Тема VI: Евклидовы и унитарные пространства

3. Метод наименьших квадратов

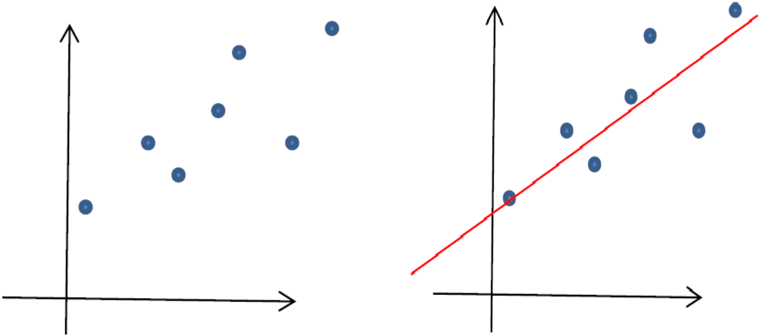
М.В.Волков

Уральский федеральный университет
Институт естественных наук и математики
кафедра алгебры и фундаментальной информатики

2024/2025 учебный год

Обсудим решение несовместных систем линейных уравнений подробнее.

Типичный источник таких систем — обработка экспериментальных данных.



Число неизвестных мало (в данном примере ищется прямая $y = ax + b$, и неизвестные — это коэффициенты a и b), а число уравнений велико (в данном примере каждая точка (x_i, y_i) задает уравнение $y_i = ax_i + b$).

Псевдорешения

Псевдорешение системы линейных уравнений $A\mathbf{x} = \mathbf{b}$ — это вектор $\mathbf{x}_0$, минимизирующий расстояние между векторами $A\mathbf{x}$ и $\mathbf{b}$.

В прошлой лекции был намечен такой план поиска псевдорешений:

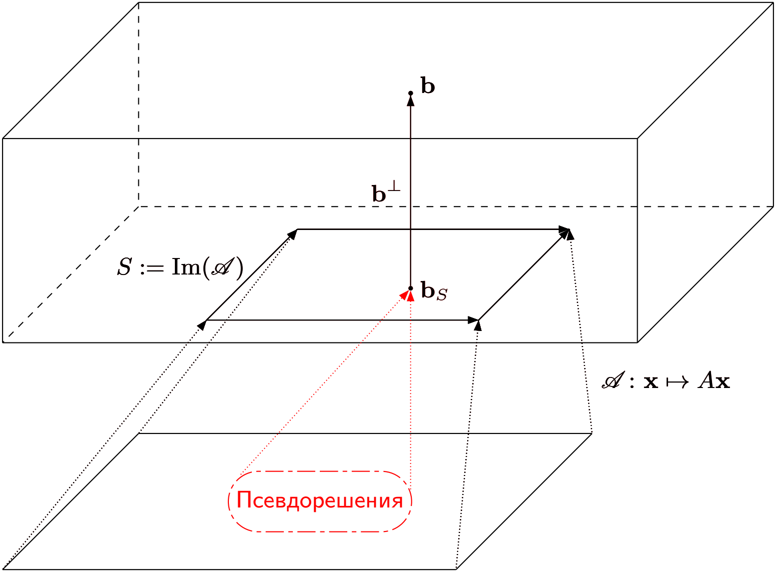
- найти ортогональную проекцию $\mathbf{b}_S$ вектора $\mathbf{b}$ на образ S линейного отображения $\mathbf{x} \mapsto A\mathbf{x}$ (т.е. на подпространство, порожденное столбцами матрицы A);
- решить совместную систему $A\mathbf{x} = \mathbf{b}_S$.

Было показано, что любое решение системы $A\mathbf{x} = \mathbf{b}_S$ действительно является псевдорешением исходной системы $A\mathbf{x} = \mathbf{b}$, так как наименьшее расстояние от вектора $\mathbf{b}$ до подпространства S есть расстояние от $\mathbf{b}$ до $\mathbf{b}_S$.

В реальных задачах ранг матрицы A равен числу неизвестных. (Например, в задаче проведения прямой $y = ax + b$ через набор точек $\{(x_i, y_i)\}$ ранг равен 1 только, если все эти точки лежат на одной вертикальной прямой.) При таком условии система $A\mathbf{x} = \mathbf{b}_S$ имеет единственное решение, а значит, имеется *единственное* псевдорешение исходной системы $A\mathbf{x} = \mathbf{b}$.

Если псевдорешение неединственно, то обычно интересуются псевдорешением наименьшей длины (*нормальное* псевдорешение). К вопросу о нормальных псевдорешениях мы вернемся позже.

Иллюстрация



Концептуально подход, описанный и проиллюстрированный выше, прост. Однако вычисление ортогональной проекции с помощью процесса Грама–Шмидта приводит к громоздким и *неустойчивым* вычислениям.

Опишем простое соображение, которое позволяет находить псевдорешения без вычисления ортогональной проекции. Его называют *методом наименьших квадратов*, поскольку речь идет о минимизации длины вектора $A\mathbf{x} - \mathbf{b}$, т.е. минимизации скалярного квадрата этого вектора.

Для определенности ограничимся случаем евклидова пространства; именно этот случай важен для практики.

Теорема (обоснование метода наименьших квадратов)

Пусть A — $k \times n$-матрица над $\mathbb{R}$, а S — образ линейного отображения $\mathbf{x} \mapsto A\mathbf{x}$ пространства $\mathbb{R}^n$ в пространство $\mathbb{R}^k$. Для произвольного вектора $\mathbf{b} \in \mathbb{R}^k$ системы линейных уравнений $A\mathbf{x} = \mathbf{b}_S$ и $A^T A\mathbf{x} = A^T\mathbf{b}$ равносильны.

Таким образом, вектор $\mathbf{x} \in \mathbb{R}^n$ будет псевдорешением системы $A\mathbf{x} = \mathbf{b}$ тогда и только тогда, когда $\mathbf{x}$ является решением системы $A^T A\mathbf{x} = A^T\mathbf{b}$.

Метод наименьших квадратов (2)

Доказательство. Подпространство S порождается образами базисных векторов пространства $\mathbb{R}^n$, т.е. столбцами матрицы A. Столбцы матрицы A — это строки матрицы A^T. Из ортогонального разложения пространства $\mathbb{R}^k$ относительно подпространства S имеем $\mathbf{b} = \mathbf{b}_S + \mathbf{b}^\perp$. Умножая это равенство слева на матрицу A^T и вспоминая формулу, выражающую скалярное произведение в евклидовом пространстве через координаты в ортонормированном базисе (т.е. формулу $\mathbf{uv} = [\mathbf{u}]^T[\mathbf{v}]$), получаем:

$$A^T\mathbf{b} = A^T(\mathbf{b}_S + \mathbf{b}^\perp) = A^T\mathbf{b}_S + A^T\mathbf{b}^\perp = A^T\mathbf{b}_S, \tag{$\star$}$$

поскольку вектор $\mathbf{b}^\perp$ ортогонален всем векторам из S.
Из $(\star)$, если $\mathbf{x} \in \mathbb{R}^n$ — решение системы $A\mathbf{x} = \mathbf{b}_S$, то

$$A^TA\mathbf{x} = A^T\mathbf{b}_S \overset{(\star)}{=} A^T\mathbf{b}.$$

Обратно, пусть $\mathbf{x} \in \mathbb{R}^n$ является решением системы $A^TA\mathbf{x} = A^T\mathbf{b}$. Возьмем произвольный вектор $\mathbf{y} \in \mathbb{R}^n$ и перемножим вектора $A\mathbf{y}$ и $A\mathbf{x} - \mathbf{b}_S$. Имеем (снова используя формулу $\mathbf{uv} = [\mathbf{u}]^T[\mathbf{v}]$)

$$(A\mathbf{y})^T(A\mathbf{x}-\mathbf{b}_S) = \mathbf{y}^TA^T(A\mathbf{x}-\mathbf{b}_S) = \mathbf{y}^T(A^TA\mathbf{x}-A^T\mathbf{b}_S) \overset{(\star)}{=} \mathbf{y}^T(A^TA\mathbf{x}-A^T\mathbf{b}) = 0.$$

Итак, вектор $A\mathbf{x} - \mathbf{b}_S$ ортогонален любому вектору из S, в частности, самому себе. Отсюда $A\mathbf{x} - \mathbf{b}_S = \mathbf{0}$ и $A\mathbf{x} = \mathbf{b}_S$. □

Отметим, что если ранг $k \times n$-матрицы A равен n, то $n \times n$-матрица $A^T A$ будет обратимой (*упражнение*). В этом случае при любой правой части $\mathbf{b} \in \mathbb{R}^k$ система $A\mathbf{x} = \mathbf{b}$ имеет единственное псевдорешение, для которого есть простая формула:

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}.$$

Метод наименьших квадратов изобрел (по его словам — в 1795 г.) и с большим успехом применял Карл Фридрих Гаусс (1777–1855).

Заключительные замечания

1. Метод наименьших квадратов работает и для унитарных пространств, т.е. для систем с комплексными коэффициентами. Единственное отличие состоит в том, что для отыскания псевдорешений системы линейных уравнений $A\mathbf{x} = \mathbf{b}$ надо решать систему $A^*A\mathbf{x} = A^*\mathbf{b}$, где A^* — *эрмитово сопряженная* матрица к матрице A. (Эрмитово сопряженная матрица получается, если исходную матрицу транспонировать и заменить каждый элемент его сопряженным: если $A = (a_{ij})_{k\times n}$, то $A^* := (\overline{a_{ji}})_{n\times k}$.)

2. Имеются и другие методы нахождения псевдорешений несовместных системы линейных уравнений, например, итерационный *метод Качмажа*.

Утверждалось, что если ранг $k \times n$-матрицы A равен n, то $n \times n$-матрица $A^T A$ будет обратимой. На самом деле, справедлив более общий факт:

Предложение

Для любой матрицы A над $\mathbb{R}$ ее ранг равен рангу матрицы $A^T A$.

Доказательство. Выше установлено, что системы линейных уравнений $A\mathbf{x} = \mathbf{b}_S$ и $A^T A\mathbf{x} = A^T\mathbf{b}$ равносильны для произвольного вектора $\mathbf{b}$. Полагая $\mathbf{b} = \mathbf{0}$, заключаем, что у однородных систем $A\mathbf{x} = \mathbf{0}$ и $A^T A\mathbf{x} = \mathbf{0}$ одно и то же пространство решений; обозначим его через R. Применяя к каждой из этих систем теорему о размерности пространства решений линейной однородной системы, получаем, что размерность R равна разности между числом неизвестных и рангом матрицы A и в то же время равна разности между числом неизвестных и рангом матрицы $A^T A$. Следовательно, эти ранги равны. $\square$

Вопрос

Верен ли аналогичный факт для матриц над произвольными полями?