



МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М. В. ЛОМОНОСОВА  
Факультет вычислительной математики и кибернетики  
Кафедра математических методов прогнозирования

---

Отчет к третьему практическому заданию по МОМО:  
Выпуклая негладкая, условная и структурная  
ОПТИМИЗАЦИЯ

Студент 517 группы:  
*Оспанов А.М.*

Москва, 2016

# 1 Введение

В данной работе будут выведены вспомогательная функция и ньютоновское направление для метода барьеров, будет реализован модуль llinreg.py с реализацией проксимального метода, субградиентного метода и метода барьеров.

Код написан на языке Python 3 с использованием библиотеки numpy.

## 2 $\phi_\tau(w^+, w^-)$ и вывод $(p_k^+, p_k^-)$

Рассмотрим задачу оптимизации следующего вида:

$$\begin{cases} f(x) \rightarrow \min, & x \in \mathbb{D} \subseteq \mathbb{R}^n \\ g_i(x) \leq 0, & i = 1, \dots, m \end{cases}$$

Тогда вспомогательная функция для метода барьеров будет иметь следующий вид:

$$\phi_\tau(x) = \tau f(x) - \sum_{i=1}^m \log(-g_i(x))$$

Учитывая общий вид, напомним вспомогательную ф-ю для задачи (5):

$$\phi_\tau(w^+, w^-) = \frac{\tau}{2n} \|X(w^+ - w^-) - y\|_2^2 + \tau \lambda \vec{1}(w^+ + w^-) - \sum_{i=1}^m (\log(w_i^+) + \log(w_i^-))$$

Далее, чтобы выписать систему линейных уравнений, задающую ньютоновское направление  $(p_k^+, p_k^-)$ , нужно посчитать градиенты и гессианы:

$$\begin{aligned} \nabla_{w^+} \phi_\tau &= \frac{\tau}{n} X^T (X(w^+ - w^-) - y) + \tau \lambda \vec{1} - \tilde{w}^+, \text{ где } \tilde{w}^+ = \left( \frac{1}{w_1^+}, \dots, \frac{1}{w_d^+} \right) \\ \nabla_{w^-} \phi_\tau &= -\frac{\tau}{n} X^T (X(w^+ - w^-) - y) + \tau \lambda \vec{1} - \tilde{w}^-, \text{ где } \tilde{w}^- = \left( \frac{1}{w_1^-}, \dots, \frac{1}{w_d^-} \right) \\ \nabla_{w^+}^2 \phi_\tau &= \frac{\tau}{n} X^T X + \text{diag}(\tilde{w}_s^+), \text{ где } [\tilde{w}_s^+]_i = \left( \frac{1}{w_i^+} \right)^2, \quad i = 1, \dots, d \\ \nabla_{w^-}^2 \phi_\tau &= \frac{\tau}{n} X^T X + \text{diag}(\tilde{w}_s^-), \text{ где } [\tilde{w}_s^-]_i = \left( \frac{1}{w_i^-} \right)^2, \quad i = 1, \dots, d \\ \nabla_{w^- w^+}^2 \phi_\tau &= -\frac{\tau}{n} X^T X \end{aligned}$$

Обозначим эти функции следующим образом:

$$A = \nabla_{w^- w^+}^2 \phi_\tau$$

$$D_+ = \text{diag}(\tilde{w}_s^+)$$

$$D_- = \text{diag}(\tilde{w}_s^-)$$

$$b^+ = -\nabla_{w^+} \phi_\tau$$

$$b^- = -\nabla_{w^-} \phi_\tau$$

Тогда:

$$\nabla_{w^+}^2 \phi_\tau = -A + D_+$$

$$\nabla_{w^-}^2 \phi_\tau = -A + D_-$$

и система уравнений в этих обозначениях выглядит следующим образом:

$$\begin{bmatrix} -A + D_+ & A \\ A & -A + D_- \end{bmatrix} \begin{bmatrix} p_k^+ \\ p_k^- \end{bmatrix} = \begin{bmatrix} b^+ \\ b^- \end{bmatrix}$$

Просуммировав два уравнения, получим:

$$p_k^+ = D_+^{-1}(b^+ + b^- - D_- p_k^-)$$

Далее подставив полученное выражение во 2е уравнение, получим:

$$(AD_+^{-1} + A - D_-)p_k^- = AD_+^{-1}(b^+ + b^-) - b^-$$

Введем следующие обозначения:

$$\frac{a}{b} = \left( \frac{a_1}{b_1}, \dots, \frac{a_d}{b_d} \right)$$

$$a^2 = (a_1^2, \dots, a_d^2)$$

Если подставить матрицы в изначальном виде, то получим следующие выражения:

$$\begin{aligned} p_k^+ &= w^+ + \left( \frac{1}{w^-} - 2\tau\lambda\vec{1} \right) (w^+)^2 - \left( \frac{(w^+)^2}{(w^-)^2} \right) p_k^- \\ \left[ -\frac{\tau}{n} X^T X \text{diag} \left( \left( \frac{w^+}{w^-} \right)^2 + \vec{1} \right) - \text{diag} \left( \frac{1}{(w^-)^2} \right) \right] p_k^- &= \\ = -\frac{\tau}{n} X^T \left[ X \left( 2w^+ - w^- + \left( \frac{1}{w^-} - 2\tau\lambda\vec{1} \right) (w^+)^2 \right) - y \right] + \tau\lambda\vec{1} - \frac{1}{w^-} \end{aligned}$$

**Оценка  $\alpha$**

$$w_k^+ + \alpha * p_k^+ \geq 0$$

$$w_k^- + \alpha * p_k^- \geq 0$$

Учитывая эти неравенства, получаем:

$$\alpha \leq -\frac{w_i^+}{p_i^+}, i \in I^+, \text{ где } I^+ = \{i | p_i^+ < 0\}$$

$$\alpha \leq -\frac{w_i^-}{p_i^-}, i \in I^-, \text{ где } I^- = \{i | p_i^- < 0\}$$

Следовательно:

$$\alpha_{max} = \min \left\{ \min_{i \in I^+} \left( -\frac{w_i^+}{p_i^+} \right), \min_{i \in I^-} \left( -\frac{w_i^-}{p_i^-} \right) \right\}$$

Если  $I^+ = I^- = \emptyset$ , то  $\alpha_{max} = 1$

Начальную точку  $(w_0^+, w_0^-)$  берем такую, чтобы она лежала далеко от границы и близко к центральному пути. Для нашей задачи подойдет точка  $(w_0^+, w_0^-) > 0$

### 3 Проксимальный метод

В проксимальном методе на шаге  $k$  решается следующая задача:

$$w_{k+1} = \arg \min_w \left\{ f(w_k) + \nabla f(w_k)^T (w - w_k) + \frac{L}{2} \|w - w_k\|_2^2 + h(w) \right\}$$

Эта задача является сепарабельной. То можно решить задачу для каждой координаты отдельно.

В нашем случае, т.е. когда  $h(x) = \lambda \|x\|_1$ , задача представима в следующем виде:

$$\frac{L}{2} w_i^2 + \delta_i w_i + \lambda |w_i| \rightarrow \min_{w_i}, \text{ где } \delta = \nabla f(w)$$

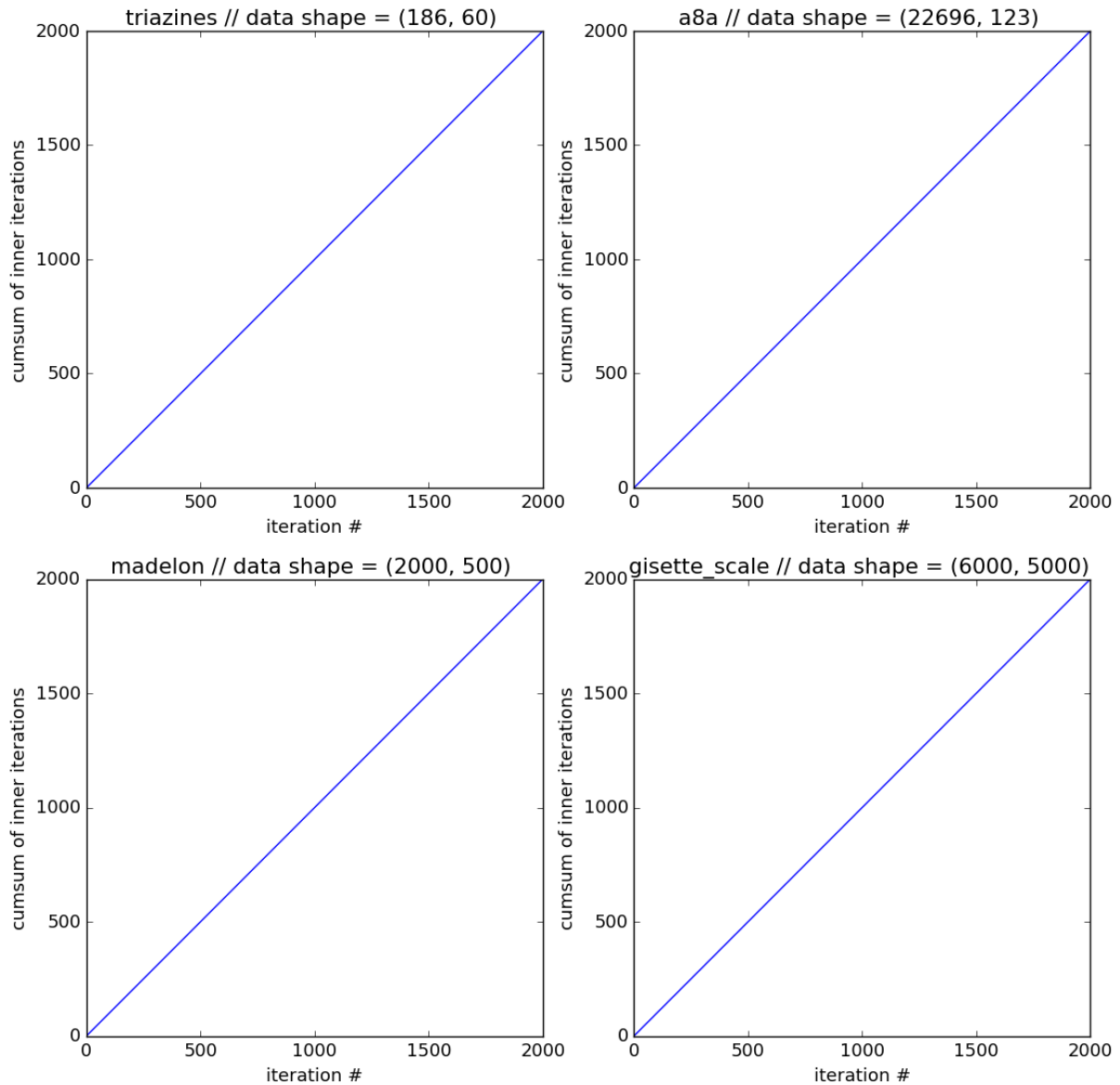
Решив эту задачу, получим:

$$w_i = \begin{cases} \frac{-\delta_i - \lambda}{L}, & \delta_i < -\lambda \\ 0, & |\delta| \leq \lambda \\ \frac{-\delta_i + \lambda}{L}, & \delta_i > \lambda \end{cases}$$

И в проксимальном и в субградиентном методах нужно брать начальную точку такую, чтобы субградиент вычислялся единственным образом. Для нашей задачи подойдет точка  $w_0 > 0$

## 4 Экспериментальное сравнение трех реализованных методов

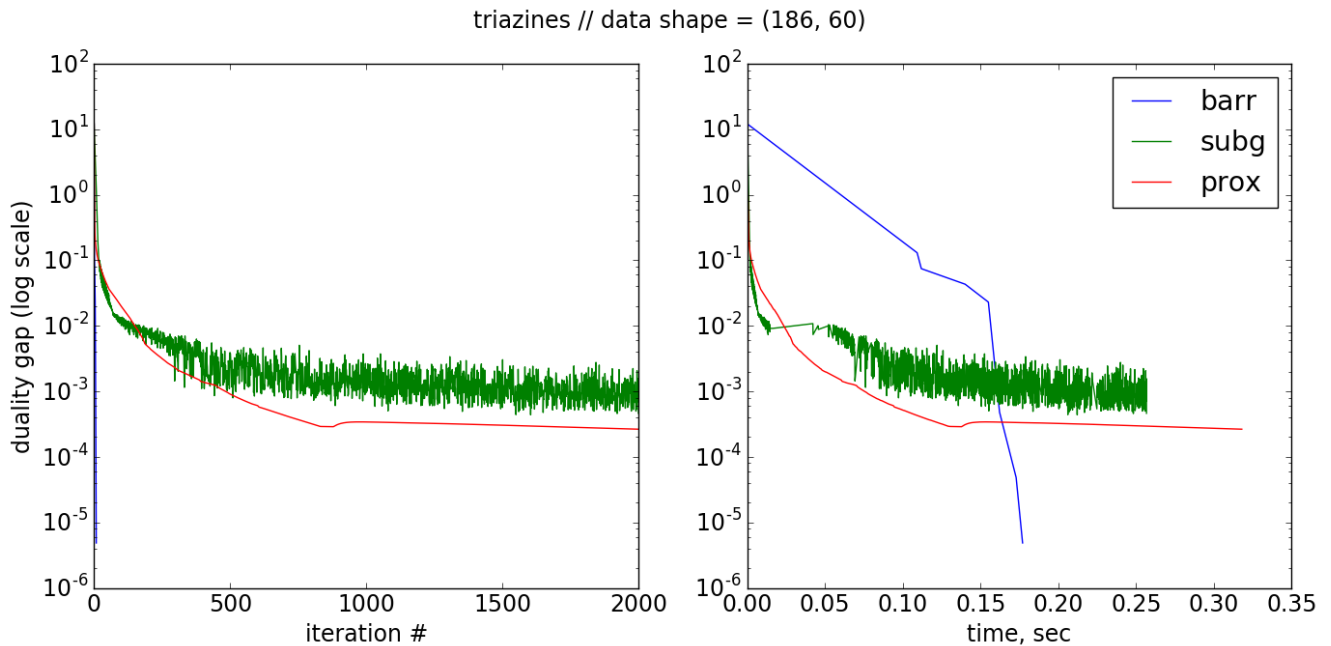
Рассмотрим график суммарного числа итераций одномерного поиска в зависимости от номера итерации.



Из графика видно, что для любых данных разных размеров суммарное число итераций одномерного поиска линейное. Т.е. в каждой итерации проксимального метода выполняется всего одна итерация одномерного поиска. Таким образом график доказывает, что внутренняя итерация в среднем выполняется не более 2 раз.

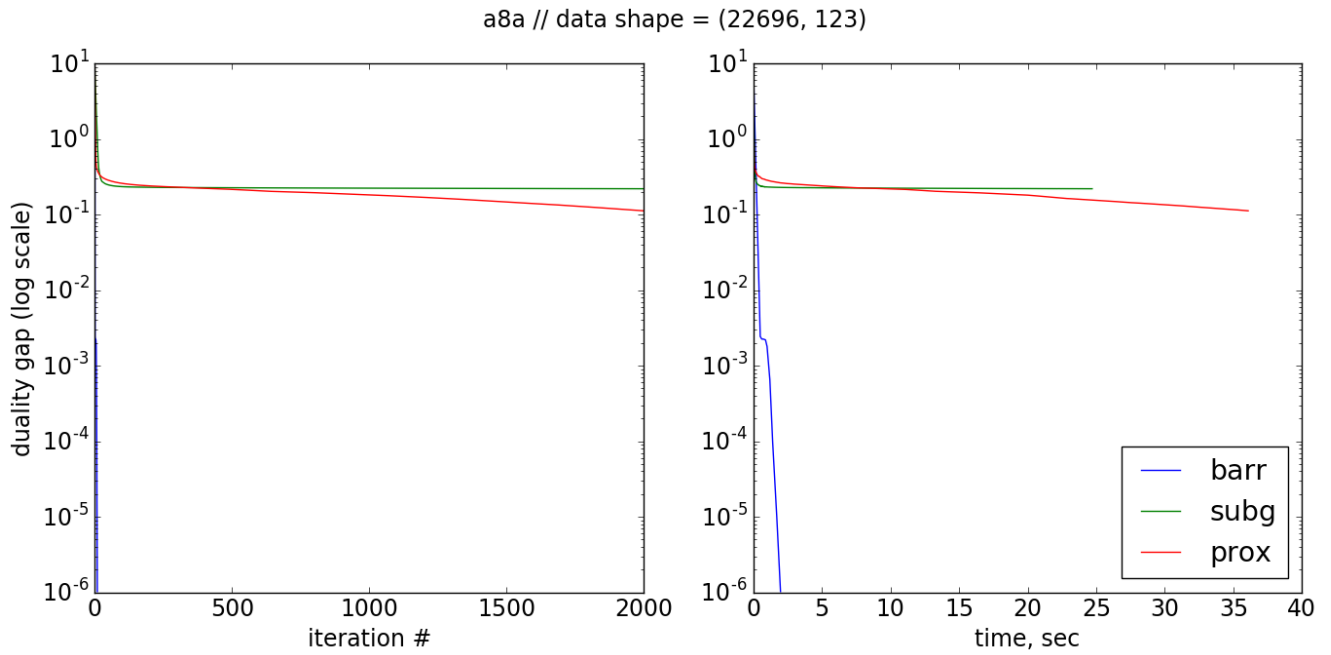
Далее рассмотрим графики зазоров для трех методов при максимальном количестве итераций равным 2000 и точности, раному  $10^{-5}$ .

Расположим графики в порядке увеличения количества признаков.

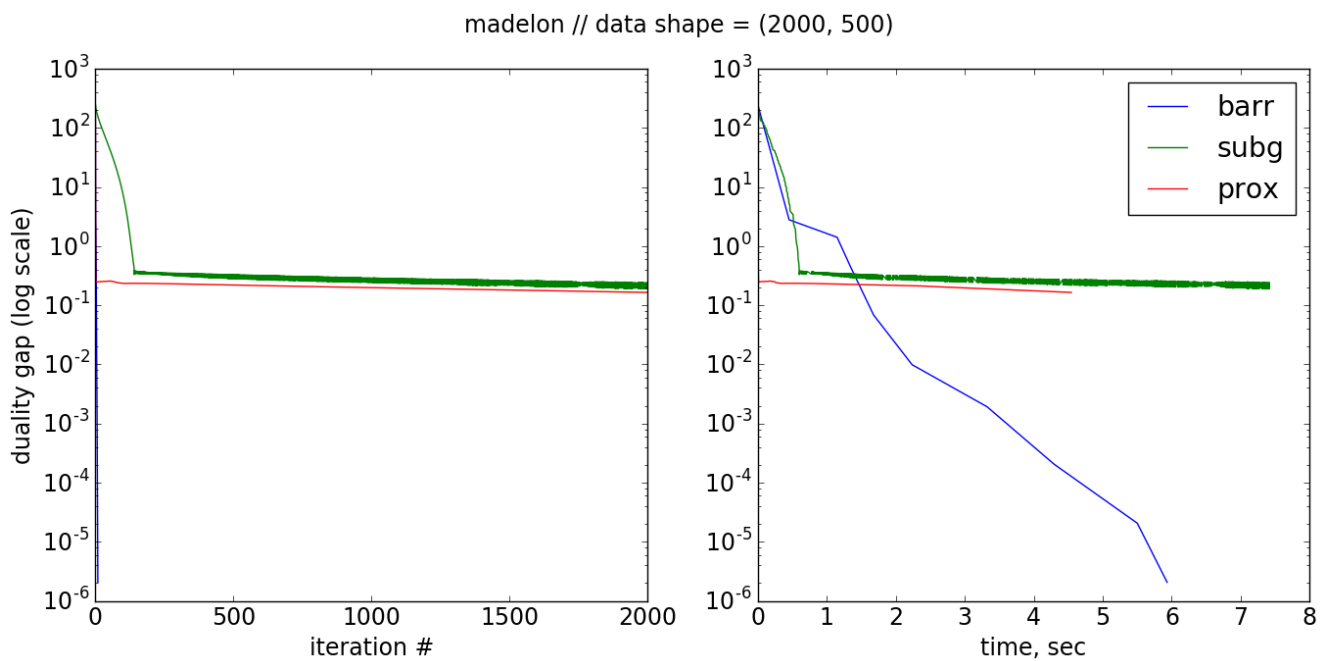


Из графика видно, что проксимальный метод сходится до определенного значения зазора – асимптоте и далее не оптимизирует функционал. Также и с субградиентным методом. Зазор в субградиентном методе в целом уменьшается, но т.к. субградиентный метод строит не обязательно релаксационную последовательность, мы видим колебания. Метод барьеров сошелся очень быстро по итерациям и до точности  $10^{-2}$  и до точности  $10^{-5}$ . Но по времени быстрее до  $10^{-2}$  сошелся метод субградиентов, а до  $10^{-5}$  сошелся только метод барьеров.

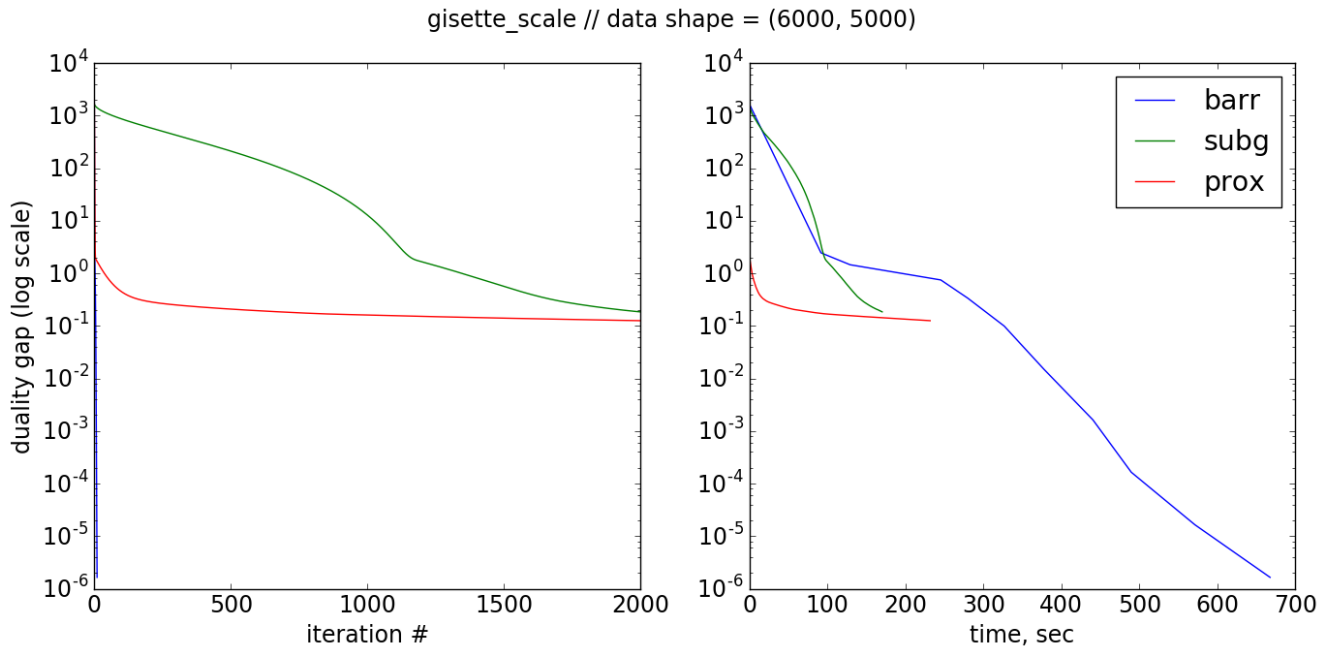
Следующая выборка интересна тем, что метод субградиентного спуска и проксимальный метод не показали хорошую точность. И соответственно быстрее всех до точности  $10^{-2}$  и  $10^{-5}$  сошелся метод барьеров и по времени и по итерациям.



Следующий случай примерно аналогичен предыдущему.



Но для данных с большим количеством признаков методы показывают чуть другие результаты. Субградиентный метод падает до ассимптоты медленнее. Но проксимальный метод уменьшает зазор также очень быстро до ассимптоты. Метод барьеров также очень быстро сходится по итерациям, но очень долго по времени.



## 5 Заключение

В работе были выведены вспомогательная функция и ньютоновское направление для метода барьеров, исследованы проксимальный метод, субградиентный метод и метод барьеров. В итоге работы мы получили следующие результаты:

1. Метод барьеров очень быстр по итерациям, но не по времени. Чем больше признаков, тем больше время работы.
2. Проксимальный и субградиентный методы имеют очень маленькую точность, равную  $10^{-1}$  в общем случае.
3. Если нужна большая точность, то лучше использовать метод барьеров, т.к. он достигает хороших точностей.
4. Если большая точность не нужна, то лучше использовать проксимальный метод, т.к. он сходится меньше чем за секунду до своей асимптоты.
5. Субградиентный оказался не нужным на фоне других методов. Этот метод не имеет ни быструю сходимость при маленьких точностях, ни большую точность.