Лабораторная работа 3: Композитная оптимизация

Белова Юлия

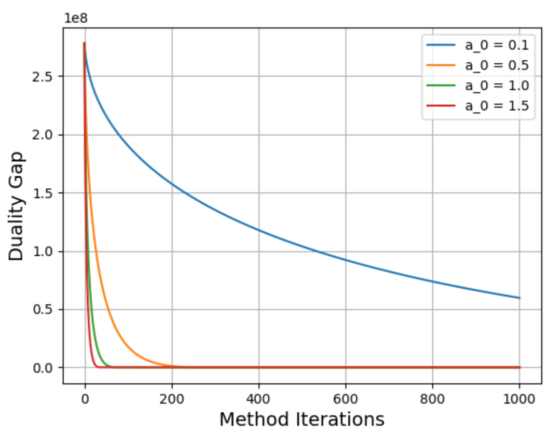
25 июня 2023 г.

Содержание

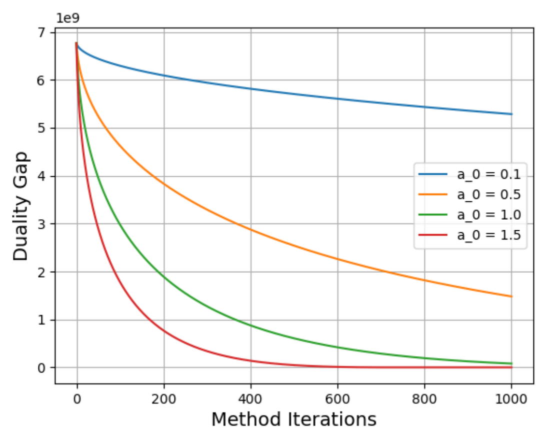
1 Эксперимент 1: Выбор длины шага в субградиентном методе

В данном эксперименте исследуется зависимость величины невязки, в нашем случае зазора двойственности, от числа итераций для различных констант в формуле для длины шага. Параметры задачи:

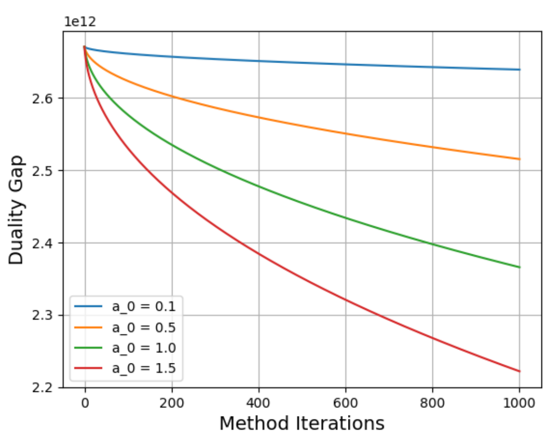
1. Lasso-задача с параметром регуляризации $\lambda = 1$
2. $n = 500$, $m = 1000$, где m - размерность пространства
3. Три типа начальных точек:
   (a) $x_0 \sim \text{uniform}(-1, 1)$
   (b) $x_0 \sim \text{uniform}(-5, 5)$
   (c) $x_0 \sim \text{uniform}(-100, 100)$
4. Перебираемые α_0: $[0.1, 0.5, 1.0, 1.5, 2.0]$



(a) Начальная точка №1



(b) Начальная точка №2



(c) Начальная точка №3

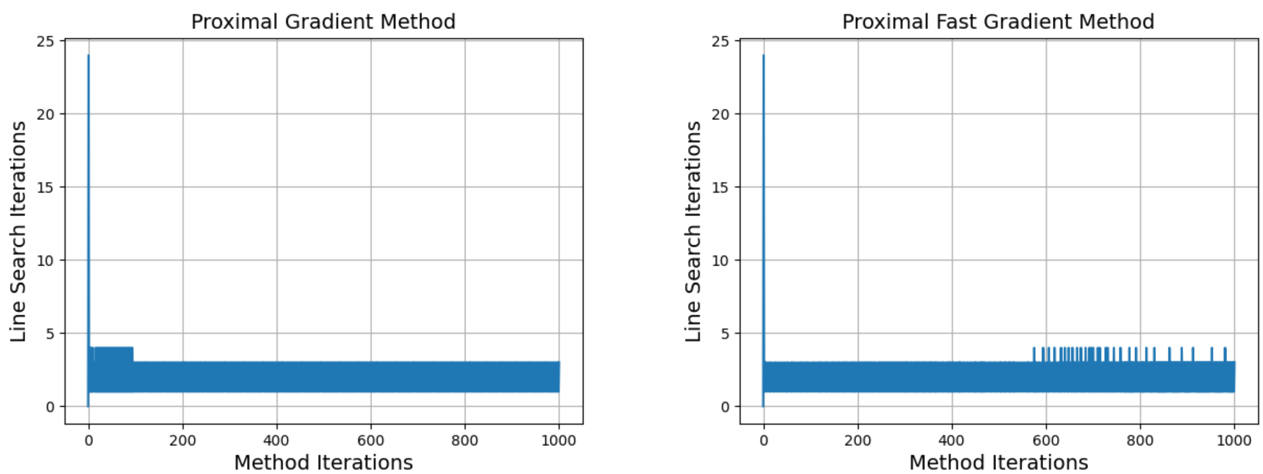
Рис. 1: Зависимость зазора двойственности от числа итераций

По результатам эксперимента можно сделать следующий вывод: чем ниже значение α_0, тем медленнее будет снижаться зазор двойственности, а значит будет медленнее сходиться метод в целом. Выбор оптимального α_0 не зависит от выбора начальной точки.

2 Эксперимент 2: Среднее число итераций одномерного поиска в градиентных методах

Исследуем, сколько итераций линейного поиска L необходимо проксимальному градиентному методу и ускоренному методу Нестерова. Параметры задачи:

1. Lasso-задача с параметром регуляризации $\lambda = 1$
2. $n = 500$, $m = 1000$, где m - размерность пространства
3. Начальная точка $x_0 = 0$



(a) (b)

Рис. 2: Зависимость числа итераций линейного поиска от количества итераций метода

Оба метода не успели сойтись за максимальное число итераций - 1000. При этом минимальный зазор двойственности меньше в быстром градиентном методе (13.323 в градиентном методе и 8.007 в быстром градиентном методе). При этом среднее число итераций линейного поиска, действительно, равно примерно двум и не зависит от числа итераций.

3 Эксперимент 3: Сравнение методов

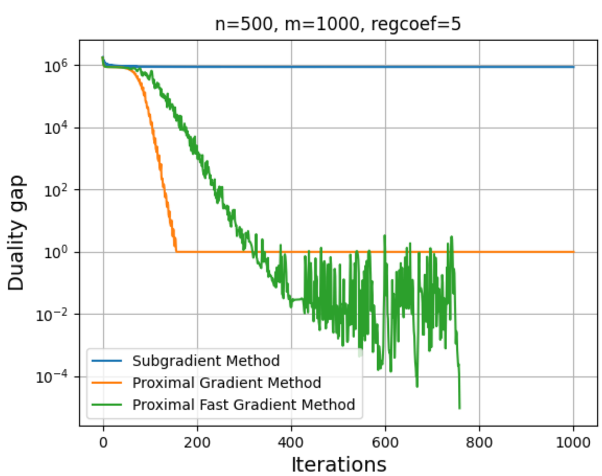
Для сравнения субградиентного метода, проксимального градиентного метода и быстрого градиентного метода на задаче Lasso были выбраны следующие параметры:

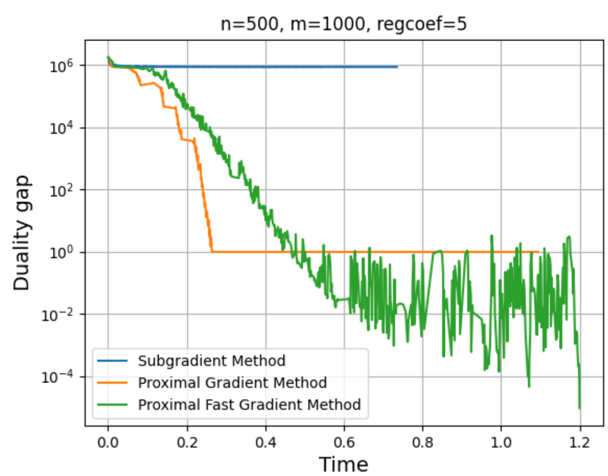
1. Количество семплов n: $[500, 1000]$
2. Размерность пространства m: $[1000, 5000]$
3. Regcoef: $[0.1, 1.0, 5.0]$

Матрица A формировалась путем перебора всевозможных пар n и m, для каждой такой матрицы проверялись различные коэффициенты $l1$ регуляризации. Всего проведено 12 экспериментов. Графики приводятся выборочно.

3.1 $n = 500$, $m = 1000$

Для данной задачи лучшей всего себя показал быстрый градиентный метод с коэффициентом регуляризации $\lambda = 5$, это единственный случай, когда удалось сойтись к решению.

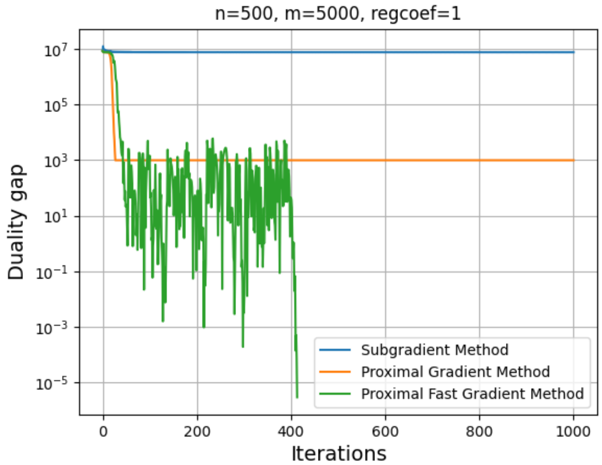


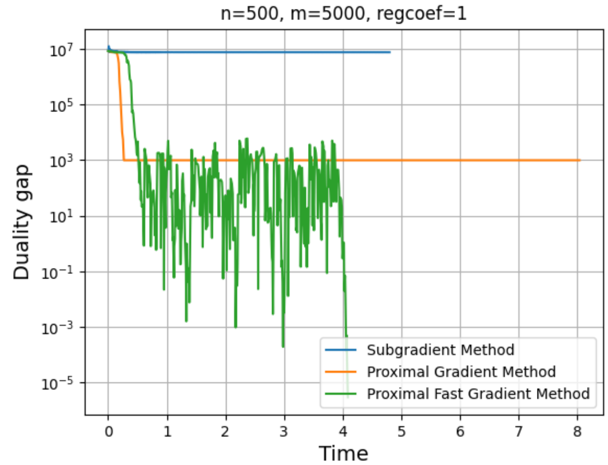


Из графиков видно, что субградиентный метод совсем не сходится (что является следствием выбора случайного субградиента, отсюда - возможно не идем в сторону уменьшения функции). Также видно, что субградиентный метод работает быстрее остальных методов. Проксимальный градиентный метод "застрял"в какой-то точке. Быстрый градиентный метод сходится скачкообразно, вероятнее всего, на это влияет то, что мы запоминаем лучшее приближение функции v_k.

3.2 $n = 500$, $m = 5000$

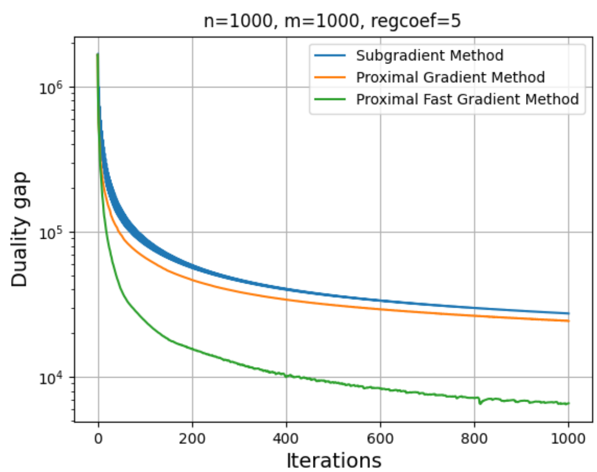
На данной задаче быстрый градиентный метод сошелся при $\lambda = 1$, $\lambda = 5$. При этом время на решение данной задачи, по сравнению с предыдущей, выросло примерно в 4 раза. Быстрому градиентному методу, однако, потребовалось меньше итераций.

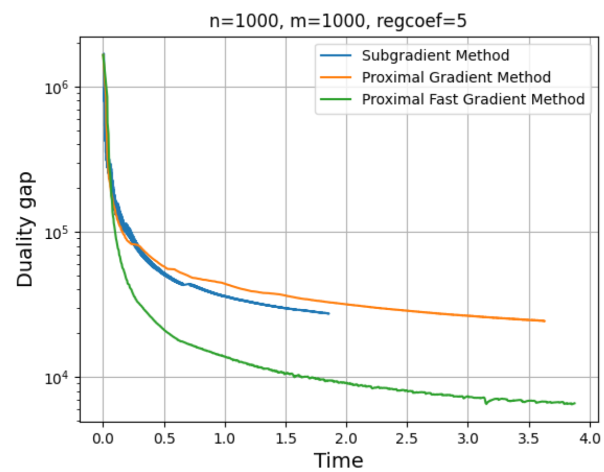




3.3 $n = 1000$, $m = 1000$

На квадратичной задаче не сошелся ни один из методов. При этом быстрый градиентный метод сходился более плавно, чем на предыдущих задачах. При этом размер зазора двойственности не сильно уменьшается в зависимости от коэффициента регуляризации.





3.4 $n = 1000$, $m = 5000$

В данной задаче сошелся к оптимуму только быстрый градиентный метод. Наблюдается все та же зависимость: чем выше коэффициент регуляризации, тем лучше сходится метод.

3.5 Выводы

1. Чем выше коэффициент регуляризации, тем больше вероятность того, что метод сойдется

2. Быстрый градиентный метод показал себя лучше всего, потому что он хоть где-то сошелся

3. Субградиентному методу очень трудно сойтись, начать уменьшать зазор двойственности из-за случайности в выборе субградиента