СИБИРСКИЙ МАТЕМАТИЧЕСКИЙ ЖУРНАЛ

Том III, № 5 Сентябрь — Октябрь 1962 г.

Л. В. КАНТОРОВИЧ

О НЕКОТОРЫХ НОВЫХ ПОДХОДАХ К ВЫЧИСЛИТЕЛЬНЫМ МЕТОДАМ И ОБРАБОТКЕ НАБЛЮДЕНИЙ*.

Введение

Имевшие место сдвиги в развитии математики и вычислительных средств должны иметь следствием коренные изменения в технике, а возможно и теории численных методов и обработки наблюдений. В той или иной форме отдельные высказываемые ниже соображения встречались в литературе, но не разрабатывались систематически. В частности, мы считаем, что существенное значение имеют следующие моменты:

1. Бо́льшая ответственность за результаты расчетов, на которых сейчас нередко базируются решения, касающиеся сложных дорогостоящих объектов современной физики и техники, наличие больших не наблюдаемых этапов при машинных вычислениях повышают требования к надежности окончательных и промежуточных данных, получаемых в процессе применения численных методов и при обработке данных наблюдений. Это обуславливает систематический переход от построения приближенных значений и результатов, к получению точных двухсторонних границ для искомых величин или, если говорить о нечисловых величинах, областей расположения искомых и наблюдаемых величин; иначе говоря возникает задача возможно более точного описания расположения этих величин в соответствующих пространствах их значений. Идеи теорий полуупорядоченных пространств и операций в них, а также некоторых других абстрактных систем объектов дают определенную теоретическую базу для реализации этой точки зрения.

2. В качестве основного аппарата в численных алгорифмах, после сведения задачи к конечно-мерной, до сих пор служили системы линейных (иногда и нелинейных) алгебраических уравнений и аналитический аппарат линейной алгебры в целом. Широко использовались также итеративные процессы характера разностных уравнений. Развитый в связи с экономической проблематикой новый математический аппарат (линейное, нелинейное и динамическое программирование) делает возможным систематическое использование в численных методах, с не меньшей эффективностью, систем линейных неравенств и новых типов итеративных процессов. Этот аппарат, в частности, существен при решении поставленной выше задачи о построении двусторонних приближений и характеристики области расположения решений. В ча-

* Работа представляет несколько дополненный текст докладов, прочитанных в мае 1962 года в Ленинградском и Новосибирском университетах и в Московском математическом обществе.

стности, он дает возможность эффективного оперирования с многогранными областями в конечномерных пространствах, описывающими характерные области расположения объектов векторного типа.

3. Систематическое использование в численных методах, при нахождении определяемых величин, и в обработке наблюдений, возможно полной (количественной и качественной), а иногда и избыточной информации о данном объекте, может существенно способствовать уточнению границ расположения объекта и его количественных характеристик. В ряде случаев ограниченное, неполное использование имеющейся информации вызывалось стремлением сократить и упростить расчеты и обработку данных, а также недостатки используемого для этих целей традиционного математического аппарата, создавшегося в другое время, в других условиях и требованиях, для других объектов. Имеющийся новый математический аппарат, а также малая трудоемкость вычислительных работ при использовании электронных машин по сравнению со сбором информации, производством наблюдений и статистических выборок, делают осуществимой и оправданной гораздо более тщательную и полную обработку информации, позволяя меньше считаться с объемом вычислительной работы.

В частности, в ряде случаев существенное использование в численных алгорифмах может найти полученная теоретическим путем информация о расположении и свойствах решения (границы самого решения и его производных и т. п.). В результате, как это уже не раз имело место в прошлом, в частности, в связи с применением функционального анализа в теории приближенных методов, для численных методов приобретает значение ряд теоретических результатов теории функций, функционального анализа, теории уравнений математической физики. При этом на сей раз они проникают еще более глубоко в численные методы, не только в теорию их исследование сходимости, оценки, общий качественный анализ методов, но и в самую структуру численных алгорифмов.

4. Наконец обратим внимание еще на несколько вопросов, связанных с влиянием машинной техники на развитие численных методов.

Широкое использование современной вычислительной техники и опыт решения с ее помощью объемных вычислительных задач привели к переоценке различных численных методов. Некоторые из них оказались скомпрометированными и отвергнутыми при этой проверке (неустойчивость, плохая обусловленность). Представляется, что такое доверие к «выводам» машинной техники в данном вопросе является неосновательным, т. к. при этой машинной проверке не учтены доступные человеку и постоянно применявшиеся при вычислениях вручную возможности внесения различных модификаций в форму применения методов, их улучшения и контроля результатов в процессе счета, позволяющие устранить подобные недостатки. Иначе говоря, не учтено, что с помощью надлежащей модификации соответствующих методов, основанной, в частности, на тех же, указанных выше идеях и средствах, некоторые из этих методов допускают реабилитацию.

Машинное проведение объемных задач, связанное с необходимостью систематического внесения вычислительных погрешностей, делает

постоянный их учет при выборе методов организации вычислительного плана и в самом процессе вычислений, важным и совершенно необходимым элементом численного анализа. В частности, это делает часто малоприемлемыми многие традиционно использовавшиеся формы записи математических выражений и осуществления преобразований, т. к. они в ряде случаев оказываются не безобидными (скажем упрощение многочлена). В то же время сами средства машинной математики и уже разработанные приемы ее использования для описания, хранения и обработки математической информации (машинный математический язык, например, схемы, величины, списки и пр.), открывают возможность систематического использования других, нетрадиционных форм записи математических выражений.

Все сказанное приводит к выводу о целесообразности ревизии всей данной области с точки зрения высказанных общих установок.

Такой пересмотр потребует проведения ряда исследований и представляет дело будущего. Мы не ставим задачей в этой статье дать даже основы такой новой теории, а хотим только на нескольких примерах проиллюстрировать те возможности, которые дают эти новые подходы.

Приводимые ниже конкретные примеры, иллюстрирующие эти общие положения, взяты по преимуществу из близких мне областей численного анализа на основании, главным образом, опыта, накопленного в практике применения численных методов в Ленинградском Отделении Математического института Академии наук СССР.

§ 1. Первые иллюстративные примеры

1. Начнем с самой элементарной задачи вычисления значения полинома в данной точке. Она становится, однако, совсем нетривиальной, если поставим ее, например, по отношению к полиному Чебышева с большим номером, скажем, $T_{51}(x)$, нормированному, со старшим коэффициентом 1. Его отклонение от нуля равно $\frac{1}{2^{n-1}} = \frac{1}{2^{50}} \approx 10^{-15}$.

При вычислении его значения мы будем иметь слагаемые порядка 1. Если мы производим вычисления с десятью знаками, то в полученном результате должно было бы пропасть 15 знаков, но фактически знаки с 11-го по 15-й должны остаться, то есть первые пять цифр в результате все неверные. Ошибка примерно в 10^5 превосходит результат! Конечно, полином Чебышева можно сосчитать достаточно точно иным путем (например, по формуле $T_n(x) = \frac{1}{2^{n-1}} \cos n \arccos x$). Такие возможности имеются и для других специальных полиномов. Однако, если дан произвольный, мало отклоняющийся от нуля полином высокой степени, то вычисление его значения представляется безнадежным. Если при этом коэффициенты известны с 10-ю значащими цифрами, то это нетрудно доказать. Действительно, в этом случае в пространстве коэффициентов мы имеем не точку, а $n+1$-мерный куб с размерами 10^{-10}, поэтому значение полинома в точке, представляющее линейную форму от коэффициентов, фактически будет меняться в пределах порядка 10^{-10} и следовательно, действительные значащие цифры многочлена получить невозможно.

Однако положение меняется, если нам известна дополнительная информация о данном многочлене. Скажем, то что он отклоняется от нуля не более, чем на $3-10^{-15}$, производная его или неопределенный интеграл не превосходят таких-то границ и т. п. Тогда уже далеко не всякая точка в упомянутом кубе пространства коэффициентов определит полином, который удовлетворяет указанным условиям. Область в пространстве коэффициентов сужается, превращается из куба в тонкий многогранник, в связи с чем значительно сужаются границы линейной формы, определяющей значение полинома, и оно может быть определено значительно точнее. Аналитически задача определения границ этой формы приводит к нахождению максимума и минимума линейной формы

$$L(a) = \sum_{i=0}^{n} a_i x_0^i \tag{1}$$

при следующих условиях:

$$\underline{a_i} \leqslant a_i \leqslant \overline{a_i}, \quad i = 0, \ldots, n,$$

$$\left| \sum_{i=0}^{n} a_i x_j^i \right| \leqslant C, \quad j = 1, \ldots, m,$$

где $\underline{a_i}$, $\overline{a_i}$ — границы коэффициентов, C — граница полинома, $x_j (j = 1, \ldots, m)$ — сетка точек, в которых записаны ограничения. Задача нахождения максимума или минимума линейной формы при линейных ограничениях есть задача линейного программирования, для решения которой имеется ряд эффективных методов.

Пример.

$$\frac{\pi}{3} T_5(x) = \frac{\pi}{3} (5x - 20x^3 + 16x^5),$$

$$P_5(x) = a_1 x + a_3 x^3 + a_5 x^5,$$

$$0{,}327 \leqslant a_1 \leqslant 0{,}328$$

$$-1{,}31 \leqslant a_3 \leqslant -1{,}30$$

$$1{,}047 \leqslant a_5 \leqslant 1{,}048$$

Границы значения $P_5(-0{,}37)$, вычисленные по схеме Горнера:

$$-0{,}062778347 \leqslant P_5(-0{,}37) \leqslant -0{,}061894882,$$

с использованием аппарата линейного программирования:

$$-0{,}062287676 \leqslant P_5(-0{,}37) \leqslant -0{,}061899476$$

2. Задача интерполирования также может рассматриваться как задача нахождения значения полинома в данной точке, принимающего значения точные или приближенные в некоторой системе точек.

Относительно коэффициентов полинома это опять задача нахождения границ линейной формы при некоторых условиях.

Задача интерполирования может ставиться и иначе, когда заранее не выбирается форма интерполирующей функции, а именно, на основании

границ функции в заданной сетке точек разыскиваются значения функции в некоторой более густой сетке при некоторых требованиях гладкости (границы для I-х, II-х, III-х разностей).

При нахождении границ искомой функции в более густой сетке нам вновь приходится решать задачу линейного программирования.

3. При решении систем линейных алгебраических уравнений наиболее часто применяются те или иные разновидности метода исключения Гаусса, при этом точность получаемых решений существенно зависит не только от степени обусловленности системы, но и от порядка, в котором ведется исключение. С нашей точки зрения более эффективным представляется другой порядок вычислений. Полученные в процессе исключения уравнения с учетом вносимых при этом погрешностей вычислений должны записываться в виде неравенства для линейных форм от верхних и нижних границ неизвестных. Поэтому в результате счета должны получаться строго устанавливаемые границы для неизвестных. Если получаются границы удовлетворительные, то процесс решения можно считать законченным. Если они не удовлетворительны, то можно пытаться строить более точные оценки за счет изменения порядка операций.

§ 2. Принципы исчисления двусторонних границ

Рассмотрения естественно вести в полуупорядоченном пространстве. В дальнейшем, как правило, имеется в виду, что X — линейное полуупорядоченное пространство или множество. Величина x задана приближенно, если указаны некоторые множества ее верхних и нижних границ, $\underline{x} \leqslant x \leqslant \overline{x}$.

Под $\underline{x}$ и $\overline{x}$ будем понимать как отдельных представителей этих множеств границ, так и сами эти множества. Будем обозначать их также $\underline{x} = \underline{\Lambda}(x)$, $\overline{x} = \overline{\Lambda}(x)$, под $\Lambda(x)$ будем разуметь любую из этих границ, но одну и ту же в пределах одного соотношения. Отметим некоторые очевидные свойства этих границ

$$\underline{\Lambda}(-x) = -\overline{\Lambda}(x),\ \overline{\Lambda}(-x) = -\underline{\Lambda}(x). \tag{1}$$

Введем оператор σ, определяемый соотношением

$$\sigma\underline{\Lambda} = \overline{\Lambda},\ \sigma\overline{\Lambda} = \underline{\Lambda},$$

тогда, очевидно, $\sigma^2 = I$, где I — тождественный оператор. Пользуясь этим обозначением, предыдущее соотношение можно записать в виде

$$\Lambda(-x) = -\sigma\Lambda(x),$$

где $\Lambda = \underline{\Lambda},\ \overline{\Lambda}$.

$$\Lambda(x_1 + x_2) = \Lambda(x_1) + \Lambda(x_2). \tag{2}$$

Это следует из правила сложения неравенств.

$$\Lambda(cx) = c\sigma^{\operatorname{sign} c}\Lambda(x), \tag{3}$$

где $c = \text{const}$, а $\operatorname{sign} c = \begin{cases} 1, & c < 0 \\ 0, & c > 0. \end{cases}$

Пусть для элементов пространств X_1 и X_2 определено произведение $x_3 = x_1 \cdot x_2$, принадлежащее некоторому полуупорядоченному пространству X_3, так что выполнены обычные свойства произведения. Тогда для границ произведения имеем следующую формулу:

$$\Lambda(x_1 \cdot x_2) = \tilde{x}_1 \cdot \tilde{x}_2 + |\tilde{x}_1| \Lambda(\alpha_2) + |\tilde{x}_2| \Lambda(\alpha_1) + \Lambda(\alpha_1 \cdot \alpha_2), \tag{4}$$

где $\tilde{x}_i = \frac{\overline{x}_i + \underline{x}_i}{2}$,

$$d_i = \frac{\overline{x}_i - \underline{x}_i}{2},$$

$$\underline{\Lambda}(\alpha_i) = -d_i, \ \overline{\Lambda}(\alpha_i) = d_i, \ i = 1, 2,$$

$$\underline{\Lambda}(\alpha_1\alpha_2) = -d_1 d_2,$$

$$\overline{\Lambda}(\alpha_1\alpha_2) = d_1 d_2.$$

Если в пространстве X для некоторых элементов определена обратная величина $\frac{1}{x}$, то имеем

$$\Lambda\left(\frac{1}{x}\right) = \Lambda\left(\frac{1}{x^+ - x^-}\right) = \frac{1}{\sigma\Lambda(x^+) + \sigma\Lambda(x^-)}, \tag{5}$$

где x^+, x^- — положительные и отрицательные части элемента x, $a\Lambda(x^+)$ и $\Lambda(x^-)$ предполагаются дизъюнктными. Отсюда получается и граница для частного

$$\Lambda\left(\frac{x_1}{x_2}\right) = \frac{1}{\underline{\Lambda}(x_2) \cdot \overline{\Lambda}(x_2)} \Lambda(x_1 \cdot x_2).$$

В частности, формула границ произведения может быть применена для случая, когда один из множителей есть x, а другой линейный оператор A из пространства X в другое полуупорядоченное пространство Y.

Поскольку этот оператор A считается точно известным, границы его совпадают с ним, то из формулы границ произведения получаем

$$\Lambda(Ax) = A\tilde{x} + |A| \Lambda(\alpha),$$

или иначе говоря

$$\Lambda(Ax) = A\frac{\overline{x} + \underline{x}}{2} + |A|\frac{\overline{x} - \underline{x}}{2} = \frac{1}{2}(A + |A|)\overline{x} +$$

$$+ \frac{1}{2}(A - |A|)\underline{x} = A_+\overline{x} - A_-\underline{x}.$$

Аналогично строится граница для $\underline{\Lambda}(Ax)$. Отсюда легко строятся оценки для решения уравнения

$$Ax = b,$$

если известен обратный оператор A^{-1}; тогда

$$x = A^{-1}b,$$

а поэтому

$$\overline{x} = [A^{-1}]_+\overline{b} - [A^{-1}]_-\underline{b}.$$

Для $\underline{x}$ имеем аналогичную формулу.

Некоторые аналогичные определения могут быть даны и для случая нелинейных полуупорядоченных пространств и нелинейных операторов с

использованием вводившихся нами в свое время мажорантного оператора для данного нелинейного. Мы не будем здесь подробно останавливаться на этом.

Отметим еще, что в случаях, когда обращение оператора невозможно или обратный оператор очень велик, например в окрестности собственного значения, целесообразно не строить границы областей, содержащих решения, а оценивать область расположения решения, используя технику линейного программирования.

Замечание. Во всех формулах для границ предполагалось, что действия над исходными границами производятся точно. Если эти действия производятся приближенно, например на машине, то формулы видоизменяются за счет дополнительного введения погрешностей этих действий. Не будем приводить записи формул для этого случая.

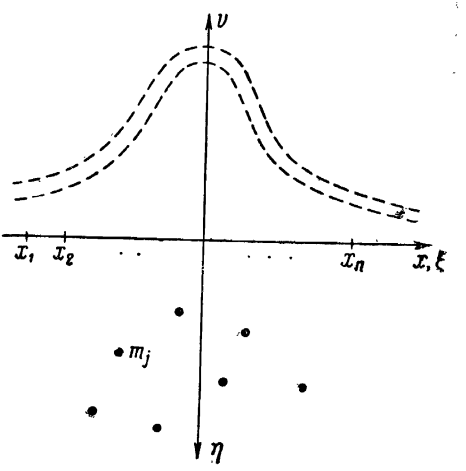
§ 3. Некоторые задачи прикладной математики

1. Задача обработки наблюдений. Обычно полученную в результате измерений избыточную систему уравнений обрабатывают по методу наименьших квадратов Гаусса. При этом происходит значительная потеря информации. По-видимому, в настоящее время более целесообразна другая техника. Уравнения, связывающие искомые величины, выписать с учетом погрешностей в форме неравенств

$$l_i - \delta \leqslant \sum_{k=1}^{n} c_{ik} x_k \leqslant l_i + \delta,$$

$$i = 1, \ldots, m,$$

и разыскивать возможные границы для x_k методами линейного программирования.



Черт. 1

2. Обратная задача теории потенциала*. По измеренным значениям гравитационного, магнитного или иного потенциала в ряде точек (рассматриваем для простоты плоский случай) нужно дать заключение о рудном теле, нарушающем поле.

Пусть m_j — сосредоточенная неизвестная масса в точке

$$(\xi_j, \eta_j), \quad j = 1, \ldots, n.$$

Тогда теоретически вычисленное значение потенциала в точке x_i будет

$$v_i^0 = \sum_{j=1}^{n} \frac{\gamma \eta_j}{(x_i - \xi_j)^2 + \eta_j^2} m_j = \sum_{j=1}^{n} a_{ij} m_j,$$

$$i = 1, \ldots, m,$$

* Задача возникла в связи с сообщением по линейному программированию, которое автор делал по предложению Э. Э. Фотиади в Институте геологии и геофизики Сибирского Отделения Академии наук СССР в марте 1962 г.

и известно, что измеренные значения v_i отклоняются от истинных не более чем на δ, то есть имеем систему линейных неравенств:

$$\left|\sum_{j=1}^{n} a_{ij}m_j - v_i\right| \leqslant \delta_i,\ i = 1, \ldots, m,$$

$$m_j \geqslant 0,\ j = 1, \ldots, n.$$

Ясно, что сами массы m_j определить невозможно. Однако некоторые функционалы от них определяются довольно удовлетворительно для тех случаев, когда точными значениями v_i они определяются однозначно. В других случаях необходимо ввести дополнительные ограничительные гипотезы на распределение масс, обеспечивающие такую однозначность. Таким образом, открывается возможность применения при решении этого класса задач ранее не применявшейся техники линейного, а также целочисленного и нелинейного программирования. Более определенное заключение об эффективности этих методов в данном вопросе может быть сделано на основе дальнейших исследований, экспериментов. Приведем один пример расчета, выполненный на ЭВМ ИМ СО АН СССР.

Пример. Измерения были произведены в 70 точках с точностью $\delta = 0{,}1$, m_j рассматривались в 36 точках (истинная масса = 1 и заполняет единичный квадрат с координатами центра тяжести 0; 1,5). В результате вычислений с помощью аппарата линейного программирования получили:

$$\max \sum_{j=1}^{36} m_j = 1{,}02, \quad \min \sum_{j=1}^{36} m_j = 0{,}98;$$

$$\max \sum_{j=1}^{36} \xi_j m_j = 0{,}11, \ \min \sum_{j=1}^{36} \xi_j m_j = -0{,}11;$$

$$\max \sum_{j=1}^{36} \eta_j m_j = 1{,}75, \ \min \sum_{j=1}^{36} \eta_j m_j = 1{,}28.$$

§ 4. Пути улучшения некоторых численных алгорифмов

1. Методы, приводящие к плохо обусловленным линейным системам. Целый ряд численных методов — методы Ритца и Галеркина, приведения к обыкновенным дифференциальным уравнениям, метод коллокаций, наискорейшего спуска весьма эффективных при нахождении первых приближений, оказываются мало эффективными при нахождении приближений более высокого порядка. При этом, в случае гладкости задачи теоретически устанавливается наличие быстрой сходимости к решению, и следовательно, существования хорошего приближения к решению в принятой форме.

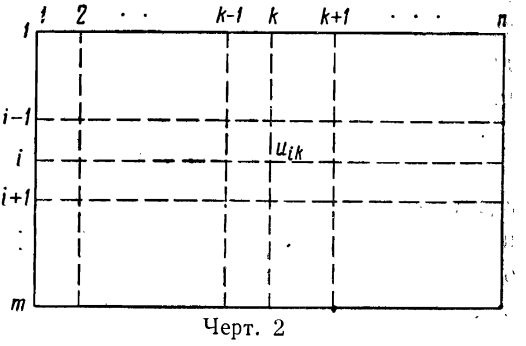
Однако для нахождения коэффициентов или других параметров получается плохо обусловленная система, которая практически неразрешима или приводит к результатам, которые не обеспечивают удовлетворительную аппроксимацию действительного решения задачи. Представляется, что эта трудность может быть снята, если использовать дополнительную информацию о решении — ограниченность, гладкость, дополнительные избыточные соотношения, характеризующие его. Это

существенно сократит область возможных значений параметров и позволит с гораздо большей точностью определить само решение или те или иные характеристики. В этом случае вновь система уравнений заменяется системой неравенств и требуется применение техники линейного программирования.

2. Методы, приводящие к неустойчивым схемам вычислений. Многие численные (разностные) методы решения дифференциальных уравнений оказались неэффективными в силу явления неустойчивости. Простейшей задачей такого рода является задача о нахождении ограниченного на бесконечности решения дифференциального уравнения 2-го порядка, у которого имеется другое — быстро растущее решение.

При численном решении неизбежно входит эта 2-я компонента, быстро искажая полученные результаты. Для устранения этих явлений предлагались различные приемы: интегрирование с другого конца, понижение порядка, факторизация, пристрелка. В связи с этим, в ряде случаев вообще отказывались от явных схем, переходя к неявным (Нейман, Ладыженская), идя, для обеспечения устойчивости, на значительное усложнение вычислительной схемы.

Нам представляется, как вполне реальный, иной путь преодоления тех же трудностей, именно, пополнение системы разностных уравнений системой ограничительных неравенств для неизвестных с надлежаще разработанной численной схемой для этой обогащенной системы. Тут могут применяться те или иные итеративные процессы, последовательно уточняющие двусторонние границы для искомых; по отношению к отдельным блокам неизвестных, а иногда и ко всей системе может применяться техника линейного программирования.



Черт. 2

Проиллюстрируем сказанное на простейшем примере решения задачи Дирихле для уравнения Лапласа в случае прямоугольника. В случае, если мы введем в качестве неизвестных значения решения конечно-разностного уравнения на линии, прилежащей к боковой стороне прямоугольника и с помощью исключения по формулам типа

$$u_{i,3} = 4u_{i,2} - u_{i,1} - u_{i-1,2} - u_{i+1,2}$$

будем пытаться последовательно исключать неизвестные, то полученная схема будет неустойчивой и не приведет к удовлетворительным результатам.

Однако, если мы дополним разностные уравнения системой неравенств вида $m \leqslant u_{i,k} \leqslant M$, где m, M границы контурной функции, то полученная система может эффективно решаться. Например, методами линейного программирования можно последовательно получать двусторонние границы для коэффициентов влияния элементов второго ряда на элементы k-го ряда.

Поступило
2. VII. 1962