1. Теоретическая часть

1.1 Функция и поиск направления:

Минимизируемая функция:

$$f_t = t\left(\frac{1}{2}||Ax-b||^2 + \lambda\left\langle 1_n, u\right\rangle\right) - \sum_{i=1}^{n}\left(\ln\left(u_i + x_i\right) + \ln\left(u_i - x_i\right)\right)$$

Градиент:

$$\frac{\partial f_t}{\partial x} = t\left(A^T Ax - \langle b, A\rangle\right) - \left(\frac{1}{u+x} - \frac{1}{u-x}\right)$$

$$\frac{\partial f_t}{\partial u} = t\lambda - \left(\frac{1}{u+x} + \frac{1}{u-x}\right)$$

Гессиан:

$$\frac{\partial^2 f_t}{\partial x^2} = tA^T A + diag\left(\frac{1}{(u+x)^2} + \frac{1}{(u-x)^2}\right)$$

$$\frac{\partial^2 f_t}{\partial x \partial u} = \frac{\partial^2 f_t}{\partial u \partial x} = diag\left(\frac{1}{(u+x)^2} - \frac{1}{(u-x)^2}\right)$$

$$\frac{\partial^2 f_t}{\partial u^2} = diag\left(\frac{1}{(u+x)^2} + \frac{1}{(u-x)^2}\right)$$

Будем искать направление, решая систему уравнений:

$$\begin{pmatrix} \frac{\partial^2 f_t}{\partial x^2} & \frac{\partial^2 f_t}{\partial u \partial x} \\ \frac{\partial^2 f_t}{\partial x \partial u} & \frac{\partial^2 f_t}{\partial u^2} \end{pmatrix} \begin{pmatrix} d^x \\ d^u \end{pmatrix} = -\begin{pmatrix} \frac{\partial f_t}{\partial x} \\ \frac{\partial f_t}{\partial u} \end{pmatrix}$$

С помощью разложения Холецкого.

1.2 Максимальная длина шага α

Вывод полностью повторяет вывод, приведенный в условии задания. Остается подставить в него границы, возникающие в задаче:

$$g_i\left(x\right) = -\left\langle e_i + e_{n+i}, x\right\rangle$$

$$g_{n+i}\left(x\right) = \left\langle e_i - e_{n+i}, x\right\rangle$$

при всех $i = 1 \dots n$. Здесь e_j - вектор длины $2n$, состоящий из 0 с 1 на j-ой позиции. Таким образом:

$$\alpha_1^{\max} = \min_{i \in I_1} \frac{-x_i - u_i}{d_i^x + d_i^u}, I_1 = \{1 \le i \le n : -d_i^x - d_i^u > 0\}$$

$$\alpha_2^{\max} = \min_{i \in I_2} - \frac{x_i - u_i}{d_i^x - d_i^u}, I_2 = \{1 \le i \le n : d_i^x - d_i^u > 0\}$$

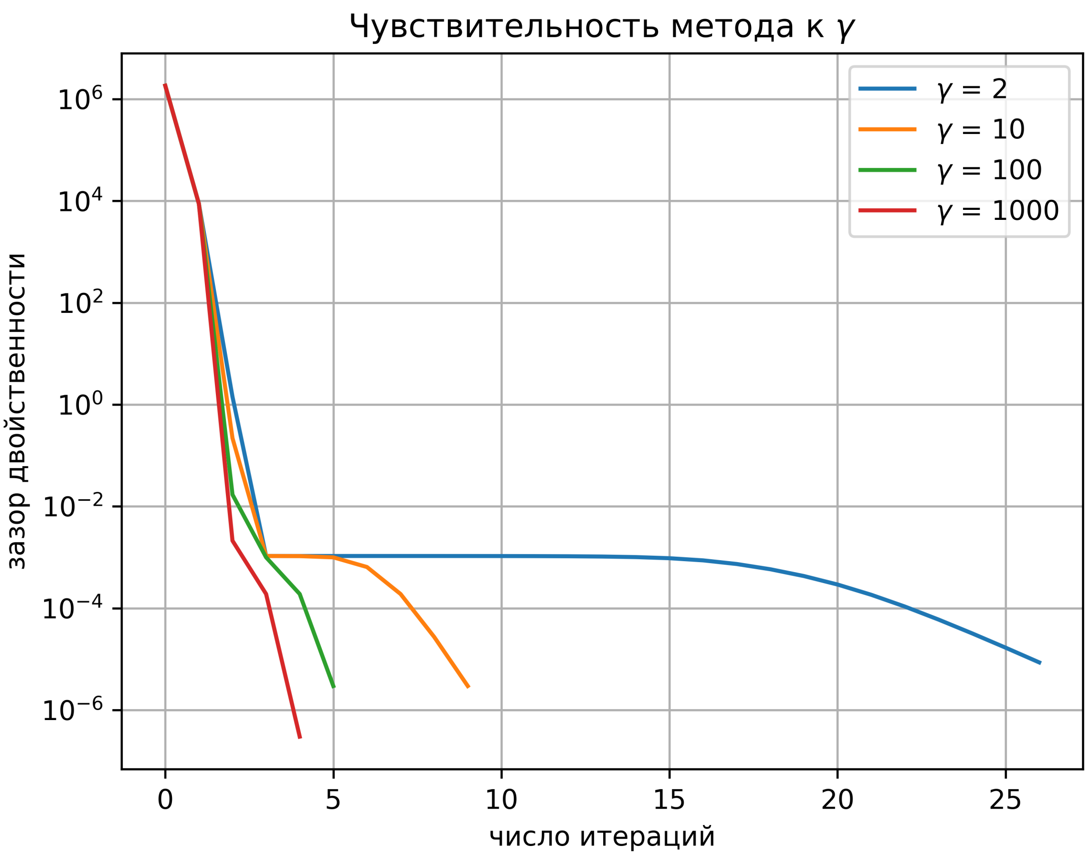
$$\alpha_0 = \min\{1, \theta\alpha_1^{\max}, \theta\alpha_2^{max}\}$$

2. Эксперименты

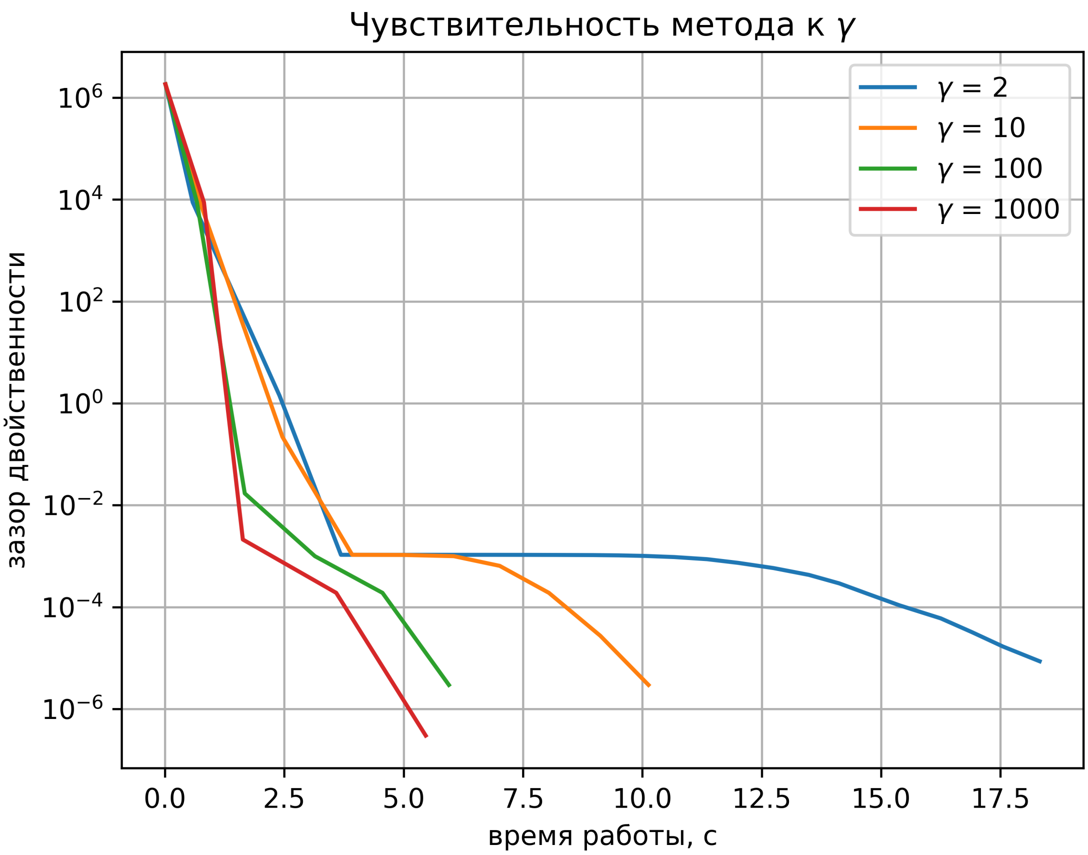
2.1 Исследование чувствительности метода к выбору γ и ϵ_{inner}

Матрица A и вектор b берутся из датасета **w8a**. Начальный вектор x_0 выбирается случайно из равномерного распределения на $[0;1]$, так как смысл LASSO в том, чтобы уменьшить координаты вектора x. А вектор u_0 выбирается равным 1_n, чтобы выполнить условия на границу в методе барьеров. Коэффициент регуляризации λ = $1/m$

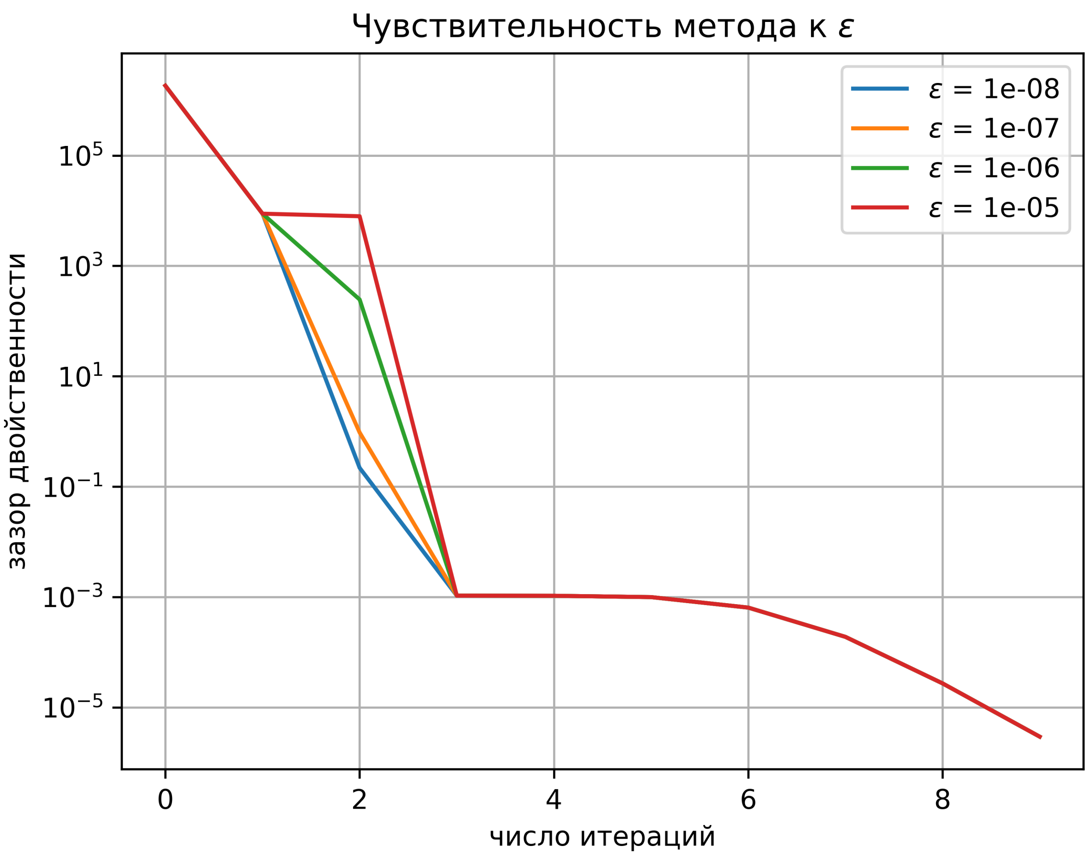
Полученные графики:



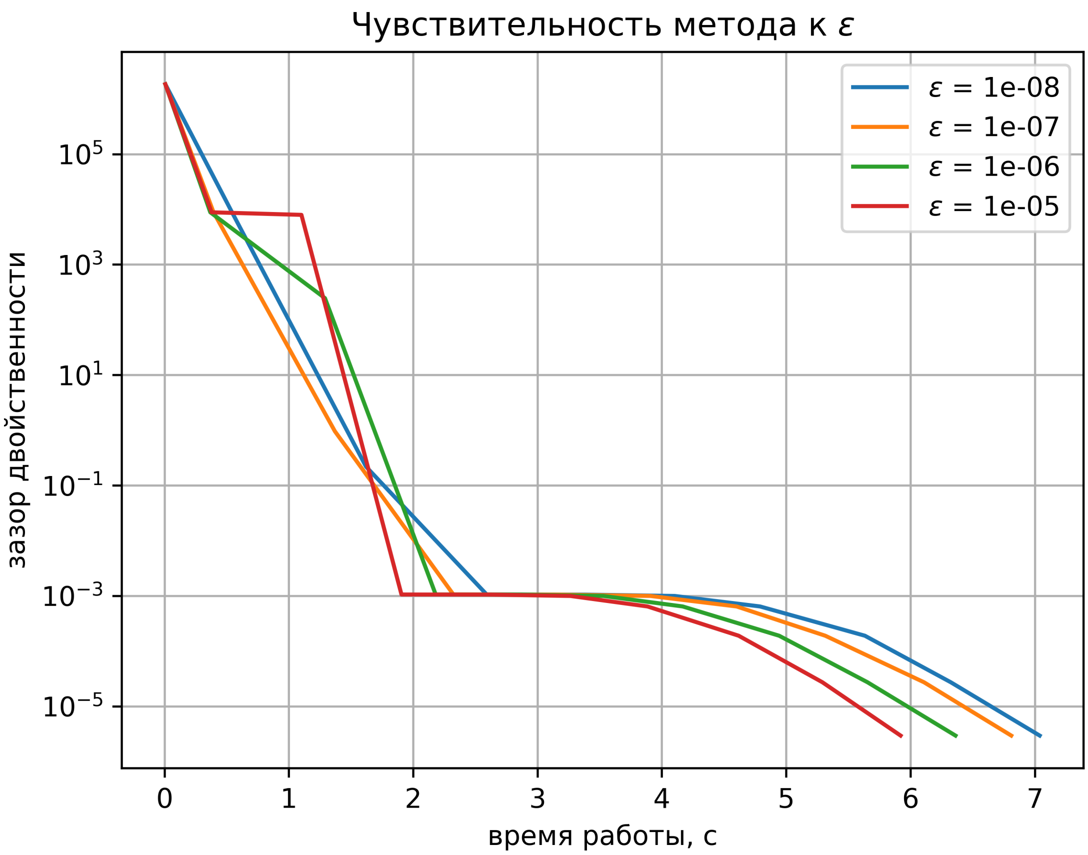
1.


Чувствительность метода к γ
γ = 2
γ = 10
γ = 100
γ = 1000
зазор двойственности
время работы, с

2.


Чувствительность метода к ε
ε = 1e-08
ε = 1e-07
ε = 1e-06
ε = 1e-05
зазор двойственности
число итераций

3.

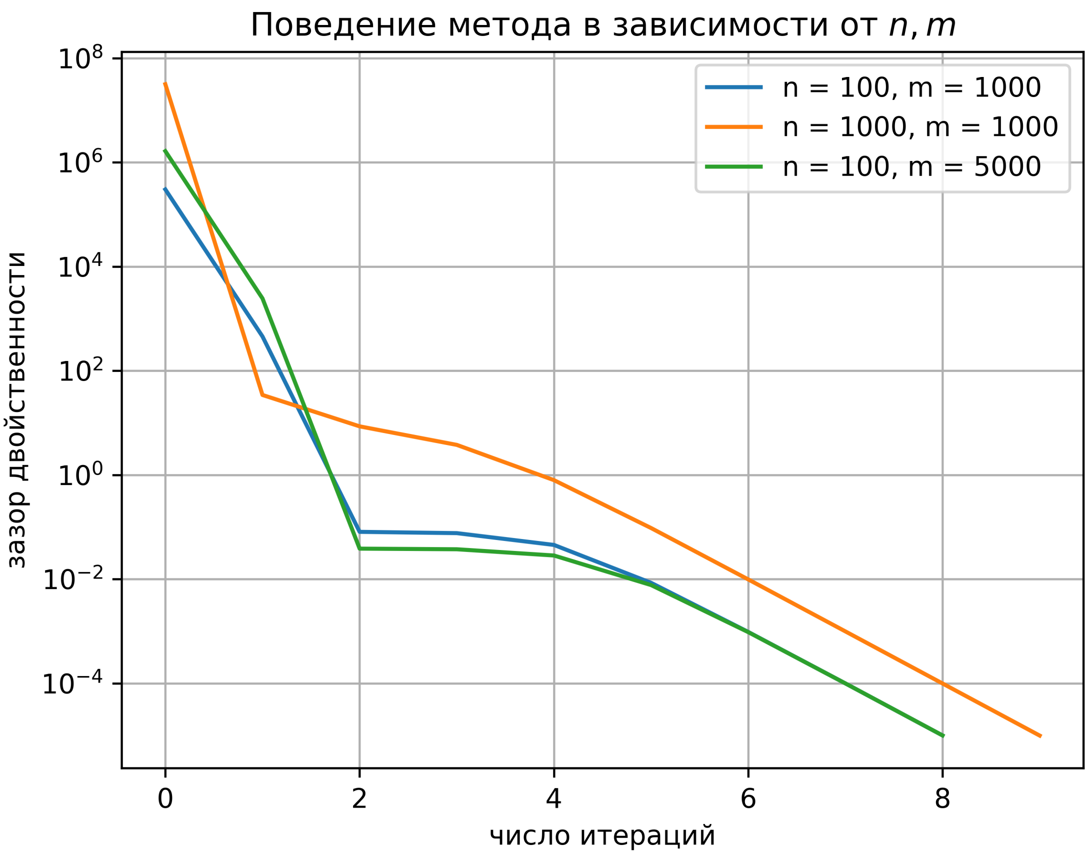


4.

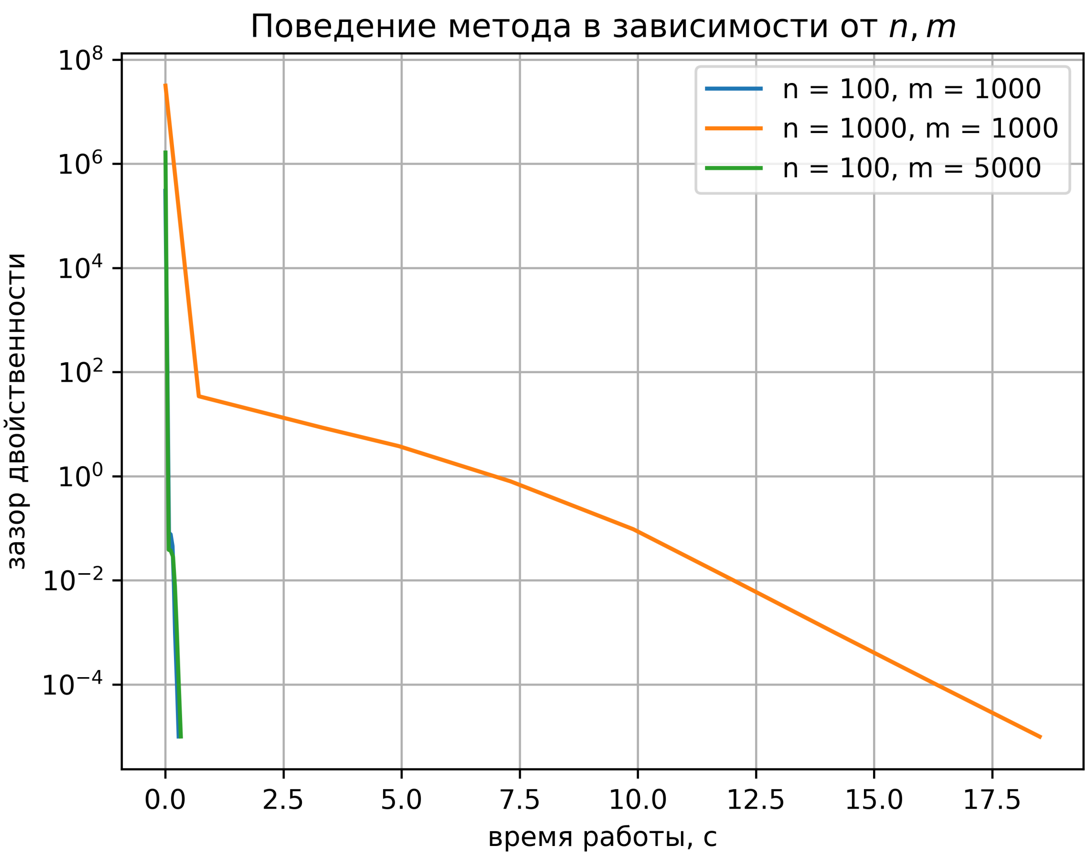
Вывод: при больших γ метод сходится быстрее как по итерациям, так и по времени работы. При высокой внутренней точности методу требуется меньше итераций для схождения, но это увеличивает временные затраты на внутренние итерации и не гарантирует ускорения сходимости всего метода.

2.2 Исследование поведения метода для различных значений размера выборки, размерности пространства и коэффициента регуляризации.

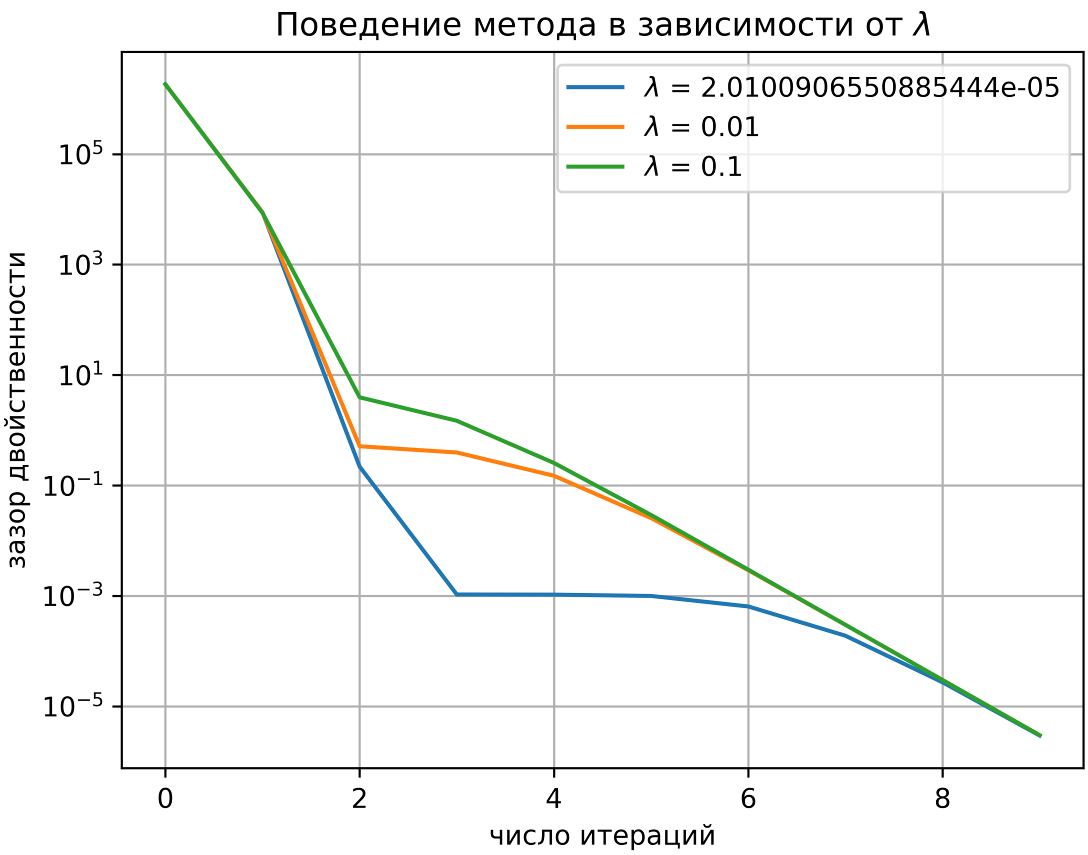
Для исследования зависимости поведения метода от λ использован датасет **w8a**. Для исследования зависимости от n, m использованы случайно сгенерированные данные. Полученные графики:


Поведение метода в зависимости от n, m
n = 100, m = 1000
n = 1000, m = 1000
n = 100, m = 5000
зазор двойственности
число итераций
10^8
10^6
10^4
10^2
10^0
10^{-2}
10^{-4}
0
2
4
6
8

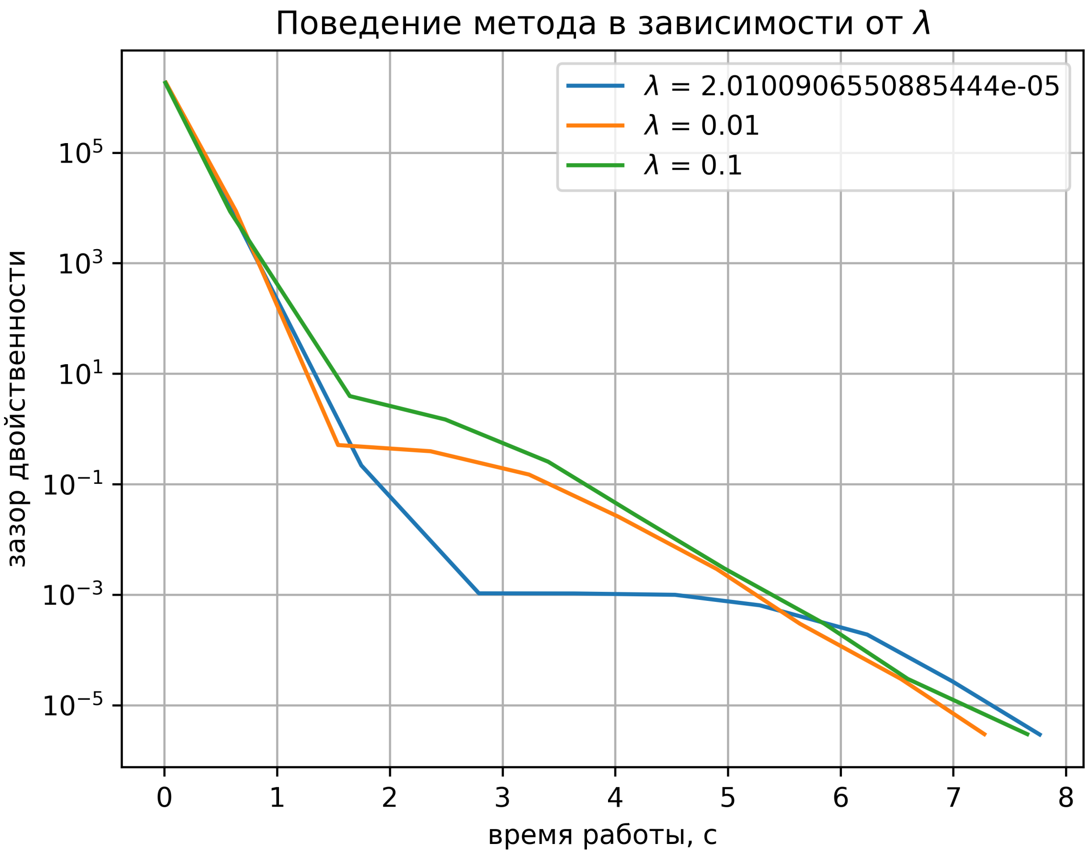
1.


Поведение метода в зависимости от n, m
n = 100, m = 1000
n = 1000, m = 1000
n = 100, m = 5000
зазор двойственности
время работы, с
10^8
10^6
10^4
10^2
10^0
10^{-2}
10^{-4}
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

2.



3.



4.

Вывод: скорость сходимости метода сильно зависит от размерности пространства и мало от числа обучающих примеров. По графику зависимости от λ видно, что при маленьких λ метод быстро достигает точности 10^{-3}, а потом медленно доходит до желаемой точности. При больших же λ метод сходится почти с одинаковой скоростью. При этом время достижения желаемой точности 10^{-5} практически не различается для всех рассмотренных λ.