

# Обусловленность и устойчивость

## Обусловленность задачи

Пусть есть абстрактная вычислительная задача.

Дано:  $x$ , найти:  $f(x)$ . Коль скоро  $x$  – это нечто, представленное в компьютере, можно сказать, что он представлен неточно, отличается от истинного на некоторое значение  $\Delta x$ . Часто бывает так, что и сами исходные данные представлены с какой-то ошибкой.

$$\Delta f = f(x + \Delta x) - f(x)$$

Необходимо исследовать, как такие ошибки влияют на значения функции  $f(x)$ .

*Абсолютная обусловленность* это предел отношения нормы приращения  $f$  к норме приращения  $x$ :

$$\nu = \lim_{\varepsilon \rightarrow 0} \sup_{||\Delta x|| < \varepsilon} \frac{||\Delta f||}{||\Delta x||}$$

Более естественным понятием является *относительная обусловленность*:

$$\mu = \lim_{\varepsilon \rightarrow 0} \sup_{||\Delta x|| < \varepsilon} \frac{||\Delta f||/||f||}{||\Delta x||/||x||}$$

Допустим, если  $\mu = 10$ , можно сказать, что одна неизвестная десятичная цифра в  $x$  приводит к тому, что в  $f$  неизвестно уже две десятичные цифры.

## Обусловленность задачи вычисления значения вещественной функции одной переменной

В случае такой задачи вычисления обусловленности несколько упрощаются:

- Абсолютная обусловленность -  $\nu = \lim_{|\Delta x| \rightarrow 0} \frac{|\Delta f|}{|\Delta x|} = |f'|$
- Относительная обусловленность -  $\mu = \lim_{|\Delta x| \rightarrow 0} \frac{|\Delta f|/|f|}{|\Delta x|/|x|} = \left| \frac{f'x}{f} \right|$

Рассмотрим значения относительной обусловленности для некоторых функций:

$$\mu_{\sqrt{x}} = 1/2$$

$$\mu_{1/x} = 1$$

$$\mu_{x^n} = n$$

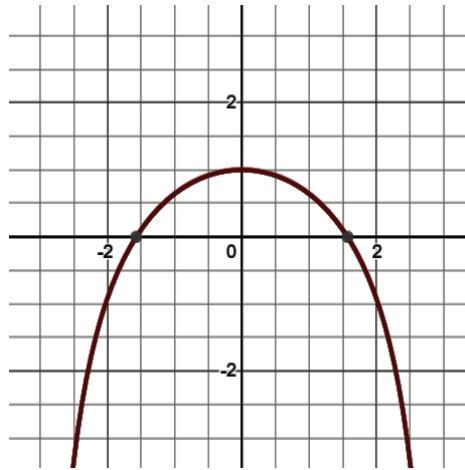
$$\mu_{e^x} = x$$

$$\mu_{\sin x} = \frac{x}{\tan x}$$

Задача вычисления корня обусловлена хорошо, что можно сказать и про  $f(x) = \frac{1}{x}$ .

Задача вычисления  $f(x) = e^x$  при больших значениях  $x$  обусловлена плохо.

График относительной обусловленности задачи вычисления функции  $f(x) = \sin x$  представлен ниже:



Наиболее хорошая обусловленность в промежутке  $[-\pi/2, \pi/2]$

Не все функции имеют удобную относительную обусловленность, например  $f(x) = \log x$ :

$$\mu_{\log x} = \frac{1}{|\log x|}$$

$$\nu_{\log x} = \frac{1}{|x|}$$

Несложно заметить, что в районе 1 число относительной обусловленности стремится к бесконечности.

В проектировании машинных алгоритмов, однако, вышеописанные соображения по поводу обусловленности не играют большой роли, потому что в компьютере нет бесконечно малых величин, а есть дискретная сетка чисел. Необходимо обеспечить верность значащих битов вычисляемой функции. Эта цель достигается средствами, порой не связанными (по крайней мере напрямую) с обусловленностью.

## Погрешности машинной арифметики

Отметим несколько аспектов того, как организуются вычисления в машинной арифметике.

Прежде всего, определим погрешность представления:

$$\delta x = \frac{\tilde{x} - x}{x} = \frac{\varepsilon_x}{x} \quad |\delta x| < \epsilon$$

$$\delta y = \frac{\tilde{y} - y}{y} = \frac{\varepsilon_y}{y} \quad |\delta y| < \epsilon$$

Предположим, необходимо посчитать  $x^2 - y^2$ . Вычислим погрешности операций, если сначала возводить  $x$  в квадрат, затем  $y$  в квадрат, а затем производить вычитание:

$$x \otimes x = (x \times x)(1 + \delta_1), |\delta_1| \leq \epsilon$$

$$y \otimes y = (y \times y)(1 + \delta_2), |\delta_2| \leq \epsilon$$

$$(x \otimes x) \ominus (y \otimes y) = (x^2(1 + \delta_1) - y^2(1 + \delta_2))(1 + \delta_3), |\delta_3| \leq \epsilon$$

И теперь можно определить относительную погрешность представления всего выражения:

$$\delta((x \otimes x) \ominus (y \otimes y)) = \frac{(x^2(1 + \delta_1) - y^2(1 + \delta_2))(1 + \delta_3) - (x^2 - y^2)}{x^2 - y^2}$$

$$\approx \frac{x^2 \delta_1 - y^2 \delta_2 + (x^2 - y^2) \delta_3}{x^2 - y^2} = \delta_3 + \delta_1 + \frac{y^2(\delta_1 - \delta_2)}{x^2 - y^2}$$

Можно сказать, что алгоритм неустойчивый, так как относительная погрешность зависит от значений  $x, y$ .

Теперь вычислим то же выражение, но используя разложение  $x^2 - y^2 = (x + y)(x - y)$ :

$$\delta((x \oplus y) \otimes (x \ominus y)) = \frac{(x+y)(1+\delta_1)(x-y)(1+\delta_2)(1+\delta_3) - (x^2 - y^2)}{x^2 - y^2} \approx \delta_1 + \delta_2 + \delta_3$$

Алгоритм можно считать устойчивым.

Фактически была решена одна и та же задача, но разными методами. Поэтому стоит говорить не только об обусловленности задачи, но и об обусловленности методов: иногда задача может быть хорошо обусловлена, а метод может оказаться неустойчивым.

## Обусловленность задачи нахождения корня функции

Дано:  $f(x) = 0$ . Найти:  $x$ .

Задача поменялась – теперь дана функция  $f(x)$ , а найти необходимо  $x$ , таким образом в определении числа обусловленности меняются местами числитель и знаменатель.

Относительная обусловленность:

$$\mu = \lim_{|\Delta x| \rightarrow 0} \frac{|\Delta x|/|x|}{|\Delta f|/|f|} = \left| \frac{f}{x f'} \right|$$

В окрестности точки, где  $f' = 0$ , численный метод, использующий  $f$  для нахождения корня, скорее всего, не сойдется из-за плохой обусловленности задачи.

Теперь рассмотрим обусловленность задачи нахождения корня многочлена.

Дано:  $P(x) = \sum_{k=0}^n a_k x^k = 0$ . Найти:  $x$

Выпишем следующую частную производную, что позволит определить относительную обусловленность:

$$\frac{dx}{da_k} = \frac{1}{P'} \frac{dP}{da_k} \Rightarrow \lim_{|\Delta a_k| \rightarrow 0} \frac{|\Delta x|/|x|}{|\Delta a_k|/|a_k|} = \left| \frac{a_k x^{k-1}}{P'} \right|$$

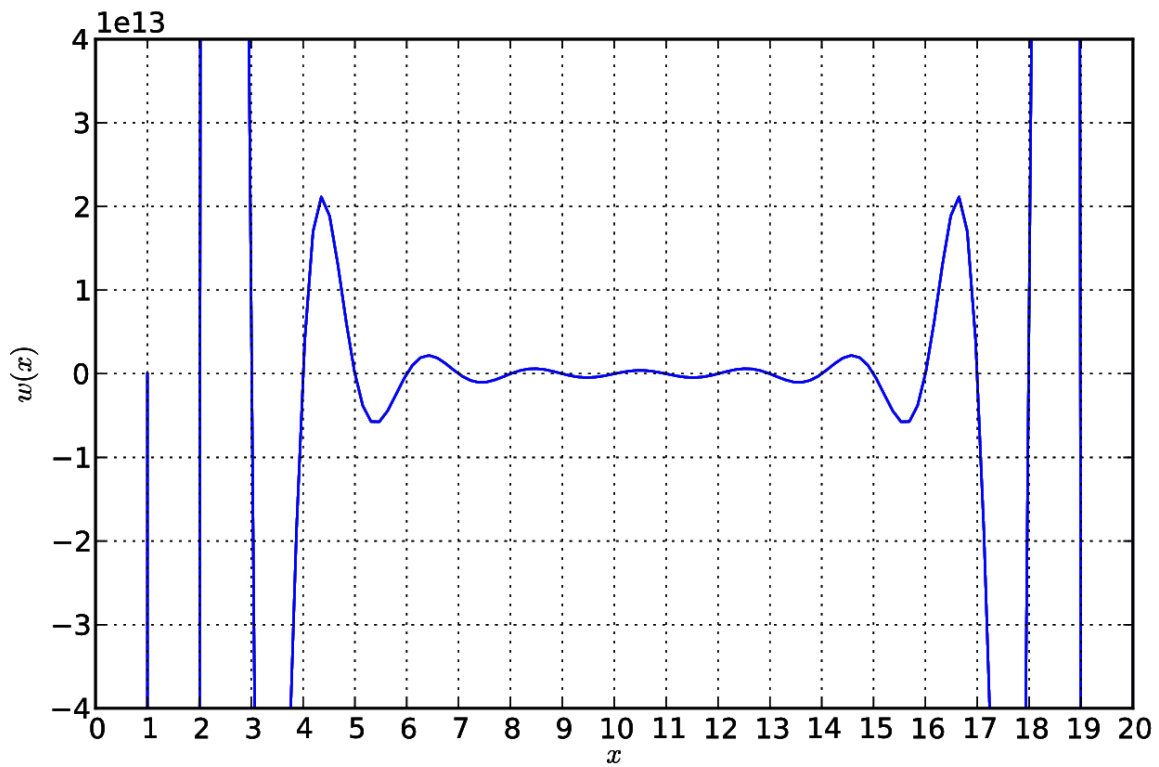
Можно показать, что чувствительность будет высокая в тех точках, где производная многочлена равна 0, то есть решение нестабильно для кратных корней.

Кратность корня – не единственная причина, по которой обусловленность задачи может быть плохой.

Известен многочлен Уилкинсона:

$$w(x) = \prod_{i=1}^{20} (x - i) = (x - 1)(x - 2) \cdots (x - 20)$$

График многочлена:



Экспериментально было показано, что многочлен очень чувствителен к малейшим изменениям коэффициента – изменение коэффициента на маленькое значение может повлечь за собой изменение корня на большие значения.

Чувствительность к коэффициентам полиномов, получающихся в результате интерполяции, как правило, также высока.

## Решение квадратного уравнения

Дано:  $ax^2 + bx + c = 0$ . Найти:  $x$ .

Решение общеизвестно:

$$x = \frac{-b \pm \sqrt{D}}{2a}, \quad D = b^2 - 4ac$$

Итак, число относительной обусловленности системы при параметрах  $a, b, c$  (в каждом случае два остальных параметра считаются фиксированными):

$$\begin{aligned} \mu_a &= \lim_{\Delta a \rightarrow 0} \frac{|\Delta x|/|x|}{|\Delta a|/|a|} = \left| \frac{xx'_a}{a} \right| = \frac{1}{2} \left| \frac{\pm b}{\sqrt{D}} - 1 \right| \\ \mu_b &= \left| \frac{xx'_a}{a} \right| = \left| \frac{b}{\sqrt{D}} \right| \\ \mu_c &= \left| \frac{xx'_c}{c} \right| = \left| \frac{2ac}{\sqrt{D}(-\sqrt{D} \pm b)} \right| \end{aligned}$$

При кратных корнях, то есть при  $D = 0$ , то получаем стремящееся к бесконечности число относительной обусловленности, что означает нестабильную работу данной формулы в случае кратных корней.

Если  $a, c$  очень малы, то в числителе для одного из корней получается практически  $-b + b$ . Получаем катастрофическое сокращение значащих битов, то есть большая часть битов обнулится, что приводит к большой неустойчивости данного численного метода. Чтобы избавиться от этого, несколько изменим формулы для вычисления корней:

Решение для  $D \approx b^2$ :

$$x_1 = \frac{2c}{-b - \operatorname{sgn}(b) \sqrt{D}}$$

$$x_2 = \frac{-b - \operatorname{sgn}(b) \sqrt{D}}{2a} = \frac{c}{ax_1}$$

Таким образом, в знаменателе не вычитаются близкие величины, значащие биты не теряются. Ответ получается с меньшей ошибкой. Для второго корня можно использовать теорему Виета. Кроме того, даже если  $a = 0$ , один из корней также можно будет найти.

## Обусловленность СЛАУ

Дано:  $A\mathbf{x} = \mathbf{b}$ . Найти  $\mathbf{x}$ .

В данной задаче тема обусловленности наиболее актуальна.

Прежде чем приступать к методам, рассмотрим математическую обусловленность задачи:

$$(A + \Delta A)(\mathbf{x} + \Delta \mathbf{x}) = \mathbf{b} + \Delta \mathbf{b}$$

Подробный вывод опустим и выпишем сразу результат:

$$\Delta A \Delta \mathbf{b} \approx 0 \Rightarrow \frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \mu \left( \frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} \right), \mu = \|A\| \cdot \|A^{-1}\| \geq 1$$

В любом случае, число обусловленности не может быть меньше 1. Числа обусловленности, исчисляемые тысячами, говорят о плохой обусловленности.

Если матрица  $A$  симметричная положительно определенная, то число обусловленности – отношение максимального и минимального собственного числа:

$$\mu = \frac{\lambda_{\max} A}{\lambda_{\min} A}$$

## Пример

Рассмотрим пример того, как работает обусловленность в СЛАУ.

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.001 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Как видно,  $x = 2, y = 0$  подходят как математическое решение системы.

Внесем возмущение в вектор  $\mathbf{b}$ :

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.001 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2.001 \end{pmatrix}$$

Решение системы:  $x = 1, y = 1$ . Получается совершенно другое решение, несмотря на небольшое возмущение во входных данных. Данная система плохо обусловлена.

Докажем это, посчитав матрицу  $A^{-1}$  и вычислив число обусловленности:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.001 \end{pmatrix}^{-1} = \begin{pmatrix} 1001 & -1000 \\ -1000 & 1000 \end{pmatrix}$$

$$\|A\|_2 \cdot \|A^{-1}\|_2 \approx 4000$$

Число обусловленности велико, что доказывает плохую обусловленность.

## Пример №2

В этом примере будет матрица  $A$ , которая дает вполне приемлемое число обусловленности:

$$A = \begin{pmatrix} 0.001 & 1 \\ 1 & 1 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} -1.001 & 1.001 \\ 1.001 & -0.001001 \end{pmatrix}, \quad \mu(A) = \|A\|_2 \cdot \|A^{-1}\|_2 \approx 3.46$$

Проведем решение СЛАУ методом Гаусса:

Первым и единственным шагом будут преобразования для приведения матрицы к верхней треугольной, с которой обусловленность уже будет плохой.

$$U = \begin{pmatrix} 0.001 & 1 \\ 0 & -999 \end{pmatrix}, \quad U^{-1} \approx \begin{pmatrix} 1000 & 1.001 \\ 0 & -0.001001 \end{pmatrix}, \quad \mu(U) = \|U\|_2 \cdot \|U^{-1}\|_2 \approx 1414$$

Таким образом, применение численного метода ведет к плохой обусловленности матрицы, которая изначально не давала плохую обусловленность.

Если воспользоваться методом  $LUP$  разложения, то такой проблемы возникать уже не будет (будет выбираться ведущий элемент в столбце, осуществляться перестановка, что исправит ситуацию):

$$PA = \begin{pmatrix} 1 & 1 \\ 0.001 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 1 \\ 0 & 0.999 \end{pmatrix}, \quad U^{-1} \approx \begin{pmatrix} 1 & -1.001 \\ 0 & 1.001 \end{pmatrix}, \quad \mu(U) = \|U\|_2 \cdot \|U^{-1}\|_2 \approx 3.46$$

Таким образом, выбор ведущего элемента влияет на обусловленность.

Это не является строгим доказательством того, что  $LUP$  разложение всегда лучше справляется с возрастанием обусловленности, но на практике чаще всего это так.

Итак, методы могут ухудшать обусловленность матрицы, не менять ее, но могут ли улучшать?

## Предобуславливатели СЛАУ

Модифицируем систему, умножив слева и справа на некоторую (невырожденную) матрицу:

$$A\mathbf{x} = \mathbf{b} \rightarrow M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}, \quad \mu(M^{-1}A) < \mu(A)$$

Если обеспечится условие меньшего значения числа обусловленности, то цель достигнута. Вопрос только в том, что это за матрица  $M$ .

$LUP$ -разложение также в некотором роде содержит предобуславливатель, в качестве него выступает матрица  $P$ :

$$PA = LU \Rightarrow PLU\mathbf{x} = P\mathbf{b}, \quad \mu(LU) \leq \mu(A)$$

Однако для нахождения  $P$  требуется прогонка метода, чего хотелось бы избежать.

Чисто формально единичная матрица является предобуславливателем, потому что неравенство чисел обусловленности нестрогое.

$$M = E : \mu(M^{-1}A) = \mu(A)$$

Также предобуславливателем может быть и сама матрица  $A$ , ведь тогда число обусловленности станет равно 1:

$$M = A : \mu(M^{-1}A) = 1$$

Но в таком случае необходимо считать матрицу  $A^{-1}$ . Вычисление обратной матрицы, во-первых, требует больших затрат, чем нахождение более простого предобуславливателя, и во-вторых, обесмысливает задачу, т. к. сама по себе обратная матрица будет содержать те самые численные ошибки, от которых мы хотели избавиться, вводя предобуславливатель.

Надо понимать, что поиск предобуславливателя – не совсем точная наука, и матрицы  $A$  бывают самые разные. Но вот некоторые из них, которые применяются в итеративных методах:

- Предобуславливатель Якоби:  $A = L + D + U$ ,  $M = D$
- Предобуславливатель Гаусса-Зейделя (GS):  $M = L + D$
- Симметричный предобуславливатель Гаусса-Зейделя (SGS):  

$$M = (L + D)D^{-1}(L + D)^T$$
- Прочие: SSOR, ILU(0), ILU(n), AMG