моделью стационарного случайного процесса.

§ 3. Фильтрация случайных ошибок в пунктах нерегулярной сети наблюдений

В § 5 гл. I изложен метод, позволяющий производить интерполяцию на регулярную сеть с высокой точностью. В то же время там же было указано, что производить подбор аппроксимирующей конструкции при $\varepsilon_0 < |\varepsilon_n|$ (ε_n — среднеквадратическая ошибка измерений) не имеет смысла, поэтому используется модифицированный метод итерационного процесса (I.95), предложенный в работе [9]:

$$\alpha_i^{(k+1)} = \begin{cases} \alpha_i^{(k)} + \delta^{(k)}, & |\delta_i^{(k)}| > \varepsilon_0 \\ \alpha_i^{(k)}, & |\delta_i^{(k)}| < \varepsilon_0. \end{cases} \quad (\text{II.56})$$

Чтобы разобраться в том, как работает этот фильтр, рассмотрим итерационный процесс (I.92) более подробно. Запишем (I.92) в следующем виде:

$$\alpha^{(k+1)}(x_i, z_1) = \alpha^{(k)}(x_i, z_1) + \delta^{(k)}(x_i, z_1), \quad (\text{II.57})$$

где

$$\delta^{(k)}(x_i, z_1) = \varphi(x_i, 0) - \frac{z_1}{\pi} \int_{-\infty}^{\infty} \frac{\alpha^{(k)}(\xi, z_1)}{(x_i - \xi)^2 + z_1^2} d\xi. \quad (\text{II.58})$$

Поскольку правая часть формулы (II.58) есть интеграл Пуассона, ясно, что после первой итерации значения $\bar{\varphi}^{(1)}(x, z)$ равны значениям $\varphi(x, z)$, пересчитанным на высоту z_1^* . Если итерационный процесс продолжать, чтобы $\delta^{(k)} \approx 0$, функция $\bar{\varphi}(x, z)$ совпадает с исходной в точках наблюдений, а при $\delta^{(k)} = \varepsilon_0$ функция $\bar{\varphi}(x, z)$ относится к некоторой промежуточной поверхности. Таким образом, как уже указывалось в предыдущем параграфе, частотная характеристика пересчета на высоту определяется выражением $e^{-|\omega|z}$, высокие частоты, свойственные ошибкам наблюдений и непредставительным характеристикам геологических признаков, будут убывать быстрее, чем более низкие, присущие полезному сигналу.

Рассмотрим частотную характеристику итерационного процесса (I.92) [11, 16]. Переходя в (II.57) и (II.58) от функций-оригиналов к Фурье образам, после некоторых преобразований

* Конечно, если $\varphi(x)$ не является гармонической функцией, такой пересчет лишен физического смысла, однако в данном случае это не имеет никакого значения.

получим

$$\delta^{(k)}(\omega, z_1) = F(\omega, 0) [1 - e^{-|\omega| z_1}]^{k*}, \quad (\text{II.59})$$

$$\bar{F}^{(k)}(\omega, 0) = F(\omega, 0) [1 - (1 - e^{-|\omega| z_1})^k]. \quad (\text{II.60})$$

Из формулы (II.60) следует, что спектр аппроксимирующей функции $\bar{\varphi}(x, z)$ может быть получен по спектру исходной функции $\varphi(x, z)$ с помощью оператора

$$\Psi^{(k)}(\omega, z_1) = 1 - (1 - e^{-|\omega| z_1})^k. \quad (\text{II.61})$$

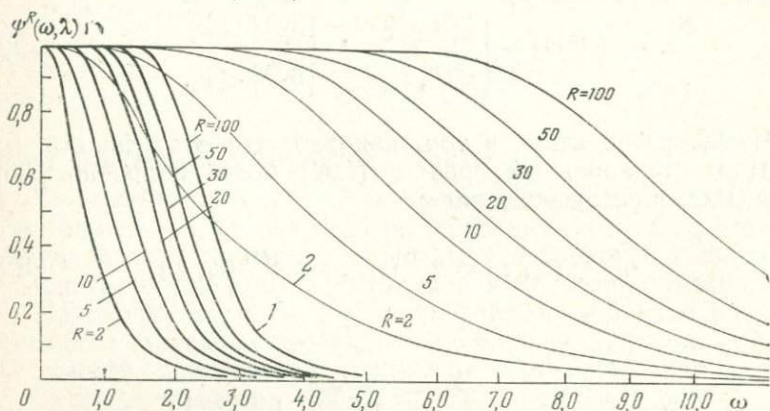


Рис. 20. Частотные характеристики итерационного подбора аппроксимирующей конструкции (I.79):
1 - $z_1/\Delta x=2,0$; 2 - $z_1/\Delta x=0,5$

Графики $\Psi^{(k)}$ представлены на рис. 20, из которого видно, что итерационный процесс (I.92) действует на исходную функцию, как двухпараметрический фильтр с последовательно расширяющейся при увеличении k и уменьшении $z_1/\Delta x$ полосой пропускания. В первую очередь пропускаются низкие частоты и лишь затем, по мере увеличения количества итераций, все более высокие. Ясно, что построение аппроксимирующей конструкции (I.79) для участков, характеризующихся большими размерами флюктуаций, большими величинами радиуса автокорреляции, требует меньшего количества итераций (при одинаковом ϵ_0). Наибольшего количества итераций потребуют, естественно, слу-

* Отсюда легко показать и сходимость итерационного процесса. При $k \rightarrow \infty$

$$\limsup_{k \rightarrow \infty} \limsup_{\omega} |\delta^{(k)}(\omega, z)| = 0,$$

следовательно,

$$\limsup_{k \rightarrow \infty} \limsup_x |\delta^{(k)}(x, z)| = 0.$$

чайные ошибки измерений. Однако заранее рассчитать и фиксировать значение k при практических расчетах нельзя. Это связано с тем, что необходимое число итераций в каждой i -ой точке для выполнения неравенства $|\delta^{(k)}| \leq \varepsilon_0$ может очень сильно различаться. Фильтр, основанный на простом ограничении числа k , будет, естественно, обладать плохими фильтрующими свойствами, и, кроме того, при этом будет нарушаться основное требование, необходимое для решения задачи интерполяции, $\bar{\varphi}(x_i) \approx \varphi(x_i)$. Учитывая, что значения

$$\Delta_i = \delta_i^{(k)} - \delta_i^{(k-1)}$$

обычно малы, так как итерационный процесс сходится сравнительно медленно, и переходя от процесса простой итерации (I.92) к процессу (II.56), будем иметь

$$\varepsilon_0 \approx |\delta^{(k)}(x_i, z)| = |\varphi(x_i, z) - \varphi^{(k)}(x_i, z)|. \quad (\text{II.62})$$

Пусть $\varepsilon_0 = |\varepsilon_n|$. Поскольку фильтр пропускает в первую очередь низкочастотные составляющие, полученные таким образом значения $\delta^{(k)}(x_i, z)$ будут обогащены высокочастотными составляющими, т. е. будут содержать в себе значительную часть случайных ошибок. Анализ получаемых значений $\delta_i^{(k)}$ полностью это подтверждает. Они обычно хорошо коррелируют со значениями $\xi(x_i)$. Их сумма всегда близка к нулю.

«Грубые» ошибки, существенно превышающие величину ε_n , при таком процессе сглаживания могут уменьшаться лишь на величину ε_0 , $\varepsilon_0 = \sqrt{\xi^2}$. Если увеличить значение ξ_0 до $(1,5-2,0) \varepsilon_n$, ошибки фильтрации при сравнительно большом уровне помех могут существенно уменьшиться. Однако при малых значениях ξ^2/\bar{f}^2 , когда ошибки интерполяции сравнимы по величине со случайными погрешностями, такое увеличение ε_0 может привести к появлению систематических ошибок. В тех случаях, когда отношения ξ^2/\bar{f}^2 и r_ξ/r_f малы, т. е. в тех случаях, когда в принципе фильтрация может быть весьма эффективной, необходимо увеличивать значение z_1 до $(1,5-2,0)\Delta x$. Очевидно, может быть развит принцип выбора ε_0 в зависимости от ξ^2/\bar{f}^2 и $H/\Delta x$. Однако при этом алгоритм (II.56) лишится универсальности, так как потребуются знать значения трудно определяемых параметров, претерпевающих значительные колебания. Поэтому при расчетах обычно принимается $\varepsilon_0 = |\varepsilon_n|$ — среднеквадратическая погрешность исходных данных с учетом ошибок интерполяции или $\varepsilon_0 = (1,5-2,0) |\varepsilon_n|$, когда требуется большая гладкость получаемых результатов.

При сглаживании графиков вручную «грубые ошибки», имеющие сравнительно небольшой радиус корреляции, легко отсеиваются. Этот процесс можно моделировать при машинной

обработке, дополнив алгоритм (II.56) следующим образом:

$$\alpha_i^{(k+1)} = \begin{cases} \alpha_i^{(k)} + \delta_i^{(k)}, & |\delta_i^{(k)}| > \varepsilon_0, \delta_i^{(k)} \in B \\ \alpha_i^{(k)}, & |\delta_i^{(k)}| > \varepsilon_0, \delta_i^{(k)} \notin B, \\ \alpha_i^{(k)}, & |\delta_i^{(k)}| < \varepsilon_0 \end{cases} \quad (\text{II.63})$$

где B — множество, состоящее из $2N+1$ последовательных значений $\delta_i^{(k)}$ одного знака.

Например, при $N=1$ значение $\delta_i^{(k)}$ будет прибавляться только тогда, когда $\delta_{i-1}^{(k)}$ и $\delta_{i+1}^{(k)}$ имеют тот же знак. При изменении значений N будет изменяться радиус автокорреляций аномалий

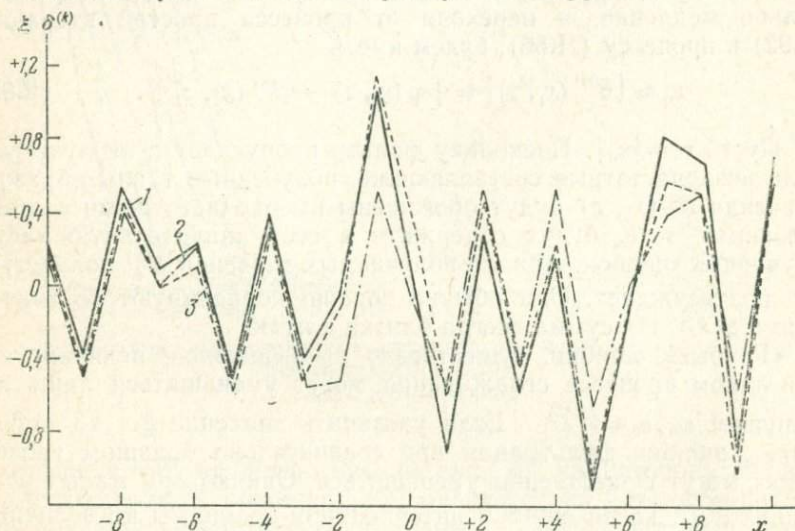


Рис. 21. Графики значений ξ_i и $\delta_i^{(k)}$:
1 — ξ_i ; 2 — $\delta_i^{(k)}$ по алгоритму (II.56), 3 — по алгоритму (II.63)

помех. В принципе, увеличив N достаточным образом, можно отфильтровать не только случайные ошибки, но и локальную составляющую полезного сигнала, выделяя тренд, и таким образом использовать этот алгоритм для проведения тренд-анализа.

Рассмотрим теоретический пример. Полезный сигнал обусловлен функцией $f(x) = U(x, H) = H/x^2 + H^2$, $H/\Delta x = 2,5$; $\xi^2/\bar{f} = 1,0$; $\bar{f}^2(-10\Delta x, 10\Delta x) = 0,6$. Значение r_ξ принято равным $1,5\Delta x$, $r_f = \pi H$, т. е. $r_\xi = 0,2r_f$ и, следовательно, для сравнения результатов можно воспользоваться коэффициентами оптимального фильтра Колмогорова — Винера, приведенными в предыдущем параграфе. Результаты расчетов представлены в табл. 6.

Результаты фильтрации случайных ошибок на теоретической модели

x_i	$f(x_i)$	ξ_i	$\varphi(x_i) = f(x_i) + \xi(x_i)$ исх.	$\hat{f}(x_i)$ (II. 38)	$\delta \hat{f}(x_i)$ (II. 38)	$f(x_i, z)$, $z = 2 \Delta x$	$\delta \hat{f}(x_i)$ (II. 38)	$\hat{f}(x_i)$ (II. 56)	$\delta_i^{(k)}$ (II. 56)	$\delta \hat{f}(x_i)$ (II. 56)	$\hat{f}(x_i)$ (II. 63)	$\delta_i^{(k)}$ (II. 63)	$\delta \hat{f}(x_i)$ (II. 63)
-10	0,06	+0,23	0,29	0,08	0,02	0,09	-0,01	0,09	0,20	0,03	0,09	0,20	0,03
-9	0,07	-0,43	-0,36	0,12	0,05	0,11	0,01	0,12	-0,48	0,05	0,13	-0,49	0,06
-8	0,09	0,53	0,62	0,18	0,09	0,13	0,05	0,17	0,45	0,08	0,19	0,43	0,10
-7	0,11	0,10	0,21	0,22	0,11	0,16	0,06	0,19	0,02	0,08	0,24	-0,03	0,13
-6	0,15	0,24	0,39	0,23	0,08	0,20	0,03	0,18	0,21	0,03	0,26	0,13	0,11
-5	0,20	-0,48	-0,28	0,26	0,06	0,25	0,01	0,19	-0,47	-0,01	0,30	-0,58	0,10
-4	0,28	0,41	0,69	0,32	0,04	0,31	0,01	0,30	0,39	0,02	0,39	0,30	0,11
-3	0,41	-0,41	0,00	0,41	0,00	0,38	0,03	0,57	-0,57	0,16	0,51	-0,51	0,10
-2	0,61	-0,05	0,56	0,52	-0,09	0,46	0,06	1,09	-0,53	0,48	0,69	-0,13	0,08
-1	0,86	1,10	1,96	0,60	-0,26	0,53	0,07	1,40	0,56	0,54	0,82	1,14	-0,04
0	1,00	0,09	1,09	0,56	-0,44	0,56	0,00	1,06	0,03	0,03	0,74	0,35	-0,26
1	0,86	-0,92	-0,06	0,45	-0,41	0,53	-0,08	0,51	-0,57	-0,35	0,56	-0,62	-0,30
2	0,61	0,29	0,90	0,35	-0,26	0,46	-0,11	0,36	0,54	-0,24	0,42	0,48	-0,19
3	0,41	-0,58	-0,17	0,25	-0,16	0,38	-0,13	0,24	-0,41	-0,17	0,30	-0,47	-0,11
4	0,28	0,17	0,45	0,17	-0,11	0,31	-0,14	-0,11	0,55	-0,39	0,20	0,25	-0,08
5	0,20	-1,15	-0,95	0,12	-0,08	0,25	-0,13	-0,41	-0,54	-0,61	0,15	-1,10	-0,05
6	0,15	-0,09	0,06	0,17	0,02	0,20	-0,03	0,06	0,00	-0,09	0,18	-0,12	0,03
7	0,11	0,82	0,93	0,23	0,12	0,16	0,07	0,54	0,39	0,43	0,24	0,69	0,13
8	0,09	0,70	0,79	0,22	0,13	0,13	0,09	0,28	0,51	0,19	0,23	0,56	0,14
9	0,07	-0,93	-0,86	0,17	0,10	0,11	0,06	-0,28	-0,58	-0,35	0,18	-1,04	0,11
10	0,06	0,77	0,82	0,16	0,10	0,09	0,07	0,26	0,56	0,20	0,17	0,65	0,11
					$\pm 0,02$		$\pm 0,07$			$\pm 0,28$			$\pm 0,13$

 $\sqrt{\delta \hat{f}^2}$

При применении фильтра (II.56) ($z_1=2,0$, $\epsilon_0=0,6$) среднеквадратическая ошибка составила 0,28, несмотря на то, что значения $\delta_i^{(k)}$ хорошо коррелируются со значениями $\xi(x_i)$ (рис. 21). Увеличение значения ϵ_0 до 1,0 (при нормальном распределении помехи в этот интервал входит примерно 90% значений ошибок) существенно уменьшает среднеквадратическую ошибку, делая ее равной ошибке оптимального фильтра $\pm 0,2$ мгл. Особенно хорошие результаты получаются при использовании фильтра (II.63): полученное значение среднеквадратической ошибки составило $\pm 0,13$ мгл. Здесь наблюдается очень хорошая корреляция значений $\delta_i^{(k)}$ и $\xi(x_i)$. Важно отметить, что систематические искажения существенны лишь для оптимальной фильтрации. Это легко видеть, сопоставляя значения $\sum_i \delta_i^{(k)}(x_i)/n$. Весьма важным обстоятельством является и то, что сумма значений $\delta_i^{(k)}$ при $\epsilon_0 \approx |\epsilon_n|$ всегда близка к нулю. Большое число примеров обобщения (II.56) приведено в работе [16].

Если наблюдения произведены с разной точностью и о величине ошибок имеются соответствующие данные, алгоритм (II.56) позволяет это учитывать. Для этого достаточно записать его в виде

$$\alpha_i^{(k+1)} = \begin{cases} \alpha_i^{(k)} + \delta_i^{(k)}, & |\delta_i^{(k)}| > \epsilon_{0,i} \\ \alpha_i^{(k)}, & |\delta_i^{(k)}| < \epsilon_{0,i} \end{cases}, \quad (\text{II.64})$$

где $\epsilon_{0,i}$ — теперь величина, характеризующая ошибку в каждом пункте наблюдений. Такая необходимость может возникнуть при составлении сводных карт по сейсмическим данным и данным бурения, при наблюдениях с приборами разной точности и т. п.

Сравнение результатов, получаемых на теоретических примерах, проведенное к тому же во многих случаях в более сложных, чем обычно, условиях (с учетом произвольных высот поверхности наблюдений) с теоретически возможными результатами применения оптимального фильтра показало их примерную эквивалентность. С практической точки зрения рассмотренный метод весьма эффективен. Он обеспечивает получение результатов примерно той же точности, что и оптимальный фильтр в идеальных условиях, однако в отличие от последнего может быть использован при произвольной сети пунктов и, что не менее важно, для его использования необходимо знать всего лишь одну, обычно известную величину — значение ϵ_n .

По-видимому, возможно дальнейшее улучшение алгоритмов (II.56) и (II.63), направленное на выделение оставшейся части случайной ошибки. Действительно, учитывая, что $\delta_i^{(k)} \approx \xi_i$, а среднее значение случайной помехи $\bar{\xi}$ близко к нулю, при использовании алгоритмов (II.56) и (II.63) будем изменять зна-

чение ε_0 до тех пор, пока не будет выполняться условие $\Sigma \delta_i^{(k)} = \bar{\delta} \approx 0$ на определенном, достаточно большом множестве точек. Значимость разности величины $\bar{\delta} - 0$ можно определять по критерию Стьюдента. Другой возможностью является поиск такого ε_0 , когда $\Sigma \delta_i^{(k)} = \min$. Это, очевидно, не только улучшит получаемые результаты, но и позволит проводить расчеты, задавая априори лишь некоторые границы изменения ε_0 , т. е. фактически без априорного определения амплитудных характеристик помехи. В этом случае параметры фильтров будут определяться автоматически — самим алгоритмом.

§ 4. Полиномиальные методы. Осреднение

В практических приложениях тренд-анализ очень часто проводится на основе известных формул точечной квадратической аппроксимации. Региональная компонента при этом заменяется полиномом $Q_m(x)$ заданной степени m , коэффициенты которого определяются по методу наименьших квадратов из условия

$$f(c_0, c_1, c_2, \dots, c_m) = \sqrt{\sum_{i=1}^n [\varphi(x_i) - Q_m(x_i)]^2} = \min. \quad (\text{II.65})$$

Основной вопрос при таком построении поверхностей тренда связан с выбором величины m . Здесь можно исходить из следующих соображений. Наиболее простым и обоснованным, но в то же время очень редко встречающимся в практике является тот случай, когда степень полинома известна из физико-геологических представлений о поведении признака $f(x)$. Например, известно, что полезная составляющая является полиномом второй степени, тогда наилучшим способом ее выделения будет построение полинома при $m=2$ по методу наименьших квадратов.

Как правило, степень полинома определяется предварительной глазомерной оценкой исходной карты. Если видно, что необходимо выделить линейный фон, то используется полином первой степени, при более сложной региональной картине — второй, третьей и более высокой степени. Окончательное заключение производится после построения карт тренда — по степени их соответствия той гипотезе, проверка которой производится с помощью тренд-анализа. Например, если предполагается, что региональное поведение геологического признака определяется некоторым типом тектонических элементов (которые уже выделены), то выбирается та карта, на которой они находят наилучшее отражение. Может ставиться и обратная задача — поиск элементов тектоники, если известно, что они должны находить отображение в поведении рассматриваемого признака.

Иногда считается, что степень полинома может быть найдена с помощью определения так называемой силы тренда, т. е. по той доле изменчивости, которую включает в себя карта тренда, по сравнению с общей изменчивостью. Сильным трендом считается тот, у которого дисперсия остатков (разностей между исходными значениями и трендом) достаточно мала [57]. Здесь уместно вспомнить, какая решается задача: если строится карта признака $\varphi(x)$, то отклонения действительно должны быть малы, согласуясь с ошибками измерений и непредставительными флуктуациями признака, если строится карта тренда, то какое значение может иметь величина остатков? Вообще неправильно делать заключения о наличии или отсутствии тренда на основании изучения соотношения дисперсий признака $\varphi(x)$, «тренда» $\hat{f}(x)$ и $\xi(x) = \varphi(x) - \hat{f}(x)$. Существование или отсутствие тренда — объективная реальность, а коэффициенты полиномов могут быть вычислены всегда, причем необходимые соотношения дисперсий можно подобрать с помощью формальных приемов — изменением степени полинома, вычислением различных разностей между полиномами степени m и $m+k$, использованием локальных приближений и т. п.

Таким образом, первостепенную роль здесь должна иметь глазомерная оценка результатов. Рассмотрим рис. 22, заимствованный из работы [57]. Прежде всего, легко видеть, что построение тренда по такому непредставительному материалу при отсутствии региональной картины процесса вряд ли имеет смысл. Достаточно было не провести измерения в небольшой северо-западной группе точек, чтобы линейный тренд исчез или, во всяком случае, стал очень слабым. Авторы на основании изучения дисперсий пришли к выводу о существовании тренда второго порядка, но по рис. 22 видно, что это совершенно необоснованное заключение. И уже совсем неправильно построение карт тренда в юго-западной и северо-восточной части участка, где нет наблюдений.

Приведем еще один пример необходимости предварительной глазомерной оценки. Допустим, что слева направо до середины участка явно виден линейный тренд с возрастанием значений; а затем они уменьшаются с той же скоростью. Если определить коэффициенты полинома первой степени сразу для всего участка, они будут близки к нулю. В то же время при предварительной глазомерной оценке сразу выяснилось бы, что необходимо сначала поделить участок на две части и определить тренд для каждой из них в отдельности.*

Очень большую роль в искажении получаемых результатов играет неравномерность расположения значений $\varphi(x_i)$. Это следует из самой сути метода наименьших квадратов, поэтому при

* В принципе, все ситуации могут быть перечислены и их анализ можно переложить на ЭВМ. Однако это дело будущего.

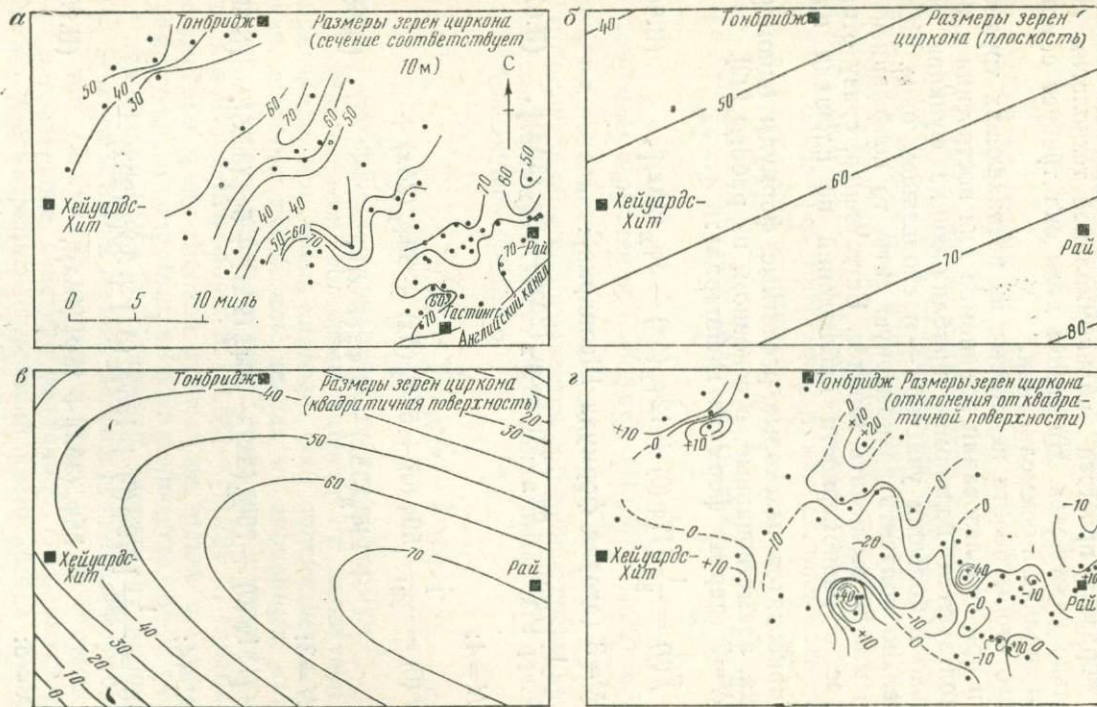


Рис. 22. Карты тренда размеров зерен циркона:

а — карта размеров зерен циркона; б — плоскость тренда первого порядка; в — поверхность тренда второго порядка; г — карта отклонений второго порядка

полиномиальном способе определения тренда необходимо отбрасывать часть наблюдений, делая их сеть близкой к равномерной. Столь же сильно метод наименьших квадратов реагирует на отдельные (большие) «выбросы» значений $\varphi(x_i)$, если число наблюдений невелико.

Таким образом, простота полиномиальной технологии является только кажущейся. На самом деле она требует самого детального анализа и осмысливания.

Для того чтобы обойти некоторые из возникающих трудностей, как и при использовании полиномов для построения карт, часто используются полиномы, приближающие функцию на сравнительно небольших участках — в скользящем окне. Если предварительно получить равномерную сеть, то коэффициенты полиномов можно вычислить заранее. Естественно, сразу отпадает вопрос о равномерности наблюдений по площади или профилю.

Ниже приводятся некоторые расчетные формулы одномерного тренда, заимствованные, в основном, из работы [22].

$m=3, N=2$ (первая формула Шеппарда):

$$\hat{f}(0) = \frac{1}{35} [17\varphi(0) + 12\varphi_{\Sigma}(\Delta x) - 3\varphi_{\Sigma}(2\Delta x)], \quad (\text{II.66})$$

$m=3, N=3$ (вторая формула Шеппарда):

$$\hat{f}(0) = \frac{1}{21} [7\varphi(0) + 6\varphi_{\Sigma}(\Delta x) + 3\varphi_{\Sigma}(2\Delta x) - 2\varphi_{\Sigma}(3\Delta x)], \quad (\text{II.67})$$

$m=3, N=4$:

$$\hat{f}(0) = \frac{1}{231} [59\varphi(0) + 54\varphi_{\Sigma}(\Delta x) + 39\varphi_{\Sigma}(2\Delta x) + 14\varphi_{\Sigma}(3\Delta x) - 21\varphi_{\Sigma}(4\Delta x)], \quad (\text{II.68})$$

$m=5, N=3$:

$$\hat{f}(0) = \frac{1}{231} [131\varphi(0) + 75\varphi_{\Sigma}(\Delta x) - 30\varphi_{\Sigma}(2\Delta x) + 5\varphi_{\Sigma}(3\Delta x)], \quad (\text{II.69})$$

$m=5, N=4$:

$$\hat{f}(0) = \frac{1}{429} [179\varphi(0) + 135\varphi_{\Sigma}(\Delta x) + 30\varphi_{\Sigma}(2\Delta x) - 55\varphi_{\Sigma}(3\Delta x) + 15\varphi_{\Sigma}(4\Delta x)], \quad (\text{II.70})$$

$m=5, N=5$:

$$\hat{f}(0) = \frac{1}{429} [143\varphi(0) + 120\varphi_{\Sigma}(\Delta x) + 60\varphi_{\Sigma}(2\Delta x) - 10\varphi_{\Sigma}(3\Delta x) - 45\varphi_{\Sigma}(4\Delta x) + 18\varphi_{\Sigma}(5\Delta x)], \quad (\text{II.71})$$

$N=10$ (формула Спенсера):

$$\begin{aligned} \hat{f}(0) = \frac{1}{350} [60\varphi(0) + 57\varphi_{\Sigma}(\Delta x) + 47\varphi_{\Sigma}(2\Delta x) + 33\varphi_{\Sigma}(3\Delta x) + \\ + 18\varphi_{\Sigma}(4\Delta x) + 6\varphi_{\Sigma}(5\Delta x) - 2\varphi_{\Sigma}(6\Delta x) - 5\varphi_{\Sigma}(7\Delta x) - \\ - 5\varphi_{\Sigma}(8\Delta x) - 3\varphi_{\Sigma}(9\Delta x) - \varphi_{\Sigma}(10\Delta x), \end{aligned} \quad (\text{II.72})$$

где $\varphi_{\Sigma}(j\Delta x) = \varphi(-j\Delta x) + \varphi(j\Delta x)$,

m — порядок аппроксимирующего полинома, $2N+1$ — число точек, участвующих в сглаживании. Начало координат $x=0$ совмещено с точкой, в которой производятся расчеты. Наряду с этими формулами при выделении региональных компонент широко применяется метод осреднения (скользящего среднего), который можно рассматривать и как полиномиальное приближение при $m=1$. В этом случае

$$\hat{f}(0) = \frac{1}{2N+1} [\varphi_0 + \varphi_{\Sigma}(\Delta x) + \varphi_{\Sigma}(2\Delta x) + \dots + \varphi_{\Sigma}(N\Delta x)]. \quad (\text{II.73})$$

С помощью осреднения и формул (II.67) — (II.72) может выделяться и локальная составляющая $-\xi(x) = \varphi(x) - \hat{f}(x)$, которая в этом случае обычно называется остаточной составляющей.

Все эти формулы могут быть записаны в виде формулы весового скользящего среднего.

$$\hat{f}(x) = \sum_{-N}^N c_j \varphi(x - j\Delta x), \quad (\text{II.74})$$

где c_j — коэффициенты соответствующих формул.

Выбор степени полинома зависит от характера функций, выделяемых при тренд-анализе. Она определяется фактически лишь опытным путем или исходя из априорных, как правило, необоснованных предположений. Кроме того, следует подчеркнуть, что коэффициенты формул (II.66) — (II.72) и, следовательно, и результаты тренд-анализа зависят не только от степени аппроксимирующего полинома, но и от количества точек, участвующих в сглаживании $(2N+1)$. Например, «короткий» полином первого или второго порядка иногда меньше сглаживает исходную функцию, чем «длинный» полином более высокой степени.

Сравнительное действие различных сглаживающих формул удобно рассматривать с позиций частотного анализа, так как в этом случае имеется возможность непосредственно сравнивать между собой операторы преобразования. Переходя от функций-оригиналов к образам Фурье, вместо формулы (II.17) получим

$$\hat{f}(\omega) = \psi(\omega) \varphi(\omega), \quad (\text{II.75})$$

где $\widehat{f}(\omega)$ и $\varphi(\omega)$ — соответственно преобразования Фурье функций $\widehat{f}(x)$ и $\varphi(x)$, $\psi(\omega)$ — частотная характеристика оператора L . В двухмерном варианте

$$\widehat{f}(u, V) = \psi(U, V) \varphi(U, V). \quad (\text{II.76})$$

Для формул типа (II.74) частотные характеристики операторов представляют собой тригонометрические суммы

$$\psi(\omega) = \sum_{-N}^N c_j e^{-i\omega j \Delta x}, \quad (\text{II.77})$$

$$\psi(U, V) = \sum_{-N_1}^{N_1} \sum_{-N_2}^{N_2} c_{k,j} e^{-i(Uk\Delta x + Vj\Delta x)}. \quad (\text{II.78})$$

Графики $\psi(\omega)$ для различных сглаживающих формул представлены на рис. 23 [10]. Из рис. 23 видно, что аппроксимирующий полином при $m=5$, $N=5$ обрезает высокие частоты сильнее, чем, например, полином более низкой степени $m=3$, но с меньшими значениями N , $N=2$, и примерно одинаково с полиномом $m=3$, $N=3$.

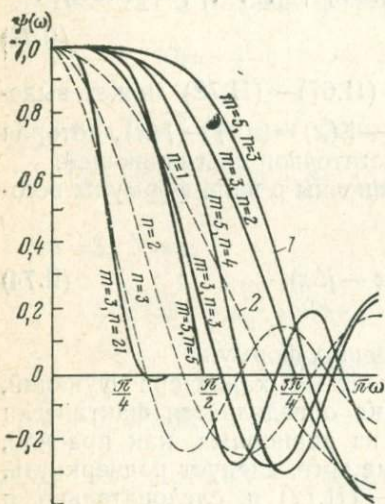


Рис. 23. Частотные характеристики операторов полиномиального тренд-анализа:

1 — для формул (II.66)–(II.72); 2 — для формулы (II.73) при различных значениях N

применение формулы Спенсера должно приводить к результатам, занимающим промежуточное положение между значениями, которые получаются при осреднении по 5-й и 7-й точкам. Такая сложная зависимость, а главное отсутствие соответствующего теоретического подхода, делает предварительное определение параметров полиномов по крайней мере неэффективным.

Метод осреднения, несмотря на свою очевидную простоту, более конструктивен,

поскольку результаты его применения и, что очень важно, их зависимость от параметра N гораздо легче предусмотреть. Интервал $(-R, R)$ (квадрат со стороной $2R$, круг радиусом R) следует выбирать таким образом, чтобы региональная (полезная) компонента искажалась сравнительно слабо при достаточно сильном подавлении локальной компоненты и

ошибок наблюдений

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2N+1} \sum_{j=-N}^N \varphi(x-j\Delta x) = \frac{1}{2N+1} \sum_{j=-N}^N [f(x-j\Delta x) + \\ &+ \xi(x-j\Delta x)] = \frac{1}{2N+1} \sum_{j=-N}^N f(x-j\Delta x) + \frac{1}{2N+1} \sum_{j=-N}^N \xi(x-j\Delta x), \end{aligned} \quad (\text{II.79})$$

где $\xi(x-j\Delta x)$ — некоррелированные значения помехи.

Если считать, что $f(x-j\Delta x) = \text{const} = f$ на интервале $(-R, R)$, то изменение величины отношения сигнал/помеха выражается следующим образом:

$$\rho_{\Phi} = \frac{D\left(\sum_{j=-N}^N f_j\right)}{D\left(\sum_{j=-N}^N \xi_j\right)} = \frac{(2N+1)^2 f^2}{\sum_{j=-N}^N D\xi_j} = 2(N+1) \frac{f^2}{\bar{\xi}^2} = K^2 \rho_0, \quad (\text{II.80})$$

где $\rho_0 = \bar{f}^2 / \bar{\xi}^2$ — отношение сигнал/помеха перед осреднением. Таким образом, отношение сигнал/помеха увеличивается в $2N+1$ раз. Такой прием называется методом накопления. Действительно, значение $\sum f_j$ увеличивается пропорционально N , а сумма $\sum \xi_j$ практически не изменяется и соответственно величина $\sum f_j / N = f$, а $\sum \xi_j / N$ быстро убывает.

Если значения ξ_i коррелированы, то величина K^2 не равна числу объектов. В этом случае она определяется из выражения

$$K^2 = \frac{2R}{\tau_0} \quad (\text{II.81})$$

в одномерном случае и

$$K^2 = \frac{4R^2}{\tau_0^2} \quad (\text{II.82})$$

в двухмерном. Отсюда ясно, что простое увеличение числа слагаемых в формуле (II.73) за счет уменьшения шага Δx с помощью интерполирования не имеет смысла. Естественно, чем больше R , тем больше K^2 и тем лучше подавляются локальные неоднородности и ошибки наблюдений. Однако на самом деле $\hat{f}(x) \neq \text{const}$, поэтому при увеличении радиуса осреднения увеличиваются систематические искажения полезного сигнала. Вместо значения $f(x)$ получается среднее значение $\bar{f}(x)$ на интервале $(-R, R)$:

$$\bar{f}(x) = \frac{1}{2N+1} \sum_{j=-N}^N f(x-j\Delta x), \quad \bar{\xi}^2 = 0. \quad (\text{II.83})$$

Допустимая величина $\delta = f(x) - \overline{f(x)}$, которая приближенно может быть легко определена, в сравнении со степенью подавления помехи и величиной $\widehat{\xi}$ фактически и определяет R .

Соотношение δ и $\widehat{\xi}$ может изменяться в зависимости от необходимой гладкости региональной компоненты $\overline{f(x)}$ при решении дальнейших задач. Если положить в основу обычный критерий минимума среднеквадратической ошибки (II.18), то значение δ надо выбирать таким, чтобы

$$\overline{\delta^2} + D\widehat{\xi} = \overline{\xi_\Phi^2} = \min. \quad (\text{II.84})$$

Оценив значения $\overline{\delta^2}$ и $D\widehat{\xi}$ для различных R , легко выбрать приближенную величину R , удовлетворяющую формуле (II.84). Особой точности здесь не требуется, так как, во-первых, небольшие колебания R сказываются не слишком сильно, а во-вторых, ∞ равно значения r_f и r_ξ всегда испытывают существенные

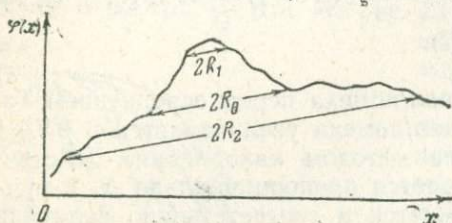


Рис. 24. Выбор величины R при выделении локальной компоненты

колебания. Необходимо ориентироваться на средние значения последних.

Локальную компоненту можно выделить, вычислив разность суммарного сигнала и региональной компоненты, полученной осреднением. Существуют и более эффективные способы, позволяющие выделять

непосредственно локальную компоненту. Они основаны на вычислении разности

$$\widehat{f}(x) = \varphi(x) - \frac{\varphi(x+R) + \varphi(x-R)}{2} \quad (\text{II.85})$$

или

$$\widehat{f}(x, y) = \varphi(x, y) - \overline{\varphi}(R),$$

где $\overline{\varphi}(R)$ — среднее значение функции $\varphi(x, y)$ на окружности радиуса R или по периметру квадрата со стороной $2R$, расстояние сторон которого от точки (x, y) равно R . И здесь, конечно, основную роль играет выбор значения R . Рассмотрим его определение на примере.

Пусть локальный максимум на фоне регионального, изображенный на рис. 24, соответствует интересующей нас структуре третьего порядка, а мелкая волнистость — ошибкам измерения и мелким неоднородностям. Уменьшение радиуса до R_1 приведет к подчеркиванию мелких особенностей исходной кривой, которые в данном случае не представляют интереса. При увеличении радиуса до R_2 значения $\widehat{f}(x)$ будут содержать значительную

долю региональной составляющей. Наиболее подходящий радиус равен приблизительно половине ширины выделяемой структуры, т. е. величина R приблизительно равна радиусу выделяемых особенностей признака или радиусу автокорреляции. Следует отметить, что истинные значения амплитуд, особенно при неудачном задании R , сильно искажаются, поэтому получаемые результаты носят преимущественно лишь качественный характер.

При выделении локальной компоненты по формуле (II.85) могут существенно сказываться случайные ошибки наблюдений и непредставительные компоненты поля. Для уменьшения их влияния можно производить расчеты по следующим формулам:

$$\hat{f}(x) = \bar{\varphi}(R_1) - \bar{\varphi}(R_2), \quad (\text{II.86})$$

где $\varphi(R_2)$ имеет тот же смысл, что и $\bar{\varphi}(R)$ в (II.85), а $\varphi(R_1)$ — среднее значение по ближайшим к (x, y) точкам, обычно

$$\bar{\varphi}(R_1) = 0,5 [\varphi(x - \Delta x/2) + \varphi(x + \Delta x/2)],$$

в одномерном случае и

$$\bar{\varphi}(R_1) = 0,25 [\varphi(x - \Delta x/2, y) + \varphi(x + \Delta x/2, y) + \\ + \varphi(x, y - \Delta y/2) + \varphi(x, y + \Delta y/2)]$$

в двумерном.

Если карты имеют четко выраженное простирание, для повышения эффективности формул можно производить осреднение не в квадрате, а в прямоугольнике, ориентируя его по простиранию. Соотношение сторон должно быть примерно равно соотношению вытянутой и короткой осей выделяемой полезной компоненты.

К определению радиуса R при осреднении можно подойти и с более формализованных позиций, но опять-таки это можно сделать лишь в тех случаях, когда известны параметры процессов $f(x)$ и $\xi(x)$. Можно, например, исходить из того, чтобы частотная характеристика осреднения при заданном значении R была в определенном смысле близка к частотной характеристике оптимального фильтра, рассчитанной по формуле (II.31) для имеющихся значений \bar{f}^2 , $\bar{\xi}^2$, r_f , r_ξ и выбранной модели автокорреляционной функции $B(\tau)$. Сравнение удобно произвести графическим способом, как это делается для построения вычислительных конструкций в работах [33, 83].

Формулы типа (II.83), (II.85) и (II.86) из-за простоты и эффективности получили большое распространение при расчетах вручную и на ЭВМ, причем первые программы с их использованием в СССР были составлены еще в 1959 г. [78]. Естественно, в настоящее время такого рода программы считаются очень простыми. В работе [10] приведены следующие формулы

для расчетов на ЭВМ, реализованные в соответствующей программе для БЭСМ-4:

$$\widehat{f}_{\text{пер}}(i\Delta x) = \frac{1}{2N+1} \sum_{j=-N}^N \varphi[(i-j)\Delta x], \quad (\text{II.87})$$

$$\widehat{f}_{\text{лок}}(i\Delta x) = \varphi(i\Delta x) - \widehat{f}_{\text{пер}}(i\Delta x), \quad (\text{II.88})$$

$$\widehat{f}_{\text{пер}}(i\Delta x, k\Delta x) = \frac{1}{(2N+1)^2} \sum_{j=-N}^N \sum_{q=-N}^N \varphi[(i-j)\Delta x, (k-q)\Delta x], \quad (\text{II.89})$$

$$\widehat{f}_{\text{лок}}(i\Delta x, k\Delta x) = \varphi(i\Delta x, k\Delta x) - \widehat{f}_{\text{пер}}(i\Delta x, k\Delta x), \quad (\text{II.90})$$

а также

$$f'_{\text{пер}}(i\Delta x) = \frac{\varphi[(i-j)\Delta x] + \varphi[(i+j)\Delta x]}{2}, \quad (\text{II.91})$$

$$\widehat{f}'_{\text{лок}}(i\Delta x) = \varphi(i\Delta x) - \widehat{f}'_{\text{пер}}(i\Delta x), \quad (\text{II.92})$$

$$\widehat{f}'_{\text{пер}}(i\Delta x, k\Delta x) = \frac{1}{8N} [\widehat{f}'_{\text{пер}}(N) - \widehat{f}'_{\text{пер}}(N-1)], \quad (\text{II.93})$$

$$\widehat{f}'_{\text{лок}}(i\Delta x, k\Delta x) = \varphi(i\Delta x, k\Delta x) - \widehat{f}'_{\text{пер}}(i\Delta x, k\Delta x), \quad (\text{II.94})$$

где Δx — шаг расчетной сети, $N=1 \div 10$, $i=0, 1, 2, \dots, 150$, $k=0, 1, 2, \dots$

В формуле (II.93) производится определение значений $\varphi(x, y)$ по периметру квадрата со стороной $2N+1$. Осредненное значение относится к центру квадрата.

Для выделения региональной составляющей используются формулы (II.87), (II.89), (II.91) и (II.93). Локальные компоненты вычисляются при помощи формул (II.88), (II.90), (II.92) и (II.94).

Остановимся теперь более подробно на оценке сравнительной эффективности методов тренд-анализа, воспользовавшись для этого результатами, полученными в работе [39]. Оценка проведена на основе расчетов по формулам (II.29) и (II.48) для трех наиболее распространенных типов автокорреляционных функций (I.26), (I.27) и (I.28). Как и ранее, обозначим $\bar{\varepsilon}^2$ средний квадрат ошибки оценки полезного сигнала $f(x)$ после применения фильтра

$$\rho_0 = \bar{f}^2 / \bar{\varepsilon}^2, \quad \rho_\Phi = \bar{f}^2 / \bar{\varepsilon}^2, \quad K^2 = \bar{\varepsilon}^2 / \bar{\varepsilon}^2.$$

Результаты вычисления наибольших значений K^2 среди формул (II.73) для различных значений N ($N \leq 10$) и (II.66) — (II.71) представлены в табл. 7. Ломаная линия указывает границу эффективности между этими формулами: слева от нее область,

Значения наибольших коэффициентов усиления K^2 формулы (II. 73) для различных значений N и формул (II. 66)–(II. 71)

r_1/r_2	\bar{f}^2/ξ^2											
	0.2	0.4	0.6	0.8	1.0	2.0	3.0	4.0	5.0	10.0	15.0	20.0
Модель (I. 26)												
2	3,66	2,40	1,89	1,70	1,59	1,20	1,18	1,14	1,10	—	—	—
4	4,36	3,13	2,68	2,45	2,30	1,95	1,80	1,67	1,56	1,41	1,34	1,29
6	5,06	3,88	3,39	3,11	2,91	2,47	2,27	2,10	2,04	1,81	1,68	1,62
8	5,86	4,58	4,03	3,71	3,49	2,95	2,70	2,54	2,45	2,11	2,01	1,92
10	6,21	5,29	4,62	4,27	4,03	3,42	3,09	2,95	2,81	2,47	2,28	2,13
Модель (I. 27)												
2	3,98	2,45	2,15	1,95	1,78	1,47	1,38	1,35	1,33	1,22	1,18	1,15
4	4,83	3,80	3,38	3,17	2,98	2,57	2,37	2,19	2,14	1,92	1,82	1,79
6	6,17	5,02	4,50	4,22	4,02	3,44	3,13	2,89	2,76	2,49	2,40	2,32
8	7,46	6,17	5,59	5,21	4,95	4,24	3,85	3,62	3,50	2,98	2,76	2,66
10	8,62	7,25	6,58	6,17	5,85	5,02	4,54	4,33	4,13	3,57	3,32	3,10
Модель (I. 28)												
2	3,89	2,41	1,84	1,57	1,26	1,02	—	—	—	—	—	—
4	4,59	3,01	2,35	2,07	1,86	1,34	1,10	—	—	—	—	—
6	5,21	3,56	2,86	2,48	2,22	1,61	1,33	1,14	1,07	—	—	—
8	5,74	4,03	3,26	2,82	2,52	1,81	1,52	1,32	1,17	—	—	—
10	6,17	4,44	3,53	3,12	2,80	2,01	1,68	1,48	1,32	1,01	—	—

Примечание. В таблице указаны границы эффективности между формулами скользящего среднего и полиномиальной аппроксимации; значения $K^2 < 1$ опущены.

где эффективнее формула скользящего среднего (II.73). Проверки соответствуют значениям $K^2 < 1$. Как видно из табл. 7, формула (II.73) является наилучшей (при правильном выборе N) среди рассматриваемых практически во всех реальных ситуациях. Полиномиальные формулы имеют преимущество лишь в тех случаях, когда значения K^2 малы и фильтрация мало эффективна. Таким образом, подтверждается вывод, сделанный ранее на основе рассмотрения частотных характеристик этих формул, о нецелесообразности применения полиномиальных приближений в тренд-анализе.

Интересно сравнить значения K^2 рассмотренных формул и оптимального фильтра, значения $K_{\text{опт}}^2$ которого представлены в табл. 8. Легко видеть, что оптимальный фильтр Колмогорова — Винера не имеет существенных преимуществ перед традиционными методами. Это особенно справедливо при $\rho_0 > 1$, $\alpha \gg 1$. Например, для модели (I.27) при $\alpha = 6$, $\rho_0 = 2$, $K_{\text{опт}}^2 = 3,67$, а $K_{\text{оср}}^2 = 3,44$; при $\alpha = 4$, $\rho_0 = 1$, $K_{\text{опт}}^2 = 3,40$, а $K_{\text{оср}}^2 = 2,98$, т. е. $K_{\text{опт}}/K_{\text{оср}} = 1,06$, и т. п. Ясно, что такие различия практически заметить нельзя. Важно, что значения $K_{\text{опт}}^2$ сильно зависят от выбора модели корреляционной функции. Как это видно из табл. 7, они изменяются в два раза и более. В то же время, как уже отмечалось выше, для обоснованного выбора модели обычно нет соответствующих данных. Фактически всегда можно выбрать любую из них. Даже при подготовке теоретических примеров для опробования фильтров из-за возникающих серьезных затруднений на это обычно не обращается никакого внимания и задаются только параметрами \bar{f}^2 , $\bar{\xi}^2$, r_f , r_ξ , поэтому коэффициенты усиления, представленные в табл. 8 и на рис. 19, могут быть лишь приближенным ориентиром. Из изложенного выше следует, что при надлежащем выборе параметра $R(N)$ метод скользящего среднего позволяет проводить тренд-анализ с точностью, близкой к оптимально возможной (в смысле среднеквадратического критерия).

В заключение необходимо отметить следующее. Несмотря на значительную математизацию формализованных методов тренд-анализа, в ряде случаев более удобным, а главное более надежным является графический метод, основанный на субъективной глазомерной оценке. При этом так называемая «субъективность исследователя» на самом деле может быть весьма полезной (если, конечно, исследователь обладает достаточными знаниями и опытом), так как он может использовать некоторые дополнительные сведения, не поддающиеся пока формализации. Рассмотрим, например, рис. 25, на котором идеализированно изображен рост локальной структуры на горизонтальной (А) и наклонной (Б) поверхностях. При удалении регионального фона с помощью формул типа (II.85) форма локальной структуры искажается. В то же время при графическом сглаживании

Значения коэффициентов усиления K^2 оптимального фильтра

r_f/r_ξ	\bar{r}^2/ξ^2											
	0,2	0,4	0,6	0,8	1,0	2,0	3,0	4,0	5,0	10,0	15,0	20,0

Модель (I. 26)

2	6,30	3,77	2,92	2,48	2,22	1,67	1,48	1,38	1,32	1,18	1,13	1,10
4	7,12	4,51	3,60	3,13	2,84	2,20	1,95	1,82	1,73	1,52	1,43	1,39
6	7,91	5,22	4,26	3,75	3,43	2,70	2,42	2,25	2,14	1,89	1,77	1,70
8	8,67	5,89	4,88	4,33	3,98	3,18	2,86	2,67	2,54	2,24	2,09	2,00
10	9,40	6,53	5,47	4,89	4,51	3,64	3,28	3,08	2,94	2,57	2,41	2,31

Модель (I. 27)

2	6,47	3,94	3,08	2,65	2,38	1,82	1,62	1,52	1,45	1,30	1,24	1,20
4	7,70	5,09	4,18	3,71	3,40	2,75	2,50	2,35	2,25	2,03	1,93	1,86
6	8,91	6,22	5,25	4,74	4,41	3,67	3,36	3,19	3,07	2,79	2,66	2,58
8	10,09	7,31	6,29	5,74	5,38	4,56	4,22	4,02	3,88	3,55	3,39	3,29
10	11,24	8,37	7,30	6,71	6,33	5,43	5,04	4,82	4,74	4,30	4,11	3,99

Модель (I. 28)

2	6,20	3,67	2,82	2,38	2,12	1,58	1,39	1,30	1,24	1,12	1,08	1,06
4	6,87	4,23	3,30	2,81	2,50	1,84	1,59	1,46	1,37	1,20	1,13	1,10
6	7,53	4,76	3,75	3,20	2,86	2,08	1,78	1,61	1,51	1,28	1,19	1,14
8	8,16	5,25	4,16	3,57	3,18	2,30	1,95	1,76	1,63	1,35	1,24	1,19
10	8,75	5,70	4,54	3,90	3,47	2,51	2,12	1,89	1,75	1,42	1,29	1,23

она сохраняется полностью. То же самое могло быть достигнуто при очень большом радиусе осреднения R . Однако ясно, что в реальных условиях, при наличии «боковых» влияний, это невыполнимо. Интересно отметить, что после снятия регионального фона своды структур уже не смещаются относительно друг друга.

Особенно важной является интуитивная способность исследователя учитывать колебания радиусов автокорреляции и амплитуд полезного сигнала и помехи, которые всегда имеют

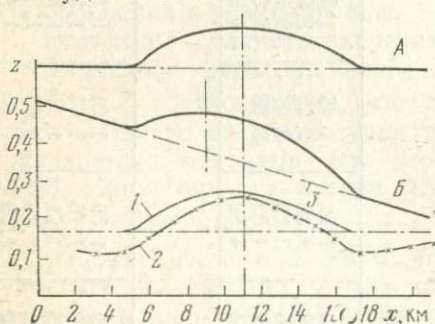


Рис. 25. Пример выделения локальной компоненты:

1 — графическим способом; 2 — с помощью формулы (II.85), $R=6$ км; 3 — региональный фон

место. При формальном подходе приходится строить несколько карт (графиков), применяя формулы с различными значениями R . Истолкование этих карт часто оказывается сложным и неудобным, а разбраковка элементов, выделяемых при помощи той или иной формулы, представляется не менее субъективной, чем графическое сглаживание.

Приведем практический пример [26]. По карте аномалий силы тяжести необходимо было

выделить локальные аномалии, сняв нелинейный региональный фон. Для этого были использованы формулы типа (II.85) и (II.86) при различных значениях R , однако удовлетворительных результатов (совпадения локальных аномалий и структур, выделяемых по данным сейсморазведки) получить не удалось. Конечно, возможно, что значения R были выбраны неудачно или что полученная серия карт была неудачно проинтерпретирована. Однако все равно это подтверждает вывод о неудовлетворительности и сложности такого подхода к выделению локальной составляющей в данном случае. В то же время опробованный авторами работы [26] графический способ разделения аномалий (рис. 26) дал очень хорошие результаты. Из рис. 26 видно, что фактически региональная компонента определялась минимальными значениями поля, исходя из ряда соображений: коррелируемости от профиля к профилю, градиентов поля, размеров локальных аномалий, кроме того, разнообразие размеров локальных аномалий не имеет при графическом сглаживании столь существенного значения.

Математические методы одновременного учета такого рода особенностей пока не разработаны, поэтому задачей дальнейших исследований является создание и алгоритмизация методов тренд-анализа, моделирующих эвристический подход и в то же

время достаточно общих и строгих с математической точки зрения. Совершенно необходим учет геологической специфики, особенностей каждой решаемой задачи, а не только общих свойств геологических признаков, отображаемых математиче-

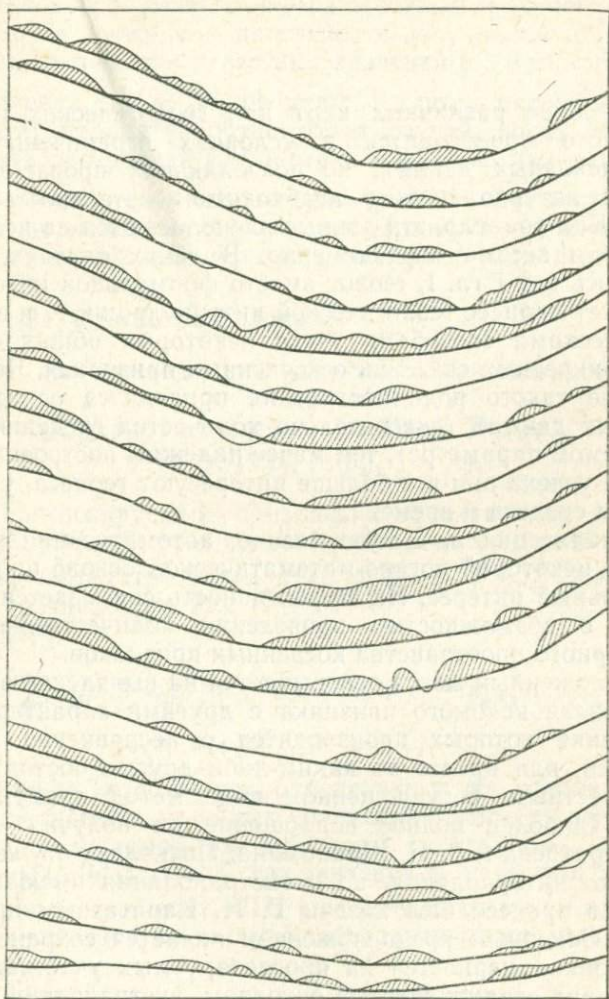


Рис. 26. Пример графического способа построения тренда

скими классами функций. Разработка таких методов позволит существенно расширить круг задач, решаемых тренд-анализом, и сделает его результаты гораздо более достоверными. В то же время не менее насущной задачей является определение круга задач, определение области применимости и внедрение в практику тех методов, которые уже разработаны.

Построение различных карт при геологических исследованиях часто производится в условиях ограниченного количества исходных данных, не позволяющих производить формальную интерполяцию с необходимой детальностью и точностью. Как правило, эти данные располагаются на исследуемой территории весьма неравномерно. В таких случаях, как уже отмечалось в § 1 гл. I, геолог вместо формальной интерполяции использует процесс геологической интерполяции, т. е. пользуется эвристическими способами учета некоторых общих соображений и конкретных сведений о косвенных признаках. Чем в большей мере такого рода построения приходится основывать на косвенных данных (относительно количества сведений о самом картируемом параметре), тем менее надежны построения, однако в случае успеха они все больше интересуют геолога, увеличивая выигрыш средств и времени.

Формализация и, соответственно, автоматизация этого процесса на некоторой логико-математической основе представляет значительный интерес. Их эффективность связывается в первую очередь с возможностью проведения количественного учета многомерного пространства косвенных признаков.

Предложенный метод основывается на выявлении закономерностей связи искомого признака с другими характеристиками, определение которых производится с несравненно меньшими затратами, или просто по каким-либо другим обстоятельствам, уже известным. В упрощенном виде метод применялся уже давно. Наиболее полное воплощение он получил в работах Г. И. Каратаева и В. И. Шрайбмана. Поскольку осуществляется не только интерполяция, но и экстраполяция имеющихся данных, этот процесс был назван Г. И. Каратаевым прогнозированием. Основным предположением является сохранение выявленных закономерностей на прогнозируемых участках. Именно этот момент ограничивает, в основном, экстраполяционные возможности метода.

Составляются уравнения вида

$$\hat{\varphi} = f(p_1, p_2, \dots, p_j, \dots, p_m),$$

где φ — прогнозируемый признак, p_j — косвенные геолого-геофизические признаки, используемые для прогноза. Например, при построении структурных карт косвенными признаками могут служить данные об отметках поверхности рельефа Земли, уже

известных структурных поверхностей, аномалии силы тяжести, магнитного поля, некоторые функции (трансформанты) от них и т. п.

Можно рассматривать два варианта решения задачи. Первый, прогнозный вариант является основным и относится к тому случаю, когда точность и детальность получаемой карты существенно ниже точности известных значений φ_i , т. е. когда совпадение значений φ_i и $\widehat{\varphi}_i$ не требуется. Второму, интерполяционно-прогнозному варианту соответствует случай, когда точность и детальность построений сравнима с точностью измеренных значений φ_i , иными словами, модуль разности $|\widehat{\varphi}_i - \varphi_i|$ должен быть меньше определенной, заранее заданной величины. Интерполяционно-прогнозный вариант рассмотрен лишь в работе [17].

§ 1. Прогнозный вариант

Обычно прогноз ведется с помощью полиномов вида

$$\widehat{\varphi} = Q(p_1, p_2, \dots, p_j, \dots, f^1(p_1), f^2(p_1), \dots, \dots, f^r(p_1) f^1(p_2), \dots), \quad (\text{III.1})$$

где φ — прогнозируемый признак, измеренный на множестве точек $\{(x_i^3, y_i^3)\} = T^3$, p_j — геолого-геофизические признаки, измеренные на множестве точек $\{(x_i, y_i)\} = T$; $f^{(r)}(p_j)$ — некоторые функции признаков p_j , $T^3 \subset T$.

Например,

$$\widehat{\varphi} = \sum_j (a_1 p_j + a_2 p_j^2 + \dots + a_k \ln p_j + a_{k+1} \ln p_{j+1} + \dots + a_{k+2} p_{j+1}^3 + \dots). \quad (\text{III.2})$$

Значения коэффициентов определяются на основе решения избыточной системы линейных уравнений по методу наименьших квадратов, получаемой на некотором «эталонном» множестве точек $T^3 = \{(x_i^3, y_i^3)\}$, $T^3 \subset T$, где значения φ_i известны. При этом вид функции в уравнении (III.1) определяется путем некоторого перебора признаков p_j и $f^r(p_j)$, обеспечивающих минимизацию величины

$$S_{\varphi p}^3 = \sqrt{\frac{\sum_i (\varphi_i - \widehat{\varphi})^2}{n - q}}, \quad (\text{III.3})$$

где n — число точек на эталоне, q — порядок уравнения.

Более внимательное отношение к выбору «наилучшего» уравнения приводит к организации опробования получаемых уравнений на некотором подмножестве эталона — контрольном множестве — т. е. тех точках $T_i^k = \{(x_i^k, y_i^k)\}$, $T^k \subset T$, которые не

использовались при построении полинома, с целью минимизации величины $S_{\varphi p}^3$. Получаемые таким образом уравнения дают обычно более надежные результаты при прогнозе.

Этот процесс описывается и приводится в различных работах на совершенно различных уровнях — от интуитивного графического до «строго» математического в терминах теории множеств и функционального анализа. Однако использование высокого теоретического уровня базируется фактически на тех же самых интуитивных представлениях о единстве эталонной и прогнозной областей $T \sim T^3$, т. е. о применимости уравнений, полученных на эталоне T^3 во всей области прогноза T , или, иначе говоря, о некоторой эквивалентности величин $s_{\varphi p}^k$ ($s_{\varphi p}^3$) и $s_{\varphi p}^{np}$. Сколько-нибудь строгих оценок точности получаемых при прогнозе результатов $s_{\varphi p}^{np}$ до сих пор нет, хотя есть хорошие примеры практического использования метода.

Рассмотрим некоторые особенности прогнозирования более подробно.

1. Подбор уравнения (III.1) ведется приближенно потому, что полиномы высоких степеней не имеют устойчивости и дают большие величины $s_{\varphi p}^{np}$. Уравнения могут быть вообще несовместными — одинаковым значениям $p_j(x_i)$ могут соответствовать разные значения φ_i . Поэтому приходится выбирать некоторую усредненную зависимость $\varphi = f(p_j)$, причем из-за наличия хорошей технологии используется метод наименьших квадратов, метрика которого здесь определяется выражением (III.3). Поэтому малость величины $s_{\varphi p}^3$ не всегда означает хорошее приближение даже на эталоне. Пусть, например, функция φ состоит из участков $\varphi = \text{const}$ и участков аномальных, имеющих значительно меньшие размеры, причем интерес представляют именно эти аномальные участки, p_j есть линейные функции координат. Тогда, в зависимости от соотношения длин аномальных и неаномальных участков, мы можем получить хорошее значение $s_{\varphi p}^3$, однако аномальные участки вовсе не будут отражены в значениях $\hat{\varphi}$. Аналогичное положение может создаваться неудачным соотношением различных частей профиля, участка на эталоне. Этот же эффект, естественно, может проявиться и при вычислении $s_{\varphi p}^k$. Отсюда ясно, что основанный лишь на рассмотрении величины $s_{\varphi p}^3$ подход к выбору признаков и к расположению точек замеров их значений на площади в ряде случаев может приводить к плохим результатам: несмотря на внешнее благополучие (малость величин $s_{\varphi p}^3$, $s_{\varphi p}^k$), прогноз может быть неудовлетворительным, причем в первую очередь это относится к наиболее интересным аномальным участкам. Следовательно, оценка возможностей прогнозирования прежде всего должна основываться на неформальном изучении каждой конкретной ситуации, на детальном рассмотрении физико-гео-

логических, геохимических и других предпосылок. Получаемые результаты полностью зависят от правильности выбора однородных в смысле возможностей использования уравнения (III.1) участков территории и правильной организации эталонного, контрольного и прогнозного множеств.

Иногда в случае больших значений $s_{\varphi p}^3$ производится формальное дробление территории на более мелкие участки. Выделяются, например, участки с одинаковыми интервалами значений φ или p_j , $f(p_j)$, или некоторой функции от нескольких признаков p_j одновременно, т. е. фактически разбиение производится с помощью некоторой формальной классификации. Количество вариантов разбиений при этом ограничивается лишь фантазией исследователя и величиной n , так как оно проводится без обучения, исходя из предвзятых соображений. Естественно, результаты такого подхода редко бывают положительными, несмотря на то, что величина $s_{\varphi p}^3$ уменьшается.

2. При изучении зависимости типа (III.1) мы не учитываем пространственное положение пунктов измерений, хотя во многих случаях имеется зависимость между измеренной величиной φ и координатами x_i , y_i . Это связано с тем, что практически невозможно (весьма трудная задача) выделить действительно однородный участок, где получаемые коэффициенты a_k хорошо выдерживаются, а прогноз идет с точностью, близкой к значениям $s_{\varphi p}^k$ *.

Итак, необходимо учитывать изменение коэффициентов в уравнении (III.1) на всей области прогноза T . В то же время обычно (и это является одним из условий возможности успешного применения прогнозирования) исходные значения φ_i обладают существенной разбросанностью по области прогноза — по территории, интервалу изменений и т. п., что позволяет в зависимости от их количества и структуры расположения производить обычную интерполяцию, т. е. в данном случае имеется в виду, что границы областей T и T^a примерно совпадают. Естественно, речь идет о таких условиях, когда точность обычной (независимо от конкретного метода) интерполяции недостаточна и желательно привлечение косвенных данных для ее повышения.

Таким образом, если рассматривать задачу как интерполяционную, ощущается необходимость привлечения дополнительной информации при интерполяции на большие расстояния. Рассматривая ее как задачу прогноза, необходимо учитывать изменение статистической связи и соответственно коэффициентов в уравнении (III.1) на рассматриваемой территории. Представ-

* Эта трудность, как правило, усугубляется еще и недостаточным количеством точек для дробления участка на более мелкие и более однородные подмножества. К тому же, при таком дроблении обычно теряется значительная доля интереса к самой постановке задачи.

ляется интересным объединить эти подходы. Для этого при построении прогнозного оператора необходимо учитывать не только значения признаков p_j , но и положение (координаты) точек наблюдений. Приблизительно это можно сделать как в работе [13], введя координаты пунктов x_i и y_i в уравнение (III.1):

$$\widehat{\varphi} = Q(x, y, \dots, f^r(x, y), p_1, p_2, \dots, p_j, f^{(1)}(p_1), \dots, f^2(p_1), f^1(p_2) \dots). \quad (\text{III.4})$$

В формуле (III.4) как бы совмещаются прогнозные и аппроксимационные полиномы. Однако и здесь из-за использования метода наименьших квадратов учитывается усредненное влияние координат. Фактически они являются некоторыми дополнительными признаками при построении аппроксимационно-прогнозного полинома. При построении оператора (III.4) степень полинома по x и y определяется минимумом величины $s_{\varphi p}^k$ путем перебора различных вариантов (см. гл. V). Рассматриваются полиномы низких степеней — обычно не выше третьей. О. В. Витвицкий предлагает сделать зависящими от координат и коэффициенты полинома a_k .

Другой возможностью учета изменения статистической связи в зависимости от x и y является предварительное вычитание «трендовой» составляющей косвенных признаков таким образом, чтобы оставшаяся часть этих признаков наилучшим образом коррелировала с φ [47]. Трендовая часть p_j определяется полиномом

$$p_j^{\text{TP}} = \sum_{i,k} a_{ki} x^k y^i, \quad (\text{III.5})$$

$$k = 0, 1, \dots, A; \quad i = 0, 1, \dots, B.$$

После вычисления значений p_j^{TP} анализируется связь φ и $p_j^{\text{OCT}} = \varphi - p_j^{\text{TP}}$. Трудности здесь заключаются в определении необходимой степени полинома в формуле (III.5).

В работе [47] рекомендуется считать оптимальным прогнозным значением p_j^{OCT} составляющую такого порядка, до которой дисперсия p_j^{OCT} не возрастает (в пределах доверительного интервала для дисперсии остаточной составляющей предыдущего порядка), а дисперсия составляющей следующего, более высокого порядка испытывает резкое увеличение. Поскольку этот критерий оказался недостаточно четким, О. В. Витвицкий предлагает другой критерий, лучше учитывающий специфику метода. При оптимальном порядке полинома (III.5) должно выполняться условие

$$|R(\varphi, p_j^{\text{TP}})| = \min, \quad (\text{III.6})$$

где R — коэффициент корреляции. По аналогии, в случаях когда число одновременно учитываемых при прогнозе признаков

больше единицы, следовало бы минимизировать величину

$$|R(\varphi; p_1^{TP}, p_2^{TP}, \dots)| = \min. \quad (III.7)$$

3. При прогнозировании предполагается определенное подобие признаков φ и p_j . Смещение экстремумов функций относительно друг друга вызывает быстрое увеличение $s_{\varphi p}^2$ и, соответственно, приводит к невозможности использования метода. На самом деле это смещение можно учесть, используя вместо (III.4) следующее уравнение:

$$\widehat{\varphi}(x, y) = Q((x - \xi_0, y - \eta_0, \dots, p_1(x - \xi_1, y - \eta_1), \dots, \\ \dots, p_2(x - \xi_2, y - \eta_2) \dots)), \quad (III.8)$$

где ξ_j, η_j — величина смещения, определяемая предварительно по максимуму функции взаимной корреляции или в результате некоторого перебора значений ξ_j, η_j в (III.8) по минимуму значения $s_{\varphi p}^2$. Эти значения могут определяться интерпретатором и просто при визуальном сопоставлении графиков или карт признаков φ и p_j .

Рассмотрим пример из работы [13]. Геолого-геофизические данные о строении Большеземельского района Тимано-Уральской нефтегазоносной провинции свидетельствуют о существовании определенной зависимости между рельефом нижнепермских карбонатных пород (сейсмический горизонт III^a), аномалиями силы тяжести Δg и нижележащими горизонтами ее осадочного чехла (сейсмические горизонты III, IV, VI). Поверхность карбонатных пород отложений нижней перми исследована наиболее полно, а ее структурный план в общих чертах совпадает с планом нижележащих границ, приуроченных к отложениям карбона и девона. С учетом данной физико-геологической ситуации была поставлена задача прогнозирования структурных поверхностей III, IV и V, которые не удается непрерывно проследить сейсморазведкой МОВ, к тому же редкая сеть профилей не обеспечивает надежных структурных построений. Ниже приводятся результаты прогнозирования горизонта IV по сведениям об отметках горизонта III^a, Δg и их координатах.

Расчеты проводились с помощью программы «Статистика» (гл. V), позволяющей автоматически производить перебор сочетаний признаков (и их трансформант) с целью выбора наилучшего полинома. Лучшим среди рассматриваемых считался полином по возможности наименьшей длины, содержащий наиболее простые функции и обеспечивающий минимум среднеквадратической ошибки прогноза на контрольном материале. Эталонными считались данные по профилям I и III, контрольными — данные по профилю II. В качестве аргументов рассматривались $H_{III}^a, \Delta g, \ln H_{III}^a, \ln \Delta g, \frac{1}{H^a}, \frac{1}{\Delta g} x, y, x^2, y^2, xy$. Наилучшие полиномы в прогнозном варианте, аппроксимационно-про-

гнозом, соответствующие уравнениям (III.1) и (III.4), а также аппроксимационный полином имеют вид

$$H_{IV} = 1,050H_{III}^a + 0,014\Delta g^* - 1,006, \quad (III.9)$$

$$R^s = 0,79, \quad s^s = \pm 140 \text{ м}, \quad s^k = \pm 150 \text{ м}.$$

$$H_{IV} = -1,301H_{III}^a + 0,004\Delta g + 0,195x - 0,225x^2 - \\ - 0,001y - 0,015y^2 - 0,001xy - 0,647. \quad (III.10)$$

$$R^s = 0,91, \quad s^s = \pm 90 \text{ м}, \quad s^k = \pm 60 \text{ м}.$$

$$H_{IV} = -0,225x + 0,065x^2 - 0,177y + 0,005y^2 + 0,028xy - 2,865, \\ R^s = 0,65, \quad S^s = \pm 170 \text{ м}, \quad S^k = \pm 160 \text{ м}. \quad (III.11)$$

Результаты прогноза по профилю II представлены на рис. 27. Из рис. 27 видно, что ошибка уравнения (III.9) на протяженных участках имеет систематический характер, причем искажения достигают значительной величины. Наибольшее искажение амплитуды регионального подъема наблюдается в северо-восточной части профиля. Интерполяция с помощью полинома второй степени (III.11) дает еще менее точные результаты. В то же время при использовании уравнения (III.10) качество прогноза значительно улучшается, систематические отклонения практически отсутствуют. Весьма интересным является редко отмечаемый случай: $s^k < s^s$. Наиболее вероятной интерпретацией этого факта является региональное изменение количественных соотношений H_{IV} , H_{III}^a , Δg от профиля I к профилю III, поэтому на профиле II, находящемся между ними, полином (III.10), учитывающий это изменение, дает более надежные результаты. В то же время, поскольку основная задача — построение карты в области между I и III профилями, контроль по профилю II является наиболее жестким.

Аналогичные в методическом смысле результаты получены при прогнозировании горизонтов III и V, а также при использовании этой методики для изучения строения локальных структур Денисовского прогиба Большеземельского района.

Следует отметить, что точность и детальность исходных материалов в данном случае таковы, что некоторое несовпадение значений φ_i и $\hat{\varphi}_i$ на эталоне не играет существенной роли.

Рассмотрим теперь, что может дать использование (III.4) в условиях примера, приведенного на рис. 28 (Ижевская структура в Верхне-Камской впадине). В данном случае стоит задача прогнозирования отметок кровли нижележащего (сакмарского) горизонта, вскрытого малым числом скважин ($n_1 = 12$) с учетом поведения кровли вышележащего (верейского) горизонта ($n_2 = 34$). Предполагается некоторое подобие этих поверхностей.

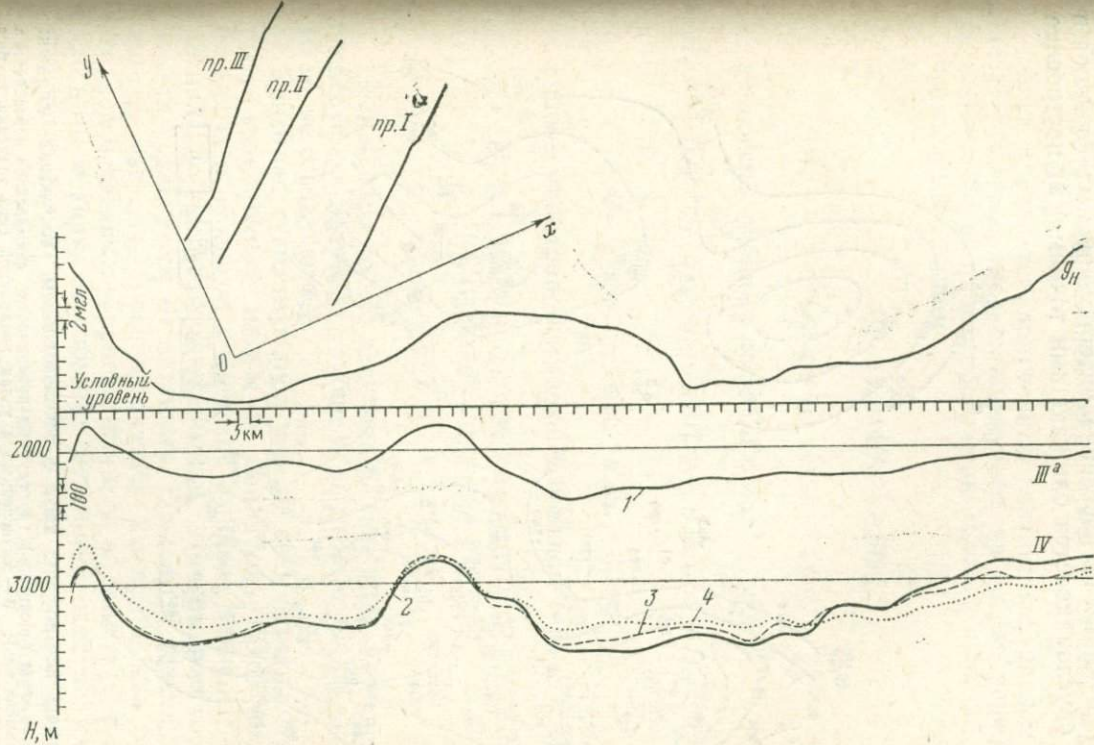


Рис. 27. Схема расположения региональных сейсмических профилей МОВ и результаты прогноза на контрольном профиле II:

1 — поверхность нижнепермских карбонатных отложений по данным МОВ; 2 — опорный сейсмический горизонт IV по данным МОВ; 3, 4 — результаты прогноза горизонта IV по уравнениям (III.10) и (III.9) соответственно

Если строить чисто интерполяционный полином $\varphi = f(x, y)$ степени $m = n_1 - 1$, то мы не будем иметь возможности использовать поведение менее глубоко залегающей поверхности. У такого полинома все точки φ_i^3 имеют равный вес, что невыгодно по двум причинам: 1) поведение интерполируемых функций в основном обуславливается ближайшими точками; 2) построение

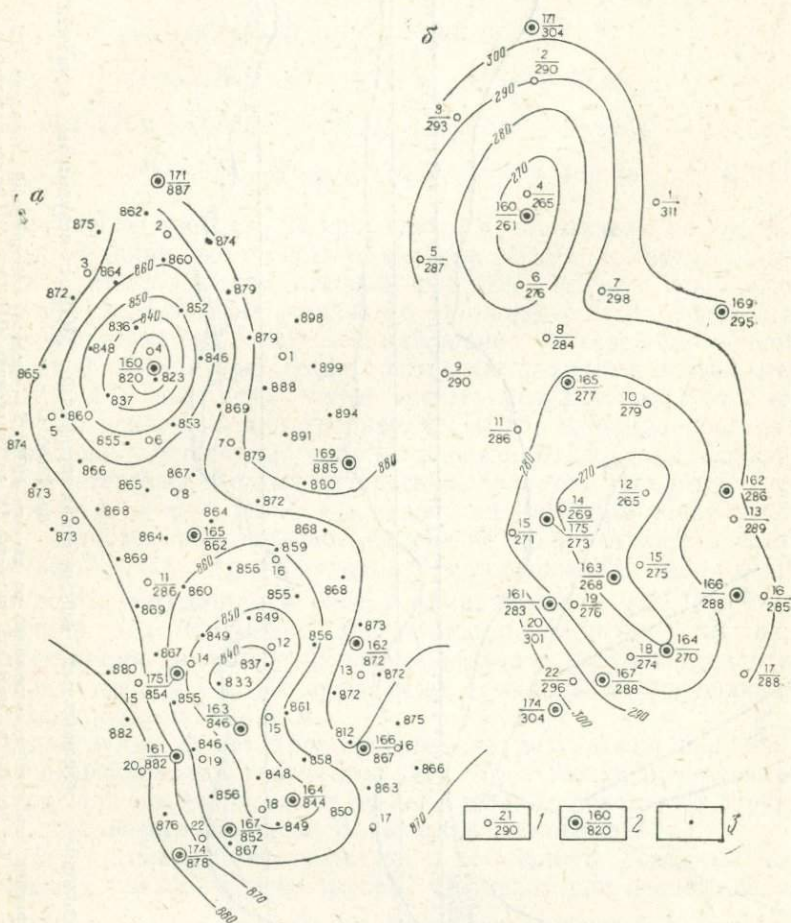


Рис. 28. Пример построения карт с использованием косвенных данных: 1, 2 — номер скважины (числитель) и глубина поверхности сакмарского и верейского горизонтов (знаменатель); 3 — вычисленные в узлах квадратной сети значения глубины поверхности верейского горизонта

полинома высокой степени приводит к возникновению осцилляции значений φ и снижению точности интерполяции. Испол-

зование же «скользящего окна» при таком незначительном количестве точек практически невозможно.

Если воспользоваться методом наименьших квадратов и перейти к полиному более низкой степени, поверхность верейского горизонта (H_2) заглаживается, но зато появляется возможность использовать известные значения отметок глубин сакмарского горизонта (H_1).

Ниже приводятся полученные результаты, причем выражение (III.14) выбрано как наилучшее на основе перебора ряда зависимостей с помощью программы «Статистика»:

прогнозный полином

$$H_2 = 1,042H_1 - 0,564, \quad (III.12)$$

$$s^3 = \pm 9 \text{ м}, \quad s^k = \pm 12 \text{ м},$$

аппроксимационный полином второй степени

$$H_2 = -0,011x - 0,005y - 0,002xy - 0,0015x^2 - 0,0009y^2 - 0,852, \quad (III.13)$$

$$s^3 = \pm 14 \text{ м}, \quad s^k = \pm 20 \text{ м},$$

аппроксимационно-прогнозный полином типа (III.4)

$$H_2 = 1,845H_1 + 0,0017x + 0,0016y + 0,0008xy + \\ + 0,0007x^2 - 0,0007y^2 + 0,3488, \quad (III.14)$$

$$s^3 = \pm 7 \text{ м}, \quad s^k = \pm 10 \text{ м}.$$

Несмотря на то что величина n_1 сравнительно невелика, прогнозируемые значения глубин в некоторых точках заметно (до 10 м) не совпадают.

Учитывая существующую уже степень изученности структуры и сравнительно малую ее амплитуду, такое несовпадение истинных и прогнозных значений слишком велико. Карта, построенная по прогнозным значениям, в некоторых точках будет заметно отличаться от истинных. Если же в точках, где значения φ_i известны, воспользоваться их истинными значениями H_2 , а в остальных — прогнозными, т. е. попытаться «натянуть» прогнозную карту на опорные (истинные) отметки, карта также получится сильно искаженной за счет возникновения зон слишком высоких и слишком низких градиентов изменения отметок H_2 . Таким образом, в этом случае желательно построить такую плавную поверхность, которая совпадала бы с известными значениями H_i достаточно хорошо. Другими словами, необходим переход к интерполяционно-прогнозному варианту.

§ 2. Интерполяционно-прогнозный вариант

В случае когда целесообразно интерполяционно-прогнозное решение задачи, необходимо выполнение условий

$$\widehat{\varphi} = Q(x, y, p_1, p_2, \dots), \quad S_{\varphi, pxy}^k \leq c' \varepsilon_i, \quad T^k = \{(x_i^k, y_i^k)\}, \quad (III.15)$$

$$|\widehat{\varphi}_i - \varphi_i| \leq \varepsilon_i, \quad T^3 = \{(x_i^3, y_i^3)\},$$

где ε_i — некоторая достаточно малая величина (например, точность наблюдений), c' таково, что использование косвенных признаков еще имеет смысл, например, $c' \varepsilon_i \leq \delta_{ii}$, где δ_{ii} — ошибка интерполяции только по значениям φ_i . Таким образом, поведение $\widehat{\varphi}$ в области между точками измерений в значительной мере определяется значениями косвенных признаков p_j , однако

разность $|\widehat{\varphi}_i - \varphi_i|$ в тех точках, где значения φ заданы, должна быть достаточно мала. Преимущества равномерного приближения (III.15) перед среднеквадратическим здесь очевидны.*

В качестве наиболее простого варианта реализации условия (III.15) можно предложить следующее.

После определения с помощью формулы (III.4) значений $\widehat{\varphi}$ значения $\delta_i = \widehat{\varphi}_i - \varphi_i$ интерполируются на множестве T и в дальнейшем используются значения $\widehat{\varphi}' = \widehat{\varphi} + \delta(x, y)$. Такую интерполяцию можно провести с помощью метода, изложенного в § 5 гл. I, который проводит ее с высокой точностью и гарантирует выполнение неравенства $\delta_i \leq \varepsilon_i$, т. е. условия (III.15). В простых случаях эта интерполяция может быть проведена и полиномами низких степеней, хотя это, даже без учета снижения точности, неудобно, поскольку выполнение неравенства (III.15) заранее не гарантируется. На этом возможности статистического подхода при решении интерполяционно-прогнозных задач, по крайней мере в его современном состоянии, в значительной мере исчерпываются.

В то же время во многих случаях достигаемая точность решения задачи является недостаточной. Возникает естественный вопрос — нельзя ли решать задачи учета косвенных признаков, используя не только статистический, но и детерминированный подход или их комбинацию. Можно ожидать, что в случае положительного результата это позволило бы резко сократить количество необходимых для интерполяции прогноза значений φ_i .

* Название «интерполяционно-прогнозный» введено здесь, так как решается интерполяционно-прогнозная задача при заданных заранее достаточно малых, хотя и не нулевых значениях ε_i .

Например, в случае построения удовлетворительной математической модели образования структур в осадочной толще такая задача могла бы решаться по весьма незначительному количеству исходных точек на эталоне (необходимых для определения параметров модели) или даже вовсе без них — только по признакам p_j . При этом, конечно, мы должны были бы знать априори о соответствии выбранной модели и реального процесса.

Весьма удобно, с принципиальной точки зрения*, эта задача решалась бы с помощью аналитического продолжения, если бы функции φ и p_j были одной и той же гармонической функцией, например измеренной на равных уровнях. Предположим, нам надо восстановить поле аномалий силы тяжести на уровне $z = -h$, где имеются измерения, полученные в скважинах с ошибкой ϵ' и, кроме того, на плоскости $z = 0$ имеются измерения по значительно более густой сети с ошибкой ϵ'' . Ясно, что значения u , измеренные на плоскости $z = -h$, могут значительно улучшить точность аналитического продолжения, проведенного на основе использования одних только значений $U(z = 0)$. Представим теперь, что то же самое можно делать и с поверхностями осадочной толщи, т. е. по относительно более изученным, мелко залегающим структурным поверхностям восстановить более глубокие, о которых имеется существенно меньшее количество данных [17].

Рассмотрение некоторых разрезов и карт показывает, что иногда изменение различных поверхностей с глубиной вполне аналогично поведению гармонических функций. К таким явлениям относится, например, вылаживание структурных поверхностей с уменьшением глубины и увеличение мощности осадков на крыльях.

Можно отметить также хорошие результаты использования для интерполяции самых различных геологических характеристик — карт структурных поверхностей, мощностей, метропроцентов, содержаний рудных залежей и т. д. — методов, разработанных для интерполяции гармонических функций (гл. I). Эти факты свидетельствуют в пользу возможности практического использования предположения о гармоничности геологических признаков в определенных ситуациях. Однако коренная разница заключается в том, что в отмеченных примерах предположение о гармоничности используется для интерполяции, а здесь предлагается его использовать и для экстраполяции. Ясно, что в этом случае требуется совершенно другой порядок близости выбранной теоретической модели к реальной, однако в

* Далее изложение ведется применительно к структурным поверхностям, хотя, возможно, его удастся распространить и на другие геологические признаки.

случае успеха и результаты будут несравненно более интересными.

Очень важным в данном случае является вопрос определения области, в которую имеет смысл производить аналитическое продолжение. В отличие от обычного в геофизике случая, когда мы имеем дело с некоторыми полями аномальных источников, геологические признаки, начиная с некоторого момента, просто исчезают или их поведение затушевывается наложением осложняющих факторов. При изучении локальных структурных поверхностей к таким осложняющим факторам, очевидно, следует отнести резкое изменение условий и причин складкообразования, осадконакопления, изменение региональных процессов, дизъюнктивные дислокации, большие перерывы в осадконакоплении и т. п. Таким образом, аналитическое продолжение возможно лишь в область, характеризующуюся некоторым единством геологического строения. Здесь имеется прямая аналогия с основным принципом статистического прогнозирования — необходимость наличия определенной эквивалентности эталонной и прогнозной областей. Очевидно, в ряде случаев «помехи» можно учесть, используя эталонные данные. С другой стороны, вычислительные трудности, возникающие из-за некорректности решаемой задачи (пересчет в нижнее полупространство), здесь, очевидно, не будут иметь серьезного значения, поскольку можно заранее ожидать, что аналитическое продолжение будет производиться лишь в полосу, непосредственно примыкающую к исходным данным, т. е. для значений $h/\Delta x < 1$, где Δx — расстояние между точками p_j .

Для проверки правильности принятого предположения можно использовать аналитическое продолжение в верхнее полупространство прогнозируемой функции как гармонической, заданной на бесконечной прямой, которое реализуется с помощью интеграла Пуассона:

$$U(x, h) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{hU(\xi, 0)}{(\xi - x)^2 + h^2} d\xi, \quad (\text{III.16})$$

где $U(\xi, 0)^*$ — значения гармонической функции на прямой $z=0$, h — высота пересчета, $U(x, h)$ — значения функции на плоскости $z=h$. Используя формулу (III.16), о пригодности выбранной модели можно судить по величине нормы

$$\| p'(x_i, 0) - \bar{p}''(x_i, 0) \|, \quad (\text{III.17})$$

т. е. по разности между известными отметками верхнего горизонта p' и значениями, полученными с помощью продолжения в

* Кроме предположения о гармоничности функции $U(\xi, 0)$, необходимо еще выполнение условия $\lim_{\xi \rightarrow \infty} U(\xi, 0) = 0$. Это условие легко выполнимо.

верхнее полупространство отметок поверхности прогнозируемого более глубокого горизонта p'' . Величина h определяется опытным путем — по минимуму нормы (III.17). Отметим, что если $U(\xi, 0) = \text{const} = c$, то $U(x, h) = c$; если $U(\xi, 0) = a\xi + c$, то и $U(x, h) = ax + c$. Таким образом, постоянные величины и линейные относительно координат функции преобразование (III.16) оставляют без изменения. Отсюда ясно, что нельзя, например, непосредственно использовать абсолютные значения глубины залегания сейсмических отражающих горизонтов, во много раз превышающие амплитуды их колебаний. Перед расчетами их необходимо привести к одному уровню или после их проведения учесть постоянную составляющую. Подобным образом необходимо учитывать и линейный региональный фон.

Необходимо отметить, что предположение о гармоничности функций p_j не является обязательным. Оно введено лишь для того, чтобы была ясна физическая интерпретация предлагаемого метода использования косвенных признаков в интерполяционно-прогнозном варианте и та трактовка его, которая излагается ниже.

Пусть, например, $p'(x_{i'})$ и $p''(x_{i''})$ — значения отметок двух структурных поверхностей, известные в n' и n'' точках, причем $x_{i'} \neq x_{i''}$. Будем считать, что значения p' относятся к прямой $z=0$, а значения p'' — к прямой $z=-h$. Положим для определенности, что $n' \gg n''$, т. е. значения p' рассматриваются как значения косвенного признака. Аппроксимируем значение p' конструкцией вида

$$p'(x) \approx \bar{p}(x) = \sum_k \alpha_k G_k(x - k\Delta x, z_1), \quad (\text{III.18})$$

где G_k определяются выражением (I.80), а значения α_k определяются уравнением (I.87) по изложенной там же методике, $z_1 = (1 \div 2)\Delta x$. Используя полученные значения α_k , вычислим значения $\bar{p}(x_{i''}, z_1 + h)$, выбирая значения h таким образом, чтобы

$$\| p''(x_{i''}) - \bar{p}(x_{i''}, z_1 + h) \| = \min_h. \quad (\text{III.19})$$

Если величина нормы (III.19) удовлетворяет неравенству (III.15), вычисляются значения $\bar{p}(x, z_1 + h)$ на всей интерполяционно-прогнозной области.

Важно отметить, что принципиальным преимуществом перед статистическим моделированием является возможность вычисления значений $\bar{p}(x, z_1 + h)$ не только в точках $x_{i''}$, но и в любых точках x_i , например, в точках равномерной сети при $\Delta x' < \Delta x$, как это сделано в примере на рис. 28.

Если неравенство (III.15) не выполняется, но величина нормы (III.19) не слишком велика, «натягивание» значений $p(x, z_1 + h)$ на значения $p''(x_{i''})$ производится повторением

процесса вычисления a_h , используя в формуле (III.18) одновременно значения $p'(x_{i'}, z_1)$ и $p''(x_{i''}, z_1+h)$, и последующим вычислением $\bar{p}(xz_1+h)$. Заметим, что величина h может быть положительной и отрицательной — в зависимости от соотношения кривизны поверхностей p' и p'' . Алгоритм метода в двухмерном варианте совершенно аналогичен.

Таким образом, для реализации интерполяционно-прогнозного процесса предлагается строить такую аппроксимационную конструкцию из гармонических функций (I.80), которая одновременно приближает значения p' и p'' , т. е. φ и p_j . Весь этот процесс можно рассматривать как построение оператора L , переводящего значения p_j в значения φ_i при выполнении условия (III.15):

$$Lp = \widehat{\varphi}, \quad |\widehat{\varphi}_i - \varphi_i| < \varepsilon_i. \quad (\text{III.20})$$

Легко убедиться, что если соответствующим образом строить численную схему аналитического продолжения, полученные результаты будут совершенно одинаковы.

Изложенная методика была опробована на примере Ижевской структуры (см. рис. 28). Параметр h в данном случае определяется с помощью пересчета верхней поверхности $p(x_{i'})$ на ряд уровней вниз и принят равным $h=0,2\Delta x$. При этом среднеквадратическая ошибка простого пересчета составила

$$\delta = \sqrt{\frac{\sum_{i=1}^{i=n} [p'(x_{i'}, 0) - \bar{p}''(x_{i''}, 0)]^2}{n-1}} = 8,8 \text{ м.}$$

Окончательный вариант интерполяции был получен путем совместной аппроксимации верхнего и нижнего горизонтов при $h=0,2\Delta x$. Точность аппроксимации в исходных точках $x_{i'}$ и $x_{i''}$ была задана равной

$$\varepsilon' = |p'(x_{i'}) - \bar{p}(x, h)| \leq 2\text{м},$$

$$\varepsilon'' = |p''(x_{i''}) - \bar{p}(x, 0)| \leq 2\text{м}.$$

Полученная карта приведена на рис. 28,а.

Весьма вероятно, что в некоторых случаях выполнение условия (III.15) возможно лишь при изменении величины h от участка к участку. В этом случае плавность «склейки» поверхностей может быть достигнута путем перекрытия этих участков и построения такой вычислительной схемы, в которой значения h плавно изменяются в каждой точке φ_i (или группе точек). Минимальное количество точек, по которому можно определить значение h , равно двум. Если необходимо использовать несколько признаков p_j , которые к тому же могут иметь другую раз-

мерность, чем φ , можно сначала вычислить $\hat{\varphi}_i$ по формуле (III.4) и затем использовать их в качестве p_j . Таким образом, предлагаемая методика организации интерполяционно-прогнозного процесса позволяет выполнять условия (III.15).

Отметим еще раз, что идея применения аналитического продолжения для прогнозирования структурных поверхностей не имеет строгого математического обоснования и базируется на визуальном сопоставлении геологических разрезов. Существует, однако, простой способ приближенной проверки (III.17) адекватности такой модели реальному процессу.

Интерполяционно-прогнозный процесс, по-видимому, всегда следует строить так, чтобы выполнялись условия (III.15). И в тех случаях, когда применение предлагаемой методики нецелесообразно, из-за большой величины нормы (III.17), имеет смысл, как это указывалось выше, проводить простую интерполяцию невязок в узлы прогнозной сети.

Подсчет запасов производится сейчас практически только по объемной формуле. С точки зрения численного интегрирования это соответствует случаю, когда вся область интегрирования состоит из единственной элементарной площадки. Точность численного интегрирования при этом наиболее низка, поэтому расчеты такого рода могут быть оправданы лишь при высоком уровне ошибок значений интегрируемой функции φ в точках наблюдений (скважинах) φ_i , $i=1, 2, 3, \dots$, существенно превышающем систематические изменения значений признака между скважинами, или слишком незакономерным поведением функции φ между узлами (x_i, y_i) , не позволяющем производить сколько-нибудь достоверную интерполяцию. Под φ здесь подразумевается функция удельных запасов, $\varphi = \sum \varphi_j$, где φ_j — коэффициент открытой пористости, плотность, содержание и т. п. признаки. Даже в том, возможном только теоретически случае, когда средние значения функции φ на площади залежи известны нам точно, ошибки вследствие использования объемной формулы могут достигать нескольких десятков процентов. В тех же реальных случаях, когда средние значения φ определяются по значениям φ_i в ограниченном числе точек, расположенных, как правило, по неравномерной сети, ошибки еще более велики (даже без учета величины ошибок измерений). Это связано с тем, что объемная формула не учитывает расположения пунктов измерений и, следовательно, все измерения учитываются с одинаковым весом. Положение еще больше осложняется в том случае, когда сначала вычисляются средние значения исходных функций $\bar{\varphi}_j$, произведение которых определяет функцию φ . Поэтому иногда делаются попытки «взвешивания» исходных значений φ_j и φ . Например, «взвешиваются» значения пористости по мощности нефтенасыщенного пласта, учитывается площадь залежи, приходящейся на i -ое измерение, и т. п., иногда применяется метод «изолиний». При этом объем расчетов резко увеличивается, а задача повышения точности решается лишь частично. В ряде случаев начинает сильно сказываться влияние субъективных оценок.

Объемная формула имеет очень простую трактовку и с позиций математической статистики. С этим, очевидно, и связано то, что сейчас имеются по сути дела лишь статистические оценки точности подсчетов запасов. Можно, однако, заметить, что, поскольку при получении оценок дисперсии $\bar{\varphi}$ и φ_j учиты-

вается только количество замеров, без какого-либо учета их расположения по площади залежи*, получаемые оценки должны быть много хуже тех, которые можно надеяться получить при детерминированном подходе.

Естественно, что проведение оценок точности с детерминированных позиций требует знания свойств функций φ_j , например, значения производных, спектральной характеристики и др. Не следует думать, что при статистическом подходе к оценке точности задача существенно упрощается. В этом случае необходимо определить тип распределения признаков, измерения должны удовлетворять требованиям однородности и независимости, должны располагаться случайным образом, и все это по заведомо недостаточному количеству измерений. Кроме того, вряд ли выполнение этих требований может соблюдаться в реальных условиях. Наоборот, скважины могут концентрироваться на определенных участках залежи, создавая превратное представление не только о дисперсии признаков, но и об их средних значениях. Независимость измерений, как правило, вообще не проверяется, поскольку сделать это, во всяком случае на этапе разведки, практически невозможно. В то же время наблюдения, как правило, заведомо зависимы. И, конечно, противоестественной является необходимость отбрасывания в этом случае части замеров (если бы удалось установить, что шаг измерений, при котором измерения можно считать независимыми, больше реального) в условиях небольшого их числа при разведке залежей нефти и газа.

В то же время при благоприятных обстоятельствах, т. е. при достаточно хорошем поведении признаков, возможна их интерполяция на всю площадь залежи без существенной потери точности относительно значений φ_i и, следовательно, возможно соответствующее повышение точности расчетов. В некоторых случаях одновременное применение эффективных методов фильтрации может существенно уменьшить влияние случайных ошибок измерений.

В связи с этим интересно рассмотреть возможность применения при подсчете запасов «высоких» категорий методов, позволяющих производить численное интегрирование с высокой точностью при нерегулярной сети исходных данных. Очевидно, не меньший интерес, чем сама возможность подсчета запасов таким образом, представляет оценка точности этих расчетов. При этом следует сразу учесть, что такого рода оценки, так же как и оценки, получаемые методами математической статистики, не могут быть полностью достоверными. Они обязательно базируются на некоторых предположениях о свойствах признаков

* Учитывая недостатки этого подхода, залежи иногда разбиваются на отдельные блоки, в которых признаки φ_j или хотя бы часть из них более однородны.

φ_j , которые мы можем установить по аналогии с детально изученными месторождениями того же типа и в определенной степени по тем данным, которые мы имеем по уже проведенным измерениям. Точность этих оценок по сути дела полностью зависит от адекватности выбранной модели и реального месторождения, в особенности от тех значений параметров, характеризующих признаки, которые мы приняли при расчетах. Можно предположить, что с накоплением опыта эти оценки будут достаточно близкими к действительности. Однако и качественная сторона этого вопроса очень важна. Нам важно четко представлять себе все, что происходит при расчетах на моделях, для того чтобы по возможности правильно и, соответственно, наиболее экономичным способом учитывать влияние тех или иных факторов, сказывающихся на точности расчетов. Наиболее важным фактором такого рода является, очевидно, шаг измерений Δx , которым фактически определяется экономическая целесообразность той или иной степени детальности разведки месторождения.

Рассмотрение этих вопросов представляется тем более актуальным, что возможность использования ЭВМ ставит, с одной стороны, вопрос об автоматизации подсчета запасов, а с другой — позволяет рассматривать методики расчетов фактически вне зависимости от объема вычислений. Естественно при этом сделать точность этих расчетов максимальной.

Далее будут даны теоретические оценки точности, оптимальные при определенных предположениях о свойствах функций φ_j и характере разведочной сети, а также оценки точности метода, изложенного в § 5 гл. I, который предлагается использовать при расчетах [16].

Оценки можно получить моделированием на детально изученных структурах и затем использовать в аналогичных условиях [25]. Однако при этом труднее уловить особенности влияния каждого фактора, приходится рассчитывать большое число вариантов. Особенно важно, чтобы адекватность условий устанавливалась лишь для достаточно общих характеристик функций, что единственно возможно в практических условиях. Фактически лишь в этом случае их можно будет конструктивно использовать.

Дальнейшее изложение проводится применительно к залежам нефти и газа, однако с учетом сделанных ограничений и при условии выбора соответствующих признаков φ_j , оно полностью переносится и на твердые полезные ископаемые.

**§ 1. Оценка точности подсчета запасов
в зависимости от величины шага наблюдений**

Как известно, задача подсчета запасов нефти заключается в вычислении следующего интеграла:

$$Q = \int_V \beta(x, y, z) m(x, y, z) \rho(x, y, z) \eta(x, y, z) dV, \quad (IV.1)$$

где η — коэффициент нефтеотдачи, m — коэффициент открытой пористости, β — коэффициент нефтенасыщенности, ρ — плотность нефти, Q — извлекаемые запасы, V — объем залежи. Будем считать, что разделение залежи на отдельные пласты-коллекторы уже произведено и расчеты проводятся для каждого пласта отдельно. Учитывая это, а также что расстояние между скважинами примерно на один-два порядка больше мощности нефтяного пласта, задача практически сводится к вычислению двойного интеграла:

$$Q = \int_S \int h(x, y) m'(x, y) \rho'(x, y) \eta'(x, y) dS, \quad (IV.2)$$

где S — площадь продуктивного поля, h — мощность нефтенасыщенной части коллектора, m' , ρ' , η' — средневзвешенные на $h(x, y)$ значения m , ρ и η . Как и ранее, будем обозначать эти функции φ_j , а их произведение — φ . Для простоты теоретические вопросы (там, где это не имеет принципиального значения) будем рассматривать для одномерного случая:

$$Q = \int_{-d}^d \varphi(x) dx, \quad (IV.3)$$

где $2d$ — длина продуктивного поля.

Необходимо получить достаточно хорошие оценки точности вычисления интегралов (IV.2), (IV.3). Воспользуемся для этого способом, с помощью которого в § 2 и § 5 гл. I были проведены оценки точности интерполяции геолого-геофизических признаков на регулярную сеть. Отметим сразу, что эти оценки в определенном смысле можно считать оптимальными и что метод, предложенный в § 5, позволяет производить практические расчеты именно с этой точностью.

Искомая величина равна

$$\int_{-d}^d \varphi(x) dx - \sum_k^k c'_k \varphi(x_k), \quad (IV.4)$$

где функция $\varphi(x)$ задана дискретно, c'_k — коэффициенты численного интегрирования.

Ясно, что при разумном выборе значений c'_k величина этой разности определяется свойствами функции $\varphi(x)$. Как уже

указывалось в § 2 гл. I, из-за отсутствия теории процесса эти свойства заранее определить нельзя. Поэтому, так же как и при оценке точности интерполяции, воспользуемся моделью $\varphi_j \in U'(H)$ и проведем аналогичный анализ.

Пусть, как и ранее (§ 2, гл. I), $\tilde{\varphi}$ — целая функция с ограниченным спектром степени σ , значения которой в узлах сети совпадают с наблюдаемыми, $\Delta x = \text{const}$, $\tilde{\varphi}(k\Delta x) = \varphi(k\Delta x)$. Поскольку функцию $\varphi(k\Delta x)$ можно восстановить точно, то, очевидно, и

$$\int_{-\infty}^{\infty} \tilde{\varphi}(x) dx = \Delta x \sum_{k=-\infty}^{\infty} \tilde{\varphi}(k\Delta x). \quad (\text{IV.5})$$

Это можно показать, воспользовавшись суммационной формулой Пуассона [18]:

$$\begin{aligned} \sum_{k=-\infty}^{\infty} \tilde{\varphi}(k\Delta x) &= \lim_{k \rightarrow \infty} \sum_{-k}^k \frac{1}{\Delta x} \int_{-\infty}^{\infty} \tilde{\varphi}(x) e^{-2k\pi i x / \Delta x} dx = \\ &= \frac{1}{\Delta x} \left[\int_{-\infty}^{\infty} \tilde{\varphi}(x) dx + 2 \sum_{1}^{\infty} \int_{-\infty}^{\infty} \tilde{\varphi}(x) e^{-2k\pi i x / \Delta x} dx \right]. \end{aligned} \quad (\text{IV.6})$$

Последний интеграл в (IV.6) есть преобразование Фурье функции $\tilde{\varphi}(x)$, причем $|\omega| = \frac{2k\pi}{\Delta x} > \sigma = \frac{\pi}{\Delta x}$, $k=1, 2, 3 \dots$ Отсюда, по определению,

$$\sum_{1}^{\infty} \int_{-\infty}^{\infty} \tilde{\varphi}(x) e^{-2k\pi i x / \Delta x} dx \equiv 0 \quad (\text{IV.7})$$

и, следовательно, равенство (IV.5) верно.

Таким образом, по аналогии с введенными в § 2 гл. I оптимальными оценками восстановления функций при вычислении запасов под оптимальной и асимптотически оптимальной величинами Δ и Δ_n будем понимать

$$|\Delta| = \left| \int_{-\infty}^{\infty} \varphi(x) dx - \int_{-\infty}^{\infty} \tilde{\varphi}(x) dx \right| \quad (\text{IV.8})$$

и, соответственно,

$$|\Delta_n| = \left| \int_{-d}^d \varphi(x) dx - \int_{-d}^d \tilde{\varphi}_n(x) dx \right|, \quad (\text{IV.9})$$

т. е.

$$|\Delta| = \left| \int_{-\infty}^{\infty} \varphi(x) dx - \Delta x \sum_{k=-\infty}^{\infty} \tilde{\varphi}(k\Delta x) \right|, \quad (\text{IV.8a})$$

$$|\Delta_n| = \left| \int_{-d}^d \varphi(x) dx - \Delta x \sum_{k=-n}^n \tilde{\varphi}_n(k\Delta x) \right|, \quad (\text{IV.9a})$$

причем

$$\begin{aligned}\tilde{\varphi}_n(k\Delta x) &= \tilde{\varphi}(k\Delta x) = \varphi(k\Delta x), \quad |k| \leq n; \\ \tilde{\varphi}_n(k\Delta x) &= 0, \quad |k| > n.\end{aligned}$$

Хотя ясно, что в реальных условиях количество отсчетов и интервал задания функции $-d \leq x \leq d$ всегда ограничен, удобнее, как будет показано ниже, сначала получить значения Δ , а затем, внося соответствующие поправки, перейти к значениям Δ_n .

Функция $\tilde{\varphi}$ является аналитической и не может быть тождественно равна нулю ни на каком интервале. Поэтому на поведение функции φ вне интервала $-d \leq x \leq d$ должны быть наложены соответствующие ограничения. Будем, например, считать, что значения $\varphi(k\Delta x) = 0$ при $-\infty < x \leq -d$ и $d \leq x < \infty$, а

$$\int_{-\infty}^{-d} \varphi(x) dx + \int_d^{\infty} \varphi(x) dx < \varepsilon, \quad (\text{IV.10})$$

где ε достаточно мало. В этом случае поведение функции за пределами интервала $-d \leq x \leq d$ не будет существенно влиять на величину Δ , т. е. Δ практически не зависит от величины ε в (IV.10). Впрочем, в каждом конкретном случае можно сделать соответствующие оценки.

Наконец, так же как и при рассмотрении ошибок интерполяции, все изложенное и, соответственно, получаемые далее результаты остаются верными и для случая нерегулярной сети, если вместо Δx использовать среднее значение $\overline{\Delta x}$, такое чтобы удовлетворялось условие (I.53).

На основании формулы (I.39) запишем

$$\begin{aligned}F(\omega) - \tilde{F}(\omega) &= F(\omega) - \sum_{-\infty}^{\infty} F(\omega + 2n\sigma) = \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\infty} \varphi(x) e^{-i\omega x} dx - \int_{-\infty}^{\infty} \tilde{\varphi}(x) e^{-i\omega x} dx \right]. \quad (\text{IV.11})\end{aligned}$$

Положим $\omega = 0$. Тогда

$$\begin{aligned}|\Delta| &= \left| \int_{-\infty}^{\infty} \varphi(x) dx - \int_{-\infty}^{\infty} \tilde{\varphi}(x) dx \right| = \\ &= \sqrt{2\pi} \left[\sum_{-\infty}^{-1} F(2n\sigma) + \sum_1^{\infty} F(2n\sigma) \right]. \quad (\text{IV.12})\end{aligned}$$

Таким образом, зная $F(\omega)$, величину Δ можно получить суммированием рядов в правой части неравенства (IV.12). Для

функции $U(x, H)$, $c=1|\Delta|$ равно

$$|\Delta| = 2\pi \sum_1^{\infty} e^{-2n\sigma H} = 2\pi \frac{e^{-2\sigma H}}{1 - e^{-2\sigma H}}, \quad (IV.13)$$

относительная ошибка

$$\mu = \frac{|\Delta|}{\sqrt{2\pi} |F(0)|} = \frac{2e^{-2\sigma H}}{1 - e^{-2\sigma H}}. \quad (IV.14)$$

При $H/\Delta x = 0,5$ величина $\mu = 0,09$; при $H/\Delta x = 0,75$ $\mu = 0,02$, а в случае $H/\Delta x \geq 1$ величина μ практически равна 0 или, другими словами, можно сказать, что для случайных процессов типа (I.28) величина μ не превышает 9%, если $\tau_0 \geq 1,6\Delta x$. Это, конечно, очень хороший результат, однако не следует забывать, что он справедлив для случая, когда измерения произведены точно и, что особенно важно при малых n , необходимо еще перейти к Δ_n .

Получим теперь оценку Δ для двухмерного случая. Пользуясь тем же подходом, что и в одномерном случае, будем иметь:

$$\begin{aligned} F(U, V) - \tilde{F}(U, V) &= F(U, V) - \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} F(U + 2p\sigma, V + 2q\sigma) = \\ &= \frac{1}{2\pi} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) e^{-i(Ux+Vy)} dx dy - \right. \\ &\quad \left. - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\varphi}(x, y) e^{-i(Ux+Vy)} dx dy \right]. \quad (IV.15) \end{aligned}$$

При $Ux = Vy = 0$

$$\begin{aligned} |\Delta| &= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) dx dy - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{\varphi}(x, y) dx dy \right| = \\ &= 2\pi \left| \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} F(2p\sigma, 2q\sigma) - F(0, 0) \right|. \quad (IV.16) \end{aligned}$$

Подставляя значение $F(U, V)$ в формулу (IV.16), окончательно получим

$$|\Delta| = 2\pi \left| \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} e^{-2H\sigma\sqrt{p^2+q^2}} - 1 \right|, \quad (IV.17)$$

$$\mu = \frac{|\Delta|}{2\pi |F(0)|} = \frac{|\Delta|}{2\pi}. \quad (IV.18)$$

Рассмотрение более медленно сходящегося, но суммируемого аналитически ряда

$$\sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} e^{-\sqrt{2} \sigma H (|p| + |q|)} \quad (\text{IV.19})$$

показывает, что при расчетах можно ограничиться значениями $|p| = |q| \geq 10$ для $H/\Delta x \geq 0,5$. При $H/\Delta x = 0,5$ значение $\mu = 0,24$, при $H/\Delta x = 0,75$ $\mu = 0,04$, а при $H/\Delta x \geq 1$ μ не превышает 0,01. Полученные результаты относятся к значениям $\tau_0/\Delta x$, равным соответственно 1,4; 2,1 и 2,8.

Найдем теперь величину Δ_n :

$$\begin{aligned} |\Delta_n| &= \left| \int_{-d}^d [\varphi(x) - \tilde{\varphi}_n(x)] dx \right| = \\ &= \left| \int_{-d}^d \left[\varphi(x) dx - \sum_{k=-n}^n \varphi(k\Delta x) \frac{\sin \sigma(x - k\Delta x)}{\sigma(x - k\Delta x)} \right] dx \right|. \quad (\text{IV.20}) \end{aligned}$$

Величину Δ_n имеет смысл определять для величины $|d|$, примерно равной радиусу автокорреляции функций $U(x, H)$, т. е. $|d| \geq \pi H$.

К сожалению, удобного аналитического выражения для Δ_n найти не удалось. Однако, чтобы выяснить качественную картину, можно поступить следующим образом:

$$\begin{aligned} \Delta_n &= \int_{-d}^d (\varphi - \tilde{\varphi}) dx - \int_{-d}^d (\tilde{\varphi} - \tilde{\varphi}_n) dx = \\ &= \Delta - \int_{-\infty}^{-d} (\varphi - \tilde{\varphi}) dx - \int_d^{\infty} (\varphi - \tilde{\varphi}) dx + \int_{-\infty}^{\infty} (\tilde{\varphi} - \tilde{\varphi}_n) dx - \\ &\quad - \int_{-\infty}^{-d} (\tilde{\varphi} - \tilde{\varphi}_n) dx - \int_d^{\infty} (\tilde{\varphi} - \tilde{\varphi}_n) dx. \quad (\text{IV.21}) \end{aligned}$$

На основании формулы (IV.5) получим

$$\begin{aligned} \Delta_n &= \Delta - \int_{-\infty}^{-d} \varphi dx - \int_d^{\infty} \varphi dx + \Delta x \sum_{k=-\infty}^{-(n+1)} \varphi(k\Delta x) + \\ &\quad + \Delta x \sum_{k=n+1}^{\infty} \varphi(k\Delta x) + \int_{-\infty}^{-d} \tilde{\varphi}_n dx + \int_d^{\infty} \tilde{\varphi}_n dx. \quad (\text{IV.22}) \end{aligned}$$

Рассмотрим полученное выражение. Легко видеть, что бесконечные суммы в (IV.22) суть вычисление двух интегралов по формуле прямоугольников. Поскольку при $|x| = d = \pi H$ функция $U(xH)$ является достаточно гладкой и, к тому же, небольшой по величине ($\sim 0,1U(0, H)$), сумма этих величин невелика. Таким образом, величина Δ_n в основном, очевидно, определяется значением Δ и суммой двух последних интегралов от $\tilde{\varphi}_n$.

Как следует из выражения (IV.20), функция $\tilde{\varphi}_n(x)$ вне $(-d, d)$ осциллирует, причем ее экстремумы убывают по абсолютной величине как $\frac{1}{x}$, а $\tilde{\varphi}_n(k\Delta x) = 0, |k| > n$. Таким образом, значение n может быть небольшим, а величина $\Delta_n(-d, d)$ должна сравнительно быстро стремиться к величине $\Delta_n(-\infty, \infty)$:

$$\begin{aligned} \Delta_{n, \infty} &= \int_{-\infty}^{\infty} [\varphi(x) - \tilde{\varphi}_n(x)] dx = \Delta - \int_{-\infty}^{\infty} [\tilde{\varphi} - \varphi_n] dx = \\ &= \Delta - \Delta x \sum_{k=-\infty}^{k=-(n+1)} \varphi(k\Delta x) - \Delta x \sum_{k=n+1}^{\infty} \varphi(k\Delta x) = \Delta - \bar{\Delta}. \end{aligned} \quad (\text{IV.23})$$

Величина $\bar{\Delta}$ для функций $U(x, H)$ довольно велика, даже если значения Δx и n выбраны так, чтобы интервал, в котором учитываются значения $\varphi(k\Delta x)$, удовлетворял условию $|x| \geq \tau_0$, т. е. $n\Delta x \geq \pi H$:

$$\bar{\Delta} \approx 2 \left[\frac{\pi}{2} - \arctg \pi \right] \approx 0,6; \quad \bar{\mu} = 0,2. \quad (\text{IV.24})$$

Здесь мы воспользовались тем обстоятельством, что суммы в (IV.23) можно приближенно заменить соответствующими интегралами. Следует подчеркнуть, что здесь просто сказывается слабое убывание при $x \geq \pi H$ функций $U(x, H)$, т. е. в этом смысле выбранная модель является неудачной. Действительно, легко видеть, что

$$\int_{-\infty}^{-d} \varphi(x) dx + \int_d^{\infty} \varphi(x) dx \approx \Delta x \left[\sum_{k=-\infty}^{k=-(n+1)} \varphi(k\Delta x) + \sum_{k=n+1}^{\infty} \varphi(k\Delta x) \right], \quad (\text{IV.25})$$

т. е. главная часть сумм в (IV.23), по сути дела, определяется поведением функции $\varphi(x)$ вне интервала $(-d, d)$, где на самом деле при подсчете запасов $\varphi(x) = 0$.

Итак, определив величину H по первым скважинам или по аналогии с соседними хорошо изученными месторождениями, можно определить возможную точность подсчета запасов. И, наоборот, задаваясь необходимой точностью подсчета запасов, можно определить необходимое значение шага Δx при разведке. Методика определения величины H изложена в § 2 гл. I. Если для аппроксимации функции φ необходимо несколько функций $U(x, H)$, оценки легко получаются для каждой аппроксимирующей компоненты. Следует отметить, что при $H/\Delta x \leq 0,5$ значения Δ быстро растут и, начиная с некоторого момента, получаемые оценки точности вычисления Q могут стать хуже статических оценок. Это вполне правомерно. Здесь отражается тот факт, что использование более точных способов оценивания требует и более надежных исходных данных. При подсчете промышленных запасов высоких категорий таких ситуаций быть не должно.

§ 2. Метод подсчета запасов на ЭВМ

В § 1 гл. IV получены оценки точности определения величины Q , справедливые в том случае, когда мы можем точно восстановить функцию $\bar{\varphi}$ или $\bar{\varphi}_n$. Однако по соображениям, рассмотренным в § 5 гл. I, использовать в этих целях ряд Котельникова или ряды аналогичного типа, но для $\Delta x \neq \text{const}$, не представляется возможным. В то же время изложенный там метод позволяет производить восстановление значений $\bar{\varphi}$ с весьма малой ошибкой δ_2 , причем $\delta_2 \rightarrow 0$, при $z_1/\Delta x \rightarrow \infty$, где z_1 — параметр вычислительной схемы, т. е. является асимптотически оптимальным. Естественно предположить, что при подсчете запасов он также будет асимптотически оптимальным*. Покажем это для $\Delta x = \text{const}$ и $U = \infty$.

Полагая в (I.99) $\omega = 0$, получаем

$$\frac{1}{\pi} \sum_{-\infty}^{\infty} \alpha_k = \frac{1 + \frac{2e^{-2\sigma H}}{1 - e^{-2\sigma H}}}{1 + \frac{2e^{-2\sigma z_1}}{1 - e^{-2\sigma z_1}}}. \quad (\text{IV.26})$$

При $z_1 \rightarrow \infty$ знаменатель в (IV.26) стремится к 1 и, следовательно,

$$\lim_{z_1 \rightarrow \infty} \sum_{-\infty}^{\infty} \alpha_k = \pi \left(1 + \frac{2e^{-2\sigma H}}{1 - e^{-2\sigma H}} \right), \quad (\text{IV.27})$$

$$|\bar{\Delta}| = 2\pi \frac{e^{-2\sigma H}}{1 - e^{-2\sigma H}}, \quad (\text{IV.28})$$

т. е. полученное значение $\bar{\Delta}$ совпадает с оптимальным значением Δ в (IV.13): $\bar{\Delta} = \Delta$, $z_1 \rightarrow \infty$.

При проведении расчетов значение z_1 обычно принимается равным $(1 \div 2)\Delta x$. Величины $\bar{\mu}$ при некоторых конечных значениях z_1 представлены в табл. 9.

Таким образом, уже при $z_1 = \Delta x$ можно считать, что $\bar{\Delta} = \Delta$. Подобным образом можно получить оценки и для двухмерного случая:

$$\frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \alpha_{p,q} = \frac{\sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} e^{-2H\sigma \sqrt{p^2+q^2}}}{\sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} e^{-2z_1\sigma \sqrt{p^2+q^2}}}. \quad (\text{IV.29})$$

* В несколько иной реализации метод уже использовался при подсчете запасов железорудных и медноколчеданных месторождений [82], однако вопросы теоретической оценки точности в этой работе не рассматривались.

Таблица 9

Значения относительных ошибок $\bar{\mu}$ определения величины Q (одномерный случай)

$z_1/\Delta x$	$H/\Delta x$		
	0.5	0.75	1.0
0,1	0,67	0,68	0,69
0,5	0,00	0,06	0,08
1,0	0,09	0,02	0,00
∞	0,09	0,02	0,00

Как и в одномерном случае, при $z_1 \rightarrow \infty$ полученное значение совпадает с оптимальным значением Δ в формуле (IV.17). В табл. 10 представлены значения $\bar{\mu}$ при $z_1=0,5; 1,0; \infty$.

Таблица 10

Значения относительных ошибок $\bar{\mu}$ определения величины Q (двухмерный случай)

$z_1/\Delta x$	$H/\Delta x$		
	0.5	0.75	1.0
0,5	0,00	0,16	0,19
1,0	0,23	0,03	0,00
∞	0,24	0,04	0,01

Таким образом, и в этом случае можно ограничиться значением $z_1/\Delta x=1$, а следовательно, метод действительно обеспечивает асимптотически оптимальное решение поставленной задачи.

В реальном случае, вследствие ограниченности размеров залежи и сложности контура нефтегазоносности, непосредственное вычисление запасов по значениям $a_{p,q} = \bar{\varphi}(x_p, y_q, z_1)$ нерацонально. Дело в том, что дисперсия $a_{p,q}$ может быть довольно большой, во всяком случае она значительно больше дисперсии значений $\bar{\varphi}(x, y)$, т. е. значений интерполирующей функции на исходной плоскости $z_1=0$. Кроме того, значения $a_{p,q}$ могут быть не равны нулю и вне $(-d, d)$. Учитывая, однако, что функция $\bar{\varphi}(x, y, z)$ — гармоническая и, следовательно,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\varphi}(x, y, z) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\varphi}(x, y, 0) dx dy, \quad (\text{IV.30})$$

полученные результаты об оптимальности расчетов по $a_{p,q}$ полностью переносятся на расчеты с помощью функции $\bar{\varphi}(x, y)$, дисперсия которой равна дисперсии соответствующей функции $\bar{\varphi}$.

Итак, процесс подсчета запасов сводится к следующей про-

цедуре. По исходным значениям φ_i определяются значения $a_{p,q}$, с помощью которых легко вычисляется достаточно густая (для точного учета границ залежи) регулярная (квадратная) сеть значений $\varphi(x, y)$, которые необходимо просто суммировать (по границам залежи — с весами, определенными пропорционально «полезной» площади каждой элементарной площадки).

§ 3. Фильтрация случайных ошибок

Эта операция при подсчете запасов довольно своеобразна. Ошибки измерений (аппаратурные) обычно невелики, во всяком случае обычно это не главная помеха по сравнению с изменчивостью признаков по площади. Помехой является в данном случае высокочастотная составляющая полезного сигнала. Ясно, что если признак на базе Δx испытывает незакономерные флуктуации, то эту колеблющуюся составляющую мы практически восстановить не можем. Возможность уменьшения шага Δx также сильно ограничена быстрым возрастанием стоимости разведочных работ. В то же время при интерполяции эти незакономерные, непредставительные характеристики признаков могут сильно исказить результаты. Поэтому необходимо по возможности уменьшить их влияние, иначе преимущества рассматриваемого здесь метода подсчета запасов перед расчетами по стандартной объемной формуле могут стать иллюзорными. Однако поскольку характеристика этой непредставительной части изменчивости признаков может быть известна лишь очень приближенно, использование «строгих» методов фильтрации, основанных на достаточно хорошем знании свойств помех, здесь вряд ли будет эффективным. Во всяком случае неточное их определение будет сводить к нулю «оптимальность» подобных фильтров, например фильтра Колмогорова — Винера. Последний, к тому же, характеризуется возникновением систематического расхождения между «теоретической» и аппроксимирующей функциями, что в случае задачи подсчета запасов особенно опасно.

При подсчете запасов рудных месторождений, где число измерений обычно велико, можно сначала производить преобразование, аналогичное тому, которое решается задачей «крайгинга», но более простое [67]: регуляризация исходных данных с помощью перехода к средним значениям признака объема v' , $v' > v$, где v — объем, характеризующий исходным значением признака φ , т. е.

$$\varphi'_i = \frac{1}{v'} \int \varphi_i(x, y, z) dx dy dz. \quad (\text{IV.31})$$

Благодаря усреднению исходных данных получаемые переменные φ' будут иметь значительно более закономерный характер, чем исходные признаки φ , причем тем более закономерный,

чем больше величина v' , S или l . Такая трансформация исходных данных позволяет построить удобные для рассмотрения карты и выделить отдельные блоки. Она является очень устойчивой, главное здесь заключается в определении оптимальных значений v' , S или l . Действительно операция вычисления скользящего среднего, к которой сводится вычисление интеграла (IV.31) при $\Delta x = \text{const}$, эквивалентна (в первом приближении) фильтрации с помощью фильтра Колмогорова — Винера [10], но при значениях v' , S или l , выбранных соответствующим образом, — на основе знания некоторых свойств полезного сигнала и помехи (значений радиусов автокорреляции и средних квадратов амплитуд). Поскольку эти свойства нам неизвестны, расчеты проводятся последовательно для все увеличивающейся базы осреднения, а наилучшие результаты выбираются качественным образом — на основе опыта и интуиции геолога, в зависимости от данной конкретной ситуации. С этих позиций введенная Матероном в рассмотрение задача оптимальной оценки среднего значения в блоке по совокупности наблюдений внутри и за пределами блока — крайгинг — имеет очень нечеткое содержание. Если блоки не выделены, то нам в первую очередь неизвестны или, вернее, плохо известны даже необходимые значения базы осреднения. Если они уже выделены, то какое значение могут иметь замеры вне этих блоков? При неравномерном опробовании технология крайгинга и осреднения резко усложняется и практически невозможна без проведения предварительной интерполяции. Очевидно, «оптимальное» вычисление «крайгинга», рассматриваемое в ряде работ, может помочь нам ликвидировать ошибки второго порядка малости, полностью оставив ошибки первого. Другое дело — вычисление оптимального «крайгинга» на всей территории залежи, но тут мы приходим к первоначальной постановке задачи оптимального подсчета запасов.

При подсчете промышленных запасов нефти и газа количество скважин измеряется обычно не более чем первыми десятками, уже один этот факт затрудняет предварительное проведение простого осреднения по площади*, кроме того, в этих условиях существенно сказывается нерегулярность сети исходных данных. Это особенно остро ставит вопрос о фильтрации случайных ошибок измерений и непредставительных характеристик признаков непосредственно в пунктах измерений одновременно с решением задачи интерполяции. Следует отметить, что последнее замечание является очень важным, так как при интерполяции на регулярную сеть без фильтрации влияние «помех» может возрасти во много раз, а их радиусы «автокорреляции» будут соответствовать особенностям размещения на площади исходной

* Как уже указывалось, речь идет о средневзвешенных по мощности значениях признаков φ_i .

сети данных, которые на регулярной сети будут утеряны. Во многих случаях положение при подсчете запасов твердых ископаемых вполне аналогично.

Сформулированным требованиям вполне отвечает метод фильтрации, изложенный в § 3 гл. II, совмещенный с только что рассмотренным методом подсчета запасов. Здесь весьма важно то обстоятельство, что в отличие от оптимальной фильтрации систематические искажения сигнала несущественны: $\sum_i \delta_i^{(k)} \approx 0$ и, следовательно, величина запасов будет претерпевать значительно меньшие искажения. Кроме того, метод легко допускает возможность учета различия величины ϵ_0 в отдельных пунктах даже при нерегулярной сети наблюдений — в зависимости от различных предположений об их качестве, характере поведения признаков, площади залежи, относящейся к нему, и т. п.

В случае когда о величине ϵ_0 нет достаточно обоснованных предположений, можно начинать с малых значений ϵ_0 , последовательно увеличивая их и следя одновременно за поведением величины $\delta_\Sigma = \sum_i \delta_i^{(k)}$ до тех пор, пока эта величина не станет существенно отличной от нуля. Фактически величиной $\sup_{\epsilon_0} \delta_\Sigma$ при условии $\sum_i p_i \delta_i^{(k)} \approx 0$ (p_i — «вес» i -ого измерения) и будет оцениваться влияние случайных компонент признака ϕ при вычислении запаса Q . Доверительные интервалы этой величины легко определить по дисперсии $\delta_i^{(k)}$. Таким образом, влияние незакономерной компоненты ($\tau_0 \approx \Delta x$) на вычисление Q при таком подходе будет минимальным и в то же время все вместе позволяет «точно» вычислять тренд.

При наличии «ураганных» или просто слишком больших значений признаков, которые обычно присущи рудным залежам, можно использовать модификацию метода, позволяющую эффективно фильтровать и такого рода помехи (§ 3, гл. II).

Изложенный способ оценки влияния случайной составляющей, естественно, в большинстве случаев будет давать существенно завышенные значения, поскольку дисперсия $\delta_i^{(k)}$ близка к дисперсии исходной случайной составляющей до проведения фильтрации. Если нам априори известны квадраты амплитуд «полезного» сигнала и «помехи», а также их радиусы автокорреляции, уровень «помех» после фильтрации может быть приближенно определен на основе использования известного выражения для теоретической ошибки фильтра Колмогорова — Винера или с помощью построения и исследования соответствующей теоретической модели. В последнем случае одновременно могут быть учтены особенности распределения замеров по площади и одновременно изучены возможности используемого метода интерполяции, в частности, предлагаемого в данной ра-

боте. Возможно, этот способ окажется наиболее эффективным из всех возможных способов оценивания точности подсчета величины запасов и одновременно проектирования разведочных сетей. По сути дела, этот подход очень близок к широко применяемому методу моделирования при расчете оптимальных сетей для добычи. Расчеты, естественно, следует повторять по мере поступления новых результатов наблюдений, последовательно уточняя параметры модели.

§ 4. Ошибки подсчета запасов по объемной формуле

Как уже указывалось выше, расчеты по объемной формуле даже при точном определении средних значений φ_j могут содержать существенные ошибки. Рассмотрим результаты исследования ошибок объемной формулы на теоретических примерах, полученные в работе [71] на примере залежей нефти.

При подсчете запасов объемным методом используется следующая формула:

$$Q_2 = S \cdot \bar{h} \cdot \bar{m} \cdot \bar{\beta} \cdot \eta \cdot \rho, \quad (IV.32)$$

где S — площадь залежи; \bar{h} — средняя нефтенасыщенная мощность; \bar{m} — средний коэффициент открытой пористости; $\bar{\beta}$ — средний коэффициент нефтенасыщенности; ρ — средняя плотность нефти в пластовых условиях; η — коэффициент нефтеотдачи, принятый в данном случае постоянным, $\rho = \text{const}$, $\eta = \text{const}$.

Средние значения признаков в формуле (IV.32) вычислялись следующим образом:

$$\bar{\varphi}_j = \frac{1}{V} \iiint_V \varphi_j(x, y, z) dV, \quad (IV.33)$$

где φ_j — признаки h , m и β .

Таким образом, вычисление интеграла (IV.1) от произведения функций m , β и ρ заменяется произведением их средних значений. Ясно, что можно подобрать такие функции φ_j , что ошибка будет сколь угодно велика. Практически, учитывая ограниченные пределы изменения их реальных значений, ошибка может достигать величины в несколько десятков процентов.

Рассмотрим теоретические примеры, в которых принято $\rho = \text{const} = 0,8 \text{ г/см}^3$, $\eta = 40\%$; изменяется коэффициент открытой пористости в пределах $5\% \leq m \leq 30\%$ и коэффициент нефтенасыщенности $5\% \leq \beta \leq 30\%$. Изменение m и β происходит в направлении оси Ox , $-5 \text{ км} \leq x \leq 5 \text{ км}$ по формулам, указанным в табл. 11.

Пример 1. Залежь имеет форму прямоугольного параллелепипеда длиной 20 км, шириной 10 км, высотой 30 м, $V = 6 \text{ км}^3$.

Пример 2. Залежь имеет форму треугольной призмы высотой 30 м, шириной 10 км, длиной 20 км, $V = 3 \text{ км}^3$.

Пример 3. Залечь имеет форму пласта, ограниченного двумя полуокружностями: сверху $R_1 = A(B^2 - x^2)$, $B = 5$ км, $A = 0,3 \cdot 10^{-2}$ км; снизу $R_2 = A(D^2 - x^2)$, $D = 4$ км; мощность пласта 30 м, длина залежи 10 км, $V = 5,42$ км³.

Т а б л и ц а 11

Ошибки подсчета запасов по объемной формуле для моделей 1, 2, 3

Пример	m	β	\bar{m}	$\bar{\beta}$	Q_1	Q_2	Δ	μ	
1	a	$x^2 + 5$	$0,2 x ^3 + 5$	13,0	22,5	38,8	28,8	10,0	26
	b	$150/30 - x^2$	$30 - x^2$	8,5	22,0	9,0	11,2	2,2	24
	c	$x^2 + 5$	$30 - 0,2 x ^3$	13,0	23,8	78,8	98,8	20,0	25
2	$x^2 + 5$	$0,2 x ^3 + 5$	9,2	15,0	13,2	16,9	3,7	22	
3	$0,16 x ^3 + 10$	$x^2 + 5$	13,8	11,9	35,2	28,3	6,9	20	

Результаты расчетов приведены в табл. 11, где приняты следующие обозначения:

Q_1 — запасы, подсчитанные по формуле (IV.1),

Q_2 — запасы, подсчитанные по формуле (IV.32),

$\Delta = |Q_1 - Q_2|$ в тыс. т, $\mu = \frac{\Delta}{Q_1}$.

Абсолютная ошибка слабо зависит от формы залежи, но прямо пропорциональна степени отклонения m и β от константы. Относительная ошибка обратно пропорциональна величине постоянных составляющих в уравнениях для m и β . Это полностью соответствует выводам из статистической трактовки объемной формулы: она верна, когда признаки φ_j , входящие в нее, независимы. Чем сильнее корреляционная связь между ними, тем больше ошибка.

Следует еще раз подчеркнуть, что полученные результаты соответствуют тому случаю, когда средние значения \bar{m} , $\bar{\beta}$ известны точно. Ошибки в их определении, естественно, приводят к увеличению ошибок величины Q .

Следует отметить, что и эти величины ошибок достаточно велики. Так, запасы категории A должны иметь ошибку $\pm 5\%$, категории B $\pm 15\%$, категории C $\pm 35\%$ [90]. В работе [90] формулируются количественные требования к результатам разведки, причем точность определения параметров залежей нефти при составлении комплексной схемы и особенно проекта разработки не менее велика и определяется величиной 5—10%.

Все это подтверждает необходимость перехода к расчетам по формуле (IV.1) в системах автоматизированного подсчета запасов на ЭВМ.

Подсчет запасов на ЭВМ уже вступил в стадию промышленного использования. Например, в Казахстанской опытной экспедиции выполняются большие объемы работ по подсчету запасов твердых полезных ископаемых на основе использования метода сеток [88]. В то же время ряд теоретических и методических вопросов и в первую очередь вопрос оценки точности подсчета запасов исследован недостаточно. Неясным, например, остается вопрос о порядке подсчета запасов. Речь идет о том, вычислять ли значение Q сразу по удельным запасам $\varphi = \sum \varphi_j$ или получать сначала регулярную сеть значений φ_j и лишь затем вычислять значения φ и Q . Здесь следует заметить, что о свойствах φ_j мы обычно имеем некоторое представление, тогда как свойства φ зависят от каждого φ_j и, следовательно, должны отражать влияние самого сложного из них. Практика расчетов показывает, что результаты могут существенно различаться. Это связано с неаддитивностью (в смысле точности) процессов интерполяции и фильтрации, с коррелированностью самих функций φ_j и их случайных компонент, что оказывает непосредственное влияние на «сложность» функции φ , а также возможность использования различного количества замеров при расчетах по каждой функции φ_j отдельно. Последнее, очевидно, говорит в пользу второго способа, так же как и то, что о свойствах φ_j мы имеем значительно более четкое представление, однако, в принципе, ввиду коррелированности φ_j , в некоторых случаях (очевидно, довольно редких) первый способ может дать более точный результат. Вопрос требует детального рассмотрения.

Судя по оценкам точности, полученным в § 2 и 5 гл. I и § 2 гл. IV, при подсчете запасов целесообразно использовать метод, основанный на приближении системой гармонических функций. Это следует не только потому, что он является асимптотически оптимальным в смысле точности, но и потому, что при его использовании имеется возможность оценивать точность расчетов.

Рассмотренный в настоящей главе метод оценки точности подсчета запасов основан на детерминированном подходе. Однако, как и при решении задачи интерполяции, полученные результаты можно трактовать и в терминах теории случайных функций, аппроксимируя исходный сигнал случайными функциями с радиусом автокорреляции τ_0 (связь которого с величинами N функций $\varphi \in U'(H)$ указана выше) и нормированной функцией автокорреляции типа (I.28).

Реализация программ обработки геологических данных на ЭВМ — весьма сложная и трудоемкая задача. Здесь сказывается множество факторов: отсутствие опыта и недостаток квалифицированных кадров, быстрая смена и разнообразие типов и поколений машин, их непрерывно увеличивающиеся количественные и качественные возможности, недостаточная формализованность казалось бы уже полностью разработанных алгоритмов. Главным фактором является большая сложность и отсутствие теории построения необходимых комплексов программ для ЭВМ, систем обработки данных, которые позволяли бы решать разнообразные классы геологических задач, таких как задачи статистической обработки, интерполяции и аппроксимации, фильтрации и тренд-анализа, прогноза и распознавания образов, подсчета запасов, автоматизации графических построений, обработки различных измерений — спектральных, химических, литологических, геофизических, хранения, поиска и выдачи различных геологических данных и др.

Большинство разработанных к настоящему времени программ для отечественных ЭВМ рассчитано на решение отдельных небольших задач, они требуют собственных стандартов подготовки данных и не могут быть использованы в комплексе. Этим во многом объясняется большой параллелизм, дублирование в их составлении. Небольшую программу проще составить заново, чем приспособляться к стандартам и возможностям уже разработанных. Еще со второй половины 60-х годов стало ясно, что простейшие программы, рассчитанные на решение отдельных задач, не совместимые информационно и не объединенные единой системой управления, не могут быть эффективными, не отвечают современным требованиям и эксплуатационным возможностям ЭВМ. Имеющийся опыт применения ЭВМ в геологии и особенно в геофизике свидетельствует о необходимости создания комплексов программ — автоматизированных систем обработки геологических данных (АСОД). Концентрация работ при создании систем позволяет существенно повысить качество программ и используемых алгоритмов.

Разработка АСОД включает в себя решение следующих основных задач:

1. Стандартизация форм сбора, кодирования и хранения данных.
2. Автоматизация сбора, хранения и выдачи данных.

3. Разработка методов математической обработки данных на ЭВМ.

4. Создание комплексов программ обработки данных и решения типовых задач.

Решение каждой из них связано с преодолением определенных трудностей, осложняющихся еще и тем, что, несмотря на очевидную взаимосвязанность задач, решение приходится проводить параллельно. Кроме того, многие вопросы здесь зависят от уровня используемой техники, которая в этой области стремительно изменяется.

Общее количество геологических данных настолько велико, что организация одновременного хранения и поиска ее остается пока неразрешимой проблемой и при использовании ЭВМ. Кроме того, сам сбор всей информации занял бы слишком много времени. Поэтому целесообразно сначала создать ряд информационно-поисковых систем по важнейшим геологическим объектам, объединяющим комплексы взаимосвязанных геологических признаков, рассчитанных на достаточно широкий круг пользователей. Это могут быть системы, включающие в себя как первичные признаки, так и вторичные, получаемые в результате определенной переработки результатов наблюдений.

Рассмотрим в качестве примера информационно-поисковую систему по залежам нефти и газа. Месторождения и залежи нефти и газа характеризуются значительным набором геологических, физических, геохимических, промыслово-геофизических и экономических признаков, количество которых измеряется сотнями. Кроме того, точность определения и сами значения многих признаков изменяются со временем. Признаки обычно подразделяются на качественные и количественные. К качественным относятся данные о приуроченности к территориально-административным единицам, положение на схемах геолого-тектонического районирования, стратиграфическая характеристика, тип залежи в соответствии с генетической, морфологической или иной классификационной схемой, тип коллектора и ловушки, режим залежи, тип подземных вод и др. Многие из перечисленных признаков не имеют строгого определения. Это обусловлено отсутствием логически хорошо обоснованных и общепризнанных классификаций. Приходится выбирать одну из возможностей — предусмотреть включение в ИПС всех известных классификаций, выбрать из них только некоторые или даже одну (последний вариант наиболее реальный), разработать специальные классификации, удобные для функционирования ИПС. Следует отметить, что при использовании автоматизированных ИПС роль некоторых типов априорных классификаций резко снижается. Во многих случаях можно производить поиск информации не по заранее введенной классификации, а задаваясь непосредственно нужным сочетанием значений элементар-

ных характеристик. Во всяком случае момент увеличения времени поиска не имеет значения, как в обычных условиях.

К количественным признакам относятся данные: о глубинах залегания кровли и подошвы продуктивных пластов, дебитах скважин, характеристики коллекторов и покрышек—пористость, проницаемость и прочие, характеристики ловушки, пластовых условий, о физико-химических свойствах нефтей—плотность, вязкость, содержание серы, смол, асфальтенов, парафина и другие, характеристики группового углеводородного состава нефти по фракциям, физико-химические свойства газов, пластовой воды, сведения о запасах и добыче и т. д. Например, в макете, принятом во ВНИГНИ, содержится свыше 200 признаков. Объем информации по всем залежам нефти и газа СССР в этом случае составит 2—3 млн. чисел, что делает возможным функционирование автоматизированной ИПС на машинах второго поколения БЭСМ-4 и Минск-32 и особенно на машинах серии ЕС. В то же время рассмотрение вопросов создания массива данных по залежам свидетельствует об огромной величине массива всех геологических фактографических данных, об огромной сложности построения информационской модели геологического процесса и трудности организации сбора весьма разнообразных данных, получаемых различными организациями в разное время.

Очевидно, рациональным является построение ИПС разного уровня, построенных по иерархическому принципу. В этом случае ИПС высшего уровня должны пополняться данными на основе ИПС более низкого уровня с помощью специальных программ переработки данных.

Автоматизированная система поисковой и статистической обработки геологических данных ВНИГНИ. Первая в отечественной практике система обработки геологических данных была реализована во ВНИГНИ в 1969 г. на ЭВМ БЭСМ-4. Ее основной принцип—единый стандарт подготовки данных для всех обрабатываемых программ, базирующийся на специальной подсистеме информационного обслуживания—автоматизированной информационно-поисковой системе. Система является универсальной. Она может быть использована при обработке любых данных—по залежам и месторождениям нефти и газа, скважинам, литологическим, физическим свойствам горных пород, на описание и возможности обработки которых накладываются некоторые общие требования и ограничения. Подробное описание системы программ дано в работе [12]. Блок-схема системы изображена на рис. 29.

Алгоритм информационно-поисковой системы. Проблема хранения и поиска данных на ЭВМ состоит из следующих основных частей:

- 1) кодирование данных;
- 2) запись и хранение данных на запоминающих устройствах ЭВМ (магнитных лентах—МЛ);

3) выделение, поиск необходимых подмассивов данных и выдача их на печать в форме, удобной для чтения, или запись в память машины для проведения дальнейших расчетов на ЭВМ.

Подготовка исходных данных. Массивы фактического материала по различным геологическим объектам (месторождениям,

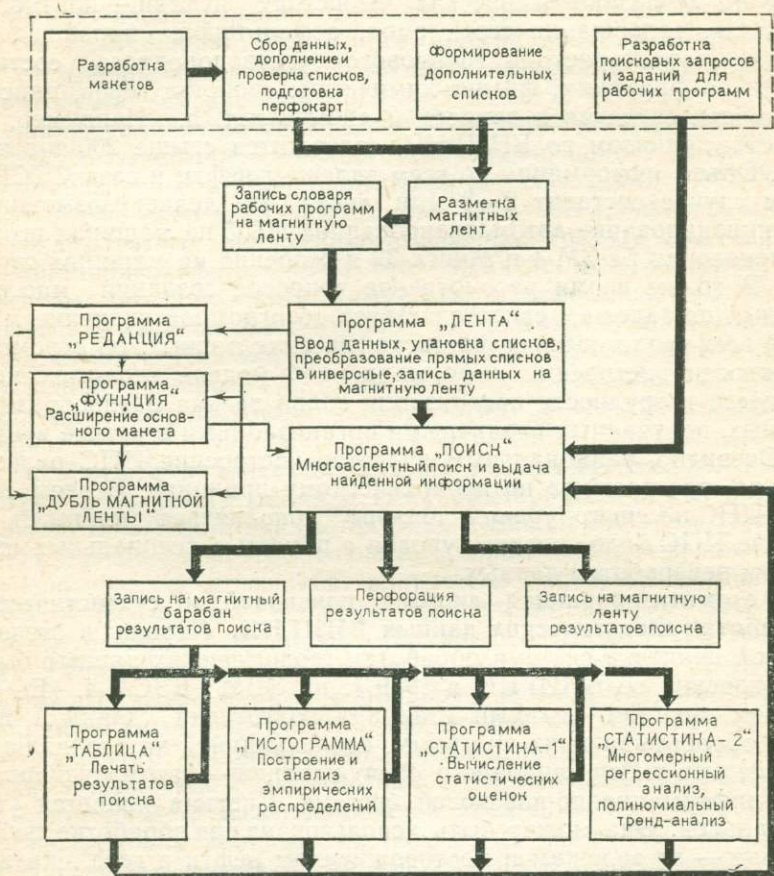


Рис. 29. Автоматизированная система поисковой и статистической обработки данных (ВНИГНИ)

залежам, скважинам, образцам пород и т. п.) создаются на базе предварительно разрабатываемых макетов. Макет представляет собой упорядоченный перечень наименований признаков p_1, p_2, \dots, p_m , характеризующих объект исследования и удовлетворяющих требованиям системы. На макет накладываются следующие ограничения:

- 1) число признаков m должно быть менее 1000, $m < 1000$;
- 2) первые 128 признаков могут быть качественными, $1 \leq m_a \leq 128$;
- 3) предусмотрено цифровое кодирование качественных признаков; причем максимальное количество цифр в коде не более семи.

Значения признаков для каждого объекта выписываются на бланки для ЭВМ в последовательности, определяемой макетом. Сначала располагаются качественные признаки. Коды качественных признаков записываются в конце строки бланка. Вслед за ними идут значения количественных признаков в стандартном нормализованном виде. Отсутствие значения пишется как «0», нулевые значения заменяются очень малым числом, например $1 \cdot 10^{-18}$.

Запись и хранение информации. Наиболее простым способом хранения является прямая форма записи, когда сохраняется исходная последовательность расположения признаков. Однако такой способ хранения, во всяком случае на МЛ ЭВМ типа М-20, невыгоден с точки зрения организации поиска — при ответе почти на любой запрос необходимо просматривать весь массив данных. Поэтому в АСОД-ВНИГНИ был принят инверсный способ хранения данных, при котором в одной или нескольких зонах хранятся списки по отдельным признакам. В этом случае расположение значений признаков по различным объектам для всех признаков одинаково и соответствует порядку списков при прямой записи, т. е. сохраняется взаимно-однозначное соответствие между номерами объектов и адресами значений признаков в зоне МЛ. При инверсной записи поиск можно сделать целенаправленным, просматривая массивы только в тех зонах, где содержатся признаки, по которым проводится поиск, и зоны, содержащие искомые признаки. Дополнительным удобством является возможность простой реализации поиска по признакам, являющимся функциями одного или нескольких исходных. С помощью программы «Функция» значения этих функций легко рассчитываются и записываются на МЛ в качестве новых признаков, по которым затем и осуществляется поиск.

Рассмотрим возможности «уплотнения» хранящихся массивов, т. е. рациональность хранения нескольких кодов в одной ячейке. Тот факт, что коды адресных признаков обычно состоят из одной, двух, реже трех цифр, создает возможность эффективного уплотнения их. Например, вся адресная часть упомянутого списка по залежам — 22 признака — уместается в четырех ячейках памяти. Соответственно этому при инверсном способе записи сокращается до четырех количество зон, содержащих адресные признаки. Поскольку наиболее часто выборки проводятся именно по адресным признакам, уплотнение существенно сказывается на скорости поиска, так как можно вести поиск сразу по всем признакам, хранящимся в одной зоне.

k_q, k_s, k_t — количество граничных пар для признаков p_q, p_s, p_t .

Признаки, по которым ведется поиск, и искомые признаки могут быть как качественными, так и количественными.

Организация поиска основана на реализации двух логических операций — логического умножения « \wedge » и логического сложения « \vee ».

В первом случае будет выбрано такое подмножество множества P , в точках которого значения каждого из признаков с номером q, s, t, \dots отвечают запросу. Во втором случае будут отбираться объекты, в которых значение хотя бы одного из признаков p_q, p_s, p_t, \dots соответствует запросу. Возможны и совместные комбинации этих вариантов. Отсюда ясно удобство инверсной записи и алгоритм поиска — просматриваются только те зоны, в которых записаны признаки, по значениям которых производится поиск p_q, p_s, p_t, \dots .

При запросе по схеме $p_q \wedge p_s \wedge p_t, \dots$ искомое подмножество должно быть пересечением таких множеств точек, каждое из которых является ответом на один из запросов p_q, p_s, p_t, \dots . Считываются зоны, содержащие признак p_q . Адреса точек, в которых значения p_q лежат в каком-либо из соответствующих запросу интервалов, запоминаются. Затем считывается признак p_s . При сравнении значений признака p_s с заданными просматриваются только те точки, адреса которых были найдены при поиске по признаку p_q , и т. д.

Работа по запросу $p_q \vee p_s \vee p_t, \dots$ отличается тем, что при поиске по каждому следующему признаку ранее найденные точки заново не просматриваются. Искомое подмножество есть сумма множеств точек, найденных по каждому из запросов p_q, p_s, p_t, \dots . В заключение проводится считывание зон, содержащих j -ые признаки, которые являлись целью поиска, и отбираются их значения в точках с ранее найденными адресами.

Поиск объектов с заданными значениями признаков осуществляется по рассмотренному алгоритму программой «Поиск». Исходными данными для программы служат массивы данных, записанные на МЛ, значения некоторых их параметров, выдаваемых программой «Лента», и запросы (задания) самой программы «Поиск». Запросы составляются в закодированном виде, однако форма их составления очень проста и удобна. Чтобы система функционировала, необходимо выполнение следующих неравенств одновременно:

$$\begin{cases} n \leq 118\,000 \\ m < 1000 \\ m \times n \leq 3 \times 10^6 \end{cases}$$

Эти ограничения связаны с возможностями памяти ЭВМ БЭСМ-4.

Результаты поиска могут быть выданы на печать, записаны на магнитный барабан (МБ) или свободную МЛ. Конкретный способ хранения результатов поиска зависит от объема получаемых данных и от целей их последующего использования.

С 1969 г. система эксплуатировалась во ВНИГНИ и затем в ряде других организаций. Число объектов по массивам изменялось в пределах от 100 до 10 000, число признаков — до 200. Обычно число поисковых запросов составляло несколько десятков, иногда до 200, по каждому из которых производилась статистическая обработка. Среднее время поиска по одному запросу не превышает минуты. Наибольший обрабатываемый массив из известных автору составил 3×10^5 кодов*. Время выборки по запросу — не более 3—5 мин. Отсюда ясно, что функционирование рассмотренной выше ИПС по залежам нефти и газа в данной автоматизированной системе является вполне реальным и эффективным. Среднее время поиска не превышает 15—20 мин.

Интересно отметить простую возможность использования «Поиска» для отыскания грубых ошибок в подготовке данных. Для этого составляются поисковые задания, где значения признаков лежат вне реального интервала их значений.

Статистическая обработка геологических данных. Решение многих геологических задач сводится к определению тесноты связи одного признака с совокупным влиянием нескольких других, определению уравнений регрессии, позволяющих наилучшим образом (обычно в смысле среднеквадратического приближения) аппроксимировать один из признаков по значениям других. При этом обычно тип зависимости и возможная степень влияния каждого признака при многомерном анализе заранее неизвестны. Точнее, именно это и представляет интерес. Обычно желательно получить достаточно тесную связь при возможно меньшем количестве признаков. Все это приводит к необходимости изучать всевозможные сочетания признаков и типов связей, к необходимости выполнения полного перебора (из M переменных по t). Однако большинство программ для многомерного регрессионного анализа имеют существенные ограничения, которые в основном сводятся к следующему:

- 1) можно оперировать лишь с исходными признаками или с функциями от них, вычисленными предварительно;
- 2) для каждого объекта должны быть известны значения всех исследуемых признаков;
- 3) программы не предусматривают какой-либо сортировки исходного материала — ни по качественным, ни по количественным характеристикам. Изменение исходной совокупности объектов и признаков производится вручную;

* Это массив по коллекторским свойствам осадочных пород (ВНИГНИ). Достоверных сведений о величине массивов, обрабатывающихся в других организациях, у автора нет.

4) малая величина m и n ;

5) в лучшем случае предусматриваются ограниченные переборы сочетаний признаков.

Все это создает значительные неудобства при использовании этих программ: производится многократный ввод исходной информации в оперативную память ЭВМ через читающее устройство, ручная сортировка исходного материала связана с многократным перфорированием, отсутствие полного перебора всевозможных сочетаний признаков не позволяет провести достаточно полный анализ имеющегося материала.

Количество всех сочетаний при большом M велико, поэтому в некоторых программах предусматривается следующее. Сначала рассматриваются парные связи и производится расстановка признаков x_j в ранжированный по величине s_{yx_j} или R_{yx_j} ряд:

$$x_1, \quad , x_2 \quad , x_3, \dots \quad (V.1)$$

$$R_{yx_1} > R_{yx_2} > R_{yx_3} > \dots$$

Далее рассчитываются уравнения регрессии типа

$$y = f(x_1, x_2), \quad (V.2)$$

$$y = f(x_1, x_2, x_3)$$

и т. д. — до получения заданной величины s_{yx} или R_{yx} . Однако! поскольку коэффициенты R_{yx_1} и R_{yx_2} — наиболее высокие, имеется большая вероятность сравнительно высокого значения $R_{x_1x_2}$. Поэтому добавление x_2 , возможно, даст меньше информации, чем добавление x_j , для которого $R_{yx_j} < R_{yx_2}$.

Рассмотрим еще один аналогичный алгоритм. После парной ранжировки рассчитываются уравнения регрессии типа

$$y = f(x_1, x'_2), \quad (V.3)$$

где x'_j — все признаки, за исключением x_1 .

Затем проводится повторная ранжировка до получения заданной величины R_{yx} или S_{yx} . В уравнении (V.3) признак x_1 фиксирован, другие сочетания признаков по два (без x_1) не рассматриваются только на том основании, что R_{yx_1} имеет максимальную величину. Такой подход ничем не обоснован, так как из условия $R_{yx_1} > R_{yx_j}$ не следует, что $R_{yx_1x'_j} > R_{yx_jx'_k}$.

Следует отметить, что построение уравнений регрессии и вычисление коэффициентов корреляции при величине t более четырех-пяти мало эффективно. Системы нормальных уравнений, как правило, становятся плохо обусловленными, коэффициенты получаемых уравнений регрессии сильно зависят от объема выборки и точности измерений. Если ограничиться указанными значениями, число сочетаний при полном переборе бу-

дет не так уж велико. Например, сумма сочетаний

$$\sum_{t=1}^5 C_5^t = 31; \quad \sum_{t=1}^4 C_{10}^t = 300.$$

Системы уравнений четвертого, пятого порядка решаются быстро, реализация полного перебора становится реальной. Поэтому в программе «Статистика» предусмотрена возможность полного перебора сочетаний признаков при вычислении многомерных уравнений регрессии и множественных коэффициентов корреляции. Предусмотрено также некоторое сокращение количества сочетаний при переборах. Сначала рассматриваются сочетания C_M^t при максимальном числе признаков t . Если при этом $R_t < R_0$, остальные сочетания признаков, входящих в t , не рассматриваются, поскольку $R_{t-1} \leq R_t$.

Предусматривается возможность работы в комплексе с информационно-поисковой системой, рассмотренной выше. При этом время на подготовку к расчетам по новому заданию (выборку исходных данных) незначительно. Количество объектов определяется числом замеров только тех признаков x_j , которые участвуют в расчетах по данному сочетанию: проводится просмотр признаков x_j по всем объектам и вычеркиваются те объекты, где замеры хотя бы по одному из признаков отсутствуют. Например, при вычислении $R_{x_j x_{j+1}}$ может оказаться n_1 точек, а при вычислении $R_{x_j x_{j+2}}$ — n_2 точек. Величина n_i выдается на печать.

Вычислительная схема программы «Статистика». Система статистической обработки, реализованная в программе «Статистика», включает в себя решение следующих вопросов:

1) вычисление оценок средних значений признака, дисперсии, центральных моментов третьего и четвертого порядка, определение типа распределения (нормальный, логнормальный);

2) вычисление множественных коэффициентов корреляции и линейных уравнений регрессии до 24-го порядка включительно. При этом имеется возможность оперировать не только непосредственно с исходными признаками, но и с их функциями (трансформантами): $\ln y$; y^2 ; y^3 ; $\frac{1}{y}$; $\ln x$; x^2 ; x^3 ; $\frac{1}{x}$, вид которых указывается исследователем. Для выбора наилучшего уравнения связи (в смысле среднеквадратического приближения) программа предусматривает возможность автоматизированного перебора сочетаний признаков и функций от них, начиная с C_M^1 до C_M^t , где $t \leq 23$, M — количество признаков, включая возможные функции от исходных, $M \leq 180$. Для каждого сочетания x_j могут быть согласно заданию рассмотрены соответствующие трансформанты от y . Этим обеспечивается широкий выбор типов зависимостей.

Определяется также величина s_{yx} , характеризующая величину разброса измеренных значений от линии регрессии:

$$s_{yx} = \sqrt{\frac{\sum_j (y_T - y_{\Phi})^2}{n - (p + 1)}}, \quad (V.9)$$

где y_T — значения, получаемые из уравнений регрессии для заданных значений x_j ; y_{Φ} — данные, получаемые из измерений при тех же значениях x_j .

В отличие от коэффициентов корреляции R величина s_{yx} позволяет определить не только наилучший тип связи (по минимальному значению s_{yx}), но и численно характеризует среднеквадратическую ошибку отклонения фактических данных от уравнения регрессии. Сравнение величины s_{yx}^2 с величиной дисперсии y — s_y^2 характеризует степень уменьшения дисперсии при использовании соответствующего уравнения регрессии. Такой анализ нагляднее анализа с использованием коэффициентов корреляции. Коэффициент корреляции может быть весьма значимым, а величина s_{yx} может слабо отличаться от s_y . В этом случае применение уравнений регрессии, например, при прогнозе, построении карт не может быть эффективным. Связь между величинами s_{yx} и R определяется следующим выражением:

$$R_{y; x_1 \dots x_j \dots x_t}^2 = 1 - \frac{s_{y; x_1 \dots x_j \dots x_t}^2}{s_y^2}, \quad (V.10)$$

отсюда

$$s_{yx} = s_y \sqrt{1 - R^2}. \quad (V.11)$$

Например, при $R=0,6$ $s_{yx}=0,8 s_y$.

Абсолютные значения коэффициентов корреляции вычисляются по формуле (V.10). Знак парных коэффициентов корреляции определяется по знаку коэффициента a_1 уравнения (V.8), оценка значимости проводится вручную. Для множественных коэффициентов корреляции используется следующая приближенная формула:

$$s_R = \frac{1 - R^2}{\sqrt{n - (t + 1)}}. \quad (V.12)$$

Результаты расчетов печатаются на АЦПУ с использованием обозначений, принятых в АЛГОЛе. На печать выдаются значения коэффициентов корреляции и коэффициентов уравнений регрессии лишь тогда, когда $R > R_0$, где R_0 определяется исходя из конкретных условий задачи и указывается в задании. В одном из вариантов программы предусмотрена печать лишь определенного количества (не более 10) «наилучших» уравнений.

Программа «Статистика» может работать совместно с программой «Поиск» и без нее. В последнем случае ввод данных производится специальной подпрограммой «Ввод». Совместная работа поисковой системы и «Статистики» возможна в двух вариантах: 1) производится счет по одному заданию к программе «Статистика» для нескольких запросов «Поиска»; 2) производится одна выборка по «Поиску» и результаты выборки подвергаются статистической обработке по нескольким заданиям программы «Статистика».

Программа «Гистограмма». По результатам поискового запроса могут быть построены гистограммы распределения признаков. Они рисуются АЦПУ в таком же виде, как и при ручных способах построения. Одновременно вычисляются и печатаются статистические оценки: моментов до 4-го порядков включительно, асимметрии и эксцесса, минимальное и максимальное значение признака, проверяется, можно ли считать распределение нормальным или логнормальным.

Программа «Таблица». Программа выдает результаты поиска с МБ на печать в виде таблицы. Значения признаков печатаются по 10 столбцов. Наименования располагаются в шапке таблицы.

Для выдачи значений качественных признаков в естественном виде на МЛ хранится макет массива с полным перечислением множества значений всех качественных признаков и их кодов. В принципе это позволяет обойтись без ручной кодировки и при вводе, однако здесь достигается во много раз меньшая экономия труда, чем при многократном выводе, причем кодировщик должен знать всю систему сокращений и принятых в макете названий значений каждого признака. Во много раз увеличивается и работа по перфорации данных на устройствах подготовки перфокарт.

Организация необходимой последовательности проработки программ осуществляется управляющим блоком системы. Он анализирует поисковый запрос и формирует команды передачи управления и чтения с МЛ, обеспечивая возможность автоматического режима проведения расчетов, что особенно важно при проведении многократной проработки обрабатываемых программ.

Требованиям этой системы обработки отвечает и система программ «Карта», рассмотренная в § 5 гл. I. По данным, полученным в результате выполнения поискового задания, может быть проведена фильтрация случайных ошибок, интерполяция на квадратную сеть с дальнейшим построением карты на графопостроителе, проведен тренд-анализ, получены данные для решения задачи подсчета запасов.

За время эксплуатации система АСОД—ВНИГНИ доказала свою эффективность и применимость к решению многих геологических задач. Следует особо отметить, что работа с ней не-

сложна, в результате ее легко осваивают даже специалисты, которые ранее никогда не пользовались ЭВМ.

При разработке системы внимание акцентировалось на создание благоприятных условий для проведения математической обработки данных. Имелось в виду и самостоятельное использование информационно-поисковой подсистемы, но ее главная задача заключалась в подготовке данных для последующей обработки, в обеспечении стандартизации массивов для всех обрабатывающих программ. Опыт показал, что этот подход полностью себя оправдал. В то же время, если рассматривать информационно-поисковую подсистему как самостоятельную систему, цель которой — многостороннее информационное обслуживание, то она, естественно, не удовлетворяет всем предъявленным требованиям. Ее дальнейшее совершенствование может быть направлено на решение следующих вопросов [14].

1. Создание управляющей программы, дающей возможность легко подключать к системе новые обрабатывающие программы и обеспечивающей последовательную автоматическую их обработку.

2. Совершенствование обслуживания: выявление ошибок в первоначальных данных и разносторонние возможности по редактированию данных на МЛ, возможность ввода качественных признаков в алфавитном виде, алфавитно-цифровой вид запроса и т. д.

3. Подключение к системе новых обрабатывающих программ, таких, как программы разделения объектов и распознавания образов, статистической обработки, различных модификаций построения поверхностей тренда, подсчета запасов и др.

4. Выдача результатов обработки в удобной и разнообразной табличной и графической форме.

Современные исследования в области создания систем обработки данных и опыт эксплуатации АСОД—ВНИГНИ позволяют сформулировать некоторые общие требования.

1. Высокоэффективные системы программ для обработки данных должны базироваться на использовании современных вычислительных комплексов, обладающих достаточной памятью, развитым математическим обеспечением и мощными средствами обмена данными между процессором и внешними устройствами, обеспечивающими работу в режиме разделения времени.

2. Структура и принципы управления системой должны обеспечивать проработку программ системы в последовательности, предусматриваемой заказчиком, возможность подключения дополнительных программ и их обновление.

3. Основой любой системы обработки должна быть развитая подсистема информационного обслуживания. Автоматизация хранения и поиска информации выделилась как самостоятельное важное звено и свелось к созданию на ЭВМ автоматизированных ИПС. Сами ИПС должны отвечать ряду требований, та-

ких, как универсальность — возможность использования при решении широкого круга задач и разнообразных макетов сбора исходных данных, компактность хранения данных, полнота и надежность поиска, экономичность, ремонтоспособность.

4. Необходимы единые стандарты данных ИПС и обрабатывающих программ для обеспечения информационной совместимости программ, исходных и выходных данных.

5. Автоматизированная система обработки данных должна быть проста в эксплуатации и доступна геологам.

Рассмотрение более широкого круга задач обработки данных и в особенности задач многостороннего информационного обслуживания приводит к необходимости создания сложных информационно-поисковых систем, рассчитанных на решение целого ряда задач АСУ — «Геология». Первостепенное значение приобретает возможность системы работать с массивами разнообразной структуры, в том числе и с массивами документального типа, а также возможность реализации разнообразных запросов справочно-информационного характера. В то же время, естественно, они могут быть использованы и для подготовки данных для последующей математической обработки. К такого рода системам (конечно, с учетом возможностей, предоставляемых ЭВМ Минск-32, на которые она рассчитана) можно отнести геологическую автоматизированную справочно-информационную систему (ГАСИС) [77]. Каждый документ в этой системе характеризуется наименованием, типом данных, форматом и другими показателями, позволяющими, по мнению авторов, не связывать систему со структурой входных данных. Это позволяет использовать при вводе уже существующие формы документов. Ввод данных осуществляется с автоматическим контролем до 40 видов ошибок. Вывод данных предусматривается в виде стандартных справок или массивов и в специальном виде — в формате пользователя. Предусмотрена запись на МЛ специальных итоговых таблиц поиска и матриц типа объект — свойство. Разработан специальный язык, который используется не только при составлении поисковых заданий, но и позволяет определять способы обмена информацией, записывать и корректировать исходные данные, определять структуры массивов.

Итак, рассмотрены подходы, в которых системы обработки данных базируются на универсальных ИПС. Это позволяет, не накладывая существенных ограничений, выдерживать стандарты в сборе, хранении и подготовке данных (или даже пользоваться уже существующими, традиционными) при решении разнообразных геологических задач, в том числе и справочно-информационных.

Возможно, однако, что это не всегда рационально. При обработке геофизических данных, например, организация хранения массивов и подготовки данных для обрабатывающих программ может быть более простой и удобной. С другой стороны,

здесь выше требования к организации вычислительного процесса, поскольку затраты машинного времени, количество обрабатываемых программ, цепочек их последовательной работы, а также звеньев в них несравненно больше. Основные вопросы заключаются в организации взаимодействия программ и создании условий для экономии труда при программировании. Большое значение имеют вопросы рационального составления программ в смысле экономии машинного времени.

Такого рода специфике, очевидно, отвечает система обработки геофизических данных АСОГ, предложенная в работе [64]. Ее общие принципы, по сути дела, те же: организационной основой АСОГ являются стандарты ввода, хранения и подготовки данных в ЭВМ, однако наряду с этим первостепенное значение придается стандартам составления программ. Эффективная реализация идей работы [64] дала основу для их дальнейшего развития в более широких масштабах — при решении задач АСУ — «Геология» [65, 66]. На основе полученного опыта В. В. Ломтадзе и С. М. Тороповым составлено и реализуется техническое задание на разработку системы оперирования данными для ЕС ЭВМ (СОД/ЕС).

Задачи СОД/ЕС:

1) стандартизация внутреннего информационного обеспечения автоматизированных систем обработки данных и управления предприятиями;

2) унификация математического обеспечения АСУ;

3) упрощение перехода на ЕС ЭВМ программистам, работающим на машинах второго поколения.

Предусматривается составление комплекса обслуживающих программ, позволяющих производить накопление данных на машинных носителях, их контроль, перекомпоновку, сортировку, объединение и разделение массивов, вывод информации на печать в виде таблиц и документов, программ ввода данных с различных специализированных устройств и вывода в виде карт и графиков на графопостроителях, включая программы интерполяции, статистического анализа программы общего назначения. Рассмотрим кратко основные определения и требования СОД/ЕС.

Структуры данных. Наименьшая единица данных — одно значение признака — называется элементом. В памяти ЭВМ элемент данных характеризуется длиной занимаемого поля и типом (числа, символы...).

Упорядоченная совокупность элементов, относящихся к одному объекту, образует запись, которая характеризуется числом образующих ее элементов. Это, например, совокупность всех значений признаков одного объекта.

Совокупность записей одного типа вместе с заголовком образует массив данных. Массивы характеризуются параметрами, полностью определяющими его длину и внутреннюю структуру.

Совокупность родственных массивов, имеющих одинаковую структуру, называется файлом. Структура файла также является стандартной. Это позволяет разработать стандартные процедуры обмена между основной памятью и запоминающими устройствами на МД и МЛ. Файл описывается именем и каталогом, в который входят данные: о числе массивов, количестве элементов в заглавии и в записи, идентификаторе шифра массивов и его длине, типе шифра и т. п. Каталог файла формируется одновременно с созданием самого файла, являясь его первым массивом. Наличие каталога освобождает программы от подготовки информации о структуре массивов и позволяет организовать прямой доступ к ним. Достаточно знать лишь адрес каталога файла в основной памяти.

Подготовка данных для рабочих программ осуществляется обслуживающими программами пускового комплекса СОД/ЕС. Это программы ввода файла с перфоносителей, перегруппировки записей файла, обмена между МД и МЛ, логического контроля файла, перекомпоновки, сортировки и внесения изменений в файл, формирование нового файла из массивов двух исходных файлов.

Для типовых процедур в рабочих и обслуживающих программах пускового комплекса СОД/ЕС предусматривается создание макроопределений Ассемблера, позволяющих упростить составление программ. Они выполняют следующие операции: вычисление таблицы адресов файлов, обмен с МЛ и МД, вычисление длины заглавия, записи адресов элементов, перекодировки элементов, сортировки в оперативной памяти, ввод подмассива с перфоносителей, печать строки. С целью составления экономичных программ программирование ведется на языке, близком к машинному, — Ассемблере.

Работа каждой программы СОД/ЕС определяется параметрическим массивом, задаваемым ключевыми словами. Назначение определенной части параметрического массива всегда одинаково: идентификатор программы; символы, определяющие режим работы программы; имена исходных файлов, номера устройств, на которых должны находиться исходные или создаваемые файлы; имя создаваемого программой файла, имя файла, после которого должен быть записан новый файл.

В СОД/ЕС любая программа может быть вызвана в основную память, подготовлена к выполнению и запущена любой рабочей программой.

Пока еще рано делать окончательные выводы о перспективности того или иного направления. Даже в отношении конкретных систем — ГасИС и СОД/ЕС нет полной определенности, поскольку отсутствует опыт их эксплуатационного использования. Ясно, что во вновь создаваемой системе программ должны быть учтены ошибки и недостатки старых.

1. Автоматизированная система обработки геологических данных — АСОД—ВНИГНИ. — В кн.: Математизация и автоматизация в геологических исследованиях, М., ВСЕГЕИ, 1972. 174 с. Авт.: В. И. Аронов, Н. Е. Беляйков, Э. Я. Невельская и др.
2. Автоматизированная система оперативной обработки данных гравиметрии и магнитометрии. Киев, «Наукова думка», 1972. 162 с. с ил. Авт.: В. И. Старостенко, Р. Г. Бас, Г. С. Бутаков и др.
3. Алгоритмы и программы для решения типовых геологических задач. Обзор. Сер. «Математические методы исследований в геологии». М., ВИЭМС, 1972. 128 с. с ил. Авт.: Т. Н. Лазарева, Р. И. Коган, Ю. П. Белов, А. Г. Мюллербек.
4. Алферова З. В., Лихачева Г. Н., Шураков В. В. Математическое обеспечение ЭВМ. М., «Статистика», 1974. 383 с. с ил.
5. Андреев Б. А., Клушин И. Г. Геологическое истолкование гравитационных аномалий. Л., Гостоптехиздат, 1962. 495 с. с ил.
6. Аронов В. И. К вопросу о редуцировании аномалий силы тяжести в горной области. — «Геофизическая разведка», 1963, № 14, с. 80—91 с ил.
7. Аронов В. И., Гордин В. М. К вопросу об интерполяции производных гравитационного потенциала в горной области. — «Изв. вузов, сер. «Геодезия и аэрофотосъемка», 1965, № 6, с. 88—93 с ил.
8. Аронов В. И. О вычислении трансформант и редукции аномалий силы тяжести на внешнюю плоскость в горном районе. — «Изв. АН СССР, сер. «Физика Земли», 1970, № 7, с. 79—84.
9. Аронов В. И. Об оптимальной фильтрации случайных ошибок в гравиметрии. — «Изв. АН СССР, сер. «Физика Земли», 1970, № 10, с. 79—85 с ил.
10. Аронов В. И., Гордин В. М. О применении тренд-анализа в геологии. — В кн.: Математические методы и ЭВМ в геологии. М., 1971, вып. 103, с. 46—70 с ил.
11. Аронов В. И., Гордин В. М., Ширгинова А. И. К вопросу о построении графиков и карт изолиний в геологии и геофизике с помощью ЭВМ. — В кн.: Математические методы и ЭВМ в геологии. ВНИГНИ, М., 1971, вып. 103, с. 71—124 с ил.
12. Аронов В. И., Невельская Э. Я., Самарин В. С. Автоматизированная система поисковой и статистической обработки геологической информации на ЭВМ. Сер. «Математические методы исследований в геологии». М., ВИЭМС, 1972. 55 с. с ил.
13. Аронов В. И., Водолазский В. Н. Повышение точности прогнозных структурных карт. Экспресс-информация, сер. «Математические методы исследований в геологии». М., ВИЭМС, 1972, № 8—9, с. 1—5 с ил.
14. Аронов В. И., Невельская Э. Я. Опыт эксплуатации системы поисковой и статистической обработки геологических данных на ЭВМ. — В кн.: Применение математических методов и ЭВМ в геологии. Алма-Ата, Изд. МГ КазССР, 1974, с. 113—116.
15. Аронов В. И. Выбор шага разведочной сети, оценка и повышение точности подсчета запасов нефти и газа на ЭВМ. Экспресс-информация, сер. «Математические методы исследований в геологии». М., ВИЭМС, 1975, № 8—9, с. 1—31.
16. Аронов В. И. Обработка на ЭВМ значений аномалий силы тяжести при произвольном рельефе поверхности наблюдений. М., «Недра», 1976. 131 с. с ил.
17. Аронов В. И., Михайлов В. О. К вопросу о прогнозировании и интерполяции на ЭВМ геологических признаков с использованием косвен-

ных данных. Экспресс-информация, сер. «Математические методы исследований в геологии». М., ВИЭМС, 1976, № 1, с. 1—19.

18. Ахизер Н. И. Лекции по теории аппроксимации. М., «Наука», 1965. 407 с. ил.

19. Ашаев А. П., Веселов В. В., Сергийко Ю. А. Аппроксимация функций вида $h=f(x, y)$ уравнением Лапласа (АУЛ-2) (расчет числовых моделей карт геологических и гидрогеологических параметров). Алмата, Изд. МГ Каз. ССР, 1972 (Казахская опытно-методическая экспедиция). 48 с. с ил.

20. Бахвалов Н. С. Численные методы. М., «Наука», 1973, т. I, 632 с. с ил.

21. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов. М., «Мир», 1974. 464 с. с ил.

22. Березин И. С., Жидков Н. П. Методы вычислений. М., Физматгиз, 1959, т. 1. 464 с. с ил.; т. 2. 1960. 620 с. с ил.

23. Беспрозванный П. А., Чернов А. П. Интерполяция в узлы регулярной сети с помощью ЭВМ. Экспресс-информация, сер. «Математические методы исследований в геологии». М., ВИЭМС, 1973, № 11—12, с. 1—20 с ил.

24. Благиных В. Л., Никулин А. В., Новоселицкий В. М. Определение относительных параметров локальных поднятий на платформенной части Пермского Прикамья. — В кн.: Геология Урала и Приуралья, 1969, № 48, с. 171—178 с ил. Пермский политехнический ин-т.

25. Борисенко З. Г., Сосон М. Н. Подсчет запасов нефти объемным методом. М., «Недра», 1973. 177 с. с ил.

26. Брюсов Б. А., Гейшерик Г. М., Епифанов А. А. О возможности выделения аномалий силы тяжести, обусловленных локальными структурами, в осадочном чехле. — «Прикладная геофизика», 1972, № 68, с. 139—146 с ил.

27. Бугаец А. Н. Вопросы обработки геологической информации с помощью ЭВМ (сообщение о 24 сессии Международного геологического конгресса). Экспресс-информация. Сер. «Математические методы в геологии». М., ВИЭМС, 1973, № 10, с. 1—25.

28. Вайнштейн Л. А., Зубаков В. Д. Выделение сигналов на фоне случайных помех. М., «Советское радио», 1960. 447 с. с ил.

29. Вистелиус А. Б. О методе обзора числовых совокупностей, спектрально-тектоническом анализе и характеристике волновых движений земной коры по И. А. Одесскому. — «Изв. АН СССР». Сер. «Геотектоника», 1975, № 1, с. 117—123.

30. Водолазский В. Н., Зубов И. П., Топорская Л. П. Статистическое моделирование структурных поверхностей осадочного чехла Большеземельской тундры. — «Нефтегазовая геология и геофизика», 1974, № 4, с. 19—23.

31. Вороновский В. Р., Максимов М. М. Система обработки информации при разработке нефтяных месторождений. М., «Недра», 1975. 230 с. с ил.

32. Габриэлянд Г. А., Пороскун В. И. Методика определения рациональной сетки скважин при разведке массивных залежей. — «Геология нефти и газа», 1974, № 8, с. 61—66 с ил.

33. Гладкий К. В., Серкерев С. А. Методика построения вычислительных схем для трансформаций аномальных полей. — В кн.: Новое в полевой геофизической разведке на нефть и газ. 1965, с. 78—88 с ил.

34. Гладкий К. В., Серкерев С. А. Определение некоторых статистических характеристик аномальных гравитационных и магнитных полей. — В кн.: Полевая геофизика. М., «Недра», 1967, вып. 68. (Тр. МИНХиГП), с. 151—157 с ил.

35. Гладкий К. В., Серкерев С. А. Определение радиуса корреляции погрешности наблюдений. — В кн.: Полевая геофизика, М., «Недра», 1971, вып. 95. (Тр. МИНХиГП), с. 131—142.

36. Глушков В. М. Введение в АСУ. Киев, «Техника», 1972. 171 с. с ил.
37. Гольдин С. В. Линейные преобразования сейсмических сигналов. М., «Недра», 1974. 351 с. с ил.
38. Гордин В. М. Способ построения асимптотически оптимальных операторов тренд-анализа типа весового скользящего среднего и его реализация на ЭВМ. Экспресс-информация. Сер. «Математические методы в геологии». 1973, № 7—8, М., ВИЭМС, с. 1—11 с ил.
39. Гордин В. М. Исследования по методике тренд-анализа геолого-геофизических характеристик. — В кн.: Математические методы и ЭВМ в геологии. М., Тр. ВНИГНИ, 1973, вып. 135, с. 55—79 с ил.
40. Гордин В. М. К вопросу о вычислении выборочных автокорреляционных функций. — В кн.: Математические методы и ЭВМ в геологии. М., Тр. ВНИГНИ, 1973, вып. 135, с. 80—89 с ил.
41. Градштейн И. С., Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. М., Физматгиз, 1962. 1100 с. с ил.
42. Демидович Б. П., Марон И. А., Шувалова Э. З. Численные методы анализа. М., «Наука», 1967. 368 с. с ил.
43. Демидович Б. П., Марон И. А. Основы вычислительной математики. М., «Наука», 1970. 664 с. с ил.
44. Демин В. И., Конторович А. Е. Применение математических методов и ЭВМ при подсчете запасов нефти и газа промышленных категорий. — Изв. СО АН СССР, сер. «Геология и геофизика», 1973, № 10, с. 3—13.
45. Дубов Р. И. Количественные исследования геохимических полей для поисков рудных месторождений. Новосибирск, «Наука», 1974. 277 с. с ил.
46. Жданов М. А., Лисунов В. Р., Гришин Ф. А. Методика и практика подсчета запасов нефти и газа. М., «Недра», 1967. 403 с. с ил.
47. Жданов М. С., Шрайбман В. И. Корреляционный метод разделения геофизических аномалий. М., «Недра», 1973. 127 с. с ил.
48. Игревский В. И. Основные принципы построения автоматизированной системы управления АСУ — Геология. М., «Советская геология», 1973, № 1, с. 3—14.
49. Заикин В. В. Об одном способе представления функций с ограниченным спектром при неперiodических отсчетах. — «Радиотехника», 1974, № 8, с. 12—15.
50. Зайдель А. Р., Лохматов Г. И., Таусон Н. Н. Математические методы при построении карт в практике геологических работ. Сер. «Нефтегазовая геология и геофизика». М., ВНИИОЭНГ, 1967. 71 с. с ил.
51. Каратаев Г. И., Ватлин Б. П., Захарова Т. Л. Методика комплексной геологической интерпретации гравитационных и магнитных аномалий. Новосибирск, «Наука», 1973. 167 с. с ил.
52. Качаев В. А., Перфильев Л. Г. Стандартные подпрограммы вычерчивания произвольных осей координат и графиков на устройстве «Атлас» для БЭСМ-4. Алма-Ата, 1971, вып. 4, (Казах, опытно-метод. эксп., серия XI, машинная графика). 52 с. с ил.
53. Китов А. И. Программирование информационно-логических задач. М., «Советское радио», 1967. 327 с. с ил.
54. Кноринг Л. Д., Деч В. Н. Оптимизация объемов работ при разведке на нефть и газ. — «Советская геология», 1973, № 6, с. 92—101 с ил.
55. Коган Р. И. Интервальные оценки запасов полезных ископаемых. М., «Недра», 1972. 142 с. с ил.
56. Королев Л. Н. Структуры ЭВМ и их математическое обеспечение. М., «Наука», 1974. 254 с. с ил.
57. Крамбейн У., Грейбилл Ф. Статистические модели в геологии. М., «Мир», 1969. 398 с. с ил.
58. Кривулин Г. С., Пустовойтов П. П. Устройство Атлас-3 для графического вывода информации из ЭВМ. Экспресс-информация. Сер.

«Региональная, разведочная и промысловая геофизика», № 7, М., ВИЭМС, 1973, с. 1—9 с ил.

59. Криницкий Н. А., Миронов Г. А., Фролов Г. Д. Программирование и алгоритмические языки. М., «Наука», 1975. 496 с. с ил.

60. Лапина М. И., Страхов В. Н. Формализованные алгоритмы фильтрации потенциальных полей. — «Изв. АН СССР, сер. «Физика Земли», 1973, № 7, с. 46—67 с ил.

61. Лапорт М. Ускоренная обработка гравиметрических карт в редукции Буге с помощью электронных счетных устройств. — «Бюлл. науч.-техн. информации», М., «Недра», 1964, № 3 (53), с. 63—68 с ил.

62. Левин Б. Р. Теоретические основы статистической радиотехники. М., «Советское радио», 1974, т. 1, 550 с.; 1968, т. 2, 502 с. с ил.

63. Литвиненко О. К., Русьянов Ю. Г. Построение геофизических карт с помощью электронных цифровых вычислительных машин. Сер. «Региональная, разведочная и промысловая геофизика». М., ВИЭМС, 1967, № 17, с. 15 с ил.

64. Ломтадзе В. В. Принципы организации автоматизированной системы обработки геофизических данных. — «Прикладная геофизика», 1973, № 72, с. 183—188.

65. Ломтадзе В. В., Торопов С. М. О стандартизации математического и информационного обеспечения АСУ. Управляющие системы и машины. Киев, «Наукова думка», 1974, № 6, с. 30—33.

66. Ломтадзе В. В., Торопов С. М. К вопросу о стандартизации математического и информационного обеспечения АСУ — Геология. — В кн.: Алгоритмы и программы, М., ВИЭМС, 1975, вып. 5, с. 31—38.

67. Марголин А. М. Оценка запасов минерального сырья. Математические методы. М., «Недра», 1974. 261 с. с ил.

68. Математический метод палеотектонического анализа платформенных структур. Иркутск, 1969 (Тр. Вычислит. центра Иркутского гос. ун-та), вып. 3, с. 157 с ил. Авт.: Г. И. Лохматов, В. Н. Евдокимова, Г. Т. Алаев и др.

69. Матерон И. С. Основы прикладной геостатистики. М., «Мир», 1968. 407 с. с ил.

70. Миллер Р. Л., Кан Дж. С. Статистический анализ в геологических науках. М., «Мир», 1965. 480 с. с ил.

71. Михайлов В. О. К вопросу о точности подсчета запасов нефти и газа объемным методом. — «Тр. ВНИГНИ, Математические методы и ЭВМ в геологии». М., 1973, вып. 135, с. 90—95.

72. Налимов В. В. Теория эксперимента. М., «Наука», 1971. 208 с. с ил.

73. Омелин В. М. Система вывода графической информации на ФОРТРАНе. Экспресс-информация. Сер. «Математические методы исследования в геологии». М., ВИЭМС, 1975, № 4, с. 19—24.

74. О некоторых путях повышения эффективности разведки нефтяных месторождений. — «Геология нефти и газа», 1975, № 8, с. 1—7. Авт.: В. В. Стасенков, Н. Е. Быков, А. Я. Фурсов и др.

75. Перфильев Л. Г., Захарченко В. В., Качаев В. А. Основные элементы математического обеспечения графопостроителя «Атлас». Экспресс-информация, сер. «Региональная, разведочная и промысловая геофизика», М., 1972, № 13, с. 1—12 с ил.

76. Перфильев Л. Г. Модернизированная программа вычерчивания карт изолиний на графопостроителе «Атлас» для ЭВМ БЭСМ-4, М-222. Вып. 9. Алма-Ата, 1974, Казахст. опытно-метод. экп., серия XI, машинная графика. 66 с. с ил.

77. Поплавский Н. Н., Коротков Б. П., Бурмистров В. Н. Принципы создания автоматизированной системы информационного обеспечения комплекса подсистем «Геология» в АСУ — «Геология». — В кн.: Повышение эффективности использования вычислительной техники на геологоразведочных работах. Киев, Изд. МГ УССР, 1975, с. 105—106.

78. Применение электронных счетных машин при интерпретации

гравитационных и магнитных аномалий. М., Гостоптехиздат, 1962. 76 с. с ил. Авт.: Р. Ф. Володарский, В. И. Аронов, Е. Г. Дьяконов и др.

79. Применение многомерного регрессионного анализа в газонефтяной геологии и геофизике. — В кн.: Математические методы в геологии. М., «Наука», 1968, с. 62—68. Авт.: М. М. Элланский, Т. Н. Трошко, Э. А. Калинина, А. П. Петров.

80. Рао С. Р. Линейные статистические методы и их применения. М., «Наука», 1968. 548 с. с ил.

81. Романова М. А. Тренд-анализ данных геологических наблюдений (основная литература). — В кн.: Вопросы математической геологии. Л., «Наука», 1968, с. 284—288.

82. Савинский И. Д. Применение математических методов и ЭВМ при выборе разведочных сетей и построении планов в изолиниях. М., «Недра», 1974. 115 с. с ил.

83. Серкеров С. А. Частотная характеристика оптимального фильтра для разделения полей. — В кн.: Новое в полевой геофизической разведке. М., 1965. с. 88—97 с ил.

84. Система программ обработки на ЭВМ гравитационных и магнитных аномалий, заданных на плоскости или произвольной поверхности. Обзор, сер. «Региональная, разведочная и промысловая геофизика». М., ВИЭМС, 1972. 93 с. с ил. Авт.: В. И. Аронов, Н. Е. Беляйков, Т. И. Ланда и др.

85. Страхов В. Н. К вопросу о точности задания гармонических функций их значениями в узлах прямоугольной сети. — «Геология и геофизика», 1964, № 3, с. 112—129.

86. Страхов В. Н. О сглаживании наблюдаемых значений потенциальных полей. — «Изв. АН СССР, сер. «геофизика», 1964, № 10, с. 1479—1493, № 11, с. 1634—1653.

87. Страхов В. Н., Лапина М. И. Метод сглаживания потенциальных полей. — «Изв. АН СССР, сер. «Физика Земли», 1967, № 8, с. 40—57 с ил.

88. Технология автоматизированного подсчета запасов месторождений твердых полезных ископаемых методами разрезов и блоков на ЭВМ БЭСМ-4. Серия VII, вып. 6, опытно-метод. экспед., Алма-Ата, 1973, 123 с. с ил.

89. Ткаченко В. В., Торопов С. М., Ломтадзе В. В. Автоматизированная система управления Восточным геофизическим трестом (АСУ ВГТ). Экспресс-информация. Сер. «Математические методы исследования в геологии». М., ВИЭМС, 1975, № 7, с. 1—29 с ил.

90. Требования к достоверности определения параметров залежей нефти в процессе разведочных работ. — «Геология нефти и газа», № 10, 1974, с. 18—22. Авт.: Н. Е. Быков, Р. А. Егоров, Е. Ф. Фролов и др.

91. Тутубалин В. Н. Теория вероятностей в естествознании. Сер. «Математика, кибернетика». М., «Знание», 1972. 48 с.

92. Харбух Дж., Бонэм-Картер Г. Моделирование на ЭВМ в геологии. М., «Мир», 1974. 319 с. с ил.

93. Харкевич А. А. Борьба с помехами. М., «Наука», 1965. 275 с. с ил.

94. Хемминг Р. В. Численные методы. М., «Наука», 1968. 400 с. с ил.

95. Хургин Я. И., Яковлев В. П. Фinitные функции в физике и технике. М., «Наука», 1971. 408 с. с ил.

96. Шрайбман В. И. Геологическое истолкование геофизической информации о глубинной структуре Туранской плиты. Автореф. докт. дисс., МИНХиГП, 1970. 48 с.

ОГЛАВЛЕНИЕ

	Стр.
Введение	3
Глава I. <i>Интерполяция геологических характеристик на регулярную сеть. Построение графиков и карт изолиний</i>	6
§ 1. Постановка задачи	7
§ 2. О точности интерполяции геологических характеристик	18
§ 3. Интерполяция алгебраическими полиномами	27
§ 4. Метод сеток	38
§ 5. Приближение гармоническими функциями	45
§ 6. Построение графиков и карт изолиний с помощью графопостроителя	66
Глава II. <i>Фильтрация случайных ошибок. Тренд-анализ</i>	72
§ 1. Постановка задачи	73
§ 2. Оптимальный фильтр Колмогорова — Винера	81
§ 3. Фильтрация случайных ошибок в пунктах нерегулярной сети наблюдений	93
§ 4. Полиномиальные методы. Осреднение	99
Глава III. <i>Прогнозирование количественных геолого-геофизических характеристик</i>	114
§ 1. Прогнозный вариант	115
§ 2. Интерполяционно-прогнозный вариант	124
Глава IV. <i>О подсчете запасов на ЭВМ</i>	130
§ 1. Оценка точности подсчета запасов в зависимости от величины шага наблюдений	133
§ 2. Метод подсчета запасов на ЭВМ	139
§ 3. Фильтрация случайных ошибок	141
§ 4. Ошибки подсчета запасов по объемной формуле	144
Глава V. <i>Обработка геологических данных на ЭВМ</i>	147
Список литературы	164

1 р. 08 н.

2068

НЕДРА