

МЕЖДУНАРОДНАЯ
КОНФЕРЕНЦИЯ
по ВЫЧИСЛИТЕЛЬНОЙ
МАТЕМАТИКЕ

Труды: Рабочие совещания

Редакторы:

Ю.И. Шокин, А.М. Федотов,
С.П. Ковалев, Ю.И. Молородов,
А.Л. Семенов, С.П. Шарый

Новосибирск, 2004

Труды Международной конференции по вычислительной математике МКВМ-2004. Рабочие совещания / Ред.: Ю.И. Шокин, А.М. Федотов, С.П. Ковалев, Ю.И. Молородов, А.Л. Семенов, С.П. Шарый. – Новосибирск: Изд. ИВМиМГ СО РАН, 2004.

В сборнике публикуются доклады, представленные на Рабочих совещаниях “Распределенные информационно-вычислительные ресурсы и математическое моделирование” и “Интервальная математика и методы распространения ограничений”, проходящих в рамках Международной конференции по вычислительной математике МКВМ-2004 (Новосибирск, 21–25 июня 2004 г.). Первое Рабочее совещание посвящено современному состоянию работ в области организации и эксплуатации распределенных информационно-вычислительных ресурсов. Второе Рабочее совещание охватывает темы, посвященные как собственно интервальной математике, так и быстро развивающимся методам интервального распространения ограничений, а также комбинации этих подходов.

Программный комитет Рабочего совещания “Распределенные информационно-вычислительные ресурсы и математическое моделирование”:

Ю.И. Шокин	ИВТ СО РАН, Новосибирск, Россия – председатель
А.М. Федотов	ИВТ СО РАН, Новосибирск, Россия – зам. председателя
С.П. Ковалев	ИВТ СО РАН, Новосибирск, Россия – уч. секретарь
И.В. Бычков	ИДСТУ СО РАН, Иркутск, Россия
О.Л. Жижимов	ОИГГМ СО РАН, Новосибирск, Россия
Ю.И. Молородов	ИВТ СО РАН, Новосибирск, Россия
В.П. Потапов	ИУУ СО РАН, Кемерово, Россия
В.В. Шайдуров	ИВМ СО РАН, Красноярск, Россия

Программный комитет Рабочего совещания “Интервальная математика и методы распространения ограничений”:

А.Л. Семенов	ИСИ СО РАН, Новосибирск, Россия – председатель
С.П. Шарый	ИВТ СО РАН, Новосибирск, Россия – зам. председателя
Ф. Бенаму	Университет Нанта, Нант, Франция
Т.П. Кашеварова	ИСИ СО РАН, Новосибирск, Россия
Р.Б. Кирфотт	Университет Южной Луизианы, Лафайетт, США
В. Крейнвич	Университет Техаса в Эль Пасо, Эль Пасо, США
А.В. Лакеев	ИДСТУ СО РАН, Иркутск, Россия

© ИВМиМГ (ВЦ) СО РАН, 2004

Редакционно-издательская группа:

Л.И. Бессильных, М.В. Зверева, А.И. Роженко

Подготовлено в макросе NCC-ITrX. Лицензия ИД № 02202 от 30.06.2000 г.
Отпечатано в ЗАО РИЦ “Прайс-курьер”, т. (8-3832) 34-22-02,
630090 Новосибирск, пр. Лаврентьева, 6

**THE INTERNATIONAL
CONFERENCE
ON COMPUTATIONAL
MATHEMATICS**

Proceedings: Workshops

Edited by

Yu.I. Shokin, A.M. Fedotov,
S.P. Kovalyov, Yu.I. Molorodov,
A.L. Semenov, S.P. Shary

Novosibirsk, 2004

Proceedings of the International Conference on Computational Mathematics ICCM-2004. Workshops / Eds.: Yu.I. Shokin, A.M. Fedotov, S.P. Kovalyov, Yu.I. Molorodov, A.L. Semenov, S.P. Shary. – Novosibirsk: ICM&MG Publisher, 2004.

This volume of Conference Proceedings presents the reports submitted to the Workshops “Distributed information-computing resources and mathematical modeling” and “Interval mathematics and interval constraint programming” being held in the framework of the International Conference on Computational Mathematics ICCM-2004 (Novosibirsk, June 21–25, 2004). The first Workshop is devoted to the state of the art in the field of organization and usage of distributed information-computing resources. The second Workshop covers a wide range of topics in the fields of interval mathematics and interval constraint programming, as well as combinations of these approaches.

Program Committee of the Workshop “Distributed information-computing resources and mathematical modeling”:

Yu.I. Shokin	ICT SB RAS, Novosibirsk, Russia – chairman
A.M. Fedotov	ICT SB RAS, Novosibirsk, Russia – vice-chairman
S.P. Kovalyov	ICT SB RAS, Novosibirsk, Russia – sci. secretary
I.V. Bychkov	ISDCT SB RAS, Irkutsk, Russia
O.L. Zhizhimov	UIGGM SB RAS, Novosibirsk, Russia
Yu.I. Molorodov	ICT SB RAS, Novosibirsk, Russia
V.P. Potapov	ICC SB RAS, Kemerovo, Russia
V.V. Shaidurov	ICM SB RAS, Krasnoyarsk, Russia

Program Committee of the Workshop “Interval mathematics and interval constraint programming”:

A.L. Semenov	IIS SB RAS, Novosibirsk, Russia – chairman
S.P. Shary	ICT SB RAS, Novosibirsk, Russia – vice-chairman
F. Benhamou	University of Nantes, Nantes, France
T.P. Kashevarova	IIS SB RAS, Novosibirsk, Russia
R.B. Kearfott	University of South Louisiana, Lafayette, USA
V. Kreinovich	University of Texas at El Paso, El Paso, USA
A.V. Lakeyev	ISDCT of SB RAS, Irkutsk, Russia

© 2004 by ICM&MG Publisher

Editorial board:

L.I. Bessil'nykh, M.V. Zvereva, A.I. Rozhenko

Prepared with NCC- \LaTeX Macro.

Содержание

Распределенные информационно-вычислительные ресурсы и математическое моделирование	1
Предисловие	2
Шокин Ю.И., Федотов А.М. <i>Высокоскоростные сети передачи данных СО РАН для GRID-систем</i>	3
Алексеев А.С., Глинский Б.М., Котелевский С.П., Кучин Н.В., Малышкин В.Э., Селихов А.В. <i>Сибирский Суперкомпьютерный Центр (ССКЦ), перспективы его развития</i>	19
Шайдуров В.В., Исаев С.В., Малышев А.В. <i>Красноярская сеть параллельных вычислений</i>	27
Азизов Р.К., Привезенцев А.И., Фазлиев А.З. <i>Организация обмена RDF-документами в распределенной информационной системе</i>	34
Бычков И.В., Гаченко А.С., Лемперт А.А., Урбанович Д.Е. <i>Разработка вычислительного сервера с удаленным доступом на примере комплекса программ по параметрической идентификации</i>	38
Городилов В.В., Залеская А.А. <i>Распределенные источники данных для создания систем поддержки принятия решений (OLAP-систем)</i>	41
Жижимов О.Л., Мазов Н.А. <i>Серверы Z39.50 ZooPARK и Z-IRBIS – новые возможности новых версий</i>	45
Завертайлов А.В., Ковалев С.П. <i>Система поддержки деятельности распределенных экспертных групп по разработке онтологий предметных областей</i>	56
Ковалев С.П., Яковченко К.Н. <i>Организация информационных порталов на основе канальной интеграции</i>	66
Кузьмин Е.В., Соколов В.А. <i>Проверка модели для вполне структурированных систем переходов автоматного типа</i>	73
Молородов Ю.И. <i>Методология разработки распределенной информационной среды по атмосферным аэрозолям Сибири</i>	86
Новоселов Д.Ю., Ковалев С.П. <i>Аспектно-ориентированный подход к созданию GRID приложений</i>	95
Попов С.Е. <i>Принципы построения систем распределенных вычислений для решения задач горного дела на примере модели энтропийного анализа функциональных показателей угольных предприятий</i>	103
Пушной Б.М. <i>О системном подходе к задачам обработки наблюдений</i>	110

Рогалев А.Н. <i>Параллельное вычисление гарантированных границ решений дифференциальных уравнений</i>	119
Соколов В.А., Чалый Д.Ю. <i>Методы исследования поведения транспортных протоколов в условиях интенсивного сетевого трафика</i>	126
Федотов А.М. <i>Концептуальные подходы к построению распределенных систем</i>	132
Интервальная математика и методы распространения ограничений	145
Предисловие	146
Воцинин А.П. <i>Задачи анализа с неопределенными данными – интервальность и/или случайность?</i>	147
Kearfott R.B. <i>Validated constraint solving – practicalities, pitfalls, and new developments</i>	159
Базаров М.Б. <i>Новые алгоритмы интервального вычисления определенных интегралов</i>	168
Джаныбеков Б.С., Шарый С.П. <i>Интервальный метод Хаусхолдера для комплексных линейных систем</i>	171
Добронец Б.С. <i>Специальные приближения множеств решений</i>	179
Долгов Ю.Г. <i>Метод глобальной оптимизации на основе метода ветвей и границ</i>	184
Ерохин В.И. <i>Необходимые и достаточные условия невырожденности интервальных матриц</i>	193
Ершов А.Г., Кашеварова Т.П. <i>Интервальная математическая библиотека, основанная на разложениях в ряды Чебышева и Тейлора</i>	201
Жилин С.И. <i>О построении зависимостей по эмпирическим данным с интервальной ошибкой</i>	210
Калинкина С.Ю., Пущков С.Г. <i>Модификация метода граничных реализаций для интервальных импульсных последовательностей смешанного типа</i>	219
Киншт Н.В., Петрунько Н.Н., Цициашвили Г.Ш. <i>О распознавании экстремальных объектов с помощью интервальных решающих правил</i>	225
Манкевич П.В. <i>Средство для решения полиномиальных ограничений в решателе Sibsalc</i>	231
Рогалев А.Н. <i>Ансамбли систем дифференциальных уравнений с интервальными данными</i>	240
Филиппенко Е.Г., Зюзин В.С. <i>Приближение решения задачи Гурса интервальными многочленами Тейлора</i>	255

Хлебалин Н.А., Пятых Д.С. <i>Моделирование систем автоматического управления с интервальной неопределенностью параметров (новые версии и возможности комплекса программ “АСИАС”)</i>	258
Шарая И.А. <i>Критерий неограниченности допустимого множества решений</i>	267
Akhmerov R.R. <i>Interval-affine Gaussian method for systems with constrained coefficients</i>	271
Herrero P., Sainz M.A., Vehí J., Jaulin L. <i>Quantified Set Inversion Algorithm</i>	272
Kubica B.J., Malinowski K. <i>An interval algorithm combining symbolic rewriting and componentwise Newton method applied to control a class of queueing systems</i>	280
Men’shikov G.G. <i>Example of Babuška, Práger and Vitásek in interval computations</i>	285
Michelucci D., Faudot D. <i>A reliable curve tracing method</i>	289
Senyo P.S., Hnatyshyn M.V. <i>Interval first-derivative-based iterative methods with high order of convergence for solving nonlinear systems of equations</i>	303
Shary S.P. <i>Krawczyk operator revised</i>	307
Sokolova S.P., Sokolova L.A. <i>Monitoring of the natural plague focus with interval and fuzzy data using immunocomputing</i>	314
Авторский указатель	315

Интервальная математика и методы распространения ограничений

Предисловие

Рабочее совещание “Интервальная математика и методы распространения ограничений” (ИМРО-2004) тесно связано со значительным научным мероприятием – Международной научно-практической конференцией *isiCAD-2004*. Главная цель конференции *isiCAD-2004* – представление исследовательских и технологических подходов и готовых результатов, которые вносят вклад в разработку программных продуктов в CAD/CAM/CAE/PDM, компьютерной графике, виртуальной реальности и компьютерном моделировании вообще, а также в других смежных с ними областях приложений. Особое внимание уделено разработке и применению технологий, связанных с так называемыми задачами с ограничениями и геометрическими решателями. С общей точки зрения конференция *isiCAD-2004* призвана “навести мосты” между указанными предметными областями и культивируемыми в них методами с целью придать их развитию дальнейший динамизм. Фактически, это междисциплинарная конференция, на которой, в частности, нашли место методы распространения ограничений (включая методы разрешения геометрических ограничений), интервальный анализ и классическая теория оптимизации.

Взаимный интерес *isiCAD-2004* и ИМРО-2004 объясняется также тем, что значительная часть задач, решение которых рассматривается на этих мероприятиях, может быть задана в виде систем ограничений, представляемых как алгебраические уравнения и неравенства. Для решения таких систем ограничений в настоящее время все чаще используются методы интервального анализа, а также интервальные методы распространения ограничений. Целью совещания ИМРО-2004 является желание собрать вместе ученых, занимающихся исследованиями в области интервального анализа и распространения ограничений для обмена мнениями и обсуждения существующих проблем. Такое общение будет способствовать взаимному обогащению идеями и приведет к более широкому использованию методов интервальной математики в методах распространения ограничений, и наоборот. Что, в свою очередь, послужит толчком к более активному применению этих двух тесно связанных и взаимно дополнительных подходов в практических приложениях.

ИМРО-2004 является уже вторым российским научным мероприятием, посвященным интервальной математике и методам распространения ограничений. Его тематика традиционно охватывает:

- решение интервальных линейных систем уравнений;
- решение нелинейных алгебраических систем уравнений;
- исследование дифференциальных уравнений и динамических систем с интервальной неопределенностью;
- приложения методов интервальной математики и распространения ограничений;
- решение оптимизационных задач с использованием методов распространения ограничений.

Труды совещания адресованы исследователям, инженерам, аспирантам и студентам, которые изучают или применяют интервальные методы и технику распространения ограничений, а также всем интересующимся этими перспективными и увлекательными областями знаний.

Д.Я. Левин, А.Л. Семенов, С.П. Шарый

Новосибирск, июнь 2004

Задачи анализа с неопределенными данными – интервальность и/или случайность?

А.П. Воцинин*

В подавляющем большинстве прикладных задач исследователь имеет дело с неточными исходными данными (табл. 1), неопределенность которых порождается различными факторами.

Таблица 1

№	Данные	Источник неопределенности и неточности
1	Результаты измерений	Вариабельность, шумы, ошибки измерения (систематические и случайные), ошибки округления
2	Прогнозные данные	Незнание, неопределенность, неполнота информации, методические ошибки, ошибки округления и дискретизации
3	Экспертные оценки	Субъективность, незнание, ошибки округления

В зависимости от источника неточности и неопределенности данных в настоящее время используются различные модели описания неопределенных данных, включая вероятностную, нечеткую и интервальную модели. Каждая из этих моделей имеет свою парадигму, опирается на соответствующий теоретический аппарат, имеет свои методы анализа и область применения.

Вероятностно-статистическая модель [1–3] основана на предположении, что изучаемая переменная x является случайной величиной с заданной функцией плотности вероятности $f(x)$. Эту модель можно отнести к классическим моделям из-за ее глубокой теоретической и методической проработанности для многих важных приложений, включая регрессионный, корреляционный, факторный и дисперсионный анализ. Вероятностно-статистическая модель до сих пор является наиболее популярной и широко используется в многочисленных приложениях. Несомненным преимуществом, облегчающим ее применение, является наличие многочисленных статистических пакетов, среди которых Статистика, Statgraph, GPSS и др.

“Нечеткая модель” [4, 5] основана на понятии нечеткого множества S , которое описывается парой – нечеткой переменной x и ее функцией принадлежности $\mu_S(x)$. Формально нечеткое множество записывается в виде

$$S = \{(x, \mu_S(x)) : x \in X, 0 \leq \mu_S(x) \leq 1\}, \quad (1)$$

где x означает возможные значения нечеткой переменной в заданной области X ; $0 \leq \mu_S(x) \leq 1$ – функция принадлежности, задающая степень принадлежности конкретного значения x нечеткому множеству S . Обычно функ-

*ФГУП “ЦНИИАТОМИНФОРМ”, Москва.

ция принадлежности $\mu_S(x)$ задается экспертным путем на основе информации об источниках неопределенности переменной x . Понятие “степень принадлежности” в некотором смысле аналогично понятию вероятности, в частности, $\mu_S(x_1) = 0$ означает, что значение x_1 наверняка не принадлежит множеству S , а $\mu_S(x_2) = 1$ – что значение x_2 достоверно принадлежит множеству S . Однако функция принадлежности в отличие от плотности вероятности $f(x)$ не удовлетворяет условию нормировки.

В *интервальной модели* [6–9] неопределенность параметра x описывается границами его возможных значений в виде $[x] = [x_{\min}; x_{\max}]$. В отличие от теории вероятности внутри интервала $[x]$ не задается никакой вероятностной меры, т. е. все значения внутри интервала предполагаются равновероятными (не путать с равновероятными).

Ниже в докладе рассматривается ряд прикладных задач, в которых применяется как вероятностно-статистическая, так и интервальная модели и анализируются ситуации, в которых интервальная модель представляется более осмысленной и корректной.

Неопределенность единичного измерения

Пусть в результате измерения получена точечная оценка \hat{x} неизвестной измеряемой величины x .

Вероятностная модель измерения последовательно реализуется в экспериментальной физике. При этом ошибка измерения рассматривается как случайная, нормально распределенная величина с нулевым средним и известной дисперсией σ_x^2 . Неопределенность переменной x для заданной доверительной вероятности (обычно ее принимают равной 0.95) описывается в форме доверительного интервала

$$\hat{x} - 2\sigma_x \leq x \leq \hat{x} + 2\sigma_x. \quad (2)$$

В метрологии в отличие от приведенного выше подхода принята интервальная модель неопределенности. Предполагается, что измерение \hat{x} получено с помощью неточного прибора с известной абсолютной ошибкой измерения Δ , которая включает как систематическую, так и случайную погрешности. (Случай известной относительной ошибки сводится к рассматриваемому). Предполагается, что для любого \hat{x} выполняется условие $|x - \hat{x}| \leq \Delta$, которое естественным образом задает интервал неопределенности в виде.

$$[x] = [\hat{x} - \Delta; \hat{x} + \Delta] = [x_{\min}; x_{\max}]. \quad (3)$$

Если есть необходимость учесть другие факторы неточности, например ошибки округления, исследователь может расширить интервал неопределенности (3).

Сравнивая две приведенных модели, можно заметить, что модель (2) не позволяет учесть факторы неопределенности, не связанные со случайной вариабельностью, включая систематические ошибки измерения, ошибки округления. Кроме того, постулируемое в вероятностной модели нормальное распределение, которое задает неограниченный диапазон величины x , на практике часто оказывается неадекватным, например, для заведомо положительных

переменных. Интервальная модель позволяет учесть любые факторы неопределенности.

Выборка повторных наблюдений

Допустим теперь, что имеется выборка $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ повторных наблюдений величины x . При анализе выборочных данных необходимо решить две задачи:

- проверить однородность выборки, т. е. отсутствие в ней выбросов;
- найти наиболее точную выборочную оценку математического ожидания x_0 величины x и оценить ее точность.

В рамках вероятностно-статистической модели в предположении, что выборка взята из нормального распределения с дисперсией σ_x^2 точечная оценка величины x и ее среднеквадратичное отклонение находится по формуле

$$\bar{x} = \sum \hat{x}_i/n, \quad \sigma_{\bar{x}} = \sigma_x/\sqrt{n}. \quad (4)$$

Считается, что оценка (4) является несмещенной и состоятельной, т. е. $\sigma_x/\sqrt{n} \rightarrow 0$, $\bar{x} \rightarrow x_0$ при $n \rightarrow \infty$. Нетрудно заметить, что в рамках модели с постулируемым нормальным распределением само понятие “выброса” становится не совсем корректным.

В рамках интервальной модели предполагается заданным интервал неопределенности каждого выборочного наблюдения

$$[x]_i = [x_i - \Delta x; x_i + \Delta x],$$

где Δx – известная ошибка измерения или выборочного наблюдения. Совокупность интервальных измерений образует “интервальную выборку”

$$\{[x]_1 \dots [x]_i \dots [x]_n\}.$$

При этом неопределенность выборочных значений определяется не только случайной изменчивостью, но и наличием систематических составляющих погрешности. Представим ошибку Δx в виде суммы

$$\Delta x = \Delta x_1 + \Delta x_2, \quad (5)$$

где значение Δx_1 определяет переменную, а Δx_2 – постоянную составляющую неопределенности.

Если выборка однородна, то каждое интервальное наблюдение содержит неизвестное истинное значение x_0 , т. е. $x_0 \in [x]_i, \forall i = 1, \dots, n$.

Если существует интервальное наблюдение $[x]_j$ такое, что оно не пересекается с другими интервальными наблюдениями, его естественно рассматривать как выброс и исключить из выборки для обеспечения ее однородности.

Введем понятие интервального выборочного среднего $[\bar{x}]$, которое определяется как пересечение выборочных интервальных измерений

$$[\bar{x}] = [\max_i(x_i^-); \min_i(x_i^+)] = [\bar{x}]. \quad (6)$$

Можно показать, что интервальное среднее $[\bar{x}]$, найденное по однородной выборке, обладает следующими свойствами:

- Средняя точка интервала $[\bar{x}]$ совпадает со статистической выборочной оценкой медианы. Ширина интервала $[\bar{x}]$ определяет точность оценки с учетом случайных и систематических составляющих.
- При отсутствии систематической составляющей ($\Delta x_2 \equiv 0$) и увеличении числа выборочных значений $[\bar{x}] \rightarrow x_0$ при $n \rightarrow \infty$. В общем случае среднее арифметическое (4) не принадлежит интервальному среднему $[\bar{x}]$. В ходе численного моделирования установлено, что при равномерном распределении выборочных измерений ширина интервального среднего $[\bar{x}]$ убывает со скоростью $1/n$, тогда как стандартное отклонение статистического среднего убывает со скоростью $1/\sqrt{n}$. Установлено также, что
- Если в выражении (5) обе составляющие ошибки не равны нулю, то

$$[\bar{x}] \xrightarrow{n \rightarrow \infty} [x_0 - \Delta_2(x); x_0 + \Delta_2(x)], \quad (7)$$

т. е. при любом числе измерений полу ширина интервального среднего будет не меньше значения Δx_2 . Это свойство делает бессмысленным увеличение числа измерений, если ширина интервального среднего соизмерима со значением Δx_2 . Этот факт обычно игнорируется при статистическом анализе и арифметическое среднее трактуется как состоятельная оценка.

Сравнивая два подхода, применяемых при анализе выборочных измерений, можно констатировать, что в рамках статистической модели неопределенность выборочных значений, не связанная со случайностью, игнорируется. При этом становится необоснованным утверждение о состоятельности оценок и возможности достижения любой заданной точности оценки выборочного среднего путем увеличения числа повторных измерений.

Ошибки косвенных измерений и эмпирических формул

Данные прямых измерений часто используются для нахождения по известным формулам так называемых косвенных измерений. Кроме того, в технических, экологических и экономических приложениях наряду с объективными законами широко используются так называемые “эмпирические модели”, которые включают экспериментальные, неточно задаваемые переменные. Обе задачи приводят к необходимости определения ошибки результирующего значения функции $y = f(x_1, x_2, \dots, x_k)$ с неточно заданными параметрами x_i .

Применение вероятностной модели для решения этой задачи сопряжено со значительными трудностями, так как в этом случае необходимо найти распределение случайной величины y при заданном совместном распределении в общем случае зависимых случайных величин $x_i, i = 1, \dots, k$. Результирующее распределение находится как свертка интегралов. Сравнительно просто находится распределение суммы независимых случайных величин. Однако, уже

для отношения даже двух переменных – это нетривиальная операция, при нормальном распределении приводящая к разрыву, в нулевой точке знаменателя.

В рамках интервального анализа исследуемая функция неточных переменных записывается в виде интервальной функции $[y] = f([x_1], [x_2], \dots, [x_k])$, конечные границы которой находятся как

$$[y] = [y_{\min}; y_{\max}] = \left[\min_{\vec{x} \in [X]} f(\vec{x}), \max_{\vec{x} \in [X]} f(\vec{x}) \right], \quad (8)$$

где $[X]$ – прямоугольная гиперпризма, образованная интервалами $[x_i]$.

Для значительного класса функций, включая функции линейные по интервальным параметрам или полиномы вида

$$[x_1]^{\beta_1} \cdot [x_2]^{\beta_2} \cdot \dots \cdot [x_k]^{\beta_k}, \quad (9)$$

границы интервала неопределенности $[y]$ могут быть записаны в аналитическом виде, в том числе и в форме, привычной для метрологов

$$[y] = [y_{\min} = y_{\text{cp}} \cdot (1 - \delta_y); y_{\max} = y_{\text{cp}} \cdot (1 + \delta_y)], \quad (10)$$

где y_{cp} – средняя точка интервала неопределенности, $\delta_y \cong \sum_{i=1}^k \beta_i \delta_{x_i}$ – относительная ошибка результата, δ_{x_i} – относительные ошибки измерения переменных x_i .

Полиномы (9) широко используются в многочисленных приложениях в качестве эмпирических моделей. При этом для одного и того же явления часто существует несколько альтернативных эмпирических формул, отличающихся как структурой, так и числом переменных. Например, для оценки пространственно-временных характеристик аварийного загрязнения водотоков предложено более 10 различных эмпирических формул, которые отличаются количеством учитываемых переменных. Входящие в них переменные, которые включают скорость течения, глубину и ширину реки, ее гидравлический радиус и уклон, коэффициент шероховатости русла и др., измеряются с большими погрешностями [16].

В этих условиях включение переменной в модель с одной стороны увеличивает ее полноту, т. е. снижает неопределенность результата, а с другой увеличивает ошибку модели, связанную с неточностью измерения переменных. Допустим, что модель (9) содержит полный перечень k переменных. Тогда если в модель включено $m < k$ переменных возникает ошибка из-за ее упрощения и общую ошибку результата можно записать в виде суммы

$$\delta_y(m) = \delta_{y1}(m) + \delta_{y2} = \delta_{y1}(m) + \sum_{i=1}^m \beta_i \delta_{x_i}, \quad (11)$$

где первое слагаемое связано с неполнотой модели, а второе – с ошибками измерения переменных. Для величины $\delta_{y1}(m)$ выполняются естественные условия $\delta_{y1}(m) = 0$ при $m = k$ и $\delta_{y1}(m) = 1$, т. е. имеет место максимальная неопределенность при $m=0$.

Допустим, что вклад каждой переменной в модели одинаков. Тогда при одинаковых ошибках измерения переменных $\delta_{x_i} = \delta_x$ и степенях β_i формула (11) приобретает простой вид

$$\delta_y(m) = (1 - m/k) + m \cdot \delta_x. \quad (12)$$

При этом зависимость общей ошибки результата за счет неполноты и неточности измерений от числа переменных имеет минимум, равный $\delta_{\min y} = 1 - (1 - k \cdot \delta_x)/(1 + k \cdot \delta_x)$ в точке $m_{\text{opt}} = k/(1 + k \cdot \delta_x)$.

Число переменных m_{opt} можно считать оптимальным для эмпирической модели, так как дальнейшее увеличение числа переменных будет приводить к увеличению общей ошибки модели. При $k = 10$ и $\delta_x = 0.05$ оптимальное число переменных для рассматриваемой модели не превышает $6 \div 7$.

Сглаживание данных: прямые, неявные и обратные функции

Наиболее часто в приложениях для сглаживания экспериментальных данных, полученных в результате n -измерений и заданных в виде $(n \times m)$ -матрицы X и $(n \times 1)$ -вектора Y используются линейно параметризованные функции вида [2, 3]

$$y = \beta_1 f_1(\mathbf{x}) + \dots + \beta_j f_j(\mathbf{x}) + \dots + \beta_m f_m(\mathbf{x}) + e, \quad (13)$$

где y – измеряемая выходная переменная, $\mathbf{x} = (x_1 \dots x_k)^T$ – вектор k входных переменных, f_j – известные базисные функции, β_j – неизвестные коэффициенты.

В рамках статистического подхода для нахождения оценок параметров обычно используется множественный регрессионный анализ, основанный на методе наименьших квадратов [1–3]. Точечные оценки параметров и их ковариационная матрица находятся по формулам как

$$\mathbf{B} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{Y}, \quad \mathbf{D}(\mathbf{B}) = \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1}, \quad (14)$$

где \mathbf{B} – вектор оценок неизвестных параметров β_j , $\mathbf{F} = \{f_j(\mathbf{x}_i)\}$ – $(n \times m)$ -матрица значений базисных функций в n опытах, \mathbf{Y} – вектор наблюдений выходной переменной.

Известно, что при нормальной аддитивной ошибке оценки (14) обладают целым рядом оптимальных свойств, в частности, являются несмещенными, эффективными и наилучшими в классе линейных оценок. Распределение оценок подчиняется нормальному закону, что позволяет при заданной доверительной вероятности α определить доверительные интервалы для оценок параметров.

Сложившаяся парадигма регрессионного анализа, восходящая к Фишеру, включает два главных постулата: разделение переменных на входные (независимые) и выходные (зависимые) переменные; присутствие аддитивных, нормально распределенных случайных ошибок только в выходной переменной. Именно при этих допущениях получены наиболее сильные результаты регрессионного анализа, связанные с оптимальностью найденных оценок, их несмещенностью, состоятельностью и т. п.

Вместе с тем, на практике существует широкий класс прикладных задач, когда один или одновременно оба постулата регрессионного анализа нарушаются. На практике модель со случайной, аддитивной нормально распределенной ошибкой является скорее исключением, чем правилом. Обычно ошибки

неизбежны как в выходной, так и во входных переменных. При этом их величина часто зависит от текущего значения переменной. Часто нет оснований считать источником неопределенности данных случайность и, следовательно, приписывать ошибкам определенную плотность вероятности. Например, при измерении неточными приборами, анализе ошибок квантования и округления, обработке изображений и т. п. более естественно описывать неопределенность в терминах интервала возможных значений.

В картографии и при обработке фотоизображений отношения между переменными описываются неявной функцией $f(x, y, b_1, \dots, b_m) = 0$. При этом отсутствует разделение переменных на входные и выходные. Делались многочисленные попытки смягчить достаточно жесткую систему приведенных выше допущений. Например в конъюнктном анализе допускается наличие ошибок как на входе так и на выходе. Однако в связи со сложностью аналитического представления функции правдоподобия в основном разработаны эвристические методы, в частности связанные с группированием данных (метод инструментальных переменных).

В рамках интервального подхода задачи сглаживания решаются в предположении, что экспериментальные наблюдения заданы в интервальной форме, при этом допускаются ошибки как во входных, так и в выходной переменной [9–15].

Поясним идею сглаживания интервальных данных на простом примере, рис. 1,а, где вертикальными жирными линиями изображены пять интервальных наблюдений величины y , полученных в эксперименте при $x \in [3; 7]$.

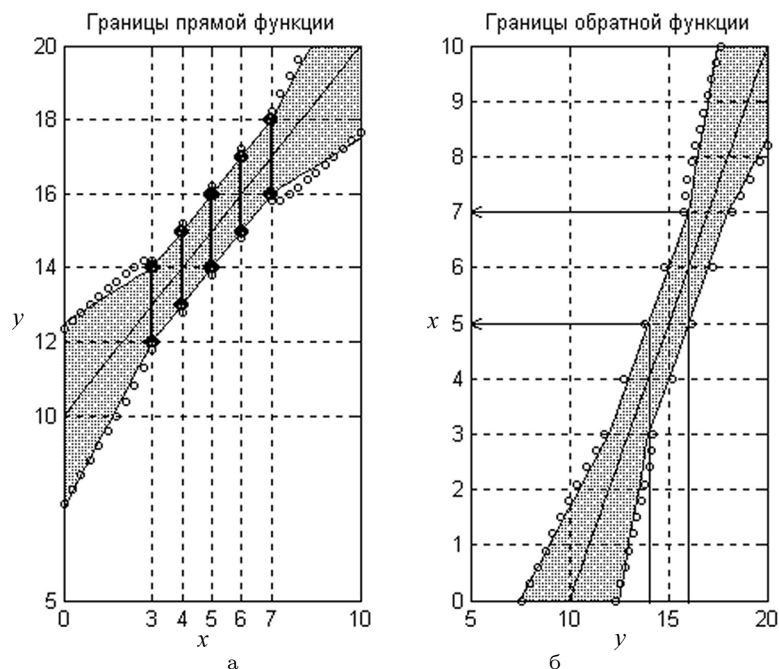


Рис. 1

Для сглаживания интервальных данных и нахождения средней линии, а также верхней и нижней границ интервала неопределенности используются линейные функции. При этом находятся в диапазоне $x \in [3; 7]$ коэффициенты уравнений границ рассчитываются методом наименьших квадратов соответственно по верхним и нижним граничным значениям интервальных наблюдений, а коэффициенты средней линии – по средним точкам интервальных наблюдений. Уравнение средней линии имеет вид $y = 10 + x$.

Очевидно, что в диапазоне $x \in [3; 7]$ через интервальные наблюдения можно провести бесчисленное множество прямых, однако существуют две “экстремальные”, одна из которых проходит через точки с координатами (3; 12) и (7; 18), а другая – через точки (3; 14) и (7; 16). Эти прямые порождают более широкий интервал неопределенности вне диапазона $x \in [3; 7]$, ширина тем больше, чем ближе значение x к краям диапазона $[0; 10]$.

Границы результирующего интервала неопределенности в диапазоне $x \in [0; 10]$ (на рис. 1,а выделен темным цветом) описываются линейными сплайн-функциями с тремя участками. Интервальный подход к данной задаче позволяет достаточно просто решить задачу аппроксимации этих сплайн-функций гладкими аналитическими функциями из класса конических сечений. Для решения задачи кроме имеющихся 5 интервальных наблюдений генерируются также дополнительные точки на участках $x \notin [3; 7]$, которые рассчитываются по найденным уравнениям прямых. Аналитические уравнения границ ищутся в форме *явной функции* [11]

$$x^2 = \beta_1 y^2 + \beta_2 yx + \beta_3 x + \beta_4 y + \beta_5. \quad (15)$$

Векторы коэффициентов уравнения (15), найденные методом наименьших квадратов для данных рис. 1,а, составили:

$$\text{для нижней границы: } \vec{\beta}_{\min} = [-1.05; 2.12; -19.16; 18.35; -78.85] \quad (16)$$

$$\text{для верхней границы: } \vec{\beta}_{\max} = [-1.05; 2.12; -24.36; 23.48; -129.84]. \quad (17)$$

Для целей прогноза уравнения (16) и (17) приводятся к каноническому виду и записываются в виде двух явных функций прогноза $y_{\min} = \varphi(\vec{\beta}_{\min}, x)$, $y_{\max} = \varphi(\vec{\beta}_{\max}, x)$. Полученные предсказанные значения, которые отмечены на рис. 1,а круглыми маркерами, обеспечивают достаточно высокую точность описания границ интервала неопределенности.

В теории измерений важное место занимает *задача калибровки*, т. е. построения градуировочной характеристики сенсора по неточным данным эксперимента. При этом для простейшего линейного случая, когда функция преобразования сенсора имеет вид линейной функции $y = \beta_1 + \beta_2 x$, необходимо найти обратную модель $x = a_1 + a_2 y$ и оценить коридор ее ошибок.

В рамках статистической парадигмы эта задача относится к классу наиболее сложных и слабо формализуемых. Ее решению посвящен отдельный раздел статистики – теория калибровки. Трудности вызываются следующими причинами. Если по данным эксперимента, найдены несмещенные статистические оценки коэффициентов β_i прямой модели, то при переходе к обратной модели это свойство теряется, кроме того, возникают теоретические сложности при по-

строении доверительного интервала. Если же рассчитываются оценки коэффициентов обратной модели, то зависимой становится переменная y , измеряемая с ошибками, что приводит к нарушению предпосылок регрессионного анализа.

В рамках интервального анализа задача решается достаточно просто даже при наличии ошибок в обеих переменных x и y , так как при наличии аналитических уравнений границ переход к обратной модели не вызывает никаких трудностей [12, 13]. Если данные рис. 1,а воспринимать как данные калибровочного эксперимента, то, имея уравнения средней линии и границ, легко получить искомую градуировочную характеристику, т.е. обратную модель и ее границы, представленные на рис. 1,б. В частности уравнение градуировочной характеристики имеет вид $x = -10 + y$. Коридор ошибок сенсора залит темным цветом.

На рис. 1,б наглядно проявляется один феномен, который не учитывается при статистическом подходе, связанный с рабочим диапазоном сенсора, в котором ошибка определена на основе исходных данных, а не по прогнозным значениям. Проясним этот момент.

Для прямой модели рис. 1,а в рабочем диапазоне переменной $x \in [3; 7]$ имела место постоянная абсолютная ошибка измерения переменной y , равная ± 1 . Для обратной модели постоянная абсолютная ошибка измерения переменной x , равная ± 1 , сохраняется лишь в диапазоне $x \in [5; 7]$, который отмечен на рис. 1,б стрелками, которому соответствует сравнительно узкий рабочий диапазон сенсора $y \in [14; 16]$. Вне этого диапазона ошибки измерения x стремительно возрастают.

Этот факт отражает объективную реальность информационной достоверности результатов внутри и вне пределов диапазонов калибровочного эксперимента.

В качестве преимущества интервального подхода к задаче калибровки необходимо отметить также тот факт, что в отличие от статистического подхода могут быть использованы оба варианта расчета: построение прямой модели и ее обращение или непосредственное построение обратной модели. Причем оба варианта приводят к одним и тем же результатам.

Оценка экономических рисков инвестиционных проектов

Для оценки эффективности инвестиционного проекта в Методических рекомендациях [17], основанных на методологии UNIDO, рекомендуется ряд интегральных показателей, среди которых: чистый доход (ЧД), чистый дисконтированный доход (ЧДД), внутренняя норма доходности (ВНД), срок окупаемости (Ток), индекс доходности дисконтированных затрат (ИДДЗ). Предполагается, что ожидаемые показатели эффективности проекта вычисляются с учетом факторов риска и неопределенности.

В соответствии с [17] *риск* для инвестиционных проектов трактуется как совокупность трех составляющих: событие, связанное с риском; ущерб (величина денежной суммы, подвергаемой риску) и *вероятность риска*. Для оценки риска инвестиционных проектов используется так называемый сценарный анализ, когда проект оценивается для конечного числа различных сценариев, каждому из которых приписывается определенная вероятность p_i . Интеграль-

ный показатель ожидаемой эффективности проекта $\Theta_{\text{ож}}$, риск неэффективности проекта P_{Θ} и средний ущерб Y_{Θ} от реализации неэффективных проектов вычисляются по формулам

$$\Theta_{\text{ож}} = \sum \Theta_i p_i, \quad P_{\Theta} = \sum p_i^-, \quad Y_{\Theta} = \sum p_i^- \Theta_i / P_{\Theta}. \quad (18)$$

Суммирование в двух последних формулах ведется только для сценариев с отрицательными интегральными эффектами Θ_j .

Учитывая, что случайность является только одним и не всегда главным источником неопределенности, которая в основном связана с незнанием, неполнотой и неточностью информации в последнее десятилетие в зарубежной литературе предложен ряд новых моделей к описанию неопределенности в терминах не вероятности, а возможности неблагоприятного события. При этом вместо классической вероятностной модели (18) находит более адекватная модель экономического риска, записываемая в виде

$$\text{Ущерб} = \text{неопределенность} \times \text{денежные потери при неблагоприятном событии}.$$

При этом для описания неопределенности предлагаются разные подходы и теоретические методы, среди которых теория нечетких множеств, робастный байесовский подход, теория неточных вероятностей, модель интервальной вероятности и модель с ограниченными вероятностями. Последние три модели в этом списке рассматриваться как симбиоз вероятностного и интервального подхода.

Ниже рассматривается интервальная модель риска [15, 16, 18], концепция которой основана на преобразовании входных интервальных показателей проекта в выходные интервальные значения интегральных критериев его эффективности.

Интервальная модель экономического риска для проектов, реализованная в программе *Интервал-Инвест* включает следующие этапы:

1. В интервальной форме $[z_{\min}; z_{\max}]$ задаются возможные границы параметров внешних условий и технико-экономических показателей и на их основе программно формируются два экстремальных сценария – пессимистический и оптимистический. В каждом из них в различных комбинациях фигурируют верхние и нижние граничные значения параметров $z_{\min}; z_{\max}$. Вводятся предельно допустимые, пороговые значения d_i для критериев эффективности. Применительно к ЧДД $d = 0$ и при ЧДД < 0 результат проекта по этому критерию трактуется как отрицательный.
2. По известным формулам рассчитываются интервальные критерии $[x_i]$ эффективности проекта. Показано, что значения критериев x_i , которые определяются через кумулятивные значения финансовых потоков, при любых комбинациях исходных данных внутри границ $z_{\min}; z_{\max}$ будут принадлежать интервалам $[x_i]$.
3. Рассчитывается риск (возможность) получения отрицательного результата, когда $x_i < d_i$ (или $x_i > d_i$ для критериев, связанных с затратами).

При расчетах риска r неблагоприятного события $x < d$ при заданном интервальном значении критерия $[x]$ используется следующий достаточно естественный подход:

- если $d \geq x_{\max}$, то $R = 1$;
- если $d \in [x]$, то риск вычисляется как отношение длины отрезка, на котором отношение выполняется, к ширине всего интервала $[x]$, т. е. по формуле

$$R = (d - x_{\min}) / (x_{\max} - x_{\min}). \quad (19)$$

Необходимо отметить, что в случае постулирования равномерного распределения на интервале $[x]$ полученные по формуле (19) значения будут совпадать со значением вероятности соответствующего события.

Ниже в табл. 2 приведены результаты расчета критериев эффективности и экономических рисков проекта АЭС, полученные по программе *Интервал-Инвест*. Пороговые значения критериев составляли: для ЧД и ЧДД $d = 0$ (не менее, чем), для ВНД $d = 0.25$ (не менее, чем), для ИДДЗ $d = 1.2$ (не менее, чем), для тарифа безубыточности $d \leq 2$ (не более, чем).

Таблица 2

Показатели эффективности / сценарии	Пессим.	Базовый	Оптим.	Риск
1. Чистый доход, млн. долл.	-1393.1	1222.2	3287.0	0.30
2. ЧДД, млн. долл.	-2373.3	-855.1	521.8	0.82
3. ВНД, %	0.0	2.5	7.5	1.00
4. ИДДЗ	0.36	0.76	1.34	0.86
5. Тариф безубыточности, цент/кВт·ч	5.53	2.98	1.87	0.53
6. Срок окупаемости, лет	Не окуп.	Не окуп.	27	
7. Дисконтированные затраты, млн. долл.	3556	2673	2484	

В экономической литературе рекомендуются следующие градации риска: $0 \div 0.1$ – минимальный; $0.1 \div 0.3$ – малый; $0.3 \div 0.4$ – средний; $0.4 \div 0.6$ – высокий; $0.6 \div 0.8$ – максимальный; $0.8 \div 1$ – критический. В соответствии с этой градацией для исследуемого в условиях неопределенности проекта АЭС критические значения риска получены по всем критериям за исключением ЧД (0.3) и тарифа безубыточности (0.53). Это позволяет сделать вывод о неэффективности проекта при сделанных допущениях относительно возможных сценариев реализации проекта и его технико-экономических характеристиках.

Заключение

Приведенные иллюстративные примеры показывают, что применение интервального анализа позволяет снять многие проблемы и методические сложности, возникающие при решении прикладных задач статистическими методами. В рамках интервального анализа неопределенность исходных данных может иметь разные источники и природу. Интервал неопределенности позволяет описать широкий класс неопределенных, неоднозначных, вариабельных и неточных исходных данных. Значения ошибок в исходных данных могут колебаться в широких пределах. Результаты, полученные с помощью в рамках

парадигмы интервального анализа, имеют ясную и четкую интерпретацию в терминах интервалов и областей неопределенности. Основной проблемой интервального анализа является корректное определение интервалов неопределенности на основе различных исходных данных при наличии различных источников неопределенности переменной. Приведенные результаты дают основание сделать заключение, что применение интервального анализа к широкому спектру прикладных задач с ограниченными ошибками и неопределенностью в данных имеет больших перспективы.

Список литературы

- [1] Вентцель Е.С. Теория вероятности. – М.: Физматгиз, 1962. – 563 с.
- [2] Химмельблау Д. Анализ процессов статистическими методами. – М.: Мир, 1973. – 948 с.
- [3] Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Финансы и Статистика, 1986. – 345 с.
- [4] Bellman G., Zadeh L. Decision making in fuzzy environment // Management Science. – 1970. – Vol. 17, № 1.
- [5] Zimmerman H. Fuzzy Sets. Decision making and expert systems. – Dordrecht: Kluwer Acad. Publishers, 1987.
- [6] Moore R. E. Interval analysis. – Englewood, Cliffs, N.Y.: Prentice-Hall, 1966.
- [7] Алефельд Г., Херцберг Ю. Введения в интервальные вычисления. – М.: Мир, 1987. – 370 с.
- [8] Калмыков С.А., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. – Новосибирск: Наука, 1986.
- [9] Шарый С.П. Интервальные алгебраические задачи и их численное решение. – Дисс. ... докт. физ.-мат. наук. – 2000. – 322 с.
- [10] Вощинин А.П., Бочков А.Ф., Сотиров Г.Р. Интервальный анализ данных как альтернатива регрессионному анализу // Заводская Лаборатория. – 1990. – № 7. – С. 76–81.
- [11] Вощинин А.П. Метод анализа данных с интервальными ошибками в задачах проверки гипотез и оценивания параметров неявных и линейно параметризованных функций // Заводская Лаборатория. – 1998. – №7.
- [12] Вощинин А.П., Скибицкий Н.В. Интервальный метод калибровки // Системы и датчики. – 2000. – № 7.
- [13] Voschinin A., Skibitsky N. Interval calibration model of multisensor systems // IDAACS'2003. – Lviv, Ukraine, 2003.
- [14] Вощинин А.П. Интервальный анализ данных: развитие и перспективы // Заводская лаборатория. – 2002. – № 1.
- [15] Voschinin A., Tyurin A. Interval identification of time series using readings with bounded errors // 5th International Scientific-Technical Conf. ProcessControl, Pordubice. – 2002.
- [16] Gorsky V., Shvetzova-Shilovskaya T., Voschnin A. Risk assessment of accident involving environmental high-toxicity substances // J. of Hazardous Materials. – 2000. – № 78.
- [17] Виленский П.Л., Лившиц В.Н., Смоляк С.А. Оценка эффективности инвестиционных проектов (теория и практика). – М.: Изд-во “Дело”, 2001.
- [18] Вощинин А.П., Тюрин А.В. Интервальные модели оценки эффективности инвестиционных проектов электроэнергетики // Тр. междунар. конф. Control 2003. – Москва, 2003.

Validated constraint solving – practicalities, pitfalls, and new developments

R. Baker Kearfott*

Abstract. Many constraint propagation techniques iterate through the constraints in a straightforward manner, but can fail because they do not take account of the coupling between the constraints. However, some methods of taking account of this coupling are local in nature, and fail if the initial search region is too large. We put into perspective newer methods, based on linear relaxations, that can often replace brute-force search by solution of a large, sparse linear program.

Robustness has been recognized as important in geometric computations and elsewhere for at least a decade, and more and more developers are including validation in the design of their systems. We provide citations to our work to-date in developing validated versions of linear relaxations.

This work is in the form of a brief review and prospectus for future development. We give various simple examples to illustrate our points.

Introduction

A very general problem, specific instances of which are important throughout CAD, as well as computing in general and operations research, is the general global optimization problem

$$\begin{aligned}
 &\text{minimize } \phi(x) \\
 &\text{subject to } c_i(x) = 0, \quad i = 1, \dots, m_1, \\
 &\quad \quad \quad g_i(x) \leq 0, \quad i = 1, \dots, m_2, \\
 &\text{where } \phi : \mathbf{x} \rightarrow \mathbb{R} \text{ and } c_i, g_i : \mathbf{x} \rightarrow \mathbb{R}, \text{ and where } \mathbf{x} \subset \mathbb{R}^n \quad (1) \\
 &\text{is the hyperrectangle (box) defined by} \\
 &\quad \quad \quad \underline{x}_{i_j} \leq x_{i_j} \leq \bar{x}_{i_j}, \quad 1 \leq j \leq m_3, i_j \text{ between } 1 \text{ and } n, \\
 &\text{where the } \underline{x}_{i_j} \text{ and } \bar{x}_{i_j} \text{ are constant bounds.}
 \end{aligned}$$

If ϕ is constant or absent, problem (1) becomes a general constraint problem; if, in addition $m_2 = m_3 = 0$, problem (1) becomes a nonlinear system of equations.

In general, and indeed, for many specific instances, it is theoretically impossible to avoid exponential time when solving problem 1; see [16]. However, a plethora of techniques that work well for a number of instances of interest have been developed (although none of these are universally practical). Particularly prominent is a set of techniques where the variables and, possibly, some of the parameters defining the objective ϕ and the constraints c_i and g_i are only approximately known initially; in such cases, we can use the relationships among the constraints and the

*University of Louisiana, Lafayette, Louisiana, USA.

objective to compute tighter bounds on the variables x_i . Calling this general idea *subdefinite computations*, Narin'yani and his group developed it starting in the early 1980's [17]; an example implementation of these ideas is UNICALC [1] of Semenov et al. This paradigm also lends itself naturally, with directed roundings, to validated computation, that is, to computation in which the results are valid not just approximately, but with the certainty of a mathematical proof.

Large numbers of researchers have developed various tools and corresponding theories for filtering the bounds on the x_i through the constraints to obtain narrower bounds, a process often described as *constraint propagation*, a burgeoning field of artificial intelligence. A succinct description of the general constraint propagation framework appears in [2]. Benhamou (ibid.) attributes the introduction of interval constraints (equivalent to subdefinite computations) within this constraint propagation framework to Cleary [4]. In the notation of [2], such interval constraint propagation depends upon *narrowing operators* $N_{\mathcal{C}}(\mathbf{x})$, where \mathbf{x} is a box and \mathcal{C} is the set of constraints¹. Application of $N_{\mathcal{C}}$ returns a narrower box $\tilde{\mathbf{x}}$, where points in \mathbf{x} that are infeasible with respect to the constraints \mathcal{C} have been eliminated². The constraint propagation community has defined various *consistency conditions*, based on how N eliminates portions of \mathbf{x} , and individual advances in the field often are in the form of new consistency conditions that can eliminate portions of \mathbf{x} that other conditions cannot.

In this paper, we focus on three consistency conditions in common use, to which we refer as follows:

- 1) basic constraint propagation,
- 2) interval Newton narrowing,
- 3) linear relaxations

Although there are other conditions (such as those found in [22] or many other places), these three represent what is included in our own work, and also are representative of three schools which have only recently begun to merge. Basic constraint propagation is a fundamental narrowing process in the constraint propagation community, while interval Newton narrowing has been the predominant process used by those in the interval analysis community doing global optimization³. Linear relaxations, developed within the global optimization community at large, although very successful at solving practical problems such as those in [21] and [5], have only recently been recognized within the interval analysis community.

In the remainder of this paper, we will give illustrative examples of these three techniques, highlighting their deficiencies and citing literature where appropriate.

¹Here, \mathcal{C} represents the constraints c , g , bound constraints of problem (1), and possibly constraints based on setting the gradient of the Lagrange function to zero, in the case that ϕ is present and non-constant.

²In the global optimization context where ϕ is present, in addition to eliminating portions of \mathbf{x} based on feasibility, we may extend the interpretation of N to include elimination of portions of \mathbf{x} if a lower bound for ϕ over the feasible portion of \mathbf{x} is less than a known upper bound on the global optimum.

³Bounds on the objective ϕ are also used extensively within this school, but mainly in a rudimentary way to reject boxes, and not to narrow boxes in the sense of a narrowing operator.

1. Basic Constraint Propagation

Our own first examination of basic constraint propagation appears in [8], where we call it “substitution-iteration.” Following our ideas in [8], basic constraint propagation can be thought of as a nonlinear version of Gauss–Seidel iteration. That is, we are given initial bounds for the variables, and solve for a variable within a constraint. Plugging in the bounds for the other variables may then give narrower bounds on that variable. If so, then we solve for the other variables in constraints in which the narrowed variable occurs, and repeat the process with these other variables. We continue this until no more variables can be significantly narrowed.

Example 1. Take the constraint system

$$c_1(x) = x_1^2 - 2x_2, \quad c_2(x) = x_2^2 - 2x_1, \quad x_1 \in [-1, 1], \quad x_2 \in [-1, 1].$$

Selecting x_2 in c_1 , we obtain $x_2 = x_1^2/2$. Plugging $x_1 = [-1, 1]$ into $x_1^2/2$, we obtain $x_2 \in [0, 0.5]$, a significant narrowing. We now solve c_2 for x_1 to obtain $x_1 = x_2^2/2$. Plugging the narrower value of x_2 into $x_2^2/2$, we obtain $x_1 \in [0, 0.125]$, obtaining a narrower value of x_1 . We can now use the first equation again to obtain an even narrower value for x_2 . After only a few iterations of this process (using INTLAB [20]), we obtain intervals for x_1 and x_2 containing $x_1 = 0$, $x_2 = 0$ and with widths less than 10^{-16} .

Just as in the classical linear Gauss–Seidel method, this process will only converge if an ordering of the constraints and variables can be found such that the corresponding Jacobi matrix is diagonally dominant.

Example 2. Take the constraint system

$$c_1(x) = x_1^3 + x_1 - x_2, \quad c_2(x) = -2x_1 - x_2, \quad x_1 \in [-.5, .5], \quad x_2 \in [-.25, .25].$$

In Example 2, there is a unique feasible point (i.e. a point at which the constraints are consistent) within the initial bounds at $x_1 = 0$, $x_2 = 0$. However, solving c_1 for x_2 as in Example 1 gives $x_2 = (x_1^3 + x_1)$. There is interval overestimation when we plug $x_1 = [-.5, .5]$ into $x_1^3 + x_1$ and use naive interval arithmetic, but that is not the only issue here, as we’ll see as we follow through the computations; we’ll use the exact range of $(x_1^3 + x_1)$ for $x_1 \in [-.5, .5]$, that is, $x_2 \in [-.625, .625]$, which is no improvement. Similarly, if we solve for x_2 in the second equation, we obtain that the range of $-2x_1$ over $x_1 \in [-.5, .5]$ is $x_2 \in [-1, 1]$, not an improvement. The only remaining alternatives for improvement are to solve for x_1 in c_1 or c_2 . Solving for x_1 in c_2 gives no improvement, but solving for x_1 in c_1 (using the cubic equation) and plugging in $x_2 \in [-.25, .25]$ gives $x_1 \in [-.237, .237]$, an improvement. Now, solving for x_2 in c_2 gives $x_2 = -2x_1 \in [-.474, .474]$, not a narrowing. Thus, the process becomes stationary at $x_1 \in [-.237, .237]$, $x_2 \in [-.25, .25]$. At this point, a constraint propagation procedure employing only this basic technique would bisect \mathbf{x} , to increase the number of boxes, a potentially expensive procedure.

The difficulty in Example 2 is that there is no ordering of the variables and constraints for which the resulting Jacobi matrix at the solution $x = (0, 0)$ is diagonally dominant. In such cases, a process that somehow decouples the system would be superior to basic constraint propagation. One such procedure is interval Newton methods.

2. Interval Newton Narrowing

Interval Newton methods are ubiquitous throughout the interval analysis literature. A good reference for the theory is [18]; in our own work, we have introduced and summarized interval Newton methods in [10]. At their most basic, interval Newton methods work on systems of n equations in n unknowns⁴. Systems with both equality and inequality constraints and optimization problems can be handled by forming the Lagrange multiplier system (or more generally, the Fritz–John equations).

An interval Newton operator is an operator of the form $\mathbf{v} = \mathbf{N}(F, \mathbf{x}, \tilde{x})$, where \mathbf{v} is an interval vector that bounds the solution set to

$$\mathbf{A}\mathbf{v} = -F(\tilde{x}), \quad (2)$$

where \mathbf{A} is some Lipschitz matrix for F or a slope matrix⁵ for F centered at \tilde{x} . (For example \mathbf{A} can consist of element-wise interval extensions of the Jacobi matrix of F over \mathbf{x} .) Generally, the bounds \mathbf{v} are obtained by first preconditioning (2) by multiplying \mathbf{A} and $-F$ by a preconditioner matrix Y , where Y serves to partially decouple the equations (and make the resulting matrix $Y\mathbf{A}$ more like the identity matrix).

In Example 2 above,

$$F(x) = \begin{pmatrix} c_1(x) \\ c_2(x) \end{pmatrix} = \begin{pmatrix} x_1^3 + x_1 - x_2 \\ -2x_1 - x_2 \end{pmatrix},$$

and an element-wise interval extension of the Jacobi matrix of F over the initial \mathbf{x} is

$$\mathbf{F}'(\mathbf{x}) \in \begin{pmatrix} [1, 1.75] & -1 \\ -2 & -1 \end{pmatrix}.$$

If the inverse of the midpoint matrix for $\mathbf{F}'(\mathbf{x})$ is used as a preconditioner matrix, then, if $\tilde{x} = (0, 0)^T$, the preconditioned system becomes

$$\begin{pmatrix} [0.8888, 1.1112] & [0.0000, 0.0000] \\ [-0.2223, 0.2223] & [0.9999, 1.0001] \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and, using the interval Gauss–Seidel method, new bounds for v are

$$v \in \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

that is, we obtain the solution sharply.

However, for some systems, multiplying by Y can result in significant overestimation due to interval dependency. We give a simple example of such a system in

⁴However, using appropriate preconditioning, interval with practical use can be devised for rectangular systems (both overdetermined and underdetermined). See, for example [9, Ch. 3] for an example of such preconditioners.

⁵See, for example, [18] for the definitions of Lipschitz matrix and slope matrix.

[15]. There, we show that such problems can sometimes be handled by preconditioning the system *symbolically*⁶.

Nonetheless, there are significant practical problems that cannot be handled even with interval Newton narrowing with symbolic preconditioning in combination with basic constraint propagation. We have found this to be particularly true in practice for systems of equations arising from applying Lagrange multipliers (or the Fritz–John conditions) to constrained optimization problems; we gave a practical example of this in [14, 13], and have studied it more thoroughly in [12]. Here, we present a simplified example of a pure constraint problem.

Example 3. Take the constraint system

$$c_1(x) = x_1^2 - x_2, \quad g_1(x) = x_2 - x_1, \quad x_1 \in [0, 1], \quad x_2 \in [0, 1].$$

One can easily verify that basic constraint propagation is unsuccessful on this problem. Although this particular system has a unique solution at $x_1 = 0, x_2 = 0$, since this is an inequality-constrained problem, it is possible to have a solution set with non-empty measure. To obtain lower and upper bounds on this solution set, we may solve the corresponding constrained optimization problems with objective functions $\min x_1, \max x_1, \min x_2$ and $\max x_2$ and with the additional constraints $-x_1 \leq 0, x_1 \leq .5, -x_2 \leq 0, x_2 \leq .5$. Considering the problem with $\min x_1$, the Fritz–John equations (as in [9, (5.7), p. 197]) can be written as

$$\begin{pmatrix} u_0 - u_1 - u_2 + u_3 + 2x_1v_1 \\ u_1 - u_4 + u_5 - v_1 \\ u_1(x_2 - x_1) \\ -u_2x_1 \\ u_3x_1 \\ -u_4x_2 \\ u_5x_2 \\ x_1^2 - x_2 \\ u_0 + u_1 + u_2 + u_3 + u_4 + u_5 + v_1^2 - 1 \end{pmatrix} = 0. \quad (3)$$

The Jacobi matrix for this system is

$$\frac{\partial F}{\partial x \partial u \partial v} = \begin{pmatrix} 2v_1 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 2x_1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 \\ -u_1 & u_1 & 0 & x_2 - x_1 & 0 & 0 & 0 & 0 & 0 \\ -u_2 & 0 & 0 & 0 & -x_1 & 0 & 0 & 0 & 0 \\ u_3 & 0 & 0 & 0 & 0 & x_1 & 0 & 0 & 0 \\ 0 & -u_4 & 0 & 0 & 0 & 0 & -x_2 & 0 & 0 \\ 0 & u_5 & 0 & 0 & 0 & 0 & 0 & x_2 & 0 \\ 2x_1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 2v_1 \end{pmatrix}.$$

Interestingly, since linear combinations of entries in any column are single-use expressions, there is no overestimation in either naive evaluation of the interval Jacobi

⁶To do symbolic arithmetic on functions, we compute within a vector space of the coefficients for a basis for the space of functions in question. In [15], we used Taylor polynomials with remainder terms, and we used the COSY system [3] for the actual implementation.

matrix or in preconditioning the resulting interval matrix. If we do not know bounds on the Lagrange multipliers u_i , $0 \leq i \leq 5$ and v_1 beforehand, then we may choose them to be the natural values $u_i \in [0, 1]$, $0 \leq i \leq 5$, $v_1 \in [-1, 1]$. The resulting interval Jacobi matrix is

$$\begin{pmatrix} [-2, 2] & 0 & 1 & -1 & -1 & 1 & 0 & 0 & [0, 2] \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 \\ [-1, 0] & [0, 1] & 0 & [-1, 1] & & 0 & 0 & 0 & 0 \\ [-1, 0] & 0 & 0 & 0 & [-1, 0] & 0 & 0 & 0 & 0 \\ [0, 1] & 0 & 0 & 0 & 0 & [0, 1] & 0 & 0 & 0 \\ 0 & [-1, 0] & 0 & 0 & 0 & 0 & [-1, 0] & 0 & 0 \\ 0 & [0, 1] & 0 & 0 & 0 & 0 & 0 & [0, 1] & 0 \\ [0, 2] & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & [-1, 1] \end{pmatrix},$$

which contains many singular matrices. Thus, an interval Newton method using this system will not, in general succeed⁷. In fact, at the solution $x_1 = 0$, $x_2 = 0$, the Fritz–John system (3) becomes

$$\begin{pmatrix} u_0 - u_1 - u_2 + u_3 \\ u_1 - u_4 + u_5 - v_1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ u_0 + u_1 + u_2 + u_3 + u_4 + u_5 v_1^2 - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

which is underdetermined; thus, the Jacobi matrix at the solution itself must be singular, and an interval Newton method cannot hope to narrow all coordinates, unless we fix one or more of the multipliers at their lower or upper bounds⁸.

3. Linear Relaxations

Problems as illustrated above in Example 3 have led us to consider *linear relaxations*, a technique by which a global optimization problem is approximated by a linear program (LP). The solution to this linear program, obtained approximately by a state-of-the-art LP solver, then gives a lower bound to the solution to the original global optimization problem. This lower bound can then be used to advantage in the constraint propagation scheme. For example, optimization problems could be formulated as we described above below Example 3, and sharper lower and upper bounds could then be obtained.

Various researchers use linear relaxations to great advantage in solving practical problems. One example of this is in the commercial global optimization software BARON, whose underlying ideas are explained in [21]. However, such software has

⁷However, conceivably, some of the coordinates could possibly be narrowed with the linear programming preconditioners of [9, Chapter 3].

⁸Actually, this is a possibility that can be investigated, although there may be too many possible values of the u 's and v 's for this to be practical, if there are many constraints.

not included validation until recently. Two years ago, Neumaier and Shcherbina [19], as well as Jansson [7] exhibited a simple technique, based on the duality gap, by which a rigorous lower bound to the minimum of an LP can be obtained from approximate values of the dual variables. This enables linear relaxations to be used in a validated context.

We explain our implementation of linear relaxations in [13] and [6], while we conduct various numerical experiments within our GlobSol environment (see [11] for a summary of GlobSol) in [12].

For Example 3 above, we may follow loosely our procedure in [13] to produce linear relaxations. A corresponding linear program for computing a lower bound on x_1 , using a single underestimator for the convex function $x_2 = x_1^2$, can be as follows:

$$\begin{aligned} & \text{minimize } x_1 \\ & \text{subject to } x_2 \leq x_1 \text{ (the overestimator),} \\ & \quad x_2 \geq .125 + .5(x_1 - .25), \\ & \quad x_2 \leq x_1 \text{ (the original constraint),} \\ & \quad x_1 \in [0, 1], x_2 \in [0, 1]. \end{aligned} \tag{4}$$

(Here, we have relaxed the constraints directly, for simplicity. In [13] and [6], we illustrate how the expressions can be parsed and relaxed automatically.) The exact minimum to this linear program is $x_1 = 0$, and the solution point is $x_1 = 0, x_2 = 0$, the solution to the original constraint problem. If we replace the objective in (4) by $-x_1$, we obtain an upper bound of 1 for x_1 , not an improvement. If we replace the objective by x_2 , we obtain 0 for the lower bound on x_2 , but we similarly obtain 0.5 for the upper bound on x_2 , an improvement. We thus obtain

$$x \in \begin{pmatrix} [0, 1] \\ [0, .5] \end{pmatrix} \subset \begin{pmatrix} [0, 1] \\ [0, 1] \end{pmatrix},$$

a significant narrowing in the second coordinate. Basic constraint propagation now converges.

This illustrates that linear underestimators are capable of making headway on problems, even pure constraint problems, for which both basic constraint propagation and interval Newton methods alone do not do well. This is especially true if there is much implicit dependency in the constraints, and the constraints are approximately linear. However, this is not always the case. For example, certain problems can be treated with basic constraint propagation or with interval Newton methods, but linear relaxations are inefficient. See [12] for further information and results on various test problems.

Summary and conclusions

Through simple examples, we have illustrated deficiencies in two techniques widely used in the past in validated constraint propagation, and we have advocated linear relaxations, a technique widely used in the non-validated global optimization community but not examined seriously until recently for validated constraint solving. Linear relaxations are not a panacea, but complement well other techniques used in the validated constraint solving community.

References

- [1] UniCalc, a novel approach to solving systems of algebraic equations / A.B. Babichev, O.B. Kadyrova, T. P. Kashevarova, A.S. Leshchenko, A.L. Semenov // *Interval Computations*. – 1993. – Vol. 2. – P. 29–47.
- [2] Benhamou F. Interval constraints, interval propagation // *Encyclopedia of Optimization* / C. Floudas and P. Pardalos, ed. – Dordrecht: Kluwer, 2001.
- [3] Berz M. Cosy infinity web page, 2000. – <http://cosy.pa.msu.edu/>.
- [4] Cleary J.G. Logical arithmetic // *Future Computing Systems*. – 1987. – Vol. 2, No. 2. – P. 125–149.
- [5] Floudas C.A. *Deterministic Global Optimization: Theory, Algorithms and Applications*. – Dordrecht: Kluwer, 2000.
- [6] Hongthong S. and Kearfott R.B. Rigorous linear overestimators and underestimators, 2004. – Preprint. – http://interval.louisiana.edu/preprints/estimates_of_powers.pdf.
- [7] Jansson Ch. A rigorous lower bound for the optimal value of convex optimization problems // *J. Global Optim.* – 2004. – Vol. 28, No. 1. – P. 121–137.
- [8] Kearfott R.B. Decomposition of arithmetic expressions to improve the behavior of interval iteration for nonlinear systems // *Computing*. – 1991. – Vol. 47, No. 2. – P. 169–191.
- [9] Kearfott R.B. *Rigorous Global Search: Continuous Problems*. – Dordrecht: Kluwer, 1996.
- [10] Kearfott R.B. Interval analysis: Interval Newton methods // *Encyclopedia of Optimization*. – Dordrecht: Kluwer, 2001. – Vol. 3. – P. 76–78.
- [11] Kearfott R.B. GlobSol: History, composition, and advice on use // *Global Optimization and Constraint Satisfaction: Lect. Notes in Comput. Sci.* – 2003. – P. 17–31.
- [12] Kearfott R.B. Empirical comparisons of linear relaxations and alternate techniques in validated deterministic global optimization, 2004. – Preprint. – http://interval.louisiana.edu/preprints/validated_global_optimization_search_comparisons.pdf.
- [13] Kearfott R.B., Hongthong S. A preprocessing heuristic for determining the difficulty of and selecting a solution strategy for nonconvex optimization, 2003. – Preprint. – http://interval.louisiana.edu/preprints/2003_symbolic_analysis_of_GO.pdf.
- [14] Kearfott R.B., Neher M., Oishi S., and Rico F. Libraries, tools, and interactive systems for verified computations: Four case studies // *Numerical Software with Result Verification* / R. Alt, A. Frommer, R.B. Kearfott, W. Luther, eds.: *Lect. Notes in Comput. Sci.* – 2004. – Vol. 2991. – P. 36–63.
- [15] Kearfott R.B. and Walster G.W. Symbolic preconditioning with Taylor models: Some examples // *Reliable Computing*. – 2002. – Vol. 8, No. 6. – P. 453–468.
- [16] Kreinovich V., Lakeyev A., Rohn J., and Kahl P. *Computational Complexity and Feasibility of Data Processing and Interval Computations*. – Dordrecht: Kluwer, 1998.
- [17] Narin'yani A.S. Intelligent software technology for the new decade // *Comm. ACM*. – 1991. – Vol. 34, No. 6. – P. 60–67.
- [18] Neumaier A. *Interval Methods for Systems of Equations*. – Cambridge: Cambridge University Press, 1990.
- [19] Neumaier V., Shcherbina O. Safe bounds in linear and mixed-integer programming // *Math. Prog.* – 2004. – Vol. 99, No. 2. – P. 283–296. – <http://www.mat.univie.ac.at/~neum/ms/mip.pdf>.

-
- [20] Rump S.M. et al. INTLAB home page, 2000. – <http://www.ti3.tu-harburg.de/~rump/intlab/index.html>.
 - [21] Tawarmalani M. and Sahinidis N.V. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, and Applications*. – Dordrecht: Kluwer, 2002.
 - [22] Van Hentenryck P., Michel L., Deville Y. *Numerica: A Modeling Language for Global Optimization*. – Cambridge: MIT Press, 1997.

Новые алгоритмы интервального вычисления определенных интегралов

М.Б. Базаров*

Аннотация. В работе предлагаются два интервальных алгоритма вычисления определенных интегралов. Первый предлагаемый алгоритм построен на основе обобщенной формулы трапеций. Вторым алгоритмом является интервальным аналогом квадратурной формулы Гаусса–Кристоффеля. Далее, представлены методики для повышения точности интервального вычисления по предложенным алгоритмам.

В данной работе предлагаются два алгоритма нахождения интервального значения интеграла вида

$$\int_a^b f(x)\rho(x) dx, \quad \rho(x) > 0, \quad (1)$$

основанные на полиномиальной аппроксимации подынтегральной функции $f(x)$. Здесь функции $f(x)$ непрерывны и имеют непрерывные производные до s -го порядка – $f^{(s)}(x) \in C^{(s)}[a, b]$ на отрезке $[a, b]$, а весовая функция непрерывна на интервале (a, b) .

Кратко напомним способ получения некоторых квадратурных формул [1, 2], так как ряд параметров, определяемых этими формулами, используются при построении соответствующих интервальных формул.

При численном интегрировании обычно заменяют $f(x)$ на такую аппроксимирующую функцию $\varphi(x, a) \approx f(x)$, чтобы интеграл от нее легко вычислялся в элементарных функциях.

Чаще всего $f(x)$ заменяют некоторым обобщенным интерполяционным многочленом

$$f(x) = \sum_{i=0}^N f(x_i)\varphi_i(x) + r(x), \quad (2)$$

где $r(x)$ – остаточный член аппроксимации. Подставляя (2) в (1), получим формулу численного интегрирования (квадратурную формулу):

$$F = \sum_{i=0}^N c_i f(x_i) + R, \quad c_i = \int_a^b \varphi_i(x)\rho(x) dx, \quad R = \int_a^b r(x)\rho(x) dx, \quad (3)$$

где величины x_i называют узлами, c_i – весами, а R – погрешностью или остаточным членом формулы. Интеграл заменяется суммой, похожей на интегральную сумму, причем узлы и коэффициенты этой суммы не зависят от функции $f(x)$.

*Навоийский государственный горный институт, Навои, Узбекистан.

Как показано в [2], квадратурные формулы (3) неустойчивы относительно ошибок округления. Эти ошибки носят случайный характер, но в среднем растут как \sqrt{N} при увеличении числа узлов. Использование аппарата интервального анализа при нахождении значений определенных интегралов устраняет неустойчивость при вычислениях.

Теперь перейдем к краткому описанию предлагаемых интервальных алгоритмов.

Первый предлагаемый алгоритм построен на основе обобщенной формулы трапеций, а второй является интервальным аналогом квадратурной формулы Гаусса–Кристоффеля.

Будем считать, что основные обозначения, понятия и факты интервального анализа известны [3, 4].

Пусть функция $f(x)$, $x \in [a, b]$, имеет интервальное расширение $\mathbf{F}(\mathbf{X})$, заданное при $\mathbf{X} \subset [a, b]$, монотонное по включению и удовлетворяет условию типа Липшица, т.е. $\text{wid}(\mathbf{F}(\mathbf{X})) \leq l \cdot \text{wid}(\mathbf{X})$, $\mathbf{X} \subset [a, b]$, где $l > 0$ – некоторая постоянная. Пусть, кроме того, функция $f^{(s)}(x)$ имеет интервальное расширение $\mathbf{F}^{(s)}(\mathbf{X})$, определенное при $\mathbf{X} \subset [a, b]$ и монотонное по включению.

Как известно, для функций малой гладкости, имеющих лишь первую и вторую производные, квадратурные формулы высокого порядка точности практического значения не имеют.

Рассмотрим обобщенную формулу трапеций

$$\int_a^b f(x) dx = \frac{h}{2}(f_0 + f_N) + h \sum_{i=1}^{N-1} f_i - \frac{h^2}{12} f''(\xi), \quad \xi \in (a, b).$$

В работе показано, что если $\mathbf{F}(\mathbf{X})$ и $\mathbf{F}^{(s)}(\mathbf{X})$ – интервальные расширения соответствующих функций $f(x)$ и $f^{(s)}$, то для приближенного значения интеграла (1) справедливо включение

$$\int_a^b f(x) dx \in h \sum_1^{N-1} F(X_i) + \frac{h}{2}(F(X_0) + F(X_N)) - \frac{h^2}{12} F''([a, b]), \quad \xi \in (a, b). \quad (4)$$

Для повышения точности получаемых интервалов в работе использована методика, которую можно назвать интервальным аналогом правила Рунге.

Параметрами формулы интегрирования (3) являются узлы и веса. Однако, строя формулы (4), мы заранее задавали узлы и по ним находили веса.

Теперь рассмотрим интервальный аналог формулы Гаусса–Кристоффеля. Формула Гаусса–Кристоффеля, которая называется формулой наивысшей алгебраической точности, имеет следующий вид:

$$\int_a^b f(x) dx = \sum_1^N c_{i,N} f(x_{i,N}) + \frac{(b-a)^{2N+1} (N!)^4}{(2N+1)((2N)!)^2} f^{(2N)}(\xi), \quad \xi \in (a, b).$$

В работе доказано включение

$$\int_a^b f(x) dx \in \sum_1^N C_{i,N} \mathbf{F}(X_{i,N}) + \frac{(b-a)^{2N+1} (N!)^4}{(2N+1)((2N)!)^2} \mathbf{F}^{(2N)}([a, b]). \quad (5)$$

Далее, для повышения точности получаемых интервалов по формуле (5) использована методика, которая является улучшенным вариантом локального адаптированного алгоритма, приведенного в [5].

Кроме того, в работе представлены результаты численных экспериментов, выполненных с предложенными алгоритмами. Алгоритмы были реализованы на языке PASCAL XSC.

Список литературы

- [1] Бахвалов Н.С. Численные методы. Ч. 1. – М.: Наука, 1975. – 631 с.
- [2] Калиткин Н.Н. Численные методы. – М.: Наука, 1978. – 512 с.
- [3] Калмыков С.А., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. – Новосибирск: Наука, 1986. – 223 с.
- [4] Kearfott R.B., Nakao M.T., Neumaier A., Rump S.M., Shary S.P., van Hentenruck P. Standardized notation in interval analysis // *Reliable Computing*. – Submitted.
- [5] Kramer, W., Wander, S. Two adaptive Gauss–Legendre type algorithms for the verified computation of definite integrals // *Reliable Computing*. – 1996. – Vol. 2, № 3. – P. 241–253.

Интервальный метод Хаусхолдера для комплексных линейных систем

Б.С. Джаныбеков, С.П. Шарый*

Аннотация. Рассматривается метод Хаусхолдера для внешнего оценивания множеств решений систем линейных алгебраических уравнений с комплексными интервальными параметрами. Для этого используется нормализованное приведение матрицы системы к верхней “треугольной” форме с помощью последовательности ортогональных преобразований.

Введение

Исследуется интервальная система линейных алгебраических уравнений (ИСЛАУ)

$$Ax = b,$$

где элементами $n \times n$ -матрицы A и n -вектора правой части b являются комплексные интервалы.

Множество решений ИСЛАУ в общем случае не имеет простого описания. Поэтому мы ограничимся нахождением интервального вектора, целиком содержащего это множество решений. Как известно, полученный вектор называется *внешней оценкой* множества решений ИСЛАУ.

Одним из способов нахождения внешней оценки множества решений ИСЛАУ является *метод Хаусхолдера* [1], его вещественный вариант известен также в отечественной литературе как *метод отражений* [2, 3]. Суть этого метода заключается в том, что матрица системы A приводится к верхней треугольной форме с помощью последовательности ортогональных преобразований. В отличие от метода Гаусса решения ИСЛАУ [4], метод Хаусхолдера дает более узкие внешние оценки множеств решений, если интервалы в матрице и правой части “не слишком широки”. Кроме того, интервальный метод Хаусхолдера позволяет получать ответ в тех задачах, к которым интервальный метод Гаусса неприменим.

Отметим, что метод Хаусхолдера ранее применялся для внешнего оценивания множеств решений ИСЛАУ с вещественными интервальными параметрами. Целью работы является обобщение метода Хаусхолдера на случай комплексных ИСЛАУ.

Чтобы сделать изложение нашей работы самодостаточным, мы подробно описываем множество комплексных интервалов и основные свойства арифметик комплексных интервалов.

*Институт вычислительных технологий СО РАН.

Обозначения:

- \mathbb{R} – множество вещественных чисел,
 \mathbb{C} – множество комплексных чисел,
 \mathbb{IR} – множество замкнутых интервалов $[a, b]$ на \mathbb{R} , $a \leq b$,
 \mathbb{IC} – множество комплексных интервалов,
 \mathbb{IC}^n – множество n -мерных векторов с элементами из \mathbb{IC} ,
 $\mathbb{IC}^{n \times n}$ – множество $n \times n$ -матриц с элементами из \mathbb{IC} .

1. Комплексные интервальные арифметики

Как отмечено выше, элементами матрицы и правой части рассматриваемой нами интервальной системы уравнений являются комплексные интервалы. Поэтому было бы целесообразным подробно рассмотреть комплексную интервальную арифметику.

Покажем, что многие свойства вещественной интервальной арифметики переносятся и на комплексный случай. Существуют два разумных подхода использования комплексных интервалов, к рассмотрению которых мы сейчас и перейдем.

1.1. Прямоугольники в качестве комплексных интервалов

Определение 1. Множество

$$\mathbf{a} = \{a_1 + ia_2 \mid a_1, a_2 \in \mathbb{IR}\} \quad (i = \sqrt{-1})$$

называется прямоугольным комплексным интервалом.

Множество всех таких комплексных интервалов обозначим через $\mathbb{IC}_{\text{rect}}$. Определенный таким образом прямоугольный комплексный интервал может быть изображен на комплексной плоскости в виде прямоугольника со сторонами, параллельными вещественной и мнимой осям. Если $a = a_1 + ia_2 \in \mathbb{C}$, то его можно рассматривать как вырожденный комплексный интервал

$$\mathbf{a} = [a_1, a_1] + i[a_2, a_2] \in \mathbb{IC}_{\text{rect}}, \quad \text{где } a_1, a_2 \in \mathbb{R},$$

а каждый элемент $\mathbf{a}_1 \in \mathbb{IR}$ – как сумму

$$\mathbf{a} = \mathbf{a}_1 + i[0, 0] \in \mathbb{IC}_{\text{rect}},$$

откуда видно, что $\mathbb{IR} \subset \mathbb{IC}_{\text{rect}}$.

Элементы $\mathbf{a} = \mathbf{a}_1 + ia_2$ и $\mathbf{b} = \mathbf{b}_1 + ib_2$ из $\mathbb{IC}_{\text{rect}}$ считаются равными, если

$$\mathbf{a}_1 = \mathbf{b}_1 \quad \text{и} \quad a_2 = b_2.$$

Определенное здесь отношение равенства рефлексивно, симметрично и транзитивно.

Теперь обобщим вещественную интервальную арифметику на комплексный случай.

Определение 2. Пусть $*$ $\in \{+, -, \cdot, : \}$ – арифметическая операция над элементами из \mathbb{IR} . Тогда если

$$\mathbf{a} = \mathbf{a}_1 + i\mathbf{a}_2 \in \mathbb{IC}_{\text{rect}}, \quad \mathbf{b} = \mathbf{b}_1 + i\mathbf{b}_2 \in \mathbb{IC}_{\text{rect}},$$

то мы полагаем

$$\mathbf{a} \pm \mathbf{b} = \mathbf{a}_1 \pm \mathbf{b}_1 + i(\mathbf{a}_2 \pm \mathbf{b}_2),$$

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}_1\mathbf{b}_1 - \mathbf{a}_2\mathbf{b}_2 + i(\mathbf{a}_1\mathbf{b}_2 + \mathbf{a}_2\mathbf{b}_1),$$

$$\mathbf{a} : \mathbf{b} = (\mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2) : (\text{sqr}(\mathbf{b}_1) + \text{sqr}(\mathbf{b}_2)) + i(\mathbf{a}_2\mathbf{b}_1 - \mathbf{a}_1\mathbf{b}_2) : (\text{sqr}(\mathbf{b}_1) + \text{sqr}(\mathbf{b}_2)),$$

где

$$\text{sqr}(\mathbf{b}_1) = \begin{cases} [\underline{\mathbf{b}}_1^2, \bar{\mathbf{b}}_1^2], & \text{если } \mathbf{b}_1 \geq 0, \\ [\bar{\mathbf{b}}_1^2, \underline{\mathbf{b}}_1^2], & \text{если } \mathbf{b}_1 \leq 0, \\ [0, \max\{\underline{\mathbf{b}}_1^2, \bar{\mathbf{b}}_1^2\}], & \text{если } 0 \in \mathbf{b}_1, \end{cases}$$

$\text{sqr}(\mathbf{b}_2)$ определяется аналогично.

Считается, что в случае деления $0 \notin \text{sqr}(\mathbf{b}_1) + \text{sqr}(\mathbf{b}_2)$. Это требование выполняется при условии $0 \notin \mathbf{b}_1 + i\mathbf{b}_2$, иначе деление не определено.

Конечно, вместо $\text{sqr}(\mathbf{b}_1) + \text{sqr}(\mathbf{b}_2)$ можно было бы взять $\mathbf{b}_1^2 + \mathbf{b}_2^2$ как $\mathbf{b}_1^2 = \mathbf{b}_1\mathbf{b}_1$ и $\mathbf{b}_2^2 = \mathbf{b}_2\mathbf{b}_2$, но тогда требование $0 \notin \mathbf{b}_1^2 + \mathbf{b}_2^2$ может оказаться невыполненным, даже если $0 \notin \mathbf{b}_1 + i\mathbf{b}_2$.

1.2. Круги в качестве комплексных интервалов

Определение 3. Пусть $a \in \mathbb{C}$ – комплексное число и $r \geq 0$. Множество

$$\mathbf{z} = \{z \in \mathbb{C} \mid |z - a| \leq r\}$$

называется круговым комплексным интервалом или кругом.

Множество всех кругов обозначим через $\mathbb{IC}_{\text{disc}}$. Круг \mathbf{z} с центром a и радиусом r будем записывать в виде $\mathbf{z} = \langle a, r \rangle$.

Комплексные числа можно рассматривать как специальные элементы из $\mathbb{IC}_{\text{disc}}$, имеющие вид $\langle a, 0 \rangle$. Ясно, что $\mathbb{C} \subset \mathbb{IC}_{\text{disc}}$.

Два круга $\mathbf{a} = \langle a, r_1 \rangle$ и $\mathbf{b} = \langle b, r_2 \rangle$ считаются равными, если $a = b$ и $r_1 = r_2$.

Это отношение равенства также рефлексивно, симметрично и транзитивно. Операции на $\mathbb{IC}_{\text{disc}}$ вводятся как обобщения операций над вещественными числами следующим образом.

Определение 4. Пусть $*$ $\in \{+, -, \cdot, : \}$ – арифметическая операция над комплексными числами. Тогда если $\mathbf{a} = \langle a, r_1 \rangle$ и $\mathbf{b} = \langle b, r_2 \rangle$, то

$$\mathbf{a} \pm \mathbf{b} = \langle a \pm b, r_1 + r_2 \rangle,$$

$$\mathbf{a} \cdot \mathbf{b} = \langle ab, |a|r_2 + |b|r_1 + r_1r_2 \rangle,$$

$$\frac{1}{\mathbf{b}} = \left\langle \frac{\bar{b}}{b\bar{b} - r_2^2}, \frac{r_2}{b\bar{b} - r_2^2} \right\rangle \quad \text{при условии, что } 0 \notin \mathbf{b},$$

$$\mathbf{a} : \mathbf{b} = \mathbf{a} \cdot \frac{1}{\mathbf{b}} \quad \text{при условии, что } 0 \notin \mathbf{b}.$$

Здесь $|a| = \sqrt{a_1^2 + a_2^2}$ обозначает модуль комплексного числа $a = a_1 + ia_2$, а $\bar{b} = b_1 - ib_2$ – сопряженное к $b = b_1 + ib_2$.

Соберем теперь вместе наиболее важные свойства операций на $\mathbb{I}\mathbb{C}_{\text{rect}}$ и $\mathbb{I}\mathbb{C}_{\text{disc}}$. Если не оговорено противное, то $\mathbb{I}\mathbb{C}$ можно понимать и как обозначение множества $\mathbb{I}\mathbb{C}_{\text{rect}}$ с операциями из определения 2, и как обозначение множества $\mathbb{I}\mathbb{C}_{\text{disc}}$ с операциями из определения 4.

Теорема 1. Пусть $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{I}\mathbb{C}$. Тогда

- I. $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$, $\mathbf{ab} = \mathbf{ba}$ (коммутативность).
- II. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$, $(\mathbf{ab})\mathbf{c} = \mathbf{a}(\mathbf{bc})$ для $\mathbf{a}, \mathbf{b}, \mathbf{c}$ из $\mathbb{I}\mathbb{C}_{\text{disc}}$ (ассоциативность).
- III. $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$ для $\mathbf{a}, \mathbf{b}, \mathbf{c}$ из $\mathbb{I}\mathbb{C}_{\text{rect}}$.
- IV. $[0, 0] + i[0, 0] \in \mathbb{I}\mathbb{C}_{\text{rect}}$ (соответственно $\langle 0, 0 \rangle \in \mathbb{I}\mathbb{C}_{\text{disc}}$) и $[1, 1] + i[0, 0] \in \mathbb{I}\mathbb{C}_{\text{rect}}$ (соответственно $\langle 1, 0 \rangle \in \mathbb{I}\mathbb{C}_{\text{disc}}$) – определенные единственным образом нейтральные элементы сложения (нуль) и умножения (единица).
- V. $\mathbb{I}\mathbb{C}$ не имеет делителей нуля.
- VI. Элемент \mathbf{z} множества $\mathbb{I}\mathbb{C}$ имеет противоположный и обратный элементы, только если $\mathbf{z} \in \mathbb{C}$ и в случае обратного элемента $\mathbf{z} \neq 0$. Однако $0 \in \mathbf{a} - \mathbf{a}$ и $1 \in \mathbf{a} : \mathbf{a}$.
- VII. $\mathbf{a}(\mathbf{b} + \mathbf{c}) \subseteq \mathbf{ab} + \mathbf{ac}$ (субдистрибутивность), $\mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{ab} + \mathbf{ac}$ для $\mathbf{a} \in \mathbb{C}$.

Особо отметим, что ассоциативный закон в общем случае не выполняется при перемножении элементов $\mathbb{I}\mathbb{C}_{\text{rect}}$.

2. Метод Хаусхолдера

Дана система интервальных уравнений

$$\mathbf{Ax} = \mathbf{b} \tag{1}$$

с невырожденной матрицей $\mathbf{A} \in \mathbb{I}\mathbb{C}^{n \times n}$ и вектором $\mathbf{b} \in \mathbb{I}\mathbb{C}^n$.

Напомним, что интервальная матрица \mathbf{A} называется невырожденной, если невырождены все матрицы A из \mathbf{A} .

Множество решений интервальной системы уравнений (1) определяется в виде:

$$\Xi(\mathbf{A}, \mathbf{b}) = \{x \in \mathbb{C}^n \mid (\exists A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}.$$

Как было уже отмечено, множество решений ИСЛАУ в общем случае не имеет простого описания, поэтому мы ограничимся нахождением интервального вектора $\mathbf{x} \in \mathbb{I}\mathbb{C}^n$, целиком содержащего это множество решений.

Пусть \mathbf{v} интервальный вектор из $\mathbb{I}\mathbb{C}^n$. Интервальное расширение нормы вектора \mathbf{v} , обозначаемое как $\text{INorm}(\mathbf{v})$, берется следующим образом:

$$\text{INorm}(\mathbf{v}) = \left(\sum_{k=1}^n \mathbf{v}_k \bar{\mathbf{v}}_k \right)^{1/2},$$

где $\bar{\mathbf{v}}_k$ – сопряженное с \mathbf{v}_k .

Если $0 \notin \mathbf{v}$, то обозначим через \mathbf{w} интервальный вектор, полученный нормированием интервального вектора \mathbf{v} :

$$\mathbf{w}_i = \frac{v_i}{\text{INorm}(\mathbf{v})}, \quad \text{если } 0 \in \mathbf{v}_i,$$

$$\mathbf{w}_i = \frac{\text{sgn}(\text{mid}(\text{Re}(\mathbf{v}_i)))}{\sqrt{1 + \left(\sum_{j \neq i}^n (\mathbf{v}_j \bar{\mathbf{v}}_j) \right) / (\mathbf{v}_i \bar{\mathbf{v}}_i)}}, \quad \text{если } 0 \notin \mathbf{v}_i,$$

где $\text{sgn}(\text{mid}(\text{Re}(\mathbf{v}_i)))$ – знак середины действительной части комплексного интервала \mathbf{v}_i для всех $i = 1, 2, \dots, n$.

Определим интервальную $n \times n$ -матрицу \mathbf{H} как

$$\mathbf{H} = \mathbf{I} - 2 \cdot (\mathbf{w} \cdot \mathbf{w}^*), \quad (2)$$

\mathbf{H} является интервальным расширением эрмитовой унитарной матрицы Хаусхолдера, порожденный интервальным вектором \mathbf{v} . Кроме того, \mathbf{H} содержит эрмитовы и унитарные матрицы отражений H , порожденные комплексными векторами $v \in \mathbf{v}$. В дальнейшем, интервальное расширение нормы вектора и интервальное расширение матрицы Хаусхолдера мы будем называть просто нормой интервального вектора и интервальной матрицей Хаусхолдера соответственно.

Пусть $\mathbf{A} = \mathbf{A}^{(1)} = (\mathbf{a}_1^{(1)} \mathbf{a}_2^{(1)} \dots \mathbf{a}_n^{(1)})$, где $\mathbf{a}_1^{(1)}, \mathbf{a}_2^{(1)}, \dots, \mathbf{a}_n^{(1)}$ – вектор-столбцы интервальной матрицы. Мы собираемся вычислить интервальную матрицу \mathbf{Q} и верхнюю треугольную интервальную $n \times n$ матрицу \mathbf{R} с помощью последовательности трансформаций Хаусхолдера. Здесь следует уточнить, что мы получим не строго треугольную интервальную матрицу \mathbf{R} , так как из-за некоторых особенностей классической интервальной арифметики, мы не можем получить 0 при выполнении арифметических операций с интервалами ненулевой ширины, но обязательным условием является включение нуля.

Обозначим через $e_1^{(p)}$ действительный p -вектор, первый компонент которого равен 1, а остальные – 0. Предположим, что $0 \notin \mathbf{a}_1^{(1)}$ и пусть $\alpha = \text{sgn}(\text{mid}(\text{Re}(\mathbf{a}_{11}^{(1)}))) \cdot \text{INorm}(\mathbf{a}_1^{(1)})$, причем условимся, что, если вещественный интервал $\text{Re}(\mathbf{a}_{11}^{(i)})$ является уравновешанным (т.е. $\text{Re}(\mathbf{a}_{11}^{(i)}) = -\text{Re}(\bar{\mathbf{a}}_{11}^{(i)})$), то $\text{sgn}(\text{mid}(\text{Re}(\mathbf{a}_{11}^{(i)}))) = +1$ для $i = 1, 2, \dots, n$.

Тогда \mathbf{H} – интервальная $n \times n$ матрица Хаусхолдера, порождаемая вектором

$$\mathbf{v} = \mathbf{a}_1^{(1)} + \alpha \cdot e_1^{(n)}.$$

Мы имеем

$$\mathbf{A}^{(1)} \subseteq \mathbf{H} \cdot \left((-\alpha \cdot e_1^{(n)}) \mathbf{w}_2^{(1)} \dots \mathbf{w}_n^{(1)} \right),$$

где $\mathbf{w}_i^{(1)} = \mathbf{H} \cdot \mathbf{a}_i^{(1)}$ для $i = 2, \dots, n$.

Первая строка матрицы \mathbf{R} :

$$\mathbf{R}_{11} = -\alpha, \quad \mathbf{R}_{1j} = \mathbf{w}_{1j}^{(1)} \quad \text{для } j = 2, \dots, n.$$

Положим $\mathbf{H}^{(1)} = \mathbf{H}$. Мы получим интервальную $(n-1) \times (n-1)$ матрицу $\mathbf{A}^{(2)}$:

$$\mathbf{A}^{(2)} = \begin{pmatrix} \mathbf{w}_{22}^{(1)} & \dots & \mathbf{w}_{2n}^{(1)} \\ \vdots & & \vdots \\ \mathbf{w}_{n2}^{(1)} & \dots & \mathbf{w}_{nn}^{(1)} \end{pmatrix}.$$

Обозначим элементы $\mathbf{A}^{(2)}$ как $\mathbf{a}_{ij}^{(2)}$, $1 \leq i \leq n-1$ и $1 \leq j \leq n-1$.

Предположим, что $0 \notin \mathbf{a}_1^{(2)}$ и пусть $\alpha = \text{sgn}(\text{mid}(\text{Re}(\mathbf{a}_{11}^{(2)}))) \cdot \text{INorm}(\mathbf{a}_1^{(2)})$. Тогда \mathbf{H} – интервальная $(n-1) \times (n-1)$ матрица Хаусхолдера, порождаемая вектором

$$\mathbf{v} = \mathbf{a}_1^{(2)} + \alpha \cdot \mathbf{e}_1^{(n-1)}.$$

Имеем

$$\mathbf{A}^{(2)} \subseteq \mathbf{H} \cdot \left((-\alpha \cdot \mathbf{e}_1^{(n-1)}) \mathbf{w}_2^{(2)} \dots \mathbf{w}_n^{(2)} \right),$$

где $\mathbf{w}_i^{(2)} = \mathbf{H} \cdot \mathbf{a}_i^{(2)}$ для $i = 2, \dots, n-1$.

Вторая строка матрицы \mathbf{R} :

$$\mathbf{R}_{22} = -\alpha, \quad \mathbf{R}_{2j} = \mathbf{w}_{1j}^{(2)} \quad \text{для } j = 3, \dots, n.$$

Положим

$$\mathbf{H}^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H} \end{pmatrix}.$$

Далее, повторяем процедуру построения матриц \mathbf{R} и \mathbf{Q} , в данном случае

$$\mathbf{Q} = \mathbf{H}^{(1)} \cdot \left(\mathbf{H}^{(2)} (\dots (\mathbf{H}^{(n-1)})) \right).$$

Интервальный метод Хаусхолдера позволяет вычислить внешнюю оценку \mathbf{x} множества решений $\Xi(\mathbf{A}, \mathbf{b})$ решением верхней треугольной интервальной $n \times n$ системы

$$\mathbf{R} \cdot \mathbf{x} = \mathbf{c}, \quad \text{где } \mathbf{c} = \mathbf{Q}^* \cdot \mathbf{b}.$$

Данный вывод следует из того, что множество решений $\Xi(\mathbf{A}, \mathbf{b}) \subset \Xi(\mathbf{R}, \mathbf{c})$ в следствии монотонности включения матриц

$$\mathbf{H}^{(1)} \subseteq \mathbf{H}^{(2)} \subseteq \dots \subseteq \mathbf{H}^{(n-1)}.$$

Далее, воспользуемся формулами обратной подстановки для нахождения компонентов интервального вектора \mathbf{x} :

$$\begin{aligned} \mathbf{x}_n &= \mathbf{c}_n / \mathbf{R}_{nn}, \\ \mathbf{x}_i &= \left(\mathbf{c}_i - \sum_{j=i+1}^n (\mathbf{R}_{ij} \mathbf{x}_j) \right) / \mathbf{R}_{ii}, \quad 1 \leq i \leq n-1. \end{aligned}$$

В конечном итоге мы получим внешнюю оценку \mathbf{x} , целиком содержащую множество решений $\Xi(\mathbf{A}, \mathbf{b})$ интервальной системы уравнений.

3. Численные эксперименты

Рассмотрим пример применения метода Хаусхолдера для нахождения внешней оценки множества решений ИСЛАУ (1) с матрицей \mathbf{A} и правой частью \mathbf{b} [4]:

$$\mathbf{A} = \begin{pmatrix} [1, 5] + i[-1, 1] & 1 \\ 25 & [-1, 1] + i[-1, 1] \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} [-1, 1] \\ [-1, 1] \end{pmatrix}.$$

Одним из условий существования решения является невырожденность интервальной матрицы \mathbf{A} :

$$\det(\mathbf{A}) = [-31, -19] + i[-6, 6].$$

В данном случае, $0 \notin \text{Re}(\det(\mathbf{A}))$, следовательно, матрица \mathbf{A} – невырождена.

Далее, определяем интервальные векторы

$$\mathbf{a}_1^1 = \begin{pmatrix} [1, 5] + i[-1, 1] \\ 25 \end{pmatrix}, \quad \mathbf{a}_2^1 = \begin{pmatrix} 1 \\ [-1, 1] + i[-1, 1] \end{pmatrix}.$$

Следовательно,

$$\alpha = \text{sgn}(\text{mid}([1, 5])) \cdot \text{INorm}(\mathbf{a}_1^{(1)}) = [25.0199, 25.5147].$$

Вычислим вектор \mathbf{v} , порождающий интервальную матрицу Хаусхолдера:

$$\mathbf{v} = \begin{pmatrix} [1, 5] + i[-1, 1] \\ 25 \end{pmatrix} + \alpha \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} [26.0199, 30.5147] + i[-1, 1] \\ 25 \end{pmatrix}.$$

Определим вектор \mathbf{w} нормированием вектора \mathbf{v} :

$$\mathbf{w} = \begin{pmatrix} [0.6583, 0.8453] + i[-0.0277, 0.0277] \\ [0.6325, 0.6925] \end{pmatrix}.$$

Далее, вычисляем интервальную матрицу Хаусхолдера по формуле (2):

$$\begin{aligned} \mathbf{H} &= I - 2 \cdot \begin{pmatrix} [0.4326, 0.7153] & [0.4164, 0.5854] + i[-0.0192, 0.0192] \\ [0.4164, 0.5854] - i[-0.0192, 0.0192] & [0.4, 0.4796] \end{pmatrix} \\ &= \begin{pmatrix} [-0.4306, 0.1348] & [-1.1708, -0.8328] - i[-0.0384, 0.0384] \\ [-1.1708, -0.8328] + i[-0.0384, 0.0384] & [0.0408, 0.2] \end{pmatrix}. \end{aligned}$$

Построим первую строку интервальной матрицы \mathbf{R} :

$$\mathbf{R}_{11} = -\alpha = [-25.5147, -25.0199].$$

$$\mathbf{R}_{12} = [-1.6398, 1.344] + i[-1.2092, 1.2092].$$

В итоге получим матрицу

$$\mathbf{A}^{(2)} = [-1.3708, -0.6328] + i[-0.2384, 0.2384],$$

следовательно,

$$\alpha = \text{sgn}(\text{mid}([-1.3708, -0.6328])) \cdot \text{INorm}(\mathbf{a}_1^{(2)}) = -[0.6328, 1.3914].$$

Вычислим последний диагональный элемент матрицы \mathbf{R} :

$$\mathbf{R}_{22} = -\alpha = [0.6328, 1.3914].$$

На этом построение матриц \mathbf{R} и \mathbf{Q} закончено, причем мы получили матрицу $\mathbf{Q} = \mathbf{H}$.

Теперь нам необходимо определить интервальный вектор \mathbf{c} :

$$\mathbf{c} = \mathbf{Q}^* \cdot \mathbf{b} = \begin{pmatrix} [-1.6014, 1.6014] + i[-0.0384, 0.0384] \\ [-1.3708, 1.3708] + i[-0.0384, 0.0384] \end{pmatrix}.$$

Далее, воспользуемся формулами обратной подстановки Гаусса для нахождения компонентов интервального вектора \mathbf{x} :

$$\begin{aligned} \mathbf{x}_2 &= \frac{\mathbf{c}_2}{\mathbf{R}_{22}} = [-2.1663, 2.1663] + i[-0.0607, 0.0607], \\ \mathbf{x}_1 &= \frac{\mathbf{c}_1 - \mathbf{R}_{12} \cdot \mathbf{x}_2}{\mathbf{R}_{11}} = [-0.2086, 0.2086] + i[-0.11, 0.11]. \end{aligned}$$

В итоге мы получим внешнюю оценку

$$\mathbf{x} = \begin{pmatrix} [-0.2086, 0.2086] + i[-0.11, 0.11] \\ [-2.1663, 2.1663] + i[-0.0607, 0.0607] \end{pmatrix}$$

целиком содержащую множество решений $\Xi(\mathbf{A}, \mathbf{b})$ интервальной системы уравнений.

Заключение

В заключении еще раз отметим, что метод Хаусхолдера ранее применялся для внешнего оценивания множеств решений ИСЛАУ с вещественными интервальными параметрами. Целью работы является обобщение метода Хаусхолдера на случай комплексных ИСЛАУ. Кроме того, интервальный метод Хаусхолдера позволяет получать ответ в тех задачах, к которым интервальный метод Гаусса неприменим, например, выше рассмотренная численная задача.

Список литературы

- [1] Bentbib A.H. Solving the full rank interval least squares problem // Applied Numerical Mathematics. – 2002. – № 41. – P. 283–294.
- [2] Калиткин Н.Н. Численные методы. – М.: Наука, 1975.
- [3] Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Наука, 2002.
- [4] Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. – М.: Мир, 1987.
- [5] Neumaier A. Interval methods for systems of equations. — Cambridge: Cambridge University Press, 1990.

Специальные приближения множеств решений

Б.С. Добронец*

Аннотация. В статье рассматриваются специальные приближения множеств решений задач с интервальными входными данными. Приближения основаны на объединении геометрических тел, определяемых вектором параметров. Приведены примеры приближения множества решений для системы линейных алгебраических уравнений и задачи Коши для системы ОДУ.

1. Введение

Интервальная математика традиционно использует для приближения множеств решений интервальные вектора или n -мерные параллелепипеды, со сторонами параллельными координатным осям.

Интервальные вектора часто дают лишь весьма приближенное представление о виде множества решения. Невозможность представить достаточно точно множество решений снижает эффективность применения интервальных методов и приводит к таким эффектам как “эффект упаковывания” (wrapping effect). Это приводит к неоправданно сильному расширению трубки решений задач Коши для систем ОДУ. В настоящее время существует постоянная потребность в повышении точности приближений множеств решений.

Стремление избежать подобных эффектов приводит к использованию алгоритмов, которые для приближения множеств решений применяют различные геометрические тела: шары, эллипсоиды, параллелепипеды, многогранники и т. п.

В работе [2] рассмотрены приближения областями, зависящими от параметров, в частности – параллелепипедами. В [3] специальные приближения представлены как объединения геометрических объектов – шаров и параллелепипедов. В работе [4] рассмотрены приближения множеств решений систем ОДУ эллипсоидами.

В данной работе продолжают исследования по построению специальных приближений для множеств решений задач с интервальными входными данными. Множества решений аппроксимируются объединением геометрических тел вдоль некоторой гладкой кривой. Такие приближения, с одной стороны, дают достаточную простоту представления и с другой – обладают достаточной гибкостью.

Рассмотрим примеры геометрических тел Ω из R^n . Шары – наиболее простые из геометрических тел, они характеризуются своим центром $x_0 \in R^n$ и радиусом R , эллипсоиды задаются так же своим центром и симметричной, положительно определенной матрицей A , параллелепипеды однозначно определяются одной из вершин x_0 и ребрами x_i . Таким образом, для задания шара

*КГТУ, Красноярск.

необходим $n + 1$ параметр, для эллипсоида – $n + n(n + 1)/2$ параметр и соответственно для параллелепипеда – $n(n + 1)$ параметр. Далее x_0 будем называть *точкой привязки* геометрического тела.

Таким образом, будем считать, что положение геометрического тела определяется точкой привязки и некоторым вектором параметров $p \in R^m$.

Рассмотрим один из возможных вариантов приближения множеств решений из R^n в виде объединения $\Omega(z, r)$ из R^{n-1}

$$\mathcal{X} \subseteq \bigcup_{\eta} \Omega(z(\eta), r(\eta)).$$

Здесь $z \in R^n$ – гладкая параметрическая кривая, значения $z(\eta)$ определяют точки привязки. Для однозначного задания будем считать, что $\Omega(z(\eta), r(\eta))$ лежит в гиперплоскости $\Gamma: (x - z(\eta), \nu(\eta))$, в частности можно положить $\nu(\eta) = z'(\eta)$.

Функции z, r , в зависимости от представления, характеризуются своими наборами параметров. Таким образом, определим множество $\Omega(p)$:

$$\Omega(p) = \bigcup_{\eta} \Omega(z(\eta), r(\eta)),$$

где $p \in R^m$ – вектор параметров.

2. Построение приближений

В каждом конкретном случае выбор геометрического тела $\Omega(z, r)$ и кривой z определяется геометрическим видом \mathcal{X} , свойствами аппроксимации и простотой использования.

Таким образом, на первом шаге зададим параметрическую кривую $z(\eta) \in R^n$, $\eta \in [\eta_1, \eta_2]$. Зафиксируем η и построим плоскость $\Gamma: (x - z(\eta), \nu)$, где $\nu = x'_0(\eta)$ – вектор касательный к кривой z . Далее на плоскости Γ зададим систему координат с началом в точке $z(\eta)$ и базисными векторами $e_i(\eta)$. Пусть $\mathcal{X}_\eta = \mathcal{X} \cap \Gamma$ и определим тело $\Omega(z, r)$ с параметрами r , такое что $\Omega(z, r) \supseteq \mathcal{X}_\eta$.

Безусловно, это самая важная часть процедуры построения приближений и ее конкретная реализация зависит от исходной задачи. В частности, при приближении множеств решений систем линейных алгебраических уравнений, можно пользоваться известными свойствами этих множеств.

Рассмотрим интервальную систему линейных алгебраических уравнений

$$Ax = b, \tag{1}$$

где $A \in R^{n \times n}$, а $b \in R^n$. Предположим так же, что интервальная матрица системы (1) регулярна.

Решением задачи (1) будем называть множество

$$\mathcal{X} = \{x \mid Ax = b, A \in \mathbf{A}, b \in \mathbf{b}\}.$$

Множество \mathcal{X} может быть описано следующим образом [1]:

$$\mathcal{X} = \{x \mid x \in R^n, \mathbf{A}x \cap \mathbf{b} \neq \emptyset\} \quad (2)$$

или

$$\{x \mid x \in R^n, 0 \in \mathbf{A}x - \mathbf{b}\}.$$

Для простоты изложения, предположим, что \mathcal{X} полностью содержится в одном квадранте и $\mathcal{X} \subset R^3$. В качестве геометрических тел выберем круги. Для определения радиуса круга введем на Γ полярную систему координат (r, φ) . Тогда $z_1 = r \cos \varphi$, $z_2 = r \sin \varphi$ и $x = z_1 e_1 + z_2 e_2$. Фиксировав φ , найдем верхнюю границу $\bar{r}(\varphi)$, при котором выполнено условие (2). Далее определим $R(\eta) = \max_{[0, 2\pi]} \bar{r}(\varphi)$ и $K(\eta) : |z| \leq R(\eta)$.

Таким образом, пара $(R(\eta), z(\eta))$ будет определять параметрическое множество $\Omega(z, R) = \bigcup_{\eta} K(\eta) \supseteq \mathcal{X}$.

Несколько иначе выглядит построение множества $\Omega(z, r)$ для задачи Коши для систем ОДУ.

Рассмотрим следующую систему:

$$\begin{aligned} x'_i &= f_i(t, x, k), & i &= 1, \dots, n, & t &\in (0, l), \\ x(0) &= x_0, \end{aligned} \quad (3)$$

где $x_0 \in R^n$ – вектор начальных значений, $x_0 \in \mathbf{x}_0$; $k \in R^m$ – вектор параметров, $k \in \mathbf{k}$; $x \in R^n$ – вектор неизвестных.

Далее будем считать, что x есть функция t, k, x_0 :

$$x = x(t, k, x_0). \quad (4)$$

Обозначим через $\mathcal{X}(t)$ множество решений системы ОДУ

$$\mathcal{X}(t) = \{x(t, k, x_0) \mid x_0 \in \mathbf{x}_0, k \in \mathbf{k}\}.$$

Рассмотрим задачу оценки в фазовом пространстве множества решений $\mathcal{X}(t)$ системы обыкновенных дифференциальных уравнений множеством $\Omega(p(t))$.

Будем считать, что начальное состояние системы (3) можно представить $x(0) \subseteq \Omega(p(0))$. Предположим, что в некоторый момент времени t нам известно множество $\Omega(p(t))$, содержащее множество решений $\mathcal{X}(t)$ исходной системы ОДУ. Рассмотрим задачу построения множества $\Omega(p(t + \tau))$:

$$\Omega(p(t + \tau)) \supseteq \mathcal{X}(t + \tau).$$

Построим приближенно $\mathcal{X}^\tau(t + \tau)$, используя методы численного интегрирования, например, метод Эйлера:

$$\mathcal{X}^\tau(t + \tau, \mathbf{k}, x_0) \supseteq \{x^\tau(t + \tau) \mid x^\tau(t + \tau) = x(t) + \tau f(t, x(t), k), x(t) \in \Omega(p(t)), k \in \mathbf{k}\},$$

Ясно, что в данном случае граница $\mathcal{X}^\tau(t + \tau)$ отличается от истинной на величину, не превышающую $O(\tau^2)$. Таким образом, зная $\Omega(p(t))$ в некоторый момент времени t , можно построить $\Omega(p(t + \tau))$.

Существует известный произвол при построении таких множеств, обычно их строят таким образом, чтобы они имели наименьший объем.

Если $\forall t, \tau > 0$ известны векторы $p(t)$ и $p(t + \tau)$, описывающие поведение $\Omega(p(t))$, то можно предельным переходом построить систему ОДУ, описывающую поведение p :

$$p'_i(t) = \lim_{\tau \rightarrow 0} (p_i(t + \tau) - p_i(t)) / \tau = g(t, p).$$

Таким образом,

$$p' = g(t, p), \quad p(0) = p_0. \quad (5)$$

3. Примеры

Пример 1. Пусть необходимо решить систему интервальных линейных алгебраических уравнений $A\mathbf{x} = \mathbf{b}$:

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} [0, 2] \\ 0 \\ [0, 2] \end{pmatrix}. \quad (6)$$

Как нетрудно убедиться, минимальный интервальный вектор, содержащий множество ее решений – $\mathbf{x} = ([0, 2], [0, 2], [0, 2])^T$.

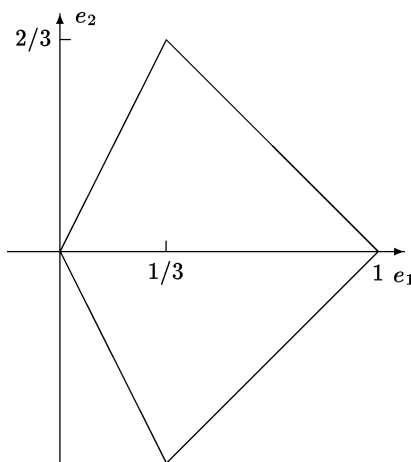
Множество решений данной системы можно описать следующим образом. Определим параметрическую кривую z :

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} t, \quad t \in [0, 1].$$

Круги в нашем случае вырождаются в отрезки, точка привязки – середина отрезка. Радиусы определяются следующим образом:

$$R(t) = \begin{cases} 2t, & t \in [0, 1/3], \\ -t + 1, & t \in [1/3, 1]. \end{cases}$$

Множество решений системы (6) лежит в плоскости векторов $e_1 = (2, 2, 2)$, $e_2 = (-1, 0, 1)$ и может быть представлено в следующем виде:



Пример 2. Рассмотрим в качестве еще одного примера линейную систему ОДУ с интервальным параметром \mathbf{k} :

$$\begin{aligned}x'_1 &= kx_2, & x'_2 &= -kx_1, & k &\in \mathbf{k} = [1, 2], \\x(0) &= x_0 \in [-\varepsilon, +\varepsilon] \times [1 - \varepsilon, 1 + \varepsilon].\end{aligned}$$

Для построения множества $\Omega(z, r)$ перейдем в полярную систему координат. В качестве параметрической кривой z выберем $r = \text{const}$, $\varphi \in [\underline{\varphi}, \overline{\varphi}]$, геометрические тела – отрезки перпендикулярные z , точки привязки – их середины. Представим

$$\Omega(\varphi(0), r(0)) = [\underline{\varphi}_0, \overline{\varphi}_0] \times [\underline{r}_0, \overline{r}_0].$$

Система ОДУ для $R, r, \underline{\varphi}, \overline{\varphi}$ выглядит следующим образом:

$$R' = 0, \quad r' = 0, \quad \underline{\varphi}' = \underline{k}, \quad \overline{\varphi}' = \overline{k},$$

с начальными данными

$$\begin{aligned}R(0) &= (\overline{r}_0 - \underline{r}_0)/2, & r(0) &= (\underline{r}_0 + \overline{r}_0)/2, \\ \underline{\varphi}(0) &= \underline{\varphi}_0, & \overline{\varphi}(0) &= \overline{\varphi}_0.\end{aligned}$$

Таким образом, хотя полученное представление множества решений отличается от оптимального, эффект упаковки полностью отсутствует.

Список литературы

- [1] Beeck H. Über die struktur und abschätzungen der lösungsmenge von linearen gleichungssystemen mit intervallkoeffizienten // Computing. – 1972. – Vol. 10. – P. 231–244.
- [2] Dobronets B.S. On some two-sided methods for solving systems of ordinary differential equations // Interval Computations. – 1992. – Vol. 1, № 3. – P. 6–19.
- [3] Добронетц Б.С., Рощина С.Л. Специальные приближения множеств решений систем ОДУ с интервальными параметрами // Вопросы математического анализа. – Красноярск: Изд. КГТУ, 2002. – Вып. 5. – С. 12–17.
- [4] Черноусько Ф.Л. Оценка фазового состояния динамических систем. Метод эллипсоидов. – М.: Наука, 1988. – 320 с.

Метод глобальной оптимизации на основе метода ветвей и границ

Ю.Г. Долгов

1. Введение

Решение задач глобальной оптимизации является, несомненно, актуальной проблемой. Подобные задачи возникают в самых различных областях человеческой деятельности (экономике, геологии, химии, биологии, физике и т. д.) Для гарантированного поиска глобального оптимума обычно пользуются интервальными методами, однако, подобные методы имеют, как правило низкую скорость сходимости.

Алгоритм, предложенный в данной статье позволяет избежать некоторых недостатков известных методов глобальной оптимизации и существенно уменьшить время решения.

Текст данной работы построен следующим образом. В п. 2 приведены определения, использованные в рамках данной работы и описываются методы, использованные в данной работе. В п. 3 предлагается описание предложенного метода и некоторые пояснения. В п. 4 представлены результаты работы алгоритма в сравнении с пакетом Numerica.

2. Основные определения, понятия и используемые методы

Приведем формальные определения ограничения и других терминов, используемых в дальнейшем.

Определение 2.1 (ограничение). *Ограничением* на множестве переменных $\{x_1, \dots, x_m\}$ будем называть подмножество S декартова произведения: $D_{x_1} \times \dots \times D_{x_m}$, где D_{x_i} – область определения переменной x_i .

Определение 2.2 (CSP – Constraint Satisfaction Problem). *CSP или задача удовлетворения ограничениям* – это тройка $P = (X, D, C)$, где $X = \{x_1, \dots, x_m\}$ множество переменных, $D = \{D_{x_1}, \dots, D_{x_n}\}$ – множество областей определения переменных (D_{x_i} – множество всевозможных значений переменной x_i), $C = \{c_1, \dots, c_k\}$ – это множество ограничений.

Определение 2.3 (решение CSP). *Решение задачи удовлетворения ограничениям* $P = (X, D, C)$ – это вектор $a = (a_1, \dots, a_n) \in D = \{D_{x_1}, \dots, D_{x_n}\}$, для которого выполняются все ограничения $c_j \in C$.

Определение 2.4 (аппроксимация). Пусть $S \subseteq \mathbb{R}$, аппроксимацией S будем называть наименьший интервальный вектор I , такой, что $S \subseteq I$ и обозначать $\text{арргох}(S)$.

Определение 2.5 (2В-совместность ограничения). Пусть c – это некоторое ограничение, определенное на множестве переменных $\{x_1, \dots, x_m\}$. Будем говорить, что c 2В-совместно, если $\forall i \in \{1, \dots, m\} D_{x_i} = \text{арргох}\{\nu_i \in D_{x_i} \mid \exists \nu_1 \in D_{x_1}, \dots, \nu_{i-1} \in D_{x_{j-1}}, \nu_{i+1} \in D_{x_{j+1}}, \dots, \nu_m \in D_{x_m}, \text{ такое что } c(\nu_1, \dots, \nu_{i-1}, \nu_i, \nu_{i+1}, \dots, \nu_m)\}$ – совместно.

Определение 2.6 (2В-совместность CSP). Задача удовлетворения ограничениям $P = (X, D, C)$ называется 2В-совместной, если все ее ограничения $c_j \in C$ 2В-совместны.

Определение 2.7 (оператор сужения). Пусть $P = (X, D, C)$ – некоторая задача удовлетворения ограничениям. С каждым ограничением $c_j \in C$ и каждой переменной $x_j \in X$ сопоставим оператор сужения $\text{red}_{c_j}^{x_j}$, который действует на множестве областей определений переменных, используемых в c_i , и результатом его применения является множество $S_j \subseteq D_j$, с удаленными из него несовместными значениями переменной x_j .

Запишем более формально: пусть ограничение c_i определено на множестве переменных $\{x_1, \dots, x_m\}$, тогда $\text{red}_{c_i}^{x_j} : P(D_{x_1}) \times \dots \times P(D_{x_m}) \rightarrow P(D_{x_j})$.

Далее приводится краткое описание методов, используемых для создания метода ветвей и сужений.

2.1. Метод распространения ограничений

Метод распространения ограничений [8] является одним из методов установления 2В-совместности, который аналогичен алгоритму АС-3. Этот алгоритм также называется алгоритмом сужения. В вычислительном плане он может быть представлен как итерационный процесс, последовательно сужающий область, гарантированно содержащую все решения. На каждом шаге итерационного процесса применяется один из операторов сужения, стягивающий область значений переменных, связанных с некоторым ограничением. Как правило, оператор сужения уточняет область определения только одной переменной. Управление итерационным процессом осуществляется по данным, то есть для исполнения на очередном шаге выбирается только тот оператор сужения, аргументы которого были изменены в процессе вычислений на предыдущем шаге.

Операторы сужения могут представляться разными способами. Например, расщеплением исходной системы уравнений на примитивные (унарные и бинарные) соотношения посредством введения дополнительных переменных. В этом случае каждое примитивное соотношение позволяет выразить каждую из входящих в него переменных через остальные. Весь набор полученных таким образом примитивных уравнений и будет задавать множество операторов сужения для исходной системы. Достоинством этого подхода является возможность решения задач с недифференцируемыми функциями.

2.2. Метод ветвей и границ

Метод ветвей и границ [6] – это общий метода поиска решения. Метод начинает работу с определения нижней и верхней границ для исходной задачи. Если верхняя и нижняя границы совпадают, то полученный результат является оптимальным значением, и метод прекращает работу. Иначе, множество переменных разбивается на несколько собственных подмножеств, объединение которых совпадает с исходным множеством. Эти подзадачи становятся потомками исходной. Далее алгоритм применяется рекурсивно к каждой из подзадач, создавая дерево подзадач. Если оптимальное решение найдено для некоторой подзадачи, то оно является достижимым для исходной задачи (не обязательно оптимальным), но так как оно достижимо, его можно использовать для обрезания ветвей у исходного дерева. Процесс поиска продолжается до тех пор, пока каждая из подзадач не будет решена или выкинута или до тех пор, пока не будет достигнут заданный порог между лучшим из найденных решений и нижней границей $f(x)$ для всех нерешенных задач.

2.3. Метод Хансена

Метод Хансена был описан в работе [3] в 1992г. Суть работы метода заключается в последовательном удалении из начальной области, подобластей, в которых не содержится глобальный минимум. Удаление происходит одним из ниже перечисленных способов:

- Удаляются подобласти, в которых градиент g функции f отличен от нуля.
- Удаляются подобласти, в которых $f > \bar{f}$, где \bar{f} – верхняя оценка глобального минимума.
- Удаляются подобласти, в которых функция f невыпукла.

Критерии остановки

Критерий А: $\text{wid}(\mathbf{x}) < \varepsilon_x$ для некоторого $\varepsilon_x > 0$.

Критерий В: $\bar{f} - \underline{f} < \varepsilon_f$ для некоторого $\varepsilon_f > 0$.

2.4. Метод моделирования обжига

Метод моделирования обжига [5, 1, 4, 7] является представителем семейства методов Монте-Карло. Он основан на физическом процессе замораживания жидкостей или рекристаллизации металлов в процессе прокаливания. В качестве оценки решения может быть выбрана точка не только не уменьшающая значения целевой функции, но и увеличивающая его. Такое поведение поиска оптимального решения помогает избежать попадания в точки локального минимума.

3. Метод ветвей и сужений

В рамках данной работы был реализован метод, объединяющий преимущества различных подходов при решении задач глобальной оптимизации.

Описание метода

Данный метод, по сути, является усовершенствованием интервального метода ветвей и границ, на каждой итерации которого применяется метод Хансена для сужения области, содержащей глобальный оптимум. Для улучшения нижней оценки функции используется численный метод симуляции обжига, объединенный с методом мултистарта, который позволяет достаточно быстро найти хорошее приближение глобального оптимума. Для применения метода Хансена используется метод распространения ограничений.

Выбор этих методов не случаен, их комбинация позволяет компенсировать слабые стороны друг друга. Приведем сводную таблицу качественных характеристик используемых методов:

Характеристики	Метод		
	ветвей и границ	Хансена	локальной оптимизации
Гарантированный поиск глобального оптимума	+	-	-
Гарантированность вычислений	+	+	-
Вычисления над интервальными типами данных	+	+	-
Небольшие временные затраты	-	+	+
Небольшое количество используемых ресурсов	-	+	+
Сужение области поиска	+	+	-

Использование комбинации методов позволяет гарантированно находить глобальный оптимум задачи, при этом использование метода локальной оптимизации для улучшения текущего рекорда значительно ускоряет работу метода ветвей и границ. Метод Хансена также позволяет откидывать области, в которых заведомо не содержится решения.

На основе результатов тестов были сделаны некоторые корректировки в используемых методах:

- в методе ветвей и границ производится деление на три части, кроме того, при выборе подзадачи выбирается та, где среднее значение функции наименьшее;
- в методе Хансена используется модернизированный метод распространения ограничений, предложенный в работе [2];
- запуск локального метода происходит только после выполнения некоторого количества итераций (это число может устанавливаться пользователем) метода ветвей и границ, что позволяет уменьшить время решения относительно простых задач. Количество стартовых точек в методе мултистарта вычисляется по формуле: $n = n_0 \cdot \dim$, где \dim – размерность задачи, а n_0 – заданное пользователем число.

GlobOpt ($X, F(x), k_{\max}, \varepsilon, n_{\text{loc}}$)

Псевдокод:

1. BEGIN
2. $k = 1$, $\text{record} = [\infty, \infty]$, $\text{SolList} = \text{null}$;
3. $T = X_0 = X$; – дерево подобластей

```

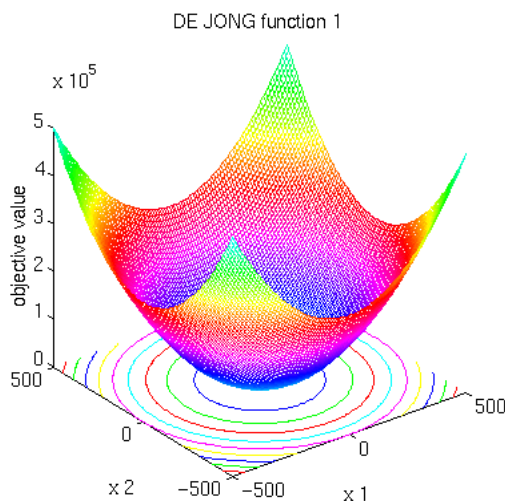
4.   WHILE  $T$  не пусто AND  $k < k_{\max}$  DO
5.     IF  $k \bmod n_{\text{loc}} = 0$  DO
6.       opt_record = LocOpt( $\varepsilon, \text{dim}$ ); ( $\text{dim}$  – размерность задачи)
7.       IF opt_record < record DO
8.         Очистить SollList;
9.       END DO
10.    END DO
11.    Взять элемент  $t_k = X_k$  из  $T$ ;
12.     $t_k = \text{HansenMeth}(t_k, \varepsilon)$ ;
13.    IF  $F(X_k) \leq \text{record}$  DO
14.      IF  $\text{wid}(F(X_k))_l < \varepsilon$  DO
15.        IF  $F(X_k) < \text{record}$  DO
16.          Очистить SollList;
17.        END DO
18.        record =  $[\min(\text{record}_l, F(X_k)_l), \min(\text{record}_r, F(X_k)_r)]$ 
19.        Добавить( $X_k, F(X_k)$ ) в SollList;
20.      ELSE DO
21.        Поделить  $X_k$  на 3 части и добавить их в  $T$ ;
22.      END DO
23.    END DO
24.    Удалить  $X_k$  из  $T$ ,  $k = k + 1$ ;
25.  END DO
26.  RETURN SollList;
27. END

```

4. Результаты численных экспериментов

Приведем сравнение результатов решения нескольких задач глобальной оптимизации методом ветвей и сужений и пакетом Numerica.

Пример 1 (Функция де Джонга).



$$f = \sum_{i=1}^n x_i^2, \quad x_i = [-10^{30}, 10^{30}].$$

Глобальный минимум:

$$f = 0, \quad x_i = [0, 0],$$

$i \in \{1, \dots, n\}$.

SibCalc	Numerica	n
0	0.10	5
0.015	0.30	20
0.015	8.00	100
0.064	40.10	200
0.156	Out of Memory	400

Пример 2.

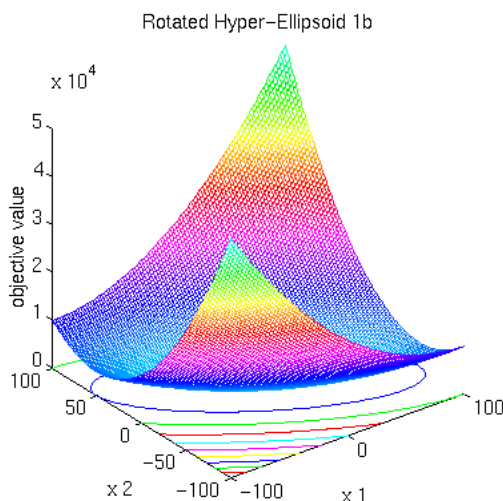
$$f = \sum_{i=1}^n \sum_{j=1}^i x_j^2, \quad x_i = [-10^{30}, 10^{30}].$$

Глобальный минимум:

$$f = 0, \quad x_i = [0, 0],$$

$i \in \{1, \dots, n\}$.

SibCalc	Numerica	n
0	0.10	5
0.015	3.70	20
0.438	Out of Memory	100
1.813	Out of Memory	200
7.953	Out of Memory	400



Пример 3 (функция Растригина).

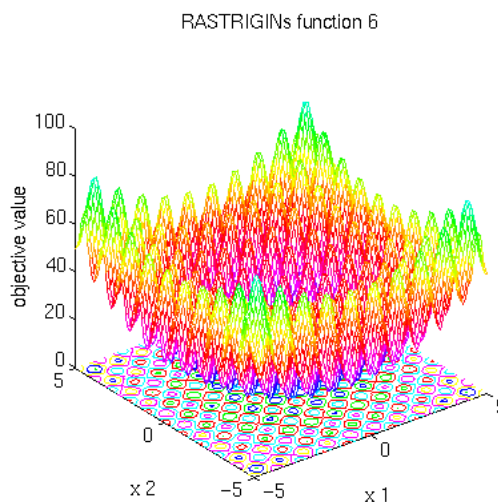
$$f = \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i)) + 10n, \quad x_i = [-10^{30}, 10^{30}].$$

Глобальный минимум:

$$f = 0, \quad x_i = [0, 0],$$

$i \in \{1, \dots, n\}$.

SibCalc	Numerica	n
0.328	0.10	5
2.578	0.30	10
11.546	0.70	15
25.141	1.20	20
58.016	1.90	25



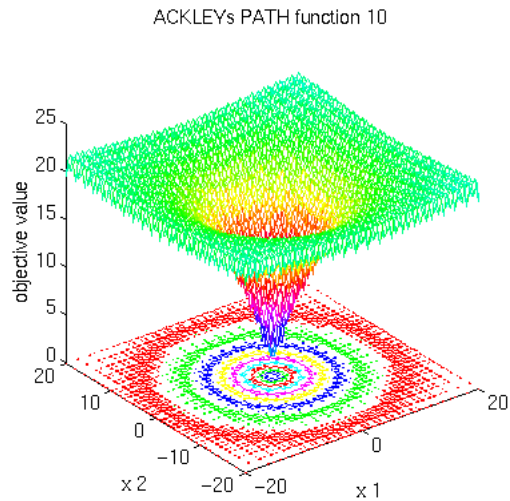
Пример 4 (функция Экли).

$$f = -20 \cdot \exp\left(-\frac{1}{5} \sqrt{\frac{1}{n}} \sum_{i=1}^n x_i^2\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + \exp(1),$$

$$x_i = [-33, 33].$$

Глобальный минимум:

$$f = 0, \quad x_i = [0, 0], \quad i \in \{1, \dots, n\}.$$



SibCalc	Numerica	n
1.578	0.40	5
15.266	1.20	10
83.546	3.60	15

Пример 5 (шестигорбый верблюд).

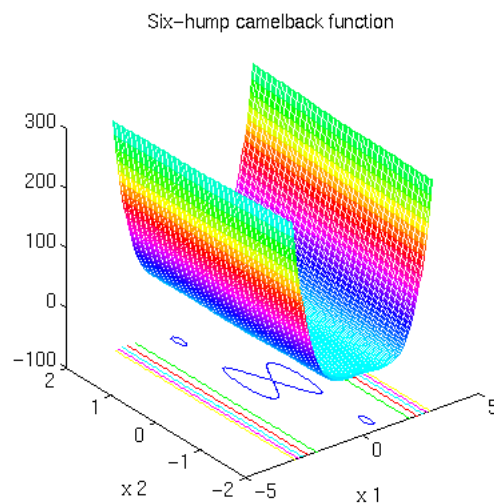
$$\left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2, \quad x_1 = x_2 = [-10^{30}, 10^{30}].$$

Глобальный минимум:

$$f = [-1.03628453, -1.03628453].$$

Первое решение: $x_1 = [-0.089842, -0.089842]$, $x_2 = [0.7126564, 0.7126564]$.

Второе решение: $x_1 = [0.089842, 0.089842]$, $x_2 = [-0.7126564, -0.7126564]$.



При решении Numerica находит только второе решение.

Время счета: SibCalc – 0.125, Numerica – 0.

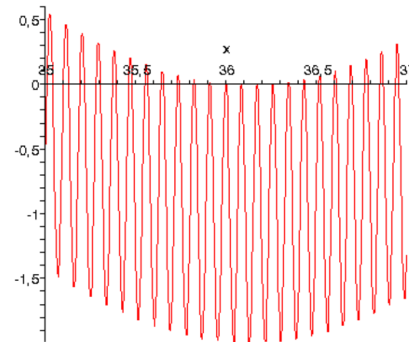
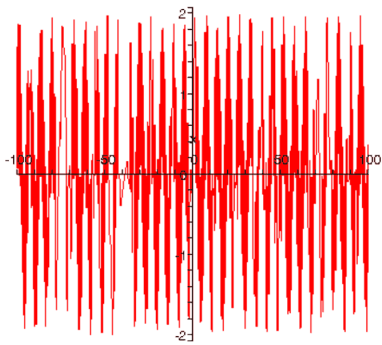
Пример 6.

$$f = \sin x + \sin x^2, \quad x = [-100, 100].$$

Глобальный минимум:

$$f = [-1.999999438712797, -1.999999438712796],$$

$$x = [36.12937493100531, 36.12937493100539].$$



Numerica находит решение:

$$f = [-1.999999438712743, -1.999999438712743],$$

$$x = [99.02767859, 99.02767859].$$

Время счета: SibCalc – 0.015, Numerica – 0.1.

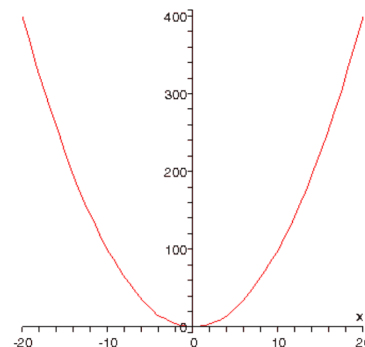
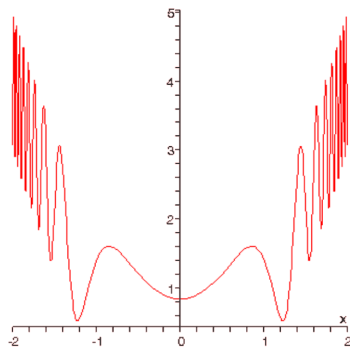
Пример 7.

$$f = x^2 + \sin(\exp(x^2)), \quad x = [-100, 100].$$

Глобальный минимум:

$$f = [0.5264768703, 0.5264768703],$$

$$x = \pm[1.22528916901954, 1.225289169019542].$$



Numerica находит решение $x = [0, 0]$.

Время счета: SibCalc – 0.015, Numerica – 0.1.

Результаты экспериментов показали, что метод ветвей и сужений позволяет гарантированно находить глобальный минимум, в отличие от пакета Numerica, который зачастую теряет корни, а в примерах 6 и 7 находит неверные значения. Кроме того, время работы алгоритма остается вполне приемлемым, он работает медленнее лишь на задачах с большим количеством локальных экстремумов (примеры 3 и 4).

5. Заключение

В рамках данной работы был проведен анализ существующих алгоритмов поиска глобального оптимума. Разработан и реализован метод глобальной оптимизации. Проведены сравнения с пакетом Numerica. По результатам сравнения можно сделать следующие выводы:

1. Созданный метод позволяет гарантированно находить решения, даже для задач с большим значением локальных минимумов.
2. Метод может реально применяться для задач с размерностями несколько десятков.
3. Скорость работы метода сравнима, а на некоторых задачах превосходит скорость работы существующих методов.

Стоит также отметить, что полученный метод может применяться для решения задач условной оптимизации. Однако, скорость работы метода при этом существенно падает из-за отсутствия возможности применять в этом случае метод Хансена.

В дальнейшем предполагается адаптировать и улучшить данный метод для решения задач условной оптимизации.

Список литературы

- [1] Cerny V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm // J. Opt. Theory Appl. – 1985. – Vol. 45. – P. 41–51.
- [2] Dolgov Yu. On strategies of the narrowing operator selection in the constraint propagation method // Lect. Notes Comput. Sci. – 2004. – Vol. 2890. – P. 431–437.
- [3] Hansen E., Global Optimization Using Interval Analysis. – New York: Dekker, 1992.
- [4] Kirkpatrick, S., Gelatt Jr.C.D., Vecchi M.P. Optimization by simulated annealing // Science. – 1983. – Vol. 220. – P. 671–680.
- [5] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., Teller E. Equation of state calculations by fast computing machines // J. Chem. Phys. – 1953. – Vol. 21. – P. 1087–1092.
- [6] Nemhauser G.L., Wolsey L.A. Integer combinatorial optimization. – New York: Wiley.
- [7] Press Wm. H., Flannery B., Teukolsky S., Vetterling Wm. // Numerical Recipes. – New York: Cambridge University Press, 1986. – P. 326–334.
- [8] Semenov A., Kashevarova T., Leshchenko A., Petunin D. Combining various techniques with the algorithm of subdefinite calculations // Proc. of the 3rd Intern. Conf. on the Practical Application of Constraint Technology ПАСТ'97. – London (England), April 1997. – P. 287–306.

Необходимые и достаточные условия невырожденности интервальных матриц

В.И. Ерохин*

Аннотация. Рассмотрены необходимые и достаточные условия невырожденности интервальных матриц с невырожденной центральной матрицей и одноранговой матрицей радиуса

Введение

Пусть

$$\mathbf{A}(A_c, \Delta_A) \triangleq [A_c - \Delta_A, A_c + \Delta_A] \quad (1)$$

интервальная матрица такая, что $A_c \in \mathbb{R}^{n \times n}$, $\text{rank } A_c = n$, $\Delta_A \in \mathbb{R}^{n \times n}$, $\text{rank } \Delta_A = 1$. Заметим, что последнее условие эквивалентно условию

$$\Delta_A = \varepsilon \cdot uv^T, \quad (2)$$

где $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, $u, v \in \mathbb{R}^n$, $u, v \geq 0$ — некоторые векторы.

В дальнейшем окажется важным выполнение условий

$$\|u\|_\infty = \|v\|_\infty = 1, \quad (3)$$

$$u, v > 0. \quad (4)$$

Таким образом, основным объектом исследования в настоящей работе является интервальные матрицы, удовлетворяющие условиям (1)–(4). Заметим, условиям (1)–(4) удовлетворяют и интервальные матрицы специального (достаточно простого) вида

$$\mathbf{A}(A_c, \varepsilon) \triangleq [A_c - \varepsilon E, A_c + \varepsilon E], \quad (5)$$

где $A_c \in \mathbb{R}^{n \times n}$ — центральная матрица, $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, $E \in \mathbb{R}^{n \times n}$ — матрица, все элементы которой — единицы. Действительно, пусть все компоненты векторов u и v — единицы. Очевидно, что при этом условия (3), (4) выполняются и, в то же время, $uv^T = E$.

Как известно (см., напр., [6]), задача проверки невырожденности интервальной матрицы общего вида является NP-трудной. В работе [8] было показано, что проверка невырожденности интервальной матрицы вида (5) для произвольного рационального $\varepsilon > 0$ и некоторой симметричной положительно определенной матрицы с неотрицательными рациональными элементами A_c также является NP-трудной. В связи с этим становится актуальной проблема получения некоторых относительно простых и практичных в вычислительном

*Борисоглебский государственный педагогический институт.

плане критериев (“verifiable conditions”), способных выполнять роль достаточных условий невырожденности интервальных матриц. Одной из последних серьезных работ, посвященных указанной теме, является, по-видимому, [7]. В то же время, автору настоящей работы удалось, опираясь на некоторые полезные свойства обобщенных матричных норм и технику матричной коррекции несовместных систем линейных алгебраических уравнений, рассмотренные в работе [1], получить некоторые новые результаты из области критериев невырожденности интервальных матриц как общего вида, так и вида (5) [2]. В частности, для интервальных матриц вида (5) были получены новые необходимые и достаточные условия невырожденности. В настоящей работе указанный результат обобщается на интервальные матрицы вида (1)–(4).

1. “Инструментарий”

Определим матричную норму $\|\cdot\|_{\varphi,\psi}$ соотношением [4]:

$$\|A\|_{\varphi,\psi} \triangleq \max_{x \neq 0} \frac{\psi(Ax)}{\varphi(x)}, \quad (6)$$

где $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $\varphi(\cdot)$ и $\psi(\cdot)$ – произвольные векторные нормы (не обязательно нормы Гельдера) соответственно в \mathbb{R}^n и \mathbb{R}^m и соотношение между m и n произвольное. Заметим, что в общем случае норма $\|\cdot\|_{\varphi,\psi}$ удовлетворяет только аксиомам векторных норм (т. е., не является мультипликативной).

Определим векторную норму $\varphi^*(\cdot)$, двойственную к заданной векторной норме $\varphi(\cdot)$ относительно скалярного произведения [4] как:

$$\varphi^*(y) \triangleq \max_{x \neq 0} \frac{|y^T x|}{\varphi(x)}, \quad (7)$$

где $x, y \in \mathbb{R}^n$ – некоторые векторы.

Вектор $y \in \mathbb{R}^n$ будем называть двойственным к ненулевому вектору $x \in \mathbb{R}^n$ относительно нормы $\varphi(\cdot)$, если выполняется условие [4]:

$$\varphi^*(y) \cdot \varphi(x) = y^T x = 1. \quad (8)$$

Следует заметить, что объекты, задаваемые формулами (6)–(8), действительно определены корректно и всегда существуют, а использование “max” вместо “sup” оправдано, поскольку соответствующие верхние грани всегда достигаются [4].

Теперь обратимся к приведенной в работе [1] лемме, обобщающей полученное А. Н. Тихоновым [3] минимальное по евклидовой норме решение системы линейных уравнений $Ax = b$ относительно неизвестной матрицы A на случай произвольной нормы $\|\cdot\|_{\varphi,\psi}$:

Лемма 1. Пусть $x \in \mathbb{R}^n$ и $b \in \mathbb{R}^m$ – произвольные векторы, $x \neq 0$. Тогда система уравнений $Ax = b$ разрешима относительно A и при этом существует решение \hat{A} из класса одноранговых матриц размера $m \times n$, минимальное по норме $\|\cdot\|_{\varphi,\psi}$, которое дается формулой

$$\hat{A} = by^T, \tag{9}$$

где $y \in \mathbb{R}^n$ – вектор, двойственный вектору x относительно нормы $\varphi(\cdot)$. При этом

$$\|\hat{A}\|_{\varphi,\psi} = \frac{\psi(b)}{\varphi(x)}. \tag{10}$$

Предположим теперь, что некоторая матрица $A_c \in \mathbb{R}^{n \times n}$ является невырожденной, что эквивалентно отсутствию нетривиальных решений соответствующей однородной системы линейных алгебраических уравнений. Зафиксируем некоторый вектор $x \in \mathbb{R}^n$, $x \neq 0$, и рассмотрим систему линейных алгебраических уравнений

$$(A_c + H)x = 0 \quad \Leftrightarrow \quad Hx = -A_c x, \tag{11}$$

где $H = (h_{ij}) \in \mathbb{R}^{n \times n}$ – некоторая матрица.

В силу леммы 1 для любого вектора $x \neq 0$ система (11) разрешима относительно матрицы H , и поэтому, в свою очередь, матрица $A_c + H$ оказывается вырожденной. При этом минимальная по $\|\cdot\|_{\varphi,\psi}$ норме матрица \hat{H} при фиксированном ненулевом векторе x имеет вид

$$\hat{H}(x) = -A_c xy^T,$$

где $y \in \mathbb{R}^n$ – вектор, двойственный вектору x относительно нормы $\varphi(\cdot)$. При этом

$$\|\hat{H}(x)\|_{\varphi,\psi} = \frac{\psi(A_c x)}{\varphi(x)} \triangleq \Phi(x).$$

Несложно показать [2], что

$$\inf_{x \neq 0} \Phi(x) = \|A_c^{-1}\|_{\psi,\varphi}^{-1}$$

и достигается для любой квадратной невырожденной матрицы A_c .

Более того, пусть $\mathfrak{A}(A_c, H)$ – множество всех вырожденных матриц вида $A_c + H$, где $A_c \in \mathbb{R}^{n \times n}$ – невырожденная матрица. Справедлива следующая

Теорема 1 [2].

$$\min_{\mathfrak{A}(A_c, H) \neq \emptyset} \|H\|_{\varphi,\psi} = \|A_c^{-1}\|_{\psi,\varphi}^{-1}. \tag{12}$$

При этом существуют такие ненулевые векторы $u, x, y \in \mathbb{R}^n$, где y – вектор, двойственный вектору x относительно нормы $\varphi(\cdot)$, что

$$\begin{aligned} x &= A_c^{-1}u, & H &= -uy^T, \\ (A_c + H) &\in \mathfrak{A}(A_c), & (A_c + H)x &\equiv 0, \end{aligned}$$

$$\|H\|_{\varphi,\psi} = \frac{\psi(u)}{\varphi(A_c^{-1}u)} = \|A_c^{-1}\|_{\psi,\varphi}^{-1}.$$

Следствие [2]. Пусть матрица A_c – невырождена и

$$\|\Delta_A\|_{\varphi,\psi} < \|A_c^{-1}\|_{\psi,\varphi}^{-1}. \tag{13}$$

Тогда интервальная матрица A является невырожденной.

Конкретный вид условия (13) зависит от вида используемых векторных норм $\varphi(\cdot)$ и $\psi(\cdot)$, которые в общем случае не обязаны быть нормами Гельдера. Но даже при использовании только гельдеровских векторных норм класс $\|\cdot\|_{\varphi,\psi}$ норм оказывается достаточно представительным. К нему, например, принадлежат такие известные и широко употребляемые на практике нормы (см., напр., [4]), как

$$\max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_1} = \|A\|_{\ell_{\infty}} = \max_{ij} |a_{ij}|, \quad (14)$$

$$\max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \|A\|_1 = \max_j \sum_i |a_{ij}|, \quad (15)$$

$$\max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2 = \lambda_{\max}^{\frac{1}{2}}(A^T A), \quad (16)$$

$$\max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_{\infty}} = \|A\|_{\infty} = \max_i \sum_j |a_{ij}|, \quad (17)$$

Очевидно, что при исследовании интервальных матриц вида (5) (с интервалами равной величины) наиболее полезной оказывается $\|\cdot\|_{1,\infty}$ норма. Ее использование позволяет получить [1] относительно простой и достаточно практичный критерий невырожденности, имеющий вид необходимого и достаточного условия.

Теорема 2 [2]. *Для того, чтобы интервальная матрица $\mathbf{A}(A_c, \varepsilon)$ с невырожденной центральной матрицей $A_c \in \mathbb{R}^{n \times n}$ была невырожденной, необходимо и достаточно, чтобы выполнялось условие*

$$\varepsilon < \|A_c^{-1}\|_{\infty,1}^{-1}. \quad (18)$$

2. Необходимые и достаточные условия невырожденности интервальной матрицы с однораговой матрицей радиуса

Приведем теперь новый, ранее не публиковавшийся результат, являющийся обобщением рассмотренной выше теоремы 2 на случай интервальной матрицы с невырожденной центральной матрицей и однораговой матрицей радиуса.

Пусть

$$\varepsilon^* \triangleq \|\text{diag}(v) \cdot A_c^{-1} \cdot \text{diag}(u)\|_{\infty,1}^{-1}. \quad (19)$$

Справедлива следующая

Теорема 3. *Для того, чтобы интервальная матрица $\mathbf{A}(A_c, \Delta_A)$ с невырожденной центральной матрицей $A_c \in \mathbb{R}^{n \times n}$ и матрицей радиуса $\Delta_A \in \mathbb{R}^{n \times n}$, удовлетворяющей условиям (2)–(4), была невырожденной, необходимо и достаточно, чтобы выполнялось условие $\varepsilon < \varepsilon^*$.*

Доказательство.

Достаточность. Пусть

$$\varphi(x) \triangleq \|\text{diag}(v) \cdot x\|_1, \quad \psi(x) \triangleq \|\text{diag}^{-1}(u) \cdot x\|_\infty. \quad (20)$$

Легко показать (см., напр., [4]), что функции $\varphi(x)$ и $\psi(x)$ являются векторными нормами. При этом функция

$$\Psi(H) \triangleq \|\text{diag}^{-1}(u) \cdot H \cdot \text{diag}^{-1}(v)\|_{1,\infty} \quad (21)$$

является $\|\cdot\|_{\varphi,\psi}$ нормой, т. е., $\Psi(H) = \|H\|_{\varphi,\psi}$. В силу теоремы 1 и формул (20), (21),

$$\min_{\mathfrak{A}(A_c, H) \neq \emptyset} \|H\|_{\varphi,\psi} = \|A_c^{-1}\|_{\psi,\varphi}^{-1} = \|\text{diag}(v) \cdot A_c^{-1} \cdot \text{diag}(u)\|_{\infty,1}^{-1} = \varepsilon^*. \quad (22)$$

В то же время, в силу формул (2), (20), (21)

$$\|\Delta_A\|_{\varphi,\psi} = \varepsilon \cdot \|\text{diag}^{-1}(u) \cdot uv^T \cdot \text{diag}^{-1}(v)\|_{1,\infty} = \varepsilon \cdot \|E\|_{1,\infty} = \varepsilon.$$

Для завершения обоснования достаточности сопоставляем формулы (18), (22) и опираемся на следствие теоремы 1.

Необходимость. Пусть интервальная матрица $\mathbf{A}(A_c, \Delta_A)$ невырождена, но при этом $\varepsilon \geq \varepsilon^*$. В соответствии с теоремой 1 можно построить матрицу H такую, что матрица $(A_c + H)$ является вырожденной, и, в то же время,

$$\|H\|_{\varphi,\psi} = \|\text{diag}^{-1}(u) \cdot H \cdot \text{diag}^{-1}(v)\|_{1,\infty} = \max_{i,j} \left| \frac{h_{ij}}{u_i \cdot v_j} \right| = \varepsilon^*. \quad (23)$$

“Сняв” символ “max” в цепочке равенств (23), ее можно продолжить как цепочку неравенств:

$$\begin{aligned} -\varepsilon^* \cdot u_i \cdot v_j \leq h_{ij} \leq \varepsilon^* \cdot u_i \cdot v_j &\Leftrightarrow -\varepsilon^* \cdot uv^T \leq H \leq \varepsilon^* \cdot uv^T \Leftrightarrow \\ A_c - \varepsilon^* \cdot uv^T \leq A_c + H \leq A_c + \varepsilon^* \cdot uv^T &\Rightarrow \\ A_c - \varepsilon \cdot uv^T \leq A_c + H \leq A_c + \varepsilon \cdot uv^T &\Leftrightarrow \\ (A_c + H) \in \mathbf{A}(A_c, \Delta_A) - \text{противоречие.} &\quad \square \end{aligned}$$

3. Обсуждение результатов

Попробуем вначале разобраться, насколько практичными (“verifiable”) могут оказаться результаты теоремы 3. Мы будем оценивать два показателя: трудоемкость вычисления параметра ε^* и чувствительность результатов этих вычислений к погрешностям в элементах векторов u, v и матрицы A_c . При этом, как очевидно, трудоемкость вычисления ε^* есть не что иное как трудоемкость проверки необходимых и достаточных условий невырожденности интервальной матрицы вида (1)–(4), а погрешность в вычислении ε^* – это погрешность указанной проверки.

Заметим, что явная формула, связывающая значение $\|\cdot\|_{\infty,1}$ нормы произвольной матрицы со значениями ее элементов, неизвестна. Если же для вычисления $\|\cdot\|_{\infty,1}$ нормы непосредственно использовать определение (6), получаем

задачу невыпуклого математического программирования, для которой не гарантирована полиномиальная трудоемкость. А так как вычисление $\|\cdot\|_{\infty,1}$ нормы – составная часть вычисления параметра ε в соответствии с формулой (19), то необходимо признать, что в общем случае для вычисления параметра ε^* не гарантирована полиномиальная трудоемкость. В то же время, можно указать класс интервальных матриц, для которых параметр ε^* может быть вычислен за полиномиальное время. Существенной особенностью интервальных матриц указанного выше класса является выполнение условия

$$\|A_c^{-1}\|_{\infty,1} = \|A_c^{-1}\|_{\ell_1}, \quad (24)$$

где, при условии $A_c^{-1} = (\alpha_{ij})$,

$$\|A_c^{-1}\|_{\ell_1} \triangleq \sum_{i,j} |\alpha_{ij}|. \quad (25)$$

Объект, задаваемый формулой (25), является мультипликативной матричной нормой [4].

Как несложно заметить, и обращение матрицы A_c , и вычисление ее $\|\cdot\|_{\ell_1}$ нормы, и проверка выполнения неравенства (18) могут быть выполнены за полиномиальное время (при стандартных предположениях о представлении входных данных и трудоемкости арифметических операций).

Таким образом, главное, что необходимо выяснить – это условия, при которых справедливо равенство (24).

В начале покажем, что справедлива следующая

Лемма 2. *Для произвольной матрицы A имеет место неравенство*

$$\|A\|_{\infty,1} \leq \|A\|_{\ell_1}. \quad (26)$$

Доказательство. В силу (6)

$$\|A\|_{\infty,1} = \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_{\infty}} = \max_{\|x\|_{\infty}=1} \|Ax\|_1. \quad (27)$$

Пусть все элементы матрицы A являются неотрицательными числами, а все компоненты вектора x – единицы. В этом случае $\|x\|_{\infty} = 1$, $\|Ax\|_1 = \sum_{i,j} |a_{ij}| = \|A\|_{\ell_1}$. Аналогичный результат получается, если вектор x составлен из чисел -1 . Легко убедиться, что при любом другом выборе вектора x , отвечающего условию $\|x\|_{\infty} = 1$, справедливо неравенство $\|Ax\|_1 < \|A\|_{\ell_1}$, из чего, в силу (27) следует, что для любой матрицы с неотрицательными элементами

$$\|A\|_{\infty,1} = \|A\|_{\ell_1}. \quad (28)$$

Теперь заметим, что равенство (28) справедливо и для любой матрицы с неположительными элементами в силу аксиомы абсолютной однородности матричной нормы.

Рассмотрим теперь общий случай. Пусть матрица A содержит положительные, отрицательные и, возможно, нулевые элементы. Тогда она может быть

представлена в виде $A = A_+ + A_-$, где A_+ – матрица, составленная из неотрицательных элементов, A_- – матрица, составленная из неположительных элементов. Используя неравенство треугольника, определение (25) и соотношения, полученные для неотрицательных (неположительных) матриц, имеем:

$$\|A\|_{\infty,1} \leq \|A_+\|_{\infty,1} + \|A_-\|_{\infty,1} = \|A_+\|_{\ell_1} + \|A_-\|_{\ell_1} = \|A\|_{\ell_1},$$

что и требовалось доказать.

Теперь обратимся к обсуждению основного результата.

Теорема 4. *Невырожденность интервальной матрицы $\mathbf{A}(A_c, \Delta_A)$ с невырожденной матрицей A_c и одноранговой матрицей Δ_A может быть проверена за полиномиальное время, если матрица $A_c^{-1} = (\alpha_{ij})$ является такой, что*

$$\alpha_{ij} = l_i \cdot |\alpha_{ij}| \cdot r_j \Leftrightarrow A_c^{-1} = \text{diag}(l) \cdot |A_c^{-1}| \cdot \text{diag}(r), \quad (29)$$

где $l, r \in \mathbb{R}^n$ – векторы, компоненты которых принимают только значения $\{-1, 1\}$, $|A_c^{-1}|$ – матрица, составленная из абсолютных величин элементов матрицы A_c .

Доказательство. Учитывая уже проделанные выше рассуждения, для обоснования утверждения теоремы достаточно показать, что в рамках предположения (29) оказывается справедливым условие (24).

Действительно, пусть $x = r$. Легко проверить, что в этом случае $\|x\|_{\infty} = 1$. В то же время, в силу (29)

$$A_c^{-1}x = \text{diag}(l) \cdot |A_c^{-1}| \cdot 1,$$

где 1 – вектор, состоящий из единиц. Но

$$\|\text{diag}(l) \cdot |A_c^{-1}| \cdot 1\|_1 = \sum_{i,j} |\alpha_{ij}| = \|A_c^{-1}\|_{\ell_1},$$

что, с учетом леммы 2, и завершает доказательство.

Обратимся ко второму вопросу, затронутому в начале данного параграфа, – вопросу о влиянии погрешностей в коэффициентах матрицы A_c и элементах векторов u, v . Следует признать, что этот вопрос еще требует исследования. Возможно, на этом пути окажется полезной рассмотренная в работе [2] теорема, позволяющая оценивать абсолютную погрешность величины ε^* для интервальной матрицы вида (5) по возмущениям центральной матрицы:

Теорема 5. *Пусть некоторой точной невырожденной матрице $A_c \in \mathbb{R}^{n \times n}$, удовлетворяющей условию*

$$\varepsilon^* = \|A_c^{-1}\|_{\infty,1}^{-1} = \|A_c^{-1}\|_{\ell_1}^{-1}$$

сопоставлена приближенная матрица $\tilde{A}_c \in \mathbb{R}^{n \times n}$ такая, что

$$\delta = \|\tilde{A}_c - A_c\|_{1,\infty} < \varepsilon^*.$$

Тогда абсолютная погрешность σ определения минимального радиуса ε^ вырожденной интервальной матрицы $\mathbf{A}(A_c, \varepsilon)$, возникающая при замене A_c на \tilde{A}_c , не превосходит числа δ .*

Благодарности. Автор благодарит своего научного руководителя д.ф.-м.н., профессора Горелика Виктора Александровича за ценные замечания и внимание к работе. Особую благодарность автор выражает д.ф.-м.н. Шарому Сергею Петровичу за то, что именно он обратил внимание автора на проблему исследования невырожденности интервальных матриц, а также любезно предоставил труднодоступные для автора полные тексты статей [5], [7] и большое количество других материалов по интервальным вычислениям.

Список литературы

- [1] Ерохин В.И. Оптимальная матричная коррекция и регуляризация несовместных линейных моделей // Дискрет. анализ и исслед. операций. Сер. 2. – 2002. – Т. 9, № 2. – С. 41–77.
- [2] Ерохин В.И. Об условиях невырожденности интервальных матриц // Электронный журнал “Исследовано в России”. – 2003. – С. 2488–2508. – <http://zhurnal.ape.relarn.ru/articles/2003/214.pdf>.
- [3] Тихонов А.Н. О нормальных решениях приближенных систем линейных алгебраических уравнений // Докл. АН СССР. – 1980. – Т. 254, № 3. – С. 549–554.
- [4] Хорн Р., Джонсон Ч. Матричный анализ. – М.: Мир. 1989.
- [5] Coxson G.E. Computing exact bounds on elements of an inverse interval matrix in NP-hard // Reliable Computing. – 1999. – Vol. 5, № 2. – P. 137–142.
- [6] Nemirovskii A. Several NP-hard problems arising in robust stability analysis // Mathematics of Control, Signals and Systems. – 1993. – Vol. 6, № 1. – P. 99–105.
- [7] Rex G., Rohn J. Sufficient conditions for regularity and singularity of interval matrices // SIAM Journal on Matrix Analysis and Applications. – 1999. – Vol. 20, № 2. – P. 437–445.
- [8] Rohn J. Checking positive definiteness or stability of symmetric interval matrices is NP-hard // Commentationes Mathematicae Universitatis Carolinae. – 1994. – Vol. 35, № 4. – P. 795–797.

Интервальная математическая библиотека, основанная на разложениях в ряды Чебышева и Тейлора

А.Г. Ершов*, Т.П. Кашеварова†

Введение

Интервальный анализ, активно развиваемый в последние три десятилетия, выводит вычислительную математику на качественно новую ступень. Интервальные методы позволяют получать гарантированные интервальные оценки множеств решений в виде многомерных брусков. Кроме того, использование интервалов дает возможность корректно учитывать погрешности, вызываемые машинными округлениями.

Представление данных в виде интервалов вызывает необходимость разработки соответствующей математической библиотеки. В настоящее время существует несколько разработок интервальных математических библиотек, однако почти каждая из них, находящаяся в открытом доступе, обладает теми или иными недостатками. Так, например, для реализации библиотеки математических функций Pascal-XSC (см. [6]) используется целочисленная арифметика, которая позволяет получать точные результаты при операциях над числами с плавающей точкой, однако использование целочисленной арифметики значительно увеличивает время вычислений.

По этой причине авторами были разработаны и реализованы алгоритмы вычисления гарантированных верхних и нижних границ элементарных математических функций с помощью направленных округлений на основе разложений в ряды Чебышева и Тейлора. Эти алгоритмы позволяют обеспечить высокую точность определения верхних и нижних границ, а также производительность, не худшую по сравнению со стандартными математическими функциями библиотеки C. Разработанная интервальная математическая библиотека была реализована в рамках работ над проектом SibCalc (см. [7]).

1. Основные определения интервального анализа

Определим интервал над полем действительных чисел \mathbb{R} следующим образом.

Определение. Множество

$$\mathbf{x} = [\underline{x}, \bar{x}] = \{x \mid x \in \mathbb{R} \cup \{-\infty, +\infty\}, \underline{x} \leq x \leq \bar{x}\}$$

*Институт систем информатики им. А.П. Ершова СО РАН, ЗАО “Ледас”, eag@sib3.ru

†Институт систем информатики им. А.П. Ершова СО РАН, toma@iis.nsk.su

будем называть интервалом на \mathbb{R} , а \underline{x} , \bar{x} , соответственно, его нижней и верхней границами. Кроме того, определим специальный интервал \emptyset , который не содержит ни одного элемента.

Если нижняя граница равна верхней, то интервал содержит только одно число, такой интервал называется вырожденным. Множество всех интервалов обозначим через \mathbb{IR} .

В данной работе мы будем использовать расширенную интервальную арифметику [1, 5]. Машинная реализация выполнения арифметических операций осуществлена с помощью вычислений с направленными округлениями.

Сложение и умножение в интервальной арифметике обладают ассоциативностью и коммутативностью, но не обладают дистрибутивностью. Для интервалов выполняется свойство, называемое субдистрибутивностью:

$$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) \subseteq \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}.$$

Для операций над интервалами выполняется основное свойство интервальных вычислений – монотонность по включению, т.е. если $\mathbf{a}^{(k)}, \mathbf{b}^{(k)} \in \mathbb{IR}$ и $\mathbf{a}^{(k)} \subseteq \mathbf{b}^{(k)}$, $k = 1, 2$, тогда для любой операции справедливо

$$\mathbf{a}^{(1)} * \mathbf{a}^{(2)} \subseteq \mathbf{b}^{(1)} * \mathbf{b}^{(2)}.$$

2. Математические основы алгоритмов вычисления элементарных функций

Интервальная библиотека математических функций является базой для всех интервальных методов, поэтому от корректности ее реализации зависит гарантированность получаемых результатов.

При разработке алгоритмов вычисления математических функций стоят две главные задачи – точность и скорость. Интервальное представление позволяет получать нижнюю и верхнюю оценки искомой функции, то есть обеспечивает гарантированность вычислений, однако здесь возникает новая проблема – получение как можно более узких интервалов, в идеале, состоящих из двух подряд идущих машиннопредставимых вещественных чисел. Время вычисления значений интервальных функций существенно влияет на эффективность вычислений в целом. Именно поэтому было принято решение использовать равномерные приближения Чебышева, поскольку они обеспечивают наименьшую абсолютную погрешность при фиксированном числе членов ряда [4].

Алгоритмы вычисления интервальных значений математических функций, реализованные в SibCalc, основаны на разложении функций в ряды Чебышева и Тейлора с корректным округлением результатов арифметических операций и оценкой остаточного члена. Для получения наиболее узких интервалов, гарантированно включающих математическое значение вычисляемой функции, был реализован ряд оригинальных приемов, позволяющих значительно уменьшить величину вычислительной ошибки. Ряды Тейлора используются только в окрестности нуля, так как в данном случае они обеспечивают более высокую относительную точность. На остальных участках разложения используются ряды Чебышева, поскольку они обеспечивают равномерное приближение и требуют меньшего числа членов ряда по сравнению с рядами Тейлора.

2.1. Многочлены Чебышева. Разложения по многочленам Чебышева обычно строятся в виде

$$f(x) = \sum_{n=0}^{\infty} a_n T_n(x), \quad x \in [-1, 1].$$

При этом x может быть некоторой линейной функцией: если необходимо рассматривать разложение функции на интервале $[a, b]$, то используется линейное преобразование [4]. В работе рассматриваются смещенные чебышевские многочлены $T_n^*(x)$, заданные на интервале $[0, 1]$. Для них существует рекуррентные формулы (см. [4]):

$$T_{n+1}^*(x) = 2(2x - 1)T_n^*(x) - T_{n-1}^*(x).$$

Для отрезка $[-1, 1]$, остаточный член разложения в ряд Чебышева $\omega_n(x) = \frac{1}{2^{n-1}} T_n(x)$ будет иметь наименьшее возможное значение $\sup |\omega_n(x)|$ среди всех многочленов данной степени. В этом случае оценка погрешности разложения $L_n(x)$ функции в ряд Чебышева примет вид (см. [2]):

$$|f(x) - L_n(x)| \leq \frac{M_{n+1}}{2^n(n+1)!}, \quad \text{где } M_{n+1} = \sup_x |f^{(n+1)}(x)|.$$

2.2. Организация машинных вычислений. Вычисление элементарной функции на всей области определения сводится с помощью соответствующих формул приведения к вычислению ее на некотором, называемом каноническим, интервале. Для уменьшения вычислительных затрат канонический интервал делится на подинтервалы длиной 2^{-k} , на каждом из которых получаем свое разложение. Такой способ деления позволяет эффективно определять подинтервал, в который попадает аргумент функции, и провести линейное преобразование приведения к каноническому для смещенных многочленов Чебышева интервалу $[0, 1]$ без привнесения ошибки округления.

Вид частичной суммы ряда Чебышева можно преобразовать следующим образом:

$$f(x) = \sum_{i=0}^n a_i T_i(x) = \sum_{i=0}^n a_i \sum_{j=0}^i T_{i,j} x^j = \sum_{k=0}^n c_k x^k, \quad c_k = \sum_{i=k}^n a_i T_{i,k},$$

получив тем самым частичную сумму обычного степенного ряда. Коэффициенты разложений в степенные ряды на каждом подинтервале вычисляются заранее с необходимой точностью и хранятся в отдельных таблицах. Такой подход незначительно увеличивает объем памяти, зато существенно увеличивает скорость вычисления функции.

При вычислении элементарных функций с помощью разложения в ряд возникает три рода погрешностей:

- 1) погрешность, вызываемая использованием вместо бесконечного ряда его конечной частичной суммы. Эту погрешность можно оценить сверху с помощью формул оценки остаточного члена;
- 2) погрешность, вызываемая использованием вместо истинных значений коэффициентов полиномов их машинных приближений;

3) погрешность округления результатов арифметических операций.

Важным фактором в организации машинных вычислений частичной суммы ряда является порядок суммирования, поскольку он определяет величину погрешности третьего рода. Так как мы знаем вид элементарной функции и ее поведение на интервале, мы можем выбрать максимально эффективный порядок суммирования. Разложение в ряд элементарных функций таково, что наибольший вклад в сумму вносят члены низкого порядка, меньший – члены более высокого порядка, а наименьший – члены всех порядков выше некоторого заранее выбранного, сумму которых мы и оцениваем сверху с помощью известных формул оценки остаточного члена. Уменьшение величины членов частичной суммы с увеличением их порядка происходит достаточно быстро, как правило, соотношение между величинами двух соседних членов меньше величины интервала разбиения (в нашей реализации имеющей порядок 10^{-2} – 10^{-3}). Оптимальным, таким образом, является суммирование частичной суммы, начиная с членов высокого порядка к членам более низкого порядка. В таком случае, ошибки округления, внесенные на более ранних этапах вычисления, становятся незначительными на более поздних этапах вычисления; в результате, общая погрешность вычисления третьего рода формируется почти исключительно при добавлении члена первого порядка и свободного члена. Заметим, что такому порядку вычисления соответствует известная и вычислительно эффективная (минимум операций с плавающей точкой) схема Горнера.

Для корректного осуществления вычисления верхней (нижней) границы элементарной функции с помощью разложения в ряд к частичной сумме ряда необходимо добавлять (отнимать) два дополнительных слагаемых: первое оценивает сверху погрешность первого рода (остаточный член ряда), второе оценивает сверху погрешность второго рода, возникающую из-за использования вместо истинных коэффициентов Чебышева и Тейлора их машинных приближений. Величина первого слагаемого имеет порядок главного члена в остатке ряда Чебышева, величина второго – порядок разности между истинным и точным значением свободного члена. Таким образом, отношение второго слагаемого к вычисляемой сумме имеет порядок машинной точности, а значит, учет этого слагаемого негативным образом влияет на ширину интервала между верхней и нижней оценками значения функции, во многих случаях не позволяя добиться желаемой ширины в один ulp (unit in last position – величина единицы в последнем разряде мантииссы). Для того, чтобы избежать этого эффекта, предлагается использовать следующие формулы вычисления верхней и нижней оценки (при условии $x \geq 0$ включены соответственно флаг округления вниз и вверх):

$$F_+(x) = c + (((((r[5] \times x + r[4]) \times x + r[3]) \times x + r[2]) \times x + r[1]) \times x + d_+),$$

$$F_-(x) = c + (((((s[5] \times x + s[4]) \times x + s[3]) \times x + s[2]) \times x + s[1]) \times x + d_-).$$

Здесь r и s – массивы, содержащие приближения коэффициентов ряда разложения с округлением вверх и вниз соответственно; $c + d_+ \geq a_0 \geq c + d_-$, a_0 – истинное значение свободного члена разложения, d_+ и d_- – два последовательно идущих вещественных числа, отношение d_+ к a_0 имеет порядок машинной точности.

При организованных таким образом вычислениях мы добиваемся того, что свободный член учитывается со значительно более высокой точностью, снижая тем самым погрешность второго рода, порядок которой становится равным произведению длины интервала разбиения на разность между истинным и точным значением коэффициента при x . Ценой за достигнутую точность вычислений является одна дополнительная операция сложения, расходы на выполнение которой не слишком значительны по сравнению с общим временем вычисления элементарной функции.

Рассмотренный выше прием имеет смысл, если наибольший по величине вклад в частичную сумму вносит именно свободный член — тогда представление его в виде большей и меньшей части оправдано. В противном случае, больший вклад вносит член при первой степени аргумента, и погрешность, вносимая им, превышает погрешность, вносимую неточным представлением свободного члена. На таких интервалах разбиения предлагается использовать следующие формулы:

$$F_+(x) = x + (((((r[5] \times x + r[4]) \times x + r[3]) \times x + r[2]) \times x + r[1]) \times x + r[0]),$$

$$F_-(x) = x + (((((s[5] \times x + s[4]) \times x + s[3]) \times x + s[2]) \times x + s[1]) \times x + l[0]).$$

Здесь мы разбиваем на два слагаемых уже член первой степени, используя в качестве большего слагаемого x и в качестве меньшего слагаемого $r[1] \times x$ (или $s[1] \times x$). Для этого есть три причины. Первая состоит в том, что разлагаемые функции ($\sin()$, $\operatorname{asin}()$, $\tan()$, $\operatorname{atan}()$, $\ln(1+x)$) в областях, где значения функции существенно меньше единицы, достаточно хорошо приближаются функцией x — а значит, величина x существенно превышает остальные члены суммы. Вторая заключается в том, что представление члена первой степени в виде суммы большей и меньшей части $c \times x$ и $d \times x$ ведет к тому, что возникает погрешность третьего рода при вычислении произведения c и x , чего можно избежать, используя в качестве большей части x . Третья причина состоит в том, что, отказываясь от лишнего умножения на коэффициент, мы экономим время вычислений.

Существенным недостатком разложения в ряд Чебышева является тот факт, что частичная сумма ряда Чебышева обеспечивает наилучшее приближение на интервале в смысле абсолютного отклонения, но не в смысле относительного отклонения (см. [3]), которое наиболее важно для расчета элементарных функций. Поэтому на интервалах, где относительная погрешность существенно превышает абсолютную погрешность, используется разложение в ряд Тейлора. Однако почти для всех элементарных функций ($\sin()$, $\operatorname{asin}()$, $\tan()$, $\operatorname{atan}()$, $\ln(1+x)$) на области их разложения в ряд такого рода интервал разбиения только один — это интервал $[0, 2^{-k}]$ (для $\ln(1+x)$ еще есть интервал $[-2^{-k}, 0]$). Заметим, что вычисление этих функций в нуле с помощью разложения в ряд Чебышева привело бы к плачевным результатам — максимальное отклонение приближения Чебышева от истинной функции достигается как раз на концах интервала приближения, и в малой окрестности нуля относительная погрешность была бы очень большой. Кроме того, есть еще один аргумент в пользу использования на этом интервале разложения в ряд Тейлора — все эти функции разлагаются в ряд Тейлора в окрестности нуля только по нечетным степеням, что существенно экономит время вычислений.

2.3. Использование одного флага округления. Реализация флагов округления на многих компьютерных платформах такова, что время, затрачиваемое на изменение флагов округления, сравнимо со временем, затрачиваемым на вычисление элементарной функции как частичной суммы ряда Чебышева или Тейлора. В связи с этим в описываемой авторами реализации флаг округления выставляется один раз, до проведения вычислений, а вычисления верхней и нижней границ функций реализованы так, что они делаются при одном и том же флаге округления. Например, если выставлен флаг округления “округлять вверх”, и $x \geq 0$, то вычисление верхней границы элементарной функции происходит по обычной формуле Горнера:

$$F(x) = ((((((r[5] \times x + r[4]) \times x + r[3]) \times x + r[2]) \times x + r[1]) \times x + r[0]),$$

где $r[i]$ – коэффициенты полинома Чебышева, посчитанные для определения верхней границы, а вычисление нижней границы происходит по следующей модифицированной формуле Горнера:

$$F(x) = -((((((-s[5]) \times x - s[4]) \times x - s[3]) \times x - s[2]) \times x - s[1]) \times x - s[0]),$$

где $s[i]$ – коэффициенты полинома Чебышева, посчитанные для определения нижней границы.

Использование функций вычисления верхних и нижних границ с одним и тем же флагом округления позволяет эффективно осуществлять большие объемы вычислений. Кроме того, во избежание влияния вызова функций математической библиотеки на дальнейшие вычисления, все функции математической библиотеки имеют дубликаты, которые дополнительно меняют флаг округления на значение “округлять вверх” перед началом вычисления элементарной функции, и устанавливают исходный флаг округления после завершения вычисления.

3. Примеры интервальных алгоритмов вычисления элементарных функций

В качестве примеров приведем алгоритмы вычисления некоторых элементарных функций.

3.1. Вычисление функции синус. Разложение синуса в ряд проводится на интервале $[0, \pi/2]$ (см. [3]). Этот интервал разбивается на подинтервалы длиной 2^{-k} , на каждом подинтервале $x \in \left[\frac{i}{2^k}, \frac{i+1}{2^k}\right]$ с помощью преобразования $t = 2^k \times \left(x - \frac{i}{2^k}\right)$ задача о разложении синуса сводится к задаче о разложении функции $\sin\left(2^{-k} \times t + \frac{i}{2^k}\right)$ на каноническом интервале $t \in [0, 1]$. Для данной длины подинтервала достаточно использовать разложение в ряд Чебышева до пятой степени включительно. Заметим, что разложение в ряд Тейлора требует для достижения такой же точности при том же числе членов разложения использования в 4–8 раз большего числа подинтервалов, что ведет к 4–8-кратному увеличению объема используемой памяти.

Пусть, к примеру, $k = 7$ и $i = 30$, т. е. мы разлагаем $\sin(x)$ на промежутке $x \in \left[\frac{30}{128}, \frac{31}{128}\right]$. Мы вычисляем верхние и нижние оценки коэффициентов Чебышева и составляем коэффициенты разложения в степенной ряд для вычислений верхней и нижней границы синуса, в результате чего получаем формулы для приведенного аргумента $t = 128 \times \left(x - \frac{30}{128}\right)$:

$$\begin{aligned} \sin_+(x) = & (((((8.097870612551672648 \cdot 10^{-3} \times t + \\ & 9.676530503240125813 \cdot 10^{-3}) \times t - 1.621099466468840788 \cdot 10^{-1}) \times t - \\ & 1.161175593052558991 \cdot 10^{-1}) \times t + 9.726596782449125067 \cdot 10^{-1}) \times t + \\ & 1.952311138366777654 \cdot 10^{-17}) + 0.232235118611511443, \end{aligned}$$

$$\begin{aligned} \sin_-(x) = & -((((((-8.097870612551670913 \cdot 10^{-3} \times t - \\ & 9.676530503240124079 \cdot 10^{-3}) \times t + 1.621099466468841066 \cdot 10^{-1}) \times t + \\ & 1.161175593052559130 \cdot 10^{-1}) \times t - 9.726596782449123957 \cdot 10^{-1}) \times t - \\ & 1.942458581633222228 \cdot 10^{-17}) - 0.232235118611511443), \end{aligned}$$

$$\sin(x) = [\sin_-(x), \sin_+(x)].$$

Вычисления значений синуса для значений аргумента, выходящих за пределы интервала $\left[0, \frac{\pi}{2}\right]$, требуют применения формул приведения тригонометрического аргумента. Основной трудностью в них является корректное вычитание полных периодов тригонометрической функции для значений аргумента, достаточно удаленных от нуля. Обычно выбирается некоторый достаточно большой интервал, на котором значение тригонометрических функций желательно вычислять с высокой точностью (в разработанной библиотеке — $[-10^9, 10^9]$), а за пределами этого интервала высокая точность вычислений не требуется. Используя специальное представление числа $2\pi = \sum_{i=0}^m p_i$, где p_i удовлетворяют неравенству $8 \times 2^{-l \times i} \geq p_i \geq 4 \times 2^{-l \times i}$ и имеют мантиссу длины l , мы можем от тригонометрического аргумента вычесть величину $2\pi n = \sum_{i=0}^m p_i n$ так, чтобы вычисление произведений $p_i n$ происходило без ошибки округления. Для этого необходимо, чтобы сумма l и $\lceil \ln n \rceil$ — соответственно разрядностей мантисс p_i и n — не превышала разрядности мантиссы машинного представления. Заметим также, что разрядность мантиссы машинного представления, удовлетворяющему стандарту IEEE-754, равна 52, а для указанного выше интервала $[-10^9, 10^9]$ величина $\lceil \ln n \rceil$ ограничена сверху 28 — таким образом, l можно выбрать равным 24. Из практических соображений длина представления 2π ограничена величиной $m = 3$.

Вычисление функции синус для интервального значения аргумента состоит в том, чтобы найти глобальные минимум и максимум на заданном интервале. Понятно, что если ширина интервального значения аргумента x больше, чем 2π , то $\sin(x) = [-1.0, 1.0]$. Если ширина интервала-аргумента меньше, чем 2π , то определяются значения аргумента x_{\min} и x_{\max} , в которых достигаются соответственно минимальное и максимальное значения функции синус на дан-

ном интервале. Это можно сделать используя свойство кусочной монотонности функции. Результатом вычисления синуса для невырожденного интервального аргумента будет интервал, нижняя граница которого равна нижней границе значения синуса в точке x_{\min} , а верхняя – верхней границе значения синуса в точке x_{\max} .

3.2. Вычисление логарифма. Разложение логарифма в ряд проводится на интервале $[2^{-1/2}, 2^{1/2}]$. Вне этого интервала мы используем формулу $x = 2^m y$ для выбора подходящего значения y из интервала разложения, вычисляем $\ln(y)$, а затем пользуемся соотношением $\ln(x) = \ln(y) + m \ln(2)$. Такое значение y всегда можно найти, поскольку отношение верхней и нижней границы интервала $[2^{-1/2}, 2^{1/2}]$ равно 2. Выбор этого интервала обладает также тем преимуществом, что он содержит значение 1.0, на котором логарифм обращается в нуль, что позволяет вычислять логарифм в окрестности 1.0 с высокой относительной точностью.

В силу того, что логарифм является монотонно возрастающей функцией, его вычисление для невырожденного интервального аргумента сводится к вычислению нижней границы на левом конце и верхней границы на правом конце интервала-аргумента.

3.3. Вычисление экспоненты. Разложение экспоненты в ряд проводится на интервале $[0, \ln 2]$. Вне этого интервала мы используем формулу $x = y + m \times \ln 2$ для выбора подходящего значения y из интервала разложения, вычисляем e^y , а затем пользуемся соотношением $e^x = 2^m e^y$. Заметим, что вычисления по последней формуле проводятся без привнесения ошибки округления.

3.4. Вычисление квадратного корня. Поскольку на современных компьютерных платформах затраты на вычисление квадратного корня малы (они примерно равны затратам на вычисление операции деления), авторами было принято решение использовать для вычисления верхних и нижних границ этой функции специальную технику. Для вычисления нижней границы мы используем значение t , возвращаемое стандартной математической библиотекой C, и проверяем, является ли оно нижней границей с помощью вычисления его квадрата $t \times t$ с округлением вверх. Если результат не превосходит аргумента, то t является нижней границей; иначе мы выполняем сдвиг: берем вместо t соседнее снизу машиннопредставимое число и повторяем описанную выше операцию. Вычисление верхней границы проводится аналогичным способом с помощью сдвига вверх. Как показали эксперименты, для вычисления границ ни разу не пришлось выполнять более одного сдвига.

Заключение

Математическая библиотека реализована как модуль, независимый от платформы (с единственным ограничением соответствия представления вещественных чисел стандарту IEEE-754), с интерфейсом на языке C, состоящий из набора функций на языке C++. Он содержит следующие функции: синус, косинус, тангенс, котангенс, арксинус, арккосинус, арктангенс, квадратный корень,

натуральный логарифм, экспонента, функция возведения в вещественную степень. Данная библиотека встроена в интервальный решатель нелинейных систем уравнений SibCalc. Для обеспечения точности результата в библиотеке используются оригинальные разработки авторов. Гарантированный интервальный результат вычисления функции для вырожденного интервального значения аргумента в 80–90% случаях состоит из двух идущих подряд машинно-представимых чисел, то есть является наилучшей интервальной оценкой. При этом обеспечивается высокая производительность вычисления элементарных функций, не уступающая производительности стандартной математической библиотеки C.

Список литературы

- [1] Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. – М.: Мир, 1987. – 260 с.
- [2] Березин И.С. Жидков Н.П. Методы вычислений. – М.: Физматгиз, 1959. – Т. 1. – 464 с.
- [3] Люк Ю. Специальные математические функции и их аппроксимации. – М., 1980. – 608 с.
- [4] Ремез У.Я. Основы численных методов чебышевского приближения. – Киев: Наукова думка, 1969.
- [5] Hansen E. Global Optimization Using Interval Analysis. – New York: Marcel, Dekker, 1992.
- [6] Клатте Р., Кулиш У., Неага М., Рац Д., Ульрих У. PASCAL-XSC. Язык численного программирования. – М.: ДМК, 2000.
- [7] Kleymenov A., Petunin D., Semenov A., Vazhev I. A Model of Cooperative Solvers for Computational Problems // Proc. of the 4th Intern. Conf. PPAM 2001, Poland, September: Lect. Notes Comp. Sci. – Vol. 2328. – P. 797–802.

О построении зависимостей по эмпирическим данным с интервальной ошибкой

С.И. Жилин*

Аннотация. В работе рассматривается задача построения линейно параметризованных зависимостей типа “вход–выход” по эмпирическим данным с интервальной ошибкой измерения выходной переменной. Предложен метод выявления наблюдений–выбросов. По результатам имитационных экспериментов проведено сравнение точечных оценок максимального правдоподобия и наименьших квадратов с точечными оценками, получаемыми путем выбора срединной точки интервальных оценок, и получен вывод о конкурентоспособности подобного нестатистического подхода.

1. Введение

Рассматриваемая в работе постановка задачи построения и анализа зависимостей по эмпирическим данным с интервальной ошибкой восходит к идее Л. В. Канторовича [5] и исследуется в [1, 3, 8–11], совпадая в них с точностью до терминологии.

Суть задачи состоит в построении по эмпирическим данным линейно параметризованной зависимости

$$y = \sum_{i=1}^n \beta_i x_i, \quad (1)$$

где $x \in \mathbb{R}^n$ – вектор входных переменных, $\beta \in \mathbb{R}^n$ – вектор параметров, подлежащих оцениванию, y – скалярная выходная переменная.

Зависимость конструируется по эмпирической информации, главное место в которой занимает таблица экспериментальных данных, полученная в N наблюдениях:

$$T = \{(y_j, x_{1j}, \dots, x_{nj}) \mid j = 1, \dots, N\}.$$

При этом предполагается, что погрешностью измерения входных переменных x_i можно пренебречь, а выходная переменная y в j -м наблюдении измеряется с предельной абсолютной погрешностью ε_j .

Ограниченность ошибки измерения выходной переменной позволяет для каждого из наблюдений записать двустороннее неравенство

$$y_j - \varepsilon_j \leq \sum_{i=1}^n \beta_i x_{ij} \leq y_j + \varepsilon_j, \quad j = 1, \dots, N. \quad (2)$$

Неравенства (2) в совокупности определяют множество B допустимых значений параметров $\beta = (\beta_1, \dots, \beta_n)$, именуемое множеством неопределенности.

* Алтайский государственный университет, Барнаул.

В качестве интервальных оценок параметров β_i часто используются проекции $[\underline{\beta}_i, \overline{\beta}_i]$ наименьшего из охватывающих множество B брусков, а в качестве точечных оценок – середины этих проекций

$$\hat{\beta}_i = (\underline{\beta}_i + \overline{\beta}_i)/2. \quad (3)$$

Границы проекций могут быть найдены решением задач линейного программирования:

$$\underline{\beta}_i = \min_{\beta \in B} \beta_i, \quad \overline{\beta}_i = \max_{\beta \in B} \beta_i, \quad i = 1, \dots, n. \quad (4)$$

В отношении множества B может также ставиться задача интервального и точечного прогноза значения выходной переменной y в точке x . Границы интервальных оценок $[\underline{y}_i(x), \overline{y}_i(x)]$ могут быть найдены решением задач линейного программирования:

$$\underline{y}_i(x) = \min_{\beta \in B} \sum_{i=1}^n \beta_i x, \quad \overline{y}_i(x) = \max_{\beta \in B} \sum_{i=1}^n \beta_i x, \quad i = 1, \dots, n. \quad (5)$$

Точечная оценка прогноза $\hat{y}(x)$ строится как середина интервальной оценки:

$$\hat{y}(x) = \frac{1}{2}(\underline{y}(x) + \overline{y}(x)). \quad (6)$$

Однако любая из указанных задач оценивания имеет смысл лишь в случае ограниченности и непустоты множества неопределенности B . Неограниченность множества B очевидным образом распознается в результате ранговых исследований матрицы наблюдений и содержательно может интерпретироваться как недостаток эмпирической информации для построения зависимости. Пустота множества B говорит о противоречивости собранной информации, одной из возможных причин которой может служить наличие выбросов среди наблюдений. В работе предлагается метод выявления выбросов, позволяющий добиться непротиворечивости исходных данных и, соответственно, непустоты множества неопределенности.

Еще одним вопросом, рассматриваемым в работе, является выяснение соотношения оценок, получаемых с помощью изложенного выше подхода, с традиционно используемыми статистическими оценками метода максимального правдоподобия (ММП) и метода наименьших квадратов (МНК) на основе имитационного эксперимента. О необходимости проведения подобного эксперимента как единственного средства сравнения методов оценивания, опирающихся на различные системы гипотез, говорилось в заметке [2]. Для определенности и краткости оценки, получаемые посредством выражений (3)–(6), далее будем называть нестатистическими.

2. Выявление выбросов

Одним из наиболее значимых с практической точки зрения свойством описанного во введении подхода является его потенциальная способность выявлять ситуации, в которых собранные для построения зависимости совокупности данных противоречивы. Индикатором наличия противоречий в данных

является пустота множества неопределенности. Основными источниками противоречий являются либо нарушение гипотезы о структуре конструируемой зависимости, либо наличие выбросов в данных. Выбор способа разрешения противоречий в конечном итоге определяется исследователем по результатам всестороннего анализа. Однако результаты такого анализа во многом зависят и от того, какой информацией располагает для этого исследователь. Настоящий раздел посвящен описанию одного из возможных подходов к получению информации, позволяющей разрешать противоречия, возникающие в случае наблюдений с выбросами.

Выброс представляет собой определенную особенность, нетипичное наблюдение по отношению к остальным данным. Это означает, что выбросы должны подвергаться особенно тщательному рассмотрению с целью выяснения причин их возникновения. Иногда выброс дает такую информацию, которую не могут дать другие наблюдения, и является результатом измерений при необычной комбинации условий. В этом случае требуется дальнейшее углубленное исследование. Однако чаще выбросы вызваны грубыми промахами при регистрации значений наблюдаемых величин. В этом случае производится исключение или целенаправленное ослабление веса наблюдения–выброса в общей информационной совокупности.

Выброс, обусловленный грубым промахом при регистрации результатов измерений, можно трактовать как наблюдение, предельная погрешность которого занижена по отношению к реальной ошибке, имевшей место при измерении. Чтобы такое наблюдение стало “правильным”, необходимо найти нижнюю границу реальной ошибки, при которой наблюдение не будет вступать в противоречие с остальными. Сравнение значения этой нижней границы и приписанной наблюдению ошибки, позволяет строить некоторые суждения относительно степени несоответствия наблюдения–выброса общей картине.

Нижние границы предельных ошибок наблюдений, при которых множество неопределенности становится непустым, можно отыскивать, решая задачу

$$\min_{\beta, w} \sum_{j=1}^N w_j, \quad (7)$$

$$y_j - w_j \varepsilon_j \leq \sum_{j=1}^N \beta_j x_j \leq y_j + w_j \varepsilon_j, \quad w_j \geq 1, \quad j = 1, \dots, N, \quad (8)$$

где w_j – масштабирующие коэффициенты, указывающие, во сколько раз необходимо растянуть исходную предельную ошибку ε_j для того, чтобы j -е наблюдение не вступало в противоречие с общей совокупностью данных. Полученные в результате решения задачи (7), (8) значения масштабирующих коэффициентов, превосходящие единицу, соответствуют наблюдениям–выбросам. Если у исследователя есть основания считать, что надежность некоторых наблюдений одинакова, то система ограничений (8) может быть пополнена равенствами вида $w_{j_1} = w_{j_2} = \dots = w_{j_k}$. В случае, когда в надежности каких-либо наблюдений исследователь уверен полностью, при решении задачи (7), (8) соответствующие им величины w_j можно положить равными единице.

Количество наблюдений, для которых масштабирующие коэффициенты w_j , полученные в результате решения задачи (7), (8), превосходят единицу, позво-

ляет судить о доле выбросов в совокупности данных. Большая доля выбросов может говорить либо о неверно выбранной структуре зависимости, либо о том, что предельные ошибки измерения занижены во многих наблюдениях (например, в результате неверной оценки точности измерительного прибора).

Любопытным представляется тот факт, что предложенный подход к решению противоречий в совокупности экспериментальных данных укладывается в рамки теории коррекции несобственных задач линейного программирования [4] и может рассматриваться как один из возможных способов параметризации несобственной задачи линейного программирования с целью поиска ее аппроксимации собственной задачей и путей коррекции с минимальными затратами.

3. Экспериментальное сравнение статистических и нестатистических оценок

Главным отличием в системах гипотез, лежащих в основании статистического и нестатистического подходов к построению и анализу зависимостей, является гипотеза о структуре ошибки.

В статистическом подходе ошибка полагается случайной величиной, описываемой некоторым законом распределения, выбираемым исследователем. На практике часто, но как показывает ряд исследований [6, 7], далеко не всегда обоснованно, закон распределения ошибки выбирается нормальным. В этом случае наиболее качественные (состоятельные и эффективные) оценки обеспечивает МНК, являющийся частной формой ММП.

Одним же из главных принципов нестатистической обработки наблюдений, определяющим все последующие алгоритмы и получаемые выводы, является равновозможность всех элементов интервала ошибки, а следовательно, и множества неопределенности B .

Описательные способности рассматриваемых методов при построении зависимости по эмпирическим данным предлагается выяснить по результатам вычислительного эксперимента, состоящего в многократном решении каждым из сравниваемых методов задачи точечного прогноза по модельным данным и выяснении стандартных отклонений прогнозных оценок от истинных модельных значений. Модельные данные предлагается генерировать путем добавления ошибки с заданным распределением к точным значениям выходной переменной при фиксированных значениях входных переменных для некоторой заранее известной зависимости. Выбор именно точечных оценок в качестве сравниваемых показателей объясняется их особой ролью для исследователей-практиков и возможностью одинаковой интерпретации, что не вполне реализуемо в отношении статистических нестатистических интервальных оценок.

Что касается выбора распределения ошибки при генерировании модельных данных, то интерес представляют ситуации “наилучшие” для каждого из сравниваемых методов, а также некоторые близкие к ним варианты. Наилучшими условиями для статистических методов являются ситуации, когда ошибка распределения подчиняется некоторому унимодальному распределению, в частности, для МНК таковым является нормальное распределение ошибки. Базовому для ММП предположению о равновозможности всех элементов мно-

жества неопределенности в вероятностных терминах наиболее адекватно соответствует равномерное распределение. Таким образом, сравнительный эксперимент предлагается провести для унимодального и равномерного распределений ошибки, а также некоторых промежуточных распределений.

3.1. Нестатистические оценки и оценки максимума правдоподобия

Сравнение нестатистического метода построения оценок с ММП проведено для семейства распределений с плотностью

$$p_{\alpha}(x) = \begin{cases} \frac{1-2\varepsilon\alpha}{\varepsilon^2}x + \frac{1-\varepsilon\alpha}{\varepsilon}, & -\varepsilon \leq x < 0; \\ \frac{2\varepsilon\alpha-1}{\varepsilon^2}x + \frac{1-\varepsilon\alpha}{\varepsilon}, & 0 \leq x \leq \varepsilon; \end{cases}$$

где ε – абсолютное значение предельной ошибки, а $\alpha \in \left[0, \frac{1}{2\varepsilon}\right]$ – параметр, определяющий степень близости распределения к треугольному. При $\varepsilon = 1$ графики функции $p_{\alpha}(x)$ для граничных и двух промежуточных значений параметра α приведены на рис. 1.

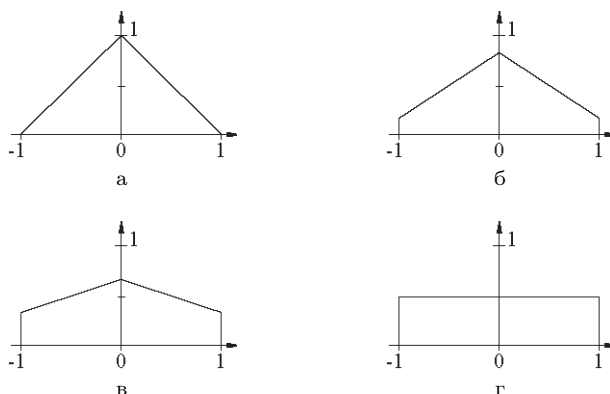


Рис. 1. Графики функции плотности $p_{\alpha}(x)$ при $\varepsilon = 1$ и (а) $\alpha = 0$, (б) $\alpha = 1/6$, (в) $\alpha = 1/3$, (г) $\alpha = 1/2$

При построении оценок ММП при распределении ошибки, близком к равномерному, возникают сложности в выборе оценки, обусловленные неединственностью максимума функции правдоподобия. Выход из этой ситуации видится в регуляризации задачи поиска максимума функции правдоподобия $L(\beta)$ путем добавления слагаемого $\delta|L(\beta)|(\beta - \beta^0)^2$, где $\delta < 0$ – постоянный весовой коэффициент (в эксперименте $\delta = -0.1$), а β^0 – известное модельное значение параметров.

В качестве модельной зависимости была выбрана функция $y = x + 1$, т. е. истинное модельное значение вектора параметров $\beta^0 = (1, 1)$. Совокупность точных значений модельной зависимости была получена путем вычисления значений выходной переменной в узлах регулярной сетки с шагом 1 на интервале $[1; 10]$. Для каждого из фиксированных значений $\alpha_m = \frac{m}{20} \frac{1}{2\varepsilon}$ ($m = 0, \dots, 20$)

5000 раз генерировалась таблица наблюдений путем добавления к точным значениям выходной переменной случайной ошибки из интервала $[-0.5; 0.5]$ с плотностью распределения $p_\alpha(x)$ и каждым из сравниваемых методов строились точечные оценки параметров зависимости, на основе которых вычислялись прогнозные значения зависимости в точке $x = 5.5$. По результатам повторений эксперимента при фиксированном m вычислялось стандартное отклонение каждого из типов оценок от истинного значения модельной зависимости в этой точке. Зависимость стандартного отклонения нестатистических оценок и оценок ММП от m приведена на рис. 2.

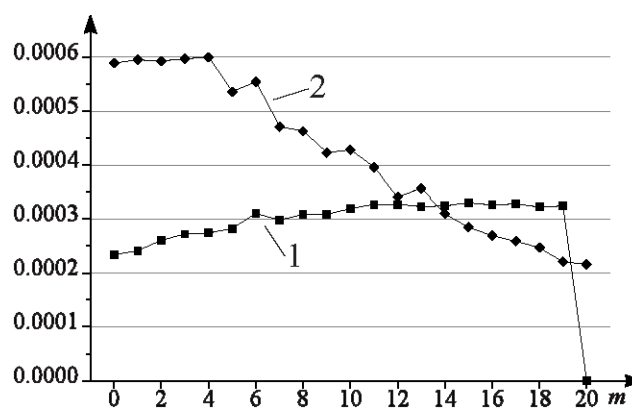


Рис. 2. Среднеквадратичные отклонения прогнозных значений от истинных для ММП (1) и нестатистического метода (2)

Сравнительный анализ стандартных отклонений нестатистического и ММП-прогноза показывает, что при распределениях погрешности, близких к “треугольным”, характер поведения ошибки прогноза соответствует известным соотношениям и закономерностям, свойственным использованным методам оценивания. Действительно, нестатистическая процедура не учитывает дополнительную информацию, связанную с характером распределения, и, соответственно, имеет большую ошибку прогноза. Но по мере приближения распределения погрешности к равномерному ошибка нестатистического прогноза снижается. Это объясняется тем, что такая ситуация становится все более соответствующей базовому для метода построения нестатистических оценок предположению о равновозможности всех элементов множества неопределенности. В то же время при приближении распределения ошибки к равномерному стандартное отклонение ММП-прогноза возрастает и с некоторого момента начинает превосходить стандартное отклонение нестатистического прогноза. Резкое падение стандартного отклонения ММП-прогноза в точке $m = 20$ объясняется возросшим относительным весом регуляризирующего слагаемого.

3.2. Нестатистические оценки и оценки наименьших квадратов

Схема и параметры экспериментов по сравнению нестатистических оценок с оценками наименьших квадратов в основном повторяет схему экспериментов, описанную в предыдущем разделе. Изменения касаются лишь количества

повторений, вида распределения ошибки и кратности наблюдений.

Количество повторений эксперимента при фиксированных параметрах распределения составляло 1000.

Семейство распределений ошибки $N_k(a, \sigma^2)$ в этом случае представляло собою нормальные распределения, усеченные на уровне k , т. е. ошибка принимает значения из интервала $[a - k\sigma, a + k\sigma]$, где a – математическое ожидание, σ – среднее квадратическое отклонение. В проведенной серии экспериментов математическое ожидание было нулевым, среднее квадратическое отклонение единичным, а k выбиралось из интервала $[0, 2; 3]$ с шагом 0,2. По мере роста k получаемые распределения принимали вид от почти равномерного до почти нормального.

По указанной схеме эксперимент проводился в полном объеме для каждого из фиксированных значений кратности наблюдений $Q = 1, 3, 9$.

Результаты эксперимента в графическом виде приведены на рис. 3.

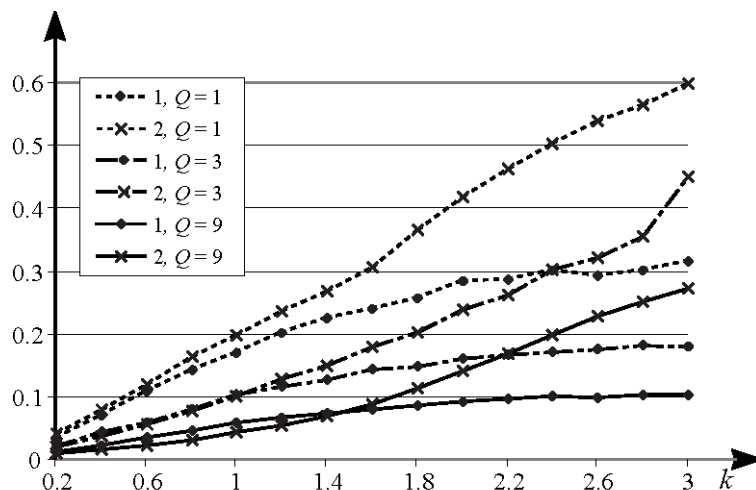


Рис. 3. Зависимость среднеквадратичных отклонений прогнозных значений от истинных для МНК (1) и нестатистического метода (2) от уровня усечения нормального распределения k и кратности наблюдений Q

Качественный анализ взаимосвязей среднеквадратичных отклонений нестатистического и МНК-прогнозов с уровнем усечения нормального распределения ошибки и кратностью измерений позволяет сделать следующие наблюдения:

1. По мере уменьшения уровня усечения нормального распределения ошибки измерений среднеквадратичные отклонения и нестатистического, и МНК-прогнозов также убывают. При этом для МНК-прогноза скорость убывания можно качественно охарактеризовать как логарифмическую или линейно-логарифмическую, в то время как для нестатистического прогноза – как полиномиальную.
2. При больших значениях k оценки МНК-прогноза более устойчивы, чем нестатистические оценки. Однако с уменьшением k их преимущество

утрачивается. Кроме того, с увеличением кратности измерений более устойчивыми становятся нестатистические оценки. Объяснение этому факту, так же как и в случае сравнения с ММП, состоит в уменьшении степени соответствия распределения ошибки измерения гипотезе о нормальности, в рамках которой МНК дает наилучшие результаты. В то же время, приближение распределения ошибки к равномерному все более соответствует одному из базовых предположений нестатистического метода о равновозможности всех элементов интервала ошибки и множества неопределенности.

3. С увеличением кратности измерений устойчивость нестатистических оценок растет несколько быстрее. Тенденция усиливается по мере уменьшения уровня усечения нормального распределения ошибки измерений, то есть по мере приближения распределения к равномерному. Этот факт свидетельствует о способности нестатистического метода неявно накапливать информацию о распределении ошибки, незадействуемую явным образом в отличие от статистических процедур оценивания.

Таким образом, результаты сравнительного анализа точечных нестатистических оценок с оценками, получаемыми методами максимального правдоподобия и наименьших квадратов позволяют сделать вывод о конкурентоспособности нестатистического подхода к построению и анализу зависимостей в случае ограниченности ошибки наблюдений несмотря на то, что статистические методы, вообще говоря, задействуют больше информации, требуя указания явным образом структуры предпочтений на интервале ошибки в виде закона распределения.

Список литературы

- [1] Белов В.М., Суханов В.А., Гузеев В.В., Унгер Ф.Г. Оценивание параметров линейных физико-химических зависимостей прямоугольником метода центра неопределенности // Изв. вузов. Физика. – 1991. – № 8. – С. 35–45.
- [2] Бородюк В.П. Комментарий I к статье А.П. Воцинина, А.Ф. Бочкова, Г.Р. Сотирова “Метод анализа данных при интервальной нестатистической ошибке” // Заводская лаборатория. – 1990. – Т. 56, № 7. – С. 81–83.
- [3] Воцинин А.П., Бочков А.Ф., Сотиров Г.Р. Метод анализа данных при интервальной нестатистической ошибке // Заводская лаборатория. – 1990. – Т. 56, № 7. – С. 76–81.
- [4] Еремин И.И. Противоречивые модели оптимального планирования. – М.: Наука, 1988. – 160 с.
- [5] Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сиб. мат. журнал. – 1962. – Т. 3, № 5. – С. 701–709.
- [6] Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, 1985. – 248 с.
- [7] Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. – 1991. – Т. 57, № 7. – С. 64–66.
- [8] Оскорбин Н.М., Максимов А.В., Жилин С.И. Построение и анализ эмпирических зависимостей методом центра неопределенности // Известия Алтайского государственного университета. – 1998. – № 1. – С. 35–38.

- [9] Спивак С.И. Детальный анализ применения методов линейного программирования при определении параметров кинетической модели // Математические проблемы химии. – Новосибирск: ВЦ СО АН СССР, 1975. – Ч. 2. – С. 35–42.
- [10] Milanese M., Belforte G. Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: linear families of models and estimators // IEEE Transactions on Automatic Control. – 1982. – Vol. 27, № 2. – P. 408–414.
- [11] Rodionova O.Ye., Pomerantsev A.L. Antioxidants activity prediction using DSC measurements and SIC data processing // II Conference on Experimental Methods in Physics of Heterogeneous Condensed Media. – Barnaul, 2001. – P. 239–246.

Модификация метода граничных реализаций для интервальных импульсных последовательностей смешанного типа

С.Ю. Калинкина, С.Г. Пушков*

Аннотация. Рассматривается проблема реализации в пространстве состояний для интервальных динамических систем с дискретным временем. Предложены методы нахождения алгебраических реализаций для полностью неотрицательных и полностью неположительных систем. Представлена модификация метода граничных реализаций для систем “смешанного” типа. Приведены иллюстрирующие численные примеры.

Введение

Одной из распространенных форм представления управляемых объектов и систем управления является представление в пространстве состояний. Задача представления информации об объекте тесно связана с проблемой реализации динамических систем. Эта проблема состоит в построении модели пространства состояний для динамической системы с известным соотношением между ее входными и выходными сигналами.

Моделируя любой динамический процесс, мы часто имеем дело с неопределенностями и неоднозначностями в данных, источники которых могут быть различны: ошибки округления, погрешности измерений из-за естественного несовершенства приборов, использование приближенных чисел и т. д. Одним из приемов, позволяющим учесть такие неопределенности, является представление параметров объекта в виде некоторых множеств. В том случае, когда эти множества представляют собой интервалы, мы имеем дело с интервальной неопределенностью.

Проблема реализации для интервальных динамических систем и виды таких систем рассматривались в [1]. А в данной работе сосредоточим внимание на нахождении алгебраических реализаций интервальных динамических систем с помощью метода граничных реализаций.

1. Проблема реализации для интервальных систем

Классическая проблема реализации состоит в определении модели в пространстве состояний для динамической системы, заданной своим поведением вход-выход. Поведение вход-выход линейной стационарной многомерной управляемой системы может быть охарактеризовано импульсной последовательностью матриц размера $p \times m$ (m – число входов, p – число выходов системы): $\{A_1, A_2, \dots\}$.

*Бийский технологический институт Алтайского государственного технического университета.

В этом случае для заданной последовательности векторов управлений (входной последовательности) $u(0), u(1), \dots \in U$ выходная последовательность векторов $y(0), y(1), \dots \in Y$ определяется соотношениями

$$y(t) = \sum_{i=1}^t A_i u(t-i), \quad t = 1, 2, \dots$$

Задача реализации для данного класса систем состоит в определении математической модели этой системы в пространстве состояний, которая описывается разностными уравнениями

$$\begin{aligned} x(t+1) &= Fx(t) + Gu(t), \\ y(t) &= Hx(t), \quad t = 0, 1, \dots, \end{aligned}$$

где $x(t)$ и $x(t+1)$ – векторы состояний в моменты времени t и $(t+1)$ соответственно. Хорошо известно [2], что задача реализации в этом случае сводится к нахождению тройки матриц (F, G, H) таких, что

$$A_i = HF^{i-1}G, \quad i = 1, 2, \dots$$

Одной из возможных формулировок задачи реализации для интервальных систем может быть такая:

Для заданной последовательности интервальных матриц размера $p \times m$

$$\{\mathbf{A}_1, \mathbf{A}_2, \dots\}, \quad \mathbf{A}_i \in \mathbb{IR}^{p \times m}, \quad i = 1, 2, \dots,$$

определить размерность n и тройку интервальных матриц $(\mathbf{F}, \mathbf{G}, \mathbf{H})$ таких, что выполняются интервальные уравнения

$$\mathbf{A}_i = \mathbf{H}\mathbf{F}^{i-1}\mathbf{G}, \quad i = 1, 2, \dots,$$

где $\mathbf{F} \in \mathbb{IR}^{n \times n}$, $\mathbf{G} \in \mathbb{IR}^{n \times m}$, $\mathbf{H} \in \mathbb{IR}^{p \times n}$, а матричные произведения выполняются справа налево, т. е. сначала вычисляется произведение $\mathbf{F}\mathbf{G}$, затем $\mathbf{F}(\mathbf{F}\mathbf{G})$ и т. д.

Под интервальной линейной стационарной динамической системой с дискретным временем (с m входами, n состояниями и p выходами) будем понимать такую систему $[\Sigma] = (\mathbf{F}, \mathbf{G}, \mathbf{H})$, динамическое поведение которой описывается уравнениями

$$x(t+1) = \mathbf{F}x(t) + \mathbf{G}u(t), \quad y(t) = \mathbf{H}x(t), \quad x(0) = x_0 \in \mathbb{R}^n, \quad t = 0, 1, 2, \dots,$$

где $u(t) \in \mathbb{R}^m$, $x(t), x(t+1) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$, и понимать как семейство математических моделей

$$x(t+1) = Fx(t) + Gu(t), \quad y(t) = Hx(t), \quad x(0) = x_0 \in \mathbb{R}^n, \quad t = 0, 1, 2, \dots,$$

матрицы (F, G, H) которых принадлежат заданным интервальным матрицам $(\mathbf{F}, \mathbf{G}, \mathbf{H})$, т. е. $F \in \mathbf{F} \in \mathbb{IR}^{n \times n}$, $G \in \mathbf{G} \in \mathbb{IR}^{n \times m}$, $H \in \mathbf{H} \in \mathbb{IR}^{p \times n}$.

Другие возможные определения интервальных динамических систем можно найти в [1].

Указанную выше задачу мы далее будем называть *задачей алгебраической реализации*, а тройку интервальных матриц $(\mathbf{F}, \mathbf{G}, \mathbf{H})$ будем называть *алгебраической интервальной реализацией*.

2. Граничные реализации

С импульсной последовательностью интервальных матриц можно связать две обычные (вещественные) импульсные последовательности, определяемые верхними и нижними границами интервальных матриц.

Для последовательности интервальных матриц

$$\{\mathbf{A}_1, \mathbf{A}_2, \dots\} = \{[\underline{\mathbf{A}}_1, \overline{\mathbf{A}}_1], [\underline{\mathbf{A}}_2, \overline{\mathbf{A}}_2], \dots\} \quad (1)$$

реализации последовательности $\{\underline{\mathbf{A}}_1, \underline{\mathbf{A}}_2, \dots\}$ будем называть нижними граничными реализациями последовательности (1), а реализации последовательности $\{\overline{\mathbf{A}}_1, \overline{\mathbf{A}}_2, \dots\}$ будем называть верхними граничными реализациями последовательности (1).

Для полностью неотрицательных интервальных систем метод граничных реализаций опирается на следующие утверждения.

Предложение 1. Если для нижней и верхней граничных реализаций одинаковой размерности $(\underline{\mathbf{F}}, \underline{\mathbf{G}}, \underline{\mathbf{H}})$ и $(\overline{\mathbf{F}}, \overline{\mathbf{G}}, \overline{\mathbf{H}})$ некоторой последовательности интервальных матриц выполняется

- 1) $\underline{\mathbf{F}}, \underline{\mathbf{G}}, \underline{\mathbf{H}}, \overline{\mathbf{F}}, \overline{\mathbf{G}}, \overline{\mathbf{H}}$ – неотрицательные;
- 2) $\underline{\mathbf{F}} \leq \overline{\mathbf{F}}, \underline{\mathbf{G}} \leq \overline{\mathbf{G}}, \underline{\mathbf{H}} \leq \overline{\mathbf{H}}$, то интервальная система $([\underline{\mathbf{F}}, \overline{\mathbf{F}}], [\underline{\mathbf{G}}, \overline{\mathbf{G}}], [\underline{\mathbf{H}}, \overline{\mathbf{H}}])$ является интервальной точной (алгебраической) реализацией этой последовательности.

Предложение 2. Если для граничных реализаций одинаковой размерности $(\underline{\mathbf{F}}, \underline{\mathbf{G}}, \underline{\mathbf{H}})$ и $(\overline{\mathbf{F}}, \overline{\mathbf{G}}, \overline{\mathbf{H}})$ некоторой последовательности интервальных матриц найдутся такие матрицы T_1 и T_2 , что выполняются неравенства

$$\hat{\underline{\mathbf{F}}} \leq \hat{\overline{\mathbf{F}}}, \hat{\underline{\mathbf{G}}} \leq \hat{\overline{\mathbf{G}}}, \hat{\underline{\mathbf{H}}} \leq \hat{\overline{\mathbf{H}}},$$

где

$$\hat{\underline{\mathbf{F}}} = T_1 \underline{\mathbf{F}} T_1^{-1}, \hat{\underline{\mathbf{G}}} = T_1 \underline{\mathbf{G}}, \hat{\underline{\mathbf{H}}} = \underline{\mathbf{H}} T_1^{-1}, \quad (2)$$

$$\hat{\overline{\mathbf{F}}} = T_2 \overline{\mathbf{F}} T_2^{-1}, \hat{\overline{\mathbf{G}}} = T_2 \overline{\mathbf{G}}, \hat{\overline{\mathbf{H}}} = \overline{\mathbf{H}} T_2^{-1}, \quad (3)$$

$\hat{\underline{\mathbf{F}}}, \hat{\underline{\mathbf{G}}}, \hat{\underline{\mathbf{H}}}, \hat{\overline{\mathbf{F}}}, \hat{\overline{\mathbf{G}}}, \hat{\overline{\mathbf{H}}}$ – неотрицательные, то интервальная система

$$([\hat{\underline{\mathbf{F}}}, \hat{\overline{\mathbf{F}}}], [\hat{\underline{\mathbf{G}}}, \hat{\overline{\mathbf{G}}}], [\hat{\underline{\mathbf{H}}}, \hat{\overline{\mathbf{H}}})$$

является интервальной алгебраической реализацией этой последовательности.

Таким образом, решение задачи реализации для интервальных полностью неотрицательных систем может быть осуществлено с помощью следующего алгоритма:

Алгоритм 1.

1. Находим нижнюю и верхнюю граничные реализации одинаковой размерности для заданной импульсной последовательности интервальных матриц.
2. Если найденные граничные реализации удовлетворяют условиям предложения 1, то соответствующая интервальная реализация и будет искомой, иначе
3. В соответствии с предложением 3, с помощью преобразований подобия (2), (3) попытаемся найти эквивалентные (с точностью до изоморфизма) граничные реализации, для которых условия предложения 1 выполняются.

Заметим, что при выполнении шага 1 данного алгоритма для вычисления граничных реализаций могут применяться, например, алгоритм Б.Л. Хо [2] или другие методы вычисления конечномерных реализаций [3, 4].

Пример 1. Рассмотрим интервальную импульсную последовательность для системы с одним входом и одним выходом:

$$\begin{aligned} \mathbf{A}_1 &= [0, 0.21], & \mathbf{A}_2 &= [0.684, 1.357], \\ \mathbf{A}_3 &= [0.599, 1.915], & \mathbf{A}_4 &= [1.072, 3.805]. \end{aligned} \quad (4)$$

Нижняя и верхняя граничные реализации имеют вид:

$$\begin{aligned} \underline{F} &= \begin{pmatrix} 0.9 & 1 \\ 0.81 & 0 \end{pmatrix}, & \underline{G} &= \begin{pmatrix} 0 \\ 0.81 \end{pmatrix}, & \underline{H} &= (1 \ 0), \\ \overline{F} &= \begin{pmatrix} 12.2 & 1 \\ -133.1 & -11 \end{pmatrix}, & \overline{G} &= \begin{pmatrix} 0.1 \\ 0 \end{pmatrix}, & \overline{H} &= (1 \ 0). \end{aligned}$$

Эта пара не удовлетворяет условиям предложения 2. Применив к этим реализациям преобразования подобия

$$T_1 = \begin{pmatrix} 1 & 0.876 \\ 0 & 1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & 6.462 \\ 0 & 1 \end{pmatrix},$$

получим пару реализаций в наблюдаемой канонической форме:

$$\begin{aligned} \hat{\underline{F}} &= \begin{pmatrix} 0 & 1 \\ 0.8 & 0.876 \end{pmatrix}, & \hat{\underline{G}} &= \begin{pmatrix} 0 \\ 0.684 \end{pmatrix}, & \hat{\underline{H}} &= (1 \ 0), \\ \hat{\overline{F}} &= \begin{pmatrix} 0 & 1 \\ 1.043 & 1.25 \end{pmatrix}, & \hat{\overline{G}} &= \begin{pmatrix} 0.21 \\ 1.357 \end{pmatrix}, & \hat{\overline{H}} &= (1 \ 0). \end{aligned}$$

Полученная нами интервальная система

$$\hat{\mathbf{F}} = \begin{pmatrix} [0, 0] & [1, 1] \\ [0.8, 1.043] & [0.876, 1.25] \end{pmatrix}, \quad \hat{\mathbf{G}} = \begin{pmatrix} [0, 0.21] \\ [0.684, 1.357] \end{pmatrix}, \quad \hat{\mathbf{H}} = ([1, 1][0, 0]).$$

отвечает условиям предложения 1 и является интервальной реализацией исходной последовательности матриц (4).

Для полностью неположительных интервальных импульсных последовательностей матриц можно использовать следующий алгоритм.

Алгоритм 2.

1. Образует симметричную относительно нуля последовательность матриц

$$\{[-\bar{A}_1, -\underline{A}_1], [-\bar{A}_2, -\underline{A}_2], \dots\}.$$

2. Для полученной таким образом последовательности матриц (которая является полностью неотрицательной) с помощью алгоритма 1 вычисляем интервальную реализацию $(\mathbf{F}, \mathbf{G}, \mathbf{H})$.
3. Строим реализацию исходной полностью неположительной последовательности, в качестве которой могут быть системы $(\mathbf{F}, -\mathbf{G}, \mathbf{H})$ или $(\mathbf{F}, \mathbf{G}, -\mathbf{H})$.

3. Реализация импульсных последовательностей смешанного типа

В случае “смешанных” интервальных систем, т. е. систем, в которых присутствуют и отрицательные и положительные элементы, можно использовать следующую модификацию метода граничных реализаций.

Разложим исходную импульсную последовательность интервальных матриц (5) на неположительную и неотрицательную последовательности:

$$\{\mathbf{A}_1, \mathbf{A}_2, \dots\} = \{\mathbf{A}_1^-, \mathbf{A}_2^-, \dots\} + \{\mathbf{A}_1^+, \mathbf{A}_2^+, \dots\}, \quad (5)$$

где \mathbf{A}_i^- и \mathbf{A}_i^+ ($i = 1, 2, \dots$) отрицательная и положительная части матрицы \mathbf{A}_i . Последовательность $\{\mathbf{A}_1^-, \mathbf{A}_2^-, \dots\}$ (которая является полностью неположительной), будем называть отрицательной частью последовательности (5). Аналогично последовательность $\{\mathbf{A}_1^+, \mathbf{A}_2^+, \dots\}$ (которая является полностью неотрицательной), будем называть положительной частью последовательности (5).

Имеет место следующий результат.

Предложение 3. Если для положительной и симметричной относительно нуля отрицательной частей последовательности матриц (5) существуют неотрицательные интервальные алгебраические реализации $(\mathbf{F}^+, \mathbf{G}^+, \mathbf{H}^+)$ и $(\mathbf{F}^-, \mathbf{G}^-, \mathbf{H}^-)$ соответственно, то интервальная система $(\mathbf{F}, \mathbf{G}, \mathbf{H})$ с блочными матрицами

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}^- & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^+ \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \mathbf{G}^- \\ \mathbf{G}^+ \end{pmatrix}, \quad \mathbf{H} = (-\mathbf{H}^- \quad \mathbf{H}^+). \quad (6)$$

является интервальной алгебраической реализацией последовательности (5).

Заметим, что в соответствии с теоретико-системной терминологией система (6) является ни чем иным как параллельной композицией систем $(\mathbf{F}^+, \mathbf{G}^+, \mathbf{H}^+)$ и $(\mathbf{F}^-, \mathbf{G}^-, -\mathbf{H}^-)$.

Пример 2. Рассмотрим интервальную импульсную последовательность для системы с одним входом и одним выходом:

$$\begin{aligned} \mathbf{A}_1 &= [-0.44, 0.22], & \mathbf{A}_2 &= [0.412, 1.504], \\ \mathbf{A}_3 &= [-0.2216, 2.2528], & \mathbf{A}_4 &= [-0.07592, 4.42156]. \end{aligned} \quad (7)$$

Образует отрицательную и положительную части этой последовательности:

$$\begin{aligned} \mathbf{A}_1^- &= [-0.44, 0], & \mathbf{A}_2^- &= [0, 0], & \mathbf{A}_3^- &= [-0.2216, 0], & \mathbf{A}_4^- &= [-0.07592, 0]; \\ \mathbf{A}_1^+ &= [0, 0.22], & \mathbf{A}_2^+ &= [0.412, 1.504], & \mathbf{A}_3^+ &= [0, 2.2528], & \mathbf{A}_4^+ &= [0, 4.42156]. \end{aligned}$$

Интервальные реализации, вычисленные с помощью алгоритмов 1 и 2, имеют вид

$$\begin{aligned} \mathbf{F}^- &= \begin{pmatrix} [0, 0] & [0, 1] \\ [0, 0.504] & [0, 0.343] \end{pmatrix}, & \mathbf{G}^- &= \begin{pmatrix} [0, 0.44] \\ [0, 0] \end{pmatrix}, & \mathbf{H}^- &= ([-1, 0] \ [0, 0]); \\ \mathbf{F}^+ &= \begin{pmatrix} [0, 0] & [1, 1] \\ [0, 0.879] & [0, 1.373] \end{pmatrix}, & \mathbf{G}^+ &= \begin{pmatrix} [0, 0.22] \\ [0, 0.412, 1.504] \end{pmatrix}, & \mathbf{H}^+ &= ([1, 1] \ [0, 0]). \end{aligned}$$

Используя предложение 3, получаем искомую реализацию последовательности (7):

$$\begin{aligned} \mathbf{F}^- &= \begin{pmatrix} [0, 0] & [0, 1] & [0, 0] & [0, 0] \\ [0, 0.504] & [0, 0.343] & [0, 0] & [0, 0] \\ [0, 0] & [0, 0] & [0, 0] & [0, 1] \\ [0, 0] & [0, 0] & [0, 0.879] & [0, 1.373] \end{pmatrix}, \\ \mathbf{G}^- &= \begin{pmatrix} [0, 0.44] \\ [0, 0] \\ [0, 0.22] \\ [0, 0.412, 1.504] \end{pmatrix}, & \mathbf{H}^- &= ([-1, 0] \ [0, 0] \ [1, 1] \ [0, 0]). \end{aligned}$$

Список литературы

- [1] Пушков С.Г., Кривошапко С.Ю. О проблеме реализации в пространстве состояний для интервальных динамических систем // Вычислительные технологии. – 2004. – Т. 9, № 1. – С. 75–85.
- [2] Калман Р., Фалб П., Арбиб М. Очерки по математической теории систем. – М.: Мир, 1971.
- [3] Пушков С.Г. О вычислении конечномерной реализации // Кибернетика и системный анализ. – 1991. – № 6. – С. 107–112.
- [4] Пушков С.Г. Конечномерные реализации импульсной характеристики, основанные на псевдообращении ганкелевой матрицы // Известия РАН. Теория и системы управления. – 2002. – № 3. – С. 5–11.

О распознавании экстремальных объектов с помощью интервальных решающих правил

Н.В. Киншт, Н.Н. Петрунько*, Г.Ш. Цициашвили†

Аннотация. Задано конечное множество объектов $\mathbf{A} = \{a_i\}$, $i = 1 \dots n$, характеризуемых определяющими параметрами $\mathbf{X} = \{x_i\}$ и экспериментально полученными интервальными свойствами $\mathbf{Y} = \{y_{ij}\}$, $i = 1, \dots, n$, $j = 1, \dots, m$. На множестве \mathbf{A} заданы подмножества характерных представителей $\{\mathbf{A}_1, \dots, \mathbf{A}_r\}$. Обсуждаются вопросы корректности, эффективности и оптимизации классификации, основанной на имеющихся экспериментальных данных.

Задачи классификации объектов на основе экспериментальных наблюдений давно и многократно привлекали к себе внимание исследователей. Особо оправдывает себя интервальный подход к проблеме при обработке ограниченного числа наблюдений и отсутствии надежд описать процессы, управляющие состоянием объектов, реальными статистическими закономерностями. Один из вариантов практически эффективного и простого интервального решения задачи обработки естественнонаучных наблюдений продемонстрировал с соавторами в своих работах Г.Ш. Цициашвили [1, 2] и др. Представляется, что подобный подход мог бы быть полезным в задачах диагностики (оценки технического состояния) высоковольтного электроэнергетического оборудования, тем более, что в настоящее время идет мощный процесс реструктуризации электроэнергетики, сопровождающийся массовым обследованием энергетических мощностей. Имея это в виду, в настоящей работе делается попытка более глубоко формализовать и (или) развить отмеченный выше подход.

Задано конечное множество объектов $\mathbf{A} = \{a_i\}$, $i = 1, \dots, n$, характеризуемых определяющими параметрами $\mathbf{X} = \{x_i\}$, которые являются основой для классификации объектов. (Для простоты в дальнейшем объекты a_i будем отождествлять с их номерами, т. е. индексами i , что не может вызвать недоразумений). Рассматривается также множество интервальных свойств объектов, задаваемых матрицей $\mathbf{Y} = \{y_{ij}\}$, $i = 1, \dots, n$, $j = 1, \dots, m$.

Считается, что интервальные значения элементов матрицы \mathbf{Y} получены экспериментальным путем и отражают либо тренд соответствующего измеряемого значения во время его наблюдения, либо несколько повторных наблюдений в серии, либо априори интервально заданные ошибки измерений.

На множестве \mathbf{A} в соответствии с некоторой классификацией заданы подмножества характерных представителей $\mathbf{A} = \cup \mathbf{A}_k$, $k = 1, \dots, r$. Им соответствуют непересекающиеся интервальные определяющие параметры:

$$\mathbf{X}_k = [\min(x_i), \max(x_i)], \quad i \in \mathbf{A}_k;$$

*Институт автоматизации и процессов управления ДВО РАН.

†Институт математики ДВО РАН.

$$\mathbf{X}_{k_1} \cap \mathbf{X}_{k_2} = \emptyset, \quad k_1 \neq k_2, \quad k_1, k_2 = 1, \dots, r.$$

Ясно, что каждому множеству \mathbf{A}_k соответствует m -мерный интервал (гиперпараллелепипед) в пространстве свойств объектов $\mathbf{Y}_k = \cup \mathbf{Y}_i, i \in \mathbf{A}_k$. Если ассоциировать множество номеров свойств объекта с множеством точек плоскости на оси абсцисс, а их значения – с ординатами, то удобно говорить также о трубке свойств (значений \mathbf{y}_i) для каждого объекта i и, соответственно, о трубках свойств \mathbf{Y}_k для характерных представителей.

В терминах используемой интервальной модели можно ставить и решать следующие задачи:

- оценивать корректность классификации;
- оценивать эффективность классификации, основанной на заданном конкретном экспериментальном материале;
- эксплуатировать классификацию, т. е. принимать практическое решение о принадлежности нового объекта ($a_i \notin \mathbf{A}$) при получении соответствующих экспериментальных данных y_i ;
- оптимизировать классификацию, т. е. пересматривать и уточнять множества характерных представителей;
- оценивать значимость каждого из свойств при решении задачи классификации;
- исследовать задачу прогнозирования поведения объектов.

Классификацию будем считать корректной, если выполняется условие

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2 \quad \exists i_1 \in \mathbf{A}_{k_1}, \quad i_2 \in \mathbf{A}_{k_2}, \quad \mathbf{y}_{i_1} \setminus (\mathbf{y}_{i_1} \cap \mathbf{Y}_{k_2}) \neq \emptyset. \quad (1)$$

Соотношение (1) постулирует принципиальную возможность классификации. При необходимости можно говорить о слабой либо сильной корректности, опираясь на свойства интервалов \mathbf{Y}_k в целом, а именно, если никакая пара интервалов не находится в отношении полного включения (слабая корректность):

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2, \quad \mathbf{Y}_{k_1} \setminus (\mathbf{Y}_{k_1} \cap \mathbf{Y}_{k_2}) \neq \emptyset, \quad (2)$$

либо не пересекаются вообще (сильная корректность):

$$\forall k_1, k_2 \in \{1, \dots, r\}, \quad k_1 \neq k_2, \quad (\mathbf{Y}_{k_1} \cap \mathbf{Y}_{k_2}) = \emptyset. \quad (3)$$

В дальнейшем, не оговаривая специально, рассматриваемые классификации будем считать корректными (например, в смысле (1)).

Оценим эффективность некоторой классификации. Для каждого $i \in \mathbf{A}$ проверим соблюдение условия принадлежности \mathbf{y}_i трубке \mathbf{Y}_k и построим множества \mathbf{B}_k , соответствующие попаданию \mathbf{y}_i в трубку \mathbf{Y}_k и непопаданию параметра x_i в подмножество \mathbf{X}_k :

$$\forall i \in \mathbf{A} \quad (\mathbf{y}_i \subset \mathbf{Y}_k) \vee (x_i \notin \mathbf{X}_k) \Rightarrow i \in \mathbf{B}_k. \quad (4)$$

В соотношении (4) подразумевается полное попадание объекта \mathbf{y}_i в трубку \mathbf{Y}_k ; однако аналогично свойствам (2, 3) можно различать слабое либо сильное (как в (4)) соответствии объекта характерным подмножествам. Можно говорить также о несоответствии объекта \mathbf{y}_i трубке \mathbf{Y}_k , если хотя бы по одному параметру y_{ij} объект не попадает в трубку свойств \mathbf{Y}_k :

$$\forall i \in \mathbf{A} \quad \exists j (y_{ij} \notin \mathbf{Y}_k) \Rightarrow i \in \mathbf{B}_k. \tag{5}$$

Теперь эффективность классификации (вероятность распознавания объектов на основе имеющихся экспериментальных данных) отражается вектором неудачно классифицированных случаев $\mathbf{b} = (b_1, \dots, b_r) = (|\mathbf{B}_1|, \dots, |\mathbf{B}_r|)$, или же, например, суммарным интегральным показателем $(1 - (b_1 + \dots + b_r)/n)$. Естественно, что для оценки эффективности классификации можно было бы покоординатно сравнивать пересечения гиперпараллелепипедов \mathbf{Y}_k и оценивать их объемы; формально здесь нет проблем, но, имея в виду обработку естественнонаучных или технических экспериментальных данных с не вполне ясными внутренними связями и физически несоизмеримыми параметрами, идеи о гиперобъеме необходимо убедительно аргументировать в каждой конкретной задаче.

При получении дополнительных экспериментальных данных $\mathbf{y}_i, i = n, \dots, p$, соответствующих новым объектам, множество \mathbf{A} расширяется, практическое решение о принадлежности объекта соответствующему подмножеству характерных представителей принимается по принципу, аналогичному (4) или (5):

$$\forall i (i = n, \dots, p) \quad (\mathbf{y}_i \subset \mathbf{Y}_k) \Rightarrow i \in \mathbf{A}_k, \tag{6}$$

причем уверенность в правильности решения обеспечивается с вероятностью $(1 - b_k/n)$.

Обратимся к иллюстративному примеру. Пусть исходные данные заданы табл. 1 (для простоты введена нормировка: $x_i \in [0, 1], y_{ij} \in [0, 1]$).

Таблица 1

A	X	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈
1	0.0000	0.5133	0.5494	0.7391	0.6429	0.8030	0.7500	0.8750	0.7273
2	0.0000	0.6467	0.9699	0.9348	0.9286	0.7121	0.5000	0.9375	0.7273
3	0.0000	0.7067	0.9699	0.7826	0.5714	1.0000	0.7500	0.6250	0.9091
4	0.0760	0.5933	0.7651	0.9130	1.0000	0.6061	0.7500	0.5015	0.5000
5	0.0770	0.5933	0.8735	0.8478	0.5714	0.7727	0.7500	0.6625	0.6818
6	0.0800	0.4667	0.6265	0.8478	0.9286	0.8636	0.7500	0.7500	1.0000
7	0.1538	0.5733	0.7711	0.4783	0.2857	0.3030	0.5000	0.8125	0.3636
8	0.1698	0.5733	0.8434	0.3043	0.0000	0.3030	0.2500	0.5625	1.0000
9	0.2308	0.0000	0.2771	0.5217	0.7143	0.3030	0.5000	0.7500	0.1818
10	0.2564	0.5733	0.3855	1.0000	0.2857	0.2424	0.5000	0.7500	0.1818
11	0.2809	0.2800	0.1566	0.5217	0.2857	0.2727	0.2500	1.0000	0.9091
12	0.3077	0.0400	0.4217	0.7826	0.1429	0.3030	0.2500	0.4375	0.4545
13	0.3468	0.4267	0.2169	1.0000	0.7143	0.0909	0.2500	0.6875	0.7273
14	0.3846	0.1333	0.6265	0.7391	0.7143	0.0000	0.2500	0.3750	0.7727
15	0.4156	0.3067	0.2169	0.3913	0.7143	0.1515	0.5000	0.3125	0.1364
16	0.5611	0.3067	0.0000	0.5217	1.0000	0.7576	0.2500	0.0000	0.6364
17	0.8851	0.8400	0.8072	0.9130	0.5714	0.4545	0.5000	0.3125	0.3182
18	0.9231	0.7067	0.3976	0.6087	1.0000	0.4545	0.7500	0.8125	0.0455
19	1.0000	1.0000	0.9880	0.3043	0.1429	0.2121	1.0000	0.4375	0.0909
20	1.0000	0.2800	0.8795	0.7391	0.2857	0.4848	0.5000	0.7500	0.2727

Пусть также дано, что $A = A_1 \cup A_2 \cup A_3 = \{1, \dots, 6\} \cup \{7, \dots, 15\} \cup \{16, \dots, 20\}$, и, следовательно, $\mathbf{X}_1 = [0, 0.08]$; $\mathbf{X}_2 = [0.1538, 0.4156]$; $\mathbf{X}_3 = [0.5611, 1.0]$. Если предполагается, что определяющий параметр \mathbf{X} может принимать любые значения на интервале $[0, 1]$, учитывая, что $[0, 1] \setminus (\mathbf{X}_1 \cup \mathbf{X}_2 \cup \mathbf{X}_3) \neq \emptyset$, необходимо

Таблица 2

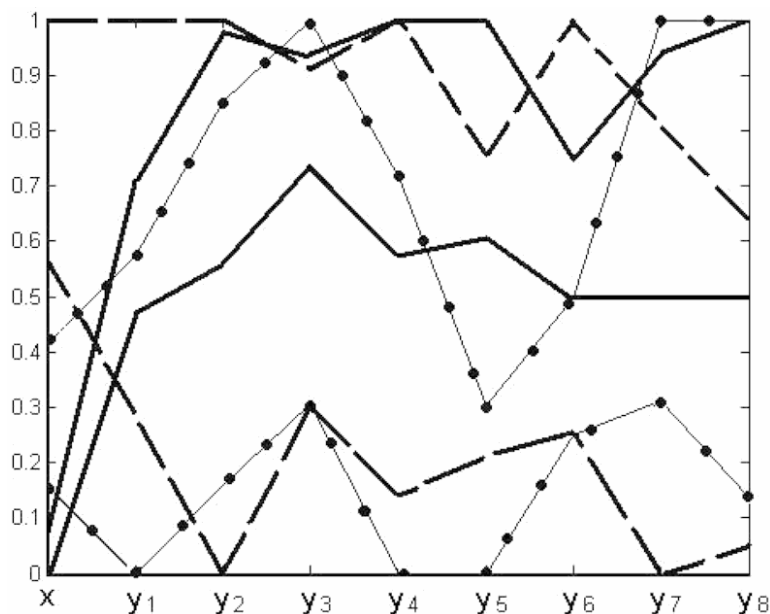
	X_k	Y_{1k}	Y_{2k}	Y_{3k}	Y_{4k}	Y_{5k}	Y_{6k}	Y_{7k}	Y_{8k}
A_1	[0, 0.1)	[0.4667 0.7067]	[0.5494 0.9699]	[0.7391 0.9130]	[0.5714 1.0000]	[0.6061 1.0000]	[0.5000 0.7500]	[0.5015 0.9375]	[0.5000 1.0000]
A_2	[0.1538, 0.4156]	[0.0000 0.5733]	[0.2771 0.8434]	[0.3043 1.0000]	[0.0000 0.7143]	[0.0000 0.3030]	[0.2500 0.5000]	[0.3125 1.0000]	[0.1364 1.0000]
A_3	[0.5611, 1.0]	[0.2800 1.0000]	[0.9880 0.0000]	[0.3043 0.9130]	[0.1429 1.0000]	[0.2121 0.7576]	[0.2500 1.0000]	[0.0000 0.8125]	[0.0909 0.6364]

соответствующим образом дополнить интервалы характерных представителей X_k , например: $X_1 = [0, 0.1)$; $X_2 = [0.1, 0.45)$; $X_3 = [0.45, 1.0]$. Впрочем, последнее представляет интерес лишь при физической интерпретации результатов.

Построив гиперпараллелепипеды Y_1, Y_2, Y_3 , (табл. 2), изобразим их также, для наглядности, в виде трубок характерных представителей (рисунок).

Проверим корректность классификации. Построив гиперпараллелепипеды Y_1, Y_2, Y_3 , выясним, что они находятся в соотношении (2) слабой корректности. Для оценки эффективности классификации построим множество неудачно классифицированных объектов: $B_1 = \emptyset, B_2 = \emptyset, B_3 = \{4, 7\}$. Таким образом, из 20 объектов неудачно классифицируются два, и эффективности классификации можно приписать значение 0.9.

Рассмотрим далее практическое использование построенной классификации. Пусть получены экспериментальные данные $A' = \{21, \dots, 26\}$ (табл. 3).



Представление гиперпараллелепипедов Y_1 (сплошные линии), Y_2 (штриховые линии), Y_3 (линии точек)

Таблица 3

A	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈
21	0.2467	0.6783	0.3913	0.2857	0.2152	0.5000	0.5625	0.4091
22	0.7933	0.1387	0.3043	0.2857	0.2121	0.5000	0.6250	0.1364
23	0.5067	0.9398	0.3448	0.2857	0.5152	0.5000	0.6250	0.3636
24	0.2400	0.6386	0.3592	0.5714	0.3030	0.2500	0.4375	0.8636
25	0.5733	0.7843	0.8783	0.5714	0.8030	0.7500	0.6875	0.5909
26	0.3067	0.3566	0.5217	0.2857	0.3030	0.5000	0.6250	0.1364

Применяя принцип (5) к объектам {21, ..., 26}, получим, что {a₂₁, a₂₄ ∈ A₂, a₂₂, a₂₃ ∈ A₃, a₂₅ ∈ A₁, а объект a₂₆ может быть отнесен как к подмножеству A₂, так и подмножеству A₃. Таким образом, объекты a₂₁, ..., a₂₅ классифицируются с уверенностью 0.9, а объект a₂₅ — с уверенностью 0.45.

Об оптимизации классификации. Можно варьировать разбиение области определения определяющего параметра X на (заданное) число r интервалов определения характерных представителей. Каждому из этих разбиений будет в соответствии с (4) сопутствовать вектор b = (b₁, ..., b_r), оптимизация классификации сводится к минимизации его нормы.

Оценивая значимость каждого из свойств j, j = 1, ..., m, при решении задачи классификации, аналогично (4) либо (5) построим вектор

$$b_j = (b_{1j}, \dots, b_{rj}) = (|B_{1j}|, \dots, |B_{rj}|);$$

$$\forall j = 1 \dots m, \quad i = 1 \dots n, \quad (y_{ij} \in Y_{kj}) \vee (x_i \notin X_k) \Rightarrow i \in B_j. \quad (7)$$

Норма вектора b_j в обратном соотношении ассоциируется со значимостью параметра j.

Наконец, остановимся вкратце на связи данной модели с задачей прогнозирования поведения объектов. Интерпретируя исходную таблицу как классификатор состояний объекта, а дополнительные экспериментальные данные — как изменение параметров некоторого объекта во времени, произведя экстраполяцию каждого из параметров, возможно на этой основе с помощью данной классификации осуществлять прогноз состояния объекта по его определяющему параметру. Так, к примеру, обратившись к табл. 4, представим себе, что данные строк 27 и 28 соответствуют наблюдению поведения некоторого объекта в паре временных точек с интервалом Δt. Нетрудно видеть, что состояние объекта характеризовалось интервалом характерного свойства X₁. Произведя элементарную, единственно возможную (при отсутствии иных дополнительных данных) интерполяцию поведения объекта, получим строку 29 в таблице свойств объекта. Очевидно, что теперь прогнозируется переход его в состояние, соответствующее X₃.

Таблица 4

A''	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈
27	0.65	0.9	0.9	0.9	0.8	0.6	0.6	0.9
28	0.5	0.8	0.8	0.7	0.6	0.65	0.62	0.7
29	0.35	0.7	0.7	0.5	0.4	0.7	0.64	0.5

Выводы. Рассматриваемая формальная модель классификации достаточно прозрачна по постановке и проста в использовании. Представляется полезным более глубоко ее формализовать.

Список литературы

- [1] Цициашвили Г.Ш. и др. Модификация метода интервальной математики применительно к прогнозу экстремальной ледовитости в Охотском море // Вестник ДВО РАН. – 2002. – № 4. – С. 111–114.
- [2] Болотин Е.И., Цициашвили Г.Ш., Голычева И.В. Некоторые аспекты и перспективы факторного прогнозирования эпидемического проявления очагов клещевого энцефалита на основе многомерного анализа временных рядов // Паразитология. – 2002. – Т. 36, № 2. – С.89–95.

Средство для решения полиномиальных ограничений в решателе **Sibcalc**

П.В. Манкевич*

Введение

Многие приложения в науке и технике требуют нахождения всех изолированных решений системы полиномиальных ограничений (СПО) на множестве вещественных чисел. В частности, многие механические системы, включающие в себя почти всех роботов-манипуляторов, описываются множеством состояний, заданным некоторой СПО. Это справедливо для любого механизма, состоящего из жестких конструкций, соединенных шарнирами. Поэтому решение СПО является центральным моментом при анализе движения механических систем и при их проектировании.

В последние два десятка лет появилось несколько методов нахождения всех изолированных корней полиномиальных систем, среди которых можно выделить интервальные методы [1–4], гомотопные методы (homotopic methods) [5], [6] и методы исключения (elimination methods) [7].

Недостатком методов исключения является их большая ресурсоемкость и зависимость от символьных вычислений [7], что делает их непригодными для решения больших полиномиальных систем.

Гомотопные методы непригодны для решения больших систем и систем ограничений большой степени в силу необходимости прохождения путей, число которых есть произведение степеней имеющихся ограничений.

Кроме указанных выше недостатков классических численных методов, их использование часто бывает неприемлемым по нескольким причинам, в том числе и из-за накопления ошибок округления. В частности, может случиться так, что накопленные ошибки округления приводят к тому, что ни одна из цифр полученного результата не будет значимой [1].

С другой стороны, интервальный подход позволяет получить гарантированное приближенное множество решений рассматриваемой проблемы [1]. В силу базовых свойств интервальных операций результирующий интервал каждой конкретной арифметической операции над интервалами содержит все возможные значения. Поэтому по окончании работы интервального метода полученный интервал всегда содержит истинное решение задачи. Если полученный интервал достаточно узкий, то он может рассматриваться как требуемое решение, в противном случае этот интервал может включать несколько решений и для их разделения может потребоваться разбиение результата на подынтервалы.

*Новосибирский государственный университет, mpv@ngs.ru.

Традиционные интервальные подходы, позволяя получить гарантированное решение, обычно работают медленнее своих численных аналогов. По этой причине все чаще создаются алгоритмы, объединяющие интервальные методы с другими подходами. Одной из наиболее успешных таких комбинаций для решения систем полиномиальных ограничений многих переменных является объединение одномерного интервального метода Ньютона с методами распространения ограничений [8].

В данной работе рассматривается реализация этого метода решения полиномиальных ограничений в решателе Sibcalc [9]. Статья организована следующим образом. Во второй главе описываются базовые понятия интервальной математики, дается понятие интервального расширения и приводятся основные определения из области распространения ограничений, необходимые для дальнейшего изложения. В следующей главе излагаются алгоритмы и методы решения поставленной задачи. Четвертая глава содержит описание численных экспериментов, в ней приводятся сравнительные результаты решения полиномиальных систем решателем Sibcalc и с помощью рассматриваемого подхода. В заключение подводятся итоги проделанной работы и описываются направления дальнейших работ.

1. Основные определения и понятия

Рассмотрим множество вещественных чисел расширенное двумя символами положительной и отрицательной бесконечности $R^\infty = R \cup \{-\infty, +\infty\}$. Также рассмотрим конечное подмножество F множества R^∞ , содержащее $\pm\infty$ и 0. На практике множество F соответствует вещественным числам в машинном представлении.

Определение 1. Интервалом $[l, u]$, где $l, u \in F$ называется множество вещественных чисел $\{r \in R \mid l \leq r \leq u\}$.

Множество интервалов обозначается символом $I(R)$. Левый и правый концы интервала $I \in I(R)$ обозначаются как l и u : $I = [l, u]$. Символы \in, \cup, \cap, \subset и т.п. понимаются в обычном теоретико-множественном смысле. Отношение порядка на множестве $I(R)$ определяется следующим образом: I_1 не превосходит I_2 , если $I_1 \subseteq I_2$.

В дальнейшем вещественные числа будут обозначаться строчными латинскими буквами (x, y, z) , интервалы – прописными (X, Y, Z) .

Определение 2. Пусть множество $S \subset R$. Замыкание S – это наименьший интервал $I \in I(R)$, содержащий S : $S \subset I$. Замыкание объединения двух интервалов I_1 и I_2 обозначается как $I = I_1 \nabla I_2$: $I_1 \cup I_2 \subset I$. Наименьший интервал, содержащий вещественное число x , обозначим как $\{x\}$.

Определение 3. Пусть $I_1, I_2 \in I(R)$. Тогда операции сложения, вычитания, умножения и деления задаются следующим образом:

$$I_1 + I_2 = \{x + y \mid x \in I_1, y \in I_2\},$$

$$I_1 - I_2 = \{x - y \mid x \in I_1, y \in I_2\},$$

$$I_1 \times I_2 = \{x \times y \mid x \in I_1, y \in I_2\},$$

$$I_1/I_2 = \{z \mid \exists x \in X, y \in Y : y \neq 0, z = x/y\}.$$

Заметим, что результатом деления интервалов может быть объединение двух интервалов:

$$\{1\}/\{x \mid x \leq 1\} = \{x \mid x < 0\} \cup \{x \mid 1 \leq x\}.$$

Определение 4. Пусть $I_1, \dots, I_n, L_1, \dots, L_m \in I(R)$. Тогда обобщенное пересечение

$$(I_1 \cup \dots \cup I_n) \& (L_1 \cup \dots \cup L_m) = (I_1 \cap L_1) \nabla \dots \nabla (I_n \cap L_m).$$

Результат такой операции может быть точнее, чем операции вида

$$(I_1 \nabla \dots \nabla I_n) \cap (L_1 \nabla \dots \nabla L_m).$$

Например, если $I = [-5, -1], I_1 = [-\infty, -3], I_2 = [3, +\infty]$, то $I \& (I_1 \cup I_2) = [-5, -3]$, в то время как $I \cap (I_1 \nabla I_2) = [-5, -1]$, т. к. $I_1 \nabla I_2 = [-\infty, +\infty]$.

Определение 5. Функция $F : I(R)^n \rightarrow I(R)$ – это интервальное расширение функции $f : R^n \rightarrow R$, если и только если

$$\forall I_1, \dots, I_n \in I(R) : r_1 \in I_1, \dots, r_n \in I_n \Rightarrow f(r_1, \dots, r_n) \in F(I_1, \dots, I_n).$$

Интервальное отношение $C : I(R)^n \rightarrow \{\text{true}, \text{false}\}$ – это интервальное расширение отношения $c : R^n \rightarrow \{\text{true}, \text{false}\}$, если и только если

$$\forall I_1, \dots, I_n \in I(R) : \{\exists r_1 \in I_1, \dots, \exists r_n \in I_n c(r_1, \dots, r_n)\} \Rightarrow C(I_1, \dots, I_n).$$

Например, интервальная функция \oplus , определенная как

$$[a, b] \oplus [c, d] = [a + c, b + d]$$

есть интервальное расширение функции сложения вещественных чисел.

Естественное интервальное расширение вещественной функции f обозначается \hat{f} , отношения c обозначается \hat{c} . Данное интервальное расширение сохраняет ограничения в том виде, в котором их задает пользователь.

Другим интервальным расширением является дистрибутивное интервальное расширение, которое приводит ограничения к другой форме записи. Основным преимуществом этого расширения является то, что оно позволяет устанавливать совместность по брусам (см. п. 2), используя интервальный метод Ньютона для функций одной переменной.

Определение 6. Ограничение c находится в дистрибутивной форме, если оно представлено в виде суммы

$$m_1 + \dots + m_k = 0,$$

где каждое слагаемое $m_i = qx_1^{e_1} \dots x_n^{e_n}$, $q \in Q$, $e_i \in N$.

Определение 7. Распределенное интервальное расширение функции f – это естественное интервальное расширение ее дистрибутивной формы. Оно обозначается как \hat{f} .

Еще одно интервальное расширение основано на разложении функции f в ряд Тейлора. Данное интервальное расширение имеет много важных свойств и исследовано многими авторами [8]. Это расширение предполагает наличие у функции непрерывных частных производных на некотором интервале.

Определение 8. Пусть ограничение c имеет вид $f(x) = 0$, f имеет непрерывные частные производные. $I = \langle I_1, \dots, I_n \rangle$ – брус и $m_i = \text{mid}(I_i)$ – центры соответствующих интервалов. Интервальное расширение Тейлора ограничения c на бресе I задается как интервальное ограничение вида

$$\hat{f}(m_1, \dots, m_n) + \sum_{i=1}^n \frac{\partial \hat{f}}{\partial x_i}(I_1, \dots, I_n)(X_i - m_i) = 0.$$

Одним из преимуществ полиномиальных систем ограничений является их простое автоматическое дифференцирование. Это значит, что при усечении каждого из измерений исходного бруса можно воспользоваться простым и эффективным интервальным методом Ньютона.

В алгоритме, рассматриваемом в данной работе, используется комбинация интервальных методов и методов удовлетворения ограничений, основным понятием которых является понятие совместности. Например, при сужении каждого отдельного измерения используется одномерный метод Ньютона с применением условия совместности по брусам, которое является приближением условия совместности по дугам [10]. Нестрого говоря, ограничение $c(x_1, \dots, x_n)$ совместно по дугам, если не одна из D_i областей определения переменных x_1, \dots, x_n не может быть больше сужена, используя проекции ограничения c . Ниже будут даны необходимые определения, используемые в методах распространения ограничений.

Определение 9. Проекция ограничения c – это пара ограничения c и индекса

$$i : \langle c, i \rangle, \quad i = 1, \dots, n.$$

Определение 10. Проекция ограничения $\langle c, i \rangle$ совместна по дугам в области $D_1 \times \dots \times D_n$, если и только если

$$D_i = D_i \cap \{r_i \mid \exists r_1 \in D_1, \dots, \exists r_{i-1} \in D_{i-1}, \exists r_{i+1} \in D_{i+1}, \dots, \exists r_n \in D_n : c(r_1, \dots, r_n)\}.$$

Определение 11. Ограничение c совместно по дугам в области $D_1 \times \dots \times D_n$, если каждая его проекция совместна в этой области. Система ограничений совместна в области $D_1 \times \dots \times D_n$, если каждое ограничение совместно в данной области.

По некоторой начальной области $D_1 \times \dots \times D_n$ алгоритм совместности по дугам вычисляет наибольшую подобласть начальной области, на которой все ограничения совместны. Такая подобласть всегда существует и единственна. Алгоритм совместности по дугам эффективен для решения дискретных комбинаторных задач. Тем не менее, условие совместности по дугам не может быть вычислено в общем случае на вещественных областях с полиномиальными ограничениями. [8]

Условие совместность по брусам является некоторой аппроксимацией совместности по дугам. Для начала введем понятие интервальной проекции ограничения.

Определение 12. Назовем интервальной проекцией ограничения $\langle C, i \rangle$ пару интервального ограничения C и индекса i ($i = 1, \dots, n$).

Определение 13. Интервальная проекция ограничения совместна по брусам на брусе $I = \langle I_1, \dots, I_n \rangle$, если и только если

$$C(I_1, \dots, I_{i-1}, [l, l^+], I_{i+1}, \dots, I_n) \ \& \ C(I_1, \dots, I_{i-1}, [u^-, u], I_{i+1}, \dots, I_n),$$

где $l = \text{left}(I_i)$, $u = \text{right}(I_i)$. Интервальное ограничение совместно по брусам на брусе I , если каждая из его интервальных проекций совместно по брусам на I . Система интервальных ограничений совместна по брусам на брусе I , если каждое интервальное ограничение совместно по брусам на I .

Определение 14. Пусть $\langle C, i \rangle$ интервальная проекция ограничения и $\langle I_1, \dots, I_n \rangle$ – брусок, тогда оператор сужения

$$\text{Red}(\langle C, i \rangle, \langle I_1, \dots, I_n \rangle) = \langle I_1, \dots, I_{i-1}, I, I_{i+1}, \dots, I_n \rangle, I \subset I_i$$

есть наибольшее подмножество, такое что интервальная проекция ограничения $\langle C, i \rangle$ совместна по брусам на брусе $\langle I_1, \dots, I_{i-1}, I, I_{i+1}, \dots, I_n \rangle$.

2. Описание подхода

В данной работе рассматривается задача нахождения всех изолированных решений системы полиномиальных ограничений

$$p_1(x_1, \dots, x_n) = 0, \quad \dots, \quad p_n(x_1, \dots, x_n) = 0$$

в интервальном брусе $I = \langle I_1^0, \dots, I_n^0 \rangle$. Эта задача может быть рассмотрена в следующей форме: найти все бруски I_1, \dots, I_k , удовлетворяющие системе вида

$$P_1(I) \ni 0, \quad \dots, \quad P_n(I) \ni 0,$$

где P_i – интервальные расширения полиномов p_i , $i = 1, \dots, n$.

Большинство интервальных методов представляют собой алгоритмы глобального поиска, состоящего из двух основных частей: усечения исходного бруса и разбиение его на два подбруса для дальнейшего их усечения.

Процедура Solve

Вход: S – система полиномиальных ограничений (интервальных расширений), I_0 – начальный брус.

Выход: Набор всех брусов удовлетворяющих системе.

Начало:

1. $I = \text{Prune}(S, I_0)$! сузить интервал
2. **if** (I – непустой) **then**
3. { **if** (I – достаточно мал) вернуть I }
4. **else**
 $\langle I_1, I_2 \rangle = \text{Branch}(I)$! разбить брус на два подбруса
 вернуть $\text{Solve}(S, I_1) \cup \text{Solve}(S, I_2)$.

Конец.

Предложение 1. Пусть C интервальное расширение ограничения c , $\langle C, i \rangle$ – интервальная проекция ограничения, $\langle I_1, \dots, I_n \rangle$ брус u

$$\langle I_1, \dots, I_{i-1}, I, I_{i+1}, \dots, I_n \rangle = \text{Red}(\langle C, i \rangle, \langle I_1, \dots, I_n \rangle).$$

Тогда

$$\{r_1 \in I_1, \dots, r_n \in I_n \ \& \ c(r_1, \dots, r_n)\} \Rightarrow r_i \in I.$$

Доказательство. Предположим, $r_1 \in I_1, \dots, r_n \in I_n$ и $r_i \notin I$. Тогда либо $r_i < l$, либо $r_i > u$, где $I = [l, u]$. Если $r_i < l$, то

$$C(I_1, \dots, I_{i-1}, \{r_i\}, I_{i+1}, \dots, I_n)$$

по определению интервального расширения и

$$C(I_1, \dots, I_{i-1}, [u^-, u], I_{i+1}, \dots, I_n)$$

по предположению. Таким образом, C удовлетворяет условию совместности по брусам на брусе $\langle I_1, \dots, I_{i-1}, I \nabla \{r_i\}, I_{i+1}, \dots, I_n \rangle$, что противоречит тому, что I задает наибольший интервал из I_i , такой что $\langle C, i \rangle$ удовлетворяет условию совместности по брусам на брусе $\langle I_1, \dots, I_{i-1}, I, I_{i+1}, \dots, I_n \rangle$. Случай $r_i > u$ доказывается аналогично. \square

Теперь переходим к следующему шагу – описанию алгоритма усечения. Данный алгоритм состоит из применения оператора сужения на каждую проекцию ограничения до тех пор, пока усечение возможно.

Процедура усечения Prune

Вход: S – система полиномиальных ограничений (интервальных расширений), I_0 – начальный брус.

Выход: Наибольший брус $I \subset I_0 : S(I)$.

Начало:

1. **Repeat**
2. $I_p = I$
3. $\text{Box_prune}(\{\langle \hat{c}, i \rangle \mid c \in S, i = 1, \dots, n\}, I)$
4. **Until** ($I = I_p$).

Конец.

Procedure Box_prune

Вход: P – множество интервальных проекций ограничений, I – входной брус.

Выход: I – усеченный брус.

- Начало:*
1. Repeat
 2. $I_p = I$
 3. $I = \bigcap \{\text{Red}(p, I) \mid p \in P\}$
 4. Until ($I = I_p$)

Конец.

Для реализации оператора сужения воспользуемся следующим утверждением.

Предложение 2. *Каждый ноль функции f на интервале I лежит также в интервале $N(F, F', I)$.*

Здесь $N(F, F', I) = I \nabla \left[\{\text{mid}(I)\} - \frac{F(\{\text{mid}(I)\})}{F'(\{\text{mid}(I)\})} \right]$. Интервальная функция одной переменной F получается из функции f , \hat{f} и интервальной проекции ограничения $\langle C, i \rangle$ следующим образом: все переменные X_k интервальной функции \hat{f} , за исключением X_i , заменяются соответствующими интервалами I_k сужаемого бокса $I = \langle I_1, \dots, I_n \rangle$. Аналогично, функция F' получается из естественного интервального расширения производной функции f .

Таким образом, оператор сужения бруса для естественного интервального расширения можно задать в виде

$$\text{Red}(\langle C, i \rangle, I) = N(F, F', I_i).$$

Для каждого из описанных выше интервальных расширений (см. определение 5) можно определить свой оператор сужения. Известно, что интервальный метод Ньютона для естественного интервального расширения хорошо зарекомендовал себя на широких интервалах. В то же время, интервальное расширение Тейлора хорошо работает на узких интервалах. Для реализации оператора сужения для распределенного интервального расширения можно воспользоваться особенностями этого интервального расширения. Так же как и при использовании естественного интервального расширения, рассмотрим функцию

$$F(X) = \tilde{f}(I_1, \dots, I_{i-1}, X, I_{i+1}, \dots, I_n).$$

Далее необходимо отыскать нули этой функции. Основная идея, используемая здесь, состоит в ограничении данной интервальной функции F двумя вещественными функциями одной вещественной переменной f_l и f_u , определенными следующим образом:

$$f_l(x) = \text{left}(F(\{x\})), \quad f_u(x) = \text{right}(F(\{x\})).$$

Заметим теперь, что усечение i -го измерения можно проводить при помощи поиска самого левого нуля интервального расширения F_l функции f_l и самого правого нуля интервального расширения F_r функции f_r .

Основным преимуществом распределенного интервального расширения является то, что функции f_l и f_r могут быть легко найдены. Функция $F(X)$ имеет вид

$$F(X) = I_1 X^{n_1} + \dots + I_p X^{n_p},$$

Тогда функция f_l определяется следующим образом:

$$f_l = \text{low}(I_1, x, n_1) + \dots + \text{low}(I_p, x, n_p),$$

где

$$\text{low}(I, x, n) = \begin{cases} \text{left}(I)x^n, & x \geq 0 \text{ или } n - \text{четное}, \\ \text{right}(I)x^n, & \text{иначе.} \end{cases}$$

Функция f_u определена как

$$f_u(x) = \text{high}(I_1, x, n_1) + \dots + \text{high}(I_p, x, n_p),$$

где

$$\text{high}(I, x, n) = \begin{cases} \text{right}(I)x^n, & x \geq 0 \text{ или } n - \text{четное}, \\ \text{left}(I)x^n, & \text{иначе.} \end{cases}$$

Далее, аналогично оператору сужения для естественного интервального расширения, можно применять интервальный метод Ньютона на данных функциях для поиска их нулей.

3. Численные эксперименты

Для решения полиномиальных систем был реализован метод PR-Solver (Polynomial Relations) с использованием описанного в данной работе подхода. Этот метод сравнивался с решателем Sibcalc [9], в котором при решении данных задач использовался метод интервального распространения ограничений и бисекция. Сравнение проводилось на компьютере с процессором Celeron 4-2000 с 640 Мб оперативной памяти и операционной системой WindowsXP. В таблице представлены результаты тестирования метода на нескольких полиномиальных системах. Приведено время решения каждой системы. Погрешность вычислений задавалась равной 10^{-8} . Времена указаны в миллисекундах.

Анализируя приведенные результаты, можно сказать, что времена в целом сопоставимы, хотя Sibcalc имеет небольшое преимущество. Это можно объяснить тем, что в PR-Solver реализованы еще не все методы, позволяющие существенно улучшать эффективность алгоритма. Также отметим, что методы

Тип	PR-Solver	Sibcalc
$\{y = x, y = -x\}$	0.0	0.0
$\{y = x^2, y = x\}$	0.02	0.0
$\{y = x^2, y = -x^2 + 1\}$	0.04	0.015
$\{x^2 + y^2 = 4, y = 0\}$	0.02	0.01
$\{x^2 + y^2 = 4, x^2 - 4x + y^2 = 0\}$	0.03	0.02
$\{y = x^2, x^2 + y^2 = 4\}$	0.06	0.045

по техническим причинам не сравнивались на задачах с большим числом переменных и ограничений, в которых рассматриваемый подход имеет наибольший выигрыш.

Заключение

В работе продемонстрирована возможность применения метода деления и усечения наряду с классическим методом Ньютона для решения полиномиальных систем. С его помощью можно решать как полиномиальные уравнения одной переменной, так и системы полиномиальных уравнений.

На данном этапе разработки реализовано только естественное интервальное расширение и ведется работа над распределенным интервальным расширением. Следующим этапом для увеличения эффективности системы будет введено интервальное расширение Тейлора. Отдельным направлением будущих работ будет комплексное сравнение реализованного подхода с другими существующими пакетами решения систем полиномиальных уравнений.

Список литературы

- [1] Hansen E. *Global Optimization Using Interval Analysis*. – New York: Marcel Dekker, 1992.
- [2] Hansen E.R., Greenberg R.I. An interval Newton method // *Appl. Math. Comput.* – 1983. – № 12. – P.89–98.
- [3] Hansen E.R., Sengupta S. Bounding Solutions of Systems of Equations Using Interval Analysis // *BIT*. – 1981. – № 21. – P. 203–211.
- [4] Neumayer A. *Interval Methods for Systems of Equations / PHI Series in Computer Science*. – Cambridge: University Press, 1990.
- [5] Morgan A.P. *Solving Polynomial Systems Using Continuation for Scientific and Engineering Problems*. – Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [6] Verschelde J., Verlinden P., Cools R. Homotopies Exploiting Newton Polytopes For Solving Sparse Polynomial Systems // *SIAM J. Num. Anal.* – 1994. – Vol. 31, № 3. – P. 915–930.
- [7] Tien-Yien Li. Solving Polynomial Systems by Polyhedral Homotopies // *Taiwanese J. Mathem.* – 1999. – Vol. 3, № 3. – P. 251–279.
- [8] Van Henteryck P., McAllester D., Kapur D. Solving Polynomial Systems Using a Branch and Prune Approach // *SIAM J. Num. Anal.* – 1997. – Vol. 34, № 2.
- [9] Kleymenov A., Petunin D., Semenov A., Vazhev I. A Model of Cooperative Solvers for Computational Problems // *Proc. 4th Intern. Conf. PPAM 2001, Poland, September / LNCS*. – Vol. 2328. – P. 797–802.
- [10] Tsang E. *Foundations of Constraint Satisfaction*. – Essex: Academic Press, 1993.

Ансамбли систем дифференциальных уравнений с интервальными данными

А.Н. Роголев*

Аннотация. Построение двусторонних и интервальных оценок множеств решений дифференциальных уравнений с неточно заданными (интервальными) данными сильно усложняется экспоненциальным ростом этих границ. Этот рост, называемый “wrapping effect”, требует разработки методов оценки, которые бы устраняли его влияние. В статье рассматриваются свойства решений ансамблей систем дифференциальных уравнений, являющихся множествами всех возможных систем, соответствующих изменениям данных задач в пределах содержащих их интервалов. Принципиальным моментом является, в частности, сохранение характеристик фазового портрета ансамбля при возмущении закона движения. Методы, основанные на аппроксимации оператора сдвига вдоль траектории, строят гарантированные границы множеств решений, асимптотически точные относительно каждой координаты. Приводятся примеры вычисленных границ множеств решений для практических задач.

Введение

В статье рассматривается задача нахождения гарантированных границ и исследования свойств множеств решений динамических систем с интервальными (неточно заданными) данными

$$\frac{dy}{dt} = f(t, y), \quad y(t^0) = y^0 \in Y^0. \quad (1)$$

Здесь Y^0 обозначает интервал в \mathbb{R}^n или прямоугольный параллелепипед. Подобные задачи связаны с принципиальной неточностью определения данных для многих классов задач, описываемых системами обыкновенных дифференциальных уравнений. Сложности, которые возникают при реализации различных методов решения задач этого класса хорошо известны. Для большинства методов наблюдается экспоненциальный рост границ решений, названный [29, 26, 28, 33, 34, 20, 21] “wrapping effect” (эффект от “обвертывания” множества точных решений в оболочку, упаковку). Хотя выбор этого термина означает попытку пояснить причины этого явления, очевидно, что полный исчерпывающий анализ таких причин отсутствует. Гарантированные методы, основанные на аппроксимации оператора сдвига вдоль траектории и символьных формулах этих приближенных решений [8–18] позволяют преодолеть влияние эффекта экспоненциального роста границ. Если правая часть системы дифференциальных уравнений удовлетворяет условию дифференцируемости по начальным данным, то это условие можно использовать для нахождения и исследования свойств пучка решений с начальными условиями, близкими к невозмущенным.

*Институт вычислительного моделирования СО РАН, Красноярск, ran@krsk.info.

1. Нестатистические ансамбли систем дифференциальных систем обыкновенных дифференциальных уравнений с интервальными данными и множества возмущений

Рассмотрим при $t = t^0$ в области $\Omega \subset \mathbb{R}^n$, зависящей от t^0 ансамбль Гиббса системы (1). Под ансамблем Гиббса [2] будет пониматься множество идентичных систем вида (1) с одинаковыми правыми частями, отличающимися друг лишь начальными значениями. Ансамблю системы (1) будет соответствовать в области Σ при каждом t ансамбль изображающих точек. Каждая из изображающих точек $y^0 \in \Sigma_{t_0}$, двигаясь по траекториям системы (1) согласно преобразованию $y(t) = Sy^0$, поместится соответственно в точку $y(t) \in \Sigma_t$. Таким образом, оператор сдвига $S(t, t_0)$ переводит область Σ_{t_0} в область $\Sigma_t \subset \mathbb{R}^n$ за время $t - t_0$. Траектории ансамбля систем (1) равномерно распределены в ансамбле, возможно по другому вероятностному закону, что будет менять постановку задачи и свойства решений этой системы. В дальнейшем в пределах этой статьи такая постановка не будет рассматриваться.

Предположим, что у нас есть зависимость, задающая плотность попадания траекторий в некоторую область фазового пространства Σ_{t_0} , обозначим эту плотность $D(t, y)$. Движение ансамбля изображающих точек $y^0 \in \Sigma_{t_0}$, перемещающихся вдоль интегральных кривых системы (1) согласно преобразованию $y = S_{t, t_0} y^0$, будем трактовать как течение некоторой непрерывной среды в \mathbb{R}^n с плотностью $D(t, y)$ в точке y в момент времени t . Так как изменение плотности $D(t, y)$ в каждом элементе фазового пространства \mathbb{R}^n обусловлено различием по величине потоков, изображающих движение точек через его границу, то функция $D(t, y)$ удовлетворяет уравнению неразрывности

$$\frac{\partial D(t, y)}{\partial t} + \sum_{i=1}^n \frac{\partial F_i(t, y)}{\partial y_i} D(t, y) = 0 \quad (2)$$

и называется обобщенным уравнением Лиувилля относительно плотности $D(t, y)$ изображающих точек системы (1). Уравнение (2) совпадает с дифференциальным уравнением последнего множителя Якоби системы (1).

Каждый из двусторонних методов имеет свои свойства, выделяющие его среди других методов, однако принципиальный момент для них – это свойство верхних и нижних границ являться решениями мажорирующих уравнений [19, 25, 37, 3].

Интервальные методы выполняются пошагово так, чтобы обеспечить на каждом шаге по времени включение всех точных решений в полученные интервальные векторы, на основе свойств монотонности относительно включения интервальных операций. Наиболее распространены методы, построенные на основе разложений по формуле Тейлора. Причинами этого является возможность единым образом оценивать величину глобальной ошибки методов рядов Тейлора, находить их области значений, и составлять для них простые программы, вычисляющие значения производных высоких порядков от правой части методами автоматического дифференцирования. Обзоры существующих интервальных методов даны в работах [30, 24, 35] (Р. Мур, Дж. Корлисс, А. Ноймайер).

Большинство методов, так или иначе основанных на разложении решений по формуле Тейлора, состоят из двух этапов исполнения, которые можно условно назвать этапами предиктор-корректор.

Первый этап (предиктор) заключается в определении априорного включения \mathbf{y}_j и величины шага h_j таких, чтобы решение задачи (1) $y(t)$ существовало на интервале $[t_j, t_{j+1}]$ при любых начальных условиях $y_{t_j} \in \mathbf{Y}_j$ и при этих условиях выполнялось включение

$$\bigcup_{y(t_j) \in \mathbf{Y}_j} (y(t, t_j, y(t_j))) \subseteq \mathbf{Y}_j$$

для всех $t \in [t_j, t_{j+1}]$.

Второй этап (корректор) состоит в уточнении полученного интервала (уменьшении его ширины), используя $\mathbf{y}(t)$.

Описываемая группа интервальных методов, основанных на использовании рядов Тейлора, находит включение $\forall t \in [t_j, t_{j+1}]$

$$y(t, t_0, \mathbf{Y}_0) \subseteq \mathbf{Y}(t) = \mathbf{Y}_j + \sum_{i=1}^{k-1} \mathbf{Y}_{j_i} (t - t_j)^i + \mathbf{Y}_{j_k}.$$

Этот подход, был начат в работах Мура [29] и является основным до настоящего времени. Были внесены существенные модификации в интервальные методы, использующие разложения в ряды Тейлора, но они не затрагивали принципиальных моментов.

Большую известность среди интервальных методов решения ОДУ приобрели методы математика Р. Лонера [27]. Это объясняется, разумеется, в первую очередь тем, что выполнена программная реализация этого метода, с помощью которой просчитаны интервальные оценки для решений некоторых систем ОДУ. Метод Лонера представляет модификацию интервального метода рядов Тейлора. Основное содержание изменений состоит в том, что на каждом шаге изменение ориентации интервального вектора (прямоугольного параллелепипеда) относительно координатных осей выполняется путем домножения на специально подобранную матрицу вращения.

Методы, основанные на интервальных формах разложений по формуле Тейлора, используют следующую технологическую цепочку доказательства включения решений. Производится запись исходного дифференциального уравнения с помощью оператора Пикара–Линделефа, при определенных условиях на коэффициенты поставленной задачи справедливы теоремы о сжимающем отображении [29].

Двусторонние и интервальные методы производят включение точных решений, поскольку в операторе Пикара–Линделефа правые части представляют многозначные (интервальные) функции

Собственно включение точного решения задачи в получаемый интервальный вектор обосновывается тем, что границы этого вектора являются фактически решениями уравнений сравнения. Это следует из включения

$$y^0(t) + \int_{t_0}^t (\mathbf{y}^{\nu-1}(s)) ds \in \int_{t_0}^t \mathbf{F}(t)(s) ds.$$

Интервальные методы решения дифференциальных уравнений и двусторонние методы основаны на мажорировании исходных уравнений с помощью уравнений сравнения и применении теорем о дифференциальных неравенствах. Теоремы о дифференциальных неравенствах [37] исследованы достаточно подробно и применяются в задачах устойчивости. Однако нам потребуются специальные формы дифференциальных неравенств.

Теорема. Пусть $u(t)$, $\bar{u}(t)$, $\underline{u}(t)$ – непрерывные на интервале $[a, b]$ и дифференцируемые на открытом интервале (a, b) векторные функции, функция $f(t, u)$ непрерывна и удовлетворяет условию Липшица

$$\|f(t, u) - f(t, v)\| = \sup_i |f_i(t, u) - f_i(t, v)| \leq L\|u - v\|, \quad L > 0, \quad t \in [a, b].$$

Если имеют место неравенства:

$$\begin{aligned} \underline{u}(a) < u(a) < \bar{u}(a); \\ \frac{d\underline{u}}{dt} - \underline{f}(t, \underline{u}, \bar{u}) \leq \frac{du}{dt} - f(t, u) \leq \frac{d\bar{u}}{dt} - \bar{f}(t, \underline{u}, \bar{u}), \end{aligned} \quad (3)$$

тогда выполняется включение

$$\underline{u}(t) < u(t) < \bar{u}(t)$$

для всех $t \in [a, b]$.

Как следствие из этой теоремы формулируется следующий важный результат. Не нарушая общности, полагаем, что в (3) выполнены равенства, и каждый член (3) равен нулю на интервале $(a, b]$. Тогда граничные функции этих неравенств являются решениями систем дифференциальных уравнений

$$\frac{d\bar{u}}{dt} = \bar{f}(t, \bar{u}, \underline{u}), \quad \frac{d\underline{u}}{dt} = \underline{f}(t, \bar{u}, \underline{u}, t), \quad (4)$$

относительно $2n$ независимых переменных. Если векторная функция $f(t, u_1, \dots, u_n)$ является монотонно неубывающей функцией относительно компонент u_α , где $\alpha \in S_i$, $S_i \subseteq S = \{1, 2, \dots, n\}$, и $\alpha \neq i$, и монотонно невозрастающей функцией относительно компонент u_β , $\beta \in \setminus S_i$, $\beta \neq i$, то справедливо равенство

$$\bar{f}_i(t, \bar{u}, \underline{u}) = f_i(t, \bar{u}_\alpha, \underline{u}_\beta, u_i), \quad \underline{f}_i(t, \bar{u}, \underline{u}) = f_i(t, \underline{u}_\alpha, \bar{u}_\beta, u_i).$$

Определение 1. Будем называть дифференциальную систему сравнения (3), включающей систему ОДУ (4), если каждая функция, задающая границу множества решений, является решением системы (или одного из уравнений, входящих в систему) (4). Переход от исходной системы (1) к системе (4) является возмущением структуры дифференциальных уравнений, определим его как структурное возмущение. Соответственно, размерность пространства R^{2n} , в котором определен вектор решений системы (4) будет называться размерностью включения.

Таким образом, использование при построении двусторонних и интервальных оценок решений систем ОДУ дифференциальных неравенств, приводит к качественным изменениям (возмущению структуры) системы обыкновенных дифференциальных уравнений.

Главная особенность классических понятий устойчивости состоит в том, что они относятся к конкретной системе и поведению ее траекторий в окрестности точки равновесия (притяжения или отталкивания). Совершенно другого подхода требует анализ поведения семейства траекторий, включая особые точки, сепаратрисы и предельные циклы, взаимное расположение которых определяет структуру семейства, этот анализ возникает при рассмотрении всех систем, “близких” к стандартной системе (1).

Согласно определению, введенного Понтрягиным и Андроновым [1], система (1) называется структурно устойчивой (грубой), если топологический характер траекторий всех близких к ней систем такой же, как у системы (1). Определенные математические трудности связаны с уточнением понятий “близкая система”, а также с конкретизацией смысла, который подразумевается, когда говорят о том, что траектория эквивалентна, или топологически подобна другой траектории. Но основная идея остается ясной, достаточно малые изменения структурно устойчивой системы должны приводить к соответственно малым изменениям в динамике ее поведения. Можно выделить два вида описания множества систем: ансамбли систем дифференциальных уравнений, имеющие одну структуру системы (вид правой части системы) и отличающиеся лишь значениями параметров, и возмущенные системы, для которых варьируются все характеристики системы – правые части, единственным ограничением является близость значений правых частей или начальных данных в некоторой метрике. Для возмущенных систем имеет место значительное изменение как видов возмущенных систем, так и изменения динамики решений. Хаотические движения не встречались у двумерных систем, и появились при переходе трехмерным. Единственно возможными установившимися движениями в двумерных системах являются устойчивые состояния равновесия и периодические движения. Однако при периодических силовых или параметрических возмущениях на осциллятор или ротатор возникают хаотические движения.

Траектории хаотических динамических систем по определению экспоненциально неустойчивы и крайне чувствительны по отношению к ошибкам округления.

Дальнейшие исследования [6] подтвердили этот результат более строго, связав экспоненциальную расходимость траекторий с энтропией Колмогорова и характеристическими показателями Ляпунова. Вычисление этих характеристик стало стандартным методом проверки на стохастичность.

Наличие хаотического аттрактора в фазовом пространстве системы обыкновенных дифференциальных уравнений служит причиной сложного поведения траекторий системы в его окрестности.

2. Символьные методы и формулы приближенных решений

Гарантированные методы, основанные на аппроксимации оператора сдвига вдоль траектории систем ОДУ, могут быть охарактеризованы следующим образом. Для всех подобных методов общим является введение определенных функциональных связей между областями всех решений системы ОДУ в начальный момент времени t^0 и при $t > t^0$ в пределах конечного интервала времени. На свойства гарантированных методов влияет характер ограничений, налагаемых на параметры, например, начальные данные системы и характер изменения во времени области всех решений.

Гарантированные методы оценки множеств решений с интервальными данными основаны на символьном представлении формул приближенных решений. Эти методы позволяют строить границы множеств решений, точно отслеживающие поведение множества всех точных решений, а также устранить влияние так называемого “wrapping” эффекта, присущего практически всем двусторонним и интервальным методам. Основные этапы исполнения гарантированных методов построения границ это:

- 1) представление приближенных решений как символьных формул (аналитических выражений), зависящих от векторов начальных данных, также являющихся символьными переменными,
- 2) покомпонентное оценивание множеств значений этих формул;
- 3) нахождение области всех возможных значений по этой формуле;
- 4) гарантированное оценивание глобальной ошибки всех значений приближенных решений.

Определение 2. Символьная формула (аналитическое выражение) – запись имен переменных и совокупности действий, которые нужно проделать в определенном порядке над значениями этих переменных, чтобы получить значение функции.

В силу этого символьный метод (аналитический метод) – запись метода преобразования символьной информации (символьных формул) на языке математического анализа.

В дальнейшем при записи символьных формул, аппроксимирующих оператор сдвига вдоль траектории, допускается включение в них числовых констант, с отложенным выполнением арифметических действий над ними.

В этом символьный метод отличается от численного алгоритма, основанного на исполнении конечной последовательности действий над конечным множеством чисел. Чтобы строить символьные формулы, аппроксимирующие оператор сдвига вдоль траектории ОДУ и позволяющие получать достаточно точные включения множеств решений (например, интервальные расширения), необходимо получить формулу, хорошо приближающую точное решение и выполнить алгоритм преобразования этой формулы. Строгое определение гарантированных методов решения дифференциальных уравнений с интервальными данными и анализ содержится [14].

В качестве символьной формулы \mathcal{F}_i метода сдвига вдоль траектории можно выбрать формулы, числовые оригиналы которых соответствуют приближенным методам решения систем обыкновенных дифференциальных уравнений (например, линейным одношаговым и многошаговым методам, коллокационным методам). Принципиальное отличие заключается в способе реализации этих методов, то есть получении числовых оригиналов символьных формул \mathcal{F}_i . Символьные формулы приближенных решений могут строиться с использованием разностных методов (одношаговых или многошаговых), коллокационных методов или иных схем, вычисляющих приближенные решения. В процессе конструирования символьной формулы требуется строить экономичные символьные формулы, т. е. такие последовательности имен переменных и действий, чтобы хранение формул в памяти машины и обработка проводилась за минимальное время и с разумными затратами памяти. Кроме того, методы должны обеспечивать близость множества значений, полученных по символьной формуле и множества значений точных решений. Указанные свойства означают, что само построение формул и их исполнение существенно изменяются по сравнению с численными алгоритмами, изучаемыми в разделе численного решения систем ОДУ.

Многие свойства численных методов, неизменно присущие практически всем численным методам решения ОДУ, такие как устойчивость, асимптотическая устойчивость и иные понятия либо переформулируются, либо включаются в другие характеристики гарантированных методов.

3. Графики множеств решений систем ОДУ с интервальными данными

В этом пункте представлены границы множеств решений нескольких систем дифференциальных уравнений, построены графики этих границ и проведено сравнение с экземплярами точных решений, принадлежащими совокупности всех точных решений, соответствующих изменению данных исходной задачи. Сравнение проводится на основе тех тестовых примеров, для которых либо возможно найти формулы точных решений, либо известны двусторонние оценки. Как правило, системы ОДУ были подобраны так, чтобы при такой постановке задач возникали определенные сложности с нахождением двусторонних и интервальных оценок. Тем не менее, найденные с помощью описанных в статье методов, числовые значения гарантированных границ решений успешно решают эти проблемы и являются включениями множеств решений, достаточно близкими к этим множествам. Для хорошей иллюстрации точности полученных вычисленных границ множеств решений приводятся графики, на которых изображены точные решения сходной системы ОДУ с начальными данными, выбранными внутри интервала начальных данных. Эти решения названы экземплярами множества точных решений, поскольку, вообще говоря, это множество заполнено континуумом экземпляров. Кроме того, экземпляры, начальные данные которых взяты с граничных точек интервала начальных данных (вершин параллелепипеда в \mathbb{R}^n), не могут считаться граничными линиями множества решений, через некоторый интервал времени экземпляры решений не будут мажорировать указанное множество.

Все полученные результаты приведены в п. 3. В статье использовались все возможности для сравнения гарантированных границ множеств решений и интервальных или двусторонних оценок, содержащихся в [22, 23, 32, 34, 35]. К сожалению, в тексты большинства опубликованных работ включена лишь величина ошибки решений систем ОДУ, посчитанных с начальными интервальными данными нулевой ширины. Фактический анализ и сравнение численных результатов можно провести, пользуясь только двумя программами решения систем ОДУ с интервальными данными, [27, 23]. Большее практическое значение имеет программа AWA [27], открытая для доступа и использования, результаты полученные авторами и разработчиками системы COSY, представлены не столь широко. По всей видимости, система COSY всегда являлась программной реализацией интервальных методов рядов Тейлора и ориентировалась на задачи небесной механики. Система AWA и была использована для попыток оценивания множеств решений систем, описанных в данной статье. Расчеты проводились для компьютера Intel(R) Celeron(R) CPU 1.8 ГГц, 256 Мб.

3.1. Сравнение гарантированных границ множеств решений с результатами, полученными двусторонними методами статьи [3]

Построим гарантированные границы множеств решений системы ОДУ

$$\frac{dy_1}{dt} = y_2, \quad \frac{dy_2}{dt} = y_1^2 + t \quad (5)$$

с начальными данными

$$y_1(0) \in [1 - \varepsilon, 1 + \varepsilon], \quad y_2(0) \in [-\varepsilon, \varepsilon], \quad (6)$$

где ε – заданное число. Эти границы, полученные в этой статье на основе аппроксимации оператора сдвига вдоль траектории, можно сравнить с границами точного решения, приведенными в [3].

Легко заметить, что в табл. 1 включены данные, полученные при ширине начального интервала двусторонней оценки, равной величине ошибки округления. Поскольку в статье [3] не оговаривалась точность, с которой проводились расчеты, можно принять машинную ошибку величиной порядка $\approx 10^{-8}, 10^{-9}$. При значении аргумента $t = 0.2$ ширина двусторонней оценки становится величиной порядка 10^{-5} , а при $t = 0.5$ ширина двусторонней оценки имеет порядок 10^{-4} . Таким образом на интервале от 0 до 0.5 точность оценки ухудшилась в 10^4 раз. Такое поведение двусторонних оценок объясняется влиянием так называемого wrapping эффекта [29, 26, 28, 23, 35]. Если бы нахождение двусторонних оценок было продолжено для больших значений аргумента, то двусторонние границы бы отклонились одна от другой на слишком большую величину, влияние wrapping эффекта стало бы катастрофическим.

В табл. 2 гарантированные границы посчитаны на интервале $t \in [0, 2]$. Начальный интервал выбирался равным $2 \cdot 10^{-3}$. При этом поведение интервала получаемых границ в точности соответствует росту границ множества точных решений. Найденные с помощью систем аналитических вычислений экземпляры точных решений свидетельствуют об этом. Время расчетов составило 2.81 с.

Производились попытки найти интервальные оценки решения система уравнений (5) также с помощью программы AWA [27]. Однако работа программы

Таблица 1. Двусторонние границы точного решения системы (5), полученные в [3]

t_n	\underline{y}_n	\bar{y}_n	\underline{z}_n	\bar{z}_n
0	1	1	0	0
0.1	1.005175	1.005175	0.105342	0.105342
0.2	1.021466	1.021479	0.226865	0.286886
0.3	1.050213	1.050240	0.3548852	0.354928
0.4	1.093018	1.093057	0.504463	0.504528
0.5	1.151795	1.151861	0.675130	0.675249

Таблица 2. Гарантированные границы множества решений системы (5), полученные на основе аппроксимации оператора сдвига вдоль траектории

t	Y_1	Y_2
0	[0.999, 1.001]	[-0.001, 0.001]
0.1	[1.00400978, 1.006230297]	[0.104081181, 0.106503486]
0.2	[1.02010864, 1.022584437]	[0.2212785040, 0.224176786]
0.3	[1.04862563, 1.05142768]	[0.35302938, 0.35647435]
0.4	[1.09115534, 1.09433262]	[0.50224070, 0.50632519]
0.5	[1.14961440, 1.153236682]	[0.67249330, 0.67633951]
1.0	[1.78397272, 1.791563517]	[2.058842205, 2.07138601]
2.0	[10.26966353, 10.316006031]	[27.22678736, 27.57170364]

была прервана, так как величина шага стала из-за катастрофического роста ширины получаемых интервалов. Время работы составило 53.17 с.

3.2. Пример гарантированного оценивания множеств решений нелинейной системы ОДУ (из справочника Камке [5])

Будем оценивать множество решений нелинейной системы ОДУ,

$$\frac{dy_1}{dt} = -y_1 y_2^2 + y_1 + y_2, \quad \frac{dy_2}{dt} = y_1^2 y_2 - y_1 - y_2, \quad \frac{dy_3}{dt} = y_2^2 - y_1^2, \quad (7)$$

приведенное в [5] система (9.25). Интегральными кривыми этой системы являются пересечения поверхностей

$$y_1^2 + y_2^2 + \ln y_3^2 = c_1, \quad y_3(y_1 y_2 - 1) = c_2. \quad (8)$$

Для этой системы ОДУ мы рассматриваем задачу с начальными данными, заданными как интервалы $y_1^0 \in [1, 1.5]$, $y_2^0 \in [2, 2.4]$, $y_3^0 \in [0.5, 1]$, на интервале $t \in [0, 2]$.

Картина поведения экземпляров интегральных кривых, начинающихся в граничных точках интервалов начальных данных, помещены в сравнении на графике: рис. 1 и расположены на них слева. Для сравнения с ними на этом же графике построены гарантированные границы множеств решений рассматриваемой задачи, размещенные справа. Из сравнения численных значений границ множеств решений и экземпляров точных решений, а также их графиков можно заключить о точности включения множеств решений для этой системы ОДУ. Время расчетов составило 2.9 с.

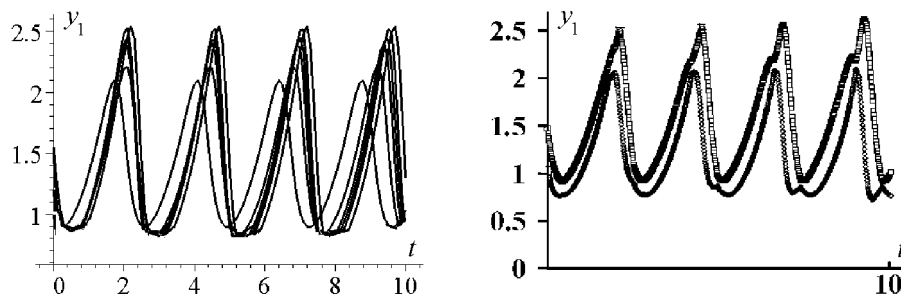


Рис. 1. Проекция экземпляров точных решений и гарантированных границ множества решений системы (7) на плоскость $t - y_1$: \square — гарантированная верхняя граница; \circ — гарантированная нижняя граница

Была произведена попытка найти интервальные границы система ОДУ (7) с помощью программы AWA [27]. Выборка числовых значений [27] свидетельствует о том, что программа прервала процесс вычислений при $t = 0.1110688557$, так как шаг стал равен 0:

```
t = 1.110688557E-001
[-4.02857339E+008, 4.02857344996879E+008] 8.06E+008 8.1E+008
[-4.02857339289200E+008, 4.02857344942050E+008] 8.06E+008 8.1E+008
[-2.50854422226110E+001, 2.69234195274611E+001] 5.21E+001 5.2E+001
h = 0.000000000E+000
```

3.3. Поведение множества решений локально неустойчивой динамической системы, приведенной в статье [35]

Динамическая система

$$\frac{dy_1}{dt} = 9.9y_1 - 7.6y_2 + 7.6, \quad \frac{dy_2}{dt} = 12.6y_1 - 9.9y_2 + 9.9 + y_1^3/7.6 \quad (9)$$

имеет две устойчивые неподвижные точки $y_1 = \pm 1.5, y_2 = 1 \pm 297/152$ и неустойчивую неподвижную точку $y_1 = 0, y_2 = 1$. Система (9) консервативна, т.е. на решениях этой системы сохраняется постоянным значение $c = 1/2(y_1^2 - 2.25)^2 + (9.9y_1 - 7.6y_2 + 7.6)^2$. Все траектории этой системы являются периодическими, и остаются ограниченными на всем интервале определения. Определитель якобиана этой системы $\det F'(y) = 3y_1^2 - 2.25$. Если $|y_1| < 1/2\sqrt{3} \approx 0.866$, то собственные значения якобиана вещественные числа, кратности 1. Тогда решения этой системы локально неустойчивы в полосе, определяемой неравенством $|y_1| \leq \frac{\sqrt{3}}{2}$, каждая орбита системы, пересекающая прямую $y_2 = 1$ проходит через две неустойчивых точки, отличающиеся на период.

Автор статьи [35] предложил сравнить как ведут себя различные программные системы при определении интервальных оценок для системы (9). Выбрав значения $y_1(0) = 2^{-16}$ и $y_2 = 1$, Neumaier приводит следующие результаты, полученные системой AWA [27]: система прервала свою работу из-за числового

переполнения при $t = 33.9101420298123$, выполнив 1566 шагов по времени через 2.6 с. Последней полученной интервальной оценкой, указанной в [35], является интервальный вектор

$$\begin{aligned} y_1(t) &= [-196.199244777380, 196.2262981935131], \\ y_2(t) &= [6039.01251175114, 6041.0421329249]. \end{aligned}$$

Далее следует аварийное прерывание программы в силу переполнения.

$$\begin{aligned} y_1(t) &= [-196.199244777380, 196.2262981935131], \\ y_2(t) &= [6039.01251175114, 6041.0421329249]. \end{aligned}$$

Результаты расчетов, полученные в этой статье выглядят так. Поскольку большой интерес представляют задачи, в которых данные изначально заданы как интервальные величины, то расчеты, выполненные описанным в этой статье гарантированным методом проводились при начальных данных

$$y_1(t) \in [-2^{-16}, 2^{-16}], \quad y_2(t) \in [1, 1].$$

Интервальный вектор включения решения с приемлемой точностью был вычислен при $t = 30.04$ и равен

$$\begin{aligned} y_1(t) &= [-3.125 \cdot 10^{-2}, 3.125 \cdot 10^{-2}], \\ y_2(t) &= [0.93750183902, 1.0325037562]. \end{aligned}$$

Границы множеств решений систем [35], проведены вплоть до точки $t = 37.2$, последнее полученное значение равно

$$\begin{aligned} y_1(t) &= [-128.1276601342, 128.1090127387], \\ y_2(t) &= [-32768.223821563, -16384.6015723511, 6041.0421329249]. \end{aligned}$$

Время расчетов составило 4.3 с.

Изменив начальные данные $y_1(0) = [9.99 \cdot 10^{-17}, 1.01 \cdot 10^{-15}]$ и $y_2 = [1, 1]$, получаем гарантированные границы для существенно большего интервала времени, вычисления продолжались вплоть до $t = 50$. Графики, изображенные на рис. 2, получены для $t \in [0, 323]$. Время расчетов составило 38 с.

Легко заметить, что траектории решений совершают сильные колебания вдоль осей y_1, y_2 , но это не приводит к огромному росту их значений и прерыванию работы программы. В силу этого результат является достаточно приемлемым и сохраняется хорошая возможность добиться высокой точности границ для задач подобного класса. Так, например, полученные границы при $t = 36.055$ равны следующим величинам

$$\begin{aligned} y_1^{3605} &\in [-6.014648438, 7.043945313], \\ y_2^{3605} &\in [-0.401733398, 0.714660645], \\ y_1^{3606} &\in [-5.312988281, 6.482421875], \\ y_2^{3606} &\in [0.092086792, 0.326522827]. \end{aligned}$$

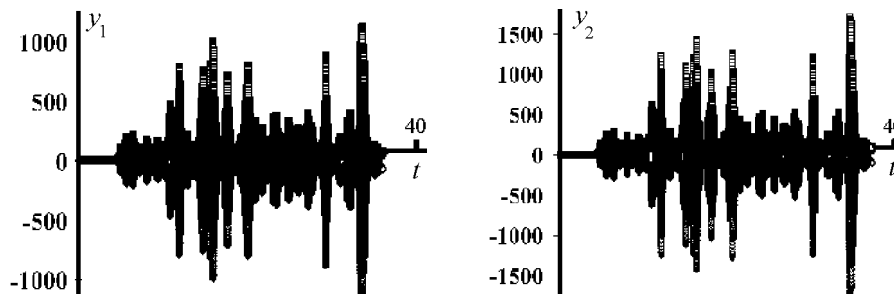


Рис. 2. Проекция гарантированных границ множества точных решений системы (9) на плоскости $t - y_1, t - y_2$ в интервале $0 < t \leq 40$

3.4. Пример гарантированного оценивания жестких систем ОДУ с неточно заданными начальными данными

Для проверки результатов выбрана система уравнений, описывающая химическую реакцию с участием восьми реагентов. Эта система была предложена Шефером [36] для объяснения роста и дифференциации растительной ткани при высоких уровнях светового облучения независимо от фотосинтеза. Готтвальд предложил использовать ее в качестве тестового примера. Соответствующие уравнения имеют вид:

$$\begin{aligned}
 \frac{dy_1}{dt} &= -1.71y_1 + 0.43y_2 + 8.32y_3 + 0.0007, \\
 \frac{dy_2}{dt} &= 1.71y_1 - 8.75y_2, \\
 \frac{dy_3}{dt} &= -10.03y_3 + 0.43y_4 + 0.035y_5, \\
 \frac{dy_4}{dt} &= 8.32y_2 + 1.71y_3 - 1.12y_4, \\
 \frac{dy_5}{dt} &= -1.745y_5 + 0.43y_6 + 0.45y_7, \\
 \frac{dy_6}{dt} &= -280y_6y_8 + 0.69y_4 + 1.71y_5 - 0.43y_6 + 0.69y_7, \\
 \frac{dy_7}{dt} &= 280y_6y_8 - 1.81y_7, \\
 \frac{dy_8}{dt} &= -y_7
 \end{aligned} \tag{10}$$

с интервальными начальными данными $y_1(0) = [1 - \delta, 1 + \delta], y_2(0) = y_3(0) = y_4(0) = y_5(0) = y_6(0) = y_7(0) = [-\delta, \delta], y_8(0) = [0.0057 - \delta, 0.0057 + \delta], t_{\text{out}} = 321.8122$.

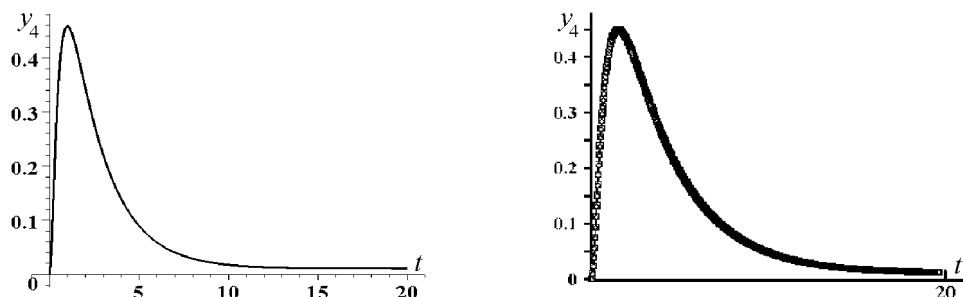


Рис. 3. Сравнение проекции на плоскости $t-y_4$ гарантированных границ множества точных решений системы (10) и экземпляров приближенных решений в интервале $0 < t \leq 20$

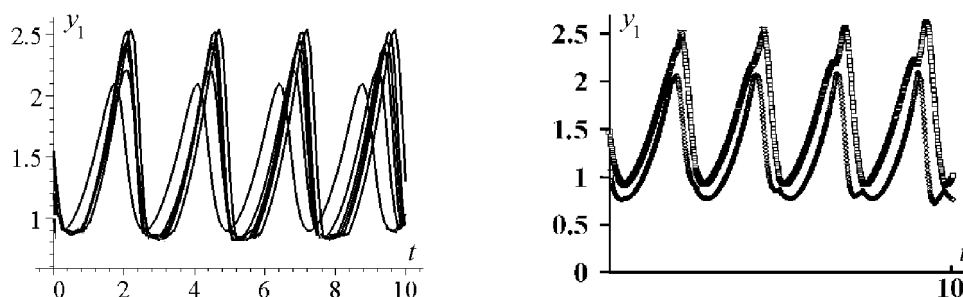


Рис. 4. Сравнение проекции на плоскость $t-y_1$ гарантированных границ множества точных решений системы (10) и экземпляров приближенных решений в интервале $0 < t \leq 321.8122$

Список литературы

- [1] Андронов А.А., Леонтович Е.А. Динамические системы первой степени негрубости на плоскости // Мат. сб. – 1965. – Т. 68, № 3. – С. 328–372.
- [2] Гиббс Дж. Б. Основные принципы статистической механики. – М.: Гостехиздат, 1944. – 215 с.
- [3] Девятко В.Н. О двустороннем приближении при численном интегрировании обыкновенных дифференциальных уравнений // Журн. вычисл. математики и мат. физики. – 1963. – Т. 3, № 2. – С. 254–265.
- [4] Калмыков С.А., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. – Новосибирск: Наука, 1986. – 221 с.
- [5] Камке Э. Справочник по обыкновенным дифференциальным уравнениям. – М.: Наука, 1971. – 576 с.
- [6] Каток А.Б., Хассельблат Б. Введение в современную теорию динамических систем. – М.: Факториал, 1999. – 768 с.
- [7] Лихтенберг А., Либерман М. Регулярная и стохастическая динамика. – Москва: Мир, 1984. – 528 с.
- [8] Новиков В.А., Рогалев А.Н. Построение сходящихся верхних и нижних оценок решений систем обыкновенных дифференциальных уравнений // Журн. вычисл. математики и мат. физики. – 1993. – Т. 33, № 2. – С. 219–231.

- [9] Рогалев А.Н., Шокин Ю.И. Исследование и оценка решений обыкновенных дифференциальных уравнений интервально-символьными методами // Вычислительные технологии. – 1999. – Т. 4, № 4. – С. 51–76.
- [10] Рогалев А.Н. Поле аппроксимаций обыкновенных дифференциальных уравнений с интервальными коэффициентами. Вопросы математического анализа. – 2001. – Вып. 4. – С. 152–170.
- [11] Рогалев А.Н. Динамика множеств решений обыкновенных дифференциальных уравнений и интервальные решения. Вопросы математического анализа. – 2001. – Вып. 5.
- [12] Рогалев А.Н. Исследование практической устойчивости при постоянно действующих возмущениях // Вычислительные технологии. – 2002. – Т. 7, ч. 5 (совм. вып. по мат. Междунар. конф. ВТММ-2002). – С. 148–150.
- [13] Рогалев А.Н. Задачи практической (интервальной) устойчивости с заданной областью предельных отклонений // Тр. пятой междунар. конф. памяти акад. А.П. Ершова: Междунар. совещ. по интервальной математике и методам расширения ограничений. – Новосибирск, 2003. – С. 90–100.
- [14] Рогалев А.Н. Гарантированные методы решения систем обыкновенных дифференциальных уравнений на основе преобразования символьных формул // Вычислительные технологии. – 2003. – Т. 8, № 5. – С. 102–116.
- [15] Рогалев А.Н. Поведение динамических систем при экстремальных возмущениях // Вычислительные технологии, 2003. – Т. 8 (совм. вып. по мат. Междунар. конф. ВТММ-2003). – С. 68–77.
- [16] Рогалев А.Н. Гарантированные оценки безопасного функционирования технических и электроэнергетических систем // Тр. Всерос. конф. с междунар. участием “Современные методы математического моделирования природных и антропогенных катастроф”. – Красноярск, 2003. – Т. 3. – С. 42–48.
- [17] Рогалев А.Н. Включение множеств решений дифференциальных уравнений и гарантированные оценки глобальной ошибки // Вычислительные технологии. – 2003. – Т. 8, № 6. – С. 80–94.
- [18] Рогалев А.Н. Границы множеств решений систем обыкновенных дифференциальных уравнений с интервальными начальными данными // Вычислительные технологии. – 2004. – Т. 9, № 1. – С. 86–93.
- [19] Чаплыгин С.А. Новый метод приближенного интегрирования дифференциальных уравнений // Чаплыгин С.А. Избранные труды. Механика жидкости и газа. Математика. – М.: Наука, 1976. – С. 307–362.
- [20] Anguelov R., Markov S. Wrapping effect and wrapping function // Reliable Computing. – 1998. – Vol. 4, № 4. – P. 311–330.
- [21] Anguelov R. Wrapping effect of the initial value problems for ODE: Applications // Reliable Computing. – 1999. – Vol. 5, № 2. – P. 143–164.
- [22] Berz M., Makino K., Hoefkens J. Verified Integration of Dynamics in the Solar System. Nonlinear Analysis: Theory, Methods and Applications. – 2001. – Vol. 47. – P. 179–190.
- [23] Berz M., Hoefkens J., Makino K. Controlling the wrapping effect in the solution of ODEs for asteroids // Reliable Computing. – 2003. – Vol. 8, № 1. – P. 21–41.
- [24] Corliss G. Survey of interval algorithms for ordinary differential equations. // Applied Mathematics and Computations. – 1989. – Vol. 31. – P. 112–120.

-
- [25] Filippov A.F. Ellipsoidal estimates for a solution of a system of differential equation // Interval Computations. – 1992. – № 2. – P. 6–17.
- [26] Lohner R. Praktikum einschließung bei Differentialgleichungen. / Wissenschaftliches Rechnen mit Ergebnisverification (Hrsg. Kulish U.), 1989. – S. 209–223.
- [27] Lohner R.J. Program AWA. – 1994. – <ftp://iamk4515.mathematik.uni-karlsruhe.de/pub/awa>.
- [28] Lohner R. On the ubiquity of the wrapping effect in the computation of error bound // SCAN-2000, Interval-2000. GAM/IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics, Sept. 18–22, 2000. – Universität Karlsruhe, Institut für Angewandte Mathematic (Germany). – P. 36.
- [29] Moore R.E. Interval analysis. – Englewood Cliffs: Prentice Hall, 1966.
- [30] Moore R.E. A survey of interval methods for differential equations // Proc. of 23-rd conf. on decision and control. – Las Vegas University, 1984. – P. 1529–1535.
- [31] Nedialkov N., Jackson K., Corliss G. Validated solutions of initial value problems for ordinary differential equations / Tech. rep. – 1998. – <http://www.cs.toronto.ed/N/reports.html>.
- [32] Nedialkov N., Jackson K. An interval Hermite-Obreschkoff Method for computing rigorous bounds on the solution of an initial value problem for an ordinary differential equation // Reliable Computing. – 1999. – Vol. 5, № 5. – P. 289–310.
- [33] Nedialkov N., Jackson K. The Design and Implementations of an Object-Oriented Validated ODE Solver / Tech. rep. – 2001. – <http://www.cs.toronto.ed/N/reports.html>.
- [34] Nedialkov N., Jackson K. A New Perspective on the Wrapping Effect // Interval Methods for Initial Value Problems for Ordinary Differential Equations. Perspectives on Enclosure Methods / Ed. by Kulisch U., Lohner R., Facius A. – Springer-Verlag, 2001. – P. 219–264 (<http://www.cs.toronto.ed/N/reports.htm/ned.scan00.ps.Z>).
- [35] Neumaier A. Taylor forms – Use and limits // Reliable computing. – 2003. – Vol. 9, № 1. – P. 43–79.
- [36] Schäfer E. A new approach to explain the “High Irradiance Responses” of photomorphogenesis on the basis of phytochrome // J. of Math. Biology. – 1975. – Vol. 2. – P. 41–56.
- [37] Walter W. Differential and integral inequalities. – Berlin: Springer, 1970. – 195 p.

Приближение решения задачи Гурса интервальными многочленами Тейлора

Е.Г. Филиппенко, В.С. Зюзин*

Аннотация. Предлагается методика нахождения коэффициентов рядов Тейлора для решения уравнения в частных производных с начальными данными, примененная для нахождения интервального решения задачи Гурса.

Аналогичная методика нахождения коэффициентов рядов Тейлора разработана для решения системы обыкновенных дифференциальных уравнений (ОДУ) [3].

Находим многочлен до n степени и оцениваем остаточный член $(n + 1)$ -порядка. Используя интервальные методы получаем гарантированные, апостериорные оценки погрешности, включая погрешности вычисления приближенного решения задачи Гурса.

В статье предлагается метод приближения решения задачи Гурса многочленами Тейлора с интервальной оценкой остаточного члена.

В общем виде задача Гурса имеет вид [1, 2]:

$$u_{xy} = f(x, y, u, u_x, u_y), \quad (1)$$

$$u(x_0, y) = \varphi_1(y), \quad y_0 \leq y \leq y_0 + h_2, \quad (2)$$

$$u(x, y_0) = \varphi_2(x), \quad x_0 \leq x \leq x_0 + h_1,$$

причем $\varphi_1(y_0) = \varphi_2(x_0)$.

Предполагается, что правая часть (1) имеет рациональный вид и непрерывные частные производные любого порядка, также предполагается дифференцируемость функций $\varphi_1(y)$ и $\varphi_2(x)$ достаточное число раз.

Решение задачи (1), (2) находим с помощью ряда Тейлора

$$u(x, y) = \sum_{i=0}^n \sum_{j=0}^n (u)_{ij} (x - x_0)^i (y - y_0)^j + R_{n+1}, \quad (3)$$

где остаточный член

$$R_{n+1} := \sum_{i=0}^{n+1} (u(x_0 + \tau_1 x, y_0 + \tau_2 y))_{i, n+1-i} (x - x_0)^i (y - y_0)^{n+1-i}, \quad (4)$$

где $x_0 \leq \tau_1 \leq x_0 + h_1$, $y_0 \leq \tau_2 \leq y_0 + h_2$,

$$(u)_{ij} := \frac{1}{i!j!} \frac{\partial^{i+j} u(x, y)}{\partial x^i \partial y^j} \Big|_{\substack{x=x_0 \\ y=y_0}}. \quad (5)$$

Для нахождения коэффициентов ряда Тейлора (3) выведены следующие рекуррентные формулы. Пусть $u(x, y)$ и $u_1(x, y)$ имеют частные производные любого порядка, тогда имеют место

* Саратовский государственный университет им. Н.Г. Чернышевского.

$$\begin{aligned}
(u_x)_{ij} &= (i+1)(u)_{i+1,j}, & (u_y)_{ij} &= (j+1)(u)_{i,j+1}, \\
(u_{xy})_{ij} &= (i+1)(j+1)(u)_{i+1,j+1}, \\
(u_{xx})_{ij} &= (i+1)(i+2)(u)_{i+2,j}, & (u_{yy})_{ij} &= (j+1)(j+2)(u)_{i,j+2}, \\
(u \pm u_1)_{ij} &= (u)_{ij} \pm (u_1)_{ij}, & (uu_1)_{ij} &= \sum_{k=0}^i \sum_{l=0}^j (u)_{kl}(u_1)_{i-k,j-l}.
\end{aligned} \tag{6}$$

Применение выведенных формул избавляет от необходимости последовательного дифференцирования правой части (1) (аналогичная методика нахождения коэффициентов рядов Тейлора для обыкновенных дифференциальных уравнений приведена в [3]).

Если нам известна двусторонняя оценка решения $(u \pm u_1)_{ij} = (u)_{ij} \pm (u_1)_{ij}$ задачи (1), (2) на области $[x_0, x_0 + h_1] \times [y_0, y_0 + h_2]$ ($\underline{u} \leq u(x, y) \leq \bar{u}$ или $u(x, y) \in [\underline{u}, \bar{u}] =: [u]$), то используя рекуррентные формулы (6) можно получить двустороннюю оценку остаточного члена (5).

В случае, когда нам неизвестна двусторонняя оценка решения $(u \pm u_1)_{ij} = (u)_{ij} \pm (u_1)_{ij}$ задачи (1), (2) в рассматриваемой области, ниже приводится вариант нахождения указанной двусторонней оценки.

Для этого аналогично как и в [2] сводим задачу (1),(2) к системе интегральных уравнений, обозначив

$$\begin{aligned}
v &:= u_x, & w &:= u_y, \\
\begin{cases} v(x, y) = T_v v := \varphi'_2(x) + \int_{y_0}^y f(x, y, u, v, w) dy, \\ w(x, y) = T_w w := \varphi'_1(y) + \int_{x_0}^x f(x, y, u, v, w) dx, \\ u(x, y) = T_u u := \varphi_2(x) + \int_{y_0}^y w dy. \end{cases} & & & (7)
\end{aligned}$$

Пусть

$$z := (v, w, u), \quad Tz := (T_v v, T_w w, T_u u) \Rightarrow z = Tz. \tag{8}$$

Предположим, что функция $f(x, y, v, w, u)$ по переменным u, v, w удовлетворяет условию Липшица

$$|f(x, y, u, v, w) - f(x, y, u_0, v_0, w_0)| \leq L \{|u - u_0| + |v - v_0| + |w - w_0|\}, \tag{9}$$

где $L > 0$ — некоторая константа.

Введем расстояние, предложенное в работе [4]

$$\rho(z, z_0) := \max_{\substack{x_0 \leq x \leq x_0 + h_1 \\ y_0 \leq y \leq y_0 + h_2}} \left\{ e^{-\alpha(x-x_0) - \beta(y-y_0)} [|u - u_0| + |v - v_0| + |w - w_0|] \right\}, \tag{10}$$

где $\alpha, \beta > 0$.

С помощью расстояния (9) показываем, что оператор T есть оператор сжатия, при соответствующем выборе $\alpha, \beta > 0$, т. е.

$$\rho(Tz, Tz_0) \leq q\rho(z, z_0), \quad \text{где } 0 < q < 1.$$

Следовательно, оператор T с расстоянием (9) обеспечивает сжатое отображение на заданной области.

Приведенная ниже теорема способствует установлению двусторонней оценки решения $(u \pm u_1)_{ij} = (u)_{ij} \pm (u_1)_{ij}$ задачи (1), (2).

Теорема. Пусть интервальный вектор $[z^0] := ([v^0], [w^0], [u^0])$ такой, что

$$[z^1] := T[z^0] \subseteq [z^0]. \quad (11)$$

Тогда задача (1), (2) на области $[x_0, x_0 + h_1] \times [y_0, y_0 + h_2]$ имеет единственное решение $u(x, y) \in [z^1] \subseteq [z^0]$.

Доказательство. Выше было установлено, что оператор T есть оператор сжатия и согласно (10) отображает замкнутую область $[z^0] := [\underline{z}^0, \bar{z}^0]$ в себя. Согласно следствию теоремы Банаха о неподвижной точке, оператор T имеет единственную неподвижную точку $z^* := (v^*, w^*, u^*)$ в $[z^1] := ([v^1], [w^1], [u^1])$. Отсюда следует утверждение теоремы. \square

Установленную оценку $[u^1] := [\underline{u}^1, \bar{u}^1]$ можно использовать для оценки остаточного члена (4), используя рекуррентные формулы (6).

Предлагаемый метод позволяет находить приближения решения $(u \pm u_1)_{ij} = (u)_{ij} \pm (u_1)_{ij}$ задачи (1), (2) в аналитическом виде, а именно, в виде многочлена Тейлора с интервальной оценкой остаточного члена. При вычислении на Pascal-XSC, получаем гарантированные двусторонние приближения точного решения с учетом погрешности округлений.

Замечания

Требование рациональности правой части уравнения (1) не ограничивает общности. В основном большой класс уравнений путем замены переменных можно свести к рациональному виду.

В формулах (6) не рассмотрен случай операции деления. Если у функции рационального вида встречается операция деления, то ее можно избежать опять путем замены переменных.

Обращаем внимание на то, что не обязательно знать точно константу Липшица $-L$ в (8). Единственное требование, чтобы $L \neq \infty$.

Список литературы

- [1] Люстерник Л.А., Соболев В.И. Элементы функционального анализа. – М.: Наука, 1965.
- [2] Соболев С.Л. Уравнения математической физики. – М.: Гос. изд-во технико-теоретической литературы, 1954.
- [3] Moore R.E. Interval analysis. – Englewood Cliffs. N.J.: Prentice-Hall, 1966.
- [4] Walter W. Gewöhnliche Differentialgleichungen. – Springer-Verlag, 1976.

Моделирование систем автоматического управления с интервальной неопределенностью параметров (новые версии и возможности комплекса программ “АСИАС”)

Н.А. Хлебалин, Д.С. Пятых*

1. Теоретические основы комплекса программ “АСИАС” (анализ и синтез интервальных автоматических систем)

Как правило, проектировщик АС имеет дело с математическими моделями (ММ) в виде систем дифференциальных уравнений (ДУ) или структурных схем с передаточными функциями (ПФ), причем коэффициенты ДУ и ПФ являются вектор-функциями некоторых параметров (объекта и регулятора). Так как коэффициенты ММ являются функциями от параметров, то такие ММ называют *параметрическим*. При этом обычно не делается дополнительных пояснений – в каких областях изменяются физические параметры (каким множествам они принадлежат). В комплексе программ (КП) “АСИАС” рассматриваются автоматические системы с интервальной неопределенностью параметров (АСИНП), у которых начальные условия и возмущения считаются точно известными, а значения *параметров* (коэффициентов математической модели АС) могут быть любыми в пределах заданных *интервалов*.

Частью АС является объект. Объекты могут быть двух видов: чисто-интервальные и функционально-интервальные. *Чисто-интервальным* называется объект, в описание которого коэффициенты, значения которых могут изменяться независимо друг от друга в пределах некоторых интервалов, входят один раз и в первой степени. *Функционально-интервальным* называется объект, коэффициенты которого являются функциями от некоторых интервально неопределенных величин.

Основа интервальной формы учета неопределенностей – интервальная математика [1]. В теории АС интервальная математика применяется примерно с 1980 г. [2]. Одной из основных задач интервального анализа является задача об оценке интервала области значений функции. Если рассмотреть некоторую интервальнозначную функцию (ИЗФ) $f(\omega, [\nu_i])$ (ν_i – параметр, изменяющийся в интервале $[\nu_i]$), у которой ω – независимая переменная, принимающая вещественные значения, то при возможных значениях параметров в зависимости от ω мы будем иметь график изменения интервала значений этой функции. Следовательно, ИЗФ представима двумя граничными функциями – мажорантой и минорантой. Если независимая переменная является комплексной величиной (в частности, чисто мнимой $j\omega$), то значения функции будут комплексным

*Электростальский политехнический институт (филиал Московского государственного института стали и сплавов). E-mail: khleb@kodos.ru.

интервалом. Они могут быть представлены в виде прямоугольных или круговых комплексных интервалов. Задача состоит в том, чтобы предложить аналитические подходы и эффективные компьютерные алгоритмы для построения мажорант и минорант. В КП “АСИАС” применяется комбинация трех групп методов: методов математического анализа (задача о наибольшем и наименьшем значении функции), методов интервального анализа, и метода угловых элементов. Фактически, это оригинальный алгоритм решения в символьном виде задачи о наибольшем и наименьшем значении функции.

2. Версии комплекса программ “АСИАС”

Комплекс программ “АСИАС” начал разрабатываться в 1980 г. в Саратовском политехническом институте на кафедре “Автоматика и телемеханика” первым автором данного доклада (таблица).

Назначение комплекса программ АСИАС – анализ устойчивости и качества автоматических систем, а также синтез регуляторов автоматических систем с интервальной неопределенностью параметров (рис. 1).

Как у нас в стране, так и за рубежом теория автоматического управления стала ареной интенсивных исследований систем с интервальными неопределенностями в параметрах. Деятельность и математиков и “управленцев” развивалась на фоне бурного развития вычислительной техники, языков программирования и крупных программных систем вообще. Появлялись многочисленные пакеты “интервальной” арифметики и разнообразное “интервальное” программное обеспечение. В данном докладе упомянуты новые разработки (известные авторам доклада) и основное внимание акцентировано на пакетах, относящихся к интервально заданным системам автоматического управления. Разумеется, представляемая информация не претендует на исчерпывающую полноту.

Одним из первых пакет программ для решения задач расчета интервальных АС был разработан М. А. Цацкиным. Наиболее известные пакеты для идентификации систем разработаны Ю. А. Меркурьевым (Рига, 1982 – для статистики АС), А. П. Воцининым, Г. Р. Сотировым и др. (INA, Москва, 1989 – статические модели), А. Ф. Бочковым и Нгуен Вьет Зунгом (INTIDE, Москва, 1992 – для статических объектов, линейных динамических объектов, нелинейных динамических объектов в виде модели Гаммерштейна и модели Винера). Программа INTIDE, в частности, была написана на языке программирования Си для персональных компьютеров, работает под управлением операционной системы MS DOS, имеет две версии (на русском и английском языках), поддерживается системой меню-окон, графики и подсказок.

Для синтеза регуляторов интервально заданных АС известны пакеты программ, созданных И. В. Дугаровой, Е. М. Смагиной (INT, Томск, 1988 – для расчета П- и ПИ-регуляторов, в том числе для многосвязных систем), А. Н. Моисеевым, С. В. Алексеевым (INTERVAL, Анжеро-Судженск, 1997 – для синтеза модального регулятора, решения задачи асимптотического слежения за командным сигналом и моделирования трубки движения).

Для анализа АСИИП разработаны комплексы программных средств Ф. М. Ахметжановым (Уфа, 1996 – по интервальным полиномам и по “труб-

Версии КП “АСИАС”

Версия/год	Место разработки	Тип ЭВМ	Программное обеспечение	Кол-во программ/ Макс. размерн. задач	Разработчики
АСИАС_{ЕС} 1980–1989	г. Саратов, Саратовский политехнический институт (СПИ), кафедра АТМ	ЕС ЭВМ	FORTRAN-IV/ ОСЕС	12/12	<i>Вед. программист:</i> доц. Хлебалин Н.А. <i>Программисты:</i> студенты кафедры “Автоматика и телемеханика” (АТМ) СПИ
АСИАС_{РС} 1990–1999	г. Электросталь, Эл-ст. филиал МИСиС (ЭФ МИСиС), кафедра АТПП	IBM PC/AT- 286...486, Pentium I	FORTRAN-V/ MS DOS (Windows 9x)	17 + xreclam/3 или 10	<i>Вед. программист:</i> асп. Лазарев А.В. <i>Программисты:</i> студенты кафедры “Автоматизация технологических процессов и произ- водств” (АТПП) ЭФ МИСиС
АСИАС_{ЕН} 1990–1994	АНГЛОЯЗЫЧНАЯ ВЕРСИЯ КОМПЛЕКСА ПРОГРАММ АСИАС _{РС}				
АСИАС₉₈ 1997–1998	ВЕРСИЯ ДЛЯ УЧЕБНОГО ПРОЦЕССА			12/3	<i>Вед. программист:</i> доц. Хлебалин Н.А. <i>Программисты:</i> Те же, что и для АСИАС _{РС}
АСИАС_М 1998 – по н.в.	г. Электросталь, ЭПИ МИСиС, кафедра АТПП	Pentium I и выше	MATLAB/ Windows 95, 98, 2000	6 / не ограничен	<i>Вед. программисты:</i> ст. ДАМ-97 Вахитов О.А., ст. ДАМ-99 Пятых Д.С. <i>Программисты:</i> Внуков, Голев, Смельцов, Михайлова, Селиверстов, Кузовахо
АСИАС_Д 2002 – по н.в.	г. Электросталь, ЭПИ МИСиС, кафедра АТПП	Pentium II и выше	Delphi/ Windows 98, 2000	3 + xreclam _Д / 2, 5, не ограничен	<i>Вед. программист:</i> ст. ВАМ-97 Голев В.В. <i>Программисты:</i> Владимирова, Пятых

кам” частотных характеристик), Н. А. Целигоровым (Ростов-на-Дону, 1999 – робастная абсолютная устойчивость для нелинейных импульсных АС с интервальной неопределенностью, интерактивный выбор параметров управляющего устройства).

Помимо этого стоит упомянуть *комплексы интервально-математических программ*, созданные А. П. Вошининым, Г. Р. Сотировым и др. (SOLI, Москва, 1989 – позволяет решать задачи линейного программирования с интервально заданной целевой функцией), А. Б. Бабичевым, О. Б. Кадыровой, Т. П. Кашеваровой, А. Л. Семеновым (UniCalk, Новосибирск, 1991 – для решения произвольных систем алгебраических и алгебро-дифференциальных соотношений), а также отдельные известные программы В. М. Нестерова (Санкт-Петербург) и С. П. Шарого (Новосибирск).

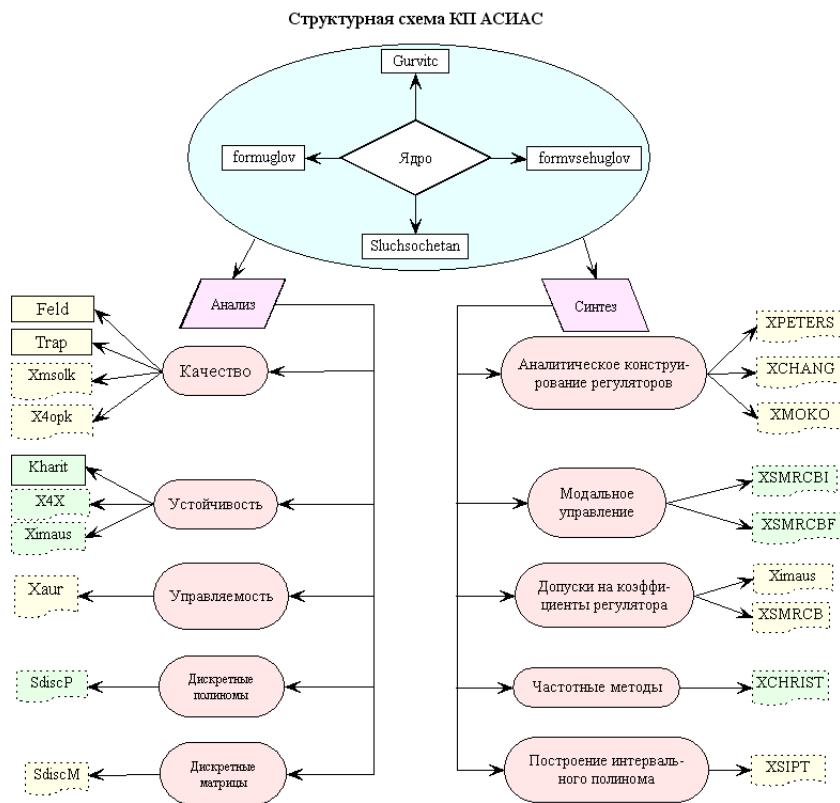


Рис. 1. Структурная схема КП АСИАС

Перейдем теперь к более подробному описанию нашего комплекса программ “АСИАС”.

На жестком диске “фортрановская” версия комплекса программ “АСИАС” хранится в одноименной папке (рис. 2).

Внутри нее находятся следующие папки (рис. 3).

Каждая из папок Asias_98, BasProg, Xtrap98, АСИАС_exe, АСИАС_архивы, АСИАС_тексты занимает около 1,4 Мб и, таким образом, может храниться на дискете. Названия папок дают представление о том, что в них находится. Для пояснения приложен файл read_me.txt. В папке BasProg находятся основные программы, отвечающие за “интервальные” вычисления, в частности – в папке XOLL.

Программы комплекса программ “АСИАС” хранятся в виде исполняемых файлов (расширение .exe) и допускают ввод данных либо с экрана, либо из файла. Каждая программа снабжена контрольным примером. Имя файла контрольного примера совпадает с именем программы и имеет расширение dat. После окончания работы программа выводит результаты в файл со своим именем и расширением rzl. Описания программ и контрольных примеров хранятся в файлах с именами, имеющими расширение chi.

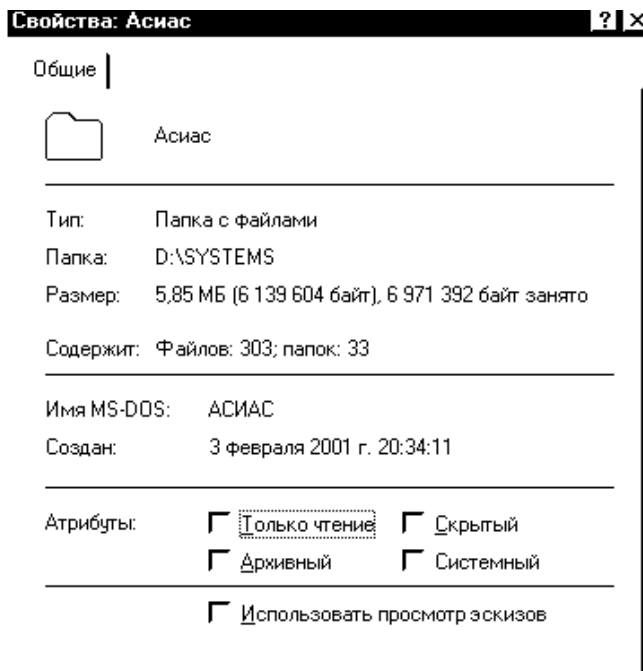


Рис. 2. Общие свойства папки “АСИАС”

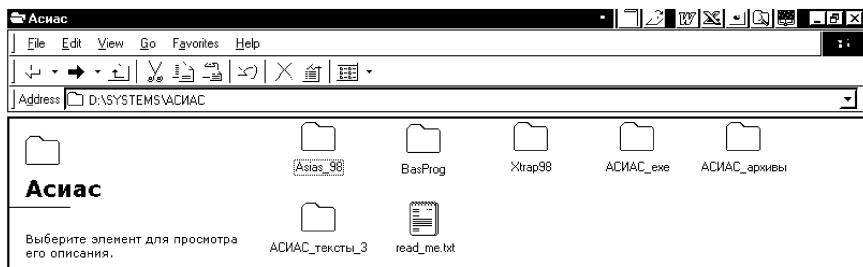


Рис. 3. Структура папки “АСИАС”

ВЫБЕРИТЕ УСТРОЙСТВО ВВОДА ДАННЫХ

0 - ЭКРАН.

1 - ФАЯП

ВВЕДИТЕ: 0**ВВЕДИТЕ РАЗМЕР МАТРИЦЫ А _1****РАЗМЕР МАТРИЦЫ А 1****ВВЕДИТЕ НИЖНИЕ ГРАНИЦЫ ЭЛЕМЕНТОВ МАТРИЦЫ А ПО СТРОКАМ**
-1**ВВЕДИТЕ ВЕРХНИЕ ГРАНИЦЫ ЭЛЕМЕНТОВ МАТРИЦЫ А ПО СТРОКАМ**
-2**НИЖНЯЯ ГРАНИЦА ИНТЕРВАЛА БОЛЬШЕ ВЕРХНЕЙ !**

-

Рис. 4. Защита от неправильно вводимых границ интервалов

Программы снабжены различными видами защит. Пример работы защиты от неправильно вводимых данных (программа ХМОКО1) показан на рис. 4.

В учебной версии комплекса программ “АСИАС” размеры решаемых задач ограничены третьим порядком, число управлений не может быть больше двух (исключением является программа ХИМАУS1, размер решаемых задач в которой равен пяти).

Новыми версиями КП АСИАС являются Delphi-версия АСИАС_D и MATLAB-версия АСИАС_M.

3. Новые возможности комплекса программ “АСИАС”

Решаемые комплексом “АСИАС” задачи относятся к робастному управлению автоматическими системами с интервальной неопределенностью параметров (АСИНП).

Задача *синтеза* состоит в проверке условий существования регулятора с постоянными (но, возможно, неопределенными, лежащими в некоторых допусках) коэффициентами по области параметрической неопределенности объекта и желаемой цели управления и в вычислении коэффициентов регулятора (если это возможно). Целью управления считается придание замкнутой системе устойчивости, а процессам, в ней протекающим, некоторых качественных показателей, определенных в виде функционалов или ограничений различного вида.

Синтез регуляторов для АСИНП проводится интервальным вариантом метода модального управления [3]. Используется способ включения интервального характеристического полинома замкнутой системы (ИХПЗС) в желаемый ИХПЗС. Достоинства и недостатки этого способа после десятилетнего опыта применения были изложены в [4]. Этот способ реализован в виде программы XSMRCB1. Для работы этой программы необходим желаемый ИХПЗС. Его синтез по желаемой области расположения корней в виде трапеции, расположенной в левой комплексной полуплоскости проводится программой XSIPT1. Соответствующая теория и алгоритм опубликованы в [5].

Под *анализом* АСИНП понимается получение оценок основных свойств АС при неопределенности параметров объекта и коэффициентов регулятора. Решение задачи анализа обычно преследует цель улучшения свойств системы, что предполагает изучение объекта для установления корректности требований к его функционированию.

Одним из этапов анализа АСИНП является определение ее управляемости. Следующим этапом анализа АСИНП является определение ее устойчивости. Если модель АСИНП задана в виде передаточной функции, то устойчивость оценивается по функционально-интервальному полиному ее знаменателя. Если модель АСИНП задана дифференциальными уравнениями, то используются критерии устойчивости функционально-интервальных матриц.

Качество АСИНП можно оценивать по корневым, частотным и временным характеристикам. КП “АСИАС” позволяет строить одно- и многосвязные области локализации нулей и полюсов АСИНП, различные виды ЧХ. Однако наиболее привлекательной является такая его возможность, как построение трубок траекторий во временной области. Для построения трубок используется

мажорантно-минорантная аппроксимация различной степени точности всего множества решений дифференциального уравнения или переменных состояния.

Система MATLAB – один из самых мощных и эффективных инструментов для создания разнообразных программных комплексов, предназначенных для решения научно-технических задач. Благодаря текстовому формату m-файлов пользователь может ввести в систему любую новую команду, оператор или функцию и затем пользоваться ими также просто, как и встроенными операторами или функциями. Благодаря именно этой особенности, а также встроенным в систему MATLAB средствам аналитических вычислений и элементам интервальных вычислений, стало возможным перевести написанные на языке Фортран программы комплекса программ “АСИАС” в систему MATLAB, создав m-файлы с текстами аналогичных программ.

Рассмотрим, для примера, как это было реализовано для программы построения уточненной области локализации корней интервального полинома (Xvicin).

Программа “Построение многоугольной области локализации корней устойчивого интервального полинома” имеет имя Xvicin. Программа функционирует в системе MATLAB версии 5.2 и выше.

Программа написана на встроенном в MATLAB m-языке.

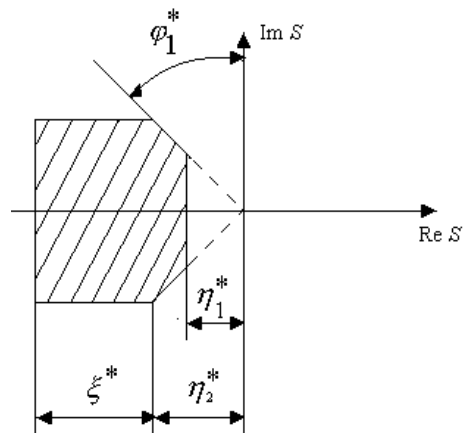


Рис. 5. Многоугольная область локализации корней интервального полинома

Программой решается задача построения многоугольной области локализации устойчивого интервального полинома.

Эта задача является частным случаем задачи об относительной устойчивости. Автоматическая система считается относительно устойчивой, если все корни ее характеристические полиномы локализованы в заданной области Λ комплексной плоскости.

Для того чтобы построить область локализации корней нужно построить все его угловые полиномы, найти корни, а затем “вокруг” этих корней построить область.

Такая область строится в данной программе по следующему алгоритму:

1. Проверка правильности введенных данных (нижние границы меньше верхних).
2. Формирование коэффициентов полиномов Харитонова.
3. Проверка устойчивости полиномов по критерию устойчивости Гурвица.
4. Если полином устойчив, то переход к п. 5, иначе сообщение о неустойчивости полинома и конец работы программы.
5. Формирование коэффициентов угловых полиномов.

6. Вычисление корней угловых полиномов и определение из их параметров области расположения корней.
7. Вывод графика области расположения корней и вывод параметров области.

Блок-схема программы представлена на рис. 6.

Входными параметрами являются:

- степень интервального полинома – n ,
- нижние границы коэффициентов интервального полинома – Gn ,
- ввод осуществляется со старшей степени,
- верхние границы коэффициентов интервального полинома – Gv ,
- ввод осуществляется со старшей степени.

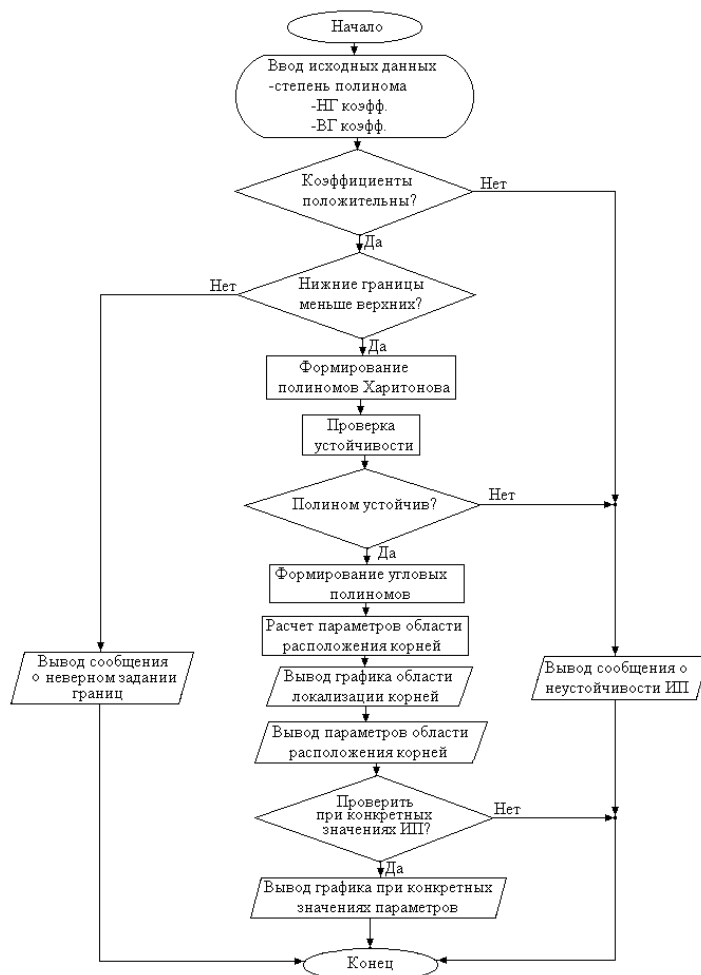


Рис. 6. Блок-схема программы Xvcisn

Данные вводятся в ответ на соответствующие запросы программы.

Число n – целое, значения коэффициентов Gn , Gv записываются в квадратных скобках и через пробел.

Выходными данными являются параметры области расположения корней:

fi – угол сектора Фи,

tet – степень устойчивости Тета,

ksi – расстояние до максимально удаленного корня Ksi ,

Ym – максимальная мнимая часть корней угловых полиномов.

Другими новыми возможностями КП “АСИАС” являются: анализ дискретных автоматических систем, коэффициентные методы анализа и синтеза АСИНП, частотные методы исследования функционально-интервальных АСИНП.

Список литературы

- [1] Шокин Ю.И. Интервальный анализ. – Новосибирск: Наука, 1981. – 112 с.
- [2] Хлебалин Н.А. Анализ асимптотической устойчивости линейных систем управления в условиях неопределенности параметров объекта / Саратовский политехн. ин-т. – Саратов, 1980. – Деп. в ЦНИИТЭИприборостроения 23.07.80, № 1370. – 9 с.
- [3] Хлебалин Н.А. Аналитический метод синтеза регуляторов в условиях неопределенности параметров объекта // Аналитические методы синтеза регуляторов: Межвуз. научн. сб. / Саратовский политехн. ин-т. – Саратов, 1981. – С. 107–123.
- [4] Хлебалин Н.А., Шокин Ю.И. Интервальный вариант метода модального управления // ДАН СССР. – 1991. – Т. 316, № 4. – С. 846–850.
- [5] Хлебалин Н.А. Построение интервальных полиномов с заданной областью расположения корней // Аналитические методы синтеза регуляторов: Межвуз. научн. сб. / Саратовский политехн. ин-т. – Саратов, 1982. – С. 92–98.

Критерий неограниченности допустимого множества решений

И.А. Шарая*

Аннотация. Получен критерий неограниченности для допустимого множества решений интервальной системы линейных алгебраических уравнений.

Договоримся использовать обозначения, предложенные в [1], и классическую интервальную арифметику (см., напр., [2]).

Для интервальной системы линейных алгебраических уравнений $\mathbf{Ax} = \mathbf{b}$, где $\mathbf{A} \in \mathbb{IR}^{m \times n}$ – интервальная матрица размерности $m \times n$, $\mathbf{b} \in \mathbb{IR}^m$ – интервальный вектор длины m , $x \in \mathbb{R}^n$ – вещественный вектор длины n) *допустимым множеством решений* или просто *допустимым множеством* будем называть, в соответствии с [3], множество

$$\Xi = \{x \mid (\forall A \in \mathbf{A})(\exists b \in \mathbf{b})(Ax = b)\}. \quad (1)$$

Определение (1) можно переписать в виде $\Xi = \{x \mid (\forall A \in \mathbf{A})(Ax \subseteq \mathbf{b})\}$ или, пользуясь свойством $\{Ax \mid A \in \mathbf{A}\} = \mathbf{Ax}$, в виде удобного для проверки и исследования *критерия принадлежности* вектора x *допустимому множеству решений*:

$$x \in \Xi \iff \mathbf{Ax} \subseteq \mathbf{b}. \quad (2)$$

Легко доказать [4, 5], что Ξ – выпуклое многогранное множество (т.е. пересечение конечного числа полупространств) в \mathbb{R}^n . Мы покажем, как по виду матрицы \mathbf{A} можно судить об ограниченности допустимого множества решений.

Докажем сначала два вспомогательных утверждения. Первое позволяет по специальному виду матрицы сказать, что допустимое множество неограничено, а второе, наоборот, показывает, что для неограниченного допустимого множества матрица имеет этот специальный вид.

Напомним, что конечное множество $\{a_j\}$ вещественных векторов называется *линейно зависимым*, если существует множество $\{c_j\}$ вещественных чисел, не все из которых равны нулю, такое, что $\sum_j a_j c_j = 0$. Заметим, что множество из одного вектора линейно зависимо тогда и только тогда, когда этот вектор нулевой.

Утверждение 1. Пусть допустимое множество Ξ непусто. Если в матрице \mathbf{A} есть линейно зависимые вещественные столбцы, то Ξ неограничено.

Доказательство. В силу критерия (2) принадлежности допустимому множеству непустота Ξ означает, что существует $\tilde{x} \in \mathbb{R}^n$ такой, что $\mathbf{A}\tilde{x} \subseteq \mathbf{b}$. Расписав произведение $\mathbf{A}\tilde{x}$ по столбцам $\mathbf{A}_{:j}$, получим

*Институт вычислительных технологий СО РАН.

$$\sum_{j=1}^n \mathbf{A}_{:j} \tilde{x}_j \subseteq \mathbf{b}. \quad (3)$$

Пусть J – множество номеров линейно зависимых вещественных столбцов $\mathbf{A}_{:j}$ матрицы \mathbf{A} . Тогда (3) можно переписать в виде

$$\sum_{j \in J} \mathbf{A}_{:j} \tilde{x}_j + \sum_{j=1, j \notin J}^n \mathbf{A}_{:j} \tilde{x}_j \subseteq \mathbf{b}, \quad (4)$$

а линейную зависимость вещественных столбцов выразить формулой

$$\sum_{j \in J} \mathbf{A}_{:j} c_j = 0, \quad (5)$$

где $c_j \in \mathbb{R}$ и $\sum_{j \in J} |c_j| > 0$.

Домножая (5) на произвольное вещественное t , добавляя к (4) и пользуясь законом дистрибутивности для вещественных чисел, получим

$$\forall t \in \mathbb{R} \quad \sum_{j \in J} \mathbf{A}_{:j} (\tilde{x}_j + tc_j) + \sum_{j=1, j \notin J}^n \mathbf{A}_{:j} \tilde{x}_j \subseteq \mathbf{b}. \quad (6)$$

Введем вектор $c = (c_1, \dots, c_n)$, дополнив множество коэффициентов линейной зависимости нулевыми для $j \notin J$. Тогда (6) можно переписать в виде

$$\forall t \in \mathbb{R} \quad \sum_{j=1}^n \mathbf{A}_{:j} (\tilde{x}_j + tc_j) \subseteq \mathbf{b},$$

что в матричной форме выглядит так:

$$\forall t \in \mathbb{R} \quad \mathbf{A}(\tilde{x} + tc) \subseteq \mathbf{b}.$$

По критерию принадлежности (2) это означает, что вместе с решением \tilde{x} в множество Ξ попадает прямая, проходящая через \tilde{x} и параллельная ненулевому вектору c . Значит, Ξ неограничено. \square

Утверждение 2. Пусть допустимое множество Ξ непусто. Если оно неограничено, то в матрице \mathbf{A} есть линейно зависимые вещественные столбцы.

Доказательство. Как упоминалось во введении, Ξ – это выпуклое многогранное множество. Если Ξ неограничено, значит, неограничено его пересечение с каким-нибудь ортантом. Тогда в этом ортанте лежит выпуклое многогранное неограниченное подмножество Ξ , из которого можно выбрать какой-нибудь луч $(\tilde{x} + tc)$, где \tilde{x} – начало луча, c – ненулевой вектор направления, $t \in \mathbb{R}^+$ – параметр, задающий точки луча.

Так как луч $(\tilde{x} + tc)$ целиком лежит в Ξ , то по критерию принадлежности (2)

$$\forall t \in \mathbb{R}^+ \quad \mathbf{A}(\tilde{x} + tc) \subseteq \mathbf{b}. \quad (7)$$

С другой стороны, луч $(\tilde{x} + tc)$ целиком лежит в одном ортанте, поэтому

$$\forall j \in \{1, \dots, n\} \quad x_j c_j \geq 0$$

и в (7) можно раскрыть скобки по правилу дистрибутивности [2, 6]:

$$\forall t \in \mathbb{R}^+ \quad \mathbf{A}\tilde{\mathbf{x}} + \mathbf{A}(tc) \subseteq \mathbf{b}. \quad (8)$$

Для произвольных интервальных векторов \mathbf{x} , \mathbf{y} , \mathbf{z} имеет место очевидное свойство

$$\left. \begin{array}{l} \mathbf{x} \subseteq \mathbf{y} \\ \mathbf{x} + \mathbf{z} \subseteq \mathbf{y} \end{array} \right\} \implies |\mathbf{z}| \leq \text{wid } \mathbf{y},$$

где $|\cdot|$ – модуль, wid – ширина интервальных векторов. Используя это свойство в (8) для $\mathbf{x} = \mathbf{A}\tilde{\mathbf{x}}$, $\mathbf{z} = \mathbf{A}(tc)$, $\mathbf{y} = \mathbf{b}$, получим, что должно выполняться неравенство

$$\forall t \in \mathbb{R}^+ \quad |\mathbf{A}(tc)| \leq \text{wid } \mathbf{b}. \quad (9)$$

Для вещественного t применимо правило дистрибутивности [2, 6]. Это позволяет переписать (9) в виде

$$\forall t \in \mathbb{R}^+ \quad |t(\mathbf{A}c)| \leq \text{wid } \mathbf{b}.$$

Положительное t можно вынести за знак модуля и разделить на него обе части неравенства. Тогда

$$\forall t \in \mathbb{R}^+ \setminus \{0\} \quad |\mathbf{A}c| \leq \frac{\text{wid } \mathbf{b}}{t}.$$

Это возможно только при $|\mathbf{A}c| = 0$, что эквивалентно

$$\mathbf{A}c = 0. \quad (10)$$

Равенство (10) означает, что

$$\sum_{j=1}^n \mathbf{A}_{:j} c_j = 0,$$

т. е. линейная комбинация столбцов интервальной матрицы \mathbf{A} с коэффициентами, соответствующими компонентам вектора c , равна нулю. Остается заметить, что ненулевые коэффициенты в этой сумме могут стоять только при вещественных столбцах матрицы. В противном случае радиус линейной комбинации

$$\text{rad}(\mathbf{A}c) = \sum_{j=1}^n |c_j| \text{rad } \mathbf{A}_{:j} \quad (11)$$

будет отличен от нуля, что противоречит (10). \square

Следствием утверждений 1 и 2 является

Критерий неограниченности. Пусть допустимое множество Ξ непусто. Оно неограничено тогда и только тогда, когда в матрице \mathbf{A} есть линейно зависимые вещественные столбцы.

Примеры применения критерия неограниченности
(для непустого допустимого множества)

Интервальная матрица A	Множество Ξ
Нет вещественных компонент	ограничено
Каждый столбец имеет незначительную компоненту	ограничено
Есть нулевые столбцы	неограничено
Есть пропорциональные вещественные столбцы	неограничено
Число вещественных столбцов, больше общего числа строк	неограничено

Замечание о виде неограниченного допустимого множества. Давайте попробуем представить, как выглядит неограниченное Ξ . Пусть $c \in \mathbb{R}^n$ – произвольный вектор коэффициентов, для которого линейная комбинация столбцов интервальной матрицы A равна нулю, т.е. $Ac = 0$. Обозначим через L пространство всех таких векторов. В силу (11) у вектора c могут отличаться от нуля только компоненты, соответствующие вещественным столбцам матрицы A . Значит, размерность пространства L равна $(p - q)$, где p – число всех вещественных столбцов, q – максимальное число линейно независимых вещественных столбцов матрицы A . Например, если в матрице A все вещественные столбцы нулевые, то размерность пространства L равна их числу.

Опираясь на доказательство утверждения 1, можно сказать, что Ξ представляет собой объединение прямых, параллельных произвольному вектору c из L . Следовательно, допустимое множество состоит из пространств, полученных параллельными сдвигами L . Для выпуклого многогранного множества это означает, что все его грани лежат в гиперплоскостях, параллельных L .

Список литературы

- [1] Kearfott R.B., Nakao M.T., Neumaier A., Rump S.M., Shary S.P., van Hentenryck P. Standardized notation in interval analysis. – <http://www.mat.univie.ac.at/~neum/software/int/>.
- [2] Neumaier A. Interval methods for systems of equations. – Cambridge: Cambridge University Press, 1990.
- [3] Шарый С.П. Алгебраический подход к анализу линейных статических систем с интервальной неопределенностью // Изв. РАН. Теория и системы управления. – 1997. – № 3. – С. 51–61 (<http://www.ict.nsc.ru/lab1.2/Shary/IzvAN.ps>).
- [4] Rohn J. Inner solutions of linear interval systems // Interval Mathematics 1985: Lect. Notes in Comput. Sci. / K. Nickel, ed. – New York: Springer Verlag, 1986. – Vol. 212. – P. 157–158.
- [5] Шайдуров В.В., Шарый С.П. Решение интервальной алгебраической задачи о допусках. – Красноярск, 1988. – (Препринт / АН СССР. Сиб. отд.-ние. ВЦ; 5).
- [6] Шарая И.А. О дистрибутивности в классической интервальной арифметике // Вычисл. технологии. – 1997. – Т. 2, № 1. – С. 71–83 (<http://www.ict.nsc.ru/lab1.2/Irene/ct97.ps>).

Interval-affine Gaussian method for systems with constrained coefficients

R.R. Akhmerov*

Abstract. In paper outer problem for interval linear algebraic systems with constrained coefficients is considered. For the case of linear constraints it is shown that the interval-affine Gaussian method can be successfully applied. The results of computation experiments using this method are given.

By the systems with constrained coefficients we mean such linear algebraic systems when there are some constraints for elements of real matrices from main interval matrix. In paper we consider case of linear constraints mainly. For example, elements of symmetric and skew-symmetric matrices are linearly constrained.

Interval-affine Gaussian method was constructed as alternative for well-known interval Gaussian method. It based on using interval-affine arithmetic and it gives a chance to noticeably improve the quality of outer enclosures of joined solution set. This improvement is consequence of interval-affine arithmetic ability to take into account dependence between arguments in arithmetical operations. It is shown this property let us easily adapt interval-affine Gaussian method for solving problem with constrained coefficients. In linear case the method takes constraints into account entirely and actually reduces problem to problem with less number of interval parameters.

The results of computation experiments using this method are given. The experiments show that taking constraints into account let us sharpen the outer enclosures of joined solution set.

*Altai State University.

Quantified Set Inversion Algorithm

Pau Herrero, Miguel A. Sainz, Josep Vehí, Luc Jaulin

Introduction

Real physical world problems frequently involve non-linear constraints over real numbers, uncertain data and badly defined parameters. This kind of problems can be expressed in terms of numerical Constraints Satisfaction Problems (CSP) [1].

Usually, quantifiers (\forall , \exists) arise in certain situations of uncertainty in the parameters of CSP. For example, universal quantification, models situations where some parameters are unknown and the mathematical model has to hold for every possible choice of these parameters. On the other hand, existential quantification models situations where some parameters can be chosen by the designer. When this quantification appears on a CSP, it can be expressed in terms of numerical Quantified Constraints Satisfaction Problems (QCSP).

The importance of solving QCSP lies on the fact that many physical problems, for example in control engineering [2–5], electrical engineering [6], mechanical engineering [7, 8], biology [9] and various others [10, 11], can be expressed under this paradigm.

Up to now, Cylindrical Algebraic Decomposition [12–14], for which a practical implementation exists [15], has been the most extended method to solve this type of problems. However, this technique is only well suited for small or middle-size problems because of its computational complexity. Moreover, it often generates huge output consisting on highly complicated algebraic expressions which are not useful for many applications and it does not provide partial information before computing the total result.

Methods that appear lately [16, 17] try to avoid some of these problems restricting oneself to approximate instead of exact solutions, using solvers based on numerical methods. However, these algorithms are also restricted to very special cases (e. g. quantified variables only occur once, only one quantifier, etc.). Recently, some of these deficiencies have been partially removed by Ratschan [18–20] but, a lot of work remains to be done before obtaining an efficient and general method. It is important to remark the important contribution on the mathematical foundations of the problem recently done by Shary in [21].

This paper describes a new reliable and efficient method, based on Modal Interval Analysis [22, 23], Set Inversion techniques [24], for the characterization of solution sets defined by numerical Quantified Constraints Satisfaction Problems (QCSP).

1. Problem Statement

A Quantified Constraint (QC) is an algebraic expression over the reals which contains quantifiers (\exists , \forall), predicate symbols (e. g., $=$, $<$, \leq), function symbols (e. g., $+$, $-$, \times , \sin , \exp), rational constants and variables $\mathbf{x} = \{x_1, \dots, x_n\}$ ranging over reals domains $\mathbf{D} = \{D_1, \dots, D_n\}$.

An example of a QC is the following one,

$$\forall x \in \mathbb{R} \quad x^4 + px^2 + qx + r \geq 0, \quad (1)$$

where x is a universally (\forall) quantified variable and p and r are free variable.

As defined in [21], a numerical constraint satisfaction problem, is a triple CSP = $(\mathbf{x}, \mathbf{D}, \mathbf{C}(\mathbf{x}))$ defined by

- (i) a set of numeric variables $\mathbf{x} = \{x_1, \dots, x_n\}$,
- (ii) a set of domains $\mathbf{D} = \{D_1, \dots, D_n\}$ where D_i , a set of numeric values, is the domain associated with the variable x_i .
- (iii) a set of constraints $\mathbf{C}(\mathbf{x}) = \{\mathcal{C}_1(\mathbf{x}), \dots, \mathcal{C}_m(\mathbf{x})\}$ where a constraint $\mathcal{C}_i(\mathbf{x})$ is determined by any numeric relation (equation, inequality, inclusion, etc.) linking a set of variables under consideration.

A solution to a numeric constraint satisfaction problem CSP = $(\mathbf{x}, \mathbf{D}, \mathbf{C}(\mathbf{x}))$ is an instantiation of the variables of \mathbf{x} for which both inclusion in the associated domains and all the constraints of $\mathbf{C}(\mathbf{x})$ are satisfied. All the solutions of a constraint satisfaction problem thus constitute the set

$$\Sigma = \{\mathbf{x} \in \mathbf{D} \mid \mathbf{C}(\mathbf{x}) \text{ is satisfied}\}. \quad (2)$$

Now suppose that the constraints $\mathbf{C}(\mathbf{x})$ depend on some parameters p_1, p_2, \dots, p_l about which we only know that they belong to some intervals P_1, P_2, \dots, P_l . Moreover, these parameters have an associated quantifier $Q \in \{\forall, \exists\}$. Taking into account the dual character of interval uncertainty, the most general definition of the set of solutions to such Quantified Constraint Satisfaction problem QCSP should have the form

$$\Sigma = \{\mathbf{x} \in \mathbf{D} \mid Q_1(p_{\sigma_1}, P_{\sigma_1}) \dots Q_l(p_{\sigma_l}, P_{\sigma_l}) \mathbf{C}(\mathbf{x})\}, \quad (3)$$

where

- Q_i are logical quantifiers \forall or \exists (in this paper, only the case of universal quantifiers preceding the existential ones will be dealt),
- $\{p_1, p_2, \dots, p_l\}$ is the set of parameters of the constraints system considered,
- $\{P_1, P_2, \dots, P_l\}$ is a set of intervals containing the possible values of these parameters, and
- $\sigma_i \in \Sigma_l$ is a permutation of the numbers $1, \dots, l$.

The sets of the form (3) will be referred to as quantified solutions sets to the numerical quantified constraints satisfaction problem QCSP = $(\mathbf{x}, \mathbf{D}, \mathbf{C}(\mathbf{x}))$.

2. Methodology

2.1. Set Inversion. One way of solving a CSP is through the characterization of its solution set by means of the Set Inversion (SI) approach.

Let CSP be a constraint satisfaction problem CSP = $(\mathbf{x}, \mathbf{D}, \mathbf{C}(\mathbf{x}))$. Set inversion aims at characterizing the set Σ of all \mathbf{x} such that \mathbf{C} is satisfied.

Remark. All constraints are considered under the form $\mathcal{C}(\mathbf{x}) := \mathbf{f}(\mathbf{x}) = \mathbf{y}$, where \mathbf{f} a continuous function from \mathbb{R}^n to \mathbb{R}^m .

Given a box \mathbf{X} (cartesian product of intervals), an algorithm which does set inversion is based on a branch-and-bound technique and the 3 followings set of rules.

Rule 1: $\forall(\mathbf{x}, \mathbf{X})\mathcal{C}(\mathbf{x}) \Leftrightarrow \mathbf{X} \subseteq \Sigma$.

This logic formula, used to prove that a box \mathbf{X} is contained in the solution set, is equivalent to the following interval computation and interval inclusions

$$\text{Out}(f(\mathbf{X})) \subseteq \mathbf{Y},$$

where $f(\mathbf{X})$ are the ranges of the function components over the interval vector \mathbf{X} and $\text{Out}(f(\mathbf{X}))$ are outer approximations of $f(\mathbf{X})$

Rule 2: $\forall(\mathbf{x}, \mathbf{X})\neg\mathcal{C}(\mathbf{x}) \Leftrightarrow \mathbf{X} \subseteq \bar{\Sigma}$.

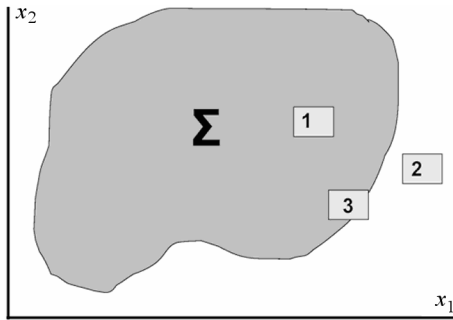


Figure 1. Solution set: 1 – solution, 2 – non-solution, 3 – undefined

This logic formula, used to prove that a box \mathbf{X} does not belongs to the solution set, is easily proved by means of the following interval computation and interval inclusions

$$\text{Out}(f(\mathbf{X})) \subseteq \bar{\mathbf{Y}}.$$

Finally, if Rule 1 and 2 are not accomplished the position of the box \mathbf{X} is undefined

Rule 3: Otherwise, \mathbf{X} is undefined.

Figure 1 shows a two dimensional example of the three possible situations corresponding to the 3 rules.

Then the algorithm which does set inversion is as follows

Algorithm SI (In: $\mathcal{C}, \mathbf{X}, \epsilon$, Out: $\text{Inn}(\Sigma), \text{Out}(\Sigma)$)

1. **if** $\text{With}(\mathbf{X}) \leq \epsilon$ **then** \mathbf{X} is undefined
2. **else if** Rule 1 is satisfied **then** \mathbf{X} is solution
3. **else if** Rule 2 is satisfied **then** \mathbf{X} is non solution
4. **else** $\text{Branch}(\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2)$
5. $\text{SI}(\mathcal{C}, \mathbf{X}_1, \epsilon)$
6. $\text{SI}(\mathcal{C}, \mathbf{X}_2, \epsilon)$

where

- ϵ : SI stops the branching procedure over \mathbf{X} when this precision is reached,
- $\text{Inn}(\Sigma)$: Inner approximation of the solution set,
- $\text{Out}(\Sigma)$: Outer approximation of the solution set.

2.2. Quantified Set Inversion via Modal Interval Analysis. Classical Set Inversion is well suited characterizing solution sets of the form (2). The problem arises when the sets are of the form (3). Then, a new algorithm for the characterization of quantified solution sets is needed. This algorithm will be referred to as Quantified Set Inversion (QSI).

Let us consider the case when the constraints are under the form $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) \stackrel{\leq}{\geq} 0$, with f a continuous function from \mathbb{R}^n to \mathbb{R} .

The main difference between the classical Set Inversion Algorithm and the quantified one lies on the used set of rules. For the proposed algorithm the following rules will be used:

Rule 1: $\forall(\mathbf{x}, \mathbf{X})\forall(\mathbf{p}_U, \mathbf{P}_U)\exists(\mathbf{p}_E, \mathbf{P}_E)\mathcal{C}(\mathbf{x}) \Leftrightarrow \mathbf{X} \subseteq \Sigma$.

This logic formula, used to prove that a box \mathbf{X} belongs to the solution set, can not be easily proved by means of classical interval computations. For this reason, Modal Interval Analysis is proposed (MIA). MIA is a powerful mathematical tool which allows the evaluation of quantified interval formulas by means of interval computations. Concretely, to evaluate the set of logic formulas, the *-semantic theorem given by MIA is used to reduce equivalently the logical formula to the interval inclusion

$$\text{Out}(f^*(\mathbf{X}, \mathbf{P}_U, \mathbf{P}_E)) \subseteq Z,$$

where \mathbf{X}, \mathbf{P}_U are proper intervals, \mathbf{P}_E improper one, $\text{Out}(f^*(\mathbf{X}, \mathbf{P}_U, \mathbf{P}_E))$ is an outer approximation of the the *-semantic extension of the continuous function f and $Z = [0, 0]$, $Z = [-\infty, 0]$ or $Z = [0, \infty]$ depending on if the constraints are under the form $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) = 0$, $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) < 0$ or $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) > 0$, respectively.

In order to obtain the second rule, used to prove that a box \mathbf{X} does not belongs to the solution set, the following implication is used:

Rule 2: $\neg(\forall(\mathbf{p}_U, \mathbf{P}_U)\exists(\mathbf{p}_E, \mathbf{P}_E)\exists(\mathbf{x}, \mathbf{X}) \mathcal{C}(\mathbf{x})) \Rightarrow \mathbf{X} \subseteq \bar{\Sigma}$.

This logical formula is, analogously, equivalent to the following interval exclusion:

$$\text{Inn}(f^*(\mathbf{X}, \mathbf{P}_U, \mathbf{P}_E)) \not\subseteq Z,$$

where \mathbf{P}_U is a proper interval, \mathbf{X}, \mathbf{P}_E improper ones, $\text{Inn}(f^*(\mathbf{X}, \mathbf{P}_U, \mathbf{P}_E))$ is an inner approximation of the the *-semantic extension of the continuous function f . and $Z = [0, 0]$, $Z = [-\infty, 0]$ or $Z = [0, \infty]$ depending on if the constraints are under the form $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) = 0$, $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) < 0$ or $\mathcal{C}(\mathbf{x}) := f(\mathbf{x}) > 0$, respectively.

Finally, if none of these rules are accomplished, the box \mathbf{X} is undefined.

Rule 3: otherwise, \mathbf{X} is undefined.

Computing the semantic extension of a continuous function f by means of any of their interpretable rational extensions provokes an overestimation of the interval evaluation, due to the multi-occurrences of variable, when the rational computations is not optimal. An algorithm, based on results of Modal Interval Analysis and branch-and-bound techniques which allows to efficiently compute an inner and an outer approximation of f^* has been recently built.

When the constraints are under the form $C(\mathbf{x}) := \mathbf{f}(\mathbf{x}) \stackrel{\leq}{\geq} 0$, with \mathbf{f} a continuous function from \mathbb{R}^n to \mathbb{R}^m and each variable existentially quantified appears in only a component function, the problem is reduced to m different problems, one for each component function.

3. Example

Given the intervals $T_1, \dots, T_m, Y_1, \dots, Y_m$, to find the inner estimation of the sets:

$$\begin{aligned}\Sigma_E &= \{ \mathbf{p} \in \mathbb{R}^2 \mid \exists(t_1, T'_1) \exists(y_1, Y'_1) \cdots \exists(t_n, T'_n) \exists(y_n, Y'_n) \\ &\quad (p_1 e^{-p_2 t_1} = y_1, \dots, p_1 e^{-p_2 t_n} = y_n) \}, \\ \Sigma_U &= \{ \mathbf{p} \in \mathbb{R}^2 \mid \forall(t_1, T'_1) \exists(y_1, Y'_1) \cdots \forall(t_n, T'_n) \exists(y_n, Y'_n) \\ &\quad (p_1 e^{-p_2 t_1} = y_1, \dots, p_1 e^{-p_2 t_n} = y_n) \}.\end{aligned}$$

We have

$$\Sigma_E = \Sigma_{E_1} \cap \dots \cap \Sigma_{E_n}, \quad \Sigma_U = \Sigma_{U_1} \cap \dots \cap \Sigma_{U_n},$$

where

$$\begin{aligned}\Sigma_{E_i} &:= \{ \mathbf{p} \in \mathbb{R}^2 \mid \exists(t_i, T'_i) \exists(y_i, Y'_i) p_1 e^{-p_2 t_i} = y_i \}, \\ \Sigma_{U_i} &:= \{ \mathbf{p} \in \mathbb{R}^2 \mid \forall(t_i, T'_i) \exists(y_i, Y'_i) p_1 e^{-p_2 t_i} = y_i \}\end{aligned}$$

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

3.1. Characterizing Σ_E . The logic formula which fulfils the points belonging to the solution set Σ_{E_i} is

$$\forall(p_1, P'_1) \forall(p_2, P'_2) \exists(t_i, T'_i) \exists(y_i, Y'_i) \quad p_1 e^{-p_2 t_i} - y_i = 0$$

which is equivalent to the following inclusion test

$$\text{Out}(f_i^*(P_1, P_2, T_i, Y_i)) \subseteq [0, 0],$$

with P_1 and P_2 proper intervals and T_i and Y_i improper ones.

The logic formula which fulfils the points not belonging to the solution set Σ_{E_i} is

$$\neg(\exists(p_1, P'_1) \exists(p_2, P'_2) \exists(t_i, T'_i) \exists(y_i, Y'_i) \quad p_1 e^{-p_2 t_i} - y_i = 0)$$

which is implied by the following exclusion test

$$\text{Inn}(f_i^*(P_1, P_2, T_i, Y_i)) \not\subseteq [0, 0],$$

with P_1, P_2, T_i and Y_i improper intervals.

Then, $\Sigma_E = \Sigma_{E_1} \cap \dots \cap \Sigma_{E_n}$.

In less than 4 seconds on a Pentium III 1 GHz, for $n = 2$, an $\epsilon = 0.05$ and the following interval domains: $\mathbf{X} = (P_1, P_2) = ([-1, 4], [-1, 1])$, $Y'_1 = [1.3, 3.3]$, $Y'_2 = [0.3, 2.3]$, $T'_1 = [2, 3]$ and $T'_2 = [3.5, 4]$, QSI generates the paving represented in figure 2, where the darker region corresponds to the solution set Σ_E , the grey region corresponds to the non solution set $\bar{\Sigma}_E$ and the white region is undefined.

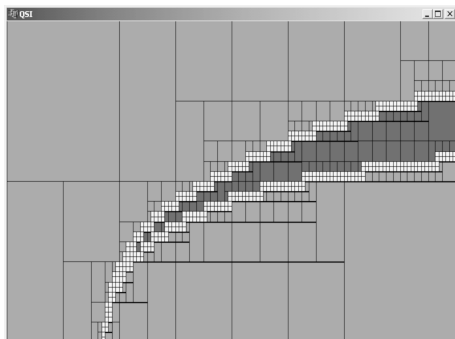


Figure 2. Σ_E paving

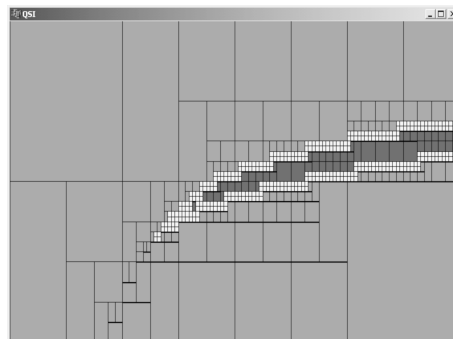


Figure 3. Σ_U paving

3.2. Characterizing Σ_U . The logic formula which fulfil the points belonging to the solution set Σ_{U_i} is

$$\forall(p_1, P'_1)\forall(p_2, P'_2)\forall(t_i, T'_i)\exists(y_i, Y_i) \quad p_1 e^{-p_2 t_i} - y_i = 0$$

which is equivalent to the following inclusion test

$$\text{Out}(f_i^*(P_1, P_2, T_i, Y_i)) \subset [0, 0], \tag{4}$$

where P_1, P_2 and T_i are proper intervals and Y_i is improper.

The logic formula which fulfil the points not belonging to the solution set Σ_{U_i} is

$$\neg(\forall(t_i, T'_i)\exists(y_i, Y'_i)\exists(p_1, P'_1)\exists(p_2, P'_2) \quad p_1 e^{-p_2 t_i} - y_i = 0)$$

which is implied by the following exclusion test

$$\text{Inn}(f_i^*(P_1, P_2, T_i, Y_i)) \not\subset [0, 0]$$

with P_1, P_2 and Y_i improper intervals and T_i proper.

Then, $\Sigma_U = \Sigma_{U_1} \cap \dots \cap \Sigma_{U_m}$.

In less than 3 seconds on a Pentium III 1 GHz, for $n = 2$, an $\epsilon = 0.05$ and the same interval domains used for the previous example, QSI generates the paving represented in figure 3, where the darker region corresponds to the solution set Σ_U , the grey region corresponds to the non solution set $\bar{\Sigma}_U$ and the white region is undefined.

4. Future Work

4.1. Reducing the complexity via Constraint Propagation. In order to reduce the complexity of the set inversion algorithm due to the branching, a narrowing operator (a contractor) for quantified constraints will be provided. This contractor, based on *constraint propagation* techniques and Modal Interval Analysis, allows the contraction of an initial box \mathbf{X} to another one \mathbf{X}' such that \mathbf{X}' still contains the solution set Σ .

The basic idea consists on decomposing the set of constraints into their primitive constraints and by means of Modal Interval Arithmetic to compute local approximations of the solution space for a given primitive constraint. These evaluation

provokes domain reduction over \mathbf{X} which are propagated through the whole set of constraints by a propagation engine.

4.2. Application on parameter identification. An application on parameter identification and its comparison with the classical interval approach used in [25] is under study.

References

- [1] Lhomme O. Consistency techniques for numerical CSPs // Intern. Joint Conf. on Artificial Intelligence (IJCAI). – 1993. – Vol. 1. – P. 232–238.
- [2] Abdallah C., Dorato P., Yang W., Liska R., Steinberg S. Applications of quantifier elimination theory to control system design // 4th IEEE Mediterranean Symp. on Control and Automation, Crete, Greece, 1996.
- [3] Jaulin L., Walter É. Guaranteed tuning, with application to robust control and motion planning // Automatica. – 1996. – Vol. 32, No. 8. – P. 1217–1221.
- [4] Jirstrand M. Nonlinear control system design by quantifier elimination // J. Symbolic Computation. – 1997. – Vol. 24, No. 2. – P. 137–152.
- [5] Dorato P. Quantified multivariate polynomial inequalities // IEEE Control Systems Magazine. – 2000. – Vol. 20, No. 5. – P. 48–58.
- [6] Sturm T. Reasoning over networks by symbolic methods // Applicable Algebra in Engineering Communication and Computing. – 2000. – Vol. 10, No. 1. – P. 79–96.
- [7] Ioakimidis N.I. REDLOG-aided derivation of feasibility conditions in applied mechanics and engineering problems under simple inequality constraints // J. Mechanical Engineering (Strojnícky Casopis). – 1999. – Vol. 50, No. 1. – P. 58–69.
- [8] Ioakimidis N.I. Quantifier elimination in applied mechanics problems with cylindrical algebraic decomposition // Intern. J. Solids and Structures. – 1997. – Vol. 34, No. 30. – P. 4037–4070.
- [9] Chauvin C., Müller M., Weber A. An application of quantifier elimination to mathematical biology // Computer Algebra in Science and Engineering. – World Scientific, 1994. – P. 287–296.
- [10] Benhamou F., Goualard F. Universally quantified interval constraints // Proc. of the Sixth Int. Conf. on Principles and Practice of Constraint Programming (CP'2000). – Singapore, 2000. – No. 1894.
- [11] Gerber R., Pugh W., Saksena M. Parametric dispatching of hard real-time tasks // IEEE Transactions on Computers. – 1995. – Vol. 44, No. 3. – P. 471–479.
- [12] Tarski A. A Decision Method for Elementary Algebra and Geometry. – Berkeley: Univ. of California Press, 1951.
- [13] Collins G.E. Quantifier elimination for real closed fields by cylindrical algebraic decomposition // 2nd GI Conf. Automata Theory and Formal Languages. – 1975. – Vol. 33. – P. 134–189.
- [14] Hong H. Simple solution formula construction in cylindrical algebraic decomposition based quantifier elimination // Proc. ISSAC'92, International Symposium on Symbolic and Algebraic Computation. – 1992. – P. 177–188.
- [15] Brown Chr. W. Quantifier Elimination by Partial Cylindrical Algebraic Decomposition. – <http://www.cs.usna.edu/~qepcad/B/QEPCAD.html>.

-
- [16] Garloff Ü., Graf B. Solving strict polynomial inequalities by Bernstein expansion // The Use of Symbolic Methods in Control System Analysis and Design. – London: The Institution of Electrical Engineers, 1999. – P. 339–352.
 - [17] Benhamou F., Goulard F. Universally quantified interval constraints // Sixth Intern. Conf. on Principles of Constraint Programming, 2000.
 - [18] Ratschan S. Quantified Constraints Under Perturbation // J. Symbolic Computation. – 2001. – Vol. 33, No. 4. – P. 493–505.
 - [19] Ratschan S. Efficient Solving of Quantified Inequality Constraints over the Real Numbers // Submitted for publication, 2003.
 - [20] Ratschan S. Approximate Quantified Constraint Solving. – <http://www.mpi-sb.mpg.de/~ratschan/AQCS/AQCS.html>.
 - [21] Shary S.P. A new technique in systems analysis under interval uncertainty and ambiguity // Reliable Computing. – 2002. – Vol. 8. – P. 321–418.
 - [22] Gardenyes E., Sainz M.Á., Jorba L., Calm R., Estela R., Mielgo H., Trepát A. Modal Intervals // Reliable Computing. – 2001. – Vol. 7, No. 2. – P. 77–111.
 - [23] Sainz M.Á., Gardenyes E., Jorba L. Interval Estimations of Solution Sets to Real-Valued Systems of Linear or Non-Linear Equations // Reliable Computing. – 2002. – Vol. 8, No. 4. – P. 283–305.
 - [24] Jaulin L., Walter E. Set inversion via interval analysis for nonlinear bounded-error estimation year // Automatica. – 1993. – Vol. 32, No. 8. – P. 1053–1064.
 - [25] Jaulin L., Walter E. Guaranteed Bounded-Error Parameter Estimation for Nonlinear Models with Uncertain Experimental Factors // Automatica. – 1999. – Vol. 35, No. 5. – P. 849–856.

An interval algorithm combining symbolic rewriting and componentwise Newton method applied to control a class of queueing systems*

Bartłomiej Jacek Kubica[†], Krzysztof Malinowski[‡]

Introduction

The idea of Componentwise Newton Operator has been presented in [9]. It seems up to now not much attention has been paid to this concept.

In this paper we present properties of the Componentwise Newton Operator and show how they can be used in numerical algorithms. It seems they are especially important when transforming the problem by some symbolic methods, based on the computation of Groebner bases.

The resulting algorithm will be applied to a problem of optimizing control rules for a queueing system.

1. Componentwise Newton method

1.1. The definition. According to [9] we define the interval componentwise Newton operator as follows.

Assume we try to solve the equation system:

$$\begin{cases} g_1(x_1, \dots, x_n) = 0, \\ \dots \\ g_n(x_1, \dots, x_n) = 0, \end{cases}$$

where $x_1 \in X_1, \dots, x_n \in X_n$.

We denote real variables by small letters and intervals by cardinal letters. Real-valued functions and its interval envelopes are denoted by the same small letters (this should not lead to any misunderstanding. The symbol $\text{mid}(X)$ denotes the midpoint of the interval X).

The componentwise Newton operator for this equation system with respect to box X , equation i and variable j is defined as:

$$N_{\text{cmp}}(X, i, j) = \text{mid}(X_j) - \frac{g_i(X_1, \dots, X_{j-1}, \text{mid}(X_j), X_{j+1}, \dots, X_n)}{\frac{\partial g_i}{\partial x_j}(X_1, \dots, X_n)}. \quad (1)$$

*This work has been supported by the Foundation for Polish Science through its stipends.

[†]Warsaw University of Technology, e-mail: bkubica@elka.pw.edu.pl

[‡]Research and Academic Computer Network (NASK), e-mail: K.Malinowski@ia.pw.edu.pl

1.2. Known properties. In [9] two simple properties of the N_{cmp} operator are proven:

Proposition 1. *All roots of the equation system belonging to the box X (if any) must lie in $X \cap N_{\text{cmp}}(X, i, j)$ for any i, j .*

Proposition 2. *If for an arbitrary i and j we have $X \cap N_{\text{cmp}}(X, i, j) = \emptyset$ then there are no roots in the box X .*

1.3. New properties. One of the most interesting and very useful properties of traditional interval Newton operators (wide literature available, e.g. [10]) is that they allow to verify not only the existence of a root in a box, but also the uniqueness of it. We omit the details due to the fact that they are well-known.

Can analogous properties be found for the componentwise method? The answer is positive, though they hold for special classes of functions only.

Theorem 1. *Assume that for a given i the function $g_i(\dots)$ depends only on x_j :*

$$g_i(x_1, \dots, x_n) = g_i(x_j). \quad (2)$$

Then if the componentwise Newton operator satisfies the condition:

$$N_{\text{cmp}}(X, i, j) \subset X_j \quad (3)$$

then there is a unique value of $x_j \in X_j$ for which $g_i(x_1, \dots, x_n) = 0$. The method has quadratic rate of convergence.

The proof is obvious because for the considered case the componentwise Newton operator is equivalent to the univariate Newton operator of a one-dimensional function.

The condition (2) is not as restrictive as it may seem; if we seek for zeros of the gradient of a separable function, it will be satisfied for all equations.

Can it be somehow applied to non-separable problems, such as the one in section 3, though? The answer is positive again (at least for polynomial equations), but we have to apply some symbolic transformations first.

2. Groebner bases

2.1. Basics. The formal definition of a Groebner basis may be found e.g. in [5] or many other works and books. We will not describe the sophisticated theory here, we just say informally that the computation of a Groebner basis of a polynomial equation system is analogous to transforming the linear equation system to its triangular form.

The meaning is as follows.

Consider an ordering of the variables, say $x_1 \prec x_2 \prec \dots \prec x_n$. We can transform the polynomial system to an equivalent one, for which the last equation will depend only on x_1 , the previous one – on x_1 and x_2 etc. (this property is called *the elimination property* of Groebner bases with lexical ordering).

Groebner bases have already been used in connection with interval analysis (see [3, 4]) and they proved useful. Interval Newton methods were more efficient on transformed equation systems, due to reduced dependency problem.

For componentwise method the problem is a bit different. We would like to have univariate equations for each of the variables, not only for one of them. We can obtain it by computing Groebner bases for different orderings and taking the univariate equations from each of them.

2.2. Conversions of Groebner bases. Unfortunately, the cost of computing a Groebner Basis may be exponential. However, the conversion of one Groebner Basis to another one (e.g. the basis for another ordering of variables) may be done by quite efficient polynomial algorithms.

There are a few methods for such conversions – the FGLM method [8], the Groebner Walk method [6] or the LLL method [2] to name a few.

They differ in many details. In our application most interesting will be the FGLM method. It is restricted to systems with zero-dimensional ideal, but it will allow us to find the univariate equation (for the “smallest” variable) *without* computing the whole basis.

This gives us a fast way to obtain univariate equations for each variable.

2.3. Application to an interval method. The resulting system of n univariate equations won't be equivalent to the primary problem; it will constitute necessary conditions only. It seems however that using such equations in addition to the original problem may increase the efficiency of the algorithm, because we can quickly reduce the dimensions of boxes by cheap iterations of univariate Newton method.

3. An example – long-run optimization of the Mendelson's queueing system

The considered example is a combination of problems from [7] and [11]. We try to optimize the long-run behaviour of a priority queueing system. The notion “long-run” means in this context that we can control not only the arrival rates $\lambda_1, \dots, \lambda_R$, but also the service rate μ .

The problem is as follows:

$$\max_{\lambda_1, \dots, \lambda_R, \mu} J = \sum_{i=1}^R V_i(\lambda_i) - \sum_{i=1}^R \lambda_i \cdot G_i(\lambda_1, \dots, \lambda_R; \mu) - C(\mu), \quad (4)$$

s.t.

$$0 \leq \lambda_i \leq \Lambda_i \quad \forall i = 1, \dots, R, \quad (5)$$

$$\sum_{i=1}^R \lambda_i < \mu. \quad (6)$$

Meaning of the parameters is as follows:

- $V_i(\cdot)$, $i = 1, \dots, R$, is the aggregate value received by the clients from class i ; authors generally assume the form $V_i(\lambda_i) = a_i \cdot \lambda - \frac{a_i}{2\Lambda_i} \cdot \lambda_i^2$,

- $G(\dots)$ is the delay cost (see below),
- $C(\mu)$ is the capacity cost (authors assume linear cost $C(\mu) = b \cdot \mu$).

The problem is known to be nonconvex and multiextremal [12].
In case when delay cost has linear structure, we can write it as:

$$G_i(\lambda_1, \dots, \lambda_R) = \delta \cdot W_i(\lambda_1, \dots, \lambda_R).$$

Mendelson and Whang claim [11] that the expected sojourn times for a non-preemptive priority queue are given by the following equation:

$$W_i(\lambda_1, \dots, \lambda_R) = \frac{\sum_{j=1}^R \lambda_j / \mu_j^2}{(1 - \sum_{j=1}^{i-1} \rho_j) \cdot (1 - \sum_{j=1}^i \rho_j)} + \frac{1}{\mu_i}, \quad (7)$$

where $\rho_j = \frac{\lambda_j}{\mu_j}$.

The motivation for the above formula may be found e. g. in [1].

For homogeneous service times for all client classes, i. e. for $\mu_1 = \dots = \mu_R = \mu$ equation (7) takes the form:

$$W_i(\lambda_1, \dots, \lambda_R) = \frac{\sum_{j=1}^R \lambda_j}{(\mu - \sum_{j=1}^{i-1} \lambda_j) \cdot (\mu - \sum_{j=1}^i \lambda_j)} + \frac{1}{\mu}. \quad (8)$$

In the optimization problem formulated above a strict inequality $\sum_{i=1}^R \lambda_i < \mu$ is present. Changing it to a non-strict inequality we risk obtaining infinite values of the function during the computations.

We can overcome this difficulty e. g. by the following substitution of the variables:

$$\mu = \sum_{i=1}^R \lambda_i + \psi \quad \psi \in [\epsilon_\psi, +\infty]$$

where ϵ_ψ is a small positive number.

In practice the upper bound on the ψ guaranteeing that the maximal value of μ is a few times bigger than the sum of maximal values of the arrival rates is enough in both cases.

The following bounds were used in numerical computations: $\psi \in [0.0005, 5.0]$.

4. Comments

The problem from the previous section is obviously non-separable. All functions may be described by polynomials, though. The developed method seems to be suitable then.

The implementation will be done in C++ and will base on C-XSC 2.0 beta3 and CToolbox 2.0 beta3 (see [14]) and the library from [13] extended by the first author. The symbolic computations use the PoSSoLib package ver. 4.99 [15].

In many cases it may be especially profitable to use the PoSSoLib polynomial coefficients in computations. The coefficients are then polynomials of “variables”

a_i, Λ_i, δ_i . Symbolic computations on them may reduce the overestimation due to occurrence of these parameters (usually known only approximately) in many equations.

Other details of the implementation and numerical results are planned to be presented during the Workshop.

References

- [1] Adan I., Resing J. Queueing Theory, 2001, the book downloadable from. – <http://www.cs.duke.edu/~fishhai/misc/queue.pdf>.
- [2] Basiri A., Faugere J.C. Changing the Ordering of Groebner Bases with LLL: Case of Two Variables. – <http://www.inria.fr/INRIA/publication/publi-ps-gz/RR/RR-4746.ps.gz>
- [3] Benhamou F., Granvilliers L. Combining Local Consistency, Symbolic Rewriting and Interval Methods. – 1994. – http://www.sciences.univ-nantes.fr/info/perso/permanents/benhamou/papers/BenGra_AISMC96.pdf.
- [4] Benhamou F., Granvilliers L. Automatic Generation of Numerical Redundancies for Non-linear Constraint Solving. – 1997. – http://www.sciences.univ-nantes.fr/info/perso/permanents/benhamou/papers/BenGra_Reliable97.pdf.
- [5] Buchberger B. Groebner Bases and Systems Theory. – 2001. – <http://www.risc.uniz.ac.at/people/buchberg/papers/2001-05-12-A.ps>.
- [6] Collart S., Kalkenbrener M., Mall D. Converting bases with the Groebner walk // J. Symbolic Computations. – 1997. – No. 24. – P. 465–469.
- [7] Dewan S., Mendelson H. User Delay Costs and Internal Pricing for a Service Facility // Management Science. – 1990. – Vol. 36, No. 12. – P. 1502–1517.
- [8] Faugere J.C., Gianni P., Lazard D., Mora T. Efficient Computation of Zero-dimensional Groebner Bases by Change of Ordering // J. Symbolic Computations. – 1993. – No. 16. – P. 329–344.
- [9] Herbort S., Ratz D. Improving the Efficiency of a Nonlinear-System-Solver Using the Componentwise Newton Method. – 1997. – <http://www.uni-karlsruhe.de/~iam/html/reports/rep9702.ps.gz>.
- [10] Kearfott R.B. An Interval Branch and Bound Algorithm for Bound Constrained Optimization // J. Global Optimization. – 1992. – No. 2. – P. 259–280. – <http://interval.louisiana.edu/preprints.html/>.
- [11] Mendelson H., Whang S. Optimal Incentive – Compatible Priority Pricing for the M/M/1 Queue // Operations Research. – 1990. – Vol. 38, No. 5. – P. 870–883.
- [12] Stidham Jr. S. Pricing and Capacity Decisions for a Service Facility: Stability and Multiple Local Optima // Management Science. – 1992. – Vol. 38, No. 8. – P. 1121–1139.
- [13] Wiechowski P. An object-oriented interval library for solving global optimization problems and nonlinear equation systems (in Polish): BEng Thesis under supervision of Andrzej Karbowski. – Warsaw, 2003.
- [14] <http://www.xsc.de/>.
- [15] <http://posso.dm.unipi.it/>.

Example of Babuška, Práger and Vitásek in interval computations

Gregory G. Men'shikov*

Abstract. Taking the well-known example by Babuška, Práger and Vitásek (“BPV example”), we show that the interval computations themselves (even their validating version) do not guarantee high quality of the computational work yet, and the stability of the “core” algorithm plays a crucial role. If such an algorithm is not suitable, then the whole procedure is fallable.

As for “BPV example”, we, first, use it to show troubles emerging in traditional computations apart from the absence of the guaranteed information about the error. Then, we apply validated interval computations to the “BPV example”, but the accuracy of results still remains low.

Trying to proceed in the interval manner without taking any care of the validation, shows that the quality of the results is not improved. On the other hand, we may lose the information about the accuracy, the latter may prove really interesting in the results.

Finally we make an decisive break by using the intersection of separate enclosures of the result under computation or a family of such results. At the same time, we can see how the validation achieved by analytical means may make the results of calculations more precise.

1. The point execution of the example

Let us consider the example of recurrent computation

$$I_n = \frac{1}{e} \int_0^1 x^n e^x dx, \quad n = 0, 1, \dots \quad (1)$$

from the book [1]. It is clear that $I_0 = 1 - e^{-1}$. To express I_n from I_{n-1} [2], we, integrating by parts, obtain the following recursive equality:

$$I_n = 1 - nI_{n-1}. \quad (2)$$

Computing I_1, I_2, \dots in accordance to the above formula on a real computer (when the computation are subject to rounding, etc.) produces after $n \approx 10 \div 15$ the result which is evidently wrong (negative), since the values I_n must be nonnegative. The analysis of of the phenomenon was fulfilled in the book [2], and its main reason turns out the subtraction of neighbouring compute values.

2. The interval (validated) execution of the example

Let us rerun the same example, in interval-validating manner at this time. Making use of the Second Theorem on compositions [2], we rewrite the right part of the recursive relation in enclosures, which leads to the inclusion:

*St. Petersburg State University, miksha@pobox.spbu.ru.

$$I_n \in 1 - n[I_{n-1}], \quad n = 0, 1, \dots \quad (3)$$

In the above, $[I_{n-1}]$ means an interval enclosure for I_{n-1} found by computer.

Hence, it is possible to use the right-hand part as $[I_n]$. Then, taking in account the majorization [1] (i. e., the auxiliary extension of the enclosure), we obtain the further modification of the relation (3):

$$[I_n] = 1 - n[I_{n-1}], \quad n = 0, 1, \dots \quad (4)$$

Its execution has been performed an interval assembler described in [2, § 17–18] while the corresponding code of the algorithm is presented in the Table 212.1 from the book [3].

The results of several steps of computations are in columns 1, 2, 3 of Table 1.

If n grows, the width of the enclosure increases even in an accelerating manner. Finally, at $n = 9$ it may be stated that further computation is senseless since drastical decreasing of the accuracy: the ratio of the width $[I_n]$ (denoted as $w(I_n)$) to I_n has the value near 0.1. Due to this, the value of $w([I_{15}])$ is large.

Table 1

n	$[I_n]$	$w([I_n])$	$[I_n]_{\text{int modif}}$	$w([I_n])_{\text{int modif}}$
0	.6321203 .6321208	4.77E-07	.6321203 .6321208	4.77E-07
1	.3678789 .36788	1.01E-06	.3678789 .36788	1.01E-06
2	.2642398 .2642424	2.56E-06	.2642398	2.56E-06
...
9	-.1697655 .3596304	5.29E-01	9.090908E-02 .1	9.09E-03
...
15	-941 903 965 832.4	1.91E+06	5.822351E-02 6.250001E-02	3.62E-03

Thus, interval-validated computation allows one to check the accuracy permanently. This property is absent in the usual, traditional computations, but the interval validated result may have abnormally large width, the enclosure being thus of low practical value.

3. The interval (non-validated) execution of the example

Table 2

n	$[I_n]$
0	.6321205
1	.3678795
2	.2642411
...	...
14	-797.5973
15	-11964.96

It may appear that the successful interval execution of the example does not give an essential narrowing of the width of the enclosure. This execution can also have a low accuracy. The results of interval execution of the same example are shown in Table 2. As before, the same program is used from Table 1. To make the results non-validated ones, we remove a majorization procedure from the computation. Specifically, we set the constant C of the majorization equal to zero [2].

Therefore, we obtain degenerated intervals. Naturally, the same effect (in a quality but without fail in a quantity) is produced by the point execution.

4. The interval-validated modification of the example

By interval analogue of the recursive formula (4), we have computed enclosures and confirmed computational instability of the process.

On the other hand, theoretical reasoning in the book [2] proves the boundness of $\{I_n\}$: the inequality is obtained:

$$\frac{1}{n+1} < I_{n-1} < \frac{1}{n}. \tag{5}$$

The corresponding close intervals $\left[\frac{1}{n+2}, \frac{1}{n+1}\right]$ present one more system of enclosures. Using them, it is possible to stop catastrophic raise of the width. Indeed, the non-empty intersection of the interval enclosures is the interval enclosure too. Thanks to this fact, we denote

$$[I_n] = (1 - n[I_{n-1}]) \cap \left[\frac{1}{n+2}, \frac{1}{n+1}\right] \tag{6}$$

instead of (3a).

It seems (in the ideal model of interval computing [3]),

$$w([I_n]) \leq w\left(\left[\frac{1}{n+2}, \frac{1}{n+1}\right]\right) = \frac{1}{n+1} - \frac{1}{n+2} = \frac{1}{(n+1)(n+2)}. \tag{7}$$

Hence, we arrive at $w(I_n) \rightarrow 0$.

Moreover, the relative width is infinitely small too:

$$\frac{w([I_n])}{|I_n|} \leq \frac{w([I_n])}{(n+1)} \leq \frac{1}{(n+2)} \rightarrow 0. \tag{8}$$

Making use of the modified interval algorithmics, let us carry out one more series of numerical experiments. Their program is formed on the basis of the above-mentioned one. Also, we run the computations for the same n as in the preceding section.

The results are contained in the columns 1, 4, 5 of Table 1, and this time the width does not increase. Moreover, it tends to relatively small values.

Thus, modifying the interval approach enables us not only to inquire into the accuracy issues, but to control the accuracy as well.

Probably, it is possible to obtain even more precise two-sided inequality for I_n . The relations of this type are suitable as a basis for further modifications of the algorithm for finding I_n .

5. The BPV example as the diverging iterative process

Sometimes, it may prove that checking out has confirmed that the point is of bad quality of the outer algorithm by point or interval execution, either validated or not.

Let us consider the BPV example as an iterative process, which correlates to the subjects of the book [1]. Here, a divergence takes place. Note that from the numerical stability viewpoint the other way is impossible.

Conclusion

Our text is primarily intended for those who are interested in the scientific computations, specifically, in their accuracy, and who teaches the interval computation in high schools.

Acknowledgements. I am very grateful to Dr. S. Shary (Novosibirsk) for his attention to this manuscript. I am also very grateful to my daughter Mrs. O. Men'shikova (St. Petersburg) for her efforts and helpful comments that facilitated the improvement of the manuscript.

References

- [1] Babuška I., Práger M., Vitásek E. Numerical Processes in Differential Equations. – Praha, Czechoslovakia: SNTL-publishers of Technical Literature, 1966. – 368 p.
- [2] Men'shikov G.G. Localizing Computing: Lectures Summary. Issue 1. Introduction to Interval-localizing Organization of Computations. – St. Petersburg: Department of Operative Poligraphy of Scientifical Research Institute of Chemistry of SPb State University, 2003. – 89 p. (in Russian).
- [3] Men'shikov G.G. Localizing Computing: Lectures Summary. Issue 2. Tasks of Compositional Computation and Problem of Rounhness their Interval-localizing Solution. – St. Petersburg: Department of Operative Poligraphy of Scientifical Research Institute of Chemistry of SPb State University, 2003. – 59 p. (in Russian).

A reliable curve tracing method

Dominique Michelucci, Dominique Faudot*

Introduction

Curve tracing is a problem encountered in CAD during the calculation of intersection between surfaces or the resolution of systems of equations by homotopy or in linear programming with the interior method. The curve is defined by $n - 1$ independent equations of n real variables $f(x) = 0$ with $f = (f_1, \dots, f_n)$ and $x = (x_1, \dots, x_n)$. Here only algebraic equations are considered. The curve is thus smooth almost everywhere. Tracing is classically done by two methods:

M1: By piecewise linear approximation. For any simplex of \mathbf{R}^n there is only one affine function $l_i = (x_1, \dots, x_n)$ which takes the same values as the function $f_i = (x_1, \dots, x_n)$ at the $n + 1$ vertices (the hypersurface in \mathbf{R}^n which equation is $f_i = (x_1, \dots, x_n) = 0$ the zero level set of which is a hyperplane inside the simplex. The intersection between the $n - 1$ hyperplanes $l_1(x_1, \dots, x_n) = \dots = l_{n-1}(x_1, \dots, x_n) = 0$ defines a line, which approaches the curve inside the simplex when this line cuts the simplex. Knowing a cut simplex, one gradually follows then the curve, from one simplex to another.

M2: By prediction and correction. Let a be a known regular point of the curve and t a vector tangent to the curve in a . The method predicts at first that a point $p_k = a + kt/\|t\|$ is close to the curve and corrects this point by some variant of the Newton method. In practice, the step of progression k is given empirically. Possible heuristics are: one gives k a priori. If the Newton's method converges in few stages and if the tangent at the new point is not too distant from t , then k was probably correct; otherwise one starts again with smaller k . We can however jump from a branch of $f(x) = 0$ to another one. One secures oneself then by using small values for k . This article proposes a reliable method to choose the step k .

An alternative of the second method considers the osculatory circle of the curve at the point a to control the step k ; this allows a better prediction but does not guarantee absolutely against the risk to jump from a branch to another as opposed to [7, 3].

Example. Let us consider the equation in \mathbf{R}^2 $c_1(x, y)c_2(x, y) = 0$ where c_1 and c_2 are two circle equations. For a point a on c_1 , the osculatory circle will be c_1 and one will deduce from them a step k and a point $p_k = a + kt/\|t\|$ independently of c_2 . It is thus enough to place c_2 judiciously (for example $p \in c_2$, c_2 tangent with the line ap , through points a and p , and $c_1 \cap c_2 = \emptyset$, so that the correction step converges towards a point of c_2 rather than of c_1).

*LE2I, Informatique, Université de Bourgogne, BP 47870, 21078 Dijon Cedex, FRANCE.
Emails: michelucci@u-bourgogne.fr; faudot@u-bourgogne.fr

1. Notations and conventions

1.1. Compatible norms, upperbounds. One uses the sup norm for the vectors $\|x\| = \max(|x_1|, \dots, |x_n|)$, x^t is the transpose of x . The norm of the matrices $\|M\|$ is $\sup \|xM\|$ for $\|x\| = 1$ and is here equal to $\max_{j=1}^n \sum_{i=1}^n |M_{i,j}|$ where $M_{i,j}$ is the element of line i column j . In other words, the norm of M is the greatest value obtained by making in each column the sum of the absolute values. One will need to compute an upper bound of the norm of a product of two matrices. One can certainly carry out the product of two matrices but the following upper bound (U_1) requires only the product of a vector by a matrix

$$\|AB\| \leq \|a\hat{B}\|, \quad \text{where } a_j = \sum_{i=1}^n |A_{i,j}| \quad \text{and } \hat{B}_{ij} = |B_{ij}| \quad (1)$$

e.g.

$$\left\| \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} -5 & -6 \\ 7 & -8 \end{pmatrix} \right\| \leq \left\| (|-1| + |3| \quad |2| + |-4|) \begin{pmatrix} |-5| & |-6| \\ |7| & |-8| \end{pmatrix} \right\|.$$

An upper bound (U_2) of $\|AB\|$ even faster and cruder is

$$\|AB\| \leq \|A\| \cdot \|B\|. \quad (2)$$

One will need an upper bound of the sum of two matrices. One can carry out the calculation of the sum but another upper bound is possible $\|A + B\| \leq \|A\| + \|B\|$.

The ball(a, R) is the set of points x such that $\|x - a\| \leq r$. Such a ball is in fact an hypercube of center a and half side r . The Jacobian of $f(x) = 0$ with $x = (x_1, \dots, x_n)$ and $f = (f_1, \dots, f_m)$ is the matrix $\frac{\partial f}{\partial x}$. Let α be a multi index, $|\alpha|$ is a notation for $\alpha_1 + \dots + \alpha_n$, $\binom{k}{\alpha}$ for $\frac{k!}{\alpha_1! \dots \alpha_n!}$ with $k \in \mathbf{N}$, x^α for $x_1^{\alpha_1} \dots x_n^{\alpha_n}$, $\frac{\partial^k f}{\partial x^\alpha}$ for $\frac{\partial^k f}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}$ with $k = |\alpha|$.

Other authors use a dual convention and write Mx for the product of a vector x by a matrix M , i. e. this dual convention regards x as a vector column. It thus uses the dual norm of ours $\|M\| = \sup_{\|x\|=1} \|Mx\| = \max_{i=1}^n \sum_{j=1}^n |M_{i,j}|$. In the same way, this convention considers a jacobian, which is the transpose of ours.

1.2. Naive Arithmetic of Intervals (IA). Naive IA [6] calculates on intervals $[v_0, v_1]$ where $v_0 \leq v_1$ are two standard floating-point numbers. We are using the usual rules of IA. The IA has for principal interest to give perfectly reliable results, contrarily to floating-point arithmetic (FPA). It does not modify the theoretical effectiveness of the methods: the elementary arithmetic operations on intervals are carried out in constant time.

But IA over-estimates the width of the intervals and does not permit the comparison of two values with overlapping intervals. The IA is today performed by software and is approximately 4 times slower than the standard FPA. One can however hope that it will be soon available on all the arithmetic processors.

IA is used in two very distinct ways:

- the intervals are sharp, at least initially. The intervals are then used only to control the rounding errors of floating point arithmetic.
- the intervals are broad, even initially. The IA is then a tool for numerical analysis by intervals. The amplitude of the intervals grows quickly during calculations. We will use IA according to this second mode in section 6 and elsewhere according to the first mode. An alternative was also proposed by [2] using affine IA.

2. The principle of our method

The curve is defined by $f(x) = 0$ with $x = (x_1, \dots, x_n)$ and

$$f(x) = (f_1(x), \dots, f_{n-1}(x)) = (0, \dots, 0).$$

The known point is $a = (a_1, \dots, a_n)$, a is a regular point: the rank $(n-1)$ of $\frac{\partial f}{\partial x}(a)$ is maximal. The tangent in a is $t = (t_1 \dots t_n)$ in other words $t \frac{\partial f}{\partial x}(a) = 0$. One supposes moreover that $\|t\| = \max(|t_1| \dots |t_n|) = 1$. To use the fixed-point theory, it is first assumed that each point of the curve $f(x) = 0$ is solution of a system of n unknown equations and n unknown variables.

Let $s_k(x) = (f(x), f_n(x, k))$, where $f_n(x, k) = 0$ is an additional equation, parameterized by the real k , k being such that $f_n(a, 0) = 0$ and measures the progression along the curve. A natural choice for the additional equation $f_n(x, k) = 0$ is $f_n(x, k) = (x - a - kt) \cdot t^t = t_1(x_1 - a_1 - kt_1) + \dots + t_n(x_n - a_n - kt_n) = 0$.

In other words, a is seen as the point of intersection between the curve $f(x) = 0$ and the hyperplane passing through a and normal with t ; this hyperplane will be translated by a vector kt to traverse the curve, close to a . The jacobian of the system $s_k(x) = 0$, independent of k , is $s' = s'_k = \frac{\partial s_k}{\partial x} = \left(\frac{\partial f}{\partial x}, t'\right)$.

For a given value of k , the corresponding point of the curve, solution of $s_k(x) = 0$ will be calculated by the quasi Newton iteration $x^{(1)} = a + kt$, $x^{(n+1)} = QN_k x^{(n)}$ where the quasi Newton function $QN(x)$ is defined by $QN_k(x) = x - s_k(x) [s'(a)]^{-1}$ (note $QN_k(a) = a + kt$). One could thus start from $x^{(1)} = a$ instead of $x^{(1)} = a + kt$.

The larger k the better, for advancing quickly along the curve; but it is also needed that the convergence of the correction step is guaranteed and fast. According to the fixed-point theory, the iteration $x^{(n+1)} = QN(x^n)$ starting from the initial point $x^{(1)}$ converges if one can find a neighborhood (actually a ball) B containing $x^{(1)}$ and satisfying the conditions C_1 and C_2 defined below.

C₁. The contractivity condition: QN is contracting in the ball B , i.e. for any couple of items x, y in B , $\|QN(x) - QN(y)\| \leq c\|x - y\|$ where $c < 1$ is the factor of contraction. One will even impose $c \leq 1/2$ to ensure that QN converges quickly: with each iteration, the distance to the solution is divided by at least 2. Since $\|QN(x) - QN(y)\| = \|QN(x) - QN(x + (y - x))\| \leq \|y - x\| \max_{z \in B} (QN'(z))$ one will impose that

$$\max_{x \in B} (\|QN'(z)\|) = \|QN'(B)\| \leq 1/2 \quad (3)$$

(contractivity condition CC).

C₂. Stability condition: the image of the ball B by QN $QN(B)$ is such that $QN(B) \subset B$. We will use in fact a condition stronger than C_2 , and thus sufficient to guarantee C_2 , and which is more easily computable. Let us suppose that r is known such that the ball $B(a, r)$ satisfies the condition of contractivity C_1 . Then a sufficient condition to satisfy C_2 is $\|QN_k(a) - a\| \leq r/2$.

Sufficient condition (SC). Indeed by assumption $x \in B = \text{ball}(a, R) \Rightarrow \|x - a\| \leq R$ and $\|QN'(B)\| \leq 1/2$ and $\|QN_k(a) - a\| \leq R/2$. Then

$$\|QN_k(x) - a\| \leq \|QN_k(x) - QN_k(a)\| + \|QN_k(a) - a\| \leq \frac{1}{2}\|x - a\| + \frac{R}{2} \leq \frac{R}{2} + \frac{R}{2} = R$$

that implies $x \in B$. However, the SC is equivalent to $k \leq R/2$.

Really, by assumption $\|QN_k(a) - a\| \leq \frac{R}{2}$. But $QN_k(a) = a + kt$. Then

$$\|QN_k(a) - a\| = \|kt\| = |k| \cdot \|t\| = |k| \leq \frac{R}{2}.$$

It is thus very simple to find k knowing R . It is enough to take $k = R/2$. This choice is probably not optimal but this very simple value is guaranteed. The entire problem is thus reduced to find R , such that in the ball $B(a, R)$ the CC (3) is guaranteed. $QN'(x)$ is the jacobian of $QN(x)$: $QN'_k = \frac{\partial QN}{\partial x} = Id - s'(x)(s'(a))^{-1}$ is independent of k and will be denoted by QN' . Defining $\sigma = x - a = (\sigma_1, \dots, \sigma_n)$ the jacobian of QN_k is

$$\begin{aligned} QN'(a + \sigma) &= Id - s'(a + \sigma)(s'(a))^{-1} = s'(a)(s'(a))^{-1} - s'(a + \sigma)(s'(a))^{-1} \\ &= -(s'(a + \sigma) - s'(a))(s'(a))^{-1} \\ &= - \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a + \sigma) - \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_{n-1}}{\partial x_1}(a + \sigma) - \frac{\partial f_{n-1}}{\partial x_1}(a) & 0 \\ \frac{\partial f_1}{\partial x_2}(a + \sigma) - \frac{\partial f_1}{\partial x_2}(a) & \dots & \frac{\partial f_{n-1}}{\partial x_2}(a + \sigma) - \frac{\partial f_{n-1}}{\partial x_2}(a) & 0 \\ \dots & \dots & \dots & 0 \\ \frac{\partial f_1}{\partial x_n}(a + \sigma) - \frac{\partial f_1}{\partial x_n}(a) & \dots & \frac{\partial f_{n-1}}{\partial x_n}(a + \sigma) - \frac{\partial f_{n-1}}{\partial x_n}(a) & 0 \end{pmatrix} (s'(a))^{-1} = -KM. \end{aligned}$$

The CC also ensures that in the ball $B(a, R)$ for given k , the system $s_k(x) = 0$ has at most one solution.

3. The naive method

The naive method calculates $QN'(a + \sigma)$ symbolically either by using the Taylor's formula, or by developing $\frac{\partial f_i}{\partial x_j}(a + \sigma) - \frac{\partial f_i}{\partial x_j}(a)$ naively. Then it multiplies K

by M , then it deduces for each column a polynomial in $r \geq \|\sigma\|$ which is an upper bound of the sum of the absolute values of the elements in this column. The largest upper bound is an upper bound of the norm of the matrix $QN'(a + \sigma)$. These n polynomials $p_i(r)$ are null at 0, and increase with r : it is easy by dichotomy to find r_i such that $p_i(r_i) = \frac{1}{2}$: $R = \min_{i=1}^n(r_i)$. The course of the naive method is illustrated on the following example.

3.1. Example:

$$n = 3, \quad f(x) = (x_1^3 - x_2^3 + 3x_1x_2^2 - 3x_1^2x_2 - x_3, x_2^2 + x_3^2 - 1) = (0, 0), \quad (4)$$

$$a = (1, 1, 0); \quad t = (1, 0, 0).$$

The auxiliary equation is $f_3(x, k) = x_1 - 1 - k$. The jacobian of s_k is

$$s'(k) = \begin{pmatrix} 3x_1^2 + 3x_2^2 - 6x_1x_2 & 0 & 1 \\ -3x_1^2 - 3x_2^2 + 6x_1x_2 & 2x_2 & 0 \\ -1 & 2x_3 & 0 \end{pmatrix}. \quad (5)$$

The inverse matrix of

$$s'(a) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 0 \end{pmatrix} \text{ is } s'(a)^{-1} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0.5 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

3.2. Naive method on Example 3.1:

$$QN'(a + \sigma) = - \begin{pmatrix} 3\sigma_1^2 + 3\sigma_2^2 - 6\sigma_1\sigma_2 & 0 & 1 \\ -3\sigma_1^2 - 3\sigma_2^2 + 6\sigma_1\sigma_2 & 2\sigma_2 & 0 \\ 0 & 2\sigma_3 & 0 \end{pmatrix} M$$

$$= - \begin{pmatrix} 0 & 0 & -3\sigma_1^2 - 3\sigma_2^2 + 6\sigma_1\sigma_2 \\ 0 & \sigma_2 & 3\sigma_1^2 + 3\sigma_2^2 - 6\sigma_1\sigma_2 \\ 0 & \sigma_3 & 0 \end{pmatrix},$$

$$\|QN'(a + \sigma)\| \leq \left\| \begin{pmatrix} 0 & 0 & |-3|r^2 + |-3|r^2 + |6|r^2 \\ 0 & r & |3|r^2 + |3|r^2 + |-6|r^2 \\ 0 & r & 0 \end{pmatrix} \right\|,$$

where $\|\sigma\| = \max(|\sigma_1|, |\sigma_2|, |\sigma_3|) \leq r$ and $\|QN'(a + \sigma)\| \leq \|(0, 2r, 24r^2)\|$.

In practice, a is not exactly on the curve, and M is not exactly the inverse of $s'(a)$. If floating intervals are used, calculations yield, very naturally, upper bounding polynomials. The constant term will not be exactly null, but a small number such as 10^{-9} or 10^{-6} . This is not awkward, and allows on the contrary to take into account and to control very naturally the numerical inaccuracy. The most important is that the upper bounding polynomials are almost null at 0.

The main disadvantage of the naive method is its cost: its use, implicit or explicit, of the Taylor development requires the calculation of all the terms $\frac{\partial^{|\alpha|} f_i(a)}{\partial x^\alpha}$. They are numerous, even when the terms f_i are sparse. This method is inspired by [1].

4. Solution using Interval Arithmetic

Let $p(\sigma)$ be one of the $n(n - 1)$ polynomials $\frac{\partial f_i}{\partial x_j}(a + \sigma) - \frac{\partial f_i}{\partial x_j}(a)$ of the matrix K . We want to find an interval around $p(\sigma)$ using $[p_1^-, p_1^+]\sigma_1 + \dots + [p_n^-, p_n^+]\sigma_n$ where p_i^- and p_i^+ are lower and upper bounds of $\frac{\partial p}{\partial x_i}(x \in \text{ball}(a, r))$. We can use other range methods but it is necessary that $\|\sigma\| = 0 \Rightarrow p(\sigma) = 0$. The intervals $[p_i^-, p_i^+]$ for $\frac{\partial p}{\partial x_i}(x \in \text{ball}(a, r))$ may be evaluated by a classical IA, for instance the naive one or an affine one if the r value is known. The computation of r gives a value of R , called $v(r)$ because it is depending on r .

Suppose in Example 3.1 that $r = 1$, i. e.

$$x \in B = \text{ball}(a, 1) = ([0, 2], [0, 2], [-1, 1]).$$

Intervals around $\frac{\partial f_1}{\partial x_1}(B)$:

$$\begin{aligned} \frac{\partial f_1}{\partial x_1}(x) &= 3x_1^2 + 3x_2^2 - 6x_1x_2 \quad \text{and} \quad \frac{\partial f_1}{\partial x_1}(a) = 0, \\ \frac{\partial^2 f_1}{\partial x_1^2}(x) &= 6x_1 + 6x_2 \Rightarrow \frac{\partial^2 f_1}{\partial x_1^2}(B) \in [-12, 12]; \\ \frac{\partial^2 f_1}{\partial x_1x_2}(x) &= -6x_1 + 6x_2 \Rightarrow \frac{\partial^2 f_1}{\partial x_1x_2}(B) \in [-12, 12]; \\ \frac{\partial^2 f_1}{\partial x_1x_3}(x) &= 0 \Rightarrow \frac{\partial^2 f_1}{\partial x_1x_3}(B) \in [0, 0]. \end{aligned}$$

Thus $\frac{\partial f_1}{\partial x_1}(B) \in 0 + [-12, 12]\sigma_1 + [-12, 12]\sigma_2$. Similarly for other derivatives. Finally, we obtain

$$\begin{aligned} QN'(a + \sigma) &= - \begin{pmatrix} [-12, 12]\sigma_1 + [-12, 12]\sigma_2 & 0 & 0 \\ [-12, 12]\sigma_1 + [-12, 12]\sigma_2 & [2, 2]\sigma_2 & 0 \\ 0 & [2, 2]\sigma_3 & 0 \end{pmatrix} M \\ &= \begin{pmatrix} 0 & 0 & [-12, 12]\sigma_1 + [-12, 12]\sigma_2 \\ 0 & [1, 1]\sigma_2 & [-12, 12]\sigma_1 + [-12, 12]\sigma_2 \\ 0 & [1, 1]\sigma_3 & 0 \end{pmatrix}, \end{aligned} \tag{6}$$

$$\|QN'(a + \sigma)\| \leq \left\| \begin{pmatrix} 0 & 0 & 24r \\ 0 & r & 24r \\ 0 & r & 0 \end{pmatrix} \right\| = \|(0, 2r, 48r)\|. \tag{7}$$

The n computed polynomials in r have degree 1 and their constant term is ideally null. Taking into account the numerical inaccuracy can make them non null, but they remain of low magnitude, say 10^{-6} . To find the value of r where such a polynomial is equal to $1/2$ is trivial.

By studying the ball $(a, r = 1)$ we find $v(1) = 1/96 = 0.0104166$. As $v(1)$ is much smaller than 1, the value of r , one can think that one may find it beneficial

to reduce the radius r of the studied ball; this will reduce the width of the intervals of $\frac{\partial^2 f_i}{\partial x_j \partial x_k}(B)$ thus will decrease the (upper bound of the) norm of QN' and will make it possible to upper bound $v(r)$.

Thus, if one starts again calculation with $r = 0.25$, one finds $v(0.25) = \frac{1}{24} = 0.04166$. While trying with $r = 0.1$, one would find $v(0.1) = \frac{1}{2 \cdot 4.8} = 0.104166$. Since this value of $v(r)$ goes out of the studied ball, of radius $r = 0.1$, this value is not safe, but however we can deduce that the value of $R = r = 0.1$ is indeed correct.

The bounds of $\frac{\partial f_i}{\partial x_1}$ for each function f_i require the evaluation by the IA of $\frac{1}{2}n(n+1)$ functions, that is to say $O(n^3)$ functions to evaluate to bound the elements of K . Note that a derivative cannot have more monomials than the initial function. When the intervals of the elements of K are calculated, several ways of upper bounding $\|KM\|$ and thus of calculating R , are possible.

We can calculate products KM . It costs $O(n^3)$ products between elements of the matrices K and M , each product costing $O(n)$ floating point operations, which makes $O(n^4)$ floating point operations for the product by M (this estimate is pessimistic since K is often sparse).

One can speed up this part, if a more pessimistic value $v(r)$ is accepted, by using U_1 (6), or even U_2 (7). This is detailed in the following section. Calculating $v(1)$ with U_1 (6):

$$\begin{aligned} \|QN'(a + \sigma)\| &= \left\| \begin{pmatrix} [-12, 12]\sigma_1 + [-12, 12]\sigma_2 & 0 & 0 \\ [-12, 12]\sigma_1 + [-12, 12]\sigma_2 & [2, 2]\sigma_2 & 0 \\ 0 & [2, 2]\sigma_3 & 0 \end{pmatrix} M \right\| & (8) \\ &\leq \left\| (24\sigma_1 + 24\sigma_2, 2\sigma_2 + 2\sigma_3, 0) \begin{pmatrix} |0| & |0| & |-1| \\ |0| & |0.5| & |0| \\ |1| & |0| & |0| \end{pmatrix} \right\| \\ &= \|(0, 2r, 48r)\| = 48r. & (9) \end{aligned}$$

Finally, we get $R = 1/96$; nothing is lost. The cost is $O(n^3)$ functions to evaluate. The cost to go from equation (8) to (9) is $O(n^2)$ operations, the product with M costs $O(n^3)$ operations.

Choosing r : $v(r)$ is a decreasing function of $x \in \mathbf{R}^+$: $v(0) = \infty$ and $v(r)$ is decreasing with r . We want r such that $R = \min(r, v(r))$ is maximum. In other words, one wants r such that $r = v(r)$: it is inevitably the sought optimum R . The simplest is to use the dichotomy. We know r_0 such that $v(r_0) > r_0$: $r_0 = 0$ is correct. One searches then r_1 such that $v(r_1) \leq r_1$:

$$r_1 := 1; \quad \text{while } (v(r_1) < r_1) \quad \text{do } \{r_0 := r_1; r_1 := 2r_1\}.$$

To determine the interval containing R , rather than to start from $[0, 1]$, one could also start from intervals found in the preceding prediction. Then we use dichotomy inside $[r_0, r_1]$ until knowing r_0 with a sufficient relative precision, for example $r_1^{(n)} - r_0^{(n)} < r_1^{(n)}/10$, we choose $R = r_0^{(n)}$. A logarithmic number of evaluations of $v(r)$ is necessary to determine R .

5. Quadratic systems

5.1. The interest. f is quadratic when the total degree of all the monomials of f_i is at most 2. This case is interesting because:

1. Any algebraic system can be reduced to a quadratic system, with the help of the addition of a logarithmic number (using repeated squaring) of variables and equations. For example, a monomial $x_1^3 x_2$ will be replaced by the monomial $y_1 y_2$ where y_1 and y_2 are two auxiliary variables, and by the two quadratic equations $y_1 - x_1^2 = 0$ and $y_2 - x_1 x_2 = 0$.

2. In a quadratic system $f(x) = 0$ all $\frac{\partial f_i}{\partial x_j}(x)$ are constant polynomials, possibly null, and thus independent of the point a : they are calculable once and for all. The matrix K is calculable once and for all, just as an upper bound of $\|K\|$. More precisely

$$K_{ij} = \frac{\partial f_j}{\partial x_i}(a + \sigma) - \frac{\partial f_j}{\partial x_i}(a) = \sum_{k=1}^n \frac{\partial^2 f_j}{\partial x_i \partial x_k} \sigma_k \Rightarrow \|K\| \leq r \|(l_1 \dots l_n)\| = r \times l(f)$$

where $r \geq \|\sigma\|$ and

$$l_i = \sum_{j=1}^n \sum_{k=1}^n \left| \frac{\partial^2 f_j}{\partial x_j \partial x_k} \right| = \sum_{j=1}^n \left| \frac{\partial^2 f_i}{\partial x_j^2} \right| + 2 \sum_{1 \leq j < k \leq n} \left| \frac{\partial^2 f_i}{\partial x_j \partial x_k} \right|.$$

l_i is the “complexity” of f_i and is 0 when $f_i(x)$ is first degree. $l(f) = \max_{i=1}^n (l_i)$ is the complexity of f . The smallest is l , the largest is R . When l is null (in the case of a linear system for example), R is infinite.

5.2. Computation of R . At least three ways of calculating R are possible:

1. If one applies the crudest U_2 , $\|QN'\| \leq \|K\| \cdot \|M\|$ with $\|K\| \leq l(f) \times r$ then $\|QN'\| \leq \frac{1}{2} \Leftrightarrow l(f)r\|M\| \leq \frac{1}{2} \Leftrightarrow r \leq R = \frac{1}{2\|M\|l(f)}$ and one obtains R very quickly. The computing time of $\|M\|$ is $O(n^2)$ and is negligible in front of the calculation of M itself, which is $O(n^3)$, $l(f)$ is a constant, calculated once and for all, in $O(n^3)^1$.

2. If upper bound (6) is applied, it is necessary to multiply the vector

$$(l_1 r \quad l_1 r \quad \dots \quad l_{n-1} r \quad 0)$$

by \hat{M} which is equivalent to multiply

$$(l_1 \quad l_1 \quad \dots \quad l_{n-1} \quad 0)$$

by \hat{M} . $O(n^2)$ floating-point operations are required. The obtained value of R is at least as good as with upper bound (7) and is often better.

3. Lastly, KM can explicitly be calculated. In the worst case, each term of K (except the last column, which is null) is a linear combination of all σ_i and the product KM requires $O(n^4)$ floating point operations. In practice, K is often sparse. This method gives a value of R at least as good as with U_1 (6), generally much better. This method is the slowest of the three. These three ways of calculating R also apply when IA is used. The costs are identical.

¹when $l(f)$ is zero, R is infinite.

5.3. Example:

$$n = 2; \quad f(x) = (x_1^2 + x_2^2 + x_3^2 - 3, x_1^2 + x_2^2 + x_3^2 - 2x_1 - 2x_2),$$

$$f'(x) = \begin{pmatrix} 2x_1 & 2x_1 - 2 \\ 2x_2 & 2x_2 - 2 \\ 2x_3 & 2x_3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2\sigma_1 & 2\sigma_1 & 0 \\ 2\sigma_2 & 2\sigma_2 & 0 \\ 2\sigma_3 & 2\sigma_3 & 0 \end{pmatrix}.$$

Then

$$\|K\| = \|(|2\sigma_1| + |2\sigma_2| + |2\sigma_3|, |2\sigma_1| + |2\sigma_2| + |2\sigma_3|, 0)\| \Rightarrow$$

$$\|K\| \leq \|(6r, 6r, 0)\| = 6r \quad \text{where} \quad (6, 6, 0) = (l_1, l_2, 0).$$

Suppose that $a = (1, 1, 1)$ then $t = (1, -1, 0)$,

$$s'(a) = \begin{pmatrix} 2 & 0 & 1 \\ 2 & 0 & -1 \\ 2 & 2 & 0 \end{pmatrix}, \quad M = \begin{pmatrix} 0.25 & 0.25 & 0 \\ -0.25 & -0.25 & 0.5 \\ 0.5 & -0.5 & 0 \end{pmatrix}.$$

First upper bound:

$$r \leq R = \frac{1}{2\|M\|l(f)} = \frac{1}{2 \cdot 1 \cdot 6} = 0.083333.$$

Second upper bound:

$$\|QN'\| \leq r\|(l_1, l_2, 0)\hat{M}\| = \|(6r, 6r, 0)\hat{M}\|$$

$$= \|(3r, 3r, 3r)\| = 3r; \quad 3r \leq \frac{1}{2} \Leftrightarrow r \leq R = \frac{1}{2 \cdot 3} = 0.166667.$$

Third upper bound:

$$\|QN'\| \leq \|KM\| = \begin{pmatrix} 2\sigma_1 & 2\sigma_1 & 0 \\ 2\sigma_2 & 2\sigma_2 & 0 \\ 2\sigma_3 & 2\sigma_3 & 0 \end{pmatrix} \leq \left\| \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\| = 0r,$$

$$0r \leq \frac{1}{2} \Leftrightarrow r \leq R = \infty.$$

Ideally (without rounding errors due to floating point arithmetic), the method detects that the curve is a straight line, i.e. R is infinite. In practice, with the use of IA to control the rounding errors, $\|KM\|$ will be found not null but very small, for instance $[0, 10^{-6}]r$ which will give finally $R = [5 \cdot 10^{-5}, \infty]$ and $R = 5 \cdot 10^{-5}$.

6. Singular or quasi-singular points

The preceding method no longer applies in the presence of a singular point, or of a quasi singular point, i.e. when the value found for R causes an underflow of the floating-point arithmetic. How to solve, or avoid, this kind of problem?

The simplest solution is to consider the piecewise linear approximation (PLA) of the curve for given parameter μ (i.e. μ is the length of the sides of simplices partitioning \mathbf{R}^n as the real curve). The algebraic formulation is just a shortcut, a

convenience to describe, in fact, its piecewise linear approximation. This principle may be extended from a system of equations to a C.S.G. tree (a binary tree, which nodes are boolean operations). The PLA (of a CSG tree or of a curve) has many good properties:

1. The PLA is mathematically well defined.
2. The topology of the PLA is simple: no singular points.
3. PLA is calculable, and even quickly calculable.
4. The topology of the PLA may differ from that of the algebraic curve, but it has the topology of an infinitesimal perturbation- and desingularized – of this one.
5. Geometrically the PLA is close to the algebraic curve: it is close with a precision of μ .
6. The PLA is piecewise linear: one can resort to exact rational arithmetic, to ensure reliability and consistency of the built boundary representation (for example lazy rational arithmetic).
7. The PLA is a priori enough for the needs of CAD/CAM.
8. The calculation of the PLA can be made perfectly reliable.

Conceptually, the M1 method is thus sufficient to advance in a reliable way along the curve. However, if μ is chosen small to approach closely the curve, M1 becomes terribly slow, even in areas where the curve is very close to a line. Thus, the presented method M2 remains interesting: it makes possible to advance quickly in the calm areas. It is only in presence of a singularity or of quasi singularity, when R is not significantly larger than μ that one resorts to the M1 method.

7. Taking into account rounding errors

Up to now it was supposed that a was exactly on the curve, i.e. $f(a) = 0$, $M = s'(a)^{-1}$. In fact with the numerical inaccuracy inherent to floating-point arithmetic, a is only very close to the curve, and M is only close to $s'(a)^{-1}$. Moreover the floating values calculated for $f(a)$, $s(a)$, $s'(a)$ and so on, are approximations. There are three possible strategies (S1, S2, or S3).

S1: This casual approach is unaware of the problem completely and uses only the usual floating-point arithmetic, hoping that all will occur well. It is probable because the contractivity condition imposed for QN is strong. In this strategy, the terms ideally null either are not calculated (which avoids a possible contradiction between practice and theory), or are calculated but regarded as null because they are too small, in short are ignored. In this strategy, the singularities or quasi singularities on the curve are detected only when R is not sufficiently tall in front of μ .

S2: The paranoiac approach considers that a , t , M , $s(a)$ and $s(x)$, $s'(a)$, $s'(x)$ are intervals, which contain the exact value. The theory of the reliable preceding curve tracing is not modified. This approach is simple and systematic; calculation by

intervals guarantees the results to 100%; certainly the terms ideally null generally cease being it, and there is no problem when they remain sufficiently small, and make it possible to control the effect of the rounding errors. If ever they become too large (for example $0 = r = \|\sigma\|$ that makes it possible to detect simply a singularity or quasi singularity on the curve. Disadvantage: to ensure that the interval of a contains the good value, one needs an additional test, which ensures us that an interval for a contains one zero (simple) of an algebraic system: there already exist such tests, which are provided by Krawckwicz–Moore operators or Segupta–Hansel operators. Another disadvantage is that the IA over-estimates their width largely, so that the convergent iterative methods converge much more slowly, or even diverge. It is possible to find functions f such that $f(x)$ converges for any x in some domain D , but $f(D)$ diverges when it is calculated by intervals (subdividing D is a solution). It may be the same when calculating $QN_k(x^{(i)})$. One can think that this problem will not arise, but a rigorous argument would be preferable.

S3: To consider that a , M , t are specific values (i. e. non-intervals), and are approximations of the correct values, and to modify consequently the theory of the curve tracing: knowing a point a near to the curve, to calculate how much one can advance in a reliable way along the curve. This approach is certainly most painful, the least generalizable, the least systematic, but it has the advantage of employing iteration QN on points and not on intervals, and of avoiding the suspicion. The following approach is developed:

The point a is not exactly on the curve. M is a matrix not exactly equal to $s'(a)^{-1}$ is not exactly such that $tf'(a) = 0$. $\|t\| = 1$ remains true nevertheless ($\|t\|$ is the max norm). CC (3) is written now $\|QN'(a + \sigma)\| \leq \frac{1}{2}$. $\|QN'(a + \sigma)\| = \|Id - s'(a + \sigma)M\|$ (where $M \approx s'(a)^{-1} = M^{-1}M - s'(a + \sigma)M = (s'(a + \sigma) + E)M - s'(a + \sigma)M$ (where $M^{-1} = s'(a + \sigma) + E = (s'(a) - s'(x) + E)M = KM + EM$ where K is the usual matrix and $EM = Id - s'(a)M$. $\|QN'(a + \sigma)\| = \|KM + EM\| \leq \|KM\| + \|EM\| = \|KM\| + \|Id - s'(a)M\|$.

However, $\|Id - s'(a)M\|$ is indeed upper bounded by a floating value $\text{up}(\|Id - s'(a)M\|)$ (where “up” is the upper bound) calculating $Id - s'(a)M$ with naive IA. To guarantee $\|QN'(a + \sigma)\| \leq 1/2$, $\|KM + EM\| \leq \frac{1}{2}$ is thus enough: If R is such that $\|\sigma\| \leq R \Rightarrow \|KM\| \leq \frac{1}{2}$ then to take account of the rounding errors, it is enough to take $\|\sigma\| \leq R' = R - \text{up}(\|Id - s'(a)M\|)$. Of course $\|Id - s'(a)M\|$ is generally negligible.

Note: it would be also possible to get an upper bound of $\|KM + Id - s'(a)M\|$ directly (with IA) and without using an upper bound of $\|KM + Id - s'(a)M\| \leq \|KM\| + \|Id - s'(a)M\|$.

Let us treat now the SC. By being unaware of the rounding errors, one saw that SC is enough to ensure $QN(B) \subset B$ and is equivalent to $k \leq \frac{R}{2}$. One takes again the reasoning in the no ideal case. It is shown at first that $QN(B) \subset B$ always ensures SC by replacing R by R' .

Suppose that $x \in B = \text{ball}(a, R')$ and $\|QN'(B)\| \leq \frac{1}{2}$ and $\|QN_k(a) - a\| \leq \frac{R'}{2}$. Then

$$\begin{aligned} \|QN(x) - a\| &\leq \|QN_k(x) - QN_k(a)\| + \|QN_k(a) - a\| \leq \frac{1}{2}\|x - a\| + \frac{R'}{2}; \\ \|QN_k(x) - a\| &\leq \frac{R'}{2} + \frac{R'}{2} = R' \Rightarrow x \in B. \end{aligned}$$

We want $\|QN_k(a) - a\| \leq \frac{R'}{2}$. Let H be the first $n - 1$ rows of M , V the last row. Then

$$\begin{aligned} QN_k(a) &= a - s(a)M = a - (f(a), -ktt^t) \begin{pmatrix} H \\ V \end{pmatrix} = a - f(a)H + ktt^tV; \\ \|QN_k(a) - a\| &= \|-f(a)H + (ktt^t)V\| \leq \|f(a)H\| + k\|tt^tV\|, \\ \|QN_k(a) - a\| &\leq \frac{R'}{2} \Leftrightarrow k \leq \frac{R'/2 - \|f(a)H\|}{\|tt^tV\|}. \end{aligned}$$

We can effectively upper bound $\|tt^tV\|$ with $\text{up}(\|tt^tV\|)$ and $\|f(a)H\|$ with $\text{up}(\|f(a)H\|)$ by calculating these expressions using IA. We obtain finally the following reliable value for k : $k \leq \frac{R'/2 - \text{up}(\|f(a)H\|)}{\text{up}(\|tt^tV\|)}$. In the ideal case ($f(a) = 0$, $ts'(a) = 0$, $M = s'(a)^{-1}$) the term $\|f(a)H\|$ is zero and $\|tt^tV\|$ is equal to 1.

Really,

$$Id = s'(a)M = (f'(a), t^t) \begin{pmatrix} H \\ V \end{pmatrix} = f'(a)H + t^tV.$$

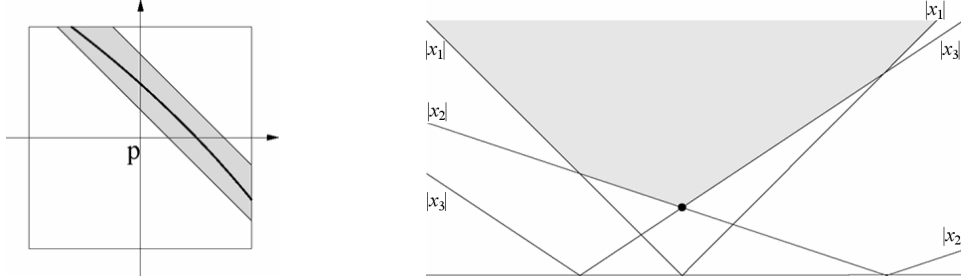
Otherwise $tf'(a) = 0$. Then $tId = tf'(a)H + tt^tV$ and $\|t\| = 1$. Then $\|tt^tV\| = 1^2$.

Obviously, it is interesting that the floating values for a, t and M are as close as possible to the ideal values. If the approximations are too bad, one obtains very small values of k , or even negative values, which means that we cannot advance in a reliable way any more, and that it is necessary to resort to the MI method.

8. Distance from a point to the curve

A related question is as follows. Let p be a point. We want to evaluate the distance between p and the curve. A solution resorts to affine arithmetic of intervals. Let us suppose that we know a radius r of a hypercube around p (a good value for r can be found by dichotomy as we saw before). We calculate $f(p + r\varepsilon)$ where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ by affine arithmetic intervals and we obtain $n - 1$ linear equations $0 = f(p + r\varepsilon) = \varepsilon A + \Delta$ where A is a $n \times (n + 1)$ matrix whose elements are sharp intervals and Δ a vector of n thick intervals. $\varepsilon A + \Delta = 0$ is the equation of a thick line, the thickness being given by the width of the intervals in the vector Δ . The vector ε with smallest Euclidean norm satisfying $\varepsilon A + \Delta = 0$ is $-\Delta(A^tA)^{-1}A^t$. The interval point on the thick line and closest to p is thus $p = -\Delta(A^tA)^{-1}A^t$.

²Note: $V = 1/(tt^t)$.



Left: An arc of curve, and a containing thick line. The center of the square is p and the half side r . The affine IA makes it possible to calculate the thick line. Right: Finding the point on the line $x = (q_1 + \lambda\mu_1, q_2 + \lambda\mu_2, q_3 + \lambda\mu_3)$ which is of minimal norm. The lines represent $|x_1|, |x_2|, |x_3|$

8.1. Example: $n = 3, f(x, y, z) = (x^2 + y^2 - 4, z), p = (1.4, 1.4, 0), r = 0.1$. Then

$$f(p + r\varepsilon) = (\varepsilon_1, \varepsilon_2, \varepsilon_3) \begin{pmatrix} 0.28 & 0 \\ 0.28 & 0 \\ 0 & 0.1 \end{pmatrix} + ([-0.08, -0.06], [0, 0]) = (0, 0). \quad (10)$$

The point $p + \varepsilon$ closest to p (with Euclidean distance) is given by $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = -\Delta(A^t A)^{-1} A^t = ([-0.08, -0.06], [0, 0])(A A^t)^{-1} A^t$.

Skipping tedious computations, the point of the curve nearest to p (in Euclidean distance) is in the box $([1.4071, 1.4071], [1.4071, 1.4071], [0, 0])$ which indeed contains the exact solution $(\sqrt{2}, \sqrt{2}, 0)$. These intervals prove that p is distant by more than 0.01515 from the curve but less than 0.02021. Here the method could obtain one lower bound (not null) distance to the curve. In the general case, the obtained lower bound can be null. It suffices that $f(p)$, calculated by intervals, contains zero, or in an equivalent way that p is inside the thick line. The preceding method extends naturally to calculation of the distance between one point and a surface, a hypersurface, and so on, and of the distance between a point p and another simple point zero of an algebraic system.

Which is the distance between p and the curve, for the max norm? The thick line is expressed in the form $x = q + \lambda\mu$ where q is a thick point, $\lambda \in \mathbf{R}$ is the parameter, along the line, and μ the directing thick line vector. The sought distance is then $d = \min \|x\| = \min(\max(|q_1 + \lambda\mu_1|, \dots, |q_n + \lambda\mu_n|))$ where $x = q + \lambda\mu$ is the only unknown. Each $|q_i + \lambda\mu_i|$ gives two lines $y - (q_i + \lambda\mu_i) = 0$ and $y + (q_i + \lambda\mu_i) = 0$ in the plane (λ, y) . The sought value of d is the coordinate y of the lowest point (i. e. of minimal y coordinate) of the convex described by $y - (q_i + \lambda\mu_i) \geq 0$ and $y + (q_i + \lambda\mu_i) \geq 0$. It is a linear programming problem in 2d. There is a traditional algorithm in $O(n)$ to find this point. The thickness of the point q is easily taken into account: one considers initially $\min(|q_i|)$, then $\max(|q_i|)$ to find one lower and upper bounds of d .

This method is generalizable to the distance (always according to the norm max) between one point and one surface (it is still a linear programming problem), or a hypersurface.

Conclusion

This paper has proposed a reliable prediction correction method for curve tracing. Several ways of computing a safe step parameter were presented, and compared on simple examples. A related question, computing a range of the distance between a point and a curve/surface/hyper surface, was also treated.

Finally, the methods presented here are compatible with all kinds of IA: the naive one, the centered one (to compute a range for $f\left(x_c \pm \frac{\omega}{2}\right)$, the latter computes $f(x_c)$ and $f'\left(x_c \pm \frac{\omega}{2}\right)$ with the naive IA), the interval affine arithmetic of Figueiredo and Stolfi [2], or the “Bernstein IA” used by Hu et al. [4], Garlof [5] (and others). The presented method detects when the current points approaches a singularity (or a almost quasi singularity, i. e. a singularity up to the finite accuracy of the computer). In such a case, one may resort to another method, for instance PLA.

References

- [1] Dedieu J.P., Yakoubsohn J.C. Two seminumerical algorithms for solving polynomial systems / Technical report. – Labo “Approximation et Optimisation”, Univ. Paul Sabatier, Toulouse, France, 1994.
- [2] de Figueiredo L.H., Stolfi J. Adaptive enumeration of implicit surfaces with affine arithmetic // Computer Graphics Forum. – 1996. – Vol. 15(5). – P. 287–296.
- [3] Faux I.D., Pratt J.M. Computational geometry for design and manufacture. – Chichester: Ellis Horwood, 1979.
- [4] Chun-Yi Hu, Takashi Maekawa, Patrikalakis N.M., Xiuzi Ye. Robust interval algorithm for surface intersections // Computer-aided Design. – 1997. – Vol. 29(9). – P. 617–627.
- [5] Garloff J., Graf B. Solving strict polynomial inequalities by Bernstein expansion // The Use of Symbolic Methods in Control System Analysis and Design / N. Munro, Ed. – London: The Institution of Electrical Engineers (IEE), 1999. – P. 339–352.
- [6] Baker R. Kearfott Rigorous Global Search: Continuous Problems. – Dordrecht: Kluwer Academic Publisher, 1996.
- [7] Luo R.C., Ma Y., Mac D.F. Allister Tracing tangential surface-surface intersections // Symposium on Solid Modeling Foundations and CAD/CAM Applications. – 1995. – P. 255–262.
- [8] Neumaier A. Interval Methods for Systems of Equations Encyclopedia of Mathematics and its Applications. 37. – Cambridge: Cambridge Univ. Press, 1990.
- [9] Taubin G. An accurate algorithm for rasterizing algebraic curves // Proc. of the Second Symposium on Solid Modeling and Applications (SMA'93). – 1993. – P. 221–230.

Interval first-derivative-based iterative methods with high order of convergence for solving nonlinear systems of equations

P.S. Senyo, M.V. Hnatyshyn*

Abstract. In the present paper we propose technique of construction and investigation of new effective iterative interval methods for solving nonlinear systems of equations. Two methods of such type and their main features are presented and compared by efficiency. The fields of their preferable application are defined.

Keywords: nonlinear system of equation, interval iterative method, order of convergence.

The whole variety of application of the interval analysis can be conventionally shared into two basic directions: the analysis of mathematical models with uncertainties without taking into account distributions of probabilities of such data and with the known borders of their change and solving of the mathematical models which we receive as a result of approximation of the considered problem by another one, methods of receiving the solution for which are known, that makes the main problem of the applied and, in particular, computational mathematics. Computing algorithms thus frequently do not converge, converge not to the solution of a problem, require information about the unknown solution (about existence, multiplicity, bifurcation, etc.). Besides at realization even entirely determined algorithms on computer some difficulties are generated also by discrete-type structure of its memory. This demands rounding, which even after a small time period of work collects a huge amount.

The second direction of application of the interval analysis frequently is applied to construction of interval methods for solving the systems of nonlinear equations. However here, on the whole, the investigations are limited to construction of various updating of the interval analogues of the Newton method [1]. It is caused by that fact, that the interval estimation of derivatives of supreme orders demands great volume of calculations and decomposition algorithms of interval expansions of functions in Taylor's series are unknown.

In a basis of construction of new interval methods of the supreme orders of convergence for solving nonlinear systems of equations

$$f(x) = 0, \tag{1}$$

where $f: R^k \rightarrow R^k$, $k \in \mathbb{N}$, we put the next ideas [2, 3]:

- Idea of “immersing” of the given problem in the wider class of problems. Among solutions of such problem there are also all solutions of a problem (1)

*Ivan Franko National University of Lviv.

which it is necessary to allocate in special manner (“to make a filtration”) with beforehand established accuracy or, at least with the greatest possible accuracy;

- Idea Runge of approximation with the greatest possible accuracy of supreme order derivatives of mapping $f(x)$ by linear combinations of values of its first derivative in corresponding points;
- Taking into account the behavior of “average” points of residual members in Lagrange form of the generalized Taylor series of mapping $f(x)$ in the case of compression of interval of decomposition into a point and ratio between these points at decomposition in Taylor’s series of mapping $f(x)$ and its first derivative.

Let

$$g_m(x, y) = f(x_n) + (\alpha_1 f'(x_n) + \alpha_2 f'(x_n + \beta_2(x - x_n)) + \dots + \alpha_m f'(x_n + \beta_m(x - x_n)))(y - x_n), \quad (2)$$

where α_i ($i = 1, 2, \dots, m$), β_j ($j = 1, 2, \dots, m$) – real coefficients, $m \geq 2$.

Mappings $f(x)$ and $g_m(x, y)$ satisfy to conditions of their decomposition in Taylor’s series in the corresponding neighborhoods of a point x_n . Coefficients α_i , β_j we shall choose so that decompositions

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2!} f''(x_n)(x - x_n)^2 + \dots + \frac{1}{(p-1)!} f^{p-1}(x_n)(x - x_n)^{p-1} + \frac{1}{p!} f^p(x_n + \theta^{(0)}(x - x_n))(x - x_n)^p, \quad (3)$$

and

$$\begin{aligned} g_m(x, x) &= f(x_n) + (\alpha_1 + \dots + \alpha_m) f'(x_n)(x - x_n) + \\ &(\alpha_2 \beta_2 + \dots + \alpha_m \beta_m) f''(x_n)(x - x_n)^2 + \dots + \\ &\frac{1}{(p-2)!} (\alpha_2 \beta_2^{p-2} + \dots + \alpha_m \beta_m^{p-2}) f^{p-1}(x_n)(x - x_n)^{p-1} + \\ &\frac{1}{(p-1)!} \left(\alpha_2 \beta_2^{p-1} f^p(x_n + \theta_2^{(1)} \beta_2(x - x_n)) + \dots + \right. \\ &\left. \alpha_m \beta_m^{p-1} f^p(x_n + \theta_m^{(1)} \beta_m(x - x_n)) \right) (x - x_n)^p, \quad (4) \\ &\theta^{(0)}, \theta_j^{(1)} \in (0, 1), \quad j = 2, \dots, m, \end{aligned}$$

would coincide with the greatest possible accuracy. It is easy to show [3], that they should be solutions of the system of equations

$$\sum_{i=1}^m \alpha_i = 1, \quad \sum_{i=2}^m \alpha_i \beta_i^s = \frac{1}{s+1}, \quad s = 1, 2, \dots, p-1, \quad (5)$$

$$\begin{cases} Y_n = x_n - \left(\frac{1}{4}f'(x_n) + \frac{3}{4}f' \left(x_n + \frac{2}{3}(X_n - x_n) \right) \right)^{-1} f(x_n), \\ X_{(n+1)} = X_{(n)} \cap Y_{(n)}, \quad n = 0, 1, \dots, \quad x_n = \text{mid}(X_n), \end{cases} \quad (6)$$

which, at carrying out corresponding enough common conditions [2] converges to solution of system (1) and has the order of convergence not less than 3.

If $m = 3$, then, according to the described above techniques for solving of system (1) we receive method [3]

$$\begin{cases} Y_n = x_n - (\alpha_1 f'(x_n) + \alpha_2 f'(x_n + \beta_2(X_n - x_n)) + \\ \alpha_3 f'(x_n + \beta_3(X_n - x_n)))^{-1} f(x_n), \\ X_{(n+1)} = X_{(n)} \cap Y_{(n)}, \quad n = 0, 1, \dots, \end{cases} \quad (7)$$

where $x_n = \text{mid}(X_n)$,

$$\alpha_1 = \frac{1}{9}; \quad \alpha_2 = \frac{16 - \sqrt{6}}{36}; \quad \beta_2 = \frac{6 + \sqrt{6}}{10}; \quad \alpha_3 = \frac{16 + \sqrt{6}}{36}; \quad \beta_3 = \frac{6 - \sqrt{6}}{10}.$$

Method (7) has some effective features, which we shall present in the next theorems.

Theorem 1. *Let mapping $f: R^k \rightarrow R^k$ is twice continuously differentiable and $x^* \in X_0$ where x^* is the solution of system (1). Then*

- a) every interval X_n , $n = 0, 1, 2, \dots$, calculated by (7), contains the solution of system (1);
- b) if all matrices $F'(X_{(n)})$, $n = 0, 1, 2, \dots$, are not singular, then $\lim_{n \rightarrow \infty} X_n = x^*$.

Theorem 2. *Let mapping $f: R^k \rightarrow R^k$ is five times continuously differentiable and $x^* \in X_0$ where x^* is the solution of system (1) and matrices $F'(X_{(n)})^{-1}$, $n = 0, 1, \dots$, exist, then sequence of intervals $\{X_{(n)}\}_{n=0}^{\infty}$ calculated by (7) converges to x^* , moreover $\omega(X_{n+1}) \leq c \cdot (\omega(X_n))^5$, where c is a positive constant.*

Here $F'(X) = \alpha_1 f'(x) + \alpha_2 f'(x + \beta_2(X - x)) + \alpha_3 f'(x + \beta_3(X - x))$, $x = \text{mid}(X)$.

To prove those both theorems, we preliminary have proved the next lemma, which also has the independent importance.

Lemma. *Let mapping $f: R^k \rightarrow R^k$ is twice continuously differentiable and x^* is the real solution of system (1), $x^* \in X_0$ and $x_0 < x^*$. Then, if*

$$\Upsilon_0 \supseteq \left[x_0, x_0 + \frac{1}{k}(x^* - x_0) \right]$$

then

$$f''(x_0 + \theta_2^{(0)}(x^* - x_0))(x^* - x_0)^2 \subset f''(x_0 + k[\theta^{(1)}](\Upsilon_0 - x_0))(\Upsilon_0 - x_0)^2,$$

where $[\theta^{(1)}] \in [0, 1]$, $\theta_2^{(0)} \in (0, 1)$, $k > 0$ is a constant.

Method (7) has advantage over method (6) not only in the order of convergence, but also in that fact, that we must not analyze intermediate intervals if they contain solution x^* of system (1) and it is no necessity to expand them, if they do not contain x^* .

In [3], we have presented methods received by the described above technique for solving system (1) with the order of convergence not less than 7 and 9 correspondingly.

References

- [1] Kalmykov S.A., Shokin Yu.I., Yuldashev Z.C. Methods of interval analysis. – Novosibirsk: Nauka, 1986.
- [2] Senyo P. New approach to construction of interval methods for solving nonlinear systems of equations // *Visnyk Lviv Univ. Ser. Mech.-Math.* – 1989. – No. 31. – P. 85–92.
- [3] Senyo P. Interval methods for solving some classes of determined problems // *Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci.* – 2003. – No. 7. – P. 97–101.

Krawczyk operator revised

Sergey P. Shary*

Abstract. For solution existence tests based on Brouwer fixed point theorem, we propose modifications that make use, first, of the idea of bicedentred interval extension of functions and, second, of the restriction of the test domain to the boundary of the box only. Being applied both separately and in combination with each other, they allow one to substantially increase the efficacy of computational procedures for verified solution of equations systems by interval techniques.

Introduction

In nonlinear analysis, the following fact is well-known

Brouwer fixed point theorem [1]. *Let $D \subseteq \mathbb{R}^n$ be a convex compact set. If $\Phi : D \rightarrow \mathbb{R}^n$ is a continuous function that maps D into itself, i. e.*

$$\Phi(D) \subseteq D, \tag{1}$$

then Φ has a fixed point x^ on D , such that $x^* = \Phi(x^*)$.*

Every system of n equations and n unknown variables

$$F(x) = 0, \tag{2}$$

$F(x) = (F_1(x), F_2(x), \dots, F_n(x))^T$,
 $x = (x_1, x_2, \dots, x_n)^T$, can be reduced to equivalent recurrent form

$$x = \Phi(x) \tag{3}$$

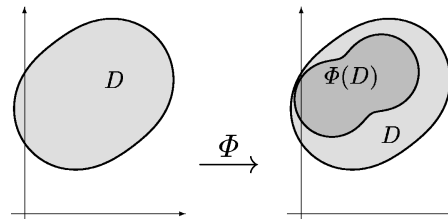


Figure 1. Illustration of Brouwer fixed point theorem

with $\Phi(x) = x - \Lambda F(x)$, where Λ is a nonsingular $n \times n$ -matrix. Because of this, Brouwer fixed point theorem and its generalizations have been often used for proving existence of solutions to equations and systems of equations. These applications were, nevertheless, mainly of theoretical character since finding the image of a set under the action of a mapping is not an easy task. It could be resolved analytically in very few special cases. The situation radically changed after appearance of interval analysis, a mathematical discipline that makes it possible to operate, on computers, sets of small and moderate constructive complexity, through establishing arithmetical and analytical operations, relations, etc., between these sets as individual entities.

*Institute of computational technologies SB RAS.

1. Interval methods for the solution of equations

In our text, the interval notation adheres to the recently adopted project of the international standard [5]. Specifically, we designate intervals and interval objects (vectors, matrices, functions) by boldface letters. \mathbb{IR} stands for classical interval arithmetic [7, 8] or its support, that is, the set of closed intervals of the real axis \mathbb{R} . \mathbb{IR}^n means the set of n -dimensional interval vectors, whose geometric images are axes aligned boxes in \mathbb{R}^n .

In interval analysis, estimating the image of a set under the action of a mapping takes a specific form, being associated with a problem of computing the so-called *interval extension* of a function.

Definition 1 [7, 8]. Interval function $\mathbf{f} : \mathbb{IR}^n \rightarrow \mathbb{IR}^m$ is called interval extension of a real function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, if

- 1) $\mathbf{f}(x) = f(x)$ for any $x \in \mathbb{R}^m$ from the domain of f ,
- 2) $\mathbf{f}(\mathbf{x})$ is inclusion monotonic, i.e., $\mathbf{x}, \mathbf{y} \in \mathbb{IR}^n$, $\mathbf{x} \subseteq \mathbf{y} \Rightarrow \mathbf{f}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{y})$.

Therefore, if $\mathbf{f}(\mathbf{x})$ is an interval extension of the function $f(x)$, then always

$$\{f(x) \mid x \in \mathbf{x}\} \subseteq \mathbf{f}(\mathbf{x}),$$

and we get an outer (by superset) estimate of the range of f over the box $\mathbf{x} \in \mathbb{IR}^n$. Constructing interval extensions of functions is one of the most important problems that interval analysis deals with, and its various aspects have been being under investigation since 1960 up to now. It makes sense to present the first result on the subject, which is often called “the main theorem of interval arithmetic”:

Theorem 1 [7, 8]. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a rational function of the arguments $(x_1, x_2, \dots, x_n) = x$. If, for a box $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$, defined is the result $\mathbf{f}(\mathbf{x})$ of substituting the intervals $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ instead of the arguments of $f(x)$ and further performing the interval arithmetic operations, then

$$\{f(x) \mid x \in \mathbf{x}\} \subseteq \mathbf{f}(\mathbf{x}),$$

that is, $\mathbf{f}(\mathbf{x})$ contains the range of values of the function $f(x)$ on \mathbf{x} .

The interval extension of the rational function $f(x)$, whose construction is described in the main theorem of interval arithmetic, is referred to as *natural interval extension* $\mathbf{f}_{nat}(\mathbf{x})$, and its values can be computed by elementary means. At the same time, using the natural interval extensions often leads to very crude results when estimating the ranges of functions. In this connection, more advanced forms of interval extensions have been developed, and one of the most popular among them is the so-called *centered form*. For a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the centered form of the interval extension looks as follows:

$$f_c(\mathbf{x}, \tilde{\mathbf{x}}) = f(\tilde{\mathbf{x}}) + \sum_{i=1}^n \mathbf{g}_i(\mathbf{x}, \tilde{\mathbf{x}})(\mathbf{x}_i - \tilde{\mathbf{x}}_i),$$

where $\tilde{\mathbf{x}} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ is a “center” point, $\mathbf{g}_i(\mathbf{x}, \tilde{\mathbf{x}})$ are intervals that depend on $\tilde{\mathbf{x}}$ and \mathbf{x} in general. In particular, $\mathbf{g}_i(\mathbf{x}, \tilde{\mathbf{x}})$ may be interval enclosures of the ranges of partial derivatives $\partial f(x)/\partial x_i$ over \mathbf{x} . The interested reader can draw the further information from the books [4, 7, 8], which expound the construction of interval extensions of functions in detail.

If the tools for computing interval extensions of functions are available, we can bypass the difficulties arising in the practical verification of the inclusion (1) in Brouwer fixed point theorem, providing that,

- first, we restrict ourselves to considering the domains D in the form of interval boxes, that is, requiring $D \in \mathbb{IR}^n$, and,
- second, we change the exact range of values of the function Φ over D (it may have a complicated shape) to its outer estimate through interval extension.

Following this way, one can derive the solution existence tests by Krawczyk, Moore, Qi, etc., which are very popular in the modern interval analysis. In the sequel, the first one of these, proposed by West German mathematician Rudolf Krawczyk in 1969 [6], is of special interest for us. It is usually introduced as follows:

Definition 2. Let some rules be defined that assign, to any box $\mathbf{x} \in \mathbb{IR}^n$, a point $\tilde{\mathbf{x}} \in \mathbf{x}$ and a real $n \times n$ -matrix Λ , while interval $n \times n$ -matrix \mathbf{G} is an enclosure for the derivative $F'(x)$ of the function $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$ over the box \mathbf{x} . The mapping

$$\mathcal{K}: \mathbb{IR}^n \times \mathbb{R}^n \rightarrow \mathbb{IR}^n,$$

defined by the rule

$$\mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}}) := \tilde{\mathbf{x}} - \Lambda F(\tilde{\mathbf{x}}) + (I - \Lambda \mathbf{G})(\mathbf{x} - \tilde{\mathbf{x}}),$$

is called interval *Krawczyk operator* for the function F .

Krawczyk operator is nothing but the centered form of the interval extension of the mapping $\Phi(x) = x - \Lambda F(x)$, which emerges in the right-hand side of the equations system (2) after it is reduced to the recurrent form (3).

The following important statements concerning Krawczyk operator are valid [6–8]:

- (i) every solution x^* of the system $F(x) = 0$ within the box \mathbf{x} is also contained in $\mathcal{K}(\tilde{\mathbf{x}}, \mathbf{x})$, so that $x^* \in \mathbf{x} \cap \mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}})$;
- (ii) if $\mathbf{x} \cap \mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}}) = \emptyset$, then the box \mathbf{x} contains no solutions of the equations system $F(x) = 0$;
- (iii) if $\mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}}) \subseteq \mathbf{x}$, then the box \mathbf{x} contains, with certainty, at least one solution of the system $F(x) = 0$;
- (iv) if $\text{int } \mathbf{x} = \{x \in \mathbb{R}^n \mid \underline{\mathbf{x}}_i < x_i < \bar{\mathbf{x}}_i \text{ for every } i\}$ is the interior of the box \mathbf{x} , $\tilde{\mathbf{x}} \in \text{int } \mathbf{x}$ and $\mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}}) \subseteq \text{int } \mathbf{x}$, then the matrix \mathbf{G} is strongly nonsingular and $\mathcal{K}(\mathbf{x}, \tilde{\mathbf{x}})$ contains exactly one solution of the system $F(x) = 0$.

Using the result of the item (i), we can reduce the box which is suspected to have a solution. The item (ii) provides us with an instrument for sifting unpromising boxes that does not have solutions. Conversely, the items (iii) and (iv), which follow from Brouwer fixed point theorem and further fine results, enables one to prove the existence of the solutions and even their uniqueness.

2. Bicentered Krawczyk operator

As far as Krawczyk operator is a centered form interval extension of the mapping $\Phi(x)$ from (3), it is amenable to all the modifications that can be applied to the centered forms in general. A promising way to improve the quality of the enclosures computed by the centered form of interval extensions is varying the center \tilde{x} , and a “final” result in the direction is

Baumann theorem [5, 8]. *Let a “cut-off function” $\text{cut} : \mathbb{R} \times \mathbb{IR} \rightarrow \mathbb{R}$ be defined as*

$$\text{cut}(x, \mathbf{x}) = \begin{cases} \bar{\mathbf{x}}, & \text{if } x \geq \bar{\mathbf{x}}, \\ \underline{\mathbf{x}}, & \text{if } x \leq \underline{\mathbf{x}}, \\ \tilde{\mathbf{x}}, & \text{otherwise.} \end{cases}$$

Let also, for a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$,

$$p_i = \text{cut} \left(\frac{\text{mid}(f'(\mathbf{x}))_i}{\text{rad}(f'(\mathbf{x}))_i}, [-1, 1] \right), \quad i = 1, 2, \dots, n,$$

and n -vectors $\tilde{\mathbf{x}} = (\tilde{x}_i)$, $\hat{\mathbf{x}} = (\hat{x}_i)$ be such that

$$\tilde{x}_i = \text{mid } \mathbf{x}_i - p_i \text{ rad } \mathbf{x}_i, \quad \hat{x}_i = \text{mid } \mathbf{x}_i + p_i \text{ rad } \mathbf{x}_i,$$

where $f'(\mathbf{x})$ is an interval enclosure of gradients for $x \in \mathbf{x}$, mid and rad are mid-point and radius of an interval. Then

- 1) the lower endpoint of the centered form, $\underline{\mathbf{f}}_c(\mathbf{x}, \tilde{\mathbf{x}})$, attains its maximum for $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}$,
- 2) the upper endpoint of the centered form, $\overline{\mathbf{f}}_c(\mathbf{x}, \tilde{\mathbf{x}})$, attains its minimum for $\tilde{\mathbf{x}} = \hat{\mathbf{x}}$,
- 3) the radius of the centered form, $\text{rad } \mathbf{f}_c(\mathbf{x}, \tilde{\mathbf{x}})$, attains its minimum for $\tilde{\mathbf{x}} = \text{mid } \mathbf{x}$.

Baumann theorem naturally begets the so-called *bicentered form* of interval extensions of functions in which we take the intersection of two ordinary centered interval extensions computed with respect to two optimal centers $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$:

$$\mathbf{f}_{bic}(\mathbf{x}) = \mathbf{f}_c(\mathbf{x}, \tilde{\mathbf{x}}) \cap \mathbf{f}_c(\mathbf{x}, \hat{\mathbf{x}}).$$

And the same idea can be applied to Krawczyk operator as well! In doing so, each separate component with the number i should be supplied with its own optimal centers \tilde{x}^i and \hat{x}^i , $i = 1, 2, \dots, n$.

An outline of the algorithm for computing bicertered form of the interval Krawczyk operator looks as follows:

1. Compute “shifts” of the centers

$$p_{ij} = \text{cut} \left(\frac{\text{mid}(\Phi'_{ij}(\mathbf{x}))_i}{\text{rad}(\Phi'_{ij}(\mathbf{x}))_i}, [-1, 1] \right), \quad i, j = 1, 2, \dots, n,$$

where $\Phi'_{ij}(\mathbf{x})$ is an interval enclosure, over \mathbf{x} , of the derivative $\partial\Phi_i/\partial x_j$ of the function $\Phi(x) = x - \Lambda F(x)$.

2. Compute the “shifted centers”

$$\check{x}_j^i = \text{mid } \mathbf{x}_j - p_{ij} \text{ rad } \mathbf{x}_j, \quad \hat{x}_j^i = \text{mid } \mathbf{x}_j + p_{ij} \text{ rad } \mathbf{x}_j, \quad i, j = 1, 2, \dots, n.$$

3. Compute the centered forms proper:

$$\begin{aligned} \mathcal{K}_i(\mathbf{x}, \check{x}^i) &= (\check{x}_i^i - \Lambda F(\check{x}^i)) + (I - \Lambda \mathbf{G})(\mathbf{x} - \check{x}^i), \\ \mathcal{K}_i(\mathbf{x}, \hat{x}^i) &= (\hat{x}_i^i - \Lambda F(\hat{x}^i)) + (I - \Lambda \mathbf{G})(\mathbf{x} - \hat{x}^i), \quad i = 1, 2, \dots, n. \end{aligned}$$

4. Compute the bicertered Krawczyk operator:

$$\mathcal{K}_{bic}(\mathbf{x}) = \mathcal{K}(\mathbf{x}, \check{x}) \cap \mathcal{K}(\mathbf{x}, \hat{x})$$

The statements (i)–(iv) that substantiate application of Krawczyk operator for testing solutions of equations will remain valid for bicertered Krawczyk operator too. However, we should pay a special attention to the choice of the “preconditioning” matrix Λ : it must not be equal to the inverse of the midpoint of the derivative enclosure matrix $F'(\mathbf{x})$, insofar as all the shifts p_{ij} are zeros in this case, and the advantages of the bicertered form are unrealized.

3. Boundary Krawczyk operator

Brouwer fixed point theorem can be substantially strengthened:

Strengthened Brouwer fixed point theorem [3, 9]. *Let $D \subseteq \mathbb{R}^n$ be a convex compact set. If $\Phi : \mathbb{R}^n \supseteq D \rightarrow \mathbb{R}^n$ is a continuous function that maps the boundary ∂D of the set D into D itself, i.e.*

$$\Phi(\partial D) \subseteq D, \tag{4}$$

then Φ has a fixed point x^ on D , such that $x^* = \Phi(x^*)$.*

The above formulation is really stronger than the original Brouwer fixed point theorem since it embraces a wider class of mappings to be verified with respect for

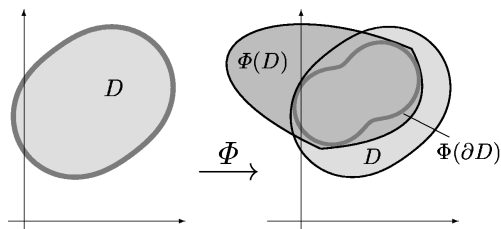


Figure 2. Illustration for strengthened Brouwer fixed point theorem

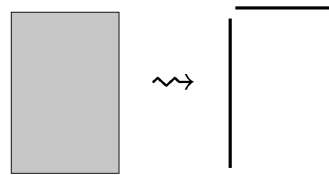


Figure 3. Only boundary to be checked instead of the whole box ...

the existence of the fixed point. It is applicable for functions that may transform D into a set stretching outside D itself, and the function's behavior only on the boundary proves to be essential (Figure 2).

Additionally, when testing the solutions by interval techniques, checking the condition (4) is much more beneficial than (1) from the accuracy reasons. The point is that, for any form of interval extension, the accuracy of the interval evaluation crucially depends on the width of the domain box. Usually, the excess width of the interval enclosure is proportional to a certain power of the width of the box, namely, the first power for the natural interval extension, the second power for centered form, and so on. Meanwhile, the boundary of a box is always “thinner” than the box itself (Figure 3).

As applied to Krawczyk operator in the property (iii), we suffice to check out its action not on the entire box \mathbf{x} , but only on its boundary made up from $2n$ pieces

$$\begin{array}{ll} (\underline{\mathbf{x}}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), & (\bar{\mathbf{x}}_1, \mathbf{x}_2, \dots, \mathbf{x}_n), \\ (\mathbf{x}_1, \underline{\mathbf{x}}_2, \dots, \mathbf{x}_n), & (\mathbf{x}_1, \bar{\mathbf{x}}_2, \dots, \mathbf{x}_n), \\ \vdots & \vdots \\ (\mathbf{x}_1, \mathbf{x}_2, \dots, \underline{\mathbf{x}}_n), & (\mathbf{x}_1, \mathbf{x}_2, \dots, \bar{\mathbf{x}}_n). \end{array}$$

Each one of the above is an $(n - 1)$ -dimensional interval box having both the dimension and width smaller than those of \mathbf{x} . Therefore, we will get a more sensitive variant of the existence test (iii) at the price of some additional labor (approximately $2n$ times larger).

The idea of bicedentred Krawczyk operator may also be implemented to result in further efficiency improvement.

References

- [1] Nikaido H. Convex Structures and Economic Theory. – New York: Academic Press, 1968.
- [2] Baumann E. Optimal centered forms // BIT. – 1988. – Vol. 28. – P. 80–87.
- [3] Kantorovich L.V., Akilov G.P. Functional Analysis: 3rd edition. – Moscow: Nauka, 1984 (in Russian).
- [4] Kearfott R.B. Rigorous Global Search: Continuous Problems. – Dordrecht: Kluwer, 1996.

-
- [5] Kearfott R.B., Nakao M.T., Neumaier A., Rump S.M., Shary S.P., van Hentenryck P. Standardized notation in interval analysis // *Reliable Computing*. – submitted (see <http://www.mat.univie.ac.at/~neum/software/int>).
 - [6] Krawczyk R. Newton-Algorithmen zur Bestimmung von Nullstellen mit Fehler-schranken // *Computing*. – 1969. – Vol. 4. – P. 187–201.
 - [7] Moore R.E. *Methods and Applications of Interval Analysis*. – Philadelphia: SIAM, 1979.
 - [8] Neumaier A. *Interval Methods for Systems of Equations*. – Cambridge: Cambridge University Press, 1990.
 - [9] Opoitsev V.I. *Nonlinear Systemostatics*. – Moscow: Nauka, 1986 (in Russian).

Monitoring of the natural plague focus with interval and fuzzy data using immunocomputing*

S.P. Sokolova, L.A. Sokolova[†]

Immunocomputing (IC) is a new method of solving complex, computationally intensive, real-life problems [1, 2], using a mathematical basis.

This paper provides a further development of the IC approach by applying interval and fuzzy techniques to infection control in natural plague focus. Natural plague foci are present in vast areas. They are characterised by different regulation mechanisms at the population, species, and community levels for the plague triad of carrier, vector and microbe.

The following concepts have been presented:

- the use of singular value decomposition of interval matrices, to illustrate the state of the plague triad at different times;
- techniques of supervised and unsupervised learning on the base interval and fuzzy data, to forecast dynamic epizootic processes at the population level for the area known as the Akdala plane;
- the use of plague risk indices (PRI) that can clearly indicate the risk of plague infection in natural plague foci;
- data fusion of large quantities of variable data relating to this complex, interacting, dynamic system, into a single general value or index, that represents all of those factors;
- the application of data fusion and mathematical models to form PRI on the base interval and fuzzy data using immunocomputing.

References

- [1] Tarakanov A.O., Skormin V.A., Sokolova S.P. Immunocomputing: Principles and Applications. – New York: Springer, 2003. – 193 p.
- [2] Sokolova S.P., Abdyllina V.Z., et al. Immune system for the gerbil natural plague focus / Ed. A.O. Tarakanov. – Almaty: PS, 2002. – 180 p.
- [3] Sokolova L. Index design by immunocomputing // Lect. Notes in Comput. Sci. – Berlin: Springer, 2003. – Vol. 2787. – P. 120–127.

*Supported by the EU Commission under INCO-contract No. ICA2-CT-2000-10048 “The plague of Central-Asia an epidemiological study focusing on space-time dynamics” and EOARD within the ISTC Project 2200p “Development of mathematical models of immune networks intended for information security assurance”.

[†]St. Petersburg Institute for Informatics and Automation.

Авторский указатель

Азизов Р.К.	34	Пушной Б.М.	110
Алексеев А.С.	19	Пятых Д.С.	258
Базаров М.Б.	168	Рогалев А.Н.	119, 240
Бычков И.В.	38	Селихов А.В.	19
Воцинин А.П.	147	Соколов В.А.	73, 126
Гаченко А.С.	38	Урбанович Д.Е.	38
Глинский Б.М.	19	Фазлиев А.З.	34
Городилов В.В.	41	Федотов А.М.	3, 132
Джаныбеков Б.С.	171	Филиппенко Е.Г.	255
Добронец Б.С.	179	Хлебалин Н.А.	258
Долгов Ю.Г.	184	Цициашвили Г.Ш.	225
Ерохин В.И.	193	Чалый Д.Ю.	126
Ершов А.Г.	201	Шайдуров В.В.	27
Жижимов О.Л.	45	Шарая И.А.	267
Жилин С.И.	210	Шарый С.П.	171
Завертайлов А.В.	56	Шокин Ю.И.	3
Залеская А.А.	41	Яковченко К.Н.	66
Зюзин В.С.	255	Akhmerov R.R.	271
Исаев С.В.	27	Faudot D.	289
Калинкина С.Ю.	219	Herrero P.	272
Кашеварова Т.П.	201	Hnatyshyn M.V.	303
Киншт Н.В.	225	Jaulin L.	272
Ковалев С.П.	56, 66, 95	Kearfott R.B.	159
Котелевский С.П.	19	Kubica B.J.	280
Кузьмин Е.В.	73	Malinowski K.	280
Кучин Н.В.	19	Men'shikov G.G.	285
Лемперт А.А.	38	Michelucci D.	289
Мазов Н.А.	45	Sainz M.A.	272
Мальшев А.В.	27	Senyo P.S.	303
Мальшкин В.Э.	19	Shary S.P.	307
Манкевич П.В.	231	Sokolova L.A.	314
Молородов Ю.И.	86	Sokolova S.P.	314
Новоселов Д.Ю.	95	Vehí J.	272
Петрунько Н.Н.	225		
Попов С.Е.	103		
Привезенцев А.И.	34		
Пушков С.Г.	219		