Основы метода сопряженных градиентов

Е.А. Нурминский

13 ноября 2022 г.

Аннотация

Эта заметка описывает классический вариант замечательного метода сопряженных градиентов для решения задачи безусловной квадратичной оптимизации. Метод не требует задания в явной форме матрицы квадратичной формы, составляющей основной объем данных задачи, в качестве промежуточных данных генерирует лишь три линейных объекта (векторы) и обеспечивает в точной арифметике нахождение оптимума за число итераций, не превосходящих размерность пространства переменных.

Понятия и обозначения

$a, b, \ldots, z$ — векторы некоторого конечномерного пространства E

$A, B, \ldots, Z$ — матрицы

ab — скалярное произведение векторов a и b

$|a| = \sqrt{aa}$ — норма вектора a

Сложение, умножение матриц и векторов на действительные числа понимается обычным образом. Умножение матриц на векторы определяется так, как это принято в стандартной линейной алгебре. Некоторые дополнительные операции будут определяться по ходу дела.

1 Постановка задачи

Задача заключается в нахождении $x^\star$ такого, что

$$\min_{x \in E} \left(\frac{1}{2} xAx + bx \right) = \frac{1}{2} x^\star A x^\star + bx^\star \tag{1}$$

для заданных квадратной матрицы A и вектора b.

Сразу ответим на некоторые очевидные вопросы, ответы на которые неочевидны.

- Матрицу A можно считать симметричной.
- Если $\frac{1}{2}xAx + bx \geq C$ для некоторого C, то такое $x^\star$ существует.
- Если матрица A положительно определена, то $x^\star$ единственно.

Если матрица A отрицательно определена на некотором линейном подпространстве L, то $q(x) = \frac{1}{2}xAx + bx$ неограничена снизу и, следовательно, решение задачи (1) не существует.

Если матрица A не является отрицательно определенной ни на каком линейном подпространстве, то решение задачи (1) может как существовать, так и нет.

Для исключения ситуаций неопределенности с существованием решения задачи (1) в дальнейшем матрица A будет считаться положительно определенной.

В данной краткой заметке представлен классический вариант замечательного метода решения задачи (1), который не требует задания в явной форме матрицы квадратичной формы, составляющей основной объем данных и в качестве промежуточных данных использует лишь три линейных объекта (векторы) — текущее приближение к оптимуму, вычисляемый градиент целевой функции и дополнительный вектор, определяющий направление поиска. Этот метод был предложен в знаменательной работе Magnus R. Hestenes и Eduard Stiefel [1] и попрежнему остается одним из ключевых вычислительных инструментом в практической оптимизации (см. например, [2]). К сожалению, в отечественной учебной литературе этому методу уделяется недостаточное внимание [1], несмотря на его практическую эффективность и множество весьма полезных теоретических идей заложенных в его обоснование. Данная заметка мотивирована стремлением заполнить этот пробел.

2 Сопряженные системы векторов

Для решения задачи (1) фундаментальное значение имеет следующее понятие.

Определение 1 *Векторы $\{g^1, g^2, \dots, g^k\}$ называются сопряженными (относительно матрицы A), если*

$$g^i A g^j = \begin{cases} 0 & \text{если } i \neq j \\ > 0 & \text{если } i = j \end{cases} \tag{2}$$

В силу этого определения векторы $\{g^1, g^2, \dots, g^k\}$ будут линейно независимыми и, следовательно, $k \leq n$, где n — размерность пространства E.

Имея в наличии произвольную систему линейно независимых векторов $\{a^1, a^2, \dots, a^k\}$, можно построить систему сопряженных векторов $\{a^1, a^2, \dots, a^k\}$ при помощи аналога процедуры ортогонализации Грамма-Шмидта, представленной Алгоритмом 1.

Data: Система линейно независимых векторов $\{a^1, a^2, \dots, a^k\}$.
Result: Система сопряженных векторов $\{g^1, g^2, \dots, g^k\}$.
Инициализация: $g^1 = a^1, \ i = 1$;
while $i < k$ **do**
Вычислить

$$g^{i+1} = \sum_{s=1}^{i} \gamma_s g^s + a^{i+1} \tag{3}$$

где

$$\gamma_s = -g^s A a^{i+1} / g^s A g^s \tag{4}$$

end

Algorithm 1: Построение сопряженной системы

Действительно, легко проверить, что при этом каждый последующий вектор g^{i+1} будет сопряжен всем предыдущим $g^s, s = 1, 2, \dots, i$:

$$g^s A g^{i+1} = \sum_{\tau=1}^{i} \gamma_\tau g^s A g^\tau + g^s A a^{i+1} = \gamma_s g^s A g^s + g^s A a^{i+1} = 0$$

для $s = 1, 2, \dots, i$.

Таким образом, рассматривая в качестве исходной системы произвольный базис пространства E можно с помощью процедуры Грамма-Шмидта построить другой базис, состоящий из линейно-независимых сопряженных векторов $g^1, g^2, \dots, g^n$. Тогда любую точку x пространства E можно будет представить в виде линейной комбинации этих векторов:

$$x = \sum_{i=1}^{n} \xi_i g^i \tag{5}$$

и это колоссальным образом упрощает задачу квадратичной оптимизации.

3 Метод сопряженных градиентов

Исключительное значение систем сопряженных векторов для решения задачи (1) имеет то, что при использовании представления (5) задача квадратичной оптимизации превращается в сумму

[1] Это можно подтвердить приблизительно 17.5 миллионами ссылок в Интернет по ключу "conjugate gradient method" и всего лишь 33.4 тысячами ссылок в русскоязычном Интернете при поиске по "метод сопряженных градиентов". Русскоязычные ссылки в `https://ru.wikipedia.org/wiki/` "Метод сопряженных градиентов" безнадежно устарели, а сама статья содержит массу ошибок.

простейших одномерных квадратичных минимизаций:

$$\begin{gathered}\min_{x \in E} \left(\tfrac{1}{2}xAx + bx\right) = \min_{\xi \in E} \left\{ \tfrac{1}{2}(\textstyle\sum_{i=1}^n \xi_i g^i)A(\sum_{j=1}^n \xi_j g^j) + \sum_{k=1}^n \xi_i b g^k \right\} = \\ \textstyle\sum_{i=1}^n \min_{\xi_i} \left(\tfrac{1}{2}\alpha_i \xi_i^2 + \beta_i \xi_i\right) = \sum_{i=1}^n \left(\tfrac{1}{2}\alpha_i \bar{\xi}_i^2 + \beta_i \bar{\xi}_i\right)\end{gathered} \tag{6}$$

где $\alpha_i = g^i A g^i > 0,\ \beta_i = bg^i$. Эта система задач может быть решена в любом порядке, параллельно или нет и по сути дела с минимальными вычислительными затратами: любой (ну, почти любой) старшекласник скажет вам, что $\bar{\xi}_i = -\beta_i/\alpha_i,\ i = 1, 2, \ldots, n$. Однако ситуация тут совсем нетривиальна: для ее решения необходимо знать $\alpha_i, \beta_i, i = 1, 2, \ldots, n$, что требует определения системы сопряженных векторов $g^1, g^2, \ldots, g^n$.

Учитывая ресурсы современной вычислительной техники можно использовать непосредственно процедуру Грамма-Шмидта, описанную выше, задав произвольный начальный базис $\{a^1, a^2, \ldots, a^n\}$. Такой метод действительно существует и называется методом сопряженных направлений.

В качестве такого исходного базиса подошел бы, например, канонический базис из координатных ортов $\{e^1, e^2, \ldots, e^n\}$, который сам по себе имеет минимальные затраты по памяти, однако процедура (3) - (4) требует использования всей последовательности сопряженных векторов, вычисленных к этому моменту. Кроме этого в формуле (4) явным образом используется матрица квадратичной формы A, что приводит к затратам памяти для задачи с n переменными к величине порядка $\frac{3}{2}n^2$, что становиться проблемой уже для задач с достаточно скромным n порядка 10^5.

Именно тут на сцену появляются градиенты и авторы метода сопряженных градиентов Hesten и Stiefel [1] нашли великолепный выход из этого положения, заметив, что для квадратичной функции $q(x) = \frac{1}{2}xAx + bx$

$$q'(\bar{x} + z) - q'(\bar{x}) = A(\bar{x} + z) + b - A\bar{x} - b = Az \tag{7}$$

для любых $\bar{x}$ и z.

Тогда основная проблемная формула (4) может быть переписана например, как

$$r^s = q'(x^s - \lambda_s g^s) - q'(x^s),\ \gamma_s = -r^s a^{i+1}/r^s g^s,\ s = 1, 2, \ldots, i; \tag{8}$$

где x^s, a^{i+1} и λ могут быть выбраны произвольно. Однако, учитывая итеративный характер метода, можно избавится от затрат памяти на вектора $x^s - \lambda_s g^s$ и положить

$$x^{s+1} = x^s - \lambda_s g^s,\ s = 1, 2, \ldots, i \tag{9}$$

если $q'(x^s - \lambda_s g^s) \neq 0$, что требуется по условиям линейной независимости. Заметим, что если это не так, то вектор x^{s+1} является решением задачи (1) и нам не о чем больше беспокоится. В противном случае (8) упрощается до

$$r^s = q'(x^{s+1}) - q'(x^s),\ \gamma_s = -r^s a^{i+1}/r^s g^s,\ s = 1, 2, \ldots, i \tag{10}$$

и у нас остается еще возможность дополнительного выбора λ_s, неявно входящего в определение x^{s+1} и вектора a^{i+1}, условием на который является линейная независимость от всех предыдущих $a^s, s = 1, 2, \ldots, i$. Конечно, при этом хорошо бы выбрать эти величины так, чтобы как можно большее число коэффициентов $\gamma_s, s = 1, 2, \ldots, i$ обратились в нуль.

Обе эти задачи решает выбор шагового множителя λ_s из условия минимума по направлению $-g^s$:

$$q(x^{s+1}) = q(x^s - \lambda_s g^s) = \min_{\lambda} q(x^s - \lambda g^s) \tag{11}$$

который единственным образом его определяет. При этом из условия оптимальности в задаче (11) $q'(x^{s+1})g^s = 0$ и соотношение (10) упрощается до

$$\gamma_s = r^s a^{i+1}/q'(x^s)g^s,\ s = 1, 2, \ldots, i \tag{12}$$

Классическим выбором для вектора a^{i+1} является $q'(x^{i+1})$ для которого выполняется условие ортогональности всем предыдущим $q'(x^s), s = 1, 2, \ldots, i$.

Тогда $\gamma_s = 0$ для $s = 1, 2, \ldots, i-1$ и соотношение (12) упрощается до

$$g^{i+1} = q'(x^{i+1}) + \gamma_i g^i,\ \gamma_i = \|q'(x^{i+1})\|^2/\|q'(x^i)\|^2.$$

Убирая все эти промежуточные объяснения, метод сопряженных градиентов можно переписать в виде алгоритмической схемы Алгоритм 2, хотя при этом все его замечательные математические обоснования остаются за скобками.

Data: n-мерное пространство переменных E, квадратичная функция $q : E \to \mathbb{R}$ с положительно определенной матрицей квадратичной формы, ее градиентный оракул $q'(x)$ и начальная точка x^0.

Result: Точка $x^\star$ предоставляющая единственный минимум функции $q(x)$:

$$q(x^\star) = \min_{x \in E} q(x).$$

Инициализация: $g^0 = q'(x^0)$, счетчик итераций $i = 0$;

while $i < n$ **do**

Решить задачу одномерной минимизации

$$q(x^{i+1}) = q(x^i - \lambda_i g^i) = \min_{\lambda} q(x^i - \lambda g^i) \tag{13}$$

Вычислить следующий сопряженный вектор:

$$g^{i+1} = q'(x^{i+1}) + \gamma g^i, \tag{14}$$

где $\gamma = \|q'(x^{i+1})\|^2 / \|q'(x^i)\|^2$.;

Увеличить счетчик итераций: $i \to i + 1$ и пр.;

end

Algorithm 2: Метод сопряженных градиентов

Список литературы

[1] Hestenes M.R., Stiefel E. Methods of Conjugate Gradients for Solving Linear Systems, Journal of Research of the National Bureau of Standards 49 (1952), 409–436.

[2] Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. Пер. с англ. — М.: Мир, 1985.

Вопросы для самостоятельного поиска ответов

1. Что такое линейное подпространство ?
2. Какая матрица называется положительно (отрицательно) определенной ?
3. Почему квадратичная функция $\frac{1}{2}xAx + bx$ с матрицей A, отрицательно определенной на некотором линейном подпространстве, неограничена снизу ?
4. Что такое размерность пространства ?
5. Проверить совпадение для всех $i = 1, 2, \ldots, k$ линейных подпространств L_i^A и L_i^G, порожденных системами векторов $\{a^1, a^2, \ldots, a^i\}$ и $\{g^1, g^2, \ldots, g^i\}$ согласно процедуре Грамма-Шмидта.