

Интервальный подход к обработке зашумленных экспериментальных данных с многократными измерениями в условиях неопределённости*

С.И. Кумков

Институт математики и механики УрО РАН, Екатеринбург
e-mail: 620219, Екатеринбург ГСП-384, ул.С.Ковалевской 16

Рассматривается обработка выборки зашумленных экспериментальных данных при малом числе значений основного аргумента (короткая выборка) и многократных измерениях в подвыборках для каждого его значения. Задан вид исследуемой зависимости с вектором её параметров. Замеры в подвыборках содержат как обычные приборные погрешности измерений, так и хаотические искажения неизвестного знака и величины. Вероятностные характеристики обоих типов возмущений полностью неизвестны, неопределённой является также величина ограничения на максимальное значение суммарных погрешностей. Поэтому невозможно обосновать применение стандартных (ГОСТ) статистических методов. Предлагается подход на основе методов интервального анализа: для заданного уровня ограничения на максимальную величину суммарной погрешности измерения строятся интервалы неопределённости замеров (ИНЗ); выделяются совместные в интервальном смысле последовательные ИНЗ (по одному для каждого значения аргумента); каждой совместной последовательности ИНЗ ставится в соответствие информационное множество (ИМ) допустимых значений параметров. Выходной результат представляется суммарным ИМ всех допустимых последовательностей замеров. В отличие от стандартных методов, где выбраковка отдельного замера выполняется по величине его отклонения от средней аппроксимирующей зависимости, в предлагаемом подходе выбраковка последовательности замеров выполняется по пустоте или недопустимости (по отношению к априорной области параметров) её ИМ.

В настоящее время широко применяются стандартные методы [1,2] обработки зашумленных экспериментальных данных. Заранее задаётся вид исследуемой зависимости (от некоторого аргумента) с вектором её параметров. Данные подходы (ГОСТ'ы) опираются на статистические методы и предположения: о представительности выборки (большое число замеров и значений основного аргумента), о нормальности распределения погрешностей измерения, о статистической однородности погрешности многократных измерений и о несмещённости замеров, а также на предположение, что значения аргумента известны точно.

Однако на практике [3,8,9], экспериментальные данные могут содержать малое число замеров или значений основного аргумента (короткая выборка), но выполняются повторные (многократные) измерения (рис.1, подвыборки – чёрные кружки, вертикальные группы) для каждого его значения. На рис.1в жирной прямой отмечена линия зависимости заданного вида, формально полученная по [1,2].

Значения аргумента полагаются известными точно, а замеры в подвыборках содержат как обычные приборные погрешности измерений, так и хаотические (неконтролируемые) искажения неизвестного знака и величины. Наличие хаотических искажений

*Работа поддержана Программой П(29)6-2 Президиума РАН и грантом РФФИ № 10-1-96-006.

подтверждается сравнением величины приборной ошибки измерений порядка 0.005 в и размахом разбросов замеров до ~ 0.08 в (например, рис.1, средняя подвыборка). Далее, анализ рассеяния замеров (рис.1д) показывает, что распределение суммарной погрешности не является нормальным, и, как правило, полностью неизвестно. Отметим, что стандартное правило " $\pm 3\sigma$ " исключения промахов (выбросов) не удаляет ни одного замера из выборки рис.1.

В таких условиях трудно или даже невозможно обосновать применение указанных стандартных методов. Практически неопределённой является также величина ограничения на максимальное значение суммарных погрешностей.

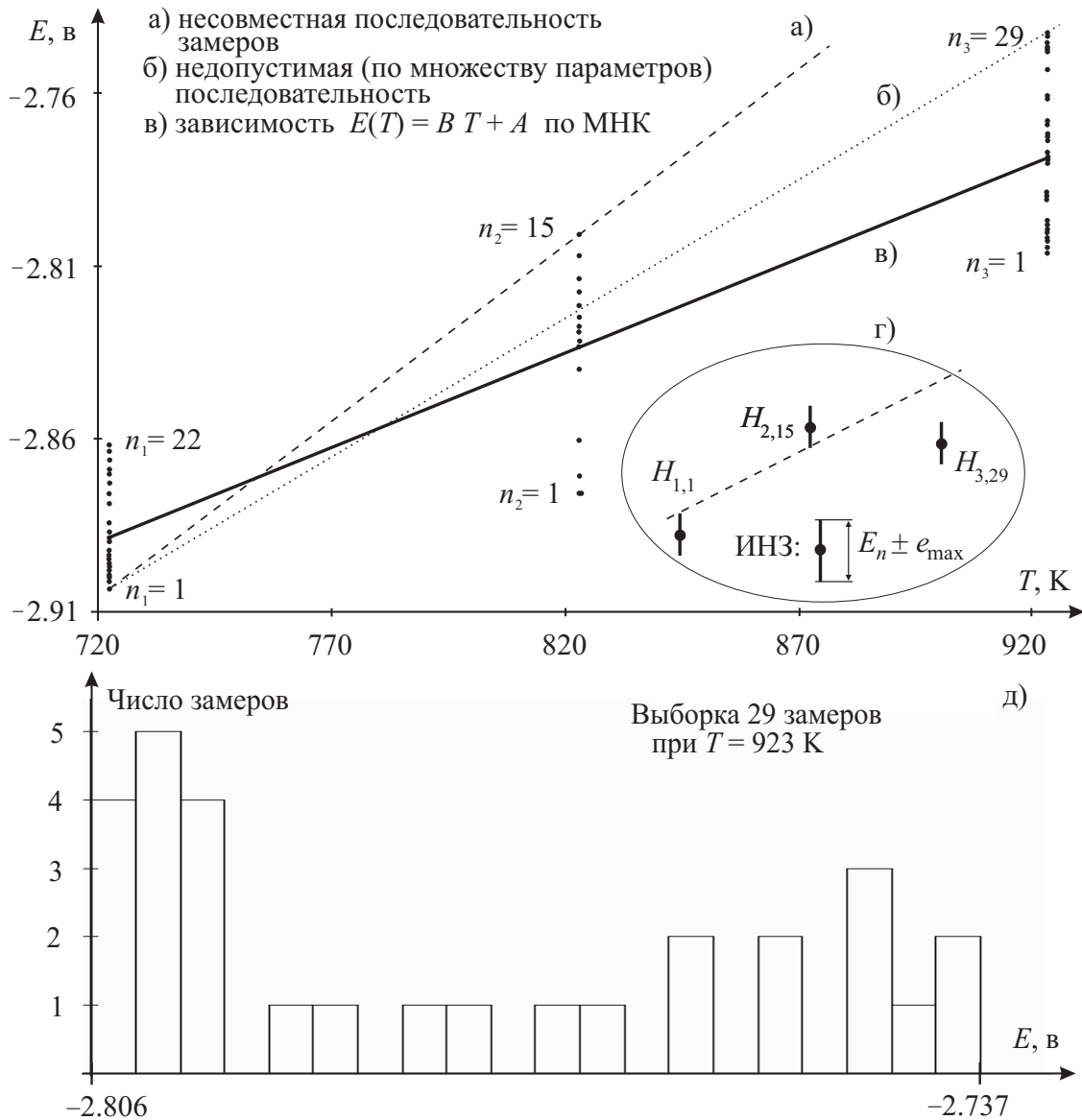


Рис. 1. Зашумленные экспериментальные данные; а) несовместная зависимость; б) совместная тройка замеров; в) линия МНК; г) несовместная тройка ИНЗ; д) типичная гистограмма рассеяния замеров

Предлагаемый подход опирается на методы интервального анализа [4–6]. В настоящей работе общие приёмы решения интервальных систем линейных уравнений конкретизируются на случай решения переопределённых интервальных систем уравнений

и неравенств. Реализуются следующие алгоритмы. Из теоретических и инженерных соображений вводится некоторый разумный приемлемый уровень ограничения на максимальную величину e_{\max} суммарной погрешности измерения. По этому уровню неопределённость измерения в каждом замере формализуется в виде его интервала неопределённости $H_{i,j}$ (ИНЗ, рис.1г, увеличенный фрагмент в овале).

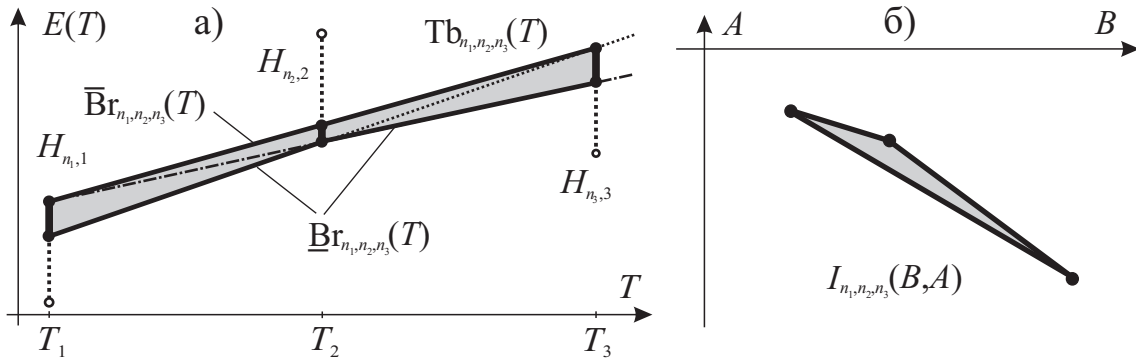


Рис. 2. Совместная тройка ИНЗ; а) трубка допустимых зависимостей; б) информационное множество параметров совместной тройки

Анализируется набор ИНЗ последовательных замеров по одному для каждого значения аргумента (рис.2а, $H_{i,j}$). Такая последовательность замеров полагается совместной в интервальном смысле, если через всех их ИНЗ можно провести хотя бы одну зависимость заданного вида. Совместная последовательность определяет пучок–трубку допустимых зависимостей (рис.2а, трубка Tb_{n_1,n_2,n_3} отмечена серой заливкой с нижней $\underline{Br}_{n_1,n_2,n_3}$ и верхней $\overline{Br}_{n_1,n_2,n_3}$ границами). Пунктиром отмечены те части интервалов неопределённости, которые оказываются несовместными в рассматриваемой последовательной тройке.

Несовместные последовательности замеров (например, рис.1г, тройка ИНЗ в овале) имеют пустое ИМ (рис.1а и г, штриховая линия).

Каждой совместной последовательности ставится в соответствие информационное множество (ИМ) I_{n_1,n_2,n_3} (рис.2б) допустимых значений параметров исследуемой зависимости. Информационное множество строится по специальной процедуре [7–9] численного решения переопределённой системы интервальных линейных неравенств тройки ИНЗ. Процедура является конструктивной, так как по ходу вычисления обнаруживает и отсеивает несовместные тройки.

Вследствие наличия хаотических возмущений в замерах набор $\{I_{n_1,n_2,n_3}\}$ непустых ИМ представляет собой совокупность (рис.3а) разбросанных и несвязных многоугольников (для выборки рис.1); величина ограничения на суммарную погрешность задавалась на уровне приборной ошибки измерений $e_{\max} = 0.005$ в. При этом уровне ограничения из возможного числа 9570 троек замеров *совместными* оказываются только 2679 троек.

Пусть имеется некоторая априорно заданная область $I^{\text{апр}}(B, A)$ значений параметров (рис.3а, прямоугольник со штриховыми границами). При этом каждое непустое ИМ может быть отброшено (как неприемлемое, рис.1б, пунктирная линия), если его пересечение с этой областью пусто, или уточнено, если его пересечение с этой областью непусто. Видно, что отбрасываться будут ИМ, соответствующие тройкам замеров с “неестественно” большими отклонениями (рис.3, многоугольники вне априорной обла-

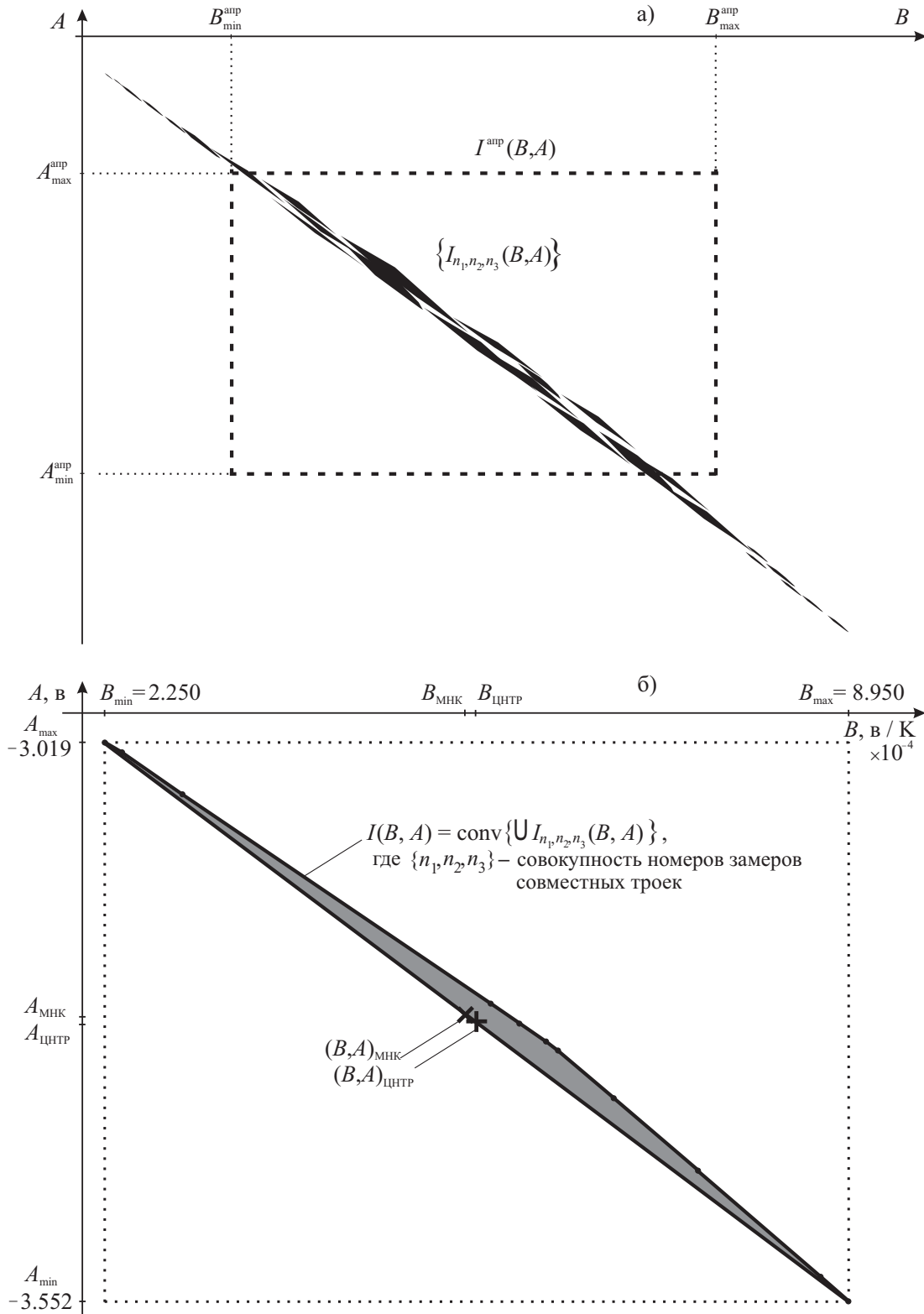


Рис. 3. Информационные множества параметров; а) набор информационных множеств совместных троек замеров и учёт априорной области параметров; б) объединённое информационное множество параметров допустимых троек (без учёта априорных данных)

сти), и в случае учета априорных данных число оставшихся *допустимых* троек уменьшается. Таким образом, предлагаемый подход позволяет *не учитывать информацию о искажённых замерах* на основе недопустимости соответствующих информационных множеств параметров.

Выходной результат оценки допустимых значений параметров представляется выпуклой оболочкой $I(B, A) = \text{conv}\{\bigcup I_{n_1, n_2, n_3}\}$ объединения всех допустимых ИМ и набором их трубок. На рис.3б показан результат обработки исходной выборки – результирующее информационное множество $I(B, A)$ (без учёта априорной информации). Это выпуклый многоугольник с 11 вершинами и линейными границами. Выходной результат получен при реальной величине ограничения $e_{\max} = 0.005$ в суммарной погрешности измерений. В качестве выходных данных потребителю выдаются следующие оценки (рис.3б):

- точное описание результирующего ИМ, набор координат его вершин $\{(B_j, A_j)\}$;
- центральная точка $(B, A)_{\text{центр}}$ результирующего ИМ (прямой крестик), для иллюстрации на рис.3б косым крестиком отмечена точка $(B, A)_{\text{МНК}}$, формально найденная стандартным методом;
- безусловные интервалы каждого параметра $\Delta B = B_{\max} - B_{\min}$ и $\Delta A = A_{\max} - A_{\min}$ (прямоугольник с пунктирными границами).

Качество обрабатываемых экспериментальных данных оценивается полуразмахом каждой из подвыборок (рис.1): ± 0.021 , ± 0.038 в и ± 0.034 в, соответственно при $T = 723$ К, $T = 823$ К $T = 923$ К. Данные оценки в $\sim 4 \div 8$ раз превышают указанный уровень приборных погрешностей 0.005 в. Эти оценки ещё раз подтверждают наличие хаотических (неконтролируемых) погрешностей измерения и говорят о плохом качестве входной обрабатываемой выборки.

Исследовалась состоятельность предлагаемого метода. На основе реальной информации, подобной приведённой на рис.1 строился модельный пример. В нем около некоторой “истинной”–модельной зависимости при заданном уровне приборной погрешности измерений e_{\max} для указанных значений аргумента T моделировались подвыборки замеров, с отклонениями не более $\pm e_{\max}$ и не содержавших хаотических искажений. По построению каждая такая подвыборка была *достоверной*, т.е. ИНЗ всех её замеров содержали соответствующее “истинное” измеряемое значение зависимости, а их пересечение было непусто. При этом результирующее информационное множество параметров, построенное по предлагаемому подходу, также было *достоверным* – содержало точку параметров “истинной”–модельной прямой.

Отметим в заключение важный момент. В стандартных методах [1,2] выбраковка отдельного замера (или нескольких замеров) выполняется по величине его отклонения от средней аппроксимирующей зависимости в случае превышения отклонением заданного порога $\pm 2\sigma$ (при доверительной вероятности 0.95) или $\pm 3\sigma$ (при доверительной вероятности 0.97). Поскольку в рассматриваемом случае из-за наличия хаотических искажений замеров статистический подход не работает, то в предлагаемом подходе:

- принципиально по-иному используется внутренняя связь между замерами в соответствии с заданным видом зависимости;
- выбраковка последовательностей замеров выполняется не по их отклонениям от некоей аппроксимирующей зависимости, а по их интервальной несовместности (пустоте их ИМ) или, для совместной последовательности, по недопустимости её информационного множества по отношению к заданной априорной области параметров.

Для линейной зависимости разработаны быстрые процедуры обработки.

Список литературы

- [1] ГОСТ 8.207-76. Государственная система обеспечения единства измерений. Прямые измерения с многократными наблюдениями. Методы обработки результатов наблюдений.
- [2] Р 40.2.028–2003. Государственная система обеспечения единства измерений. Рекомендации по построению градуировочных характеристик. Оценивание погрешностей (неопределённости) линейных градуировочных характеристик при использовании метода наименьших квадратов.
- [3] Лукьянова М.Я., Осипенко А.Г., Маершин А.А., Кормилицин М.В., Кумков С.И., Валишин М.Ф. Задача обработки экспериментальных данных по хронопотенциометрии кюрия в расплаве $3\text{LiCl}-2\text{KCl}$ // Доклады III Международной Пирохимической Конференции, 29.11–03.12. 2009, Димитровград, Россия, Исследовательский институт атомных реакторов.
- [4] ЖОЛЕН Л., КИФЕР М., ДИДРИ О., ВАЛЬТЕР Э. Прикладной интервальный анализ. Москва–Ижевск, Регулярная и хаотическая динамика, 2007.
- [5] ШАРЫЙ С.П. Конечномерный интервальный анализ. Электронная книга: <http://www.nsc.ru/interval/Library/InteBooks/SharyBook.pdf>
- [6] ХАНСЕН Э., УОЛСТЕР ДЖ.У. Глобальная оптимизация с помощью методов интервального анализа. Москва–Ижевск, Регулярная и хаотическая динамика, 2010.
- [7] Кумков С.И. Разработка совместного Российского – ISO стандарта (методики) обработки измерительной информации в условиях неопределённости ошибок измерений и малого числа наблюдений (на основе методов интервального анализа) // Доклады Всероссийского (с международным участием) Совещания по интервальному анализу и его приложениям "ИНТЕРВАЛ-06", 1–4 июля 2006 г., Санкт–Петербург, Санкт–Петербургский Государственный Университет, 2006, стр. 63–67, и сайт <http://www.ict.nac.ru/interval/Conferences/Interval-06>
- [8] Кумков С.И. Обработка экспериментальных данных ионной проводимости расплавленного электролита методами интервального анализа // Расплавы, № 3, 2010, С. 86–96.
- [9] ПОТАПОВ А.М., КУМКОВ С.И., САТО Y. Обработка экспериментальных данных по вязкости при одностороннем смещении ошибок измерения // Расплавы, № 3, 2010, С. 55–70.