Децентрализованная оптимизация в условиях Поляка-Лоясиевича

Николай Савельев

Московский физико-технический институт

Курс: Моя первая научная статья
(практика, В.В. Стрижов)/Группы 774, 794, весна 2020
Консультант: А. Н. Безносиков

Цель исследования

Цель работы

Предложить метод решения больших систем уравнений, производящий вычисления на децентрализованных системах.

Метод решения

Сведение задачи решения системы уравнений к задаче децентрализованной ограниченной оптимизации, для которой возможен переход к задаче композитной оптимизации, но уже без ограничений.

Список литературы

1. *Hamed Karimi, Julie Nutini, and Mark W. Schmidt* Linear convergence of gradient and proximal-gradient methods under the polyak- lojasiewicz condition.CoRR, abs/1608.04636, 2016.
2. *Xinlei Yi, Shengjun Zhang, Tao Yang, Karl H. Johansson, Tianyou Chai* Linear Convergence for Distributed Optimization Under the Polyak-Łojasiewicz Condition. 2019.
3. *А. В. Гасников* Универсальный метод для задач стохастической композитной оптимизации. 2016.
4. *А. В. Гасников.* Современные численные методы оптимизации, метод универсального градиентного спуска. 2018.

Постановка задачи: ограниченная оптимизация

Требуется решить систему уравнений:

$$\begin{cases} f_1(x) = 0, \\ f_2(x) = 0, \\ \dots \\ f_m(x) = 0, \end{cases} \qquad f_i : \mathbb{R}^n \to \mathbb{R}, \; i = 1, \dots, m.$$

Эквивалентная задача оптимизации:

$$g(x) := \sum_{i=1}^{m} f_i^2(x) \to \min_{x \in \mathbb{R}^n} . \tag{EQ}$$

Задача оптимизации в децентрализованном виде:

$$\begin{aligned} \min_{x_i \in \mathbb{R}^n} \quad & \sum_{i=1}^{m} f_i^2(x_i), \\ \text{s.t.} \quad & x_1 = x_2 = \dots = x_m, \end{aligned} \tag{DC}$$

где $x_1, \dots, x_m$ – векторы решений на каждом из вычислителей.

Постановка задачи: композитная оптимизация

Пусть G – граф связи вычислителей децентрализованной системы; $K \in \mathbb{R}^{n \times n}$ – матрица Кирхгофа графа G; $W = K \otimes I_n \in \mathbb{R}^{nm \times nm}$ – матрица для контроля совпадения решений; $X \in \mathbb{R}^{nm}$ – вектор из $x_1, \dots, x_m$; $F(X) = \sum\limits_{i=1}^{m} f_i^2(x_i)$.

Эквивалентная запись задачи (DC):

$$\begin{aligned} \min_{X \in \mathbb{R}^{nm}} \quad & F(X), \\ \text{s.t.} \quad & W^{1/2} X = 0. \end{aligned} \tag{LO}$$

При смягчении жестких условий возможен переход к задаче композитной оптимизации:

$$\min_{X \in \mathbb{R}^{nm}} \quad F(X) + R||W^{1/2}X||, \tag{CO}$$

где R – оптимизируемый параметр.

Предположения

Предположение 1

Граф G связный.

Предположение 2

Множество решений задачи (EQ) непусто.

Предположение 3

Градиент каждой из функций $f_1, \ldots, f_m$ удовлетворяет условию Липшица.

Предположение 4

Функция g удовлетворяет условию Поляка-Лоясиевича, то есть для всех $x \in \mathbb{R}^n$ выполнено $\frac{1}{2}||\nabla g(x)||^2 \geq \nu(g(x) - g^*)$, где $g^* = \min_{x \in \mathbb{R}^n} g(x)$, $\nu > 0$ – некоторая константа.

Децентрализованный градиентный метод

Вернемся к задаче:

$$\min_{X \in \mathbb{R}^{nm}} \quad F(X) + R||W^{1/2}X||. \tag{CO}$$

Пусть B_i – матрица из столбцов матрицы $W^{1/2}$, с номерами от $i \cdot n$ до $(i+1) \cdot n - 1$ включительно.

Шаг децентрализованного метода для i-го вычислителя

$x_i^{k+1} = x_i^k - h_i \cdot (f_i(x_i^k) \cdot \nabla f_i(x_i^k) + R \cdot B_i^T W^{1/2} X^k)$,
x_i^0 – произвольное начальное приближение,
$h_i > 0$ – фиксированный шаг.

Теорема

В предположениях 1-4 децентрализованный градиентный метод для задачи (CO) имеет линейную скорость сходимости.

Численный эксперимент

Сгенерированные данные

Рассматривается система линейных уравнений:

$$Ax = b,$$

где $A \in \mathbb{R}^{9\times 9}$ – положительноопределенная матрица с заданным числом обусловленности $\mu(A)$,
$b \in \mathbb{R}^9$ – произвольный вектор. a_i – i-я строка матрицы A.

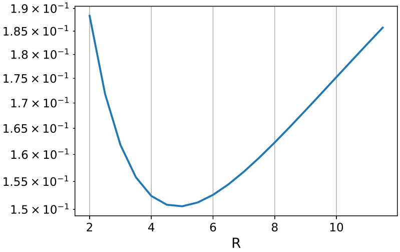
Цель эксперимента

Сравнить предложенный метод с градиентным спуском при разных числах обусловленности матрицы A.

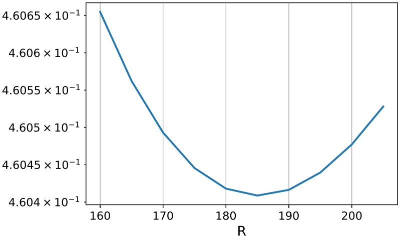
Шаг децентрализованного метода для i-го вычислителя

$$x_i^{k+1} = x_i^k - h_i \cdot ((a_i^T x_i^k - b_i) \cdot a_i + R \cdot B_i^T W^{1/2} X^k),$$
$$x_i^0 = 0;\ h_i = 1/\lambda_{max}(a_i a_i^T + R \cdot B_i^T B_i).$$

$\mu(A) = 10$:



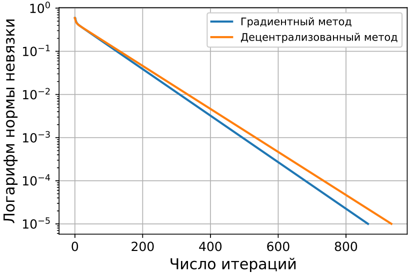
$\mu(A) = 82$:

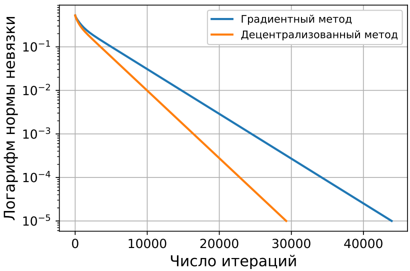


По вертикальной оси обоих графиков отложены логарифмы норм невязок решения после 1000 итераций децентрализованного метода на каждом вычислителе.

Сравнение с градиентным методом

$\mu(A) = 10$: $\mu(A) = 82$:





Предложенный децентрализованный метод менее чувствителен к обусловленности матрицы A, чем градиентный метод.

Заключение

1. Задача решения системы уравнений представлена в виде задачи композитной оптимизации.
2. Предложен децентрализованный метод решения задачи композитной оптимизации.
3. Децентрализованный метод сравнивается с градиентным методом.
4. Децентрализованный метод эффективнее градиентного метода для линейных систем с плохо обусловленной матрицей.