



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

**Бахвалов П.А.**

Нестационарный метод  
геометрического корректора  
и его использование для  
оценки точности конечно-  
объемной схемы на  
неструктурированных сетках

**Рекомендуемая форма библиографической ссылки:** 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. 28 с. doi:[10.20948/prepr-2016-122](https://doi.org/10.20948/prepr-2016-122)  
URL: <http://library.keldysh.ru/preprint.asp?id=2016-122>

**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**

**Нестационарный метод геометрического  
корректора и его использование  
для оценки точности конечно-объемной схемы  
на неструктурированных сетках**

**Москва — 2016**

## **Бахвалов П. А.**

Нестационарный метод геометрического корректора и его использование для оценки точности конечно-объемной схемы на неструктурированных сетках

В работе предлагается нестационарный метод геометрического корректора для исследования точности простейшей конечно-объемной схемы на гладких решениях. Этот метод является обобщением метода геометрического корректора (D. Bouche и др.). С помощью предложенного метода показывается, что простейшая конечно-объемная схема обладает первым порядком точности при блочном измельчении сетки.

**Ключевые слова:** метод конечных объемов, геометрический корректор, аппроксимация и точность, неструктурированная сетка

## **Pavel Alexeevich Bakhvalov**

Unsteady geometric corrector and the error estimate of finite volume method on unstructured meshes

We offer the unsteady geometric corrector for analysing the accuracy of the finite volume scheme on unstructured meshes. This method is a generalization of the geometric corrector method presented by D. Bouche et al. Using this technique we prove that finite volume scheme converges with the first order under the block refinement of the mesh.

**Key words:** finite volume method, consistency and accuracy, geometric corrector, unstructured meshes

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

## **Оглавление**

Введение . . . . .	3
Простейшая конечно-объемная схема . . . . .	4
Об истории исследования точности конечно-объемной схемы . . . . .	7
Геометрический корректор . . . . .	8
Нестационарный геометрический корректор . . . . .	10
Учёт периодических условий . . . . .	15
Блочное измельчение . . . . .	19
Пример Петерсона . . . . .	21
Численное исследование . . . . .	23
Заключение . . . . .	26
Список литературы . . . . .	27

## Введение

Теоретический анализ точности линейных разностных схем на неструктурированных сетках является достаточно сложной задачей. Если на равномерных декартовых сетках порядок точности, как правило, совпадает с порядком аппроксимации, то уже в одномерном случае на неравномерных схемах порядок точности может превосходить порядок аппроксимации на порядок старшей производной в уравнении. Эффект превосходства порядка точности над порядком аппроксимации обычно называют сверхсходимостью (supraconvergence). Впервые этот эффект был обнаружен А. А. Самарским для одномерного уравнения конвекции-диффузии [1].

Простейшая конечно-объёмная схема, которую обычно называют конечно-объёмной схемой первого порядка, обладает свойством сверхсходимости. Её аппроксимационная ошибка имеет величину  $O(1)$ , но при некоторых разумных ограничениях на сетку численное решение сходится к точному со скоростью  $O(h^{1/2})$  в интегральной норме [2] и со скоростью  $O(h^{1/2-\varepsilon})$  в максимальной норме [3]. Однако стратегия измельчения сетки может внести свои коррективы. Если сетка треугольная и её измельчение проводится разбиением каждого треугольника на 4 подобных (получаемые таким образом последовательности сеток будем называть асимптотически структурированными), то численное решение сходится к точному со скоростью  $O(h)$  [4]. Краткий обзор работ по исследованию точности простейшей конечно-объёмной схемы будет представлен ниже.

В [4] для исследования точности простейшей конечно-объёмной схемы был предложен метод геометрического корректора. Геометрический корректор представляет собой поправку к оператору проецирования решения на пространство сеточных функций, с учётом которой схема приобретает первый порядок аппроксимации. С его помощью было показано, что на асимптотически структурированных последовательностях сеток решение по простейшей конечно-объёмной схеме сходится к точному с первым порядком.

Настоящая работа представляет собой развитие метода геометрического корректора. Будет введено понятие нестационарного геометрического корректора, которое отличается от предложенного в [4] наличием зависимости коэффициентов геометрического корректора от времени.

Также в настоящей работе вводится блочный способ измельчения расчётной сетки в прямоугольной области. Он заключается в том, что расчётная сетка масштабируется в целое число раз по линейному размеру, после чего вся расчётная область регулярным образом заполняется уже готовыми фрагментами неструктурированной сетки. Конечно, блочное измельчение сетки не имеет практического применения в промышленных расчётах.

С помощью нестационарного метода геометрического корректора будет

доказано, что на при блочном измельчении, как и на асимптотически структурированных сетках, численное решение сходится к точному со скоростью  $O(h)$ . Также будет показано, что при блочном измельчении нестационарный геометрический корректор является автомодельным, то есть точность схемы можно оценивать не по скорости убывания нормы геометрического корректора при измельчении сетки, а по её росту со временем на фиксированной сетке.

Обобщение предложенной технологии оценки точности решения на разностные схемы более высокой точности будет предложено в следующей работе.

## Простейшая конечно-объёмная схема

Рассмотрим уравнение переноса

$$\frac{\partial u}{\partial t} + \mathbf{a} \cdot \nabla u = 0 \quad (1)$$

в некоторой многоугольной расчётной области  $G$ , где  $\mathbf{a}$  – постоянный во времени и пространстве вектор. На входной границе расчётной области задаётся условие  $u(t, \mathbf{r}) = u_b(t, \mathbf{r})$ ,  $\mathbf{r} \in \partial G$ , начальные условия  $u(0, \mathbf{r}) = u_0(\mathbf{r})$ .

Пусть в рассматриваемой области задана неструктурированная сетка из выпуклых элементов  $\{K_j, j = 1, \dots, N_T\}$ . Обозначим через  $|K_j|$  площадь  $j$ -го элемента. Пусть  $\mathbf{N}_{j,k}$  – вектор, равный по длине ребру, разделяющему треугольники  $K_j$  и  $K_k$ , перпендикулярный этому ребру и направленный в направлении ячейки  $K_k$ . Если  $K_j$  и  $K_k$  не соединены ребром, то  $\mathbf{N}_{j,k} = 0$ .

Пусть  $N_0(j)$  – множество элементов, соединённых с  $K_j$  ребром,  $N_b(j)$  – множество граничных рёбер расчётной области, инцидентных  $K_j$  (далее будем обозначать их  $jk$ ), и  $N(j) = N_0(j) \cup N_b(j)$ . Значения функции  $u_k$  при  $k \in N_b(j)$  будем считать равными  $u_b(\mathbf{r}_{jk})$ , где  $\mathbf{r}_{jk}$  – центр ребра  $jk$ . Обозначим через  $N^+(j) \subset N(j) = \{k : \mathbf{a} \cdot \mathbf{N}_{j,k} > 0\}$  и  $N^-(j) \subset N(j) = \{k : \mathbf{a} \cdot \mathbf{N}_{j,k} < 0\}$ ,  $N_0^+(j) = N_0(j) \cap N^+(j)$ ,  $N_0^-(j) = N_0(j) \cap N^-(j)$ . Все обозначения выбраны в соответствии с [4].

Простейшая конечно-объёмная схема (обычно называемая конечно-объёмной схемой первого порядка) для уравнения (1) имеет вид

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \frac{1}{|K_j|} \left( \sum_{k \in N^+(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} u_j^n + \sum_{k \in N^-(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} u_k^n \right) = 0. \quad (2)$$

Для краткости формулу (2) также будем записывать в общем виде

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + \sum_{k \in N(j)} L_{jk} u_k^n = 0. \quad (3)$$

Обозначим через  $\mathbf{r}_j$  центр масс треугольника  $K_j$  и через  $\mathbf{r}_{j,k}$  – центр ребра, разделяющего треугольники  $K_j$  и  $K_k$ . Докажем простое утверждение, которое будет использоваться в ходе работы.

**Утверждение 1.** Для любого элемента  $K_i$  выполняются равенства

$$\sum_{j \in N(i)} \mathbf{a} \cdot \mathbf{N}_{ij} = 0, \quad \sum_{j \in N(i)} (\mathbf{a} \cdot \mathbf{N}_{ij}) \mathbf{r}_{ij} = \mathbf{a} |K_i|.$$

Используя формулу Остроградского-Гаусса, для сеточного элемента  $K_i$  получим

$$\begin{aligned} \sum_{j \in N(i)} \mathbf{a} \cdot \mathbf{N}_{ij} &= \mathbf{a} \cdot \oint_{\partial K_i} \mathbf{n} \, dl = 0, \\ \sum_{j \in N(i)} (\mathbf{a} \cdot \mathbf{N}_{ij}) \mathbf{r}_{ij} &= \oint_{\partial K_i} (\mathbf{a} \cdot \mathbf{n}) \mathbf{r} \, dl = \iiint_{K_i} (\mathbf{a} \cdot \nabla) \mathbf{r} \, dv = \mathbf{a} \iiint_{K_i} dv = \mathbf{a} |K_i|, \end{aligned}$$

что и требовалось доказать.

Чтобы сформулировать определение аппроксимации, нужно ввести оператор, сопоставляющий функции её сеточный аналог. Обычно для конечно-объёмных схем полагается, что  $(\Pi(f))_j^n$  есть интегральное среднее функции  $f$  по треугольнику  $j$  в момент времени  $t^n$ . Также можно положить, что  $(\Pi(f))_j^n$  есть точечное значение функции  $f$  в центре масс элемента  $K_j$ . Будем называть такие операторы соответственно интегральным и точечным.

Ошибка аппроксимации схемы (3) в смысле оператора  $\Pi$  определяется равенством

$$\epsilon_j^n = \frac{(\Pi u)_j^{n+1} - (\Pi u)_j^n}{\Delta t} + \sum_{k \in N(j)} L_{jk} (\Pi u)_k^n, \quad (4)$$

где  $u$  – точное решение дифференциальной задачи. Очевидно, что разные операторы  $\Pi$  приводят к разным формам аппроксимационной ошибки.

**Определение 1.** Будем говорить, что схема (3) аппроксимирует уравнение (1) с порядком  $O(h^\delta)$  на некоторой последовательности сеток в смысле оператора  $\Pi$  при некотором ограничении на  $\Delta t$ , если выполняется условие  $\epsilon_j^n = O(h^\delta)$ , где  $\epsilon_j^n$  определено формулой (4).

Примем за характерный шаг сетки  $h$  максимальную длину сеточного ребра. Предположим, что сетки удовлетворяют двум естественным условиям: 1) объёмы ячеек подчиняются соотношению  $|K_j| > k_1 h^2$  и 2) число рёбер у каждого

элемента ограничено:  $|N(j)| \leq M$ . Первое условие, в частности, не допускает стремления к нулю минимального угла треугольника.

Аппроксимационная ошибка схемы (2) в смысле как точечного, так и интегрального оператора  $\Pi$  имеет вид

$$\begin{aligned} \epsilon_j^n = & \frac{1}{|K_j|} \left( \sum_{k \in N^+(j)} (\mathbf{a} \cdot \mathbf{N}_{j,k}) (\mathbf{r}_j - \mathbf{r}_{j,k}) + \right. \\ & \left. + \sum_{k \in N^-(j)} (\mathbf{a} \cdot \mathbf{N}_{j,k}) (\mathbf{r}_k - \mathbf{r}_{j,k}) \right) \cdot \nabla u(t^n, \mathbf{r}_j) + O(h + \Delta t). \end{aligned} \quad (5)$$

Действительно, применим определение (4) к схеме (2). Обозначим через  $u_{jk}$  значение функции в центре ребра  $jk$ . Получим

$$\begin{aligned} \epsilon_j^n = & \frac{(\Pi u)_j^{n+1} - (\Pi u)_j^n}{\Delta t} + \frac{1}{|K_j|} \left( \sum_{k \in N^+(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} (\Pi u)_j^n + \sum_{k \in N^-(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} (\Pi u)_k^n \right) = \\ = & \frac{(\Pi u)_j^{n+1} - (\Pi u)_j^n}{\Delta t} + \frac{1}{|K_j|} \sum_{k \in N(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} u_{jk}^n + \\ & + \frac{1}{|K_j|} \left( \sum_{k \in N^+(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} ((\Pi u)_j^n - u_{jk}^n) + \sum_{k \in N^-(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} ((\Pi u)_k^n - u_{jk}^n) \right). \end{aligned}$$

Первое слагаемое с ошибкой  $O(\Delta t)$  аппроксимирует интегральное среднее по ячейке  $K_j$  от  $\partial u / \partial t(\mathbf{r}_j)$ , а второе – с ошибкой  $O(h^3)/|K_j| = O(h)$  аппроксимирует интеграл по границе ячейки  $K_j$  от  $(\mathbf{a} \cdot \mathbf{n})u$ . Записывая линейное приближение для  $u$ , с учётом уравнения (1) получаем формулу (5).

Очевидно, что  $\epsilon_j^n$  является величиной порядка  $O(1)$ , то есть простейшая конечно-объёмная схема (2) не обладает свойством аппроксимации.

Схема (2) устойчива при условии на шаг по времени. Это условие имеет вид  $\Delta t < \tau^*$ , где

$$\tau^* = \min_j \frac{1}{L_{jj}} = \min_j \frac{|K_j|}{\sum_{k \in N^+(j)} \mathbf{a} \cdot \mathbf{N}_{jk}}. \quad (6)$$

При этом условии схема является монотонной, то есть все коэффициенты матрицы  $I - (\Delta t)L$  неотрицательны.

## Об истории исследования точности конечно-объёмной схемы

Конечно-объёмный метод на неструктурированных сетках стал применяться для решения гиперболических уравнений с 90-х годов, когда на них была перенесена схема MUSCL, предложенная Колганом и ван Лиром для одномерных задач. Тогда же возник вопрос о точности конечно-объёмных схем. Исследование точности метода, конечно, было начато с простейшей конечно-объёмной схемы, применённой к уравнению переноса. Мы не будем в этом обзоре касаться вопроса о точности на негладких решениях, предполагая, что начальные и граничные условия достаточно гладкие и согласованы между собой. Также мы будем предполагать, что сетки, получающиеся при измельчении, удовлетворяют двум естественным условиям: 1) объёмы ячеек подчиняются соотношению  $k_1 h^2 < |K_i| < k_2 h^2$  и 2) число рёбер у каждого элемента ограничено:  $|N(j)| \leq M$ .

Задолго до применения конечно-объёмных схем на треугольных сетках к ним уже применялся метод Галёркина с разрывными базисными функциями. Он был предложен для решения задачи о переносе нейтронов в 1973 году [5]. Хотя эта задача стационарная, как будет видно ниже, это не играет роли. В [6] была доказана естественная оценка, что этот метод обладает точностью  $O(h^k)$  в норме  $L_2$ , где  $k$  – порядок использованного полинома. При  $k = 0$  метод Галёркина с разрывными базисными функциями сводится к простейшей конечно-объёмной схеме. Однако первые же численные эксперименты показывали, что оценка  $O(h^k)$  не является оптимальной: реальная сходимость оказывалась более быстрой.

Исследования в области улучшения оценок точности простейшей конечно-объёмной схемы начались в 1986 году работой С. Johnson [7], в которой была доказана сходимость в норме  $L_2$  со скоростью  $O(h^{1/2})$ . Также в работе [7] была показана справедливость оценки  $O(h^{1/2})$  в норме  $L_\infty$  на асимптотически структурированных треугольных сетках.

В 1988 году G. Richter в [8] доказал сходимость с оптимальной скоростью  $O(h)$  на сетках специального вида, состоящих из параллельных слоёв толщиной в один треугольник, причём при условии, что углы между направлением рёбер треугольника и направлением переноса отделены от нуля. Это довольно экзотическое условие, по всей видимости, играет существенную роль: нечто похожее будет наблюдаться в последующих работах и для сеток общего вида.

В 1990 году Т. Peterson в [9] доказал оценку численной ошибки  $O(h^{1/4})$  в норме  $L_\infty$ , а в 1991 [10] им же был приведён пример последовательности сеток, на котором сходимость в  $L_2$  и  $L_\infty$  нормах происходит в точности со скоростью  $Ch^{1/2}$ . Таким образом, было доказано, что оценка в норме  $L_2$ , доказанная в [7], неулучшаема. Вопрос об оптимальной оценке в норме  $L_\infty$  пока оставался открытым.



В 1996 году В. Cockburn [11] начал исследовать точность конечно-объемной схемы на нестационарном уравнении переноса и доказал сходимость со скоростью  $O(h^{1/2})$  в норме  $L_1$ . В 2003 году В. Despres [12] доказал порядок  $O(h^{1/4})$  в норме  $L_2$  и  $O(h^{1/4-\varepsilon})$  в норме  $L_\infty$ . В 2004 году он же [2] при помощи другой техники уточнил оценку в норме  $L_2$  до неулучшаемой  $O(h^{1/2})$ .

В 2005 году F. Pascal и др. [4] предложили метод геометрического корректора для исследования точности численных методов, при помощи которого доказали оценку  $O(h)$  в норме  $L_\infty$  на асимптотически структурированных треугольных сетках. Оценка  $O(h)$  также была получена для неравномерной сетки из параллелограммов, а также для треугольной сетки, полученной из неё разрезанием каждого параллелограмма одной диагональю с сохранением структуры. В этой же работе в численном эксперименте было показано, что в норме  $L_\infty$  численный порядок точности получается около  $1/2$ .

Численное исследование было продолжено в работе [13]. Сетка полностью перестраивалась генератором Gmsh [14] в квадрате с периодическими условиями по обоим направлениям. При этом было замечено, что падение точности до  $O(h^{1/2})$  в  $L_\infty$  наблюдается только в том случае, если направление скорости переноса совпадает с направлением одного из рёбер канала. В  $L_1$  численный порядок точности оставался близок к 1. В 2011 году те же авторы рассмотрели контрпример Т. Peterson из [10] и методом геометрического корректора показали [15], что при изменении направления переноса порядок точности также становится равным 1. В дальнейших работах эти же авторы обобщили свои оценки на случай уравнения переноса с переменными коэффициентами и скалярного нелинейного уравнения [16], а также системы уравнений.

Оптимальная оценка сходимости конечно-объемной схемы на неструктурированных сетках в норме  $L_\infty$ ,  $O(h^{1/2-\varepsilon})$ , была доказана только в 2007 году В. Merlet [3]. Позже этот же результат был подтверждён F. Delarue и F. Lagoutiere при помощи стохастических методов [17].

Мы привели только работы, посвящённые анализу простейшей конечно-объемной схемы для гладких решений. Другие работы доказывали оценки для более широкого класса начальных и граничных условий; оценки также обобщались на случай переменных коэффициентов и нелинейных уравнений. Отдельным вопросом был анализ точности схем типа MUSCL. Вся совокупность этих работ могла бы быть предметом большого обзора.

## Геометрический корректор

Понятие геометрического корректора было введено в работе [4] для простейшей конечно-объемной схемы (2) применительно к скалярному уравнению переноса (1).

Согласно определению, данному в [4], геометрический корректор есть на-

бор векторов  $\Gamma_j$ , удовлетворяющих системе уравнений

$$\begin{aligned} & \sum_{k \in N^+(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} \Gamma_j + \sum_{k \in N_0^-(j)} \mathbf{a} \cdot \mathbf{N}_{j,k} \Gamma_k = \\ & = \sum_{k \in N^+(j)} (\mathbf{a} \cdot \mathbf{N}_{j,k}) (\mathbf{r}_{j,k} - \mathbf{r}_j) + \sum_{k \in N_0^-(j)} (\mathbf{a} \cdot \mathbf{N}_{j,k}) (\mathbf{r}_{j,k} - \mathbf{r}_k). \end{aligned} \quad (7)$$

Смысл этого определения можно пояснить следующим образом. Рассмотрим пространственную дискретизацию общего вида (3). Как было сказано выше, разные операторы  $\tilde{\Pi}$  приводят к разным выражениям для аппроксимационной ошибки.

Введём оператор  $\tilde{\Pi}$  следующим образом:

$$\left( \tilde{\Pi}(f) \right)_j^n = f(t^n, \mathbf{r}_j) + \Gamma_j \cdot \nabla f(t^n, \mathbf{r}_j), \quad (8)$$

где  $\Gamma_j$  – геометрический корректор – определяется из условия, чтобы в смысле оператора  $\tilde{\Pi}$  схема обладала точностью на линейной функции, а, следовательно, первым порядком аппроксимации. Другими словами, при подстановке линейной функции  $f(t, \mathbf{r}) = \mathbf{r} - \mathbf{a}t$  аппроксимационная ошибка в смысле оператора  $\tilde{\Pi}$  должна быть равна 0:

$$\frac{\left( \tilde{\Pi}(\mathbf{r} - \mathbf{a}t) \right)_j^{n+1} - \left( \tilde{\Pi}(\mathbf{r} - \mathbf{a}t) \right)_j^n}{t^{n+1} - t^n} + \sum_{k \in N(j)} L_{jk} \left( \tilde{\Pi}(\mathbf{r} - \mathbf{a}t) \right)_k^n = 0. \quad (9)$$

Суммирование в этой формуле ведётся по  $k \in N(j)$ . При этом при  $k \in N^+(j)$  коэффициенты  $L_{jk} = 0$ . В схему входят рёбра на входной границе  $k \in N_b^-(j)$ . Следовательно, линейная функция  $f(t, \mathbf{r})$  должна быть продолжена и на граничные рёбра. Граничные условия для  $\Gamma$  определим как

$$\Gamma_j^n = 0, \quad j \in \bigcup_i N_b(i).$$

Подставляя в (9) вид оператора  $\tilde{\Pi}$  (8), получаем

$$\sum_{k \in N(j)} L_{jk} (\mathbf{r}_k + \Gamma_k) = \mathbf{a} \quad \forall j = 1, \dots, N_T.$$

Таким образом, применение оператора  $L$  к геометрическому корректору должно совпадать с ошибкой аппроксимации на линейной функции. Учитывая точность простейшей конечно-объёмной схемы на константе, можно записать

$$\sum_{k \in N_0(j)} L_{jk} \Gamma_k = \mathbf{a} - \sum_{k \in N(j)} L_{jk} (\mathbf{r}_k - \mathbf{r}_j). \quad (10)$$

Это условие совпадает с определением (7) за исключением того, что суммирование в (7) в последней сумме проводится по  $k \in N_0^-(j)$ , а формула (10) приводит к суммированию в ней по  $k \in N^-(j)$ . Это отличие сказывается на дальнейших рассуждениях.

В [4] была доказана следующая теорема.

**Теорема.** Пусть  $\Gamma$  – решение уравнения (7), и пусть  $\|\Gamma\| = O(h^\delta)$ , где  $0 < \delta \leq 1$ . Тогда разность численного решения и точного сходится к нулю со скоростью  $O(h^\delta)$ .

Утверждение этой теоремы остаётся справедливым, и если  $\Gamma$  – решение уравнения (10). Таким образом, порядок малости геометрического корректора определяет порядок малости численной ошибки.

Отметим, что метод геометрического корректора может быть также применён к исследованию полудискретной аппроксимации  $du/dt + Lu = 0$  и к стационарному уравнению переноса, аппроксимируемому схемой  $\lambda u + Lu = 0$ .

## Нестационарный геометрический корректор

Метод геометрического корректора при оценке точности решения задачи (2) обладает двумя недостатками. Во-первых, возникают сложности при исследовании задачи с периодическими граничными условиями, поскольку при этом матрица  $L$  является вырожденной. Во-вторых, оценка ошибки, получаемая при помощи метода геометрического корректора, оказывается ненулевой уже при  $t = 0$ . Это связано с тем, что начальные условия, как правило, проецируются на сетку при помощи оператора  $\Pi$ , а не  $\tilde{\Pi}$ .

Мы подойдём к методу геометрического корректора с другой стороны. Геометрический корректор есть коэффициент при первой производной в ошибке численного решения, см. (8). Возьмём это наблюдение за исходный пункт нашего исследования.

Рассмотрим оператор

$$\left(\tilde{\Pi}(f)\right)_j^n = f(t^n, \mathbf{r}_j) + \Gamma_j^n \cdot \nabla f(t^n, \mathbf{r}_j), \quad (11)$$

отличающийся от (8) тем, что  $\Gamma$  теперь зависит от времени. Как и в случае стационарного геометрического корректора, будем определять  $\Gamma$  из условия точности схемы (3) на линейной функции в смысле оператора  $\tilde{\Pi}$ .

Подставим функцию  $u(\mathbf{r}, t) = \mathbf{r} - \mathbf{a}t$  в оператор (11) и затем в разностную схему (3):

$$\frac{1}{\Delta t} (\Gamma_j^{n+1} - \Gamma_j^n) - \mathbf{a} + \sum_{k \in N(j)} L_{jk} ((\mathbf{r}_k - \mathbf{a}t) + \Gamma_k^n) = 0.$$

Пользуясь точностью схемы на константе, мы можем записать окончательное уравнение для  $\Gamma$ :

$$\frac{\Gamma_j^{n+1} - \Gamma_j^n}{\Delta t} + \sum_{k \in N_0(j)} L_{jk} \Gamma_k^n = \mathbf{a} - \sum_{k \in N(j)} L_{jk} (\mathbf{r}_k - \mathbf{r}_j). \quad (12)$$

Уравнение (12) дополняется начальным условием  $\Gamma_j^0 = 0, j = 1, \dots, N_T$ . Из (12) следует, что  $\Gamma_j^n$  полагается равным 0 при  $k \in N_b(j)$ . Отметим, что если вместо этого предположить, что  $\Gamma^n$  не зависит от  $n$ , то уравнение (12) переходит в определение геометрического корректора (10).

**Определение 2.** Будем называть решение уравнения (12), где коэффициенты  $L_{jk}$  соответствуют схеме (2), нестационарным геометрическим корректором для простейшей конечно-объёмной схемы (2).

Таким образом, величина нестационарного геометрического корректора  $\Gamma$  подчиняется уравнению с тем же оператором, что и сеточная функция, но с ненулевой правой частью. Особо отметим, что эта правая часть не зависит ни от времени, ни от решения. В краткой форме будем записывать уравнение (12) в виде

$$\frac{\Gamma^{n+1} - \Gamma^n}{\Delta t} + L\Gamma^n = \mathbf{f}, \quad (13)$$

где при записи матрично-векторного произведения  $L\Gamma$  будем предполагать  $N_T$ -мерные вектора и матрицу  $L$  размера  $N_T \times N_T$ .

**Теорема 2.** Пусть решение уравнения (1) дифференцируемо и имеет липшицевы производные с константой Липшица  $\mathbb{L}$ . Пусть схема (3) точна на константе и её оператор  $L_h = \{L_{jk}\}$  удовлетворяет условию  $|L_{jk}| \leq C_1/h$  и имеет финитный носитель:  $L_{jk} = 0$  при  $|\mathbf{r}_k - \mathbf{r}_j| > C_2 h$ . Пусть шаг по времени  $\Delta t = \tau(h)$  выбирается таким образом, что схема (3) с оператором  $L_h$  является устойчивой и  $\tau(h) \leq C_3 h$ . Пусть решение  $\Gamma$  уравнения (12) имеет порядок  $O(h^\delta)$ ,  $0 < \delta \leq 1$ , на некоторой последовательности сеток. Тогда отличие численного решения по схеме (3) от точного удовлетворяет оценке

$$|u_j^n - u(t^n, \mathbf{r}_j) - \Gamma_j^n \cdot \nabla u(t^n, \mathbf{r}_j)| \leq Ch^\delta \mathbb{L}, \quad (14)$$

где  $C$  не зависит от  $h, \tau$  и  $\mathbb{L}$ . Следовательно, схема (3) имеет порядок точности  $O(h^\delta)$ .

Доказательство. Представим численное решение в виде

$$u_j^n = \left( \tilde{\Pi}(u) \right)_j^n + \epsilon_j^n = u(t^n, \mathbf{r}_j) + \Gamma_j^n \cdot \nabla u(t^n, \mathbf{r}_j) + \epsilon_j^n. \quad (15)$$

Далее индекс  $n$  будем для простоты иногда опускать. Подставляя выражение (15) в схему (2), получим уравнение для  $\epsilon$ :

$$\begin{aligned} & \frac{u(t + \Delta t, \mathbf{r}_j) - u(t, \mathbf{r}_j)}{\Delta t} + \frac{\mathbf{\Gamma}_j^{n+1} \cdot \nabla u(t + \Delta t, \mathbf{r}_j) - \mathbf{\Gamma}_j^n \cdot \nabla u(t, \mathbf{r}_j)}{\Delta t} + \\ & + \frac{\epsilon_j^{n+1} - \epsilon_j^n}{\Delta t} + \sum_{k \in N(j)} L_{jk} (u(t^n, \mathbf{r}_k) + \mathbf{\Gamma}_k^n \cdot \nabla u(t^n, \mathbf{r}_k) + \epsilon_k^n) = 0. \end{aligned}$$

Отсюда

$$\begin{aligned} & \frac{\epsilon_j^{n+1} - \epsilon_j^n}{\Delta t} + \sum_k L_{jk} \epsilon_k^n = - \sum_k L_{jk} (u(t, \mathbf{r}_k) + \mathbf{\Gamma}_k^n \cdot \nabla u(t, \mathbf{r}_k)) - \\ & - \frac{u(t + \Delta t, \mathbf{r}_j) - u(t, \mathbf{r}_j)}{\Delta t} - \frac{\mathbf{\Gamma}_j^{n+1} \cdot \nabla u(t + \Delta t, \mathbf{r}_j) - \mathbf{\Gamma}_j^n \cdot \nabla u(t, \mathbf{r}_j)}{\Delta t}. \end{aligned} \quad (16)$$

Учтём, что  $|\nabla u(t, \mathbf{r}_j) - \nabla u(t, \mathbf{r}_i)| \leq C_2 \mathbb{L}h$ ,  $|L_{jk}| \leq C_1/h$  и  $\|\mathbf{\Gamma}\| = O(h^\delta)$ , преобразуем первое слагаемое в правой части равенства:

$$\begin{aligned} & \sum_k L_{jk} (u(t, \mathbf{r}_k) + \mathbf{\Gamma}_k^n \cdot \nabla u(t, \mathbf{r}_k)) = \sum_k L_{jk} u(t, \mathbf{r}_k) + \sum_k L_{jk} \mathbf{\Gamma}_k^n \cdot \nabla u(t, \mathbf{r}_k) + O(\mathbb{L}h^\delta) = \\ & = \sum_k L_{jk} u(t, \mathbf{r}_k) - \left( \frac{\mathbf{\Gamma}_j^{n+1} - \mathbf{\Gamma}_j^n}{\Delta t} - \mathbf{a} + \sum_k L_{jk} (\mathbf{r}_k - \mathbf{r}_j) \right) \cdot \nabla u(t, \mathbf{r}_j) + O(\mathbb{L}h^\delta) = \\ & = \sum_k L_{jk} (u(t, \mathbf{r}_k) - (\mathbf{r}_k - \mathbf{r}_j) \cdot \nabla u(t, \mathbf{r}_j)) - \left( \frac{\mathbf{\Gamma}_j^{n+1} - \mathbf{\Gamma}_j^n}{\Delta t} - \mathbf{a} \right) \cdot \nabla u(t, \mathbf{r}_j) + O(\mathbb{L}h^\delta) = \\ & = \sum_k L_{jk} (u(t, \mathbf{r}_j) + O(\mathbb{L}h^2)) - \left( \frac{\mathbf{\Gamma}_j^{n+1} - \mathbf{\Gamma}_j^n}{\Delta t} - \mathbf{a} \right) \cdot \nabla u(t, \mathbf{r}_j) + O(\mathbb{L}h^\delta) = \\ & = - \left( \frac{\mathbf{\Gamma}_j^{n+1} - \mathbf{\Gamma}_j^n}{\Delta t} - \mathbf{a} \right) \cdot \nabla u(t, \mathbf{r}_j) + O(\mathbb{L}h^\delta) + O(\mathbb{L}h). \end{aligned}$$

При написании последнего равенства использовалась точность схемы на константе. Подставляя полученный результат в (16),

$$\begin{aligned} & \frac{\epsilon_j^{n+1} - \epsilon_j^n}{\Delta t} + \sum_k L_{jk} \epsilon_k^n = \\ & = - \left( \frac{u(t + \Delta t, \mathbf{r}_j) - u(t, \mathbf{r}_j)}{\Delta t} + \mathbf{a} \cdot \nabla u(t, \mathbf{r}_j) \right) - \end{aligned}$$

$$-\Gamma_j^{n+1} \cdot \frac{(\nabla u(t + \Delta t, \mathbf{r}_j) - \nabla u(t, \mathbf{r}_j))}{\Delta t} + O(\mathbb{L}h^\delta) + O(\mathbb{L}h) = O(\mathbb{L}h^\delta).$$

Таким образом, схема (16) аппроксимирует уравнение  $\partial\epsilon/\partial t + \partial\epsilon/\partial x = 0$ ,  $\epsilon(0, x) = 0$ , с ошибкой  $O(\mathbb{L}h^\delta)$ . Точным решением этого уравнения является  $\epsilon(t, x) = 0$ . Устойчивость имеется по предположению данного утверждения. Следовательно, по теореме Лакса решение уравнения (16) удовлетворяет оценке  $\epsilon = O(h^\delta\mathbb{L})$ .

Утверждение доказано.

**Замечание 1.** Теорема 2 показывает, что геометрический корректор представляет собой коэффициент при первой производной в ошибке решения. Ошибка решения за вычетом геометрического корректора, умноженного на первую производную, пропорциональна второй производной.

**Замечание 2.** Оценка (14), предположительно, является грубой, по меньшей мере, для простейшей конечно-объёмной схемы. Это предположение основано на анализе размерностей. В силу устойчивости коэффициенты  $\Gamma^n$  должны расти со временем не быстрее чем линейно. Следовательно, не быстрее чем линейно должно расти и выражение  $\|u^n - u(t^n, \mathbf{r}) - \Gamma^n \cdot \nabla u(t^n, \mathbf{r})\|$ . При этом она пропорциональна второй производной от решения. Следовательно, оценка этой величины должна иметь вид

$$\|u^n - u(t^n, \mathbf{r}) - \Gamma^n \cdot \nabla u(t^n, \mathbf{r})\| \leq th\mathbb{L},$$

то есть при  $\delta < 1$  слагаемое  $\Gamma^n \cdot \nabla u(t^n, \mathbf{r})$  должно быть не только слагаемым при старшей производной в ошибке решения, но и асимптотически наибольшим по величине слагаемым. При  $\delta = 1$  оно может иметь такой же порядок, как и  $\epsilon^n$ . Однако это требует более аккуратной оценки, чем этого можно добиться при помощи теоремы Лакса.

**Утверждение 3.** Пусть при  $\tau \leq \tau^*$  все собственные значения оператора перехода  $(I - \tau L)$  по модулю не превосходят единицы. Тогда при  $\tau < \tau^*$  все собственные значения оператора  $I - \tau L$ , кроме единицы, лежат строго внутри единичного круга.

**Доказательство.** Собственные вектора матрицы  $I - \tau^* L$  и  $I - \tau L$ , очевидно, совпадают. При  $\tau < \tau^*$  запишем

$$I - \tau L = (I - \tau^* L) \frac{\tau}{\tau^*} + \left(1 - \frac{\tau}{\tau^*}\right) I.$$

Любое собственное значение  $I - \tau L$  есть выпуклая комбинация с ненулевыми коэффициентами собственного значения  $I - \tau^* L$  и единицы, следовательно,

оно или лежит строго внутри круга  $\{|\lambda| < 1\}$ , или равно единице. Утверждение доказано.

**Утверждение 4.** *Решение уравнения (12) для конечно-объёмной схемы (2) при условии  $\Delta t < \tau^*$  имеет конечный предел.*

Доказательство. Перепишем равенство (13) в виде

$$\Gamma^{n+1} = (I - (\Delta t)L)\Gamma^n + \Delta t\mathbf{f}. \quad (17)$$

Как было показано выше, при  $\Delta t \leq \tau^*$  все элементы матрицы  $I - \Delta tL$  неотрицательны. Её строковые суммы равны единице для всех элементов, не имеющих пересечения со входной границей, и не превосходят единицы для элементов, пересекающихся со входной границей. Отсюда следует, что  $\|I - \Delta tL\|_\infty \leq 1$  при  $\Delta t \leq \tau^*$ . Таким образом, собственные значения  $I - \tau^*L$  лежат внутри единичного круга  $\{|\lambda| \leq 1\}$ . По утверждению 3 любое собственное значение  $I - \Delta tL$  при  $\Delta t < \tau^*$  либо лежит строго внутри круга  $\{|\lambda| < 1\}$ , либо равно единице. Покажем, что последний случай невозможен.

В самом деле, это означает, что найдётся нетривиальный вектор  $e$ , такой, что  $Le = 0$ . Этот вектор есть стационарное решение простейшей конечно-объёмной схемы с нулевыми входными граничными условиями. Пусть  $U^{n+1} = (I - \Delta tL)U^n$ . Легко видеть, что для любого элемента  $K_j$ , имеющего пересечение со входной границей,  $|U_j^{n+1}| < \max_k |U_k^n|$ . Таким образом, стационарное решение  $U$  должно быть равно нулю во всех элементах, расположенных на входной границе. Повторяя это рассуждение для их соседей и так далее, получаем, что стационарное решение может быть только нулевым. Таким образом, мы показали, что единица не является собственным значением  $I - \Delta tL$  и все собственные значения лежат строго внутри единичного круга  $\{|\lambda| < 1\}$ . Обратим внимание, что  $\|I - \Delta tL\|$  при этом будет равна 1, за исключением вырожденного случая, при котором каждая ячейка имеет входное граничное ребро.

Используя жорданову нормальную форму матрицы  $L$ , из условия  $\{|\lambda| < 1\}$ , где  $\lambda$  – собственное значение матрицы  $I - \Delta tL$ , легко получить, что решение уравнения (17) имеет конечный предел

$$\Gamma^\infty = \left( \sum_{m=0}^{\infty} (I - \Delta tL)^m \right) \Delta t\mathbf{f} = L^{-1}\mathbf{f}.$$

Этот предел является решением уравнения (10), который определяет стационарный геометрический корректор. Утверждение доказано.

**Утверждение 5.** Пусть решение уравнения (12) имеет конечный предел, и в некоторой норме выполняется  $\|I - (\Delta t)L\| \leq 1$ . Тогда для любого  $n$  выполняется  $\|\Gamma^n\| < 2\|\Gamma^\infty\|$ .

Преобразуем уравнение (12) к виду

$$\Gamma^{n+1} - \Gamma^\infty = (I - (\Delta t)L)(\Gamma^n - \Gamma^\infty),$$

$$\Gamma^0 = 0.$$

При  $\Delta t < \tau^*$  выполняется  $\|\Gamma^{n+1} - \Gamma^\infty\| \leq \|\Gamma^n - \Gamma^\infty\| \leq \dots < \|\Gamma^0 - \Gamma^\infty\|$ . Отсюда  $\|\Gamma^n\| < 2\|\Gamma^\infty\|$ . Утверждение доказано.

Таким образом, для простейшей конечно-объёмной схемы с условиями Дирихле на входной границе нестационарный геометрический корректор является ограниченной величиной и имеет тот же порядок, что и стационарный. При  $t \rightarrow \infty$  нестационарный геометрический корректор стремится к стационарному.

## Учёт периодических условий

Вычисление нестационарного геометрического корректора (равно как и стационарного для стационарной задачи) требует решения системы уравнений с тем же разностным оператором, что и сама разностная схема. Поэтому в общем случае его использование не даёт никакого преимущества относительно сравнения численного решения с точным на последовательности сеток. Однако при наличии периода в сетке правая часть системы уравнений также является периодической, благодаря чему достаточно ограничиться решением задачи в подобласти размером с один пространственный период. Это позволяет существенно упростить оценку точности.

При наличии периодических граничных условий по одному или нескольким направлениям матрица системы уравнений  $L$  в уравнении (10) становится вырожденной. Действительно, почти любая численная схема, в том числе (2), производную от константы считает нулём. Поэтому обычный метод геометрического корректора не подходит для исследования задач с периодическими условиями. Однако нестационарный подход (12) остаётся применимым.

Периодические граничные условия для уравнения переноса можно интерпретировать как периодичность решения, заданного в бесконечной области (которая может как совпадать со всем пространством, так и не совпадать). Обозначим за  $N^0$  множество ячеек одного блока, а за  $N^\infty$  – множество ячеек во всей расчётной сетке в бесконечной области.

Из уравнения (12) видно, что вектор нестационарного геометрического корректора одинаков во всех ячейках, совмещающихся при совмещении блоков.



Поэтому можно записать уравнение, аналогичное (12), но учитывающее наличие периодических условий. Тогда

$$\frac{\Gamma_i^{n+1} - \Gamma_i^n}{\Delta t} + \sum_{j \in N^0} \hat{L}_{ij} \Gamma_j^n = \mathbf{f}, \quad \mathbf{f} = \mathbf{a} - \sum_{j \in N^\infty} L_{ij} (\mathbf{r}_j - \mathbf{r}_i), \quad (18)$$

где  $i \in N^0$ .  $\hat{L}_{ij}$  определяется как

$$\hat{L}_{ij} = \sum_k L_{ik}, \quad (19)$$

где сумма берётся по всем ячейкам, совмещаемым с  $j$  при наложении блоков пространственной трансляцией. Важно, что сумма в правой части не может быть сведена к сумме по  $N^0$ , так как вектор  $\mathbf{r}$ , в отличие от векторов геометрического корректора, является линейной функцией во всём пространстве и, как следствие, различается в разных блоках.

Матрица  $L$  имеет нулевую сумму в каждой строке, так как аппроксимация производной точна на константе. Для простейшей конечно-объёмной схемы диагональные элементы матрицы  $L$  положительны, а внедиагональные – отрицательны. Это же свойство сохраняется и для матрицы  $\hat{L}$ , если только сечение сеточного блока характеристикой не является отрезком, в случае чего матрица  $\hat{L}$  или некоторые её строки становятся нулевыми. Поэтому при  $\Delta t < \tau^*$  все коэффициенты матрицы  $I - (\Delta t)\hat{L}$  неотрицательны; строковые же суммы её равны 1.

**Утверждение 6.** *Рассмотрим простейшую конечно-объёмную схему (2) в связной области  $G$  с периодическими граничными условиями на всех границах (исключая, быть может, границы, параллельные направлению характеристики). Пусть не существует характеристической линии  $\mathbf{r} = \mathbf{r}_0 + \mathbf{a}\xi$ , проходящей в точности по рёбрам сетки в области  $G$ . Тогда нулевое собственное значение имеет кратность 1 и соответствует собственному вектору, все элементы которого равны между собой.*

Доказательство. Вектор, все элементы которого равны между собой, очевидно, под действием простейшей конечно-объёмной схемы обращается в ноль. Нужно доказать то, что для любого другого вектора  $u \in \mathbb{R}^{N^0}$  выполняется условие  $\hat{L}u \neq 0$ .

Рассмотрим вектор  $u$ , такой что  $\hat{L}u = 0$ . Пусть  $M \subset N^0$  – набор ячеек, в которых вектор  $u$  достигает своего максимального значения  $u^{\max}$ . Рассмотрим некоторую ячейку  $j \in M$ . Предположим, что  $\exists k \in N^-(j) : u_k < u^{\max}$ . Однако

сумма элементов в строке матрицы  $\hat{L}$  равна 0, и поэтому из условия  $\hat{L}u = 0$  следует

$$u_j = -\frac{1}{L_{jj}} \sum_{k \in N^-(j)} L_{jk} u_k \leq \max_{k \in N^-(j)} u_k.$$

Так как

$$-\frac{1}{L_{jj}} \sum_{k \in N^-(j)} L_{jk} = 1,$$

то равенство достигается, только когда все  $u_k$ , по которым ведётся суммирование, совпадают с  $u^{\max}$ . Следовательно, условие  $u_k < u^{\max}$  не выполняется. Таким образом, для всех  $j \in M$  выполняется  $N^-(j) \subset M$ . Следовательно, множество ячеек  $M$  может быть отделено от всех ячеек с  $u < u^{\max}$  только рёбрами, параллельными вектору  $\mathbf{a}$ . Но это противоречит условию. Следовательно, ячеек, в которых значение меньше  $u^{\max}$ , не существует.

Таким образом, у матрицы  $\hat{L}$  существует единственный (с точностью до множителя) собственный вектор, соответствующий нулевому собственному значению. Все компоненты этого вектора равны между собой. Утверждение доказано.

Введём обозначение  $v_i = |K_i|$ , где  $|K_i|$  – объём  $i$ -го сеточного элемента, и  $v = \{v_i\}$  – вектор размера  $|N^0|$ . Поскольку конечно-объёмная схема является консервативной,  $v$  является левым собственным вектором оператора  $\hat{L}$ , соответствующим нулевому собственному значению. По только что доказанному утверждению при выполнении его условий этот собственный вектор единственный с точностью до множителя.

**Утверждение 7.** Пусть  $\mathbf{f}$  определяется равенством (18), где  $L_{ij}$  – коэффициенты простейшей конечно-объёмной схемы. Тогда  $v^T \mathbf{f} = 0$ .

Доказательство.

$$v^T \mathbf{f} = \sum_{i \in N^0} v_i \mathbf{f}_i = \sum_{i \in N^0} \mathbf{a} v_i - \sum_{i \in N^0} \sum_{j \in N^\infty} (L_{ij} v_i) (\mathbf{r}_j - \mathbf{r}_i).$$

Поскольку  $L_{ij}$  при  $i \neq j$  отлично от нуля только при  $j \in N^-(i)$ , в сумму каждое ребро  $ij$  входит по одному разу. При этом  $L_{ij} = \mathbf{a} \cdot \mathbf{N}_{ij} / v_i$ . Среди граничных рёбер блока учитываются только те, для которых внутренняя нормаль имеет острый угол с направлением  $\mathbf{a}$ . Таким образом,

$$v^T \mathbf{f} = \mathbf{a} \sum_{i \in N^0} v_i - \sum_{i \in N^0} \sum_{j \in N^-(i)} (\mathbf{a} \cdot \mathbf{N}_{ij}) (\mathbf{r}_j - \mathbf{r}_i).$$

Обозначим за  $\mathbf{r}_{ij}$  центр масс ребра  $ij$  и представим

$$\mathbf{r}_j - \mathbf{r}_i = -(\mathbf{r}_{ij} - \mathbf{r}_j) + (\mathbf{r}_{ij} - \mathbf{r}_i).$$

Тогда

$$v^T \mathbf{f} = \mathbf{a} \sum_{i \in N^0} v_i - \sum_{i \in N^0} \sum_{j \in N^-(i)} (\mathbf{a} \cdot \mathbf{N}_{ij})(\mathbf{r}_{ij} - \mathbf{r}_i) + \sum_{i \in N^0} \sum_{j \in N^-(i)} (\mathbf{a} \cdot \mathbf{N}_{ij})(\mathbf{r}_{ij} - \mathbf{r}_j).$$

В силу периодичности выполняется

$$\sum_{i \in N^0} \sum_{j \in N^-(i)} (\mathbf{a} \cdot \mathbf{N}_{ij})(\mathbf{r}_{ij} - \mathbf{r}_j) = \sum_{i \in N^0} \sum_{j \in N^+(i)} (\mathbf{a} \cdot \mathbf{N}_{ji})(\mathbf{r}_{ij} - \mathbf{r}_i).$$

Поменяем местами  $i$  и  $j$  в правой части этого равенства и воспользуемся равенством  $\mathbf{N}_{ji} = -\mathbf{N}_{ij}$ . Тогда

$$v^T \mathbf{f} = \sum_{i \in N^0} \left( \mathbf{a} v_i - \sum_{j \in N(i)} (\mathbf{a} \cdot \mathbf{N}_{ij})(\mathbf{r}_{ij} - \mathbf{r}_i) \right).$$

В силу утверждения 1 выражение в скобках равно 0. Утверждение доказано.

**Утверждение 8.** Пусть шаг по времени удовлетворяет условию  $\Delta t < \tau^*$ . Тогда нестационарный геометрический корректор для простейшей конечно-объёмной схемы, в том числе при наличии периодических граничных условий, ограничен при  $t \rightarrow \infty$ .

Доказательство. Рассмотрим уравнение (18). Преобразуем матрицу  $\hat{L}$  к жорданову виду:  $\hat{L} = S^{-1}JS$ . Введём обозначение  $\mathbf{g}_i^n = \sum_j S_{ij} \Gamma_j^n$ . Тогда

$$\mathbf{g}^{n+1} = (I - \tau J) \mathbf{g}^n + S \mathbf{f} \Delta t. \quad (20)$$

Рассмотрим вначале случай связной области при отсутствии характеристических линий, полностью проходящих по рёбрам сетки. Так как нулевое собственное значение в этом случае является простым, соответствующее ему уравнение в системе (20) отщепляется. Пусть для определённости оно первое. Оно имеет вид  $\mathbf{g}_1^{n+1} = \mathbf{g}_1^n + v^T \mathbf{f} \Delta t$ . По только что доказанному утверждению  $v^T \mathbf{f} = 0$ . Из начальных условий  $\mathbf{g}^0 = S \Gamma^0 = 0$ . Отсюда  $\mathbf{g}_1^n = 0$  для всех  $n$ .

Все остальные собственные значения матрицы  $\hat{L}$  отличны от нуля. В силу неотрицательности элементов матрицы  $I - \Delta t \hat{L}$  простейшей конечно-объёмной схемы при  $\Delta t \leq \tau^*$  собственные значения матрицы  $I - \tau^* \hat{L}$  (а равно матрицы

$I - \tau^* J$ ) по модулю не превосходят единицы. Из утверждения 3 следует, что при  $\Delta t < \tau^*$  собственные значения матрицы  $I - \Delta t L$ , отличные от единицы, по модулю строго меньше единицы. Поэтому, независимо от размера жордановой клетки, соответствующие компоненты решения будут ограничены с ростом  $n$ .

Рассматривая теперь общий случай, можно заметить, что система уравнений (18) для несвязных областей, а равно для областей, разрезанных характеристиками по сеточным линиям, распадается на полностью независимые подсистемы уравнений. Для каждой из них применимо приведённое выше рассуждение. Утверждение доказано.

Таким образом, решение уравнения (18) для геометрического корректора удовлетворяет оценке  $\|\Gamma\| \leq C$ , где константа  $C$  зависит от расчётной сетки, но не от времени.

## Блочное измельчение

Исследователям, которым приходилось изучать поведение разностных схем на задаче Римана о распаде разрыва, знаком тот факт, что качество решения на вдвое более подробной сетке получается примерно таким же, как и при расчёте на вдвое большее время. Это связано с автомодельностью решения задачи Римана. Если сетка структурированная и равномерная, то отличие между решением на вдвое более подробной сетке и на вдвое большее время может определяться точностью интегрирования по времени.

На неравномерной или неструктурированной сетке такое сравнение напрямую невозможно, поскольку узлы сетки не переходят в узлы при гомотетии относительно начального положения разрыва. Эту проблему можно устранить, если определить специальную стратегию измельчения расчётных сеток.

**Определение 3.** Пусть расчётная область является квадратом и сетка на противоположных рёбрах квадрата совпадает. Блочным измельчением будем называть такую последовательность сеток, которая получается гомотетией исходной сетки в  $k$  раз и заполнением расчётной области полученными сеточными блоками. При этом будем считать, что характерные шаги старой и новой сетки отличаются в  $k$  раз.

Применим метод нестационарного геометрического корректора на задаче с периодическими условиями к исследованию точности конечно-объёмной схемы при блочном измельчении расчётной сетки.

При наличии периодических граничных условий по всем направлениям нестационарный геометрический корректор, очевидно, одинаков во всех блоках. Поэтому при решении уравнения (18) при формальном измельчении сетки можно задавать периодические граничные условия на границах одного блока. Это позволяет ввести обозначение  $\Gamma_j^n(\Delta t, h)$ ,  $j \in N^0$ , для коэффициентов

геометрического корректора. Таким образом, можно считать, что вектора геометрического корректора на последовательности сеток, получаемых блочным измельчением, принадлежат одному пространству.

**Утверждение 9.** Пусть дана некоторая сетка с характерным шагом  $h_0$ , и  $\tau^*$  определено на ней равенством (6). Рассмотрим последовательность сеток, характеризуемых шагом  $h = h_0 n / N$ , получаемых её блочным измельчением. Рассмотрим простейшую конечно-объёмную схему (2) при  $\tau(h) = \nu \tau^* h / h_0$ ,  $\nu < 1$ . Тогда нестационарный геометрический корректор для неё удовлетворяет равенству

$$\frac{\ln \|\Gamma^n(\tau(h_0), h_0)\| - \ln \|\Gamma^N(\tau(h), h)\|}{\ln(h_0/h)} = 1 - \frac{\ln \|\Gamma^N(\tau(h_0), h_0)\| - \ln \|\Gamma^n(\tau(h_0), h_0)\|}{\ln(N/n)}. \quad (21)$$

В этом выражении норма любая.

Доказательство. Рассмотрим уравнение (12). Оператор  $\hat{L}_h$  простейшей конечно-объёмной схемы (2), очевидно, удовлетворяет условию

$$\hat{L}_h = \hat{L}_{h_0} \frac{h_0}{h}.$$

Из уравнения (12) также видно, что его правая часть не меняется при блочном измельчении сетки. Таким образом, можно записать равенство

$$\begin{aligned} \Gamma^{N+1}(\tau(h), h) &= \left( I - \tau(h) \frac{h_0}{h} \hat{L}_{h_0} \right) \Gamma^N(\nu \tau^*, h) + \tau(h) \mathbf{f} = \\ &= \left( I - \tau(h_0) \hat{L}_{h_0} \right) \Gamma^N(\tau(h), h) + \tau(h_0) \frac{h}{h_0} \mathbf{f}. \end{aligned}$$

Сравнивая это выражение с аналогичным для  $h = h_0$  и учитывая начальные условия  $\Gamma^0(\Delta t, h) = 0$ , получаем

$$\Gamma^N(\tau(h), h) = \frac{h}{h_0} \Gamma^N(\tau(h_0), h_0).$$

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

$$\ln \|\Gamma^N(\tau(h), h)\| = \ln \frac{h}{h_0} + \ln \|\Gamma^N(\tau(h_0), h_0)\|.$$

Вычтем из каждой части  $\ln \|\Gamma^N(\tau(h_0), h_0)\|$  и поделим на  $\ln(h/h_0)$ , в результате получим искомое выражение. Утверждение доказано.

Доказанное утверждение означает, что геометрический корректор является автомодельным при блочном измельчении. Левая часть (21) есть численный порядок точности, посчитанный по сеткам с шагами  $h$  и  $h_0$  при  $t = n\tau(h_0) = N\tau(h)$ . Дробь в правой части есть логарифмическая производная решения по времени. Это позволяет определять порядок сходимости не по серии расчётов на измельчающихся сетках, а по росту ошибки в пределах одного расчёта.

**Теорема 10.** *При блочном измельчении сетки при  $\tau(h) = \nu\tau^*h/h_0$ ,  $\nu < 1$ , простейшая конечно-объёмная схема обладает первым порядком точности для нестационарного уравнения переноса.*

**Доказательство.** Рассмотрим последовательность блочно-неструктурированных сеток с характерными шагами  $h \rightarrow 0$ , получаемых блочным измельчением некоторой исходной сетки с характерным шагом  $h_0$ .

В силу утверждения 8 геометрический корректор на фиксированной сетке для простейшей конечно-объёмной схемы при наличии периодических условий имеет предельное значение. В силу утверждения 5 для любого  $\tau$  выполняется оценка  $\|\Gamma^n(\tau, h_0)\| \leq 2\|\Gamma^\infty(h_0)\|$ . Таким образом,

$$\|\Gamma^n(\Delta t, h)\| \leq h \frac{2\|\Gamma^\infty(h_0)\|}{h_0}.$$

Поскольку при блочном измельчении сетки максимальное расстояние между центрами соседних элементов имеет порядок  $O(h)$ , выполняются все условия теоремы 2. Применение этой теоремы доказывает искомое утверждение.

При помощи аналогичного рассуждения можно показать, что для начально-краевой задачи в полубесконечной области, если расчётная сетка составляется из бесконечного числа одинаковых сеточных блоков, то величина геометрического корректора (стационарного и нестационарного) остаётся ограниченной с удалением от границы расчётной области. В общем случае при сохранении качества элементов, как видно из следующего примера, величина стационарного геометрического корректора может расти со скоростью  $\sqrt{L}$ , где  $L$  – размер области.

## Пример Петерсона

В [10] для стационарного уравнения переноса

$$\alpha \cdot \nabla u + \beta u = f \tag{22}$$

был приведён пример задачи и расчётной сетки, в зависимости от стратегии измельчения которой наблюдался разный порядок сходимости. В работе

рассматривался разрывный метод Галёркина с кусочно-линейными и кусочно-постоянными базисными функциями. В последнем случае он совпадает с простейшей конечно-объёмной схемой; рассмотрением этого случая мы и ограничимся.

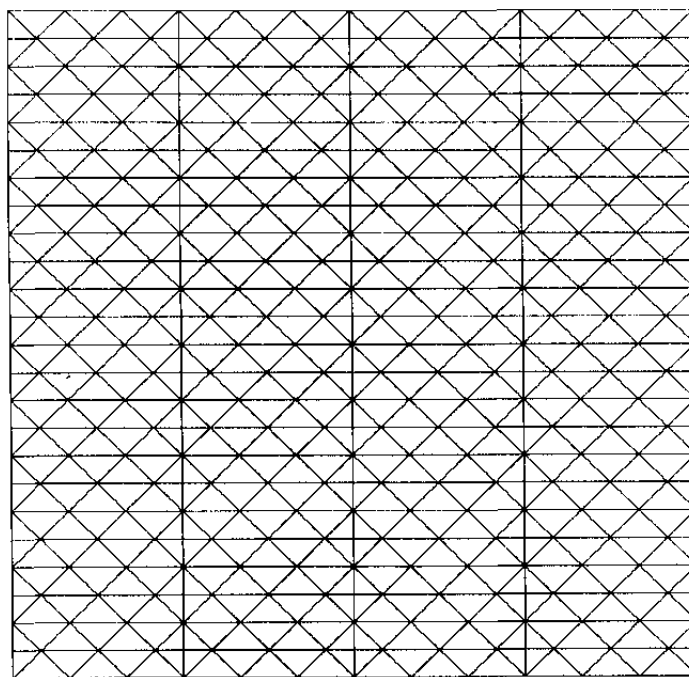


Рис. 1. Расчётная сетка для примера Петерсона

Расчётная область, использованная Петерсоном в [10], представляла собой единичный квадрат, а расчётная сетка в нём изображена на рис. 1. Число подобластей, отрезаемых на сетке вертикальными линиями, обозначено за  $m$ , причём  $m$  могло или быть постоянным, или увеличиваться при измельчении сетки. В уравнении (22) было задано  $\alpha_x = 0$ ,  $\alpha_y = 1$ ,  $\beta = 0$ ,  $f = 0$ . Граничные условия были заданы в виде  $u(x,0) = x^2$ .

В [10] было показано, что при  $m = 1/hc$  порядок сходимости равен 1, тогда как при  $m \sim h^{-\sigma}$ , где  $\sigma = \{0, 0.25, 0.5, 0.75\}$ , порядок сходимости в максимальной норме снижался до  $1/2$ .

Измельчение сетки с условием  $m = 1/hc$  является частным случаем рассматриваемого в настоящей работе блочного измельчения. Если же при измельчении сетки увеличивать размер блока, то геометрический корректор будет расти как  $\sqrt{l}$ , где  $l$  – размер расчётной области, а решение уравнения для нестационарного геометрического корректора будет расти со временем как  $\sqrt{t}$ . Это эквивалентно сходимости с порядком  $1/2$  в максимальной норме. Таким образом, утверждение, доказанное в настоящей работе, согласуется с примером Петерсона.

## Численное исследование

Для проверки теоретических результатов, доказанных в настоящей работе, рассмотрим уравнение

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial y} = 0$$

в единичном квадрате с периодическими граничными условиями при  $u(t, x, 0) = u(t, x, 1)$ . При  $x = 0$  и  $x = 1$  задание граничных условий не требуется. Будем рассматривать простейшую конечно-объемную схему (2). Начальные условия зададим в виде  $u(0, x, y) = x^2 \cos(2\pi y)$ , что приводит к точному решению  $u(t, x, y) = x^2 \cos(2\pi(y - t))$ . Если не оговорено иное, расчёт будем проводить до времени  $t = 1$ . Будем использовать расчётную сетку, предложенную Петерсоном (см. рис. 1).

Пусть  $\Pi$  есть точечный оператор проектирования, то есть  $(\Pi u)_j^n$  есть значение функции  $u$  в момент времени  $t^n$  в центре масс треугольника  $K_j$ , а  $\tilde{\Pi}$  – оператор проектирования, заданный формулой (8).

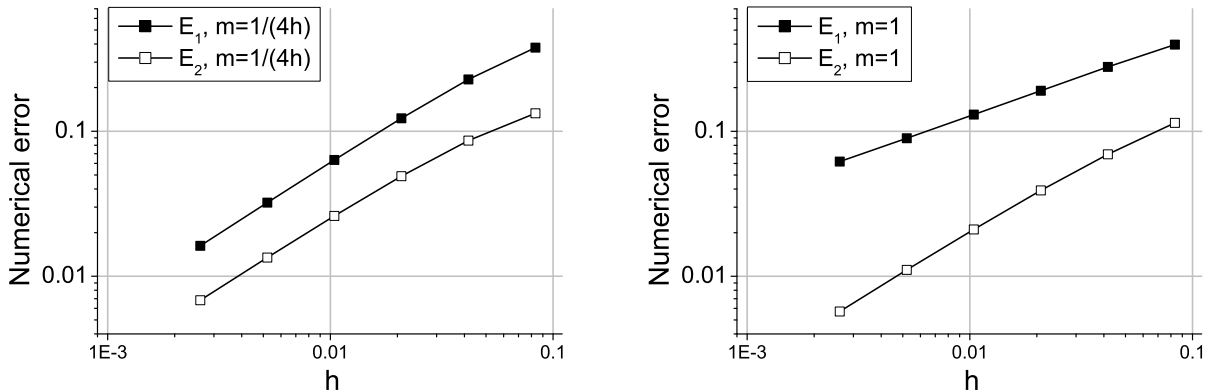


Рис. 2. Численная ошибка в смысле операторов  $\Pi$  (слева) и  $\tilde{\Pi}$  (справа)

Вначале покажем, что старший член численной ошибки определяется величиной геометрического корректора, как это предсказывается теоремой 2. Для этого на рис. 2 на последовательности измельчающихся сеток отложим величины  $E_1$  и  $E_2$  на момент времени  $t = 1$ , где

$$E_1 = \|u_h - \Pi u\|_\infty, \quad E_2 = \|u_h - \tilde{\Pi} u\|_\infty = \|u_h - \Pi(u + \Gamma \cdot \nabla u)\|_\infty,$$

где нормы вычисляются по временному слою, соответствующему  $t = 1$ . Слева на рис. 2 изображены результаты при фиксированном размере блока ( $m = 1/(4h)$ ), справа – при  $m = 1$ . Из графиков можно сделать следующие выводы. При блочном измельчении простейшая конечно-объемная схема обладает первым порядком точности; численная ошибка в смысле оператора  $\tilde{\Pi}$  также



является величиной первого порядка малости по  $h$ . При  $m = 1$  простейшая конечно-объёмная схема показывает половинный порядок точности. Однако нестационарный геометрический корректор верно описывает старший по величине член численной ошибки, и ошибка за вычетом слагаемого  $\Gamma \cdot \nabla u$  имеет величину  $O(h)$ .

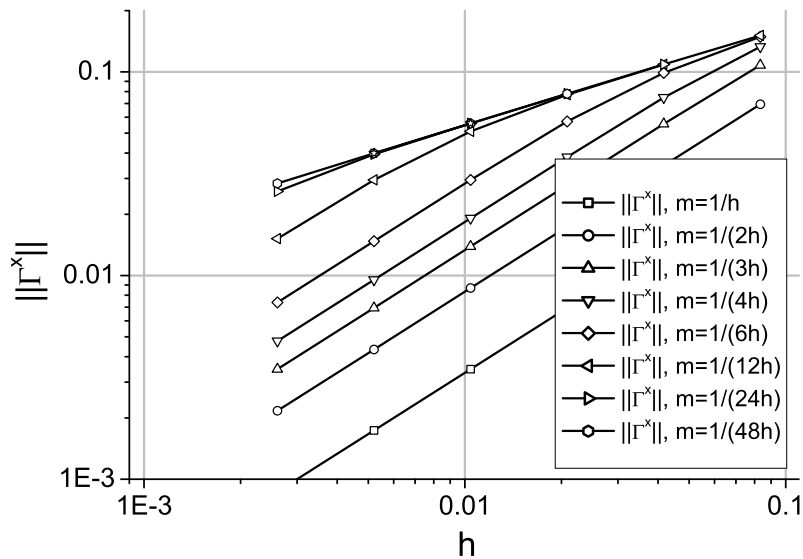


Рис. 3. Зависимость  $\|\Gamma^x\|$  от  $h$  при блочном измельчении для разных размеров блоков

Теперь исследуем поведение геометрического корректора при  $m = 1/(Mh)$  (то есть при блочном измельчении) в зависимости от выбора ширины блока  $M$ . Зависимость  $\|\Gamma^x\|$  и  $\|\Gamma^y\|$  от  $h$  представлена на рис. 3 и рис. 4 соответственно. Видно, что  $\|\Gamma^y\|$  пропорционален  $h$  и не зависит от  $m$ . Можно также заметить, что нестационарный геометрический корректор примерно за 10 итераций по времени выходит на стационарное решение с машинной точностью (эти данные мы для краткости не приводим), то есть при  $t = 1$   $y$ -компонента нестационарного геометрического корректора совпадает с  $y$ -компонентой стационарного геометрического корректора.

Полностью иная картина имеет место для  $\|\Gamma^x\|$ . При любом фиксированном числе  $M$  на момент времени  $t = 1$  величина  $\|\Gamma^x\|$  имеет порядок  $O(h)$ . Таким образом, простейшая конечно-объёмная схема действительно показывает первый порядок точности при блочном измельчении, как это предписывает теорема 10. Однако, во-первых, коэффициент при  $h$  существенно зависит от размера блока и при  $M \gg 1$  пропорционален  $M$ . А во-вторых, чем больше размер блока, тем сильнее нужно измельчать сетку, чтобы выйти на асимптотическое по-

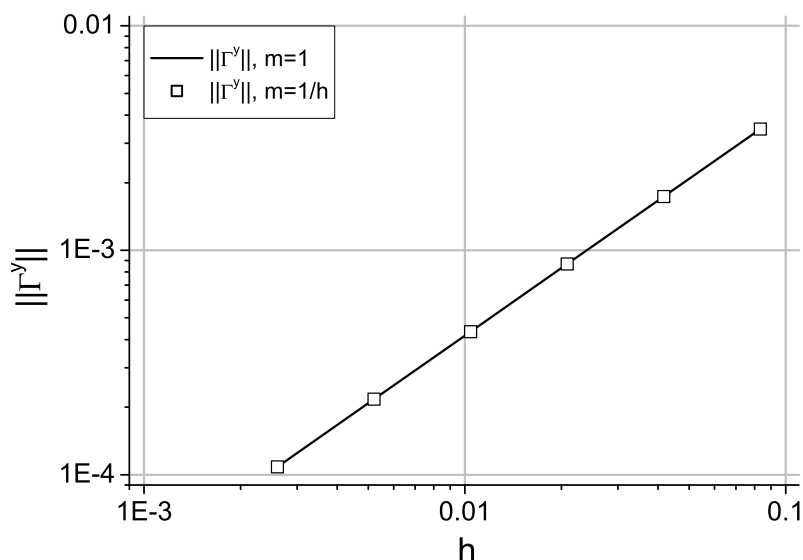


Рис. 4. Зависимость  $\|\Gamma^y\|$  от  $h$  при блочном измельчении для разных размеров блоков

ведение ошибки. Поэтому если размер блока является достаточно большим, то формальный порядок точности при блочном измельчении остаётся равным 1, но на практике блочное измельчение мало чем отличается от произвольного измельчения, где наблюдается сходимость с порядком  $O(h^{1/2})$ . Следовательно, ограниченность размера блока при измельчении сетки является существенным условием.

Полученный результат, в частности, показывает некорректность обоснования высокой точности схем численными экспериментами на блочной сетке из маленьких “плохих” блоков, как это было сделано автором в [18] для конечно-объёмной схемы повышенной точности.

Наконец, покажем автомодельность геометрического корректора, предсказываемого теоремой 9. Для этого выведем зависимость нормы геометрического корректора от времени при разных  $h$ . Слева на рис. 5 отложим случай  $M = 4$  (маленький размер блока), справа – случай  $M = 40$  (большой размер блока). Из графика видно, что поведение геометрического корректора является автомодельным. По горизонтальной оси в логарифмическом масштабе отложено отношение времени к характерному шагу сетки:  $t/h$ , по вертикали, также в логарифмическом масштабе, –  $\|\Gamma^x\|$ . Маркерами отмечены значения, посчитанные на сетке с шагом  $h = 1/40$ , а линиями – посчитанные на сетке с шагом  $h = 1/80$ .

Рис. 5 показывает, что характерное время, за которое норма геометрического корректора выходит на стационарное значение (а следовательно, зависи-

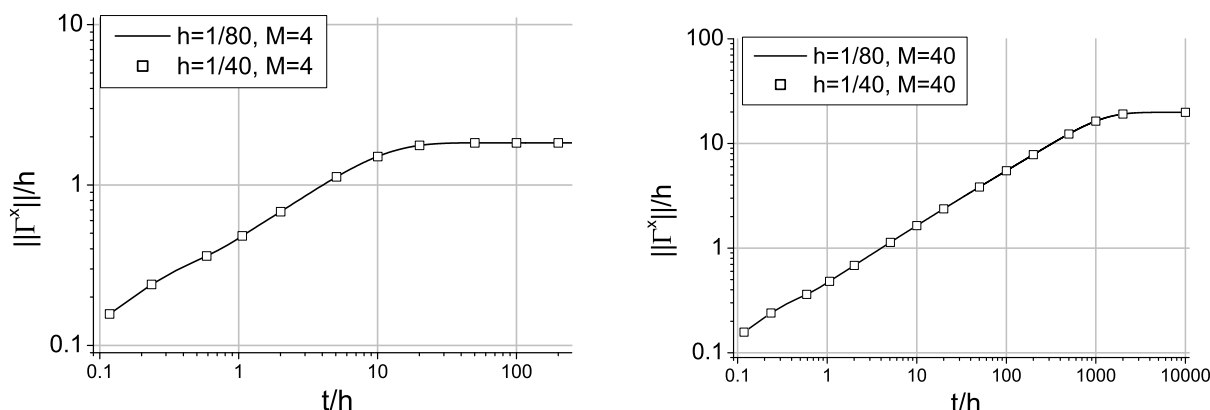


Рис. 5. Зависимость геометрического корректора от времени при  $M = 4$  (слева) и  $M = 40$  (справа)

мость ошибки от шага сетки выходит на асимптотику  $O(h)$ , может быть очень большим. Оно никак не связано с прохождением блока расчётной сетки распространяющейся по ней физической волной, в двумерном случае может быть много больше него и быстро расти с размером блока. В широком диапазоне времени и мелкости сеток для  $M = 40$  вместо выхода нормы геометрического корректора на стационарное значение наблюдается её рост со скоростью  $O(\sqrt{t})$ . Поэтому, согласно утверждению 9, численный порядок точности должен быть равен  $1/2$ . Это согласуется с численными результатами: на рис. 3 видно, что при  $M = 48$  ( $m = 1/(48h)$ ) наблюдается сходимость численного решения к точному с порядком  $1/2$ .

## Заключение

В работе был предложен нестационарный метод геометрического корректора для исследования точности разностных схем для уравнения переноса на гладких решениях. Показано, что член при старшей производной в разности численного решения и точного описывается нестационарным геометрическим корректором, и он может быть найден из решения уравнения (12). Порядок малости этого члена совпадает с порядком малости всей численной ошибки.

С помощью разработанного метода показано, что при блочном измельчении расчётной сетки простейшая конечно-объёмная схема обладает первым порядком точности. Этот теоретический результат подтверждён в численном эксперименте. Однако последний показывает, что коэффициент при  $h$  в оценке точности существенно растёт с ростом размера блока. В результате блочное измельчение при достаточно большом размере блока может иметь кажущуюся асимптотику, показывающую сходимость со скоростью  $O(h^{1/2})$ . Однако эта кажущаяся асимптотика гораздо лучше отражает реальную точность схемы, чем

настоящая.

Таким образом, при блочном измельчении и малом размере блока простейшая конечно-объёмная схема обладает свойством сверхсходимости. Оценка точности на сетке, составленной из малых одинаковых блоков, ничего не говорит о точности схемы на неструктурированной сетке. Блочное измельчение при большом размере блока, напротив, пригодно для оценки точности схем на неструктурированной сетке при периодических граничных условиях.

При блочном измельчении нестационарный геометрический корректор является автомодельным, то есть точность схемы можно оценивать не по скорости убывания нормы геометрического корректора при измельчении сетки, а по её росту со временем на фиксированной сетке. Этот результат остаётся справедливым независимо от размера блока. Таким образом, для оценки скорости сходимости численного решения к точному может быть достаточно проведения расчёта на одной сетке, а не на последовательности сеток.

В следующей работе мы обобщим предложенную технологию оценки точности решения на рёберно-ориентированные схемы.

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

## 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. Despres B. An explicit a priori estimate for a finite volume approximation of linear advection on non-cartesian grids // SIAM Journal on Numerical Analysis. 2004. Vol. 42. P. 484–504.
3. Merlet B. Linf and L2-error estimates for a finite volume approximation of linear advection // SIAM Journal on Numerical Analysis. 2007. Vol. 46. P. 124–150.
4. 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.
5. Reed W. H., Hill T.-R. Triangular mesh methods for the neutron transport equation: Tech. Rep.: 479: Los Alamos Scientific Laboratory, 1973.
6. Lesaint P., Raviart P.-A. Mathematical aspects of finite elements in partial differential equations / Ed. by C. de Boor. Academic Press, New York, 1974.

7. Johnson C., Pitkaranta J. An analysis of the Discontinuous Galerkin method for a scalar hyperbolic equation // *Mathematics of computation*. 1986. Vol. 46. P. 1–26.
8. Richter G. An optimal-order error estimate for the Discontinuous Galerkin method // *Mathematics of Computation*. 1988. Vol. 50. P. 75–88.
9. Peterson T. Convergence properties of the discontinuous Galerkin method for a scalar hyperbolic equation. Ph.D. thesis: Cornell University. Ithaca, NY, 1990.
10. Peterson T. A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation // *SIAM J. Numer. Anal.* 1991. Vol. 28. P. 133–140.
11. 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.
12. Despres B. Lax theorem and finite volume schemes // *Mathematics of Computation*. 2003. Vol. 73. P. 1203–1234.
13. Pascal F. On supra-convergence of the finite volume method for the linear advection // *ESAIM: PROCEEDINGS*. 2007. Vol. 18. P. 38–47.
14. Geuzaine C., Remacle J.-F. Gmsh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities. 1997.
15. Bouche D., Ghidaglia J.-M., Pascal F. Theoretical analysis of the upwind finite volume scheme on the counter-example of Peterson // *ESAIM: Mathematical Modelling and Numerical Analysis*. 2010. Vol. 44. P. 1279–1293.
16. Bouche D., Ghidaglia J.-M., Pascal F. Error estimate for the upwind finite volume method for the nonlinear scalar conservation law // *Journal of Computational and Applied Mathematics*. 2011. Vol. 235. P. 5394–5410.
17. Delarue F., Lagoutiere F. Probabilistic analysis of the upwind scheme for transport equation // *Arch. Rational Mech. Anal.* 2011. Vol. 199. P. 229–268.
18. Бахвалов П.А., Козубская Т.К. Экономичная формулировка схем с квазиодномерной реконструкцией переменных // *Препринты ИПМ им. М.В.Келдыша*. 2013. № 89. С. 1–16. URL: <http://library.keldysh.ru/preprint.asp?id=2013-89>.