Исследование вариационных задач для квадратичных функционалов: доказательный вычислительный эксперимент

В. А. Шишкин
Пермский государственный университет
e-mail: vsh1791@mail.ru

Конструктивный подход к исследованию вариационных задач для квадратичных функционалов основан на замене исходной задачи некоторой близкой к ней более простой, позволяющей использовать как фундаментальные положения общей теории, так и возможности современных вычислительных систем. Сначала делается попытка достоверно установить существование решения и только после этого (если доказана разрешимость исходной задачи) строится приближённое решение с гарантированной оценкой точности, для чего применяются методы доказательных вычислений, основанные на использовании арифметики рациональных и интервальных чисел.

1. Постановка задачи

Рассмотрим задачу минимизации квадратичного функционала

$$\mathcal{I}(x) = \sum_{i=1}^{N} \langle T_{1i}x, T_{2i}x\rangle_{\mathbf{H}} + \langle F_0, x\rangle_{\mathbf{X}} \tag{1a}$$

при ограничениях

$$p_i(x) = \langle \pi_i, x\rangle_{\mathbf{X}} - \alpha_i = 0, \quad i = 1, \ldots, n_1, \tag{1b}$$

$$q_i(x) = \langle \varkappa_i, x\rangle_{\mathbf{X}} - \beta_i \leqslant 0, \quad i = 1, \ldots, n_2. \tag{1c}$$

Здесь T_{1i}, T_{2i}, $i = 1, \ldots, N$, — линейные ограниченные операторы, действующие из заданного банахова пространства $\mathbf{X}$ в вещественное сепарабельное гильбертово пространство $\mathbf{H}$; $\langle f, x\rangle_{\mathbf{X}}$ — значение линейного ограниченного функционала $f \in \mathbf{X}^*$. Предполагаем, что система ограничений совместна и невырождена.

В качестве операторов T_{1i}, T_{2i}, $i = 1, \ldots, N$, могут использоваться тождественные операторы, дифференциальные операторы любого порядка, интегральные операторы, а также операторы сосредоточенного и распределённого отклонения аргумента.

Предположим, что на основе некоторой однозначно разрешимой краевой задачи

$$\delta x = z, \qquad \langle \pi_i, x\rangle_{\mathbf{X}} = \alpha_i, \quad i = 1, \ldots, m, \qquad m \leqslant n_1, \tag{2}$$

можно построить изоморфизм $\mathbf{X} \simeq \mathbf{H} \times \mathbb{R}^m$. Тогда с помощью метода редукции [1, стр. 182] задача (1a)–(1c) сводится к задаче минимизации в гильбертовом пространстве

H:

$$\mathcal{I}_1(z) = \frac{1}{2}\langle Qz, z\rangle_{\mathbf{H}} + \langle f_0, z\rangle_{\mathbf{H}} + \psi_0, \tag{3a}$$

$$g_i(z) = \langle \gamma_i, z\rangle_{\mathbf{H}} - a_i = 0, \quad i = 1, \ldots, n_1 - m, \tag{3b}$$

$$h_i(z) = \langle \eta_i, z\rangle_{\mathbf{H}} - b_i \leqslant 0, \quad i = 1, \ldots, n_2. \tag{3c}$$

$x = \Lambda z + Y\alpha'$ — решение задачи (2) ($\alpha' = \operatorname{col}\{\alpha_1, \ldots, \alpha_m\}$). Заметим, что оператор Q является самосопряжённым по построению:

$$Q = \sum_{i=1}^{N} \Lambda^*(T_{1i}^* T_{2i} + T_{2i}^* T_{1i})\Lambda$$

При решении гладких оптимизационных задач с ограничениями в виде равенств и неравенств можно использовать метод множителей Лагранжа [2, стр. 252]. Функция Лагранжа задачи (3a)–(3c) имеет вид

$$\mathcal{L}(z, \lambda_1, \lambda_2) = \mathcal{I}_1(z) + \langle \lambda_1, g(z)\rangle_{\mathbb{R}^{n_1-m}} + \langle \lambda_2, h(z)\rangle_{\mathbb{R}^{n_2}} = \frac{1}{2}\langle Qz, z\rangle_{\mathbf{H}} - \langle f(\lambda), z\rangle_{\mathbf{H}} + \psi(\lambda). \tag{4}$$

Известно [2, стр. 252-253], что если $\hat{z}$ доставляет (локальный) минимум в задаче (3a)–(3c), то найдутся такие не равные одновременно нулю множители Лагранжа $\hat{\lambda}_1, \hat{\lambda}_2$, что будут выполнены

1. условие стационарности функции Лагранжа по z: решение задачи $\hat{z}$ должно быть корнем уравнения
$$Qz = f(\hat{\lambda}); \tag{5}$$
2. условие согласования знаков: $\hat{\lambda}_2 \geqslant 0$;
3. условие дополняющей нежёсткости
$$\hat{\lambda}_{2i} h_i(\hat{z}) = 0, \quad i = 1, \ldots, n_2. \tag{6}$$

Предположим, что оператор Q можно записать в виде разности двух самосопряжённых операторов $A_1 - A_2$, где оператор A_1 имеет ограниченный обратный, а A_2 является компактным оператором. Тогда уравнение (5) можно переписать в виде

$$z - Kz = y(\lambda), \tag{7}$$

где оператор $K = A_1^{-1}A_2$ — компактный и самосопряжённый; $y(\lambda) = A_1^{-1}f(\lambda)$.

Пусть $\hat{z}$ и $(\hat{\lambda}_1, \hat{\lambda}_2)$ удовлетворяют необходимым условиям существования решения. Тогда достаточным условием разрешимости задачи (3a)–(3c) будет положительная определённость оператора Q [2, стр. 293–294]. Известно, что самосопряжённый оператор Q положительно определён тогда и только тогда, когда его спектр содержит только положительные числа: $\sigma(Q) \subset (0, \infty)$. Так как, по предположению, $Q = I - K$, то для проверки положительной определённости требуется оценить верхнюю границу спектра оператора K:

$$\max \sigma(K) < 1 \quad \Leftrightarrow \quad Q > 0. \tag{8}$$

Таким образом, *если разрешимо параметрическое уравнение* (7) *и верхняя граница спектра оператора K меньше единицы, то задача* (1a)–(1c) *имеет решение*

$$x = \Lambda z + Y\alpha'.$$

2. Конструктивное исследование условий разрешимости

Даже в простых задачах вид оператора K обычно является слишком сложным, чтобы можно было получить точное решение. Применение *конструктивной математики* [6, стр. 561] позволяет привести исходную задачу к виду, подходящему для использования доказательных вычислений.

Рассмотрим случай, когда $\mathbf{H} = \mathbf{L}_2[a, b]$ — пространство суммируемых на $[a, b]$ с квадратом функций, K — интегральный оператор Гильберта–Шмидта, и, следовательно, (7) — интегральное уравнение Фредгольма второго рода с симметричным ядром и правой частью, зависящей от параметра.

Заменим уравнение

$$z(t) - \int_a^b K(t, s)z(s)\,ds = f(t), \qquad t \in [a, b] \tag{9}$$

близким к нему интегральным уравнением с вырожденным ядром

$$z(t) - \int_{\tilde{a}}^{\tilde{b}} \tilde{K}(t, s)z(s)\,ds = \tilde{f}(t), \qquad t \in [\tilde{a}, \tilde{b}], \tag{10}$$

где $a \leqslant \tilde{a}$, $\tilde{b} \leqslant b$ и $\tilde{a}, \tilde{b} \in \mathbb{Q}$ — рациональные числа; $\tilde{K}\colon \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ и $\tilde{f}\colon \mathbb{Q} \to \mathbb{Q}$.

Для повышения точности приближения можно разбить $[\tilde{a}, \tilde{b}]$ сеткой, узлы которой тоже являются рациональными числами, и строить приближения в виде сумм финитных функций — аппроксимаций на заданных подобластях.

Заменим $K(t, s)$ и $f(t)$ отрезками ряда Фурье по заданной системе ортогональных функций. В качестве базиса аппроксимации удобно использовать ортогональные многочлены Лежандра или систему ортогональных функций Радемахера–Уолша, так как элементы этих систем попарно ортогональны с единичным весом, а также коэффициенты многочленов Лежандра и узлы перемены знака функций Радемахера–Уолша являются рациональными числами, что пригодится при программной реализации метода. Коэффициенты рядов Фурье для функций $K(t, s)$ и $f(t)$, вообще говоря, могут быть иррациональными числами, поэтому в ходе аппроксимации они заменяются приближёнными рациональными значениями.

Для каждой из областей, на которых вычисляются приближения для ядра и правой части, также оцениваются сверху абсолютные погрешности $|K(t, s) - \tilde{K}(t, s)|$ и $|f(t) - \tilde{f}(t)|$. Определение точных верхних граней является, вообще говоря, достаточно сложной задачей. Однако для наших целей достаточно использовать относительно грубые оценки. Главное, чтобы эти оценки были *гарантированы*.

Для решения интегрального уравнения с вырожденным ядром можно использовать стандартные методы, сводящиеся к решению системы линейных алгебраических уравнений $AZ = B$. Заметим, что всегда можно построить такой приближённый оператор $\tilde{K}$, чтобы уравнение (10) имело решение, но существование обратного оператора для $I - \tilde{K}$ ещё не гарантирует обратимости $I - K$. Для проверки существования будем использовать теорему об обратном операторе [3, стр. 212]: если оператор $I - \tilde{K}$ имеет обратный, то обратимы и все операторы, для которых выполняется неравенство

$$\|K - \tilde{K}\| < \frac{1}{\left\|(I - \tilde{K})^{-1}\right\|}.$$

Норма оператора $(I - \tilde{K})^{-1}$ может быть вычислена точно, а норму в левой части неравенства можно гарантированно оценить сверху. Если оператор $I - K$ обратим, то (теоретически) всегда можно построить настолько точное приближение, чтобы неравенство было выполнено.

Для оценки верхней границы спектра оператора K можно использовать теорему Вейля [5, стр. 257]: соответствующие собственные числа двух компактных самосопряжённых операторов K и $\tilde{K}$ отстоят друг от друга не более чем на $\|K - \tilde{K}\|$. Спектральное множество оператора $\tilde{K}$ совпадает с множеством собственных чисел матрицы $\tilde{A} = E - A$ (E — единичная матрица). По построению элементы матрицы A являются рациональными числами, следовательно, коэффициенты характеристического многочлена $\det(\tilde{A} - \nu E)$ также будут рациональными числами. Для оценки значения наибольшего корня можно использовать решение уравнения методом Ньютона, при котором последовательность приближений монотонно сходится к решению сверху.

3. Компьютерная реализация

3.1. Доказательные вычисления

Для того, чтобы результаты исследования были *доказательны*, требуется гарантировать точность результатов компьютерных вычислений. Стандартные средства не позволяют контролировать точность вычислений, поэтому для расчётов используются пакеты программ, реализующие интервальную арифметику, а также арифметику рациональных чисел.

Так как программная реализация рассмотренного подхода писалась для выполнения под управлением операционной системы `GNU/Linux`, то для проведения рациональных вычислений была выбрана библиотека программ `GNU MP`[1]. Существенно упростила написание кода на `C++` возможность использования классов–обёрток из `<gmpxx.h>`.

Интервальные вычисления, в отличие от арифметики рациональных чисел, могут быть реализованы как чисто программно, так и с использованием арифметических операций с аппаратно реализованным направленным округлением. Такие возможности предоставляют, в частности, современные процессоры фирм Intel и AMD (при отключенной поддержке потоковых вычислений `SSE2`). Для выполнения интервальных вычислений можно применять шаблонный класс `interval<T>` из библиотеки `boost`[2]. Простую реализацию интервальной арифметики можно написать и самостоятельно, используя `<fenv.h>`.

Известно, что выполнение последовательности рациональных вычислений приводит к быстрому росту длин числителя и знаменателя результата. Пусть, например, требуется оценить сверху наибольшее собственное число симметричной $N \times N$-матрицы, элементы которой представлены рациональными числами. Сначала с помощью подходящего численного метода вычисляются (точно!) коэффициенты характеристического многочлена

$$p(\sigma) = (-1)^N(\sigma^N - q_1\sigma^{N-1} - \cdots - q_N).$$

Выбрав в качестве начального значения $\sigma^{(0)} = 1 + \max_i |q_i|$ [4, стр. 10, теорема 1.2],

[1] `http://gmplib.org`
[2] `http://www.boost.org`

будем использовать для решения метод Ньютона

$$\sigma^{(k+1)} = \sigma^{(k)} - \frac{p(\sigma^{(k)})}{p'(\sigma^{(k)})}. \tag{11}$$

Проверка показывает, что даже при относительно небольших (десятки) размерностях матрицы вычисления вследствие «разрастания» числителя и знаменателя $\sigma^{(k)}$ выполняются *очень* медленно.

Возможным решением является комбинирование рациональных и интервальных вычислений: на каждой итерации значение (11) вычисляется с помощью рациональной арифметики, затем результат приближается интервальным числом с границами в виде чисел с плавающей точкой (пакет `mpfr`[3] — вычисления с плавающей точкой, с многократной точностью и корректным округлением), после чего верхняя граница интервала снова преобразуется в рациональное число. При этом уменьшение точности результата на каждой итерации компенсируется высокой скоростью работы с рациональными числами, числитель и знаменатель которых имеют небольшую длину.

3.2. Параллельные вычисления

Параллельные вычисления могут существенно снизить время решения задачи. В рассматриваемом подходе очевидными кандидатами на распараллеливание являются

- аппроксимация функций на подобластях;
- операции линейной алгебры;
- оценка нормы резольвентного оператора.

Для программной реализации алгоритмов, использующих параллельные вычисления, применялась библиотека `Open MPI`[4]. При программировании адаптивных алгоритмов численного интегрирования также использовалась библиотека `pthread` (стандарт `POSIX Threads`). В процессе вычислений количество потоков соответствует числу процессорных ядер (настраивается автоматически в начале работы).

После выпуска фирмой nVidia видеокарт с архитектурой Fermi, аппаратно поддерживающих вычисления с числами с плавающей запятой двойной точности, появилась возможность использовать для расчётов технологию CUDA. К сожалению, похоже, что в процессорах видеокарт GeForce 4xx и 5xx нет аппаратной реализации вычислений с направленным округлением. Однако массивное распараллеливание, возможно, компенсирует затраты на программно реализованную интервальную арифметику.

Список литературы

[1] Азбелев Н. В., Максимов В. П., Рахматуллина Л. Ф. Элементы современной теории функционально–дифференциальных уравнений. Методы и приложения. М.: Институт компьютерных исследований, 2002. 384 с.

[2] Алексеев В. М., Тихомиров В. М., Фомин С. В. Оптимальное управление. М.: Наука, Гл. ред. физ.-мат. лит., 1979. 432 с.

[3] `http://www.mpfr.org`

[4] `http://www.open-mpi.org`

[3] Канторович Л. В., Акилов Г. П. Функциональный анализ. М.: Наука, Гл. ред. физ.-мат. лит., 1977. 744 с.

[4] Прасолов В. В. Многочлены. М.: МЦНМО, 2001. 336 с.

[5] Рисс Ф., Сёкефальви-Надь Б. Лекции по функциональному анализу. М.: Мир, 1979. 569 с.

[6] Kushner B. A. The constructive mathematics of A. A. Markov // The Mathematical Association of America. Monthly. Vol. 113, no. 6. Pp. 559–566.