Лабораторная работа 3:
Метод барьеров

Выполнила Симакова Елизавета,
студентка 1 курса магистратуры ВБИБ НИУ ВШЭ Санкт-Петербург

24 июня 2025 г.

Содержание

1 Теоретическое задание

1. **Минимизируемая функция имеет вид:**

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

Система уравнений для ньютоновского направления $d_k = (d^x, d^u)$. Ньютоновское направление находится из решения системы:

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

Компоненты системы линейных уравнений.
Градиент:

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

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

Гессиан:

$$\frac{\partial^2 f_t}{\partial x^2} = tA^\top A + \operatorname{diag}\left(\frac{1}{(u_i+x_i)^2} + \frac{1}{(u_i-x_i)^2}\right)$$

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

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

Для эффективного решения системы линейных уравнений можно воспользоваться методом Холецкого, так как гессиан симметричен и положительно определён в области допустимых точек. Достоинства данного метода: стабильность, скорость. Недостатки: требует строгой положительной определённости.

2. **Максимальная допустимая длина шага** α

$$\alpha_1^{\max} = \min_{i\in I_1} \frac{-(x_i+u_i)}{d_i^x + d_i^u}, I_1 = \{i \mid -d_i^x - d_i^u > 0\}$$

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

$$\alpha_{\max} = \min\left(\alpha_1^{\max}, \alpha_2^{\max}\right).$$

$$\alpha_0 = \min\left(1, \theta\alpha_1^{\max}, \theta\alpha_2^{\max}\right),$$

3. **Выбор начальной точки** (x_0, u_0).

$$x_0 = \mathbf{0}_n, \quad u_0 = c * \mathbf{1}_n,$$

где $\mathbf{1}_n$ — вектор из единиц, $c > 0$.

2 Эксперимент

Описание эксперимента

Целью данного эксперимента является исследование поведения метода логарифмических барьеров для задачи LASSO при различных параметрах алгоритма и характеристиках задачи.

Эксперимент состоит из двух частей:

1. Чувствительность метода к параметрам.

   Исследуется влияние следующих гиперпараметров на сходимость метода:

   (a) параметр γ, определяющий скорость роста параметра t_k в барьерной функции;

   (b) параметр $\varepsilon_{\text{inner}}$, задающий точность при решении вспомогательной задачи методом Ньютона.

2. Поведение метода при различных характеристиках задачи.

   Изучается, как изменяется поведение метода при варьировании:

   (a) размерности пространства n,

   (b) размера выборки m,

   (c) коэффициента регуляризации λ.

Основные параметры эксперимента

Тестовая задача: задача LASSO-регрессии с синтетическими данными. Матрица A, вектор b со значениями из стандартного нормального распределения, $\lambda = \frac{1}{m}$. Начальная точка:

$$x_0 = \text{равномерно из } [0,1]^n, \quad u_0 = \mathbf{1}_n.$$

В экспериментах варьируются следующие параметры:

1. $\gamma \in \{1, 5, 10, 50, 100\}$,
2. $\varepsilon_{\text{inner}} \in \{10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}\}$,
3. размерность пространства $n \in \{10, 50, 100, 200, 500\}$,
4. число наблюдений $m \in \{20, 100, 200, 500, 1000\}$,
5. коэффициент регуляризации $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$.

Для каждого набора параметров метод барьеров запускается один раз.

Результаты визуализируются в виде логарифмических графиков зависимости зазора двойственности от количества итераций и от времени.

Результаты эксперимента

Ниже представлены результаты эксперимента в виде 10 графиков.

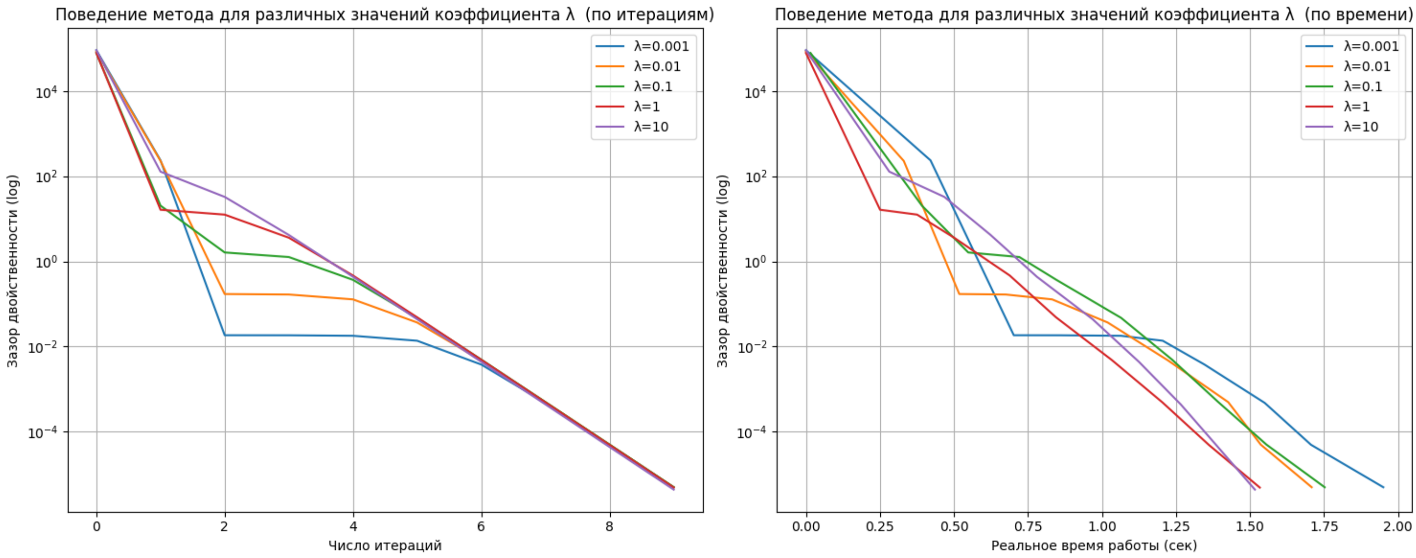

Рис. 1: Влияние коэффициента регуляризации λ на поведение метода. При малых значениях λ (например, $\lambda = 0.001$) метод быстро достигает высокой точности, но медленно желаемой. При бОльших значениях методу требуется меньше времени на достижение заданной точности. При этом количество итераций равное, время при разным параметрах отличается в пределах 0.5 секунды.

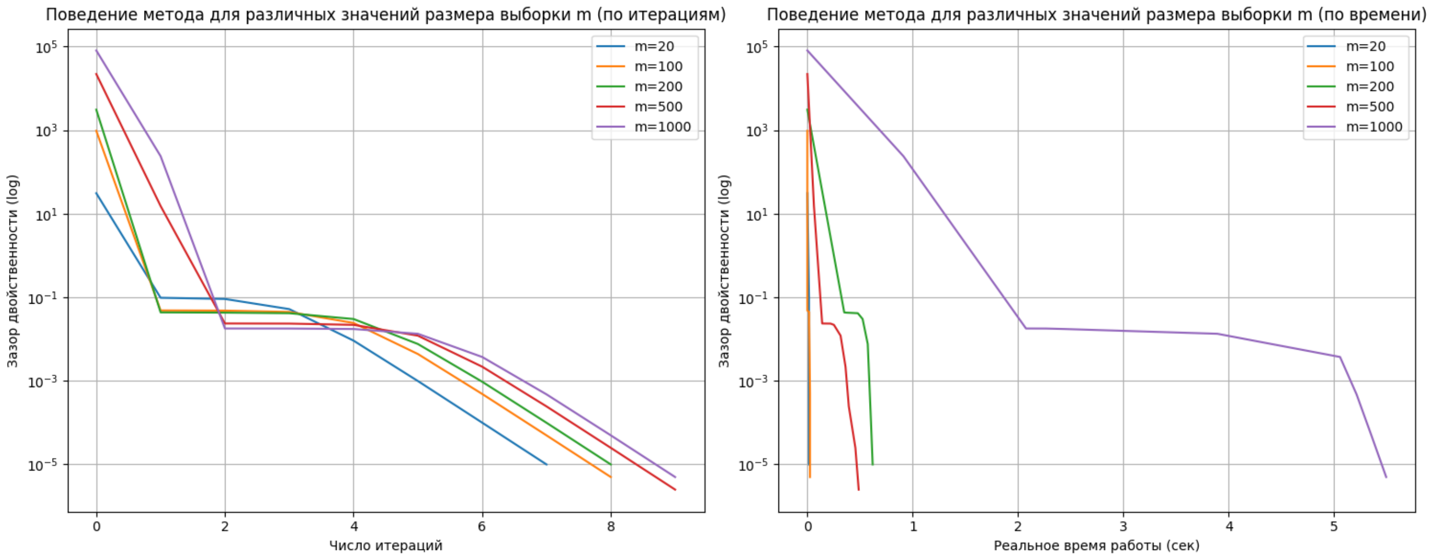

Рис. 2: Влияние размера выборки m на сходимость метода. На левом графике видно, что по мере роста m поведение становится более плавным и равномерным, число итераций увеличилось незначительно. На правом графике видно, что увеличение m существенно растягивает реальное время работы: при $m = 20$–100 метод завершает весь цикл за доли секунды, в то время как при $m = 1000$ требуется уже порядка 5–5.5 секунд.

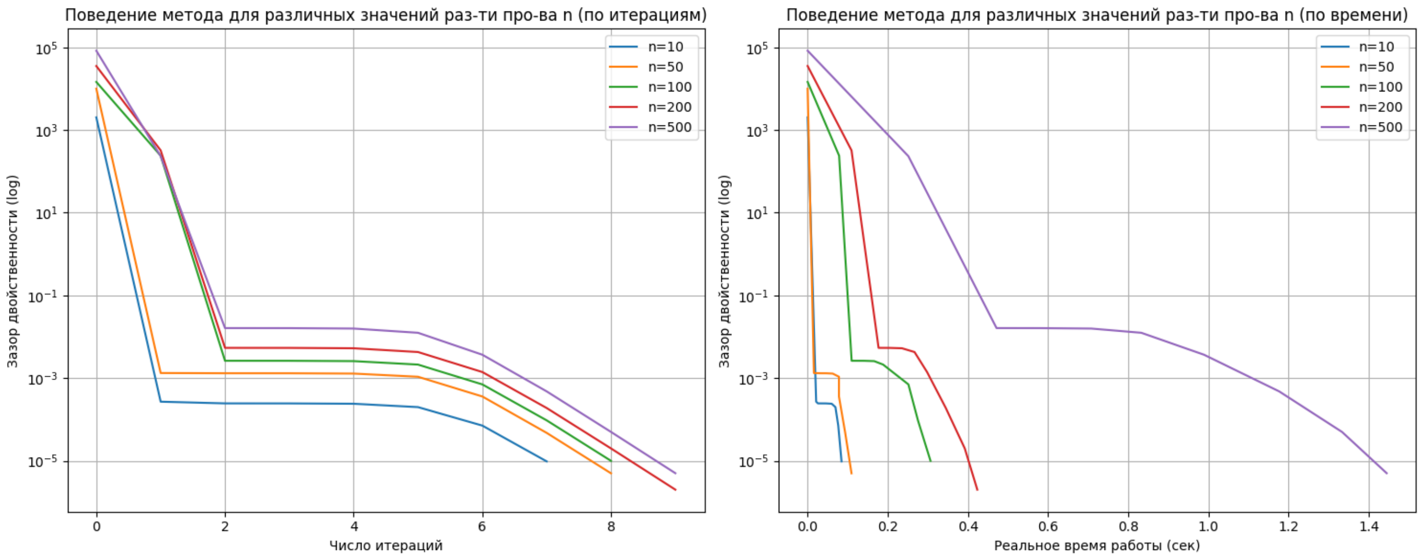

Рис. 3: Влияние размерности пространства n на поведение метода. На левом графике видно, что при увеличении размерности пространства n методу требуется больше итераций для достижения заданной точности. На правом графике эффект ещё более заметен: рост размерности существенно увеличивает затраты времени. Особенно резко возрастает время при переходе от $n = 200$ к $n = 500$, что указывает на возрастающую вычислительную сложность метода при увеличении размерности задачи.

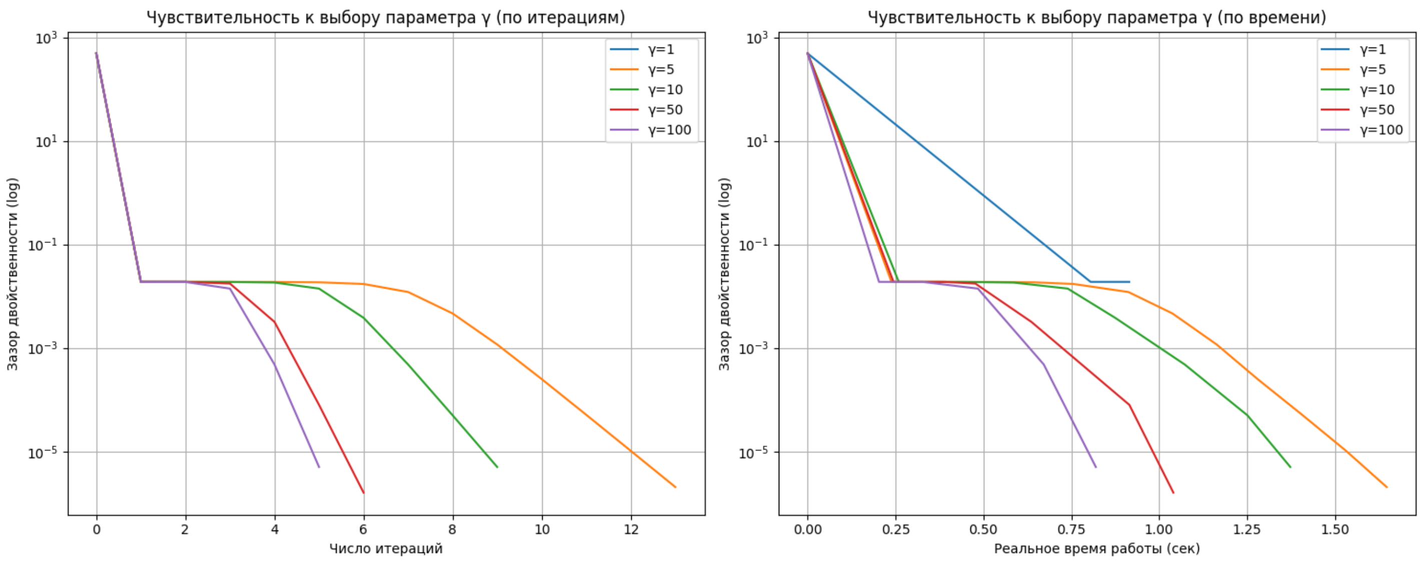

Рис. 4: Чувствительность метода к выбору параметра γ. На левом графике видно, что слишком малые значения γ ($\gamma = 1$) приводят к плохой точности метода, тогда как увеличение γ ее увеличивает. Большие значения γ ($\gamma = 100$) улучшают поведение метода по времени (правый график): быстрое уменьшение зазора по числу итераций. Оптимальным значением по времени и по точности в данном эксперименте оказалось значение $\gamma = 50$.

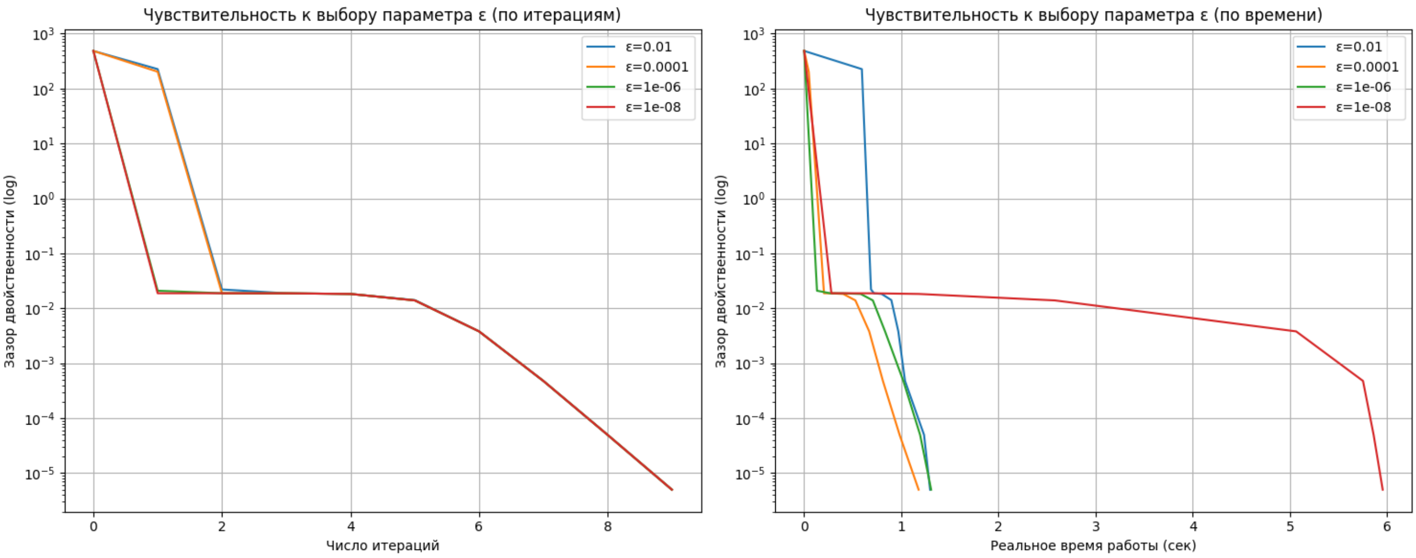

Рис. 5: Чувствительность метода к точности решения внутренней задачи $\varepsilon_{\text{inner}}$. На левом графике видно, что при уменьшении $\varepsilon_{\text{inner}}$ поведение метода практически не меняется — все графики практически совпадают. Однако на правом графике видно, что уменьшение $\varepsilon_{\text{inner}}$ существенно увеличивает время работы метода. Таким образом, чрезмерно высокая точность решения внутренней задачи не даёт выигрыша в числе итераций, но приводит к избыточным затратам времени. Для практических целей разумно выбирать умеренные значения $\varepsilon_{\text{inner}}$.

Проанализировав полученные результаты можно сделать следующие наблюдения и выводы:

1. **Влияние коэффициента регуляризации λ.** Так как увеличение параметра λ приводит к большей точности, то оптимальным значением при данном условиях является λ порядка 10.

2. **Влияние размера выборки m.** Метод достигает большей точности при больших значениях m, но увеличивается скорость работы. В данном эксперименте лучше себя показало значение $m = 500$.

3. **Влияние размерности пространства n.** Увеличение n приводит к росту числа итераций, необходимых для достижения заданного зазора, и ещё более заметному росту времени работы. Задача очень сильно зависит от размерности пространства.

4. **Чувствительность к параметру γ.** Слишком малые γ (1) дают медленную сходимость, тогда как увеличение до $\gamma = 50$–100 ускоряет процесс. Оптимум в наших условиях лежит около $\gamma = 50$–100.

5. **Чувствительность к точности внутренней задачи $\varepsilon_{\text{inner}}$.** Снижение $\varepsilon_{\text{inner}}$ практически не влияет на число итераций, но значительно увеличивает время решения и не даёт выигрыша в сходимости. Значение $\varepsilon_{\text{inner}=0.01}$ также не дает выиграша по времени. Поэтому целесообразно выбирать умеренные значения $\varepsilon_{\text{inner}} \sim 10^{-4}$–$10^{-6}$.