Десингуляризация матричных множеств

Хрульков Валентин Андреевич

Аспирант

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: khrulkov.v@gmail.com

Одним из часто встречающихся множеств в задачах оптимизации является множество матриц ограниченного ранга:

$$M_r := \{A \in \mathbb{R}^{n\times m}, \operatorname{rank} A \le r\}. \tag{1}$$

Это множество, однако, не является многообразием – все точки $A \in M_r$ такие, что $\operatorname{rank} A < r$ являются критическими. Таким образом, корректность работы таких стандартных методов оптимизации, как градиентный спуск или метод Ньютона, сложно или невозможно доказать, так как касательное расслоение M_r в особых точках не определено. Для решения этой проблемы мы рассматриваем так называемую десингуляризацию $\widetilde{M_r}$ – множество, заданное уравнениями:

$$\widetilde{M_r} = \{(A, Y) \in \mathbb{R}^{n\times m} \times Gr(m-r, m)\} \tag{2}$$

$$AY = 0.$$

Очевидно, что

$$(A, Y) \in \widetilde{M_r} \Rightarrow A \in M_r,$$

и проекция

$$\pi : (A, Y) \to A,$$

является корректно определенным (сюръективным) отображением.

С вычислительной точки зрения, однако, более удобно определить $\widetilde{M_r}$ следующим образом:

$$\widetilde{M_r} = \{(A, Y) \in \mathbb{R}^{n\times m} \times \mathbb{R}^{m\times(m-r)}\}, \tag{3}$$

$$AY = 0, Y^\top Y = I.$$

и именно этими формулами мы пользуемся в дальнейшем. Имеет место следующая теорема

Теорема 1. _Множество $\widetilde{M_r}$ является гладким многообразием._

Преимуществом работы с $\widetilde{M_r}$ является то, что касательное пространство корректно определено во всех точках и, как показывается в доказательстве Теоремы 1, обладает малой 'кривизной' – константа Липшица для функции проектора на касательное пространство не превосходит 2.

Теперь предположим, что дана некоторая задача оптимизации

$$F(A) \to \min, A \in M_r.$$

Естественным образом ее можно рассмотреть и на $\widetilde{M_r}$:

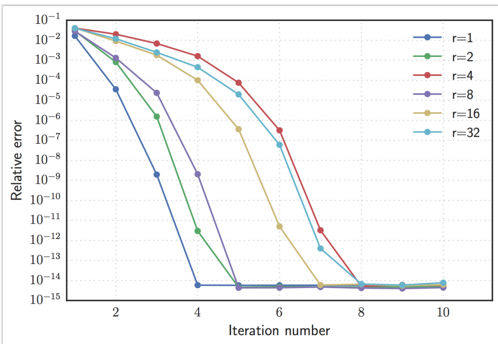
$$F(A) \to \min, (A, Y) \in \widetilde{M_r} \tag{4}$$

Для решения (4) мы применяем метод Ньютона к задаче с ограничениями (из формулы (3)):

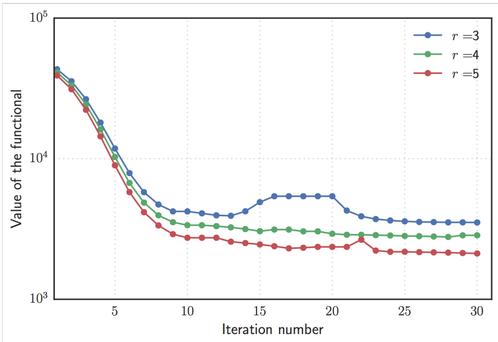
$$\mathcal{L}(A, Y, \Lambda_1, \Lambda_2) = F(A) + \langle \Lambda_1, AY \rangle + \frac{1}{2} \langle \Lambda_2, Y^\top Y - I \rangle,$$

Λ_1, Λ_2 - множители Лагранжа. Полученный алгоритм мы записываем в терминах сингулярного разложения матриц, что приводит к сильному сжатию данных для матриц малого ранга. Y в финальных формулах отсутствует. Результаты тестирования метода для различных функционалов, показывающие его квадратичную сходимость представлены ниже.

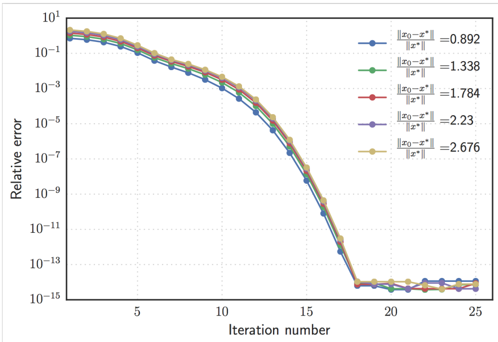
Иллюстрации



Функционал аппроксимации для матрицы 1000×1700 из набора данных MovieLens 100K - начальное условие близко к точному решению



Функционал заполнения для матрицы 1000×1700 из набора данных MovieLens $100K$, метод Ньютона с trust region модификацией



Функционал аппроксимации для матрицы 1000×1700 из набора данных MovieLens 100K, метод Ньютона с trust region модификацией

Литература

1. Khrulkov V. , Oseledets I. Desingularization of bounded-rank matrix sets// https://arxiv.org/abs/1612.03973