

ИСПОЛЬЗОВАНИЕ ИНСТРУМЕНТОВ АНАЛИЗА БОЛЬШИХ ДАННЫХ ПРИ РЕШЕНИИ ЗАДАЧ ГАЗОВОЙ ДИНАМИКИ И АКУСТИКИМ.Д. Калугин¹В.Н. Корчагова^{1,2}М.В. Крапошин¹И.К. Марчевский^{1,2}

iliamarchevsky@bmstu.ru

В.С. Морева²¹ Институт системного программирования им. В.П. Иванникова РАН,
Москва, Российская Федерация² МГТУ им. Н.Э. Баумана, Москва, Российская Федерация**Аннотация**

Рассмотрены вопросы обработки и анализа больших объемов данных, возникающих при численном моделировании процессов гидрогазодинамики и акустики в крупных промышленных установках. Представлены оценки объема обрабатываемых данных на примере задачи расчета течения и акустических нагрузок при старте ракеты-носителя. Для сокращения используемого объема памяти предложено применять технологию обработки и сжатия данных *Proper Orthogonal Decomposition*, с помощью которой выявляют наиболее «энергоемкие» моды исходных данных. Возможности этого подхода, реализованного с применением фреймворка *Apache Spark*, продемонстрированы на тестовой задаче распространения акустических волн в одномерной и двумерной постановках. Объем данных, необходимый для восстановления решения с приемлемой погрешностью порядка 1 %, удалось сократить в 30 раз в одномерном случае и в 160 раз в двумерном

Ключевые слова

Обработка больших данных, собственное ортогональное разложение, газовая динамика, акустика

Поступила в редакцию 10.08.2017

© МГТУ им. Н.Э. Баумана, 2018

Работа выполнена при финансовой поддержке Минобрнауки России (идентификатор RFMEFI60714X0090, соглашение № 14.607.21.0090)

Введение. Задачи вычислительной газовой динамики и акустики, даже в упрощенных постановках, являются одними из наиболее вычислительно сложных. Это связано с особенностями соответствующих математических моделей, включающих в себя нелинейные уравнения и определяющие соотношения. Еще более сложен расчет течений среды в реально существующих промышленных установках, моделирование протекания газа через элементы инженерных конструкций и т. п.

Расчетные сетки в актуальных в настоящее время задачах такого типа могут содержать сотни миллионов и даже миллиарды ячеек, число выполняемых шагов по времени может иметь тот же порядок.

Использование современных высокопроизводительных вычислительных комплексов, имеющих в составе тысячи вычислительных ядер, может позволить решить подобные задачи [1], однако даже простое сохранение результатов расчетов в узлах/ячейках сетки с некоторым шагом по времени требует огромных затрат дискового пространства, а их последующая обработка — еще и значительных объемов оперативной памяти, которыми располагают лишь многопроцессорные кластеры. При этом во многих случаях пространственное разрешение получаемых результатов (которое определяется расчетной сеткой) оказывается явно излишним для практических целей при их обработке и интерпретации, но совершенно необходимым для получения качественно и количественно правильного решения.

В связи с этим наиболее перспективно использовать методы приближенного представления результатов расчетов, основанных на их «малоранговых» аппроксимациях, обзор которых можно найти в работе [2]. Среди наиболее часто используемых выделим такие методы, как *Proper Orthogonal Decomposition (POD)* [3, 4], *Dynamic Mode Decomposition (DMD)* [5]. Суть таких методов сводится к приближенной замене искомого распределения некоторой физической величины (заданной непрерывно или дискретно в узлах/ячейках сетки и аналогично по времени)

суммой вида $\Psi(t, \vec{r}) \approx \sum_{k=0}^m a_k(t) \varphi_k(\vec{r})$, где набор скалярных функций $\{\varphi_k(\vec{r})\}_{k=0}^m$

имеет смысл пространственного базиса, заданного соответственно в области решения задачи или в виде сеточных функций. Естественным требованием является малое число базисных функций по сравнению с числом узлов пространственной сетки и числом временных слоев, в которых задана аппроксимируемая величина, обеспечивающее тем не менее качественное воспроизведение величины Ψ .

Желательным также является условие ортогональности базиса $\{\varphi_k(\vec{r})\}_{k=0}^m$ в смысле гильбертова пространства L_2 или конечномерного пространства \mathbb{R}^n (в этом случае значения величины Ψ в узлах/ячейках сетки будем рассматривать как элементы n -мерного вектора). Последнее требование, кроме прочего, существенно упрощает процедуру вычисления коэффициентов $a_k(t)$.

Всем перечисленным требованиям отвечает известная технология анализа и сжатия больших объемов данных *Proper Orthogonal Decomposition* [3, 4], использование которой при решении задач газовой динамики и акустики рассмотрено в настоящей работе.

Моделирование газодинамики и акустики старта ракеты-носителя. Примером актуальной прикладной задачи, в которой использование технологий анализа больших данных оправданно и востребованно, может стать проблема численного моделирования газодинамических процессов, сопровождающих старт ракеты-носителя (РН) и воспроизведение уровня акустического воздействия на РН и объекты инфраструктуры стартового комплекса (рис. 1). В такой задаче требуется определить механические и тепловые нагрузки на поверхности газоотводного лотка (газохода) и стартового стола, спектр акустических нагрузок вдоль корпуса ракеты и ее головной части.

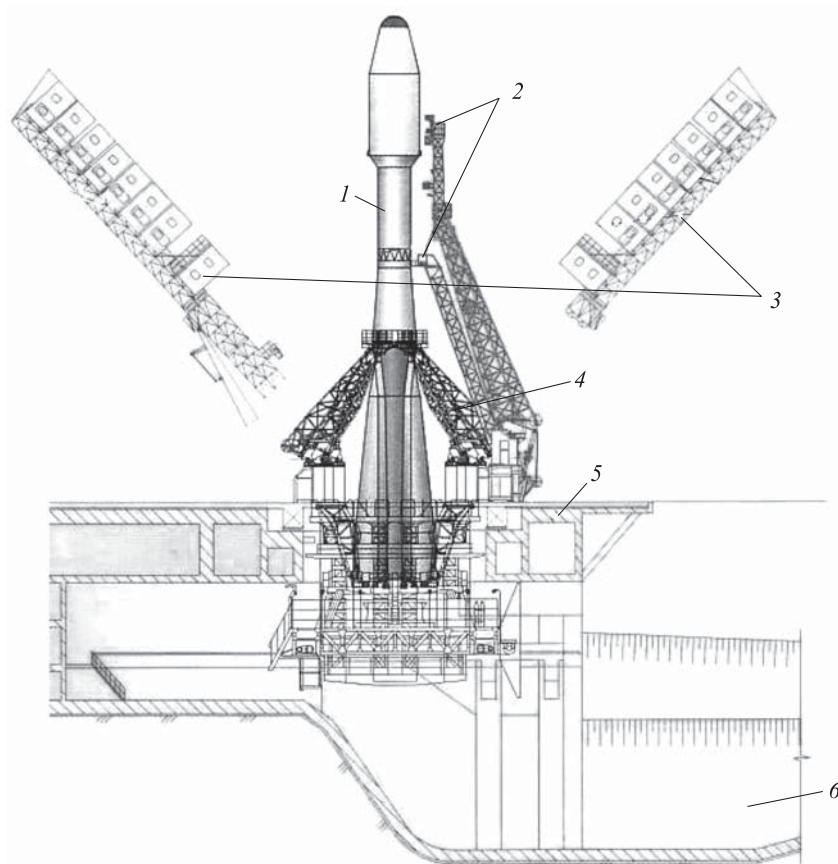


Рис. 1. Схема стартового комплекса РН «Союз»:

1 — РН с космической головной частью; 2 — кабель-заправочные мачты; 3 — фермы обслуживания с площадками обслуживания; 4 — фермы-опоры; 5 — стартовое сооружение; 6 — газоотводный лоток

Актуальность указанной проблемы сложно переоценить в связи с возросшим интересом к ракетам тяжелого класса (*Atlas-V*, *Falcon-9*, *Antares*, *Ariane-5*, «Энергия»), для которых проблема подавления акустического шума стоит наиболее остро, особенно с учетом возникающих из условий географии стоимости и сложности работ по ограничению габаритов стартовых сооружений.

Очевидно, что в силу сложности и уникальности объекта полномасштабное физическое моделирование невозможно. Для повышения точности результатов физического масштабного моделирования требуется проведение исследований для различных масштабов (1:50, 1:5 и т. д.), что приводит к существенным финансовым затратам и увеличению времени проектирования.

Следовательно, численное моделирование газовой динамики и акустики процессов старта РН во многом становится единственным инструментом априорного анализа проектируемых стартовых комплексов.

В таких задачах сложность газодинамического моделирования обусловлена разнообразием происходящих процессов и многомасштабностью задачи. Объем

пространства моделирования определяется габаритами стартового стола и сооружений, составляющих порядка сотни метров по каждому направлению, в то время как характерный размер расчетной сетки по пространству может составлять около 0,01 м. Таким образом, число неизвестных в задаче будет составлять величину порядка $10^{12} \dots 10^{13}$, что делает предъявляемые требования к вычислительным мощностям невыполнимыми в обозримом будущем.

Для решения таких задач используют гибридный метод, предполагающий разделение области моделирования на так называемые ближнее и дальнее поля (рис. 2).

В полностью нелинейной постановке газодинамическая задача решается только в ближнем поле (см. рис. 2); в остальной части расчетной области решаются линеаризованные уравнения, сводящиеся к уравнению распространения акустических колебаний. Решение задачи на границе первой области (см. рис. 2) является исходными данными для задачи во второй области. Для обеспечения надлежащего качества решения требуется оценить размер ячеек сетки — сеточное разрешение. Оно определяется путем решения модельных задач, имеющих известное аналитическое решение или тщательно исследованных экспериментально и численно. Применительно к рассматриваемой задаче необходимо такое сеточное разрешение, которое позволит корректно воспроизводить газодинамические процессы истечения сверхзвуковых струй из сопел РН и распространение акустических волн давления.

При моделировании истечения сверхзвуковых струй основная задача — воспроизведение ударно-волновых структур (ячеек расширения-сжатия, называемых иногда «бочками») и их взаимодействия с турбулентным слоем смешения. В истекающей из сопла сверхзвуковой струе можно выделить два участка [6]:

- 1) область невязкого истечения газа (2–3 «бочки» от среза сопла);
- 2) область турбулентного смешения струи с внешней средой.

Обширное сравнение различных численных методов с отечественными и зарубежными экспериментальными данными проведено в работе [6]. По результатам сравнения сделан вывод, что для получения удовлетворительного соответствия, по крайней мере на начальном (невязком) участке, требуется сеточное разрешение не менее 160 ячеек на диаметр сопла.

Еще один достаточно распространенный набор тестов описан в работе [7]. В экспериментальной серии рассмотрено истечение сверхзвуковой струи в широ-

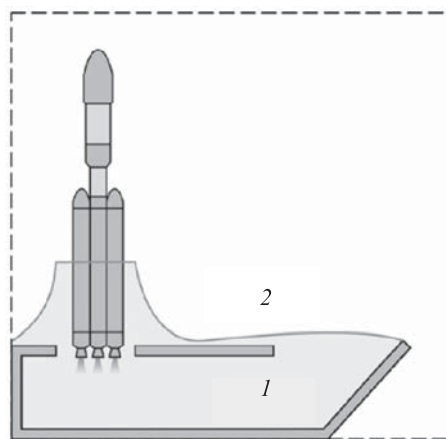


Рис. 2. Принципиальная схема вычислительного эксперимента с использованием гибридного метода расчета акустических нагрузок:

1, 2 — ближнее и дальнее поля

ком диапазоне режимов — от перерасширенных до недорасширенных истечений струи. Каждый режим характеризуется параметром истечения $\beta = \sqrt{M^2 - 1}$, где M — число Маха на срезе сопла при изоэнтропном расширении.

Для численного решения задачи в ближнем поле был выбран гибридный метод решения уравнений газовой динамики, позволяющий моделировать течения в широком диапазоне чисел Маха [8, 9]. Моделирование проведено от среза сопла в двумерной (осесимметричной) постановке для двух режимов [7]: $\beta = 1,1$ и $\beta = 2,0$. В расчетах использованы четыре сетки с сеточным разрешением 20, 40, 80 и 160 ячеек на диаметр сопла (сетки 1–4 соответственно) и модель турбулентности $k-\omega$ SST с пристеночными функциями. Результаты моделирования приведены на рис. 3.

Согласно результатам моделирования (рис. 3), решение удовлетворительно согласуется с экспериментом для первых двух «бочек» при сеточном разрешении от 80 ячеек на диаметр. Выбрав за основную сетку 4 и задавшись характерным размером диаметра сопла около 1,5 м, получим, что максимальный шаг по пространству, необходимый для воспроизведения сверхзвуковых струй, равен $h_{\max} = 0,009$ м.

Еще одним процессом, требующим корректного воспроизведения, является распространение акустических волн давления, генерируемых пульсациями сверхзвуковых струй и их взаимодействием с элементами конструкции. В рамках принятого подхода (метод контрольного объема, второй порядок аппроксимации по пространству) для корректного моделирования акустических явлений с учетом численной вязкости и диссипации необходимо по крайней мере 20 ячеек сетки на длину волны. Таким образом, если требуется разрешить распространение низкочастотных акустических волн в диапазоне частот до $f_{\max} = 1000$ Гц в воздушной среде с температурой 293 К (нормальные условия), то получаем, что размер ячейки сетки должен составлять:

$$h_{\max} = \frac{c}{20 f_{\max}} = \frac{\sqrt{\gamma \tilde{R} T}}{20 f_{\max}}; \quad h_{\max} \approx \frac{343 \text{ м/с}}{20 \cdot 1000 \text{ Гц}} \approx 0,017 \text{ м.}$$

Действительная температура газа в газоходе и на выходе из него намного выше — порядка тысячи градусов, что означает увеличение максимально допустимого размера ячейки примерно в 2 раза ($h_{\max} \approx 0,035$ м) в горячем потоке.

Следовательно, в рассматриваемой задаче основное ограничение на характерный размер ячеек сетки в ближнем поле накладывает необходимость разрешения пульсаций сверхзвуковой струи, а не воспроизведение акустических волн. Приняв за максимальный размер шага по пространству минимальное из найденных значений ($h = 0,009$ м), можно оценить число ячеек сетки, необходимое для проведения газодинамического расчета в газоходе (ближнем поле) при старте ракеты, которое приблизительно составляет $3,5 \cdot 10^9$ при объеме начального участка газохода около 3000 м^3 .

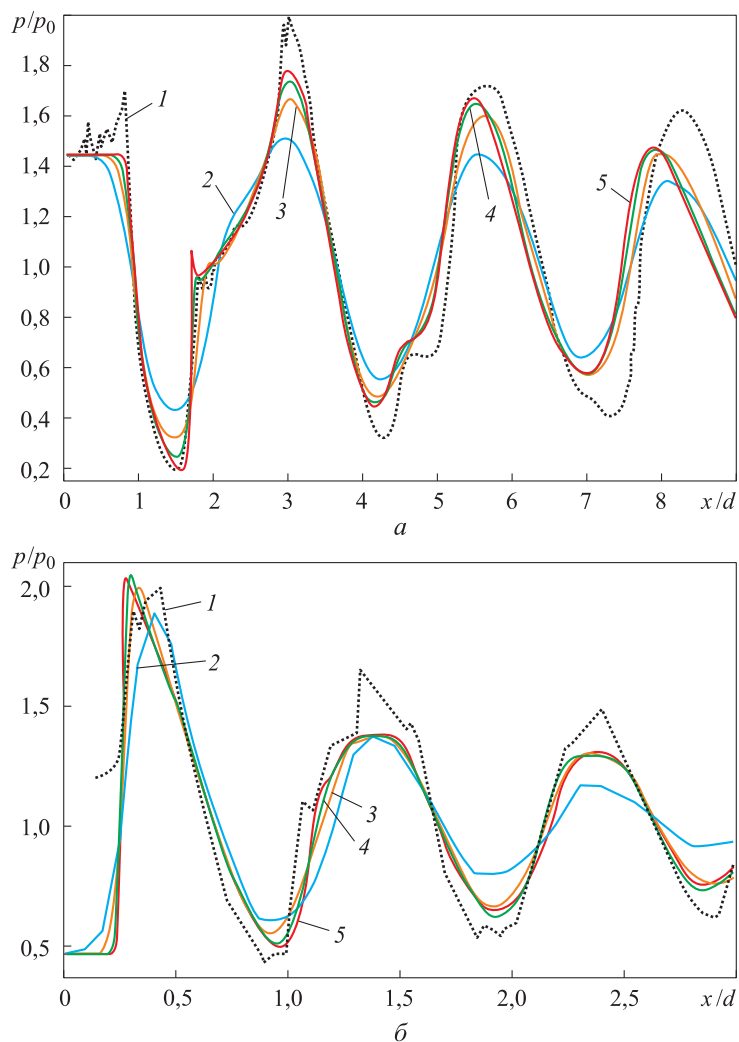


Рис. 3. Сравнение расчетного и экспериментального распределений давления вдоль оси струи для случаев недорасширенного (а) и перерасширенного (б) истечений струи из сверхзвукового сопла:

1 — эксперимент; 2–5 — расчет для сеток 1–4

Согласно выполненным оценкам шага по пространству, число элементов на границе ближнего и дальнего полей составляет приблизительно 10^7 . Оценим объем хранимых данных на этой поверхности: физическое время моделирования может составлять примерно 1 с; шаг записи, необходимый для разрешения частоты в 1 кГц, должен быть по крайней мере $1/2000$ с. Полагая, что для хранения одного числа с плавающей точкой требуется 8 байт, получаем, что общий объем данных только для хранения давления в ячейках поверхностной сетки на границе ближнего и дальнего полей составит около 150 Гб для одного расчета, что делает невозможным обработку и дальнейшее использование этих данных без предварительного сжатия.

При выполнении оценочных расчетов, если использовать достаточно грубую сетку и максимально уменьшить размер зоны ближнего поля, число ячеек на ее поверхности будет приблизительно 10^5 . Это снижает потребность в памяти до 1,5 Гб, что все равно неприемлемо для практических целей.

Описание метода *POD*. Пусть имеется некоторый объем расчетных данных, заданных в узлах сетки в последовательные моменты времени t_1, t_2, \dots, t_N . Предположено, что эти данные соответствуют результатам решения нестационарной, к примеру, газодинамической задачи, сохраненным с некоторым шагом (как правило, существенно превышающим шаг по времени в нестационарном расчете). При исследовании поля скалярной величины исходные данные представим в виде набора n -мерных векторов (n — число ячеек сетки) $\vec{x}^i, i = 1, \dots, N$. Для обработки векторных и (или) нескольких скалярных величин векторы \vec{x}^i представимы как $(p \cdot n)$ -мерные векторы, в которых подряд записаны все компоненты решения во всех ячейках.

Для построения базиса $\{\varphi_k(\vec{r})\}_{k=1}^N$, который в таком случае представляет собой набор векторов $\{\vec{\varphi}_k\}_{k=1}^N$, выполняются следующие операции.

1. Вычисление вектора средних (по времени) значений решения во всех ячейках: $\vec{x}^{aver} = \frac{1}{N} \sum_{i=1}^N \vec{x}^i$.

2. Составление симметричной положительной матрицы ковариаций размерностью $N \times N$ с компонентами $R_{ij} = \frac{1}{N} (\vec{x}^i - \vec{x}^{aver}, \vec{x}^j - \vec{x}^{aver})$.

3. Определение собственных чисел σ_k и собственных векторов \vec{v}_k матрицы R . При этом важно, чтобы все собственные числа были отсортированы по убыванию.

4. Задание векторов базиса $\{\varphi_k(\vec{r})\}_{k=1}^N$ как $\vec{\varphi}_k = \sum_{j=1}^N (\vec{v}_k)_j \vec{x}^j$. Эти векторы всегда являются ортогональными; для удобства их можно нормировать.

Для вычисления коэффициентов разложения $\{a_k(t_i)\}_{k=1}^N$, соответствующих решению в i -й момент времени, применяют стандартную процедуру разложения по ортогональному базису: $a_k(t_i) = \frac{(\vec{x}^i, \varphi_k)}{(\varphi_k, \varphi_k)}$. Тогда решение в i -й момент времени можно представить в виде суммы $\vec{x}^i = \vec{x}^{aver} + \sum_{k=1}^N a_k(t_i) \vec{\varphi}_k$. Особенностью метода *POD*, используемого для построения базиса, является тот факт, что этот базис является оптимальным в смысле [3]:

векторы базиса $\vec{\varphi}_i$ таковы, что

$$\sum_{i=1}^N (\vec{x}^i, \vec{\varphi}_1)^2 \rightarrow \max \text{ при } (\vec{\varphi}_1, \vec{\varphi}_1) = \text{const};$$

$$\sum_{i=1}^N (\bar{x}^i, \bar{\varphi}_2)^2 \rightarrow \max \text{ при } (\bar{\varphi}_2, \bar{\varphi}_2) = \text{const и } (\bar{\varphi}_2, \bar{\varphi}_1) = 0 \text{ и т. д.}$$

Подобный выбор базиса обеспечивает выполнение следующего свойства: для любого значения числа $m = 1, \dots, N$ базисных векторов сумма квадратов погрешностей аппроксимации $\sum_{i=1}^N \left\| \bar{x}^i - \left(\bar{x}^{aver} + \sum_{k=1}^m a_k(t_i) \bar{\varphi}_k \right) \right\|^2$ принимает наименьшее значение по всем возможным способам выбора ортогональных базисных векторов $\{\bar{\varphi}_k\}_{k=1}^m$. Это означает, что каждый последующий вектор базиса выбирается так, чтобы удерживать максимально возможный объем «информации» об исходном решении.

На практике достаточно ограничиться использованием сравнительно небольшого числа m ($m \ll N$) базисных векторов, при этом относительная погрешность аппроксимации решения будет определяться величиной

$$\varepsilon = \sqrt[4]{1 - \frac{\sum_{j=1}^m \sigma_j^2}{\sum_{j=1}^N \sigma_j^2}}.$$

Таким образом, метод *POD* включает в себя методiku априорной оценки погрешности аппроксимации.

Использование метода *POD* при решении одномерной задачи. Рассмотрим одномерную задачу о моделировании состояния совершенного нетеплопроводного невязкого сжимаемого газа, занимающего область $x \in [0; L]$. В начальный момент времени газ покоится, давление в нем непостоянно и задано функцией

$$p(x, 0) = p_0 + p_1 \frac{x(x-L/3)(x-L)}{L^3} \sin \frac{5\pi x}{L},$$

где $p_1/p_0 = 0,01$.

По приведенному распределению с использованием уравнения состояния совершенного газа вычислялось распределение плотности. На границах расчетной области моделировалась твердая стенка. Задача решена в пакете *OpenFOAM* гибридным методом [9, 10]; использована равномерная расчетная сетка из 1000 ячеек; длина расчетной области $L=1$, давление $p_0=1$, температура $T_0=1$, параметры газа: индивидуальная газовая постоянная $\tilde{R}=0,4$; показатель адиабаты $\gamma=1,4$. К моменту времени $t=5$ возмущения в такой среде проходят расстояние $3,75L$, т. е. успевают несколько раз отразиться от стенок.

Сохранение результатов проведено с шагом $\Delta t = 0,01$ с, таким образом, решение представляло собой $N=500$ векторов $\bar{x}^i \in \mathbb{R}^n$, $n=1000$. В результате применения метода *POD* была построена матрица ковариаций R размерностью 500×500 , модули первых 100 ее собственных чисел показаны на рис. 4, а.

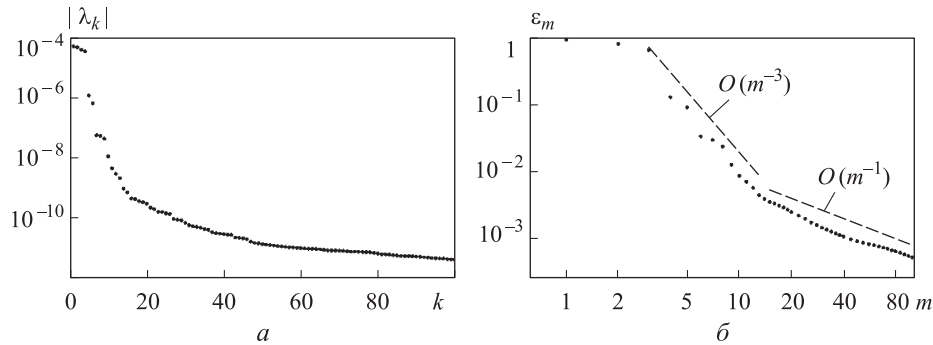


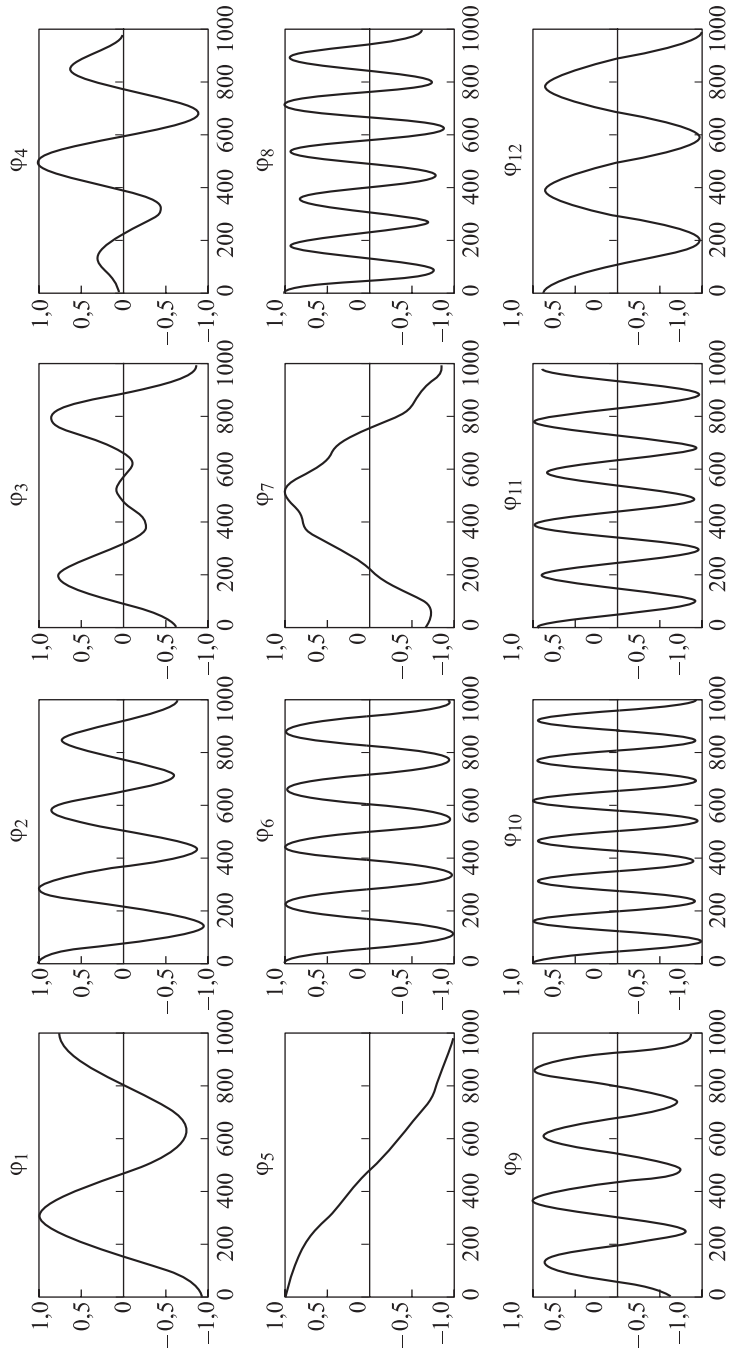
Рис. 4. Модули 100 первых собственных чисел матрицы ковариаций для модельной задачи (а) и оценка относительной погрешности воспроизведения решения (б) для различного числа базисных векторов (в двойном логарифмическом масштабе)

Приведенная выше оценка точности метода *POD* позволяет заранее оценить число базисных векторов $\bar{\varphi}_i$, необходимых для качественного воспроизведения решения. Значения ε для различного числа m базисных векторов приведены на рис. 4, б.

Учет первых трех базисных векторов в разложении *POD* практически не ведет к снижению погрешности аппроксимации решения; добавление следующих 10 векторов дает резкое повышение точности (возрастает кубически с ростом числа базисных векторов); добавление в базис последующих векторов $\bar{\varphi}_i$ приводит к существенно более медленному (линейному по m) убыванию погрешности. При использовании 12 базисных функций оценка погрешности *POD*-аппроксимации дает значение $\varepsilon \approx 0,0056$, что соответствует погрешности около 0,5 %. Графики 12 первых базисных векторов (форм) приведены на рис. 5, а.

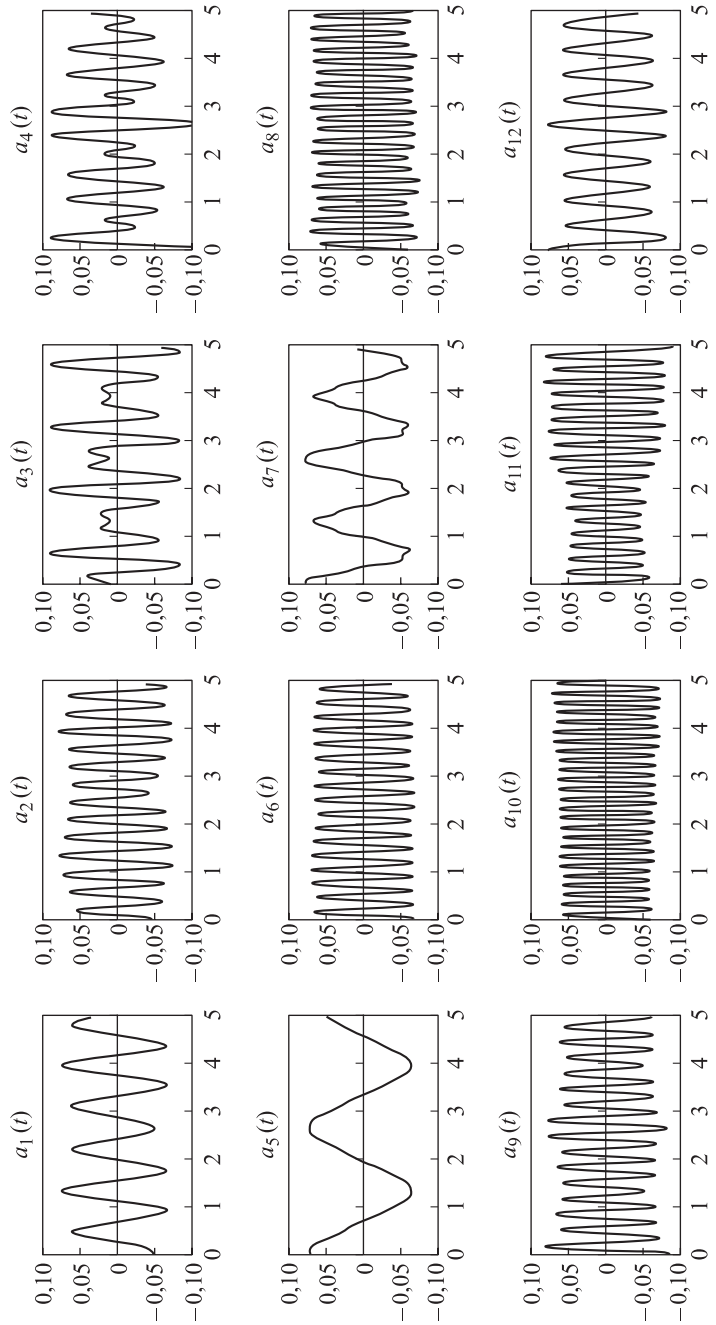
Коэффициенты $a_k(t_i)$ при базисных векторах представляют собой числовые последовательности, содержащие $N = 500$ значений. Графики коэффициентов при первых 12 базисных векторах (формах) показаны на рис. 5, б.

Непосредственное хранение значений коэффициентов a_k во все расчетные моменты времени t_i , $i = 1, \dots, N$, нецелесообразно. Для их «сжатия» можно предложить использование дискретного преобразования Фурье. Для каждой последовательности a_k вычисляют модули коэффициентов при гармониках, задают некоторый уровень «отсечки», к примеру, на уровне 0,01 или 0,001 от величины наибольшего по модулю коэффициента, и все меньшие коэффициенты Фурье обнуляются. Таким образом, необходимо хранить лишь значения ненулевых (комплексных) коэффициентов, а также их номера. Общий объем данных, которые требуется хранить, составляет в рассматриваемой задаче $(m+1)n + \sum_{k=1}^m 2,5p_k$, где p_k — число ненулевых коэффициентов с учетом того, что для хранения целых чисел — номеров коэффициентов — в общем случае требуется 4 байта.



a

Рис. 5 (начало). Первые 12 базисных векторов (форм) в разложении *POD*-решения (а) (для наглядности формы показаны нормированными на единицу)



б

Рис. 5 (окончание). Коэффициенты при первых 12 базисных векторах (формах) (б) в разложении POD-решения

Применительно к этой задаче эффективность сжатия решения с помощью метода *POD* составляет 31,2 раза при уровне «отсечки» 0,0001 и 35,6 раза при уровне 0,001. При этом наибольшая относительная погрешность восстановления решения во все моменты времени в шаровой (евклидовой) норме составляет 0,8 и 3,4 % соответственно.

Отметим, что процедуру «сжатия» коэффициентов при формах *POD* также можно проводить с помощью вейвлет-преобразования [11]; при удачном выборе вейвлета она работает эффективнее по сравнению с преобразованием Фурье. Вместе с тем такое преобразование актуально, как правило, лишь при решении одномерных задач. В дву- и трехмерных задачах объем памяти, требуемой для непосредственного хранения коэффициентов при «существенных» формах *POD*, значительно меньше, чем для хранения самих форм. При этом нет необходимости выполнять обратное преобразование Фурье, что облегчает реконструкцию решения.

Использование метода *POD* при решении двумерной задачи. Рассмотрим задачу, аналогичную описанной во введении, применительно к моделированию акустической нагрузки при старте РН. Примем, что контрольная поверхность содержит 200 000 ячеек, время моделирования составляет 20 с, шаг сохранения результатов в виде значений давления в ячейках сетки на контрольной поверхности — $\Delta t = 0,004$ с. Задача решалась в пакете *OpenFOAM* гибридным методом [8, 9]. Результаты выполнения метода *POD* в виде модулей первых 100 собственных чисел и оценки погрешности с учетом различного числа форм показаны на рис. 6.

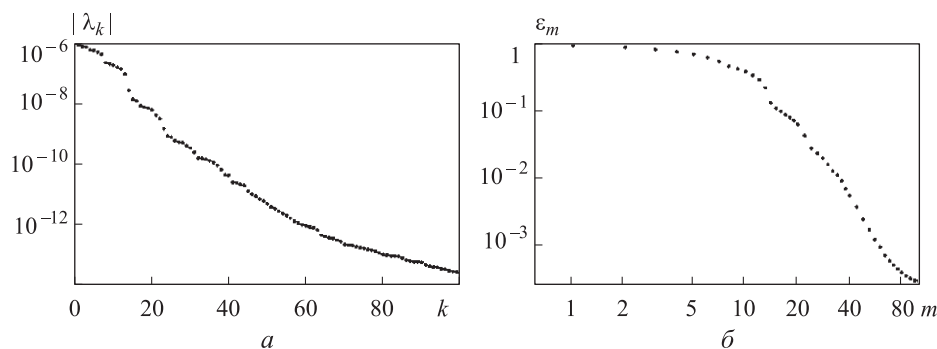


Рис. 6. Модули первых 100 собственных чисел (а) и оценка относительной погрешности аппроксимации решения с учетом различного числа форм (б)

Опираясь на результаты, полученные при решении одномерной задачи, целесообразно ограничиться учетом первых 30 форм, оценка относительной погрешности составляет при этом $\varepsilon \approx 0,02$. Для выполнения преобразования *POD* и сохранения первых 30 форм с помощью разработанного программного обеспечения [12] с использованием фреймворка *Apache Spark* на персональной ЭВМ с двухядерным процессором и 8 Гб оперативной памяти требуется 62 мин, что существенно меньше времени выполнения расчета (11 ч 17 мин на 10 ядрах кла-

стера *HP BL2×220c G7*). Дополнительное «сжатие» коэффициентов при формах с помощью преобразования Фурье нецелесообразно. При этом объем данных, подлежащих хранению, после преобразования *POD* сократился с 8 Гб до примерно 50 Мб, т. е. в 160 раз, что дает возможность эффективно их обрабатывать и использовать для решения задачи распространения акустических волн в дальнем поле.

Заключение. Исследована эффективность применения технологии обработки и сжатия данных *POD* при численном моделировании процессов гидродинамики и акустики в рамках решения практических задач большой размерности. Показано, что результаты расчетов можно эффективно сжимать, обеспечивая приемлемую погрешность восстановления решения порядка 1 %, сократив при этом объем сохраняемых данных в 30 раз в одномерном случае и в 160 раз в двумерном.

ЛИТЕРАТУРА

1. *Bieder U., Fauchet G., Calvin C.* High performance Large Eddy Simulation of turbulent flows around PWR mixing grids // The 16th IEEE Int. Conf. on High Performance Computing and Communications. Paris, 2014. 5 p.
2. *Quarteroni A., Rozza G.* Reduced order methods for modeling and computational reduction. Springer, 2014. 334 p.
3. *Holmes P., Lumley J.L., Berkooz G., Rowley C.W.* Turbulence, coherent structures, dynamical systems and symmetry. Cambridge University Press, 2012. 402 p.
4. *Kerschen G., Golinval J.-C., Vakakis A.F., Bergman L.A.* The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview // *Nonlinear Dynamics*. 2005. Vol. 41. Iss. 1-3. P. 147–169. DOI: 10.1007/s11071-005-2803-2
5. *Schmid P.J.* Dynamic mode decomposition of numerical and experimental data // *J. Fluid Mech.* 2010. Vol. 656. P. 5–28.
6. *Tosh A., Liever P., Owens F., Liu Y.A.* High-fidelity CFD/BEM methodology for launch pad acoustic environment prediction // 18th AIAA/CEAS Aeroacoustics Conf. 2012. Colorado Springs, Paper AIAA 2012–2107.
7. *Кудимов Н.Ф., Сафронов А.В., Третьякова О.Н.* Прикладные задачи газодинамики и теплообмена в энергетических установках ракетной техники. М.: Изд-во МАИ, 2014. 168 с.
8. *Seiner J.M., Norum T.D.* Experiments of shock associated noise on supersonic jets // AIAA 12th Fluid and Plasma Dynamics Conf. Williamsburg, Virginia, 1979. P. 79–152. DOI: 10.2514/6.1979-1526 URL: <https://arc.aiaa.org/doi/10.2514/6.1979-1526>
9. *Kraposhin M., Bovtrikova A., Strijhak S.* Adaptation of Kurganov — Tadmor numerical scheme for applying in combination with the PISO method in numerical simulation of flows in a wide range of Mach numbers // *Procedia Computer Science*. 2015. Vol. 66. P. 43–52. DOI: 10.1016/j.procs.2015.11.007
10. *Menter F.R., Kuntz M., Langtry R.* Ten years of industrial experience with the SST turbulence model // *Turbulence, Heat and Mass Transfer*. 2003. No. 4. P. 625–632.
11. *Добешин И.* Десять лекций по вейвлетам. Ижевск: НИЦ «РХД», 2001. 464 с.

12. *Kalugin M.D., Strijhak S.V.* Implementation of POD and DMD methods in Apache Spark framework for simulation of unsteady turbulent flow in the model combustor // Proc. 7th European Congress on Computational Methods in Applied Sciences and Engineering. Crete, Greece, 2016. P. 857–864.

Калугин Михаил Дмитриевич — научный сотрудник Института системного программирования им. В.П. Иванникова РАН (Российская Федерация, 109004, Москва, ул. Александра Солженицына, д. 25).

Корчагова Виктория Николаевна — старший лаборант Института системного программирования им. В.П. Иванникова РАН (Российская Федерация, 109004, Москва, ул. Александра Солженицына, д. 25), ассистент кафедры «Прикладная математика» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Крапошин Матвей Викторович — канд. физ.-мат. наук, старший научный сотрудник Института системного программирования им. В.П. Иванникова РАН (Российская Федерация, 109004, Москва, ул. Александра Солженицына, д. 25).

Марчевский Илья Константинович — канд. физ.-мат. наук, старший научный сотрудник Института системного программирования им. В.П. Иванникова РАН (Российская Федерация, 109004, Москва, ул. Александра Солженицына, д. 25), доцент кафедры «Прикладная математика» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Морева Виктория Сергеевна — канд. физ.-мат. наук, доцент кафедры «Прикладная математика» МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1).

Просьба ссылаться на эту статью следующим образом:

Использование инструментов анализа больших данных при решении задач газовой динамики и акустики / М.Д. Калугин, В.Н. Корчагова, М.В. Крапошин, И.К. Марчевский, В.С. Морева // Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки. 2018. № 3. С. 32–47. DOI: 10.18698/1812-3368-2018-3-32-47

USING BIG ANALYTICS TOOLS IN PERFORMANCE OF GAS DYNAMICS AND ACOUSTICS TASKS

M.D. Kalugin¹

V.N. Korchagova^{1,2}

M.V. Kraposhin¹

I.K. Marchevsky^{1,2}

V.S. Moreva²

iliamarchevsky@bmstu.ru

¹ **Ivannikov Institute for System Programming, Russian Academy of Sciences, Moscow, Russian Federation**

² **Bauman Moscow State Technical University, Moscow, Russian Federation**

Abstract

The paper centers around big data processing and analytics, as well as big data compression when doing numerical simulations of hydrodynamic and acoustic processes in large industrial applications. One of the typical cases is the simulation of supersonic turbulent jets and high-intensity acoustic loads in rocket lift-off process. We give some estimates of the memory amount for correct simulations of gas dynamic and acoustic processes in this case. Moreover, we describe one of the most common technologies of big data compression and analytics — *Proper Orthogonal Decomposition*. This technology was implemented using the *Apache Spark* framework for large-scale data processing. The necessary memory for the storing of the results of simulation of acoustic pressure propagation can be reduced by 30 times for 1D case and 160 times for 2D case

Keywords

Big data processing, Proper Orthogonal Decomposition, gas dynamics, acoustics

Received 10.08.2017

© BMSTU, 2018

The work was supported by the Ministry of Education and Science of Russia (identifier RFMEFI60714X0090, agreement no. 14.607.21.0090)

REFERENCES

- [1] Bieder U., Fauchet G., Calvin C. High performance Large Eddy Simulation of turbulent flows around PWR mixing grids. *The 16th IEEE Int. Conf. on High Performance Computing and Communications*. Paris, 2014. 5 p.
- [2] Quarteroni A., Rozza G. *Reduced order methods for modeling and computational reduction*. Springer, 2014. 334 p.
- [3] Holmes P., Lumley J.L., Berkooz G., Rowley C.W. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge University Press, 2012. 402 p.
- [4] Kerschen G., Golinval J.-C., Vakakis A.F., Bergman L.A. The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear Dynamics*, 2005, vol. 41, iss. 1-3, pp. 147–169. DOI: 10.1007/s11071-005-2803-2
- [5] Schmid P.J. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.*, 2010, vol. 656, pp. 5–28.
- [6] Tosh A., Liever P., Owens F., Liu Y.A. High-fidelity CFD/BEM methodology for launch pad acoustic environment prediction. *18th AIAA/CEAS Aeroacoustics Conf.* 2012. Colorado Springs, Paper AIAA 2012–2107.
- [7] Kudimov N.F., Safronov A.V., Tret'yakova O.N. *Prikladnye zadachi gazodinamiki i teploobmena v energeticheskikh ustanovkakh raketnoy tekhniki* [Applied problems of gas dynamics and heat exchange in energetic plants of rocket technique]. Moscow, MAI Publ., 2014. 168 p.
- [8] Seiner J.M., Norum T.D. Experiments of shock associated noise on supersonic jets. *AIAA 12th Fluid and Plasma Dynamics Conf.* Williamsburg, Virginia, 1979, pp. 79–152. DOI: 10.2514/6.1979-1526 Available at: <https://arc.aiaa.org/doi/10.2514/6.1979-1526>
- [9] Kraposhin M., Bovtrikova A., Strijhak S. Adaptation of Kurganov — Tadmor numerical scheme for applying in combination with the PISO method in numerical simulation of flows in a wide range of Mach numbers. *Procedia Computer Science*, 2015, vol. 66, pp. 43–52. DOI: 10.1016/j.procs.2015.11.007

[10] Menter F.R., Kuntz M., Langtry R. Ten years of industrial experience with the SST turbulence model. *Turbulence, Heat and Mass Transfer*, 2003, no. 4, pp. 625–632.

[11] Daubechies I. Ten lectures on wavelets. SIAM, 1992. 377 p.

[12] Kalugin M.D., Strijhak S.V. Implementation of POD and DMD methods in Apache Spark framework for simulation of unsteady turbulent flow in the model combustor. *Proc. 7th European Congress on Computational Methods in Applied Sciences and Engineering*. Crete, Greece, 2016, pp. 857–864.

Kalugin M.D. — Research Fellow, Ivannikov Institute for System Programming, Russian Academy of Sciences (Alexander Solzhenitsyn ul. 25, Moscow, 109004 Russian Federation).

Korchagova V.N. — Senior Lab Technician, Ivannikov Institute for System Programming, Russian Academy of Sciences (Alexander Solzhenitsyn ul. 25, Moscow, 109004 Russian Federation), Assistant, Department of Applied Mathematics, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Kraposhin M.V. — Cand. Sc. (Phys.-Math.), Senior Research Fellow, Ivannikov Institute for System Programming, Russian Academy of Sciences (Alexander Solzhenitsyn ul. 25, Moscow, 109004 Russian Federation).

Marchevsky I.K. — Cand. Sc. (Phys.-Math.), Leading Research Fellow, Ivannikov Institute for System Programming, Russian Academy of Sciences (Alexander Solzhenitsyn ul. 25, Moscow, 109004 Russian Federation), Assoc. Professor, Department of Applied Mathematics, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Moreva V.S. — Cand. Sc. (Phys.-Math.), Assoc. Professor, Department of Applied Mathematics, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, str. 1, Moscow, 105005 Russian Federation).

Please cite this article in English as:

Kalugin M.D., Korchagova V.N., Kraposhin M.V., Marchevsky I.K., Moreva V.S. Using Big Analytics Tools in Performance of Gas Dynamics and Acoustics Tasks. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Estestv. Nauki* [Herald of the Bauman Moscow State Tech. Univ., Nat. Sci.], 2018, no. 3, pp. 32–47 (in Russ.). DOI: 10.18698/1812-3368-2018-3-32-47