ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 140 за 2016 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Бахвалов П.А.

Метод нестационарного корректора для анализа точности линейных разностных схем для уравнения переноса

Рекомендуемая форма библиографической ссылки: Бахвалов П.А. Метод нестационарного корректора для анализа точности линейных разностных схем для уравнения переноса // Препринты ИПМ им. М.В.Келдыша. 2016. № 140. 32 с. doi:10.20948/prepr-2016-140
URL: http://library.keldysh.ru/preprint.asp?id=2016-140

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.КЕЛДЫША
Р о с с и й с к о й а к а д е м и и н а у к

П. А. Бахвалов

Метод нестационарного корректора для анализа точности линейных разностных схем для уравнения переноса

Москва — 2016

Бахвалов П. А.

Метод нестационарного корректора для анализа точности линейных разностных схем для уравнения переноса

В работе вводится понятие нестационарного корректора для широкого класса линейных разностных схем для однородного уравнения переноса с постоянной скоростью на неравномерных и неструктурированных сетках. Нестационарный корректор подчиняется тому же уравнению, что и численное решение, но с ненулевой правой частью. На примерах показывается, что его анализ позволяет устанавливать превышение порядка точности над порядком аппроксимации и специфику поведения ошибки решения при большом времени счёта.

Ключевые слова: геометрический корректор, аппроксимация и точность, неравномерная сетка, неструктурированная сетка

Pavel Alexeevisch Bakhvalov

Unsteady corrector method for accuracy analysis of linear numerical schemes for transport equation

We introduce the concept of unsteady corrector for a wide class of linear numerical schemes for the transport equation with constant velocity on non-uniform and unstructured meshes. The unsteady corrector is governed by the same equation as the numerical solution but with nonzero right-hand side. We show that its analysis allows to obtain supraconvergence and long-time simulation accuracy properties.

Key words: geometric corrector, consistency and accuracy, non-uniform mesh, unstructured mesh

Оглавление

Введение 3
Обозначения 6
Учёт многостадийности методов Рунге-Кутты 9
Нестационарный корректор 11
Уравнение нестационарного корректора 12
Нестационарный корректор и точность схемы 16
Периодические условия 20
Автомодельность главного корректора 21
Зависимость главного корректора от времени 23
Корректор, следующий за главным 25
Примеры 26
Заключение 30
Список литературы 31

Введение

Настоящая работа посвящена исследованию точности линейных разностных схем на неструктурированных сетках на примере однородного уравнения переноса с постоянной скоростью. Наиболее простым и понятным способом оценки точности разностных схем является анализ их аппроксимационной ошибки. Если аппроксимационная ошибка имеет порядок $O(h^k)$ и схема является устойчивой, то численное решение сходится к точному со скоростью, по крайней мере, $O(h^k)$. Однако такая оценка на неравномерных и неструктурированных сетках зачастую является грубой. Хорошо известно, что консервативные схемы, если они точно аппроксимируют производные от полиномов k-го порядка (k-exact schemes), показывают порядок точности в пределах от k до $k+1$ (при аппроксимации уравнений второго порядка – до $k+2$). Эффект превосходства порядка точности над порядком аппроксимации называют сверхсходимостью (supraconvergence).

Впервые эффект сверхсходимости был обнаружен А. Н. Тихоновым и А. А. Самарским в [1]. Для одномерного уравнения конвекции-диффузии они рассмотрели две схемы, одна из которых обладала первым порядком аппроксимации, а другая вообще не аппроксимировала уравнение, но обе они обладали вторым порядком точности. Для обеих схем было доказано, что их точность определяется не интегральной, а негативной нормой аппроксимационной ошибки. Последнее справедливо для многих консервативных схем для одномерного уравнения переноса $\partial u/\partial t + \partial u/\partial x = 0$. В частности, схема $(u_j^{n+1} - u_j^n)/\tau + 2(u_j^n - u_{j-1}^n)/(x_{j+1} - x_{j-1}) = 0$ не обладает свойством аппроксимации в максимальной и интегральной нормах, но обладает первым порядком аппроксимации в негативной норме и, благодаря этому, первым порядком точности в максимальной норме [2]. Этот пример также будет рассмотрен в настоящей работе. Схема на основе разделённых разностей [3] обладает первым порядком аппроксимации в максимальной и интегральной нормах, но вторым порядком аппроксимации в негативной норме и вторым порядком точности в максимальной норме (доказательство последнего факта было проведено при техническом ограничении на отношение соседних шагов). Метод Галёркина с разрывными базисными функциями k-го порядка обладает k-м порядком аппроксимации в интегральной норме, но $(k+1)$-м порядком аппроксимации в негативной норме и $(k+1)$-м порядком точности [4]. Таким образом, если негативная норма аппроксимационной ошибки имеет величину $O(h^{k+1})$, то можно ожидать, что порядок точности тоже будет равен $k+1$. В [5] упоминалась гипотеза, что консервативные схемы, обладающие k-м порядком аппроксимации на неравномерных сетках и вырождающиеся в схемы $(k+1)$-го порядка аппроксимации на равномерных, на любых сетках обладают $(k+1)$-м порядком точности. Эту гипотезу легко опровергнуть искусственным добавлением к схеме

вязкости с коэффициентом порядка $O(h^k)$, исчезающим на равномерной сетке. Однако для большинства используемых консервативных схем на неравномерных сетках в одномерном случае $(k+1)$-й порядок сходимости действительно наблюдается.

На неструктурированных сетках эффект сверхсходимости является более слабым, чем в одномерном случае. Он достаточно подробно исследован для конечно-элементых схем, в особенности для метода Галёркина с разрывными базисными функциями. Классическим результатом является сходимость этого метода в норме L_2 со скоростью $O(h^{k+1/2})$ [6] при ограниченности снизу минимального угла элемента. Для $k = 0$ и $k = 1$ в [7] показано, что эта оценка не улучшаема. Для схем повышенной точности, основанных на конечно-объёмном подходе, вопрос точности на неструктурированных сетках главным образом рассматривался в контексте монотонизации (см. [8] и библиографию к ней), а эффект сверхсходимости изучен слабо. Тем не менее, он не является тривиальным. Хотя в численных расчётах конечно-объёмные схемы, точные на полиномах k-го порядка, часто показывают порядок, близкий к $k+1$, порядком точности $k+1$ они не обладают. Это было продемонстрировано, в частности, в [9], где было проведено подробное численное исследование точности рёберно-ориентированных схем (обладающих аппроксимационной ошибкой $O(h)$) на сетках специального вида. Было продемонстрировано, что их наблюдаемый порядок точности при сохранении качества элементов лежит между 1 и 2 и зависит от стратегии измельчения сетки.

Теоретическое обоснование свойств сверхсходимости для конечно-элементных схем в большинстве работ проводилось на основе постпроцессинга решения. Этот метод, в частности, был использован в [6] для доказательства порядка точности $(k+1/2)$ разрывного метода Галёркина на неструктурированных сетках. В [4] разрывный метод Галёркина был рассмотрен на равномерных одномерных сетках. В частности, в случае кусочно-линейных базисных функций было показано, что численная ошибка решения складывается из двух слагаемых. Одно из них гладкое по пространству, имеет порядок $O(h^3)$ и растёт линейно со временем, второе – быстро осцилирующее, имеющее порядок $O(h^2)$ и ограниченное поведение во времени. Постпроцессингом последнее слагаемое могло быть исключено из решения. Анализируя представленные в [4] результаты, можно заметить, что огибающая осциллирующей части численного решения и гладкая часть расходятся друг относительно друга по фазе на четверть периода и, следовательно, связаны с разными производными от точного решения дифференциальной задачи. Метод постпроцессинга также был применён для исследования этой же задачи в [10] и [11].

С эффектом сверхсходимости тесно связан ещё один эффект, а именно, различие в точности схемы при малом и большом временах счёта (long-time

simulation accuracy). На равномерной сетке для конечно-разностных и конечно-объёмных схем ошибка решения, как правило, пропорциональна ошибке аппроксимации, накапливается от шага к шагу и, таким образом, линейно растёт со временем. Иное поведение наблюдается на неравномерных и неструктурированных сетках. Ошибка во времени может достаточно быстро достичь некоторого значения порядка h^k, после чего расти со скоростью, имеющей более высокий порядок малости по шагу сетки: $e(t,h) \sim h^k + th^{k+l}$, $l > 0$. Это, в частности, наблюдается для метода Галёркина с разрывными базисными функциями на одномерных равномерных сетках. В [10] для $k = 1$ была доказана оценка ошибки $||e(\cdot,t)||_{L^2} \leq C_1 th^{5/2} + C_2 h^2$, хотя численные эксперименты показывают более высокую точность: $||e(\cdot,t)||_{L^2} \leq C_1 th^3 + C_2 h^2$ [12].

С постпроцессингом решения тесно связан метод геометрического корректора, предложенный в [13] для исследования точности простейшей конечно-объёмной схемы на неструктурированных сетках. Геометрический корректор представляет собой поправку к оператору проектирования непрерывной функции на пространство сеточных функций, с учётом которой схема приобретает первый порядок аппроксимации. Величина этой поправки определяет точность решения по рассматриваемой схеме. Для коэффициентов геометрического корректора было выписано уравнение, которое для простейшей конечно-объёмной схемы решалось маршевым методом. С помощью этого метода было показано, что на асимптотически структурированных последовательностях сеток решение по простейшей конечно-объёмной схеме сходится к точному в максимальной норме с первым порядком, а не с половинным, как в общем случае.

В [14] на основе метода геометрического корректора было введено понятие нестационарного геометрического корректора, отличающееся от предложенного в [13] наличием зависимости коэффициентов корректора от времени. Если уравнение переноса рассматривается в бесконечной области, то нестационарный геометрический корректор обладает свойством автомодельности: норма корректора на более мелкой сетке совпадает с нормой корректора при увеличенном времени. Таким образом, анализируя зависимость коэффициентов корректора от времени, можно по одному расчёту оценивать порядок точности схемы при измельчении сетки. С помощью этого факта был доказан первый порядок точности простейшей конечно-объёмной схемы при блочном измельчении.

В настоящей работе нестационарный метод геометрического корректора обобщается на широкий класс схем повышенной точности, включающий конечно-объёмные и конечно-элементные схемы. Рассматривается приближение численного решения конечным рядом по степеням производных от точного решения. Для коэффициентов этого ряда выписывается разностное уравнение, оператор в котором совпадает с оператором рассматриваемой разностной

схемы. Это уравнение может решаться как теоретически (в простых случаях), так и численно. На периодических сетках последнее позволяет более экономично получать оценки точности численного метода, чем при расчёте на последовательностях сеток. Также метод нестационарного корректора позволяет устанавливать превышение порядка точности над порядком аппроксимации и повышенную точность схемы при большом времени счёта.

Обозначения

Пусть $\boldsymbol{a}$ – некоторый вектор, постоянный во времени и пространстве. Рассмотрим уравнение переноса

$$\frac{\partial u}{\partial t} + \boldsymbol{a} \cdot \nabla u = 0 \tag{1}$$

в некоторой области $G \in \mathbb{R}^d$ с начальными условиями $u(0, \boldsymbol{r}) = u_0(\boldsymbol{r})$, $\boldsymbol{r} \in G$, и граничными условиями на входной границе $u(t, \boldsymbol{r}) = u_b(t, \boldsymbol{r})$, $r \in (\partial G)^-$.

Рассмотрим некоторую линейную полудискретную аппроксимацию уравнения (1). Обозначим через M^0 множество её неизвестных и через M^b – число входящих в схему значений, определяемых граничными условиями. Пусть $M = M^0 \cup M^b$. Под численным решением на фиксированный момент времени будем понимать вектор в пространстве $\mathbb{R}^M$. Запишем эту полудискретную схему в виде

$$\frac{du_j}{dt} + \sum_{k \in M} L_{jk} u_k = 0, \quad j \in M^0. \tag{2}$$

Для решения системы уравнений (2) будем использовать метод Рунге-Кутты. Хотя рассуждения, приводимые в настоящей работе, применимы для произвольного метода, для простоты записи будем рассматривать только явные S-стадийные схемы следующего вида:

$$\frac{u_j^{n,s+1} - u_j^n}{\omega_{s+1,S} \Delta t} + \sum_{k \in M} L_{jk} u_k^{n,s} = 0, \quad j \in M^0, \quad s = 0, \ldots, S-1, \tag{3}$$

где $u^{n,0} = u^n$, $u^{n+1} = u^{n,S}$, а коэффициенты $\omega_{s,S}$ определяются формулой

$$\omega_{s,S} = \frac{1}{S - s + 1}, \quad s = 1, \ldots, S. \tag{4}$$

Такие схемы обладают порядком точности S для решения систем линейных ОДУ. Далее для общности записи положим $\omega_{0,S} = 0$ и будем использовать обозначения $t^n = n\Delta t$, $t^{n,s,S} = (n + \omega_{s,S})\Delta t$.

Для того, чтобы связать разностную задачу с дифференциальной и задать начальные и граничные условия, нужно определить оператор проектирования Π, сопоставляющий временному сечению функции f вектор $\Pi(f) \in \mathbb{R}^M$. Чтобы избежать лишних скобок в формулах, будем обозначать компоненты вектора $\Pi(f)$ через $\Pi_j(f)$, $j \in M$.

Определим оператор Π следующим равенством:

$$\Pi_j(f) = \int\limits_{\mathbb{R}^d} f(\boldsymbol{r}) d\mu_j, \tag{5}$$

где μ_j – некоторая неотрицательная мера, удовлетворяющая условию нормировки

$$\int\limits_{\mathbb{R}^d} d\mu_j = 1, \quad j \in M, \tag{6}$$

и носитель которой лежит в некотором шаре радиуса $R/2$:

$$\text{supp}\mu_j \subset B_{R/2}(\boldsymbol{c}_j), \quad j \in M. \tag{7}$$

Кроме того, значения в $j \in M^b$ определяются только граничными условиями:

$$\text{supp}\mu_j \subset (\partial G)^-, \quad j \in M^b. \tag{8}$$

Из определения (5)–(8) очевидны три следствия:

- $\Pi(f)$ является линейным по f;
- константа проецируется точно: $\Pi_j(1) = 1$;
- $\text{supp}\mu_j \subset B_R\left(\Pi_j(\boldsymbol{r})\right)$.

Схема (2) дополняется начальными условиями $u_j(0) = \Pi_j(u(0, \cdot\,))$, $j \in M$, и граничными условиями $u_j(t) = \Pi_j(u(t, \cdot\,))$, $j \in M^b$. При использовании явных методов Рунге-Кутты (3) эти условия записываются в виде

$$u_j^0 = \Pi_j(u(0, \cdot\,)), \quad j \in M. \tag{9}$$

$$u_j^{n,s} = \Pi_j(u(t^{n,s,S}, \cdot\,)), \quad j \in M^b. \tag{10}$$

Условиям (5)–(8) удовлетворяют, в частности, следующие операторы:

- оператор взятия точечного значения от функции в узле расчётной сетки;
- оператор взятия точечного значения от функции в центре масс сеточного элемента;
- оператор интегрального среднего от функции по сеточному элементу;

- оператор взятия точечного значения в некоторой точке от ортогональной L_2-проекции функции на пространство полиномов некоторой размерности в рамках сеточного элемента.

Использование первого оператора характерно для рёберно-ориентированных схем, второй иногда используется для конечно-объёмных схем низкого порядка, третий применяется для конечно-объёмных схем с полиномиальной реконструкцией, четвёртый – для метода Галёркина с разрывными базисными функциями.

В [13] для простейшей конечно-объёмной схемы было введено понятие геометрического корректора $\mathbf{\Gamma}$ как набора векторов $\mathbf{\Gamma}_j$, $j \in M^0$, при котором эта схема обладает первым порядком аппроксимации в смысле оператора $\tilde{\Pi}$, определённого равенством $(\tilde{\Pi} f)_j^n = f(t^n, \boldsymbol{r}_j + \mathbf{\Gamma}_j)$, где $\boldsymbol{r}_j$ – центр масс j-го элемента расчётной сетки. При этом в [13] предполагалось, что $\mathbf{\Gamma}$ не зависит от времени; в [14] это предположение было снято, и геометрический корректор был назван нестационарным. Смысл слова «геометрический» объясняется тем, что применение оператора $\tilde{\Pi}$ вместо оператора взятия точечного значения в центре масс равносильно смещению положения точек, в которых определены значения искомой функции, при сохранении коэффициентов схемы. Поскольку вводимые ниже операторы проектирования в общем случае будут лишены такой наглядной геометрической интерпретации, мы не будем употреблять слово «геометрический».

В настоящей работе будем рассматривать операторы проектирования $\tilde{\Pi}$ общего вида, сопоставляющие гладкой функции f вектор с компонентами $(\tilde{\Pi} f)_j^{n,s}$, $j \in M$, $n \in \mathbb{N} \cup \{0\}$, $s = 0, \ldots, S$. При этом будем предполагать, что $(\tilde{\Pi} f)_j^{n,S} = (\tilde{\Pi} f)_j^{n+1,0}$. Конкретное условие гладкости будет уточняться для каждого оператора.

Определение 1. *Пусть u – точное решение дифференциальной задачи (1) при некоторых начальных и граничных условиях. Ошибкой аппроксимации схемы (3) на временном подслое n, s на функции u в смысле оператора $\tilde{\Pi}$ будем называть вектор $\epsilon^{n,s}$ с компонентами*

$$\epsilon_j^{n,s} = \frac{(\tilde{\Pi} u)_j^{n,s+1} - (\tilde{\Pi} u)_j^n}{\omega_{s+1,S} \Delta t} + \sum_{k \in M} L_{jk} (\tilde{\Pi} u)_k^{n,s}. \tag{11}$$

Будем называть схему точной на полиноме порядка q (например, константе, линейной, квадратичной функции) в смысле оператора $\tilde{\Pi}$, если для соответствующих функций u, являющихся решениями уравнения (1), при всех n, s, j выполняется $\epsilon_j^{n,s} = 0$.

Учёт многостадийности методов Рунге-Кутты

Основной задачей настоящей работы является обобщение понятия нестационарного корректора на схемы высокого порядка аппроксимации. Предположим, что полудискретная схема (2) аппроксимирует пространственную производную точно на полиномах порядка q, то есть для всех решений $u(t, \boldsymbol{r})$ уравнения (1), являющихся полиномами порядка q, выполняется условие

$$L\Pi u = \Pi(\boldsymbol{a} \cdot \nabla u) = -\Pi\left(\frac{\partial u}{\partial t}\right) = -\frac{d}{dt}\Pi u. \tag{12}$$

Например, точность на линейной функции означает $L\Pi r^i = a^i$, а точность на квадратичной функции означает $L\Pi(r^i r^j) = a_i \Pi r^j + a_j \Pi r^i$.

Тогда для того, чтобы разностная схема (3) также была точна на полиномах порядка q, нужно задавать число стадий $S \geq q$. Однако при этом равенство $u_j^{n,s} = \Pi_j(u(t^{n,s,S}, \cdot))$ будет выполняться только при $s = 0$, и в смысле оператора Π (см. определение 1) схема (3) всё равно будет точна только на линейной функции, что затруднит дальнейший анализ.

Чтобы исправить этот недостаток, определим оператор $\tilde{\Pi}^0_{q,S}$, сопоставляющий q раз дифференцируемой функции $f(t, \boldsymbol{r})$ сеточную функцию $(\tilde{\Pi}^0_{q,S} f)_j^{n,s}$ следующим образом:

$$\left(\tilde{\Pi}^0_{q,S} f\right)_j^{n,s} = \Pi_j\left(f(t^{n,s,S}, \cdot)\right) + \sum_{r=1}^{\min\{q,S\}} \gamma_{r,s,S} (\Delta t)^r \Pi_j \left(\frac{\partial^r f(t^{n,s,S}, \cdot)}{\partial t^r}\right), \tag{13}$$

где коэффициенты γ определяются из двух условий:

- на целых временных слоях γ равны нулю: $\gamma_{r,0,S} = \gamma_{r,S,S} = 0$;
- для функций $f(t, \boldsymbol{r})$, являющихся полиномами q-го порядка, выполняется

$$\frac{(\tilde{\Pi}^0_{q,S} f)_j^{n,s+1} - (\tilde{\Pi}^0_{q,S} f)_j^0}{\omega_{s+1,S} \Delta t} - \left(\tilde{\Pi}^0_{q,S}\left(\frac{\partial f}{\partial t}\right)\right)_j^{n,s} = 0, \quad s = 0, \ldots, S-1, \tag{14}$$

при $S \geq q$.

Условие (14) означает, что если оператор L удовлетворяет условию (12) на полиномах порядка q, то схема (3) должна быть точной в смысле оператора $\tilde{\Pi}^0_{q,S}$ при аппроксимации по времени методом Рунге-Кутты при $S \geq q$.

Очевидно, что $\gamma_{r,s,S}$ не зависит от q. Покажем, каким образом они определяются, на примере $r = 1$ и $r = 2$. Ниже штрихом будем обозначать производные по времени. Символом $u(t^{n,s,S})$ будем обозначать соответствующее временное сечение функции $u(t, \boldsymbol{r})$.

Всюду в настоящем разделе будет подразумеваться интегрирование в интервале от t^n до t^{n+1}, поэтому индекс n будем опускать. Также зафиксируем число стадий метода Рунге-Кутты и всюду кроме окончательных выражений будем опускать индекс S.

Рассмотрим вначале случай $r = 1$. Пусть $u(t, \boldsymbol{r})$ – линейное решение (1). Подставим вид оператора $\tilde{\Pi}^0_{1,S}$ (13) в условие (14).

$$\frac{\Pi u(t^{s+1}) + \Delta t\, \gamma_{1,s+1} \Pi u'(t^{s+1}) - \Pi u(t^0)}{\omega_{s+1} \Delta t} - (\Pi u'(t^s) + \Delta t \gamma_{1,s} \Pi u''(t^s)) = 0.$$

Учитывая, что u является линейной функцией времени, получаем

$$\gamma_{1,s+1,S} = 0, \quad s = 0, \ldots, S-1.$$

Рассмотрим теперь случай $r = 2$. Пусть $u(t, \boldsymbol{r})$ – квадратичная функция, являющаяся решением (1). Подставим вид оператора $\tilde{\Pi}^0_{2,S}$ (13) в условие (14). Учитывая, что $\gamma_{1,s} = 0$, имеем

$$\frac{\Pi u(t^{s+1}) + (\Delta t)^2 \gamma_{2,s+1} \Pi u''(t^{s+1}) - \Pi u(t^0)}{\omega_{s+1} \Delta t} - \Pi u'(t^s) = 0.$$

Используя постоянство u'' и линейность u', получаем отсюда

$$\Pi u(t^{s+1}) = \Pi u(t^0) + \omega_{s+1} \Delta t \Pi u'(t^0) + (\omega_{s+1}\omega_s - \gamma_{2,s+1})(\Delta t)^2 \Pi u''(t^0).$$

Сравнивая с тейлоровским разложением, имеем

$$\gamma_{2,s+1,S} = \omega_{s,S}\omega_{s+1,S} - \frac{(\omega_{s+1,S})^2}{2} =$$

$$= \begin{cases} -(2S^2)^{-1}, & s = 0 \\ -(S-s-1)/[2(S-s)^2(S-s+1)], & s = 1, \ldots, S-1. \end{cases} \qquad (15)$$

Выполнение условия $\gamma_{r,S,S} = 0$ свидетельствует о том, что рассматриаемый двухстадийный метод Рунге-Кутты точен на квадратичных полиномах.

Коэффициенты γ более высокого порядка выражаются последовательно через предыдущие. При $S \geq r \geq 2$, $s > 0$ рекуррентная формула имеет вид

$$\gamma_{r,s+1,S} = \gamma_{r-1,s,S}\, \omega_{s+1,S} - \sum_{a=1}^{r-1} \gamma_{a,s+1,S} \frac{1}{(r-a)!} \left(\omega_{s+1,S} - \omega_{s,S}\right)^{r-a} -$$

$$-\frac{1}{r!} \left((\omega_{s+1,S} - \omega_{s,S})^r - (-\omega_{s,S})^r \right).$$

Поскольку она в явном виде нигде использоваться не будет, мы оставим её без доказательства. При $r > S$ формально положим $\gamma_{r,s,S} = 0$.

Нестационарный корректор

Понятие нестационарного корректора можно определить для произвольной схемы вида (3), (9), (10). Всюду далее мы будем предполагать, что рассматриваемая схема точна на константе. В силу свойств (5)–(6) оператора Π условие точности на константе выражается равенством

$$\sum_{k\in M} L_{jk} = 0, \quad \forall j \in M^0.$$

Отметим, что это условие точности может не выполняться, например, для некоторых конечно-разностных схем в криволинейных координатах.

Пусть $m = (m_1, \ldots, m_d)$ – мультииндекс: $m_i \geq 0$, $|m| = m_1 + \ldots + m_d$, $m! = m_1! \ldots m_d!$. Введём обозначения

$$\boldsymbol{r}^m = x_1^{m_1} \ldots x_d^{m_d}, \quad D^m = \frac{\partial^{|m|}}{\partial x_1^{m_1} \ldots x_d^{m_d}}.$$

Определение 2. *Пусть Π – некоторый оператор проектирования вида (5)–(8). Пусть $q \in \mathbb{N}$. Рассмотрим оператор $\tilde{\Pi}_{q,S}$, сопоставляющий q раз дифференцируемым функциям $f(t, \boldsymbol{r})$ сеточную функцию $(\tilde{\Pi}_{q,S} f)_j^{n,s}$ следующим образом:*

$$\left(\tilde{\Pi}_{q,S} f\right)_j^{n,s} = \left(\tilde{\Pi}^0_{\min\{q,S\},S} f\right)_j^{n,s} + \sum_{0<|m|\leq q} (C^m)_j^{n,s}\, \Pi_j \left(D^m f(t^{n,s}, \cdot)\right) =$$

$$= \Pi_j(f(t^{n,s,S}, \cdot)) + \sum_{r=1}^{\min\{q,S\}} \gamma_{r,s,S} (\Delta t)^r \Pi_j \left(\frac{\partial^r f(t^{n,s,S}, \cdot)}{\partial t^r}\right) +$$

$$+ \sum_{0<|m|\leq q} (C^m)_j^{n,s}\, \Pi_j \left(D^m f(t^{n,s}, \cdot)\right). \qquad (16)$$

Обозначим через $(\boldsymbol{C}^p)_j^{n,s}$ совокупность наборов коэффициентов $(C^m)_j^{n,s}$, $|m| = p$. Будем называть набор коэффициентов $(\boldsymbol{C}^p)_j^{n,s}$, нестационарным корректором порядка p для схемы (3), если он удовлетворяет двум условиям:

- *схема (3) точна на любом решении уравнения (1), являющемся полиномом порядка p, в смысле оператора $\tilde{\Pi}_{q,S}$, $q \geq p$;*
- $(\boldsymbol{C}^p)_j^{n,s} = 0 \quad \forall j \in M^b$.

Если схема точна на полиномах порядка q в смысле оператора $\tilde{\Pi}^0_{q,S}$, то нестационарные корректоры порядка q и ниже тождественно равны нулю. То есть для схем, точных на константе, вычисление нестационарного корректора нужно начинать с первого порядка, для схем, точных на линейной функции – со второго и т. д.

В настоящей работе мы будем рассматривать нестационарные корректоры первого, второго и третьего порядков. Введём для них разные обозначения: нестационарный корректор первого порядка обозначим за $\mathbf{\Gamma}$, второго порядка – за $\boldsymbol{\alpha}$, третьего порядка – за $\boldsymbol{\beta}$, то есть $\mathbf{\Gamma} = \boldsymbol{C}^1$, $\boldsymbol{\alpha} = \boldsymbol{C}^2$, $\boldsymbol{\beta} = \boldsymbol{C}^3$.

Уравнение нестационарного корректора

Выведем уравнения для коэффициентов нестационарного корректора для схемы (3). Предполагая, что схема точна на константе, начнём с корректора первого порядка. Всюду в настоящем разделе будет подразумеваться интегрирование в интервале от t^n до t^{n+1}, поэтому индекс n будем опускать. Также зафиксируем число стадий метода Рунге-Кутты и будем опускать индекс S.

По определению, нестационарный корректор 1-го порядка для схемы (3) есть такой набор коэффициентов, при котором эта схема точна на линейном решении уравнения (1) в смысле оператора $\widetilde{\Pi}_{q,S}$, $q \geq 1$. Для получения уравнения на $\mathbf{\Gamma}$ подставим функцию $u(t, \boldsymbol{r}) = \boldsymbol{r} - \boldsymbol{a}t$ в оператор (16) и далее в определение аппроксимации (11) для схемы (3). Имеем

$$\frac{\Pi_j(\boldsymbol{r} - \boldsymbol{a}(t + \omega_{s+1}\Delta t)) + \mathbf{\Gamma}_j^{s+1} - \Pi_j(\boldsymbol{r} - \boldsymbol{a}t) - \mathbf{\Gamma}_j^0}{\omega_{s+1}\Delta t} +$$

$$+ \sum_{k \in M} L_{jk} \left(\Pi_k(\boldsymbol{r} - \boldsymbol{a}(t + \omega_s \Delta t) + \mathbf{\Gamma}_k^s\right) = 0.$$

Пользуясь точностью схемы на константе, получаем

$$\frac{\mathbf{\Gamma}_j^{s+1} - \mathbf{\Gamma}_j^0}{\omega_{s+1}\Delta t} = \boldsymbol{a} - \sum_{k \in M} L_{jk} \left(\mathbf{\Gamma}_k^s + \Pi_k(\boldsymbol{r})\right). \tag{17}$$

Легко убедиться, что подстановка функции $u(t, \boldsymbol{r}) = c(\boldsymbol{r} - \boldsymbol{a}t) + d$ даст это же уравнение. Уравнение (17) дополняется граничными условиями $\mathbf{\Gamma}_j^s = 0$ при $j \in M^b$. На нулевой стадии $\mathbf{\Gamma}$ есть значение на n-м временном шаге, а на стадии S – значение на $(n+1)$-м шаге.

Теперь предположим, что схема (3) точна на полиномах до $(p-1)$-го порядка включительно в смысле (12), но не точна на полиномах p-го порядка. Тогда нестационарные корректоры до $(p-1)$-го порядка включительно тождественно равны нулю, а нестационарный корректор $\boldsymbol{C}^p$ отличен от нуля. Для краткости далее будем называть его **главным** корректором. Предположим, что $S \geq p$. Для получения уравнения главного корректора рассмотрим функцию

$$f = \frac{1}{m!}(\boldsymbol{r} - \Pi_j(\boldsymbol{r}) - \boldsymbol{a}(t - t^{n,s}))^m, \tag{18}$$

причём $|m| = p$. Подставим её в формулу (11) и приравняем к нулю аппроксимационную ошибку. Получаем

$$\frac{(\tilde{\Pi}_p f)_j^{s+1} - (\tilde{\Pi}_p f)_j^0}{\omega_{s+1}\Delta t} + \sum L_{jk}(\tilde{\Pi}_p f)_k^s = 0.$$

Поскольку

$$(\tilde{\Pi}_p f)_k^{\tilde{s}} = (\tilde{\Pi}_p^0 f)_k^{\tilde{s}} + (C^m)_k^{\tilde{s}},$$

имеем

$$\frac{(C^m)_j^{s+1} - (C^m)_j^0}{\omega_{s+1}\Delta t} + \sum_{k\in M} L_{jk}(C^m)_k^s = (f^m)_j^s,$$

$$(f^m)_j^s = -\frac{(\tilde{\Pi}_p^0 f)_j^{s+1} - (\tilde{\Pi}_p^0 f)_j^0}{\omega_{s+1}\Delta t} - \sum_{k\in M} L_{jk}(\tilde{\Pi}_p^0 f)_k^s.$$

Используя определение (14) оператора $\tilde{\Pi}_p^0$, преобразуем первое слагаемое:

$$(f^m)_j^s = -\left(\tilde{\Pi}_p^0 \partial_t f\right)_j^s - \sum_{k\in M} L_{jk}(\tilde{\Pi}_p^0 f)_k^s.$$

Здесь и далее обозначено $\partial_t^n f = \partial^n f / \partial t^n$. Теперь используем первую часть (13) определения оператора $\tilde{\Pi}_{|m|}^0$:

$$(f^m)_j^s = -\Pi_j \left(\partial_t f(t^s)\right) - \sum_{k\in M} L_{jk}\Pi_j f(t^s) +$$

$$+ \sum_{r=2}^{p} \gamma_{r,s}(\Delta t)^r \left(-\Pi_j \partial_t^{r+1} f(t^s) - \sum_{k\in M} L_{jk}\Pi_k \partial_t^r f(t^s) \right),$$

Функция $\partial_t^r f$, $r > 0$, является полиномом порядка не выше $p-1$, поэтому по условию точности оператора L на ней (12) выражение в скобках равно нулю. Поэтому

$$(f^m)_j^s = -\Pi_j \partial_t f(t^s) - \sum_{k\in M} L_{jk}\Pi_j f(t^s) = \Pi_j(\boldsymbol{a}\cdot\nabla f(t^s)) - \sum_{k\in M} L_{jk}\Pi_j f(t^s).$$

Таким образом, уравнение главного корректора имеет вид

$$\frac{(\boldsymbol{C}^p)_j^{n,s+1} - (\boldsymbol{C}^p)_j^n}{\omega_{s+1}\Delta t} + \sum_{k\in M} L_{jk}(\boldsymbol{C}^p)_k^{n,s} = (\boldsymbol{f}^p)_j, \quad j \in M^0, \tag{19}$$

где $(\boldsymbol{f}^p)_j$ – не зависящий от времени набор компонент

$$(f^m)_j = \Pi_j \left((\boldsymbol{a}\cdot\nabla)\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right) - \sum_{k\in M} L_{jk}\Pi_k\left(\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right), \quad |m| = p. \tag{20}$$

Видно, что уравнение главного корректора является аппроксимацией методом Рунге-Кутты системы обыкновенных дифференциальных уравнений

$$\frac{d\boldsymbol{C}^p}{dt} + L\boldsymbol{C}^p = \boldsymbol{f}^p, \quad \boldsymbol{C}^p(0) = 0,$$

правая часть (20) которой не зависит от времени и представляет собой ошибку аппроксимации в смысле оператора Π схемы (2) на нормированном полиноме порядка p. Частным случаем уравнений (19)–(20) при $p = 1$ является уравнение линейного корректора (17).

Теперь рассмотрим нестационарный корректор $(p+1)$-го порядка, предполагая, что чисто стадий метода Рунге-Кутты не меньше, чем $p+1$. Подставим многочлен (18), где $|m| = p+1$, в формулу (11) и приравняем к нулю аппроксимационную ошибку:

$$\frac{(\tilde\Pi_{p+1}f)_j^{s+1} - (\tilde\Pi_{p+1}f)_j^0}{\omega_{s+1}\Delta t} + \sum L_{jk}(\tilde\Pi_{p+1}f)_k^s = 0. \tag{21}$$

По определению оператора $\tilde\Pi_{p+1}$ выполняется

$$(\tilde\Pi_{p+1}f)_k^{\tilde s} = (\tilde\Pi_{p+1}^0 f)_k^{\tilde s} + \sum_{|i|=1, i<m} (C^{m-i})_k^{\tilde s}\Pi_k\left(\boldsymbol{r} - \Pi(\boldsymbol{r}) - \boldsymbol{a}(t^{\tilde s} - t^s)\right)^i + (C^m)_k^{\tilde s} =$$

$$= (C^m)_k^{\tilde s} + (\tilde\Pi_{p+1}^0 f)_k^{\tilde s} + \sum_{|i|=1, i<m} (C^{m-i})_k^{\tilde s}\left((\Pi_k(\boldsymbol{r}^i) - \Pi_j(\boldsymbol{r}^i)) - \boldsymbol{a}^i(t^{\tilde s} - t^0) + \boldsymbol{a}^i(t^s - t^0)\right).$$

Подставляя это выражение в (21), получаем

$$\frac{(C^m)_j^{s+1} - (C^m)_j^0}{\omega_{s+1}\Delta t} + \sum_{k\in M} L_{jk}(C^m)_k^s = (\tilde f^m)_j^s, \tag{22}$$

$$(\tilde f^m)_j^s = \Pi_j\left((\boldsymbol{a}\cdot\nabla)\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right) - \sum_{k\in M} L_{jk}\Pi_k\left(\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right) +$$

$$+ \sum_{|i|=1, i<m}\left((C^{m-i})_j^{s+1}\boldsymbol{a}^i + \sum_{k\in M} L_{jk}(C^{m-i})_k^s(\Pi_k(\boldsymbol{r}^i) - \Pi_j(\boldsymbol{r}^i))\right) -$$

$$- \frac{\omega_s}{\omega_{s+1}} \sum_{|i|=1, i<m}\left((C^{m-i})_j^{s+1} - (C^{m-i})_j^0\right)\boldsymbol{a}^i = 0. \tag{23}$$

При выводе (22)–(23) используются те же соображения, что и при получении уравнения главного корректора, и учтено, что $\gamma_{1,s,S} = 0$. В случае $\boldsymbol{C}^p = 0$ формулы (22)–(23), очевидно, сводится с системе (19)–(20).

В общем случае нестационарный корректор $(p+1)$-го порядка также подчиняется разностному уравнению с матрицей, взятой из разностной схемы (3), с однородными начальными и граничными условиями и с ненулевой правой частью. Однако эта правая часть зависит от главного корректора и, следовательно, от времени. Сравнивая уравнение (22)–(23) с аппроксимацией S-стадийным, $S \geq p+1$, методом Рунге-Кутты системы ОДУ

$$\frac{d\boldsymbol{C}^{p+1}}{dt} + L\boldsymbol{C}^{p+1}(t) = \tilde{f}^m(\boldsymbol{C}^p(t)), \quad \boldsymbol{C}^{p+1}(0) = 0, \tag{24}$$

можно заметить, что отличие между ними обращается в ноль при постоянных во времени коэффициентах главного корректора. Таким образом, если $\boldsymbol{C}^p$ перестаёт зависеть от времени, то для анализа роста во времени коэффициентов $\boldsymbol{C}^{p+1}$ достаточно анализировать поведение полудискретного уравнения (24). Этот факт существенно упрощает анализ точности разностной схемы при больших временах счёта.

Для нестационарного корректора 2-го порядка (22)–(23) принимает вид

$$\begin{aligned}
&\frac{\boldsymbol{\alpha}_j^{s+1} - \boldsymbol{\alpha}_j^0}{\omega_{s+1}\Delta t} - \Pi_j(\boldsymbol{r}) \otimes \left(\sum_k L_{jk} \left(\boldsymbol{\Gamma}_k^s + \Pi_k(\boldsymbol{r}) \right) \right) - \boldsymbol{a} \otimes \boldsymbol{\Gamma}_j^s + \\
&\quad + \sum_k L_{jk} \left(\boldsymbol{\alpha}_k^s + \Pi_k(\boldsymbol{r}) \otimes \boldsymbol{\Gamma}_k^s + \Pi_k \left(\frac{\boldsymbol{r} \otimes \boldsymbol{r}}{2} \right) \right) = \\
&\quad = -\frac{\omega_s}{\omega_{s+1}} \boldsymbol{a} \otimes (\boldsymbol{\Gamma}_j^{s+1} - \boldsymbol{\Gamma}_j^0) + \boldsymbol{a} \otimes (\boldsymbol{\Gamma}_j^{s+1} - \boldsymbol{\Gamma}_j^s).
\end{aligned}$$

где $(\boldsymbol{a} \otimes \boldsymbol{b})^{xx} = a^x b^x$, $(\boldsymbol{a} \otimes \boldsymbol{b})^{xy} = a^x b^y + a^y b^x$ и т. д.

При $S < q$ приведённые выше формулы неприменимы. В частности, при $S = 1$ для нестационарного корректора 2-го порядка получается уравнение

$$\begin{aligned}
&\frac{\boldsymbol{\alpha}_j^{n+1} - \boldsymbol{\alpha}_j^n}{\Delta t} = -\sum_k L_{jk} \left(\boldsymbol{\alpha}_k^n + \Pi_k(\boldsymbol{r}) \otimes \boldsymbol{\Gamma}_k^n + \Pi_k \left(\frac{\boldsymbol{r} \otimes \boldsymbol{r}}{2} \right) \right) + \\
&+ \boldsymbol{a} \otimes \boldsymbol{\Gamma}_j^n + (\Pi_j(\boldsymbol{r}) - \boldsymbol{a}\Delta t) \otimes \sum_k L_{jk} \left(\boldsymbol{\Gamma}_k^n + \Pi_k(\boldsymbol{r}) \right) + \frac{\boldsymbol{a} \otimes \boldsymbol{a}}{2} \Delta t.
\end{aligned}$$

Зависимость правой части от шага по времени связана с тем, что использование явной схемы первого порядка для интегрирования по времени вносит ошибку, пропорциональную второй производной от решения.

Нестационарный корректор и точность схемы

Теорема 1. *Рассмотрим уравнение переноса (1) и некоторую схему (3). Предположим, что выполнены следующие условия.*

1. *Решение $u(\boldsymbol{r},t)$ уравнения (1) q раз дифференцируемо и имеет липшицевы q-е пространственные производные с константой Липшица $\mathbb{L}$.*
2. *Π – некоторый оператор проектирования вида (5)–(8), причём радиус, входящий в формулу (7), удовлетворяет условию $R \le C_1 h$.*
3. *Коэффициенты схемы (3) удовлетворяют условию $|L_{jk}| \le C_2/h$.*
4. *Количество ненулевых коэффициентов в каждой строке матрицы L схемы (3) не превосходит $C_3 : \forall j \in M^0\ |\{k : L_{jk} \ne 0\}| \le C_3$.*
5. *Оператор L локализован в области диаметром порядка h: $L_{jk} = 0$ при $|\Pi_k(\boldsymbol{r}) - \Pi_j(\boldsymbol{r})| > C_4 h$.*
6. *Решение неоднородного уравнения*

$$\frac{u^{n,s+1} - u^n}{\omega_{s+1,S}\Delta t} + Lu^{n,s} = f^{n,s}, \quad s = 0, \ldots, S-1$$

   с нулевыми начальными и граничными условиями удовлетворяет оценке $||u^{n,0}||_p \le K_n \max\limits_{n' \le n,s} ||f^{n',s}||_p$.
7. *Схема (3) точна на константе: $\forall j \in M^0 \sum L_{jk} = 0$.*
8. *Интегрирование по времени ведётся методом Рунге-Кутты q-го порядка или выше.*

Пусть $\{\boldsymbol{C}^m\}$, $|m| = q$, – нестационарный корректор порядка q, соответствующий схеме (3) и оператору Π. Тогда выполняется оценка

$$||u_j^n - (\tilde{\Pi}_{q,S} u)_j^{n,0}||_p \le$$

$$\le K_n \tilde{C} \mathbb{L} \left((h + |\boldsymbol{a}|\Delta t)^q + \sum_{0<|m|\le q} ||(C^m)^n||_p (h + |\boldsymbol{a}|\Delta t)^{q-|m|} \right) \left(1 + \frac{\Delta t |\boldsymbol{a}|}{h} \right), \tag{25}$$

где константа $\tilde{C} = (1 + 2(C_1 + C_4)^{q+1}) C_2 (2 + C_3)$, а $(\tilde{\Pi}_{q,S} u)_j^{n,0}$ определяется формулой (16) с учётом $\gamma_{r,0,S} = 0$:

$$(\tilde{\Pi}_{q,S} u)_j^{n,0} = \Pi_j(u(t^n, \cdot)) + \sum_{0<|m|\le q} (C^m)_j^{n,0} \Pi_j(D^m u(t^n, \cdot)).$$

Доказательство. Рассмотрим произвольную $j \in M$ и произвольный момент времени $t^{n,s,S}$. Представим функцию $u(t,\boldsymbol{r})$ в виде

$$u(t,\boldsymbol{r}) = p_{j,n,s}(t,\boldsymbol{r}) + g_{j,n,s}(t,\boldsymbol{r}),$$

где $p_{j,n,s}(t,\boldsymbol{r})$ – полином порядка q от t и $\boldsymbol{r}$, представляющий собой первые $q+1$ членов разложения функции $u(t,\boldsymbol{r})$ в ряд Тейлора около точки $(t^{n,s},\Pi_j(\boldsymbol{r}))$:

$$p_{j,n,s}(t,\boldsymbol{r}) = \sum_{0\le|m|\le q} \frac{1}{m!}\left(\boldsymbol{r}-\Pi_j(\boldsymbol{r})-\boldsymbol{a}\left(t-t^{n,s,S}\right)\right)^m D^m u(t^{n,s,S},\Pi_j(\boldsymbol{r})).$$

Этот полином также является решением (1). Поскольку p – полином порядка q, то $g_{j,n,s} = u - p_{j,n,s}$ является q раз дифференцируемой функцией с липшицевыми q-ми производными с константой Липшица $\mathbb{L}$. Кроме того, $D^m g_{j,n,s}(t^{n,s,S},\Pi_j(\boldsymbol{r})) = 0$ при $|m| \le q$. Следовательно, при $|m| \le q$ имеет место оценка

$$\begin{gathered}|D^m g_{j,n,s}(t,\boldsymbol{r})| = \left|D^m g_{j,n,s}(t^{n,s,S},\boldsymbol{r}-\boldsymbol{a}(t-t^{n,s,S}))\right| \le \\ \le \mathbb{L}\left|\boldsymbol{r}-\boldsymbol{a}(t-t^{n,s,S})-\Pi_j(\boldsymbol{r})\right|^{q+1-|m|}.\end{gathered} \tag{26}$$

Отсюда в силу свойств оператора Π и неравенства треугольника следует

$$\begin{gathered}|\Pi_k\left(D^m g_{j,n,s}(t,\ \cdot\ )\right)| \le \\ \le \mathbb{L}\int\left(|\boldsymbol{r}-\Pi_j(\boldsymbol{r})|+\left|\boldsymbol{a}(t-t^{n,s,S})\right|\right)^{q+1-|m|} d\mu_k \le \\ \le \mathbb{L}\left(R+|\Pi_k(\boldsymbol{r})-\Pi_j(\boldsymbol{r})|+|\boldsymbol{a}|\,|t-t^{n,s,S}|\right)^{q+1-|m|},\end{gathered}$$

и при $L_{jk}\neq 0$ имеет место

$$|\Pi_k\left(D^m g_{j,n,s}(t,\ \cdot\ )\right)| \le \mathbb{L}\left((C_1+C_4)h+|\boldsymbol{a}|\,|t-t^{n,s,S}|\right)^{q+1-|m|}.$$

Поскольку временные производные выражаются через пространственные, для них получается аналогичная оценка. Отсюда в силу определения оператора $\tilde{\Pi}_q$ имеем

$$\begin{gathered}\left|(\tilde{\Pi}_q(g_{j,n,s}))_k^{n,\tilde{s}}\right| \le \mathbb{L}(1+(C_1+C_4)^{q+1})\times \\ \times\left((h+|\boldsymbol{a}|\Delta t)^{q+1}+\sum_{0<|m|\le q}(h+|\boldsymbol{a}|\Delta t)^{q+1-|m|}|(C^m)_j^{n,s}|\right).\end{gathered} \tag{27}$$

Далее, рассмотрим выражение

$$Y_j^m = \frac{\Pi_j\left(D^m g_{j,n,s}(t^{n,s+1,S},\ \cdot\ )\right)-\Pi_j\left(D^m g_{j,n,s}(t^{n,0,S},\ \cdot\ )\right)}{\omega_{s+1,S}\Delta t}$$

В силу свойств оператора Π имеем

$$|Y_j^m| \le \sup_{\boldsymbol{r}\in B_R(\Pi_j(\boldsymbol{r}))} \left| \frac{D^m g_{j,n,s}(t^{n,s+1,S},\boldsymbol{r}) - D^m g_{j,n,s}(t^{n,0,S},\boldsymbol{r})}{t^{n,s+1,S} - t^{n,0,S}} \right|. \tag{28}$$

При $|m| < q$ функция $D^m g_{j,n,s}$ дифференцируема, поэтому можно воспользоваться теоремой Лагранжа:

$$|Y_j^m| \le \sup_{\boldsymbol{r}\in B_R(\Pi_j(\boldsymbol{r}))} \left| D^m \frac{\partial g_{j,n,s}}{\partial t}(t+\xi,\boldsymbol{r}) \right|,$$

где $0 \le \xi \le \omega_{s+1}\Delta t$. Выражая временную производную через пространственные и пользуясь оценкой (26), получаем

$$|Y_j^m| \le |\boldsymbol{a}|\mathbb{L}(R+|\boldsymbol{a}|\Delta t)^{q-|m|}. \tag{29}$$

При $|m| = q$ в силу липшицевости функции $D^m g_{j,n,s}$ напрямую из (28) имеем $|Y_j^m| \le |\boldsymbol{a}|\mathbb{L}$, поэтому (29) верна при $|m| \le q$. Подставляя оценку (29) в определение оператора $\tilde{\Pi}_q$, получаем оценку

$$\begin{gathered} \left| \frac{(\tilde{\Pi}_q g_{j,n,s})_j^{n,s+1} - (\tilde{\Pi}_q g_{j,n,s})_j^n}{\omega_{s+1,S}\Delta t} \right| \le \\ \le (C_1)^q \mathbb{L} \left((h+|\boldsymbol{a}|\Delta t)^q + \sum_{0<|m|\le q} (h+|\boldsymbol{a}|\Delta t)^{q-|m|} |(C^m)_j^{n,s}| \right). \end{gathered} \tag{30}$$

Поскольку по определению нестационарного корректора q-го порядка в смысле $\tilde{\Pi}_q$ схема (3) точна на полиномах порядка q, то $\epsilon(p_{j,n,s}) = 0$, и $\epsilon_j^{n,s}(u) = \epsilon_j^{n,s}(g_{j,n,s})$. По определению $\epsilon_j^{n,s}(g_{j,n,s})$ запишем

$$\epsilon_j^{n,s}(u) = \epsilon_j^{n,s}(g_{j,n,s}) = -\frac{(\tilde{\Pi}_q g_{j,n,s})_j^{n,s+1} - (\tilde{\Pi}_q g_{j,n,s})_j^n}{\omega_{s+1,S}\Delta t} - \sum_{k\in M} L_{jk} (\tilde{\Pi}_q g_{j,n,s})_k^{n,s}. \tag{31}$$

Подставляя (27) и (30) в (31) и учитывая, что $|L_{jk}| \le C_2/h$, получаем

$$\left|\epsilon_j^{n,s}(u)\right| = \left|\epsilon_j^{n,s}(g_{j,n,s})\right| \le$$

$$\le \tilde{C}\mathbb{L} \left((h+|\boldsymbol{a}|\Delta t)^q + \sum_{0<|m|\le q} (h+|\boldsymbol{a}|\Delta t)^{q-|m|} |(C^m)_j^{n,s}| \right) \left(1 + \frac{|\boldsymbol{a}|\Delta t}{h} \right).$$

Для любого $p \in \mathbb{N} \cup \{\infty\}$ из этого следует оценка

$$||\epsilon^{n,s}(u)||_p \le \tilde{C}\mathbb{L}\left((h + |\boldsymbol{a}|\Delta t)^q + \right.$$
$$\left. + \sum_{0<|m|\le q} (h + |\boldsymbol{a}|\Delta t)^{q-|m|} ||(C^m)^{n,s}||_p \right) \left(1 + \frac{|\boldsymbol{a}|\Delta t}{h}\right). \tag{32}$$

Теперь представим численное решение в виде

$$u_j^{n,s} = \left(\tilde{\Pi}_q(u)\right)_j^{n,s} + \varepsilon_j^{n,s}. \tag{33}$$

Подставляя выражение (33) в схему (3), получим уравнение для ε:

$$\frac{\varepsilon_j^{n,s+1} - \varepsilon_j^n}{\omega_{s+1}\Delta t} + \sum_k L_{jk}\varepsilon_j^{n,s} = -\epsilon_j^{n,s}(u), \tag{34}$$

где $\epsilon_j^{n,s}(u)$ – аппроксимационная ошибка функции u схемой (3) в смысле оператора $\tilde{\Pi}_q$, см. определение 1. Из условия 6 следует, что

$$||u^{n,s} - (\tilde{\Pi}_q(u))^{n,s}||_p = ||\varepsilon^{n,s}(u)||_p \le K_n||\epsilon^{n,s}(u)||_p.$$

Подставляя оценку (32) для $||\epsilon^{n,s}(u)||_p$, полагая $s = 0$ и учитывая, что на целых временных слоях $\gamma_{r,0,S} = \gamma_{r,S,S} = 0$, получаем искомую оценку (25).

Теорема доказана.

Прямым следствием этой теоремы является следующая.

Теорема 2. *Рассмотрим последовательность сеток с $h \to 0$. Пусть шаг по времени удовлетворяет условию $\Delta t \le \sigma h$. Пусть выполняются все условия теоремы 1, причём константы не зависят от h. Тогда для всех $r \le q$ отличие численного решения от точного удовлетворяет оценке*

$$||u^n - \Pi u(t^n)||_p = O\left(h^r + \sum_{0<|m|\le r} ||(C^m)^n||_p \sup_{\boldsymbol{r}\in G} |D^m u(t^n, \boldsymbol{r})|\right). \tag{35}$$

Оценка (35), как правило, является оптимальной, тогда как оценка (25) – не оптимальной для большинства схем. Неточность оценки (25) связана с тем, что использование оптимальной оценки для $||\epsilon||_p$ при умножении на K не даёт оптимальной оценки на $||\varepsilon||_p$. Фактически, при этом мы ограничиваем устанавливаемый порядок точности порядком аппроксимации.

Периодические условия

Рассмотрим уравнение переноса (1) в d-мерном кубе с периодическими граничными условиями по одному или нескольким направлениям. Запишем их в виде

$$u(t, \boldsymbol{r}) = u(t, \boldsymbol{r} + \boldsymbol{b}_i), \quad i = 1, \ldots, \tilde{d}, \tag{36}$$

где N_p – число линейно независимых направлений, по которым задаются периодические условия, а $\boldsymbol{b}_i \in \mathbb{R}^{\tilde{d}}$ – соответствующие периоды.

Интерпретируем задачу (1), (36) как задачу с периодическими начальными данными в $\mathbb{R}^d$. Пусть множество неизвестных M в пространстве представляется в виде $M = M^0 \times \mathbb{Z}^{\tilde{d}}$, и их индексы представляются в виде $j = [\xi, \eta]$, где $\xi \in M^0$ – индекс неизвестной внутри блока, а $\eta \in \mathbb{Z}^{\tilde{d}}$ – индекс блока. Будем также использовать запись $\xi(j)$ и $\eta(j)$ для разложения индекса неизвестной на её индекс в блоке и индекс блока.

Периодические условия задаются в виде

$$u_j^{n,s} = u_k^{n,s}, \quad \xi(j) = \xi(k). \tag{37}$$

Условию периодичности также должен подчиняться проектор Π. Поэтому дополнительно к (5)–(8) предположим, что при $\xi(j) = \xi(k)$ мера μ_j является пространственной трансляцией меры μ_k на вектор $\sum_i (\eta_i(j) - \eta_i(k))\boldsymbol{b}_i$. То есть для любой непрерывной функции f выполняется

$$\int f\left(\boldsymbol{r} + \sum_i (\eta_i(j) - \eta_i(k))\boldsymbol{b}_i\right) d\mu_j = \int f(\boldsymbol{r}) d\mu_k. \tag{38}$$

Записанные выражения можно пояснить следующим образом. Пусть расчётная область покрыта неструктурированной сеткой, причём сетка на противолежащих гранях совпадает. Запишем на этой сетке схему (3), предполагая, что во всех копиях исходного сеточного блока она записывается одинаковым образом. Если рассматривать схемы с определением одной переменной на каждой сеточной ячейке, то M^0 будет множеством сеточных ячеек одного блока. Если же рассматривать схемы с определением одной переменной на узел, например, в трёхмерном случае, то M^0 будет содержать набор внутренних узлов блока, набор внутренних узлов одной грани из каждой пары противолежащих, набор внутренних узлов одного ребра из каждой четвёрки параллельных и один из восьми угловых узлов куба.

Определим квадратную матрицу $\hat{L}$ размерности $M^0 \times M^0$ равенствами

$$\hat{L}_{jk} = \sum_{m \in M:\ \xi(m) = \xi(k)} L_{jm}. \tag{39}$$

Тогда схему общего вида (3) при наличии периодических условий можно записать в матричном виде

$$\frac{u^{n,s+1}-u^n}{\omega_{s+1,S}\Delta t}+\hat{L}u^{n,s}=0,\quad s=0,\ldots,S-1, \tag{40}$$

где $\omega_{s,S}$ определено формулой (4).

Определим понятие блочного измельчения.

Определение 3. *Рассмотрим уравнение переноса (1) с периодическими условиями по всем направлениям. Рассмотрим последовательность разностных задач вида (3) с матрицами L_K, $K=1,\ldots,\infty$. Пусть в каждой задаче периодические условия реализуются соотношениями (37)–(38). Будем называть эту последовательность задач блочным измельчением, если выполняются следующие три условия.*

1. *Матрицы L_K удовлетворяют соотношениям $(L_K)_{jk}=KL_{jk}$, $j,k\in M$.*
2. *Операторы проектирования имеют вид*

$$(\Pi_K)_j(f)=\int\limits_{\mathbb{R}^d} f\left(\frac{\boldsymbol{r}}{K}\right)d\mu_j,$$

   где μ_j – мера, входящая в оператор проектирования Π (5) при $K=1$.
3. *Шаг по времени определяется формулой $\Delta t=\tau(K)=\tau(1)/K$.*

Очевидно, что при блочном измельчении периодические условия для задачи с матрицей L_K и оператором Π_K задаются как

$$u_j^{n,s}=u_k^{n,s},\quad \xi(j)=\xi(k),\ \eta(j)=\eta(k)\bmod K. \tag{41}$$

Блочное измельчение можно проиллюстрировать следующим образом. Пусть задана некоторая неструктурированная сетка в d-мерном кубе, такая что сетка на противоположных гранях куба совпадает. Уменьшим данный куб в K раз по всем направлениям и заполним исходную расчётную область K^d блоками исходной сетки. При этом удалим дублируемые сеточные примитивы, лежащие на склеиваемых гранях кубических блоков.

Автомодельность главного корректора

При выполнении условий (37)–(38) нестационарный корректор также подчиняется условию периодичности $(C^m)_j^{n,s}=(C^m)_k^{n,s}$, $\xi(j)=\xi(k)$. Поэтому, например, уравнение для главного корректора (19) теперь записывается в виде

$$\frac{(\boldsymbol{C}^p)^{n,s+1}-(\boldsymbol{C}^p)^n}{\omega_{s+1}\Delta t}+\hat{L}(\boldsymbol{C}^p)^{n,s}=\boldsymbol{f}^p, \tag{42}$$

где $\boldsymbol{f}^p$ определено формулой (20). Однако в (20) суммирование не может быть напрямую сведено к суммированию по $k \in M^0$, поскольку вектор $\boldsymbol{r}$, его степени и, как следствие, выражение под знаком суммы не являются периодическими функциями.

При блочном измельчении периодические условия на решение задаются формулой (41). При этом периодом является количество неизвестных в исходном d-мерном кубе: $K^d|M^0|$. Однако нестационарный корректор периодичен с периодом $|M^0|$, то есть одинаков во всех блоках. Таким образом, можно считать, что вектора главного корректора при блочном измельчении принадлежат одному тому же пространству $\mathbb{R}^{M^0}$.

Введём обозначение $(\boldsymbol{C}_K^p)_j^n$, $j \in M^0$, для коэффициентов главного корректора для задачи K при $s = 0$ (т. е. на целых временных слоях). Докажем утверждение об автомодельности главного корректора.

Утверждение 3. *Рассмотрим последовательность задач, являющихся блочным измельчением. Пусть $n \in \mathbb{N}$ – произвольное натуральное число, и $N = Kn$. Тогда главный корректор удовлетворяет равенству*

$$\frac{\ln \left\|(\boldsymbol{C}_1^p)^n\right\| - \ln \left\|(\boldsymbol{C}_K^p)^N\right\|}{\ln K} = p - \frac{\ln \left\|(\boldsymbol{C}_1^p)^N\right\| - \ln \left\|(\boldsymbol{C}_1^p)^n\right\|}{\ln(N/n)}. \tag{43}$$

В этом выражении p – порядок главного корректора, то есть минимальный порядок полинома, на котором схема не точна, а норма любая.

Доказательство. Рассмотрим уравнение (42). В целом для шага по времени при будет выполняться равенство

$$(\boldsymbol{C}_K^p)^{n+1} = \left(I + \sum_{k=1}^{S} \frac{\left(-\Delta t \hat{L}\right)^k}{k!} \right) (\boldsymbol{C}_K^p)^n + \left(\sum_{k=1}^{S} \frac{\left(-\Delta t \hat{L}\right)^{k-1}}{k!} \right) (\Delta t) \boldsymbol{f}_K^p. \tag{44}$$

По условию выражение $(\Delta t)\hat{L}$ не зависит от K, поэтому матрицы в обеих скобках не зависят от K. Обозначим их за M_1 и M_2. Из (20) видно, что при блочном измельчении коэффициенты вектора $\boldsymbol{f}_K^p$ пропорциональны $(1/K)^{p-1}$. Тогда можно записать

$$(\boldsymbol{C}_K^p)^{N+1} = M_1 (\boldsymbol{C}_K^p)^N + \frac{1}{K^p} M_2 \boldsymbol{f}_1^p.$$

Сравнивая это выражение с аналогичным для $K = 1$ и учитывая начальные условия $(\boldsymbol{C}_K^p)^0 = 0$, получаем

$$(\boldsymbol{C}_K^p)^N = K^{-p} (\boldsymbol{C}_1^p)^N.$$

Возьмём от обеих частей этого выражения норму и затем логарифм. Получим

$$\ln \left\| (\boldsymbol{C}_K^p)^N \right\| = -p \ln K + \ln \left\| (\boldsymbol{C}_1^p)^N \right\|.$$

Вычтем из каждой части $\ln ||(\boldsymbol{C}_1^p)^n||$ и поделим на $-\ln K$, в результате получим искомое выражение (43). Утверждение доказано.

Доказанное утверждение означает, что главный корректор является автомодельным при блочном измельчении. Левая часть (43) есть численный порядок точности, посчитанный по сетке и её измельчению в K раз при $t = n\tau(1) = N\tau(K)$. Дробь в правой части есть разностная логарифмическая производная нормы главного корректора по времени. Это позволяет определять порядок сходимости не по серии расчётов на измельчающихся сетках, а по росту нормы главного корректора в пределах одного расчёта.

Отметим, что условие автомодельности выполняется и для полудискретного уравнения, аппроксимацией методом Рунге-Кутты которого является (19)–(20). Действительно, оно имеет вид

$$\frac{d}{dt}\hat{\boldsymbol{C}}_K^p(t) + \hat{L}_K \hat{\boldsymbol{C}}_K^p(t) = \boldsymbol{f}_K^p. \tag{45}$$

Учитывая, что при блочном измельчении $\boldsymbol{f}_K^p = K^{1-p}\boldsymbol{f}_1^p$ и $L_K = KL_1$, система уравнений (45) заменой переменных $t' = Kt$ сводится к

$$\frac{d\hat{\boldsymbol{C}}_K^p}{dt'}(t') + \hat{L}_1 \hat{\boldsymbol{C}}_K^p(t') = K^{-p}\boldsymbol{f}_1^p,$$

то есть к уравнению (45) с подстановкой $K = 1$ и правой частью, домноженной на K^{-p}. Отсюда следует, что решение $\hat{\boldsymbol{C}}^p$ уравнения (45) при блочном измельчении является автомодельным:

$$\hat{\boldsymbol{C}}_K^p(t) = K^{-p}\hat{\boldsymbol{C}}_1^p(tK).$$

Взяв норму от этого выражения и логарифмическую разностную производную от неё, можно получить равенство, аналогичное (43).

Зависимость главного корректора от времени

Запишем уравнение для главного корректора (44) в виде

$$(\boldsymbol{C}^p)^{n+1} = (I - \Delta t \bar{L})(\boldsymbol{C}^p)^n + \Delta t \bar{\boldsymbol{f}}^p. \tag{46}$$

Пусть λ – собственное значение матрицы $I-(\Delta t)\bar{L}$. Предположим, что шаг по времени Δt выбирается таким образом, что схема (3) является устойчивой, причём все значения матрицы $\hat{L}$, отличные от нуля, лежат строго внутри области устойчивости используемого S-стадийного метода Рунге-Кутты. То есть выполняется одно из двух условий: $\lambda = 1$ или $|\lambda| < 1$. При этом $\lambda = 1$ только в том случае, когда соответствующее собственное значение матрицы $\hat{L}$ равно нулю.

Рассмотрим уравнение вида (46). Преобразовав матрицу $\bar{L}$ к жорданову виду, можно записать его в виде

$$\boldsymbol{g}^{n+1} = (I - (\Delta t)J)\boldsymbol{g}^n + S\bar{\boldsymbol{f}}^p \tag{47}$$

где $\boldsymbol{g}^n = S(\boldsymbol{C}^p)^n$, $\bar{L} = S^{-1}JS$.

При $\lambda = 1$ норма клетки матрицы $(I - (\Delta t)J)^n$ растёт как n^{s-1}, где s – размер жордановой клетки. Поэтому жордановы клетки размера выше единицы, соответствующие $\lambda = 1$, не допускаются условием устойчивости. Таким образом, матрица J состоит из клеток двух видов:

1) $\lambda = 1$ любой кратности, но соответствующее инвариантное подпространство имеет базис из собственных векторов;

2) $|\lambda| < 1$ без ограничения на кратность и структуру.

Собственные значения, меньшие единицы по модулю, приводят к ограниченному поведению соответствующих компонент решения (47). Единичные собственные значения приводят к линейному росту решения уравнения (47) со временем. Однако если компоненты вектора $S\bar{\boldsymbol{f}}^p$ в соответствующем инвариантном подпространстве равны 0, то решение уравнения (47) является ограниченным.

Таким образом, могут реализоваться один из двух вариантов.

1. Существует ненулевая компонента вектора $S\boldsymbol{f}^p$, соответствующая нулевому собственному значению матрицы $\hat{L}$. Тогда решение уравнения (46) при достаточно больших n растёт линейно со временем, и стационарного решения не существует. В этом случае порядок точности равен $p-1$, и эффекта сверхсходимости не наблюдается.

2. Все компоненты вектора $S\boldsymbol{f}^p$, соответствующие нулевым собственным значениям матрицы $\hat{L}$, равны 0. Тогда стационарное решение уравнения (46) существует, но не единственно. Например, если x – решение, то $x + c$, где c – собственный вектор матрицы $\hat{L}$, соответствующий нулевому собственному значению, – тоже решение. Решение уравнения (46) имеет конечный предел при $n \to \infty$. Следовательно, наблюдается эффект сверхсходимости: порядок точности равен p.

Корректор, следующий за главным

Продолжим рассмотрение блочного измельчения. Предположим, что главный корректор при $n \to \infty$ имеет конечный предел $(\boldsymbol{C}_K^p)^\infty$. Тогда из уравнения (43) следует, что $||\boldsymbol{C}_K^p|| = O(K^{-p})$.

Рассмотрим корректор порядка $p+1$. Уравнение, определяющее корректор $\boldsymbol{C}^{p+1}$, имеет вид (22)–(23),

$$\frac{(\boldsymbol{C}^{p+1})_j^{n,s+1} - (\boldsymbol{C}^{p+1})_j^n}{\omega_{s+1}\Delta t} + \sum_{k\in M} \hat{L}_{jk}(\boldsymbol{C}^{p+1})_k^{n,s} = (\tilde{\boldsymbol{f}}^{p+1})_j^{n,s}, \quad j \in M^0, \tag{48}$$

где $(\tilde{\boldsymbol{f}}^{p+1})_j^{n,s} = (\tilde{\boldsymbol{g}}^{p+1})_j + (\tilde{\boldsymbol{h}}^{p+1})_j^{n,s}$, и вектора $\tilde{g}$ и $\tilde{h}$ имеют компоненты

$$(\tilde{g}^m)_j = \Pi_j\left((\boldsymbol{a}\cdot\nabla)\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right) - \sum_{k\in M}\hat{L}_{jk}\Pi_k\left(\frac{(\boldsymbol{r}-\Pi_j(\boldsymbol{r}))^m}{m!}\right) +$$
$$+ \sum_{|i|=1,i<m}\left((C^{m-i})_j^\infty \boldsymbol{a}^i + \sum_{k\in M}\hat{L}_{jk}(C^{m-i})_k^\infty(\Pi_k(\boldsymbol{r}^i) - \Pi_j(\boldsymbol{r}^i))\right),$$

$$(\tilde{h}^m)_j^{n,s} = \sum_{|i|=1,i<m}((C^{m-i})_j^{n,s+1} - (C^{m-i})_j^\infty)\boldsymbol{a}^i +$$
$$+ \sum_{|i|=1,i<m}\sum_{k\in M}\hat{L}_{jk}((C^{m-i})_k^{n,s} - (C^{m-i})_k^\infty)(\Pi_k(\boldsymbol{r}^i) - \Pi_j(\boldsymbol{r}^i)) -$$
$$-\frac{\omega_s}{\omega_{s+1}}\sum_{|i|=1,i<m}\left((C^{m-i})_j^{n,s+1} - (C^{m-i})_j^n\right)\boldsymbol{a}^i = 0.$$

Оба вектора, $\tilde{\boldsymbol{g}}^{p+1}$ и $(\tilde{\boldsymbol{h}}^{p+1})^n$, имеют порядок $O(K^{-p})$. В силу блочности измельчения разность $||(\boldsymbol{C}^p)^{n,s} - (\boldsymbol{C}^p)^\infty||$ стремится к нулю с экспоненциальной скоростью. Поэтому $||(\tilde{\boldsymbol{h}}^{p+1})^n||$ экспоненциально стремится к нулю при $n \to \infty$. Представим $\boldsymbol{C}^{p+1} = \boldsymbol{C}_g^{p+1} + \boldsymbol{C}_h^{p+1}$, где $\boldsymbol{C}_g^{p+1}$ и $\boldsymbol{C}_h^{p+1}$ – решения уравнений вида (48) с соответствующими правыми частями. Оценку $||\boldsymbol{C}_h^{p+1}|| = O(K^{-p-1})$ получаем напрямую.

Рассмотрим $\boldsymbol{C}_g^{p+1}$. Если существует ненулевая компонента вектора $S\tilde{\boldsymbol{g}}_{p+1}$, соответствующая нулевому собственному значению матрицы $\hat{L}$, то корректор порядка $p+1$ растёт линейно со временем и $||\boldsymbol{C}_g^{p+1}|| = O(K^{-p})$. Если такой компоненты не существует, то корректор порядка $p+1$ ограничен во времени. В этом случае, пользуясь рассуждениями, аналогичными утверждению 3, получаем $||\boldsymbol{C}_g^{p+1}|| = O(K^{-p-1})$.

Таким образом, если при заданном Δt все собственные значения $\hat{L}$ лежат внутри области устойчивости метода Рунге-Кутты заданного порядка, корректоры каждого следующего порядка имеют порядок малости по K не ниже, чем максимум из предыдущих. Поэтому может реализоваться одна из следующих альтернатив.

- Главный корректор растёт линейно со временем. Тогда $||\boldsymbol{C}^p|| = O(K^{1-p})$.
- Главный корректор ограничен во времени, но корректор, следующий за главным, растёт линейно со временем. Тогда $||\boldsymbol{C}^p|| = O(K^{-p})$.
- Главный и следующий за ним корректоры ограничены во времени. Тогда $||\boldsymbol{C}^p|| = O(K^{-p})$, $||\boldsymbol{C}^{p+1}|| = O(K^{-p-1})$. Следовательно, старший член ошибки решения, растущий со временем, будет определяться корректором порядка $p+2$ или выше и иметь порядок, как минимум, $O(K^{-p-1})$. В этом случае разностная схема при длительном счёте при блочном измельчении будет иметь порядок $O(K^{-p-1})$, тогда как при коротком времени счёта – $O(K^{-p})$.

Примеры

Ввиду ограниченного объёма работы мы ограничимся тремя простыми примерами для одномерного уравнения переноса

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0.$$

1. Будем считать, что проектор П точечный. Рассмотрим схему с направленной разностью первого порядка на равномерной сетке:

$$\frac{u_j^{n+1} - u_j^n}{\tau} + \frac{u_j^n - u_{j-1}^n}{h} = 0.$$

Эта схема является точной на линейной функции, поэтому $\Gamma \equiv 0$. Так как задача одномерная, нестационарный корректор 2-го порядка является скалярной величиной и определяется уравнением (42). Поскольку сетка равномерная, размер блока $|M^0|$ равен 1, и в силу определения (39) выполняется $\hat{L} = 0$. Правая часть в этом уравнении равна $(f^2)_j = (h - \tau)/2$ для всех узлов сетки. При $\tau = h$ правая часть зануляется, и $\alpha = 0$, что согласуется с тем, что схема является точной. При $\tau < h$ вектор f^2 является собственным вектором матрицы $I - \tau\hat{L}$, соответствующим единичному собственному значению. Как следствие, корректор второго порядка одинаков во всех сеточных узлах и растёт линейно со временем. Схема с направленными разностями имеет первый порядок аппроксимации и первый порядок точности.

Для рассмотренной в этом примере задачи оператор проектирования $\tilde{\Pi}_2$ записывается в виде

$$(\tilde{\Pi}_2 f)_j^{n,0} = f(t, x_j) + t\frac{h-\tau}{2} f''(t, x_j). \tag{49}$$

Он имеет наглядную геометрическую интерпретацию. Для этого запишем (49) в виде

$$(\tilde{\Pi}_2 f)_j^{n,0} \approx \frac{1}{2\epsilon} \int\limits_{x_j-\varepsilon}^{x_j+\varepsilon} f(x)dx, \quad \varepsilon = \sqrt{3t^n(h-\tau)}.$$

Таким образом, оператор $(\tilde{\Pi}_2 f)_j^{n,0}$ есть оператор, усредняющий решение по отрезку длиной порядка $2(3t(h-\tau))^{1/2}$.

2. Будем считать, что проектор Π точечный. Рассмотрим простейшую конечно-объёмную схему на неравномерной сетке:

$$\frac{u_j^{n+1} - u_j^n}{\tau} + 2\frac{u_j^n - u_{j-1}^n}{x_{j+1} - x_{j-1}} = 0. \tag{50}$$

Эта схема не обладает свойством аппроксимации, поэтому для её анализа методом нестационарного корректора нужно выписать уравнение для корректора первого порядка. Имеем уравнение (42) с правой частью

$$f_j = 1 - 2\frac{x_j - x_{j-1}}{x_{j+1} - x_{j-1}}.$$

Пусть сетка периодична с периодом N. Матрица $\hat{L}$ схемы (50) имеет собственное значение $\lambda = 0$ кратности 1. Соответствующий правый собственный вектор имеет вид $(1, \ldots, 1)^T$, а левый – $v = (x_1 - x_{-1}, \ldots, x_N - x_{N-2})$. Вычислим vf.

$$vf = \sum_{j=0}^{N-1} \left(1 - 2\frac{x_j - x_{j-1}}{x_{j+1} - x_{j-1}}\right)(x_{j+1} - x_{j-1}) =$$

$$= \sum_{j=0}^{N-1} ((x_{j+1} - x_{j-1}) - 2(x_j - x_{j-1})) = \sum_{j=0}^{N-1}(x_{j+1} - x_j) - \sum_{j=0}^{N-1}(x_j - x_{j-1}) =$$

$$= (x_N - x_0) - (x_{N-1} - x_{-1}) = 0.$$

Последнее равенство имеет место в силу периодичности сетки. Таким образом, компонента правой части, соответствующая $\lambda = 1$, нулевая, корректор первого порядка является величиной, ограниченной со временем. Схема (50) не обладает аппроксимацией, но обладает первым порядком точности при блочном

измельчении. Рассматривая корректор 2-го порядка, можно заметить, что он растёт линейно со временем и тоже является величиной 1-го порядка малости по h. Поэтому отличия в поведении схемы на длительном времени от поведения на малом времени не наблюдается.

Пример (50) взят из [2], а в случае блочного измельчения неструктурированной сетки он был разобран в [14]. Отметим, что в одномерном случае для простейшей конечно-объёмной схемы требование блочности измельчения излишне, и схема обладает первым порядком точности при произвольном измельчении. В многомерном случае условие блочности является существенным.

3. Рассмотрим P1-метод Галёркина с разрывными базисными функциями на равномерной сетке в одномерном случае. Будем считать, что интегрирование по времени проводится не менее чем с третьим порядком точности. Пусть узлы сетки имеют координаты $x_j = jh$. Пусть на отрезке $[x_j, x_{j+1}]$ функция представляется линейной функцией $u(x) = u_{j,L}\phi_{j,L}(x) + u_{j,R}\phi_{j,R}(x)$, где базисные функции определены равенствами $\phi_{j,L}(x) = x_{j+1} - x$, $\phi_{j,R}(x) = x - x_j$. Тогда полудискретная аппроксимация, получаемая разрывным методом Галёркина, записывается в виде

$$\begin{pmatrix} h/3 & h/6 \\ h/6 & h/3 \end{pmatrix} \frac{d}{dt} \begin{pmatrix} u_{j,L} \\ u_{j,R} \end{pmatrix} + \begin{pmatrix} -u_{j-1,R} + (u_{j,L} + u_{j,R})/2 \\ u_{j,R} - (u_{j,L} + u_{j,R})/2 \end{pmatrix} = 0.$$

Обращая матрицу, получаем

$$\frac{du_{j,L}}{dt} + \frac{1}{h}(-4u_{j-1,R} + 3u_{j,L} + u_{j,R}) = 0,$$

$$\frac{du_{j,R}}{dt} + \frac{1}{h}(2u_{j-1,R} - 3u_{j,L} + u_{j,R}) = 0,$$

Схема точна на линейной функции, поэтому будем рассматривать корректоры начиная со второго порядка. Блок содержит две переменные $u_{0,L}$ и $u_{0,R}$, а матрица $\hat{L}$ и вектор f_2 имеют вид

$$\hat{L} = \begin{pmatrix} 3/h & -3/h \\ -3/h & 3/h \end{pmatrix}, \quad f_2 = \begin{pmatrix} -h/2 \\ h/2 \end{pmatrix}.$$

Здесь вектор f_2 приведён для точечного оператора проектирования: $(\Pi f)_{j,L} = f(x_j)$, $(\Pi f)_{j,R} = f(x_{j+1})$. Можно использовать интегральный оператор проектирования, сопоставляющий функции f коэффициенты разложения её L_2-проекции на пространство линейных функций в рамках одного отрезка:

$$\Pi_{j,L}(u) = \frac{6}{h^2} \int\limits_{x_j}^{x_{j+1}} u(x) \left(x_j + \frac{2}{3}h - x \right) dx,$$

$$\Pi_{j,R}(u) = \frac{6}{h^2} \int\limits_{x_j}^{x_{j+1}} u(x) \left(x - x_j - \frac{1}{3}h \right) dx.$$

Тогда также получаем $f_2 = (-h/2, h/2)^T$.

Матрица $\hat{L}$ имеет два собственных значения, 0 и 6/h. Поэтому при достаточно малом шаге по времени собственные значения матрицы $I - (\Delta t)\bar{L}$ по модулю не превосходят единицы. Левый собственный вектор матрицы $\hat{L}$ имеет вид $v = (1,1)$, следовательно, компонента вектора f_2, соответствующая $\lambda = 0$, нулевая. Поэтому корректор второго порядка ограничен во времени. Он стремится к величине α^∞, определяемой равенствами $\hat{L}\alpha^\infty = f_2$, $v\alpha^\infty = 0$, и имеющей вид

$$\alpha^\infty = \begin{pmatrix} -h^2/12 \\ h^2/12 \end{pmatrix}$$

Следовательно, рассматриваемая схема имеет второй порядок точности.

Однако продолжим исследование и рассмотрим полудискретное уравнение для нахождения геометрического корректора третьего порядка, подставляя α_k^∞ вместо $\alpha_k(t)$:

$$\frac{d\beta_j}{dt} + \sum_{k \in M^0} \hat{L}_{jk}\beta_k = -\sum_{k \in M} L_{jk} \left((\Pi_k(x) - \Pi_j(x))\alpha_k^\infty + \Pi_k \left(\frac{(x - \Pi_j(x))^3}{6} \right) \right).$$

Подставляя явный вид коэффициентов L_{jk} и предельные значения α^∞, получаем, что правая часть в этом уравнении равна $\tilde{f} = (-1/4, 1/4)^T h^2$ в случае точечного оператора проектирования и $\tilde{f} = (-9/20, 9/20)^T h^2$ в случае интегрального. Следовательно, компонента $\tilde{f}$, соотвествующая нулевому собственному значению матрицы $\hat{L}$, равна нулю. Поэтому коэффициенты нестационарного корректора 3-го порядка β_j будут ограниченными во времени и иметь порядок $O(h^3)$. Следовательно, все следующие корректоры также будут иметь порядок не ниже, чем $O(h^3)$. А значит, старший член численной ошибки, растущей со временем, будет иметь порядок $O(h^3)$. Таким образом, при малом времени счёта P1-метод Галёркина с разрывными базисными функциями на равномерной сетке в одномерном случае обладает вторым порядком точности, тогда как на большом времени счёта – третьим. Это согласуется с известными результатами [12].

Заключение

В настоящей работе был предложен метод нестационарного корректора для исследования точности разностных схем для уравнения переноса. Он является обобщением метода геометрического корректора, предложенного для простейшей конечно-объёмной схемы в [13] и далее развитого автором в [14]. Предлагаемый метод применим к широкому классу разностных схем и заключается в приближении ошибки решения конечным рядом по степеням производных от точного решения дифференциальной задачи. Наборы коэффициентов при производных именуются нестационарными корректорами.

Нестационарный корректор зависит от коэффициентов уравнения, но не зависит от конкретного вида начальных и граничных условий. Поэтому если расчётная сетка является периодической, а уравнение – однородным в пространстве, то наборы коэффициентов нестационарного корректора одинаковы в каждом сеточном блоке. При малых размерах блоков это позволяет существено упростить оценку точности как при теоретическом, так и при численном исследовании.

Главный (то есть первый отличный от нуля) нестационарный корректор, как правило, имеет наименьшую скорость убывания в асимптотике при измельчении сетки. Если схема точна на полиномах порядка k, то главный корректор, как правило, имеет величину от h^k до h^{k+1}. Он подчиняется тому же уравнению, что и численное решение, но с однородными начальными и граничными условиями и ненулевой правой частью, не зависящей от времени.

При наличии периодических условий по всем направлениям и блочном измельчении расчётной сетки главный корректор является автомодельным, то есть его поведение при измельчении сетки связано с ростом во времени на фиксированной сетке. Если величина главного корректора в некотором диапазоне мелкости сеток и некотором диапазоне времени растёт со временем как t^δ, $0 \leq \delta \leq 1$, то в этих же диапазонах численный порядок точности схемы будет равен $k + 1 - \delta$. Таким образом, при помощи нестационарного корректора описывается эффект сверхсходимости (превышений порядка точности над порядком аппркосимации).

При блочном измельчении разностные схемы внутри области устойчивости обладают целым порядком точности, то есть либо k, либо $k + 1$. Но шаг сетки, на котором ошибка выходит на асимптотическое поведение, быстро убывает с ростом размера блока. В результате, как продемонстрировано на примере простейшей конечно-объёмной схемы в [14], если размер блока достаточно большой, то свойства схемы при блочном измельчении отражают её свойства при произвольном измельчении.

Хотя главный корректор обычно имеет наименьшую скорость убывания в асимптотике, как показано, например, в [3], на реальных сетках соответству-

ющий член ряда не всегда является наибольшим по величине слагаемым. В частности, первый член ряда для многих схем на неравномерных сетках растёт со временем ограниченно и при больших временах уступает по величине слагаемые в ошибке, растущим со временем. Для оценки следующих членов ряда нужно определить геометрические корректоры более высокого порядка, чем главный. Они подчиняются уравнению с тем же разностным оператором, что и численное решение, однако правая часть зависит от корректоров меньшего порядка. Если первые n членов ряда растут со временем ограниченно, а $(n+1)$-й имеет более высокий порядок малости по h, то наблюдается повышенный порядок точности на длинном времени счёта.

Таким образом, метод нестационарного корректора позволяет выделить в составе численной ошибки слагаемые, имеющие различное поведение во времени и при измельчении сетки. С его помощью описываются эффекты сверхсходимости (supraconvergence) и повышенный порядок на длинном времени счёта (long-time simulation accuracy).

Метод нестационарного корректора напрямую обобщается на гиперболические системы уравнений. Также он может быть применён к уравнениям с переменными коэффициентами, однако при этом теряется свойство автомодельности.

Автор выражает благодарность М. Д. Сурначёву за внимательное прочтение работы и содержательные замечания к ней.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 16-31-60072 мол-а-дк.

Список литературы

1. Тихонов А. Н., Самарский А. А. Однородные разностные схемы на неравномерных сетках // Журнал вычислительной математики и математической физики. 1962. Т. 2. С. 812–832.

2. Pascal F. On supra-convergence of the finite volume method for the linear advection // ESAIM: proceedings. 2007. Vol. 18. P. 38–47.

3. Бахвалов П. А., Козубская Т. К. Структура ошибки консервативного 4-точечного конечно-разностного оператора дифференцирования на неравномерных сетках // Препринты ИПМ им. М.В.Келдыша. 2014. № 74. С. 1–32. URL: http://library.keldysh.ru/preprint.asp?id=2014-74.

4. Enhanced accuracy by post-processing for finite element methods for hyperbolic equations / Cockburn B., Luskin M., Shu C.-W. et al. // Mathematics of Computation. 2003. Vol. 72. P. 577–606.

5. Supra-convergence schemes on irregular grids / Kreiss H.-O., Manteuffel T. A., Wendroff B. et al. // Mathematics of Computation. 1986. Vol. 47. P. 537–554.

6. Johnson C., Pitkaranta J. An analysis ot the discontinious Galerkin method for a scalar hyperbolic equation // Mathematics of computation. 1986. Vol. 46. P. 1–26.

7. Peterson T. A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation // SIAM Journal on Numerical Analysis. 1991. Vol. 28. P. 133–140.

8. Cockburn B., Gremaud P.-A. A priori error estimates for numerical methods for scalar conservation laws. Part III: multidimensional flux-splitting monotone schemes on non-cartesian grids // SIAM Journal on Numerical Analysis. 1991. Vol. 28. P. 133–140.

9. Бахвалов П. А. Численная оценка порядка точности рёберно-ориентированных схем для уравнения переноса на сетках специального вида // Препринты ИПМ им. М.В.Келдыша. 2016. № 105. С. 1–32. URL: http://library.keldysh.ru/preprint.asp?id=2016-105.

10. Cheng Y., Shu C.-W. Superconvergence and time evolution of discontinuous Galerkin finite element solutions // Journal of Computational Physics. 2008. 11. Т. 227, № 22. С. 9612–9627.

11. Yang Y., Shu C.-W. Analysis of optimal superconvergence of discontinuous Galerkin method for linear hyperbolic equations // SIAM Journal on Numerical Analysis. 2012. Т. 50, № 6. С. 3110–3133.

12. Shu C.-W. Superconvergence and long time evolution of DG method. Lecture notes in CEAA2012. 2012.

13. Bouche D., Ghidaglia J.-M., Pascal F. Error Estimate and the Geometric Corrector for the Upwind Finite Volume Method Applied to the Linear Advection Equation // SIAM Journal on Numerical Analysis. 2006. Vol. 43. P. 557–603.

14. Бахвалов П. А. Нестационарный метод геометрического корректора и его использование для оценки точности конечно-объемной схемы на неструктурированных сетках // Препринты ИПМ им. М.В.Келдыша. 2016. № 122. С. 1–32. URL: http://library.keldysh.ru/preprint.asp?id=2016-122.