

Московский государственный университет им. М. В. Ломоносова  
Факультет вычислительной математики и кибернетики

Отчет по заданию

**Методы оптимизации для  
для  $L_1$ -регуляризованной линейной регрессии**

Выполнил студент 317 группы  
Измайлов Павел Алексеевич

Москва, 5 ноября 2015

# Содержание

<b>1</b>	<b>Описание проделанной работы</b>	<b>3</b>
<b>2</b>	<b>Рассматриваемая задача и ее эквивалентные формулировки</b>	<b>3</b>
2.1	Эквивалентная формулировка задачи и двойственная задача . . . . .	3
2.1.1	Связь решений прямой и двойственной задачи . . . . .	4
2.2	Эквивалентная формулировка в виде задачи гладкой условной оптимизации . . . . .	5
2.3	Оценка на зазор между решениями прямой и двойственной задачи . . .	5
<b>3</b>	<b>Рассматриваемые методы</b>	<b>6</b>
3.1	Прямой метод барьерных функций . . . . .	6
3.1.1	Эффективный метод решения системы 10 . . . . .	7
3.2	Прямо-двойственный метод внутренней точки . . . . .	8
3.2.1	Эффективный метод решения системы 14 . . . . .	9
3.3	Проксимальный метод . . . . .	10
<b>4</b>	<b>Эксперименты</b>	<b>11</b>
4.1	Небольшие задачи . . . . .	11
4.2	Задачи средних размеров . . . . .	12
<b>5</b>	<b>Выводы</b>	<b>13</b>

# 1 Описание проделанной работы

В данном отчете содержатся результаты экспериментов, проведенных мной в соответствии со вторым заданием по спецкурсу «методы оптимизации в машинном обучении», а также вывод всех требуемых формул. Мной были реализованы прямой метод барьерных функций, прямо-двойственный метод и проксимальный метод для задачи  $L_1$ -регуляризованной логистической регрессии. Для обоих методов внутренней точки (прямой метод и прямо-двойственный метод) был предложен способ эффективного решения системы уравнений для выбора очередного направления оптимизации.

## 2 Рассматриваемая задача и ее эквивалентные формулировки

В данной работе сравниваются результаты работы различных методов оптимизации на задаче  $L_1$ -регуляризованной линейной регрессии. Целевая функция этой задачи имеет вид

$$F(w) = \frac{1}{2} \|t - Xw\|^2 + \lambda \|w\|_1, \quad (1)$$

где  $(X, t) = \{x_n, t_n\}_{n=1}^N$  — набор данных,  $x_n \in \mathbb{R}^D$  — вектор признаков  $n$ -го объекта,  $t_n \in \mathbb{R}$  — значение неизвестной функции (которую требуется приблизить) на этом объекте,  $w \in \mathbb{R}^D$  — вектор весов,  $\lambda \geq 0$ .

### 2.1 Эквивалентная формулировка задачи и двойственная задача

Рассмотрим следующую эквивалентную формулировку рассматриваемой задачи.

$$\begin{aligned} \frac{1}{2} z^T z + \lambda \|w\|_1 \rightarrow \min_{z, w}, \\ Xw - t = z. \end{aligned} \quad (2)$$

Данная задача является негладкой задачей выпуклой оптимизации. Выведем двойственную к ней задачу.

Функция Лагранжа данной задачи имеет вид

$$L(z, w, \mu) = \frac{1}{2} z^T z + \lambda \|w\|_1 + \langle \mu, Xw - t - z \rangle.$$

Минимизируем эту функцию по переменной  $z$ . Заметим, что по переменной  $z$  функция  $L(z, w, \mu)$  квадратична, причем матрица квадратичной формы положительно определена (это единичная матрица). Поэтому у нее существует единственный минимум при фиксированных  $\mu, w$  в точке  $z$ , где градиент по  $\nabla_z L(z, w, \mu) = 0$ .

$$0 = \nabla_z (L(z, w, \mu)) = z - \mu \Rightarrow z = \mu.$$

Теперь проведем минимизацию по переменной  $w$ . Заметим, что функцию  $L$  можно переписать в виде

$$\begin{aligned} L(z, w, \mu) &= \lambda \|w\|_1 + \langle \mu, Xw \rangle + \varphi(z, \mu) = \\ &= \sum_{d=1}^D (\lambda |w_d| + c_d w_d) + \varphi(z, \mu), \end{aligned}$$

где  $c = X^T \mu$ ,  $\varphi(z, \mu)$  — некоторая функция от  $z$  и  $\mu$ . Если найдется такой индекс  $d$ , что  $|c_d| > \lambda$ , то для вектора  $w = (0, 0, \dots, 0, k, 0, \dots, 0)$ , в котором только  $n$ -ый элемент отличен от нуля будем иметь

$$L(z, w, \mu) = \lambda |k| + c_n k + \varphi(z, \mu).$$

Эта функция не ограничена снизу, т.к. при  $\text{sgn}(k) = -\text{sgn}(c_n)$  имеем

$$L(z, w, \mu) = |k|(\lambda - |c_n|) + \varphi(z, \mu) \xrightarrow{|k| \rightarrow \infty} -\infty.$$

Если же  $\lambda \geq \|X^T \mu\|_\infty$ , то имеем

$$L(z, w, \mu) = \sum_{d=1}^D (\lambda |w_d| + c_d w_d) + \varphi(z, \mu) \geq \varphi(z, \mu) = L(z, 0, \mu).$$

Таким образом, минимум по переменной  $w$  достигается при  $w = 0$ .

Отметим, что  $\frac{1}{2} \mu^T \mu + \langle \mu, -t - \mu \rangle = -\frac{1}{2} \mu^T \mu - \mu^T t$ .

Итак, двойственная задача имеет вид

$$\begin{aligned} -\frac{1}{2} \mu^T \mu - \mu^T t &\rightarrow \max_{\mu}, \\ \|X^T \mu\|_\infty &\leq \lambda. \end{aligned} \tag{3}$$

### 2.1.1 Связь решений прямой и двойственной задачи

Выведем формулы, выражающие связь между решениями прямой задачи 2 и двойственной задачи 3.

Заметим, что в прямой задаче все ограничения имеют вид линейных равенств. При таких ограничениях значения двойственной и прямой задачи совпадают, причем если  $\mu^*$  — решение двойственной задачи, то найдутся такие  $z^*, w^*$ , что  $L(z^*, w^*, \mu^*) = \inf_{z, \mu} L(z, w, \mu^*)$ , и  $z^*, w^*$  — решение прямой задачи.

Итак, пусть  $\mu^*$  — решение двойственной задачи. Из вывода двойственной задачи видно, что мы имеем следующие ограничения на  $z, w$ :

$$\begin{cases} z = \mu \\ w_d = 0, \quad \forall d : |c_d| < \lambda \\ \text{sgn}(w_d) \neq \text{sgn}(c_d), \quad \forall d : |c_d| < \lambda \\ Xw - t = z \end{cases} \tag{4}$$

Эта система и выражает зависимость, связывающую решения прямой и двойственной задачи.

## 2.2 Эквивалентная формулировка в виде задачи гладкой условной оптимизации

Выпишем еще одну эквивалентную формулировку задачи 1.

$$\begin{aligned} \frac{1}{2} \|t - Xw\|^2 + \sum_{d=1}^D u_d &\rightarrow \min_{w, u}, \\ -u_d &\leq w_d \leq u_d, \quad \forall d = 1, \dots, D, \\ u_d &\geq 0, \quad \forall d = 1, \dots, D. \end{aligned} \quad (5)$$

Данная задача является гладкой задачей условной оптимизации.

## 2.3 Оценка на зазор между решениями прямой и двойственной задачи

Вернемся к задачам 2 и 3. Пусть  $w^*, z^*$  — решение прямой задачи, а  $\mu^*$  — соответствующие множители Лагранжа. Как было показано выше,  $Xw^* - t = z^* = \mu^*$ . Рассмотрим точку

$$\hat{\mu}(w) = \frac{\lambda(Xw - t)}{\|X^T(Xw - t)\|_\infty}. \quad (6)$$

Отметим, что при любом значении  $w$  данная точка является допустимой в двойственной задаче. Покажем, что если  $\|X^T t\|_\infty > \lambda$ , то выполняется  $\hat{\mu}(w^*) = \mu^*$ .

Заметим, что максимум квадратичной формы

$$-\frac{1}{2} \mu^T \mu - \mu^T t$$

достигается при  $\mu = -t$ .

Но  $-t$  не является допустимой точкой, так как иначе выполнялось бы  $\|X^T t\|_\infty \leq \lambda$ . Следовательно, градиент целевой функции не обращается в 0 внутри допустимого множества, поэтому оптимальное значение  $\mu^*$  лежит на границе этого множества, т.е.  $\|X^T \mu^*\|_\infty = \lambda$ . Подставим  $w = w^*$  в формулу 6.

$$\hat{\mu}(w^*) = \frac{\lambda(Xw^* - t)}{\|X^T(Xw^* - t)\|_\infty} = \mu^* \frac{\lambda}{\|X^T \mu^*\|_\infty} = \mu^*.$$

Итак  $\hat{\mu}(w^*) = \mu^*$ . При этом заметим, что отображение  $\hat{\mu}$  непрерывно, то есть

$$\lim_{w \rightarrow w^*} \hat{\mu}(w) = \mu^*.$$

Последнее соотношение позволяет построить следующую оценку на невязку в прямой задаче, которая будет использоваться в критериях останова для методов, описываемых ниже.

$$F(w) - F(w^*) \leq \frac{1}{2} \|t - Xw\|^2 + \lambda \|w\|_1 + \frac{1}{2} \mu^T \mu + \mu^T t. \quad (7)$$

### 3 Рассматриваемые методы

В данном разделе приводятся краткие описания рассматриваемых методов оптимизации.

#### 3.1 Прямой метод барьерных функций

Прямой метод барьерных функций применяется к задаче 5. В этом методе ограничения-неравенства заменяются на логарифмические штрафные функции, которые прибавляются к целевой функции.

$$F_\tau(w, u) = \frac{1}{2} \|t - Xw\|^2 + \lambda \sum_{d=1}^D u_d - \frac{1}{\tau} \sum_{d=1}^D \log(u_d - w_d) - \frac{1}{\tau} \sum_{d=1}^D \log(u_d + w_d) \rightarrow \min_{u, w}. \quad (8)$$

Будем решать эту задачу методом Ньютона. Вычислим градиент.

$$\nabla_w F_\tau(w, u) = -X^T t + X^T X w + \frac{1}{\tau} \frac{1}{u - w} - \frac{1}{\tau} \frac{1}{u + w},$$

$$\nabla_u F_\tau(w, u) = [\lambda, \lambda, \dots, \lambda]^T - \frac{1}{\tau} \frac{1}{u - w} - \frac{1}{\tau} \frac{1}{u + w},$$

где  $\frac{1}{u-w} = \left[ \frac{1}{u_1-w_1}, \dots, \frac{1}{u_D-w_D} \right]^T$ ,  $\frac{1}{u+w} = \left[ \frac{1}{u_1+w_1}, \dots, \frac{1}{u_D+w_D} \right]^T$ .

Теперь вычислим Гессиан. Обозначим

$$A = \frac{1}{\tau} \begin{bmatrix} \frac{1}{(u_1-w_1)^2} + \frac{1}{(u_1+w_1)^2} & 0 & \dots & 0 \\ 0 & \frac{1}{(u_2-w_2)^2} + \frac{1}{(u_2+w_2)^2} & \dots & 0 \\ 0 & \dots & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{(u_D-w_D)^2} + \frac{1}{(u_D+w_D)^2} \end{bmatrix},$$

$$B = \frac{1}{\tau} \begin{bmatrix} \frac{1}{(u_1+w_1)^2} - \frac{1}{(u_1-w_1)^2} & 0 & \dots & 0 \\ 0 & \frac{1}{(u_2+w_2)^2} - \frac{1}{(u_2-w_2)^2} & \dots & 0 \\ 0 & \dots & \ddots & 0 \\ 0 & 0 & \dots & \frac{1}{(u_D+w_D)^2} - \frac{1}{(u_D-w_D)^2} \end{bmatrix}.$$

Тогда Гессиан

$$\nabla^2 F_\tau(w, u) = \begin{bmatrix} X^T X + A & B \\ B & A \end{bmatrix}. \quad (9)$$

Шаг метода имеет вид

$$\theta_{n+1} = \theta_n - \alpha_n \nabla^2 F_\tau(\theta_n)^{-1} \nabla F_\tau(\theta_n),$$

где  $\theta_n = (w_n, u_n)$  — приближение, полученное методом на  $n$ -ой итерации. Параметр длины шага  $\alpha_n$  подбирается исходя из правила Армико и ограничений  $|w| \leq |u|$ ,  $u \geq 0$ . В качестве начального приближения используется  $w = [0, \dots, 0]^T$ ,  $u = [1, \dots, 1]^T$ .gr

Опишем схему метода.

1. Выбираем начальное значение параметра  $\tau$ , а также значение параметра  $\nu$ .
2. Для заданного  $\tau$  решаем задачу оптимизации 8.
  - (а) Выбираем начальное приближение  $\theta_0 = (w_0, u_0)$ , являющееся внутренней точкой для допустимого множества задачи 5. Полагаем  $k = 0$ .
  - (б) Вычисляем очередное направление оптимизации, решая систему

$$\nabla^2 F_\tau(w_k, u_k)d_k = \nabla F_\tau(w_k, u_k). \quad (10)$$

Эффективный способ решения этой системы будет описан ниже.

- (с) Выбираем длину шага  $\alpha_k$ , удовлетворяющую правилу Армико, с помощью процедуры backtracking.
- (д) Обновляем приближение

$$\theta_{k+1} = \theta_k - \alpha_k d_k.$$

- (е) Проверяем выполнение условия  $\|\nabla F_\tau(w_{k+1}, u_{k+1})\| < \varepsilon_{center}$ . Если оно выполняется, то кладем  $w_\tau = w_k$  и переходим к пункту 3. Иначе увеличиваем значение  $k$  на 1 и переходим к пункту (б).

3. Проверяем выполнение условия

$$\frac{1}{2}\|t - Xw_\tau\|^2 + \lambda\|w_\tau\|_1 + \frac{1}{2}\hat{\mu}(w_\tau)^T \hat{\mu}(w_\tau) + \hat{\mu}(w_\tau)^T t < \varepsilon_{gap},$$

где  $\hat{\mu}(\cdot)$  — отображение, определенное в 6 (см. оценку 7). Если оно выполняется, то устанавливаем  $w_{opt} = w_\tau$  и останавливаем работу метода. Иначе, увеличиваем  $\tau$  в  $\nu$  раз и переходим к пункту 2.

### 3.1.1 Эффективный метод решения системы 10

Воспользуемся дополнением Шура для построения матрицы, обратной, к  $\nabla^2 F_\tau(w)$ . В результате получим

$$\nabla^2 F_\tau(w)^{-1} = \begin{bmatrix} S^{-1} & -S^{-1}BA^{-1} \\ -A^{-1}BS^{-1} & A^{-1} + A^{-1}BS^{-1}BA^{-1} \end{bmatrix},$$

где  $S$  — дополнение по Шуру  $\nabla^2 F_\tau/A$ :

$$S = X^T X + A - BA^{-1}B.$$

Рассмотрим теперь произведение

$$\nabla^2 F_\tau(w)^{-1} \begin{bmatrix} dw \\ du \end{bmatrix} = \begin{bmatrix} S^{-1}dw - S^{-1}BA^{-1}du \\ -A^{-1}BS^{-1}dw + A^{-1}du + A^{-1}BS^{-1}BA^{-1}du \end{bmatrix}. \quad (11)$$

Здесь через  $dw, du$  обозначены соответственно  $\nabla_w F_\tau(w, u)$  и  $\nabla_u F_\tau(w, u)$ .

Заметим, что обращение диагональных матриц  $A, B$  не составляет труда. Поэтому сложность данного метода определяется сложностью решения двух линейных систем с матрицами размера  $D \times D$ :

$$Sx_w = d_w, Sx_u = BA^{-1}du.$$

Через решения  $x_u, x_w$  этих систем, очередное направление оптимизации, т.е. произведение 11, выражается как

$$\begin{bmatrix} x_w - x_u \\ -A^{-1}Bx_w + A^{-1}du + A^{-1}Bx_u \end{bmatrix}.$$

### 3.2 Прямо-двойственный метод внутренней точки

Прямо-двойственный метод внутренней точки применяется к задаче 3. В этом методе одновременно производится оптимизация как по прямым, так и по двойственным переменным. Перепишем задачу 3 в виде

$$\begin{aligned} \frac{1}{2}\mu^T\mu + \mu^T t &\rightarrow \min_{\mu} \\ X^T\mu &\leq \lambda \\ X^T\mu &\geq -\lambda \end{aligned} \quad (12)$$

Рассмотрим возмущенную систему Каруша-Куна-Такера для задачи 12.

$$\begin{aligned} \nabla_{\mu}(\frac{1}{2}\mu^T\mu + \mu^T t + \langle \gamma^1, X^T\mu - \lambda \rangle + \langle \gamma^2, -X^T\mu - \lambda \rangle) &= \mu + t + X(\gamma^1 - \gamma^2) = 0, \\ \gamma^1 &\geq 0, \gamma^2 \geq 0, \\ \text{diag}(\gamma^1)(X^T\mu - \lambda) &= -\frac{1}{\tau}, \\ \text{diag}(\gamma^2)(-X^T\mu - \lambda) &= -\frac{1}{\tau}. \end{aligned} \quad (13)$$

Здесь  $\gamma^1, \gamma^2$  — векторы двойственных переменных. Линеаризованный вариант этой системы имеет вид

$$\begin{aligned} &\begin{bmatrix} I & X & -X \\ \text{diag}(\gamma^1)X^T & \text{diag}(X^T\mu - \lambda) & 0 \\ -\text{diag}(\gamma^2)X^T & 0 & \text{diag}(-X^T\mu - \lambda) \end{bmatrix} \begin{bmatrix} d_{\mu} \\ d_{\gamma^1} \\ d_{\gamma^2} \end{bmatrix} = \\ &= - \begin{bmatrix} \mu + t + X(\gamma^1 - \gamma^2) \\ \text{diag}(\gamma^1)(X^T\mu - \lambda) + \frac{1}{\tau}e \\ \text{diag}(\gamma^2)(-X^T\mu - \lambda) + \frac{1}{\tau}e \end{bmatrix} = - \begin{bmatrix} r_{dual}(\mu, \gamma^1, \gamma^2) \\ r_{center}(\mu, \gamma^1, \gamma^2) \end{bmatrix} = -r(\mu, \gamma^1, \gamma^2), \end{aligned} \quad (14)$$

где  $e$  — вектор из единиц.

Решение  $d_{\mu}, d_{\gamma^1}, d_{\gamma^2}$  СЛАУ 14 является очередным направлением прямо-двойственного метода. Для выбора длины шага методом backtracking неточно решается задача одномерной минимизации для функции  $r$ :

$$\|r(\mu + \alpha d_{\mu}, \gamma^1 + \alpha d_{\gamma^1}, \gamma^2 + \alpha d_{\gamma^2})\| \leq (1 - \alpha\rho)\|r(\mu, \gamma^1, \gamma^2)\|,$$



где  $\rho$  — параметр, задаваемый пользователем. При выборе начального значения  $\alpha$  учитывается, что точка  $(\mu + \alpha d_\mu, \gamma^1 + \alpha d_{\gamma^1}, \gamma^2 + \alpha d_{\gamma^2})$  должна быть внутренней.

Опишем схему работы метода.

1. Выбираем начальное значение  $\tau$  и значение параметра  $\nu$ . Выбираем начальное приближение  $\mu_0, \gamma_0^1, \gamma_0^2$ , удовлетворяющее условиям

$$\|X^T \mu\|_\infty < \lambda,$$

$$\gamma^1, \gamma^2 \geq 0.$$

Устанавливаем  $k = 0$ .

2. Для текущего приближения  $\mu_k, \gamma_k^1, \gamma_k^2$  решая линейную систему 14 находим направление оптимизации  $d_k$ . Эффективный метод решения этой системы изложен ниже.
3. С помощью процедуры backtracking находим длину шага  $\alpha_k$ , удовлетворяющую правилу Армико.
4. Обновляем приближение

$$\begin{bmatrix} \mu_{k+1} \\ \gamma_{k+1}^1 \\ \gamma_{k+1}^2 \end{bmatrix} = \begin{bmatrix} \mu_k \\ \gamma_k^1 \\ \gamma_k^2 \end{bmatrix} - \alpha_k d_k.$$

Увеличиваем значение  $k$  на 1.

5. Проверяем условие

$$\begin{aligned} & \|r_{dual}(\mu_k, \gamma_k^1, \gamma_k^2)\| < \varepsilon_{feas}, \\ & -(\gamma_k^1)^T (X^T \mu - \lambda) + (\gamma_k^2)^T (X^T \mu + \lambda) < \varepsilon_{gap}. \end{aligned}$$

Если они выполняются, то устанавливаем  $\mu = \mu_k$ , вычисляем  $w$  исходя из 4. Иначе увеличиваем  $\tau$  в  $\nu$  раз и переходим к пункту 2.

### 3.2.1 Эффективный метод решения системы 14

Перепишем матрицу системы 14 в виде

$$M = \begin{bmatrix} I & C \\ D & E \end{bmatrix},$$

где  $C = [X, -X]$ ,  $D = [X \text{diag}(\gamma^1), -X \text{diag}(\gamma^2)]^T$ ,  $E = \begin{bmatrix} \text{diag}(X^T \mu - \lambda) & 0 \\ 0 & \text{diag}(-X^T \mu - \lambda) \end{bmatrix}$ .

Заметим, что матрицы  $I, E$  диагональны. Воспользуемся дополнением по Шуру.

$$M^{-1} = \begin{bmatrix} I & C \\ D & E \end{bmatrix}^{-1} = \begin{bmatrix} I + CS^{-1}D & -CS^{-1} \\ -S^{-1}D & S^{-1} \end{bmatrix},$$

где  $S = E - DC$  — дополнение по Шуру  $M/E$ .

Рассмотрим произведение обратной матрицы на правую часть системы. Обозначим  $\gamma = \begin{bmatrix} \gamma^1 \\ \gamma^2 \end{bmatrix}$ ,  $r_d = r_{dual}(\mu, \gamma)$ ,  $r_c = r_{center}(\mu, \gamma)$ . Имеем

$$\begin{bmatrix} d_\mu \\ d_\gamma \end{bmatrix} = M^{-1} \begin{bmatrix} r_d \\ r_c \end{bmatrix} = \begin{bmatrix} r_d + CS^{-1}Dr_d - CS^{-1}r_c \\ -S^{-1}Dr_d + S^{-1}r_c \end{bmatrix}.$$

Остается найти решения линейных систем  $Sx_d = Dr_d$ ,  $Sx_c = r_c$ . Решение системы выражается через них следующим образом.

$$\begin{bmatrix} d_\mu \\ d_\gamma \end{bmatrix} = \begin{bmatrix} r_d + Cx_d - Cx_c \\ -x_d + x_c \end{bmatrix}.$$

Таким образом, удалось свести решение исходной системы размеров  $(N + 2D) \times (N + 2D)$  к решению двух систем размеров  $2D \times 2cD$ .

### 3.3 Проксимальный метод

Итерации проксимального метода для негладкой задачи 1 имеют вид

$$w_{k+1} = \arg \min_w \left( \|t - Xw_k\|^2 + \partial(\|t - Xw_k\|^2)^T(w - w_k) + \frac{L}{2}\|w - w_k\|^2 + \lambda\|w\|_1 \right),$$

где  $\partial f$  обозначает субградиент функции  $f$  (по переменной  $w_k$ ). В данном случае аргументом субградиента является гладкая функция  $\|t - Xw_k\|^2$ , и он совпадает с градиентом от этой функции. Исключая слагаемые, не зависящие от  $w$ , окончательно получаем

$$w_{k+1} = \arg \min_w \left( \frac{L}{2}\langle w, w \rangle - L\langle w, w_k \rangle + 2\langle w, X^T X w_k \rangle - 2\langle X^T t, w \rangle + \lambda\|w\|_1 \right). \quad (15)$$

Заметим, что задача 15 является сепарабельной, и соответствующие ей одномерные задачи могут быть записаны как

$$\frac{L}{2}w_d^2 + b_d w_d + \lambda|w_d| \rightarrow \min_{w_d}, \quad (16)$$

где  $b_d = 2X^T X w_{k,d} - 2(X^T t)_d - Lw_{k,d}$ .

Задача 16 имеет аналитическое решение

$$w_d = \begin{cases} \frac{-b_d + \operatorname{sgn}(b_d)\lambda}{L}, & \text{если } |b_d| > \lambda, \\ 0, & \text{если } |b_d| \leq \lambda. \end{cases} \quad (17)$$

Опишем схему метода.

1. Устанавливаем  $k = 0$ , выбираем начальное приближение  $w_0$ . Устанавливаем значение параметра  $L$ .
2. Вычисляем вектор  $b_k = 2X^T X w_k - 2X^T t - L w_k$ . Вычисляем новое приближение  $w_{k+1}$  исходя из формулы 17. Увеличиваем значение  $k$  на 1.
3. Проверяем для приближения  $w_k$  условие

$$\frac{1}{2} \|t - X w_k\|^2 + \lambda \|w_k\|_1 + \frac{1}{2} \hat{\mu}(w_k)^T \hat{\mu}(w_k) + \hat{\mu}(w_k)^T t < \varepsilon_{gap},$$

где  $\hat{\mu}(\cdot)$  — отображение, определенное в 6 (см. оценку 7). Если оно выполняется, то устанавливаем  $w_{opt} = w_k$  и останавливаем работу метода. Иначе переходим к пункту 2.

## 4 Эксперименты

В данном разделе приводятся результаты проведенных экспериментов по сравнению трех реализованных методов. Методы сравнивались по скорости сходимости: изучалась зависимость невязки по функции от времени и от номера итерации. Заметим, что прямо-двойственный метод и прямой метод барьеров обладают сверхлинейной скоростью сходимости, поэтому по итерациям они обычно работают существенно лучше проксимального метода. Однако итерации проксимального метода проще, и выполняются на компьютере быстрее, поэтому нас в первую очередь будет интересовать именно сравнение по времени работы.

В проксимальном методе скорость сходимости (и вообще наличие сходимости) зависит от выбора параметра  $L$ . Во всех экспериментах этот параметр подбирался опытным путем, чтобы добиться оптимальной работы метода.

### 4.1 Небольшие задачи

В данном разделе приводятся результаты работы методов на наборах данных `abalone` и `bodyfat`.

Набор данных `bodyfat` состоит из 252 объектов с 14 признаками. Набор данных `abalone` содержит 4177 объектов с 8 признаками.

Результаты работы методов по итерациям приведены на рис. 1. Видно, что в обоих случаях методы внутренней точки показали близкие результаты, а проксимальный метод показал значительно худший результат.

Графики работы методов по времени приведены на рисунке 2. Видно, что проксимальный метод за счет более быстрых итераций обгоняет другие методы на ранних итерациях. Также видно, что по времени методы внутренней точки показали разные результаты. На наборе данных `abalone` прямо-двойственный метод существенно проиграл по времени прямому методу логарифмических барьеров, а на наборе `bodyfat` — выиграл. Это можно объяснить тем, что соотношение числа признаков и числа

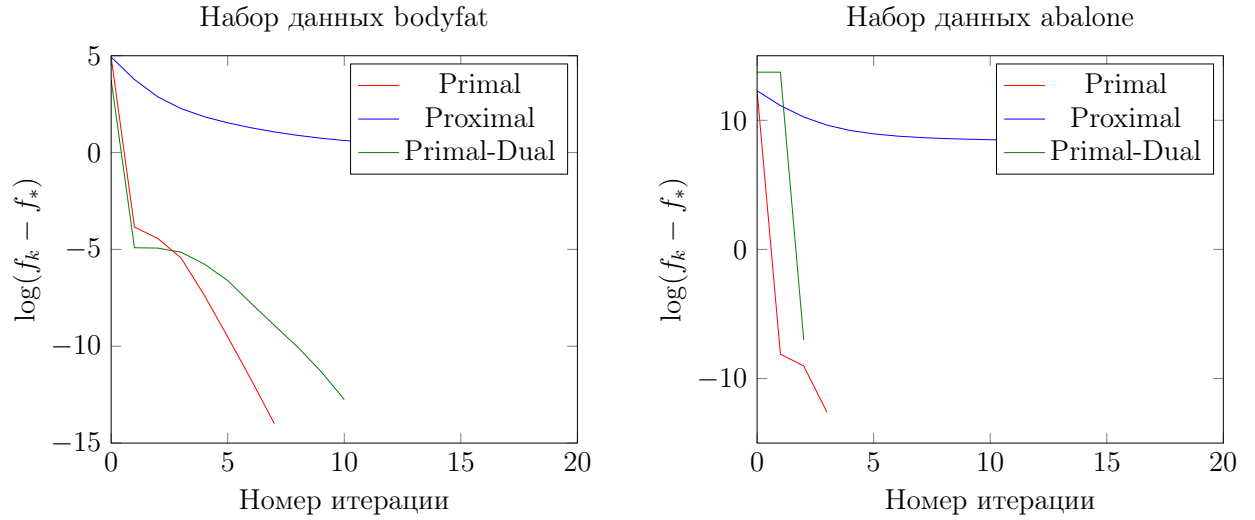


Рис. 1: Сравнение методов по итерациям на небольшой задаче.

объектов в этих задачах существенно различается. Размерность задачи, решаемой в прямом методе для набора данных abalone составляет 8, в то время как размерность задачи прямо-двойственного метода — 4193.

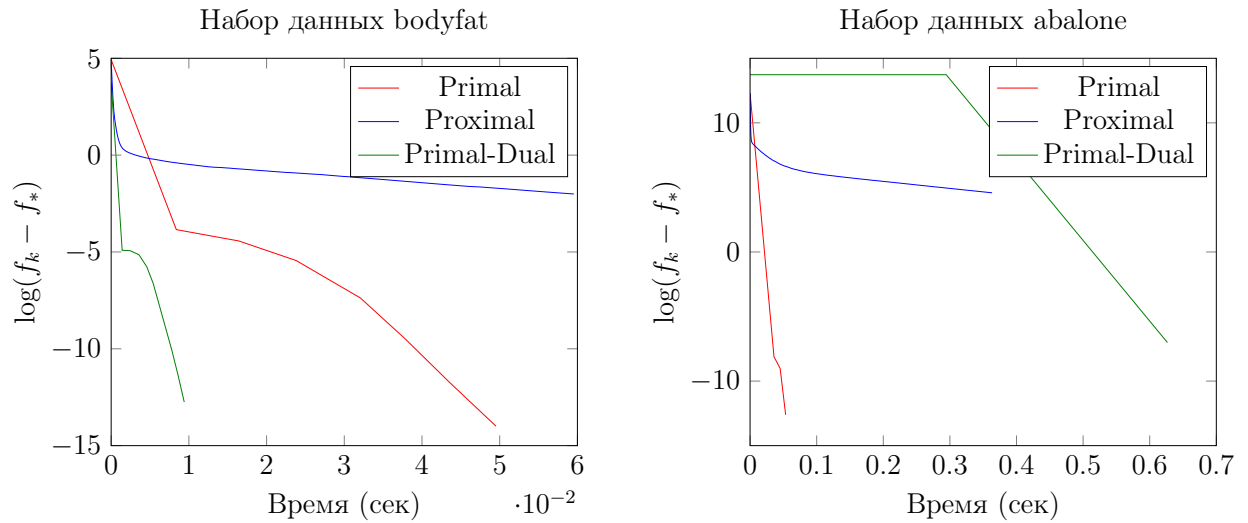


Рис. 2: Сравнение методов по времени на небольшой задаче.

## 4.2 Задачи средних размеров

В данном разделе приводятся результаты работы методов на наборе данных cрусsmall, состоящем из 8192 объектов с 12 признаками. Результаты работы методов приведены

на рис. 3.

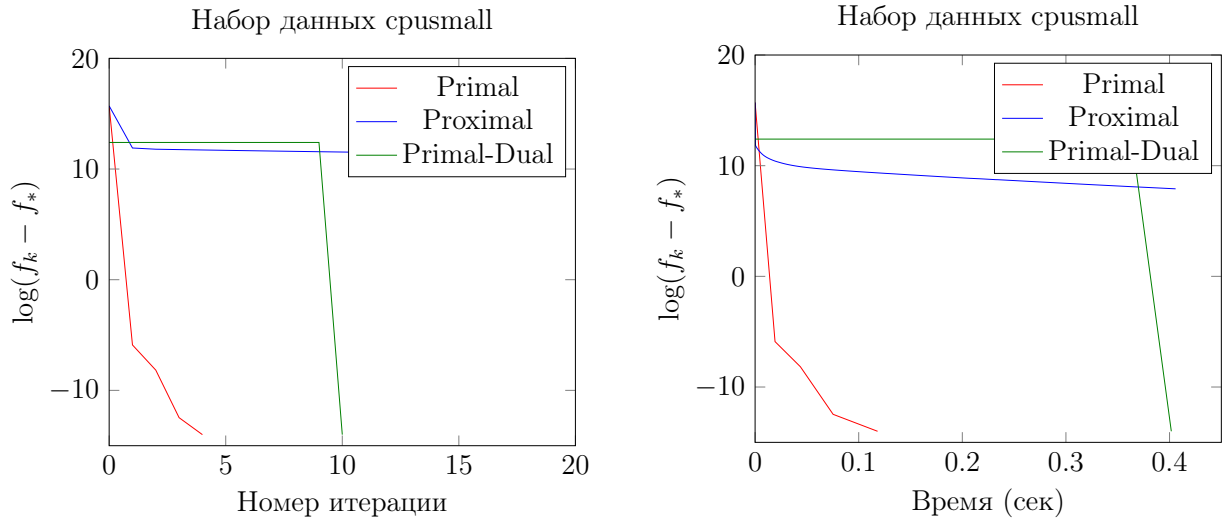


Рис. 3: Сравнение методов на задаче средних размеров.

Видно, что прямой метод логарифмических барьеров вновь оказался лучше других методов по времени. Проксимальный метод опережает другие методы на ранних итерациях, но вскоре начинает уступать им. Прямо-двойственный метод уступает прямому методу как по итерациям, так и по времени.

## 5 Выводы

По результатам проведенных экспериментов можно сделать ряд выводов.

Во-первых, проксимальный метод для задачи  $L_1$ -регуляризованной линейной регрессии имеет смысл использовать только в том случае, когда использование методов внутренней точки невозможно (например, из-за того, что их итерации становятся слишком долгими при больших размерах данных), либо если есть возможность сделать всего одну-две итерации этих методов.

Во-вторых, прямо-двойственный метод может работать лучше прямого метода барьерных функций для задач, в которых размерность признакового пространства сравнима с размером выборки. Если же размер выборки намного больше размерности признакового пространства, то имеет смысл использовать прямой метод.