2.1 Эксперимент: Зависимость числа итераций метода сопряженных градиентов от числа обусловленности и размерности пространства

Задача:

исследовать, как зависит число итераций от числа обусловленности $k \geq 1$ оптимизируемой функции и размерности пространства n оптимизируемых переменных. Сравнить результат метода сопряженных градиентов с результатом градиентного спуска

Рисунок 1. Зависимость числа итераций от размерности пространства n и числа обусловленности k. Метод сопряженных квадратов

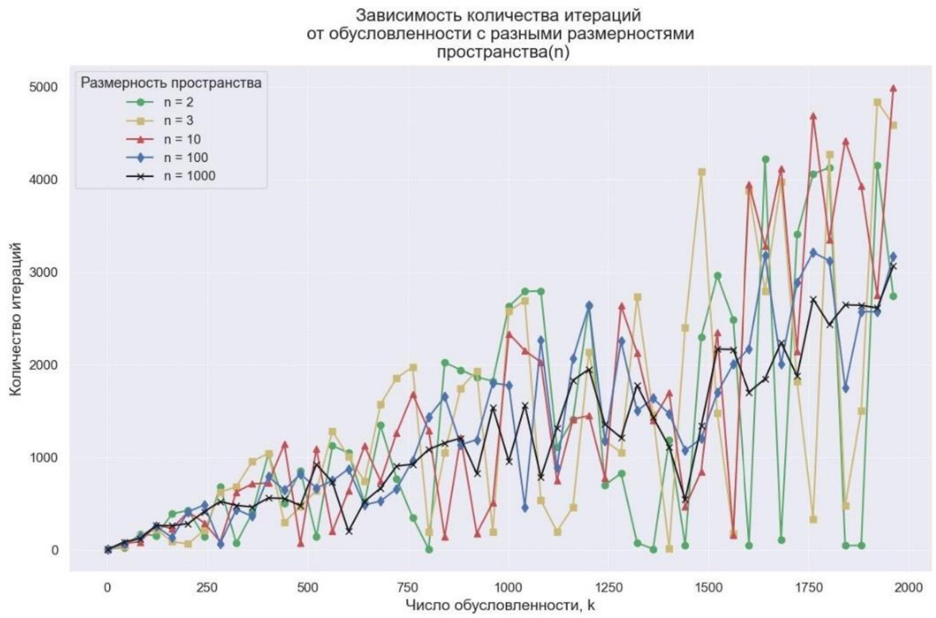

Рисунок 2. Зависимость числа итераций от размерности пространства n и числа обусловленности k. Градиентный спуск

Вывод:

Для низкоразмерных пространств n = 2, 3, 10 метод сопряженных градиентов показывает стабильность и эффективность, что выражается в минимальном изменении количества итераций при изменении числа обусловленности. Для высоких размерностей вроде n = 1000 количество итераций значительно возрастает из-за вычислительной сложности. Сама производтельность варьируется в зависимости от числа обсусловенности(k), так как обычно больше k – больше итераций, особенно в высокоразмерных пространствах. Метод сопряженных градиентов лучше как при маленькой размерности n = 2, 3, 10, так и при большой n = 100, 1000. Он требует меньше итераций для сходимости по сравнению с градиентным спуском независимо от размерности пространства

2.2 Эксперимент: Выбор размера истории в методе L-BFGS.

Задача:

исследовать как влияет размер истории в методе L-BFGS на поведение метода. Выбрано несколько вариантов размера истории [1, 5, 10, 20, 50, 100]

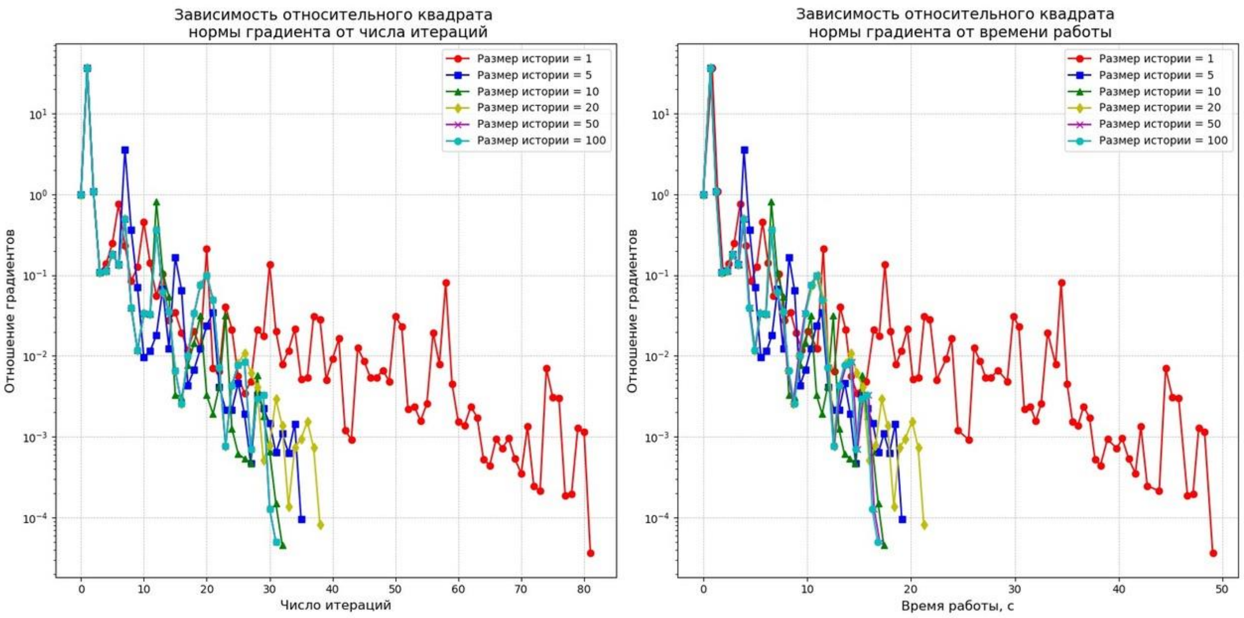

Рисунок 3. Зависимости в полулогарифмическом масштабе относительной нормы градиентов от времени работы и числа итераций. Датасет gissete

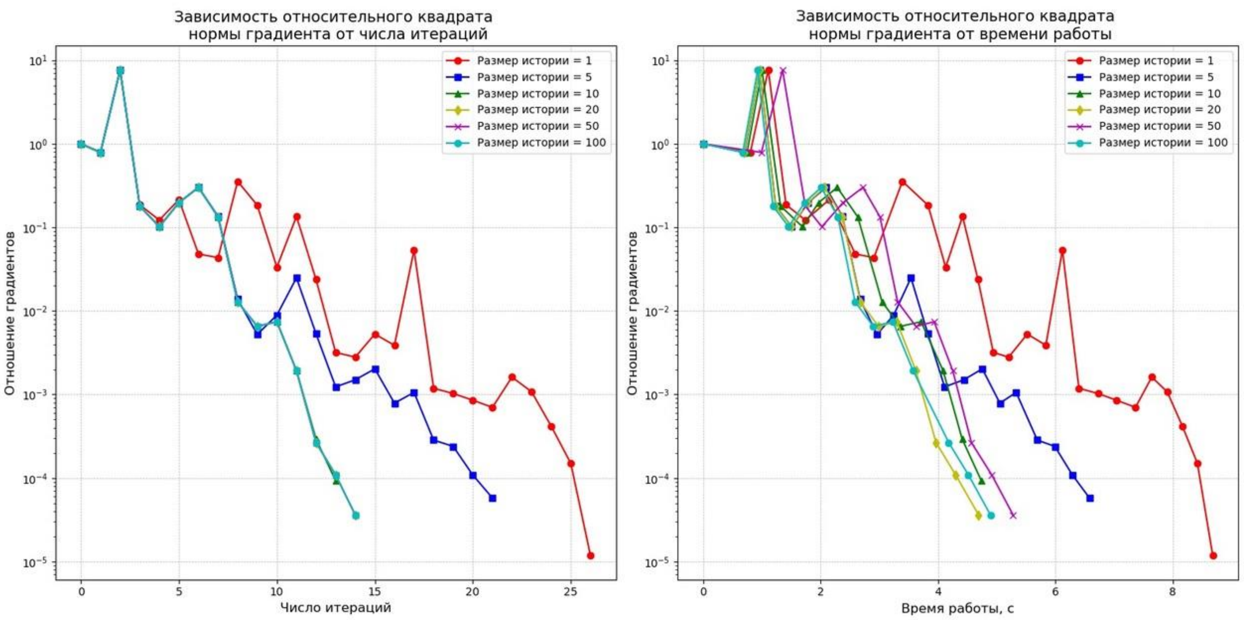

Рисунок 4. Зависимости в полулогарифмическом масштабе относительной нормы градиентов от времени работы и числа итераций. Датасет news20

Вывод:

Во-первых, чем больше история - тем быстрее работает алгоритм: это видно из графиков, где увеличение размера истории позволяет быстрее снижать норму градиента как по числу итераций, так и по времени работы. Также большая история приводит к меньшему числу резких скачков. Это делает алгоритм более предсказуемым и стабильным. Увеличение размера истории позволяет достигать сходимости за меньшее количество итераций. Это видно из того, что при большом размере истории норма градиента снижается быстрее и более плавно

2.3 Эксперимент: Сравнение методов на реальной задаче логистической регрессии

Задача:

Сравнить усеченный метод Ньютона, метод L-BFGS и градиентный спуск на реальной задаче логистической регрессии. Сравнить методы между собой

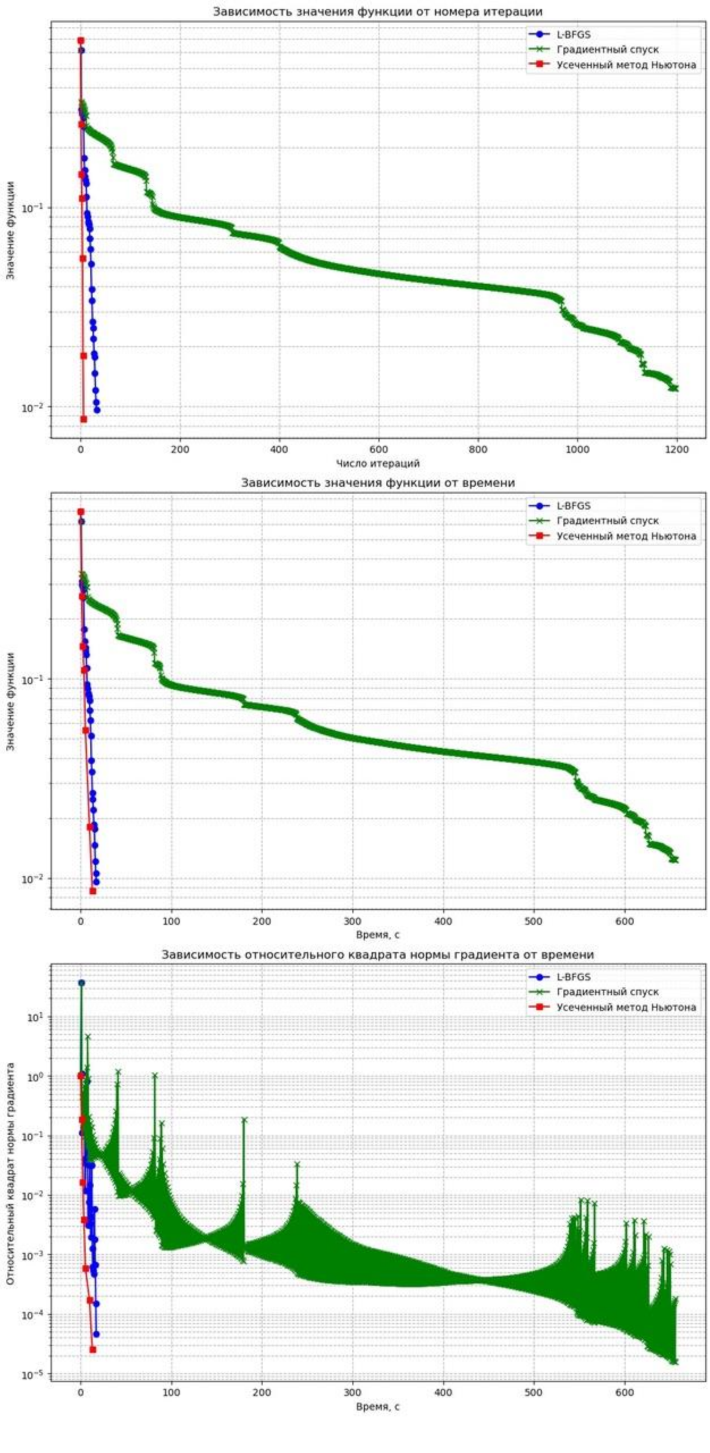

Рисунок 5. Зависимости в логарифмическом масштабе по одной оси н.град. по времени и числу итераций. gissete

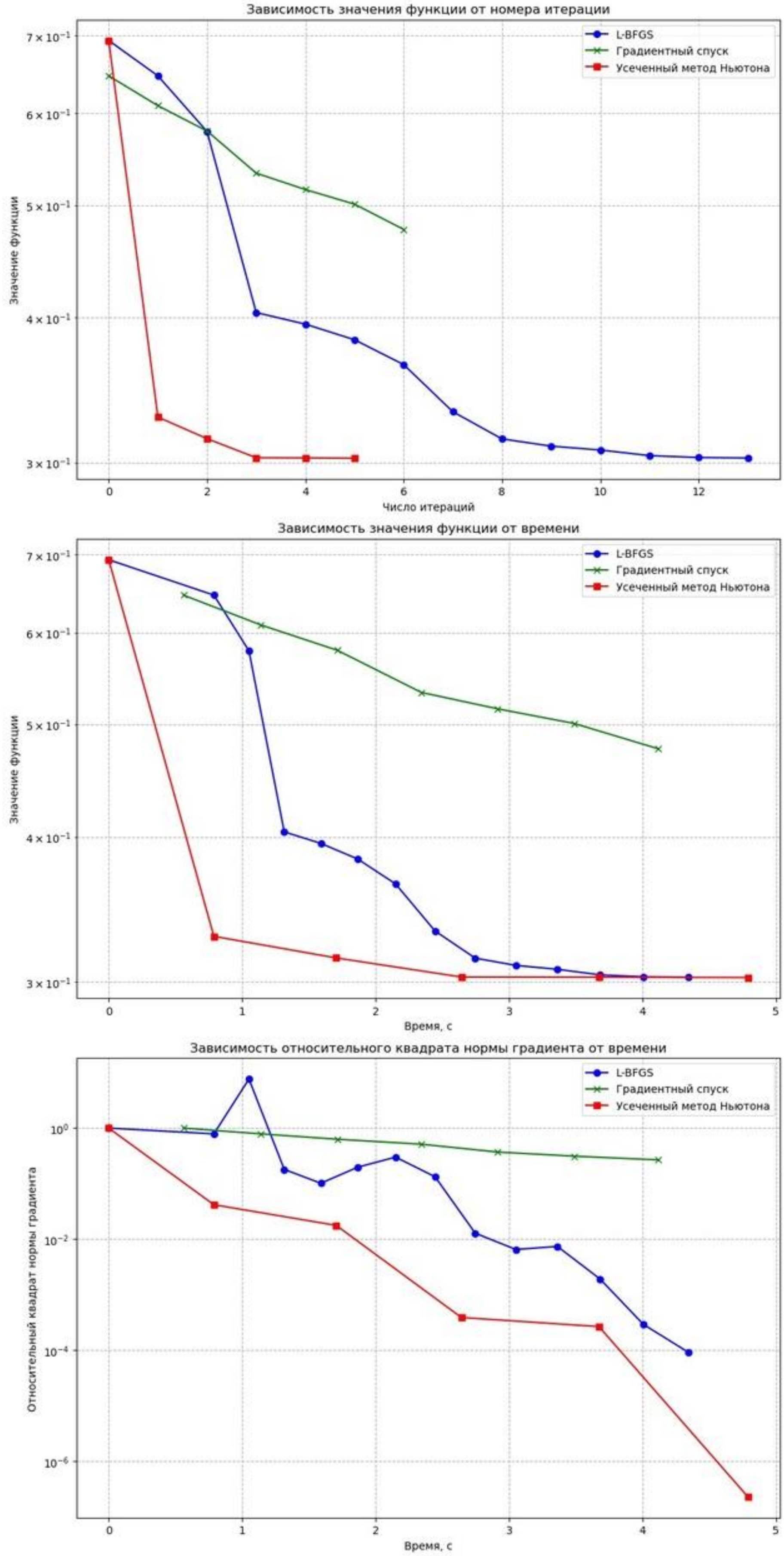

Рисунок 6. Зависимости в логарифмическом масштабе по одной оси н.град. по времени и числу итераций. News20

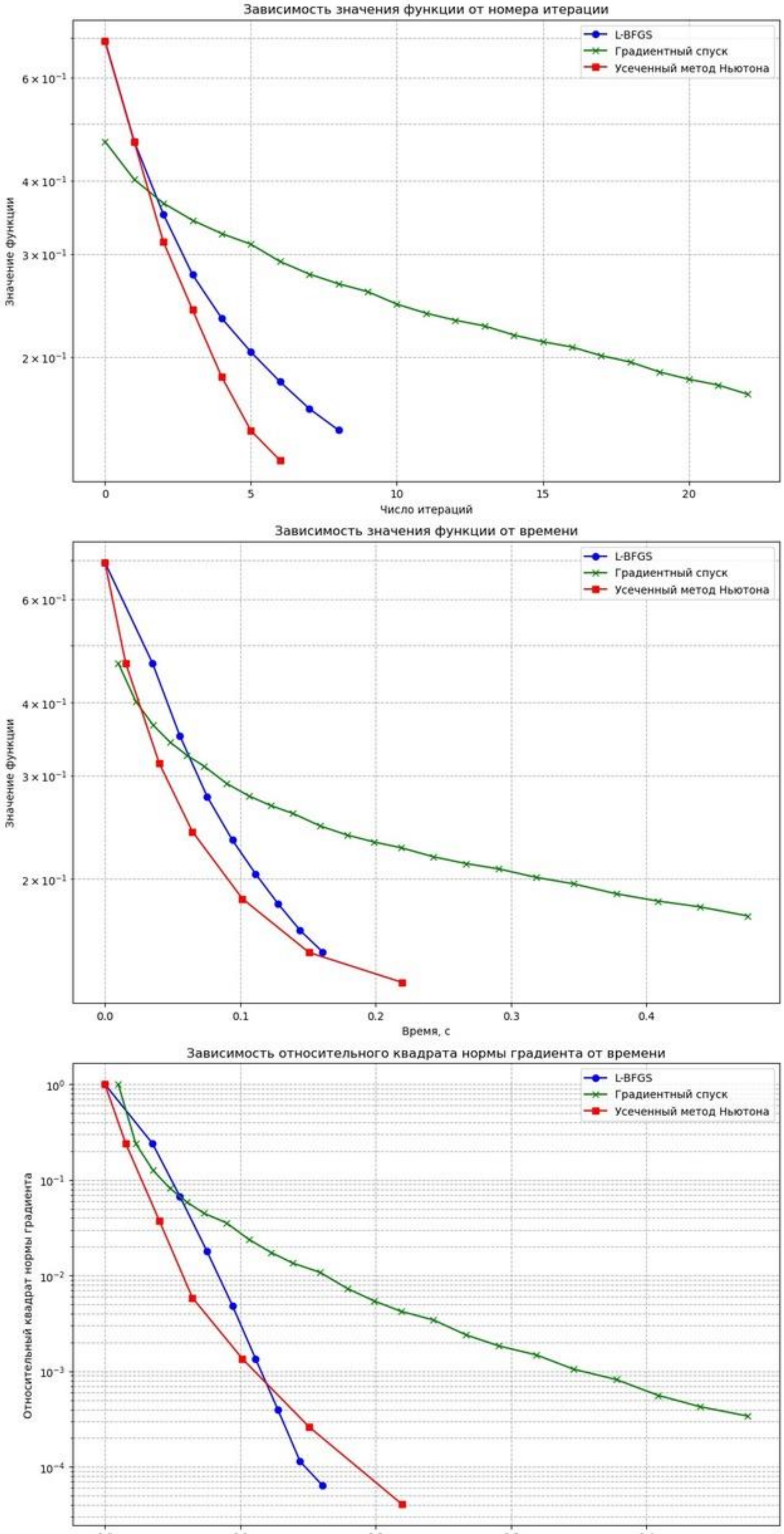

Рисунок 7. Зависимости в логарифмическом масштабе по одной оси н.град. по времени и числу итераций.w8a

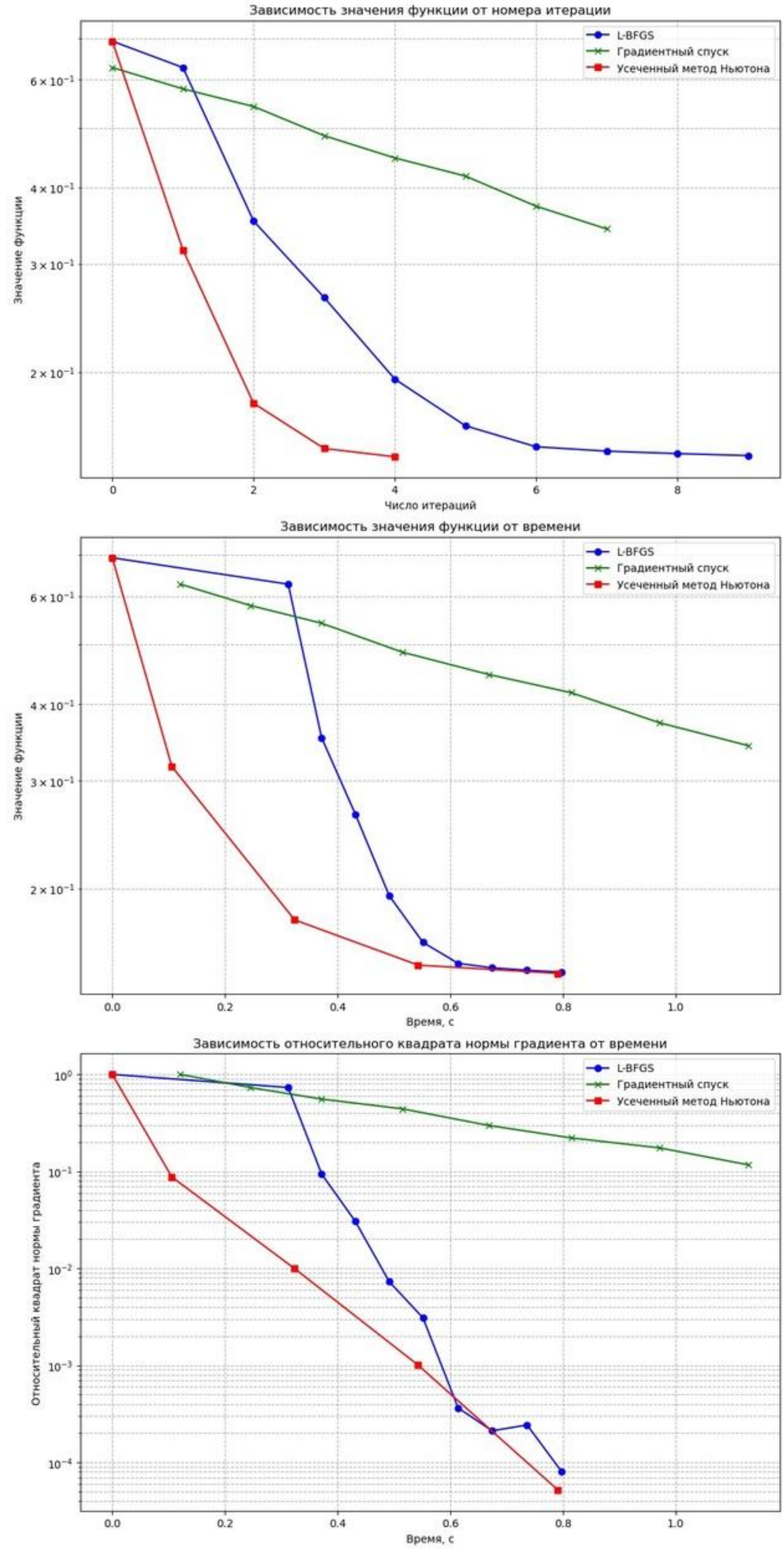

Рисунок 8. Зависимости в логарифмическом масштабе по одной оси н.град. по времени и числу итераций. realsim

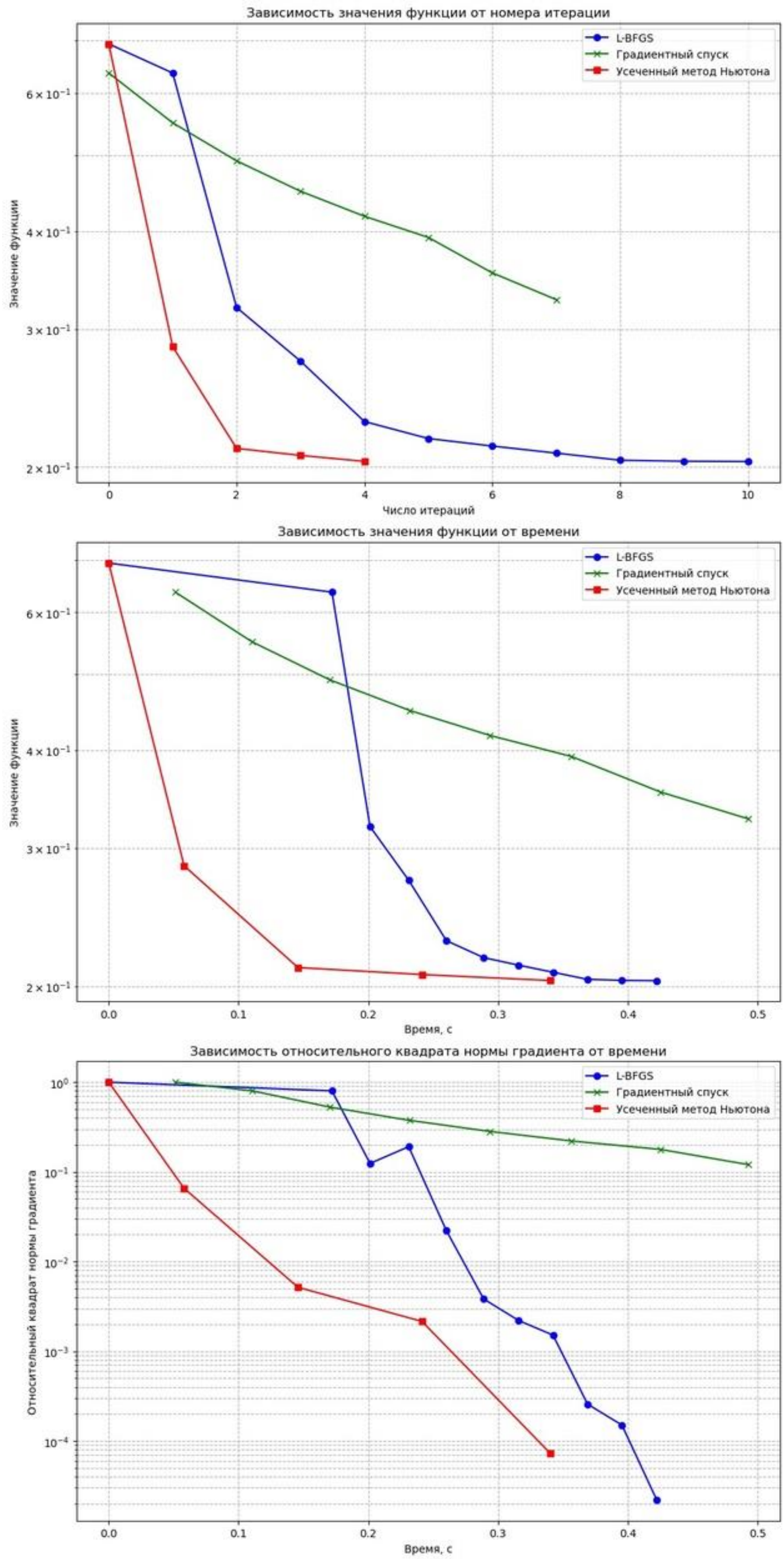

Рисунок 9. Зависимости в логарифмическом масштабе по одной оси н.град. по времени и числу итераций.rcv

Вывод:

На датасете gissete усеченный метод Ньютона (красный) сходится за минимальное количество итераций (менее 5). Метод L-BFGS (синий) требует около 10, а градиентный спуск (зелёный) требует в разы больше итераций. Схожая картина наблюдается и на других 4 датасетах. Можно заключить, что усеченный метод Ньютона является лучшим выбором, так как он показывает наилучшую скорость и стабильность сходимости. Метод L BFGS является тоже хорошим вариантом, который работает лучше, чем градиентный спуск, но уступает усеченному методу Ньютона