С. В. Курочкин, О свойствах устойчивости методов дифференциальной прогонки, *Ж. вычисл. матем. и матем. физ.*, 2000, том 40, номер 11, 1611–1614

ЖУРНАЛ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКОЙ ФИЗИКИ, 2000, том 40, № 11, с. 1611–1614

УДК 519.624.2

О СВОЙСТВАХ УСТОЙЧИВОСТИ МЕТОДОВ ДИФФЕРЕНЦИАЛЬНОЙ ПРОГОНКИ[1)]

© 2000 г. С. В. Курочкин

(117967 Москва, ГСП-1, ул. Вавилова, 40, ВЦ РАН)

Поступила в редакцию 27.07.99 г.
Переработанный вариант 27.04.2000 г.

Исследуется вопрос о том, какими свойствами устойчивости к малым возмущениям входных данных должна обладать краевая задача для системы линейных обыкновенных дифференциальных уравнений, чтобы ее можно было решать каким-либо вариантом метода дифференциальной прогонки.

Методы дифференциальной прогонки (см. [1]–[3, гл. IX]) давно зарекомендовали себя как хорошее средство численного решения краевых задач для систем линейных обыкновенных дифференциальных уравнений, особенно в тех случаях, когда коэффициенты уравнений содержат большие параметры, особенности на концах интервала и т.п. Несмотря на сохраняющиеся различия в оценках (вплоть до решительной критики в [4, с. 602]), в целом этот класс методов положительно принят специалистами и вошел в их инструментарий, что зафиксировано уже и в учебниках. Безусловно, решающую роль в этом сыграла 40-летняя успешная практика применения методов дифференциальной прогонки к решению конкретных задач. Что же касается формального обоснования применимости методов этого класса, то в совокупности строгих результатов остается существенный пробел, устранить который и имеет целью настоящая заметка. В центре внимания при этом будет следующий вопрос: какие именно свойства краевой задачи “отвечают” за успешность применения к ней метода дифференциальной прогонки.

1. Рассмотрим краевую задачу общего вида для системы линейных обыкновенных дифференциальных уравнений:

$$\mathbb{y}'(x) = A(x)\mathbb{y}(x) + \mathbb{f}(x), \quad a \le x \le b, \tag{1}$$

$$\varphi_a \mathbb{y}(a) = \gamma_a, \tag{2а}$$

$$\varphi_b \mathbb{y}(b) = \gamma_b. \tag{2б}$$

Здесь $\mathbb{y}(\cdot)$ есть n-мерный вектор-решение, $A(\cdot)$ и $\mathbb{f}(\cdot)$ – зависящие от x $(n \times n)$-матрица и n-вектор, φ_a и φ_b суть $[(n-k)\times n]$- и $(k \times n)$-матрицы полного ранга, $0 < k < n$, γ_a и γ_b – векторы соответствующего размера. Не ограничивая общности, будем считать, что строки матрицы φ_a ортонормированы, и то же – для φ_b. Краевое условие (2а) задает k-мерную (аффинную) плоскость в $\mathbb{R}^n$, а уравнение (1) “переносит” это условие во все точки отрезка $[a, b]$; аналогично – для условия (2б). В методах дифференциальной прогонки этот перенос осуществляется явным образом (решением двух задач Коши для специальных нелинейных прогоночных уравнений), при этом, грубо говоря, различные варианты метода прогонки отличаются друг от друга способом записи подпространства. Значение решения (1), (2) в произвольной точке $x \in [a, b]$ затем находится из системы линейных алгебраических уравнений

$$\begin{pmatrix}\varphi_1(x)\\ \varphi_2(x)\end{pmatrix}(\mathbb{y}) = \begin{pmatrix}\gamma_1(x)\\ \gamma_2(x)\end{pmatrix}, \tag{3}$$

где $\varphi_1(x)$ и $\gamma_1(x)$ (соответственно, $\varphi_2(x)$ и $\gamma_2(x)$) получены переносом (2а) (соответственно, (2б)) в точку x.

[1)]Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 99-01-00331).

Поскольку уравнения прогонки решаются численно, матрицы $\varphi_i(x)$ и векторы $\gamma_i(x)$, $i = 1, 2$, определяются с некоторой погрешностью (величина которой зависит от свойств жесткости системы (1)). В работах [5] (для случая одного уравнения 2-го порядка), [6] (для варианта прогонки [2]) и [7] показано следующее: если вариант метода прогонки таков, что матрицы $\varphi_i(x)$ нигде не являются плохо обусловленными (т.е. нужно следить за тем, чтобы базис в переносимом подпространстве не сильно отличался от ортонормированного), то указанная погрешность может быть реализована точным переносом краевых условий для системы (1), в которой матричной функции $A(x)$ и вектор-функции $\mathbf{f}(x)$ придано небольшое (в равномерной норме) возмущение. Отсюда обычно делается качественный вывод: если исходная задача была устойчивой относительно малых возмущений величин, определяющих эту задачу, то устойчив и рассматриваемый метод ее решения. Однако такой вывод не вполне обоснован по следующей причине. Система (3) также решается лишь приближенно, при этом (см., например, [8, теорема 21.41]) в решение $\mathbf{y}(x)$ вносится погрешность, величина которой на несколько порядков больше величины $\|\Phi\|\|\Phi^{-1}\|p^{-t}$, где

$$\Phi = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \end{pmatrix},$$

p – основание машинной арифметики, t – число разрядов мантиссы. Предположим, что прогонка выполняется корректно, так что строки матриц $\varphi_1(x)$, $\varphi_2(x)$ образуют приблизительно ортонормированные базисы в соответствующих подпространствах. Тогда $\|\Phi\|$ – величина порядка единицы, а

$$\|\Phi^{-1}\| = [\operatorname{gap}(\operatorname{span}(\varphi_1(x)), \operatorname{span}(\varphi_2(x)))]^{-1}, \text{ где } \operatorname{gap}(L, M) = \min_{\mathbf{u} \in L, \mathbf{v} \in M} \sin(\widehat{\mathbf{u}, \mathbf{v}})$$

есть так называемый зазор между подпространствами L и M. Таким образом, мы вынужденно приходим к следующей постановке вопроса: пусть исходная задача (1), (2) "устойчива"[2]; верно ли, что тогда зазор между подпространствами, полученными в результате переноса краевых условий в произвольную промежуточную точку отрезка $[a, b]$, не может быть очень маленьким? Заметим, что в приведенной формулировке не участвуют никакие вычислительные процедуры и суть вопроса об устойчивости методов дифференциальной прогонки не в том, что нужно устойчиво осуществлять прогонку (это делать нужно, и известно, как это делается), а в том, правильна ли сама идея явного переноса подпространств, соответствующих краевым условиям.

Прежде чем перейти к изложению полученных результатов, отметим, что эквивалентный вопрос исследуется в работе [6] применительно к прогонке [2] (технически он ставится иначе, поскольку в [6] используется обратная прогонка) и на него дан положительный ответ в предположении, что sup-нормы некоторых матричных функций (типа функции Грина) не очень велики. Аналогично, в работах [9], [10], где исследуются свойства дихотомии решений задач (1), (2), получена оценка снизу на зазор между подпространствами через так называемую константу устойчивости задачи (1), (2) по $\mathbf{f}$. Эта константа определяется как максимум функции Грина на квадрате, что соответствует операторной норме, когда для $\mathbf{f}$ берется L_1-норма. Но здесь сути задачи соответствует не L_1-, а sup-норма для A и $\mathbf{f}$, так как малые погрешности вносятся на всем промежутке интегрирования. Так как sup-норма на отрезке не оценивается через L_1-норму, то из устойчивости задачи (1), (2) не следует, что константа устойчивости не является неприемлемо большой (см. ниже Пример 2). Таким образом, результаты [6], [9], [10] оставляют поставленный здесь вопрос открытым.

2. Желаемый результат должен формулироваться в следующем виде: если задача (1), (2) устойчива относительно малых изменений величин, ее определяющих, то в любой точке отрезка интегрирования зазор между перенесенными подпространствами оценивается снизу величиной существенно большей, чем относительная точность машинных вычислений (и тогда метод дифференциальной прогонки для этой задачи устойчив). Данными задачи (1), (2) являются: матричная функция $A(x)$, вектор-функция $\mathbf{f}(x)$, матрицы φ_a, φ_b и векторы γ_a, γ_b. Чем меньше их будет участвовать в формулировке, тем сильнее получится результат; кроме того, предстоит уточнить, понимается ли малость их изменений в относительном или абсолютном смысле. Покажем сначала, что из устойчивости только по краевым условиям не следует, что зазор не может быть сколь угодно мал.

[2] Или "хорошо обусловлена" – о точном смысле этого предположения см. далее; здесь важно отметить, что формально из него не следует никаких фактов, касающихся взаимного расположения подпространств.

Пример 1: $\mathbf{y}(x) = (y_1(x), y_2(x))^{\mathrm{T}}$, $A(x) = \mathrm{diag}(0, 2x/(x^2 + \delta))$, $f(x) \equiv 0$, краевые условия $y_2(-1) = 0$, $y_1(1) - y_2(1) - 1 = 0$. За счет выбора малого δ угол между перенесенными подпространствами в точке $x = 0$ можно сделать сколь угодно малым, при этом задача остается хорошо обусловленной по φ_i, γ_i.

Следующий пример показывает, что зазор между подпространствами может быть малым и у краевой задачи (1), (2), устойчивой относительно возмущений $\mathbf{f}$ и A, если они малы в абсолютном смысле.

Пример 2: $\mathbf{y}(x) = (y_1(x), y_2(x))^{\mathrm{T}}$; $A(x) = C\mathrm{diag}(-\mu, \mu)C^{-1}$, $C = \begin{pmatrix} 1 & 1 \\ 0 & \delta \end{pmatrix}$, $-\delta y_1(-1) + y_2(-1) = 0$, $y_2(1) = 0$, μ – большое, δ – малое положительные числа (однородная задача рассматривается как неоднородная). Зазор между подпространствами постоянен по x и равен δ. Покажем сначала, что имеет место устойчивость по $\mathbf{f}$. Функция Грина $G(x, s)$ имеет вид

$$C^{-1}G(x,s)C = \begin{cases} \mathrm{diag}(e^{\mu(s-x)}, 0), & x \ge s, \\ \mathrm{diag}(0, -e^{\mu(x-s)}), & x < s. \end{cases}$$

При сколь угодно малом δ за счет выбора достаточно большого μ можно добиться, чтобы норма соответствующего интегрального оператора была сколь угодно мала, что будет означать устойчивость по $\mathbf{f}$. В связи с упомянутыми выше результатами [9], [10] заметим, что при этом “константа устойчивости”, которая равна норме следа функции Грина, будет величиной порядка δ^{-1}, т.е. очень большой.

Покажем теперь устойчивость задачи при малых (в абсолютном смысле, в равномерной норме) возмущениях матричной функции $A(x)$. Пусть

$$\tilde{A}(x) = A(x) + W(x), \quad \sup_{x \in [a,b]} \|W(x)\| = \varepsilon.$$

Ситуацию удобнее проследить в базисе из собственных векторов матрицы C. Пусть $\hat{y}_1$, $\hat{y}_2$ – новые координаты. Сделав “замену Прюфера” $\varphi = \arctan(\hat{y}_1/\hat{y}_2)$, получим для φ уравнение прогонки

$$\varphi'(x) = -2\mu\cos\varphi(x)\sin\varphi(x) + \hat{u}_{12}(x)\cos^2\varphi(x) + [\hat{u}_{11}(x) - \hat{u}_{22}(x)]\cos\varphi(x)\sin\varphi(x) - \hat{u}_{21}(x)\sin^2\varphi(x),$$

где $\hat{u}_{ij}$ – компоненты матрицы U в новом базисе, причем их абсолютные величины оцениваются сверху величиной порядка $\varepsilon\delta^{-1}$. Теперь очевидно, что при достаточно большом μ траектории прогонки левого краевого условия будут близки к невозмущенной траектории $\varphi \equiv 0$. Для правого краевого условия рассуждение аналогично.

Комбинируя оба примера, можно получить краевую задачу с малым зазором между подпространствами, соответствующими перенесенным краевым условиям, но устойчивую по φ_i, γ_i и (в смысле абсолютно малых возмущений) по $\mathbf{f}$ и A.

Положительный результат получается только в том случае, если допускаются возмущения A и $\mathbf{f}$, малые по отношению к их нормам. Такая постановка вопроса совершенно естественна с вычислительной точки зрения, поскольку по ходу вычислений, как правило, возникают погрешности, малые именно в относительном, а не абсолютном смысле.

Предложение. *Пусть решение задачи* (1), (2) *устойчиво относительно малых возмущений всех входящих в нее данных, причем малость возмущений* $A(x)$ *и* $\mathbf{f}(x)$ *понимается в следующем смысле:*

$$\tilde{A}(x) = A(x) + W(x), \quad \varepsilon = \sup_{x \in [a,b]} \|W(x)\| / \sup_{x \in [a,b]} \|A(x)\|$$

мало́ (соответственно – для $\mathbf{f}$*). Тогда существует равномерная по* x *оценка снизу для зазора между перенесенными подпространствами краевых условий (и, следовательно, верхняя оценка для* $\|\Phi^{-1}\|$*).*

Доказательство (от противного). Пусть при малом $\delta > 0$ в некоторой точке $x_0 \in [a, b]$ в подпространстве, полученном переносом (однородного) левого краевого условия в точку x_0, найдется вектор $\mathbf{u}$, а в перенесенном справа – вектор $\mathbf{v}$ такие, что $\sin(\widehat{\mathbf{u}, \mathbf{v}}) < \delta$. Считаем, что $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$.

Существует ортогональное преобразование Q пространства $\mathbb{R}^n$, переводящее $\mathbf{u}$ в $\mathbf{v}$ и такое, что $\|I - Q\| < \delta$ (I – тождественное преобразование). Рассмотрим возмущенную краевую задачу

$$\mathbf{y}'(x) = \tilde{A}(x)\mathbf{y}(x) + \tilde{\mathbf{f}}(x), \tag{1'}$$

$$\tilde{\varphi}_a\mathbf{y}(a) = \tilde{\gamma}_a, \quad \varphi_b\mathbf{y}(b) = \gamma_b, \tag{2'}$$

где

$$\tilde{A}(x) = \begin{cases} QA(x)Q^{-1}, & x \le x_0, \\ A(x), & x > x_0, \end{cases} \quad \tilde{\mathbf{f}}(x) = \begin{cases} Q\mathbf{f}(x), & x \le x_0, \\ \mathbf{f}(x), & x > x_0, \end{cases} \quad \tilde{\varphi}_a = \varphi_a Q^{-1}.$$

Здесь (при малых δ)

$$\sup_{x \in [a,b]} \|\tilde{A}(x) - A(x)\| / \sup_{x \in [a,b]} \|A(x)\| \le 3\delta,$$

$$\sup_{x \in [a,b]} \|\tilde{f}(x) - f(x)\| / \sup_{x \in [a,b]} \|f(x)\| \le \delta,$$

$\|\tilde{\varphi}_a - \varphi_a\| \le \delta$ (строки φ_a ортонормированы). Таким образом, задача (1)', (2)' получена малым возмущением задачи (1), (2). В то же время (однородные) краевые условия, будучи перенесены в точку x_0, пересекаются (по вектору $\mathbf{v}$), т.е. задача (1)', (2)' "сидит на спектре". Следовательно, задача (1), (2) не была устойчивой, что и требовалось доказать.

Окончательный вывод можно сформулировать так: если решение краевой задачи (1), (2) устойчиво к малым изменениям краевых условий и (относительно) малым изменениям коэффициентов системы (1), то эта задача может быть решена с помощью дифференциальной прогонки. В более слабых предположениях гарантировать это не представляется возможным. Далее, если выбран такой вариант метода прогонки, в котором векторы, задающие положение подпространства, при интегрировании прогоночных уравнений сохраняют норму порядка единицы и не "слипаются" (например, ортогональная прогонка), то полученный численный метод так же хорошо устойчив, как и исходная задача.

Автор благодарит рецензента статьи за ценные замечания и уточнения.

СПИСОК ЛИТЕРАТУРЫ

1. *Абрамов А.А.* О переносе граничных условий для систем линейных обыкновенных дифференциальных уравнений (вариант метода прогонки) // Ж. вычисл. матем. и матем. физ. 1961. Т. 1. № 3. С. 542–545.
2. *Годунов С.К.* О численном решении краевых задач для систем линейных обыкновенных дифференциальных уравнений // Успехи матем. наук. 1961. Т. 16. № 3. С. 171–175.
3. *Бахвалов Н.С.* Численные методы. М.: Наука, 1975.
4. *Бабенко К.И.* Основы численного анализа. М.: Наука, 1986.
5. *Абрамов А.А.* Вариант метода прогонки // Ж. вычисл. матем. и матем. физ. 1961. Т. 1. № 2. С. 349–351.
6. *Кузнецов С.В.* Решение краевых задач для обыкновенных дифференциальных уравнений // Тр. Ин-та матем. СО АН СССР. Вычисл. методы линейной алгебры. Новосибирск: Наука, 1985. Т. 6. С. 85–110.
7. *Babuska I., Majer V.* The factorization method for the numerical solution of two point boundary value problems for linear ODEs // SIAM J. Numer. Analys. 1987. V. 24. № 6. P. 1301–1334.
8. *Форсайт Дж., Моулер К.* Численное решение систем линейных алгебраических уравнений. М.: Мир, 1969.
9. *van Loon P.M.* Continuous decoupling transformations for linear boundary value problems // CWI Tracts, 52. Amsterdam: Centrum voor Wiskunde et Informatica, 1988.
10. *de Hoog F., Mattheij R.M.M.* On dichotomy and well conditioning in boundary value problems // SIAM J. Numer. Analys. 1987. V. 24. № 1. P. 89–105.