Определения и формулировки

1. Показать, что направление антиградиента - направление наискорейшего локального убывания функции.
2. Метод градиентного спуска.
3. Наискорейший спуск.
4. Липшицева парабола для гладкой функции.
5. Размер шага наискорейшего спуска для квадратичной функции.
6. Характер сходимости градиентного спуска к локальному экстремуму для гладких невыпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
7. Характер сходимости градиентного спуска для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
8. Характер сходимости градиентного спуска для гладких и сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
9. Связь спектра гессиана с константами сильной выпуклости и гладкости функции.
10. Связь числа обусловленности матрицы квадратичной функции с параметрами сильной выпуклости и гладкости функции.
11. Условие Поляка-Лоясиевича (градиентного доминирования) для функций.
12. Сходимость градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
13. Связь PL-функций и сильно выпуклых функций.
14. Привести пример выпуклой, но не сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
15. Привести пример сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
16. Привести пример выпуклой негладкой задачи линейных наименьших квадратов (возможно, с регуляризацией).
17. Нижние оценки для минимизации гладких выпуклых функций с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
18. Нижние оценки для минимизации гладких сильно выпуклых (PL) функций с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
19. Отличие ускоренной и неускоренной линейной сходимости для методов первого порядка.
20. Формулировка метода тяжелого шарика (Поляка) для квадратичной функции. Характер сходимости. Оптимальные гиперпараметры.
21. Формулировка метода тяжелого шарика (Поляка) для гладкой выпуклой/сильно выпуклой функций. Характер сходимости в терминах $\mathcal{O}$ от числа итераций метода.
22. Ускоренный градиентный метод Нестерова для выпуклых гладких функций. Характер сходимости в терминах $\mathcal{O}$ от числа итераций метода.
23. Ускоренный градиентный метод Нестерова для сильно выпуклых гладких функций. Характер сходимости в терминах $\mathcal{O}$ от числа итераций метода.
24. A-сопряженность двух векторов. A-ортогональность. Скалярное произведение $\langle \cdot, \cdot \rangle_A$.
25. Процедура ортогонализации Грама-Шмидта.
26. Метод сопряженных направлений.
27. Метод сопряженных градиентов.
28. Зависимость сходимости метода сопряженных градиентнов от спектра матрицы.
29. Характер сходимости метода сопряженных градиентов в терминах $\mathcal{O}$ от числа итераций метода.
30. Метод Поляка-Рибьера.
31. Метод Ньютона.
32. Сходимость метода Ньютона для квадратичной функции.
33. Характер сходимости метода Ньютона для сильно выпуклых гладких функций - куда и как

сходится.
34. Демпфированный метод Ньютона.
35. Идея квазиньютоновских методов. Метод SR-1.
36. Проекция.
37. Достаточное условие существования проекции точки на множество.
38. Достаточное условие единственности проекции точки на множество.
39. Метод проекции градиента.
40. Критерий проекции точки на выпуклое множество (Неравенство Бурбаки-Чейни-Гольдштейна).
41. Проекция как нерастягивающий оператор.
42. Метод Франк-Вульфа.
43. Характер сходимости метода проекции градиента для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
44. Характер сходимости метода проекции градиента для гладких сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
45. Характер сходимости метода Франк-Вульфа для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
46. Характер сходимости метода Франк-Вульфа для гладких сильно выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
47. Субградиент. Субдифференциал.
48. Субградиентный метод.
49. Характер сходимости субградиентного метода для негладких выпуклых Липшицевых функций в терминах $\mathcal{O}$ от числа итераций метода.
50. Характер сходимости субградиентного метода для негладких сильно выпуклых Липшицевых функций в терминах $\mathcal{O}$ от числа итераций метода.
51. Стратегия выбора шага для субградиентного метода в случае выпуклой Липшицевой функции.
52. Нижние оценки для негладкой выпуклой оптимизации с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
53. Нижние оценки для негладкой сильно выпуклой оптимизации с помощью методов первого порядка в терминах $\mathcal{O}$ от числа итераций метода.
54. Проксимальный оператор.
55. Оператор проекции как частный случай проксимального оператора.
56. Характер сходимости проксимального градиентного метода для гладких выпуклых функций f в терминах $\mathcal{O}$ от числа итераций метода.
57. Характер сходимости проксимального градиентного метода для гладких сильно выпуклых функций f в терминах $\mathcal{O}$ от числа итераций метода.
58. Аналитическое выражение для $\text{prox}_{\lambda\|x\|_1}$.
59. Аналитическое выражение для $\text{prox}_{\frac{\mu}{2}\|x\|_2^2}$.
60. Проксимальный оператор как нерастягивающий оператор.
61. Характер сходимости ускоренного проксимального градиентного метода для гладких выпуклых функций f в терминах $\mathcal{O}$ от числа итераций метода.
62. Метод стохастического градиентного спуска.
63. Идея мини-батча для метода стохастического градиентного спуска. Эпоха.
64. Характер сходимости стохастического градиентного спуска для гладких выпуклых функций в терминах $\mathcal{O}$ от числа итераций метода.
65. Характер сходимости стохастического градиентного спуска для гладких PL-функций в терминах $\mathcal{O}$ от числа итераций метода.
66. Характер работы стохастического градиентного спуска с постоянным шагом для гладких PL-функций.
67. Основная идея методов уменьшения дисперсии.
68. Метод SVRG.

69. Метод SAG.
70. Метод Adagrad.
71. Метод RMSProp.
72. Метод Adadelta.
73. Метод Adam.
74. Метод AdamW.
75. Метод Shampoo.
76. Метод Muon.
77. Как сравниваются методы в AlgoPerf Benchmark
78. Идея проекции функции потерь нейронной сети на прямую, плоскость.
79. Grokking.
80. Double Descent.
81. Взрыв/Затухание градиентов при обучении глубоких нейронных сетей.
82. Идея gradient checkpointing.
83. Идея аккумуляции градиентов.
84. Зачем увеличивать батч при обучении больших нейросетевых моделей. Warmup.
85. Идея cooldown фазы для построения расписания learning rate. В чём преимущество по сравнению с cosine scheduler?
86. Дифференциальное уравнение градиентного потока.
87. Характер сходимости траектории градиентного потока для выпуклых функций в терминах $\mathcal{O}(t)$.
88. Характер сходимости траектории градиентного потока для PL-функций в терминах $\mathcal{O}(t)$.
89. Дифференциальное уравнение Нестеровского ускоренного градиентного потока.
90. Метод двойственного градиентного подъема.
91. Связь константы сильной выпуклости f и гладкости f^*.
92. Идея dual decomposition.
93. Метод двойственного градиентного подъема для линейных ограничений-неравенств.
94. Метод модифицированной функции Лагранжа.
95. Метод ADMM.
96. Формулировка задачи линейных наименьших квадратов с ℓ_1 регуляризацией в форме ADMM.
97. Формулировка задачи поиска точки на пересечении двух выпуклых множеств в форме ADMM.

Теоремы с доказательствами

1. Теорема сходимости градиентного спуска для гладких выпуклых функций.
2. Теорема сходимости градиентного спуска для гладких PL функций.
3. Теорема сходимости градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
4. Теорема о нижней оценке для минимизации гладких выпуклых функций с помощью методов первого порядка.
5. Вывод ускоренного метода для квадратичной функции с помощью полиномов Чебышёва.
6. Теорема о сходимости метода тяжелого шарика для сильно выпуклой квадратичной задачи.
7. Доказательство сходимости метода сопряженных градиентов и вывод формул метода (В этом вопросе необходимо доказать за какое количество шагов сходится метод, как выбираются направления и почему в A-ортогонализации достаточно хранить лишь предыдущий шаг метода, а не все предыдущие).
8. Теорема сходимости метода Ньютона для сильно выпуклых функций с липшицевым гессианом.
9. Вывод формул обновления оценок обратного гессиана и гессиана квазиньютоновских методов SR-1, DFP, BFGS.
10. Теорема о сходимости метода проекции градиента для выпуклой гладкой функции.

11. Теорема о сходимости метода Франк-Вульфа для выпуклой гладкой функции.
12. Теорема о сходимости субградиентного метода для выпуклых Липшицевых функций. Стратегии выбора шага для сходимости. Как обеспечить сходимость с постоянным шагом, задаваемым заранее? Как обеспечить сходимость с убывающим шагом?
13. Теорема о сходимости субградиентного метода для сильно выпуклых Липшицевых функций.
14. Теорема о сходимости проксимального градиентного для выпуклой гладкой функции f.
15. Теорема о сходимости проксимального градиентного для сильно выпуклой гладкой функции f.
16. Теорема о сходимости стохастического градиентного спуска в гладком PL-случае.
17. Теорема сходимости траектории градиентного потока для выпуклых и PL-функций.