

# ЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ

## Метод наименьших квадратов

Константин Ловецкий  
Октябрь 2012

Кафедра систем телекоммуникаций

# История метода наименьших квадратов

Метод возник и разработан в эпоху великих географических открытий. Гауссу ([Carl Friedrich Gauss](#)) приписывают создание основ метода наименьших квадратов в 1795 году (18-ти лет от роду). Однако впервые (в 1805) результаты были опубликованы Лежандром ([Legendre](#)).

Наиболее ранняя демонстрация мощи метода была продемонстрирована в 1801 году, когда был снова обнаружен астероид Ceres на основе расчетов Гаусса по методу наименьших квадратов.

Гаусс опубликовал метод в работе 1809 года. Независимо от него метод опубликован также американцем [Robert Adrain](#) в 1808.

# История метода наименьших квадратов

- Наиболее ранний анализ систем линейных алгебраических уравнений приводится в древней китайской книге «Девять глав арифметики», предположительно написанной за 200 лет до нашей эры. В самом начале VIII-ой главы формулируется следующая задача.
- *Три меры хорошего зерна, две меры посредственного и одна плохая продаются за 39 доу. Две меры хорошего, три меры посредственного и одна плохая продаются за 34 доу; одна мера хорошего, две посредственного и три плохого зерна продаются за 26 доу. Требуется определить, сколько стоит одна мера хорошего, посредственного и плохого зерна соответственно.*
- Сегодня задача может быть переформулирована следующим образом: Найти решение системы линейных алгебраических уравнений

$$3x + 2y + z = 39,$$

$$2x + 3y + z = 34,$$

$$x + 2y + 3z = 26,$$

- 
- где  $x$ ,  $y$  и  $z$  представляют собой цены мер хорошего, посредственного и плохого зерна соответственно.

# История метода наименьших квадратов

- Метод решения задачи, предложенный древними китайцами, заключался в следующем.
- Разноцветные бамбуковые палочки, представляющие коэффициенты системы уравнений, помещались в соответствующие ячейки «матричной счетной доски» и построчно перемещались в соответствии с некоторыми эмпирическими правилами. Их техника «счетных досок» и набор правил перемещения цветных палочек сначала попала в Японию, а затем добралась и до Европы.
- Здесь цветные палочки были заменены цифрами, а счетные доски трансформировались в записи на листах бумаги. В Европе этот алгоритм стал известен как метод исключения (неизвестных) Гаусса, получившим свое имя в честь великого немецкого математика Карла Гаусса, активно применявшим этот метод при решении многих практических задач.

1820 watercolor caricatures of the French mathematicians **Adrien-Marie Legendre** (left) and **Joseph Fourier** (right) by French artist **Julien-Leopold Boilly**, watercolor portrait numbers 29 and 30 of *Album de 73 Portraits-Charge Aquarelle's des Membres de l'Institute*.



# Карл Фридрих Гаусс



Carl Friedrich Gauss (1777–1855), painted by [Christian Albrecht Jensen](#)

# Постановка задачи

Цель состоит в подборе параметров пробной функции, описывающей экспериментальный набор данных. Простой набор данных состоит из  $n$  точек (пар данных),  $(x_i, y_i), i = 1, \dots, n$ , где  $x_i$  - независимые переменные и  $y_i$  - зависимые, полученные в результате наблюдений. Целевая функция задается в виде  $f(x, \beta)$ , где  $m$  параметров подгонки сосредоточены в векторе  $\beta$ .

Цель состоит в отыскании таких параметров модели, которые «наилучшим» образом аппроксимируют набор данных. Метод наименьших квадратов минимизирует сумму,  $S$ , квадратов невязок

$$S = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n (y_i - f(x_i, \beta))^2.$$

Пример. Аппроксимация с помощью линейной функции.

$$f(x, \beta) = \beta_0 + \beta_1 x.$$

# Обобщенная полная проблема наименьших квадратов

Generalized Total Least Squares

CHENGXIAN XU

Xian Jiaotong University, Xian, China

Одним из важных приложений нелинейного метода наименьших квадратов является задача подбора коэффициентов нелинейной модели. При этом обычно предполагается, что ошибки задания независимых переменных отсутствуют или малы. Однако во многих экспериментах могут присутствовать значительные ошибки измерения независимых переменных.

# Обобщенная полная проблема наименьших квадратов

Обычная формулировка проблемы наименьших квадратов:

Выбираем функцию, описывающую моделируемое явление:  $y = \varphi(x, t)$

и хотим подобрать параметры  $x \in R^n$  модели таким образом, чтобы функция аппроксимировала измеренные значения

$y_1, \dots, y_m$ ,  $i = 1, \dots, m$ , в заданных точках  $t_1, \dots, t_m$

# Метод наименьших квадратов

Минимизировать

$$\begin{aligned} f(x, \tau) &= \frac{1}{2} \sum_{j=1}^m \left[ w_j \left( \varphi(x, \tau_j) - y_j \right)^2 \right] = \\ &= \frac{1}{2} r^T W r \end{aligned}$$

**отыскав оптимальные значения параметров**  $x \in R^n$

Здесь весовые коэффициенты задаются диагональной матрицей

$$W = \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & w_m \end{pmatrix}$$

# Обобщенный метод наименьших квадратов

**Найти оптимальные значения параметров  $x \in R^n$   
и  $\tau_j, j = 1, \dots, m$ , минимизировав**

$$\begin{aligned} f(x, \tau) &= \frac{1}{2} \sum_{j=1}^m \left[ w_j (\varphi(x, \tau_j) - y_j)^2 + v_j (\tau_j - t_j)^2 \right] = \\ &= \frac{1}{2} \left[ r^T W r + e^T V e \right] \end{aligned}$$

**Весовые коэффициенты задаются диагональными матрицами**

$$W = \text{diag}(w_1, \dots, w_m), V = \text{diag}(v_1, \dots, v_m),$$

$$w_j \geq 0, v_j \geq 0, j = 1, \dots, m$$

$$r_j = \varphi(x, \tau_j) - y_j, e_j = \tau_j - t_j, j = 1, \dots, m$$

# Решение по методу Ньютона

Обобщенная проблема может быть решена с помощью любого метода минимизации нелинейной функции  $f(x, \tau)$  по  $(n+m)$  переменным  $(x, \tau)$  не используя специальной структуры функции. Однако прямое использование таких методов не является эффективным.

Предположим, что  $r_j(x, \tau)$ ,  $j = 1, \dots, m$ , и, следовательно,  $f(x, \tau)$  дважды непрерывно дифференцируемы:

$$\nabla f = \begin{bmatrix} \nabla_x f \\ \nabla_\tau f \end{bmatrix} = \begin{bmatrix} AWr \\ Ve + DWr \end{bmatrix},$$

$$\nabla^2 f = \begin{pmatrix} \nabla_{xx}^2 f & \nabla_{x\tau}^2 f \\ \nabla_{\tau x}^2 f & \nabla_{\tau\tau}^2 f \end{pmatrix}.$$

# Решение по методу Ньютона

где

$$\nabla_{xx}^2 f = AWA^T + \sum_{j=1}^m w_j r_j \nabla_{xx}^2 r_j,$$

$$\nabla_{x\tau}^2 f = AWD + \sum_{j=1}^m w_j r_j \nabla_{x\tau}^2 r_j,$$

$$\nabla_{\tau\tau}^2 f = V + DWD + \sum_{j=1}^m w_j r_j \nabla_{\tau\tau}^2 r_j,$$

$$A = [\nabla_x r_1, \dots, \nabla_x r_m],$$

$$D = \text{diag} \left[ \frac{\partial r_1}{\partial \tau_1}, \dots, \frac{\partial r_m}{\partial \tau_m} \right].$$

# Решение по методу Ньютона

Матрица  $\nabla_{\tau\tau}^2 f$  является диагональной с элементами

$$v_j + w_j \left( \frac{\partial r_j}{\partial \tau_j} \right)^2 + w_j r_j \frac{\partial^2 r_j}{\partial \tau_j^2}.$$

Метод Ньютона решения этой задачи приводит к решению системы линейных уравнений

$$\begin{bmatrix} \nabla_{xx}^2 f & \nabla_{x\tau}^2 f \\ \nabla_{\tau x}^2 f & \nabla_{\tau\tau}^2 f \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \tau \end{bmatrix} = - \begin{bmatrix} \nabla_x f \\ \nabla_\tau f \end{bmatrix},$$

для определения очередного приближения

$$x_+ = x + \delta x, \quad \tau_+ = \tau + \delta \tau$$

# Приближенные методы Ньютона

При разработке методов решения полной задачи о наименьших квадратах важно учитывать специфическую структуру целевой функции  $f(x)$  и ее  $f'(x)$  производных, в частности, тот факт, что переменные  $x$  и  $\tau$  могут считаться независимыми.

Для полиномиальной аппроксимации измеренных данных можно рассматривать приближенные методы Ньютона. Эти методы используют вторые производные целевой функции  $f(x, \tau)$  и  $f'(x, \tau)$  по каждому из аргументов, но пренебрегают

вычислением смешанных производных  $\nabla_{x\tau}^2 f$  и  $\nabla_{\tau x}^2 f$ .

# Приближенные методы Ньютона

Задача оптимизации называется сепарабельной, если ее оптимизация по одним переменным много проще, чем по другим. Обобщенная полная проблема наименьших квадратов относится к таким задачам.

Рассмотрим метод, использующий необходимые условия первого порядка для разделения переменных  $x$  и  $t$ , после чего задача решается с использованием метода Ньютона.

Предположим, что аппроксимирующая функция  $\varphi(x, t)$  является полиномом вида

$$\varphi(x, t) = \sum_{i=1}^n x_i p_i(t),$$

где  $p_i(t), i = 1, \dots, n$ , - множество ортогональных полиномов.

# Приближенные методы Ньютона

В этом случае внедиагональные элементы матрицы  $\nabla_{xx}^2 f$  обращаются в ноль и обе матрицы  $\nabla_{xx}^2 f$  и  $\nabla_{\tau\tau}^2 f$  становятся диагональными. Предполагая, что элементы матриц  $\nabla_{x\tau}^2 f$  и  $\nabla_{\tau x}^2 f$  пренебрежимо малы, получим диагональную аппроксимацию матрицы Гессе: 
$$\begin{bmatrix} \nabla_{xx}^2 f & 0 \\ 0 & \nabla_{\tau\tau}^2 f \end{bmatrix}.$$

Решение системы метода Ньютона имеет вид:

$$\delta x_i = -\frac{\sum_{j=1}^n w_j r_j p_i(\tau_j)}{\sum_{j=1}^n w_j p_i(\tau_j)^2}, i = 1, \dots, n,$$

$$\delta \tau_j = -\frac{v_j e_j + w_j \frac{\partial \varphi(x, \tau_j)}{\partial \tau_j} r_j}{v_j + w_j \left( \frac{\partial \varphi(x, \tau_j)}{\partial \tau_j} \right)^2 + w_j r_j \frac{\partial^2 \varphi(x, \tau_j)}{\partial \tau_j^2}}, j = 1, \dots, m.$$

# Приближенные методы Ньютона

Полиномы  $p_i(t), i = 1, \dots, n$ , ортогональные на множестве  $\tau_j, j = 1, \dots, m$ , могут быть получены по рекуррентным формулам

$$p_1 = 1, \quad p_2 = t - \alpha_1,$$

$$p_i(t) = (t - \alpha_{i-1})p_{i-1}(t) - \beta_{i-1}p_{i-2}(t), \quad i = 3, \dots, n.$$

где

$$\alpha_{i-1} = \frac{\sum_{j=1}^m w_j \tau_j p_{i-1}(\tau_j)^2}{\sum_{j=1}^m w_j p_{i-1}(\tau_j)^2},$$

$$\beta_{i-1} = \frac{\sum_{j=1}^m w_j \tau_j p_{i-1}(\tau_j) p_{i-2}(\tau_j)}{\sum_{j=1}^m w_j p_{i-2}(\tau_j)^2}.$$

# Приближенные методы Ньютона

Итерации приближенного метода Ньютона начинаются с начальной точки  $x^{(1)} = 0$  и  $\tau_j^{(1)} = t_j, j = 1, \dots, m$ . Затем на каждой итерации вначале вычисляются ортогональные на множестве точек  $\tau_j^{(k)} = t_j, j = 1, \dots, m$  полиномы  $p_i(t), i = 1, \dots, n$ , а затем определяется

очередное приближение к решению  $x^{(k+1)} = x^{(k)} + \delta x^{(k)}, \tau^{(k+1)} = \tau^{(k)} + \delta \tau^{(k)}$ .

Процесс продолжается до тех пор, пока не будет достигнуто решение с заданной точностью.

Дополнительное свойство этого метода заключается в том, что при необходимости решение может быть представлено в виде степеней параметра

$$\varphi(x, t) = \sum_{i=1}^n c_i t^{i-1} = \sum_{i=1}^n x_i p_i(t)$$

# Приближенные методы Ньютона

Коэффициенты  $c_i$  вычисляются по формулам

$$c_i = \sum_{k=i}^n x_k \sigma_{i+1,k+1}, \quad 1 \leq i \leq n,$$

$$\sigma_{i,k} = \begin{cases} 1 & i = k, \\ 0 & i > k; i, k < 2, \end{cases}$$

$$\sigma_{i+1,k+1} = \sigma_{i,k} - \alpha_{k-1} \sigma_{i+1,k} - \beta_{k-1} \sigma_{i+1,k-1}, \quad i < k.$$