

О ПРИБЛИЖЕННОМ РЕШЕНИИ ПЛОХО ОБУСЛОВЛЕННЫХ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

В данной статье приводятся алгоритмы двух методов решения плохо обусловленных систем линейных алгебраических уравнений (метода сингулярного разложения и метода регуляризации) и делается сравнительный анализ их на решении конкретного примера.

Во многих прикладных задачах решение дифференциальных, интегральных и интегро–дифференциальных уравнений сводится к системе линейных алгебраических уравнений (СЛАУ), которая часто оказывается плохо обусловленной. Решение таких систем сопряжено с определенными трудностями.

Пусть имеем плохо обусловленную СЛАУ

$$Ax = f, \quad (1)$$

где

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad f = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}. \quad (2)$$

В прикладных задачах значения элементов матрицы A и столбца свободных членов f обычно задаются приближенно, так как эти значения получаются после обработки данных измерений и округления в процессе вычислений. Когда система является переопределенной ($m > n$), или вырожденной ($\det A = 0$), или близкой к вырожденной ($\det A \approx 0$), то при малых возмущениях матриц A и f ошибка в решении сильно возрастает, т.е. система оказывается неустойчивой. Если рассматривается приближенно заданная система (\tilde{A}, \tilde{f}) , то любая система (A, f) , удовлетворяющая условию

$$\|\tilde{A} - A\| \leq \Delta, \quad \|\tilde{f} - f\| \leq \delta,$$

эквивалентна данной системе по точности. Поэтому мы должны рассматривать всю совокупность систем, эквивалентных в этом смысле.

Решение любой системы (A, f) из этой совокупности является допустимым решением рассматриваемой задачи. Если вырожденные системы имеют решения, то они могут сильно отличаться по норме. Поэтому возникает необходимость ввести в постановку задачи дополнительное условие отбора допустимых решений. В качестве такого условия можно использовать наряду с условием минимума невязки $\|Ax - f\|$ решения системы (1) также условие минимума нормы вектора x .

Мы рассмотрим два метода решения плохо обусловленных систем. Первый из них основан на алгоритме сингулярного разложения матрицы системы и называется SVD–метод [2].

Для любой матрицы A и любых двух ортогональных матриц U и V рассматривается матрица Σ , определяемая соотношением

$$\Sigma = U^T \cdot A \cdot V. \quad (3)$$

Если u_j и v_j являются столбцами матриц U и V соответственно, то компоненты матрицы Σ равны

$$\sigma_{ij} = u_i^T \cdot A \cdot v_j.$$

За сингулярным разложением скрывается та идея, что надлежащим выбором матриц U и V можно обратить большинство элементов σ_{ij} в нуль; более того, можно даже

сделать матрицу Σ диагональной с неотрицательными диагональными элементами. Исходя из этого, сингулярным разложением действительной $m \times n$ – матрицы A считается всякая ее факторизация вида $A = U \cdot \Sigma \cdot V^T$, где U и V – ортогональные $m \times m$ – и $n \times n$ – матрицы соответственно; Σ – диагональная $m \times n$ – матрица, у которой $\sigma_{ij} = 0$ при $i \neq j$ и $\sigma_{ii} = \sigma_i \geq 0$. Величины σ_i называются сингулярными числами матрицы A , а столбцы матрицы U и V – левыми и правыми сингулярными векторами. Цель алгоритма состоит в том, чтобы найти ортогональные матрицы U и V такие, что матрица Σ из (3) была диагональной.

Алгоритм SVD состоит из двух этапов. Первый этап заключается в приведении матрицы A к двухдиагональной форме, т.е. к матрице, у которой ненулевыми могут быть лишь элементы диагонали и первой наддиагонали. Второй этап – это итерационный процесс, в котором наддиагональные элементы уменьшаются до пренебрежимо малой величины, что дает желаемую диагональную матрицу.

На первом этапе сначала введем нули в первый столбец a_1 матрицы A . Полученную матрицу обозначим A_1 и возьмем вектор u_1 , который получается из a_1 прибавлением к первой компоненте числа

$$\|a_1\| = (a_1^T a_1)^{\frac{1}{2}}.$$

Тогда

$$u_1 = \begin{pmatrix} a_{11} + \|a_1\| \\ a_{21} \\ \dots \\ a_{n1} \end{pmatrix}.$$

Далее вычисляем

$$\beta_1 = \frac{\|u_1\|^2}{2}, \quad U_1 = E - \beta_1^{-1} u_1 u_1^T, \quad A_2 = U_1 A_1,$$

где E – единичная матрица. Преобразованная матрица A_2 вычисляется столбец за столбцом:

$$U_1 a_1 = (E - \beta_1^{-1} u_1 u_1^T) a_1 = a_1 - \beta_1^{-1} u_1 u_1^T a_1 = a_1 - u_1 \quad (\text{так как } \beta_1^{-1} u_1^T a_1 = 1).$$

При таком преобразовании в первый столбец вводятся желаемые нули и его длина не изменится. Вычисляем остальные столбцы:

$$U_1 a_2 = a_2 - \beta_1^{-1} u_1 u_1^T a_2,$$

$$U_1 a_3 = a_3 - \beta_1^{-1} u_1 u_1^T a_3,$$

... ..

Столбцы A_2 получаются вычитанием кратных вектора u_1 из столбцов A_1 . Эти кратные порождаются скалярными произведениями, а не отдельными элементами, как в гауссовском исключении.

Ортогональность преобразующей матрицы U_1 проверяется непосредственно:

$$\begin{aligned} U_1^T U_1 &= (E - \beta_1^{-1} u_1 u_1^T)^T (E - \beta_1^{-1} u_1 u_1^T) = E - 2\beta_1^{-1} u_1 u_1^T + \beta_1^{-2} u_1 u_1^T u_1 u_1^T = \\ &= E - 2\beta_1^{-1} u_1 u_1^T + 2\beta_1^{-1} u_1 u_1^T = E. \end{aligned}$$

Ортогональность матрицы U_1 гарантирует, что длина каждого столбца в U_1 равна длине соответствующего столбца в A_1 .

Прежде чем вводить дальнейшие нули под диагональю, преобразованием вида

$$A_3 = A_2 V_1$$

получают нули в первой строке. Нули, уже стоящие в первом столбце, не должны быть испорчены, а длину первой строки нужно сохранить. На этом шаге в первой строке

должно быть получено $n-2$ нуля.

Преобразование порождается первой строкой A_2 . Последовательно вычисляются величины:

$$\gamma_2 = \frac{\|v_1\|^2}{2},$$

$$V_1 = E - \gamma_1^{-1} v_1 v_1^T,$$

$$A_3 = A_2 V_1.$$

Строка матрицы A_3 с номером i получается по формуле:

$$a_i^T V_1 = a_i^T (E - \gamma_1^{-1} v_1 v_1^T) = a_i^T - (\gamma_1^{-1} a_i^T v_1) v_1^T.$$

Таким образом, для получения строки матрицы A_3 следует из соответствующей строки A_2 вычитать подлежащее кратное v_1^T . Поскольку первая компонента v_1^T нулевая, то нули первого столбца A_2 сохраняются в A_3 . Так как V_1 ортогональная, то длина каждой строки в A_3 равна длине соответствующей строки в A_2 .

Далее можно добиться новых нулей под диагональю, не испортив полученных ранее:

$$A_4 = U_2 A_3, \quad A_5 = U_3 A_4, \dots$$

В итоге получается искомая двухдиагональная матрица. Прямое использование ортогональных преобразований не позволяет получить какие-либо новые нули. Для общего порядка n нужно n преобразований U и $n-2$ преобразований V для получения двухдиагональной матрицы.

Второй этап алгоритма *SVD* представляет собой итерационный процесс. Каждый шаг уменьшает величину наддиагональных элементов. В конечном счете эти элементы становятся сравнимыми с ошибками округлений и их можно отбросить. Каждый итерационный шаг начинается с двухдиагональной матрицы, которая преобразуется в другую двухдиагональную матрицу с меньшими внедиагональными элементами. Процесс выполняется посредством операций со строками и столбцами матрицы A , а не фактических матричных умножений. В итоге получается диагональная матрица

$$\sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_m \end{pmatrix},$$

где $\sigma_1, \sigma_2, \dots, \sigma_m$ – сингулярные числа.

Теперь, используя сингулярное разложение матрицы A , линейную систему (1) перепишем в виде

$$U \Sigma V^T x = f,$$

откуда

$$\Sigma z = D, \tag{4}$$

где $z = V^T x$ и $D = U^T f$.

Система уравнений (4) диагональная и она может распадаться на две или три подсистемы в зависимости от значений размерностей m и n и ранга k , т.е. количества ненулевых сингулярных чисел:

$$\sigma_j z_j = d_j, \text{ если } j \leq n \text{ и } \sigma_j \neq 0,$$

$$0 \cdot z_j = d_j, \text{ если } j \leq n \text{ и } \sigma_j = 0,$$

$$0 = d_j, \text{ если } j > n.$$

Вторая подсистема пуста, если $k = n$; третья пуста, если $n = m$. Отсюда следует, что уравнения совместны и решение существует в том и только в том случае, когда $d_j = 0$ при $\sigma_j = 0$ или $j > n$.

Второй метод основан на поиске нормального решения системы (1) с точностью,

соответствующей точности задания элементов матрицы A и столбца f . Нормальным решением системы (1) называется ее решение $x^0 = (x_1^0, x_2^0, \dots, x_n^0)^T$, норма которого является наименьшей среди норм всех решений x этой системы, другими словами, нормальное решение наиболее близко к началу координат.

Поскольку задача решения системы (1) в данном случае является некорректной, решение ищется методом регуляризации А.Н. Тихонова [3], согласно которому нормальное решение системы (1) доставляет минимум следующему функционалу

$$M^\alpha[A, f, x] = \|Ax - f\|^2 + \alpha \|x\|^2, \quad (5)$$

где α – некоторый числовой коэффициент, называемый параметром регуляризации. В подробной записи функционал (5) имеет вид

$$M^\alpha[A, f, x] = \sum_{i=1}^n (\sum_{j=1}^n a_{ij}x_j - b_i)^2 + \alpha \sum_{j=1}^n x_j^2. \quad (6)$$

Можно доказать [4], что при любом фиксированном $\alpha > 0$ неотрицательный функционал (6) достигает своего минимального значения в единственной точке $x^\alpha = (x_1^\alpha, x_2^\alpha, \dots, x_n^\alpha)^T$ пространства E^n .

Функционал $M^\alpha[A, f, x]$ зависит от n переменных x_1, x_2, \dots, x_n . Применяя необходимое условие минимума функции многих переменных $\partial M^\alpha / \partial x_k = 0$, получаем СЛАУ

$$\sum_{i=1}^n (\sum_{j=1}^n a_{ij}x_j - f_i) + \alpha x_k = 0, \quad k = 1, 2, \dots, n$$

или

$$\sum_{j=1}^n A_{kj}x_j = B_k, \quad k = 1, 2, \dots, n, \quad (7)$$

где

$$A_{kj} = \sum_{i=1}^n a_{ik}a_{ij}, \quad k \neq j; \quad A_{kk} = \sum_{i=1}^n a_{ik}^2 + \alpha; \quad B_k = \sum_{i=1}^n a_{ik}f_i.$$

При решении вырожденной системы методом регуляризации получается вектор, отличный от искомого вектора x^0 . Поэтому можно предложить алгоритм, вносящий поправку в полученное решение. Суть алгоритма состоит в минимизации наибольшей невязки. Подставим полученное решение в систему (1) и находим невязки уравнений. Пусть i -ое уравнение имеет наибольшую невязку. Тогда для этого уравнения образуем функционал

$$J^\alpha = \left(\sum_{j=1}^n a_{ij}x_j - f_i \right)^2 + \alpha \sum_{j=1}^n x_j^2$$

и минимизируем его. Получаем систему

$$\frac{\partial J^\alpha}{\partial x_k} = a_{ik} \left(\sum_{j=1}^n a_{ij}x_j - f_i \right) + \alpha x_k = 0, \quad k = 1, 2, \dots, n,$$

или

$$\sum_{j=1}^n A_{kj}x_j = B_k, \quad k = 1, 2, \dots, n, \quad (8)$$

где

$$A_{kj} = a_{ik}a_{ij}, \quad k \neq j; \quad A_{kk} = a_{ik}^2 + \alpha; \quad B_k = a_{ik}f_i.$$

Подставляя решение системы (8) в (1) и повторяя процедуру до получения минимальной невязки, приходим к нормальному решению системы (1).

Алгоритмы методов сингулярного разложения и регуляризации испытывались на

решении системы (1) с квадратной матрицей A размерности 10×10 и правой частью f , где

$$A = \begin{pmatrix} 0.01 & 0.01 & 0.009 & -0.008 & 0.008 & 0.009 & 0.007 & 0.007 & 0.009 & 0.008 \\ 0.09 & 0.01 & 0.009 & 0.009 & -0.009 & 0.008 & 0.007 & 0.007 & 0.008 & 0.009 \\ 0.08 & 0.009 & 0.008 & 0.008 & 0.008 & 0.007 & 0.007 & 0.008 & 0.009 & 0.008 \\ 0.01 & 0.009 & 0.007 & -0.001 & 0.007 & 0.007 & -0.008 & 0.008 & 0.009 & 0.008 \\ 0.01 & 0.009 & 0.007 & 0.007 & 0.008 & -0.008 & 0.009 & 0.009 & 0.008 & 0.007 \\ 0.01 & 0.009 & 0.007 & 0.007 & -0.008 & 0.008 & 0.009 & 0.009 & 0.008 & 0.007 \\ 0.01 & 0.009 & 0.007 & -0.008 & 0.008 & 0.007 & 0.008 & -0.008 & -0.007 & 0.007 \\ 0.009 & 0.009 & 0.007 & 0.008 & 0.008 & 0.008 & 0.008 & 0.008 & 0.007 & -0.007 \\ 0.01 & 0.009 & 0.007 & 0.008 & 0.008 & 0.008 & 0.008 & 0.008 & -0.007 & -0.007 \\ 0.01 & 0.009 & 0.007 & 0.008 & 0.008 & 0.008 & 0.008 & 0.008 & 0.007 & -0.007 \end{pmatrix}$$

$$f = (0.046 \ 0.059 \ 0.094 \ -0.031 \ 0. \ -0.016 \ 0.122 \ 0.017 \ 0.046 \ 0.018)^T.$$

Матрица A является плохо обусловленной (с числом обусловленности $cond(A) \approx 22300$) и вырожденной ($det A = -2, 7 \cdot 10^{-21}$), поэтому система (1) имеет бесконечное множество решений. Решение, полученное методом сингулярного разложения

$$x_{SVD} = (1, -2, 1, -1, 3, 2, 4, -3, -2, 1)^T$$

имеет норму $\|x_{SVD}\| \approx 7, 071$. При решении системы методом регуляризации нужно задавать начальное приближение и параметр регуляризации. Решения, полученные методом регуляризации при начальных приближениях $x^{(0)}=f$, $x^{(0)}=0$, $x^{(0)}=1$, $x^{(0)}=2$ и при различных значениях параметра регуляризации $\alpha=0,1 \div 10$, практически совпадают:

$$x_{рег} \approx (1,037; -0,769; -0,639; -1,200; 2,972; 1,972; 4,095; -2,853; -1,999; 0,938)^T$$

и имеют норму $\|x_{рег}\| \approx 6, 762$.

Вопрос о том, какое решение является предпочтительным, зависит от физического содержания задачи и от того, какими свойствами мы его наделяем. В данном случае оба решения минимизируют невязку системы (1) (практически невязка равна нулю), но второе решение (полученное методом регуляризации) имеет меньшую норму и поэтому в этом смысле является более предпочтительным.

Литература:

1. Тихонов А.Н., Костомаров Д.П. Вводные лекции по прикладной математике. – М.: Наука, 1984. – 192 с.
2. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. –М.: Мир, 1980. – 279 с.
3. Тихонов А.Н., Арсенин В.Я., Методы решения некорректных задач. –М.: Наука, 1974. – 287 с.
4. Ильин В.А., Позняк Е.Г. Линейная алгебра. –М.: Наука, 1984.