ДОКЛАДЫ АКАДЕМИИ НАУК, 2017, том 473, № 2, с. 138–141

МАТЕМАТИКА

УДК 517.9, 518.12

О СЕМЕЙСТВАХ ВЫСОКОТОЧНЫХ МУЛЬТИОПЕРАТОРНЫХ АППРОКСИМАЦИЙ ПРОИЗВОДНЫХ, ИСПОЛЬЗУЮЩИХ ДВУХТОЧЕЧНЫЕ ОПЕРАТОРЫ

 А. И. Толстых

Представлено академиком РАН Ю.Г. Евтушенко 14.05.2016 г.
Поступило 27.06.2016 г.

Приводится новое семейство мультиоператорных аппроксимаций производных четных и нечетных порядков с обращением двухточечных операторов. Формулируются теоремы существования и единственности мультиоператоров формально произвольных порядков, исследуются их спектральные свойства. В качестве примера рассматривается схема для модельного гиперболического уравнения с мультиоператорной аппроксимацией 36-го порядка. Представлены оценки точности и сходимости численных решений тестовой задачи.

DOI: 10.7868/S0869565217080059

1. Ниже рассматриваются новые варианты мультиоператорных аппроксимаций первых и высших производных, основанные на суперпозициях операторов двухточечных разностей и обращаемых двухточечных операторов. Такие аппроксимации обладают большой простотой и экономичностью реализации, а также возможностью получения очень высоких порядков. В качестве примеров их применения приводятся схемы для уравнений переноса, в которых порядки аппроксимации пространственных производных достигают 36.

Метод увеличения порядков аппроксимаций сеточных операторов, описанный в [1], основан на увеличении числа базисных операторов одной и той же структуры в их линейных комбинациях, названных мультиоператорами. В общем виде эта идея выглядит следующим образом. Пусть имеется однопараметрическое семейство операторов $L_h(s)$ компактных аппроксимаций некоторого линейного оператора L на сетке

$$\omega_h: \quad x_j = jh, \quad j = 0, \pm 1, \pm 2, \ldots, h = \text{const},$$

в которых, по крайней мере, обращаемые операторы зависят от параметра c. Зафиксируем M значений параметра c, $c = c_i$, $i = 1, 2, \ldots, M$, и построим линейную комбинацию базисных операторов $L_h(c_i)$ вида

$$L_M(c_1, c_2, \ldots, c_M) = \sum_{i=1}^{M} \gamma_i L_h(c_i), \quad \sum_{i=1}^{M} \gamma_i = 1, \tag{1}$$

в которой коэффициенты γ_i являются неизвестными линейной системы, полученной из условий обращения в ноль $M - 1$ членов низших порядков в погрешности аппроксимации мультиоператора L_M. В случае разрешимости этой системы при любых M различных значениях параметра имеет место существование и единственность мультиоператора любого порядка.

К настоящему времени имеются различные варианты семейств компактных аппроксимаций $L_h(c)$, генерирующие базисные операторы для различных формул численного анализа (например, формул численного дифференцирования, интерполяции и экстраполяции, квадратурных формул и т.д.) [2].

2. Будем рассматривать пространство U_h сеточных функций $u_j, j = 0, \pm 1, \pm 2, \ldots$, определенных на сетке ω_h. Следуя [3], определим на этой сетке левые и правые операторы

$$R_l(c) = I + c\Delta_-, \quad R_r(c) = I - c\Delta_+, \tag{2}$$

где I – единичный оператор, а $\Delta_- = I - T_{-1}$, $\Delta_+ = T_1 - I$, $T_k u_j = u_{j+k}$ суть левые и правые двухточечные разности.

Т е о р е м а 1. *Пусть* $c > -\frac{1}{2} + \delta$, $\delta > 0$. *Тогда обратные операторы* $R_l(c)^{-1}$ *и* $R_r(c)^{-1}$ *существуют.*

Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра “Информатика и управление” Российской Академии наук, Москва
Московский физико-технический институт (государственный университет), Долгопрудный Московской обл.
E-mail: tol@ccas.ru

Выполнение условия для параметра c будет предполагаться в дальнейшем. Используя операторы (2), построим однопараметрические семейства левых и правых компактных аппроксимаций вида

$$L_l(c) = \frac{1}{h} R_l(c)^{-1} \Delta_-, \quad L_r(c) = \frac{1}{h} R_r(c)^{-1} \Delta_+. \tag{3}$$

Пусть U_h есть гильбертово пространство сеточных функций с суммируемыми квадратами и скалярным произведением $u_h = (u_j, j = 0, \pm 1, \pm 2, ...)$ и $v_h = (v_j, j = 0, \pm 1, \pm 2, ...)$, определенным как $(u_h, v_h) = h \sum_{j=-\infty}^{\infty} u_j v_j$. Тогда имеют место операторные неравенства $L_l(c) > 0$, $L_r(c) < 0$, $L_l(c)^* = -L_r(c)$.

Пусть U есть пространство функций $u(x)$, обладающих производными до того порядка, который необходим для справедливости разложений в ряды Тейлора, используемых в дальнейшем. Рассмотрим разложения в окрестности $x = x_j$ действий операторов $L_l(c)[u]_j$, $L_r(c)[u]_j$, $u \in U$, где $[u]_j = u(x_j)$. Используя процедуру деления рядов, можно установить, что они с точностью до остаточного члена имеют вид

$$\begin{gathered} L_l(c)[u]_j = [u']_j + \sum_{k=1} p_k(c) h^k [u^{(k+1)}]_j, \\ L_r(c)[u]_j = [u']_j + \sum_{k=1} (-1)^k p_k(c) h^k [u^{(k+1)}]_j, \end{gathered} \tag{4}$$

где $p_k(c)$ суть полиномы степени k, а количество слагаемых в суммах зависят от запаса гладкости функции $u(x)$.

Общая процедура вычисления производных методом бегущего счета и возможные граничные условия приведены в [3]. В данном случае ее реализация для каждого оператора может быть осуществлена с использованием двух операций умножения и одной операции сложения, приходящихся на узел сетки.

3. Выберем M различных значений $c_1, c_2, ..., c_M$ параметра c, удовлетворяющих условиям $c_i > -\frac{1}{2} + \delta$, $\delta > 0$. Используя в качестве базисных операторов $L_l(c_i)$ и $L_r(c_i)$, можно построить левые и правые мультиоператоры вида (1), имеющие погрешность аппроксимации $O(h^M)$:

$$\begin{gathered} L_{M,l}(c_1, c_2, \ldots, c_M) = \sum_{i=1}^{M} \gamma_i L_l(c_i), \\ L_{M,r}(c_1, c_2, \ldots, c_M) = \sum_{i=1}^{M} \gamma_i L_r(c_i) \end{gathered}$$

при условии, что коэффициенты γ_i определяются из системы

$$\sum_{i=1}^{M} \gamma_i = 1, \quad \sum_{i=1}^{M} p_k(c_i) \gamma_i = 0, \quad k = 1, 2, \ldots, M-1. \tag{5}$$

Система (5) сводится к системе с матрицей Вандермонда, элементами которой являются степени c_i. Это доказывает существование и единственность левых и правых мультиоператоров. Более того, соответствующую обратную матрицу, которую можно использовать для получения коэффициентов γ_i, можно выписать в явном виде.

Отмечая верхними индексами (0) и (1) соответственно самосопряженные и кососимметричные составляющие операторов, легко усмотреть, что $L_{M,l}^{(1)} = L_{M,r}^{(1)}$, $L_{M,l}^{(0)} = -L_{M,r}^{(0)}$, однако $L_{M,l}^{(0)}$ (или $L_{M,r}^{(0)}$) необязательно является положительным оператором. Его положительность, необходимая при построении устойчивых схем для уравнений переноса, может быть достигнута путем выбора значений c_i. Для упрощения исследований мультиоператоров и, в частности, выбора значений параметра, как и в случае всех построенных ранее мультиоператоров, предполагается, что c_i распределены по заданному закону между их минимальными и максимальными значениями $c_{\min}$ и $c_{\max}$. В этом случае положительность $L_{M,l}^{(0)}$ определяется положительностью действительной части его фурье-образа $F_{M,l}^{(0)}(\alpha; c_{\min}, c_{\max})$, $\alpha = kh$, где k – переменная фурье-преобразования. При предположении о том, что $c_{\min}$, $c_{\max}$ обеспечивают положительность левого или правого мультиоператора, мультиоператоры образуют пары операторов, которые могут быть использованы в “противопотоковых” схемах для уравнений переноса.

4. Левые и правые мультиоператоры $L_{M,l}$ и $L_{M,r}$ в некоторых случаях могут оказаться полезными как “односторонние” аппроксимации производных, использующие значения функций в узлах, расположенных либо слева, либо справа от рассматриваемого узла. Однако построение таких мультиоператоров с очень высокими порядками аппроксимации наталкивается на неудобство в виде ухудшения обусловленности матриц Вандермонда с увеличением числа M значений параметра.

Аппроксимации производных с очень высокими порядками более выгодно получать, используя базисные операторы, генерируемые кососимметричной частью $L_l(c)$ и $L_r(c)$. В этом случае разложения в ряды Тейлора содержат только четные степени h. Идя таким путем, воспользуемся однопараметрическим семейством $L_0(c) = \frac{L_l + L_r}{2}$ и построим кососимметричный мультиоператор

$$L_M(c_{\min}, c_{\max}) = \sum_{i=1}^{M} \gamma_i L_0(c_i), \quad c_{\min} > -\frac{1}{2} + \delta, \quad \delta > 0, \tag{6}$$

где $\sum_{i=1}^{M} \gamma_i = 1$, $\sum_{i=1}^{M} p_{2k}(c_i)\gamma_i = 0$, $k = 1, 2, \ldots, M - 1$. Разрешимость системы для γ_i следует из возможности ее сведения к матрице Вандермонда, содержащей степени $y_i = c_i(1 + c_i)$. Таким образом, имеет место

Теорема 2. *Мультиоператор* (6), *имеющий порядок аппроксимации* $2M$ *при любых заданных значениях* M *и* $c_{\min}$, $c_{\max}$, *существует и единствен*.

5. Мультиоператор $L_M(c_{\min}, c_{\max})$ не содержит диссипативного механизма, обеспечивающего устойчивость и демпфирование паразитических осцилляций численных решений при его использовании, например, в уравнениях механики жидкости и газа. В этом случае в аппроксимацию первых производных без существенного понижения порядка аппроксимации можно ввести самосопряженный мультиоператор, основанный на однопараметрическом семействе $L_1(b) = L_l(b) - L_r(b)$. Выберем некоторый набор значений b_i, $i = 1, 2, \ldots N$, и зададимся целью построить самосопряженный мультиоператор D_N вида

$$D_N(b_1, b_2, \ldots, b_N) = \sum_{i=1}^{N} \beta_i L_1(b_i), \quad \sum_{i=1}^{N} \beta_i = 1, \tag{7}$$

обращающий в ноль коэффициенты при h, h^3, h^5, …, h^{2N-3} в разложении в ряд для $D_N[u]_j$, так что $D_N[u]_j = O(h^{2N-1})$.

Теорема 3. *Самосопряженный мультиоператор* (7), *действие которого на достаточно гладкую функцию оценивается как* $O(h^{2N-1})$ *при любых заданных значениях* N *и* b_i, $i = 1, 2, \ldots N$, $b_i \neq b_j$, $b_i > -\frac{1}{2} + \delta$, $\delta > 0$, *существует и единствен*.

6. Однопараметрические семейства компактных аппроксимаций $L_0(c)$ и $L_1(b)$ можно использовать для построений мультиоператорных аппроксимаций заданного порядка высших производных. Рассмотрим сначала аппроксимации производных четных порядков. Для этого образуем системы вида

$$\sum_{i=1}^{N} \beta_i = 1, \quad \sum_{i=1}^{N} p_{2k-1}(b_i)\beta_i = \delta_{kl}, \quad k = 1, 2, \ldots, N - 1, \quad l = 1, 2, \ldots, N - 2, \tag{8}$$

где δ_{kl} есть символ Кронекера. Построив мультиоператор

$$D_N(b_1, b_2, \ldots, b_N) = \sum_{i=1}^{N} \beta_i L_1(b_i), \quad \sum_{i=1}^{N} \beta_i = 1, \tag{9}$$

где β_i суть решения l-й системы из (8), получим аппроксимацию производной четного порядка

$$[u^{(2l)}]_j = h^{-(2l-1)} D_{N,l}[u]_j + O(h^{2(N-l)}), \quad l = 1, 2, \ldots, N - 2.$$

В частности, для второй производной ($l = 1$) имеем $u'' = h^{-1} D_{N,1} + O(h^{2N-2})$.

При построении мультиоператорных аппроксимаций производных нечетных порядков непосредственное обобщение системы в определении мультиоператора (6) невозможно, поскольку генерирующее семейство $L_0(c)$ заведомо определяет аппроксимацию первой производной. Введем новое однопараметрическое семейство вида $L_2(c) = L_0(c) + c(\Delta_+ + \Delta_-)/2h$. Легко усмотреть, что разложение в точке x_j для модифицированного таким образом оператора можно записать в виде $L_2(c) = \sum_{k=1} q_{2k-1}(c) h^{2(k-1)} [u^{(2k)}]_j$, где $q_k(c)$ суть полиномы степени k. Задавшись значениями параметра, образуем систему

$$\sum_{i=1}^{M} \gamma_i = 1, \quad \sum_{i=1}^{M} q_{2k-1}(c_i)\gamma_i = \delta_{km}, \quad k = 1, 2, \ldots, M - 1, \quad m = 1, 2, \ldots, M - 2. \tag{10}$$

Определив мультиоператор $L_{M,m}(c_1, c_2, \ldots, c_M) = \sum_{i=1}^{M} \gamma_i L_m(c_i)$, где γ_i суть решения m-й системы из (10), получим аппроксимацию $(2m - 1)$-й производной

$$[u^{(2m-1)}]_j = h^{-2(m-1)} L_{M,m}[u]_j + O(h^{2(M-m)}), \quad m = 1, 2, \ldots, M - 2.$$

В случае $m = 1$ имеем аппроксимацию первой производной с погрешностью $O(h^{2M-2})$, что дает худшую оценку по сравнению с оценкой $O(h^{2M})$ в случае мультиоператора L_M.

7. Мультиоператоры (6) и (7) могут быть использованы при построении устойчивых схем для уравнений и систем гиперболического типа, а также для уравнений механики жидкости и газа. Например, полудискретизованная схема для модельного уравнения $u_t + f(u)_x = 0$ в безындексной форме может быть записана в виде

$$u_t + L_M f(u) + C D_N u = 0, \quad C \geq 0, \quad u \in U_h. \tag{11}$$

Для устойчивости схемы (11) в приближении “замороженных коэффициентов” ($f(u) = au$, $a = \text{const}$) достаточно потребовать положительности $D_N(b_{\min}, b_{\max})$. Будем рассматривать значения $b_{\min}$,

Таблица 1. Точность численных решений

N	16	32	64	128
L_{36}	$1.3 \cdot 10^{-3}$	$1.8 \cdot 10^{-6}$	$1.1 \cdot 10^{-9}$	$2.5 \cdot 10^{-17}$
k_c		9.5	10.7	24.4
WENO-5	$1.25 \cdot 10^{-2}$	$1.20 \cdot 10^{-3}$	$9.5 \cdot 10^{-5}$	$3.31 \cdot 10^{-6}$
k_c		3.4	3.6	4.9

$b_{\max}$ в схеме (11) как допустимые, если условие положительности выполняется.

Наличие оператора D_N не является необходимым для устойчивости схемы (11) и в некоторых случаях можно положить $C = 0$. Однако его присутствие может способствовать подавлению паразитических осцилляции решений, возникающих вследствие отсутствия монотонности схем (в смысле С.К. Годунова) при полной дискретизации схемы (11).

Для увеличения точности описания мелких деталей решений (например, в задачах вычислительной аэроакустики) путем выбора $c_{\min}$, $c_{\max}$ можно оптимизировать аппроксимации производных таким образом, чтобы фурье-образ сеточного оператора был близок к фурье-символу оператора дифференцирования в как можно более широком диапазоне безразмерных волновых чисел kh. Возможная процедура оптимизации описана в [4].

Используя широко распространенный тест, рассмотрим периодическую задачу Коши для упомянутого выше модельного уравнения с начальными данными $u(0, x) = 1.5 + 0.5\sin\pi x$, $x \in [-1, 1]$, положив в схеме (11) $f(u) = \dfrac{u^2}{2}$, $C = 0$. Интегрирование по времени с оператором L_M 36-го порядка ($M = 18$) осуществлялось методом Рунге–Кутты четвертого порядка с шагом по времени, вносящим меньшие погрешности, чем аппроксимация пространственной производной.

В табл. 1 приведены погрешности численных решений при $t = 0.3$ в сеточной норме C при различном числе N узлов сетки и локальные порядки сходимости k для схемы (11), обозначенной как L_{36}. Точное решение вычислялось с необходимой точностью в результате решения нелинейного алгебраического уравнения. Для сравнения представлены также погрешности в случае популярной схемы 5-го порядка WENO-5 (детали и ссылки см. в [4]).

Как следует из табл. 1, хотя локальный порядок для рассмотренного диапазона числа узлов сетки оказывается меньшим формального, схема L_{36} при $N = 128$ оказывается на 11 порядков более точной, чем схема WENO-5.

Оптимизированные схемы 36-го порядка позволяют правильно воспроизводить коротковолновые гармоники на больших интервалах времени при решении уравнений переноса. В [5] была предложена тестовая задача Коши для рассмотренного выше модельного уравнения при $f(u) = u$ с начальными данными в виде волнового пакета с коротковолновым наполнением $u(0, x) = [2 + \cos(\alpha x)]\left[\exp\left(-\ln 2\left(\dfrac{x}{10}\right)^2\right)\right]$. В [5] предлагалось оценить близость численных решений к точному при $t = 800$, $h = 1$, $\alpha = kh = 1.7$. Оказалось, что схема (11) позволяет сохранять эту близость для $\alpha = 2.2$ до $t = 15000$.

Работа выполнена в ВЦ ФИЦ “ИУ” РАН при финансовой поддержке Российского научного фонда, проект 14–11–00775.

СПИСОК ЛИТЕРАТУРЫ

1. *Толстых А.И.* // ДАН. 1999. Т. 366. № 3. С. 319–322.
2. *Толстых А.И.* Высокоточные компактные и мультиоператорные аппроксимации для уравнений в частных производных. М.: Наука, 2015. 320 с.
3. *Толстых А.И.* // ДАН. 2005. Т. 403. № 2. С. 172–177.
4. *Липавский М.В., Толстых А.И.* // ЖВМиМФ. 2013. Т. 53. № 4. С. 455–468.
5. *Tam C.K.W.* NASA/CPCP-2004-2159. 2004.