

## Определения и формулировки

1. Положительно определённая матрица.
2. Евклидова норма вектора.
3.  $p$ -норма вектора.
4. Норма Фробениуса матрицы.
5. Спектральная норма матрицы.
6. Скалярное произведение двух векторов.
7. Скалярное произведение двух матриц, согласованное с нормой Фробениуса.
8. Собственные значения матрицы. Спектр матрицы.
9. Связь спектра матрицы и её определённости.
10. Спектральное разложение симметричной матрицы.
11. Сингулярное разложение матрицы.
12. Связь определителя и собственных чисел квадратной матрицы.
13. Связь следа и собственных чисел квадратной матрицы.
14. Линейная сходимость последовательности.
15. Сублинейная сходимость последовательности.
16. Сверхлинейная сходимость последовательности.
17. Квадратичная сходимость последовательности.
18. Тест корней для определения скорости сходимости последовательности.
19. Тест отношений для определения скорости сходимости последовательности.
20. Градиент функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
21. Гессиан функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
22. Якобиан функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .
23. Аппроксимация Тейлора первого порядка  $f_{x_0}^I(x)$  функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
24. Аппроксимация Тейлора второго порядка  $f_{x_0}^{II}(x)$  функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
25. Связь дифференциала функции  $df$  и градиента  $\nabla f$  для функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
26. Связь второго дифференциала функции  $d^2 f$  и гессиана  $\nabla^2 f$  для функции  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ .
27. Унимодальная функция.
28. Метод дихотомии.
29. Метод золотого сечения.
30. Условие достаточного убывания (Армихо) для неточного линейного поиска.
31. Условия Гольдштейна для неточного линейного поиска.
32. Условия Вульфа для неточного линейного поиска.
33. Forward chain rule: формула  $\frac{\partial f}{\partial t}$  через  $\frac{\partial x_i}{\partial t}$  для  $f = f(x_1(t), \dots, x_n(t))$ .
34. Backward chain rule: формула  $\frac{\partial L}{\partial v_k}$  через  $\frac{\partial L}{\partial x_i}$  для скалярной  $L$  и векторного узла  $v_k$ .
35. Идея gradient checkpointing для обратного режима AD.
36. Аффинное множество. Аффинная комбинация. Аффинная оболочка.
37. Конус. Выпуклый конус. Коническая комбинация. Коническая оболочка.
38. Выпуклое множество. Выпуклая комбинация. Выпуклая оболочка.
39. Любые 2 операции с множествами, сохраняющие выпуклость.
40. Сумма Минковского.
41. Выпуклая функция.
42. Строго выпуклая функция.
43. Надграфик функции.
44. Множество подуровня (Лебега) функции.
45. Любые 2 операции с функциями, сохраняющие выпуклость.
46. Условие Поляка-Лоясиевича (градиентного доминирования).
47. Привести пример выпуклой, но не сильно выпуклой задачи линейных наименьших квадратов

- (возможно, с регуляризацией).
48. Привести пример сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
  49. Дифференциальный критерий выпуклости первого порядка.
  50. Дифференциальный критерий выпуклости второго порядка.
  51.  $\mu$ -сильно выпуклая функция.
  52. Дифференциальный критерий сильной выпуклости первого порядка.
  53. Дифференциальный критерий сильной выпуклости второго порядка.
  54. Необходимые условия безусловного локального экстремума.
  55. Достаточные условия безусловного локального экстремума.
  56. Общая задача математического программирования. Функция Лагранжа.
  57. Теорема Каруша-Куна-Таккера: необходимые условия для задачи математического программирования.
  58. Условие Слейтера.
  59. Задача выпуклого программирования.
  60. Лагранжева двойственная функция.
  61. Двойственная задача.
  62. Слабая двойственность. Зазор двойственности.
  63. Сильная двойственность.
  64. Задача линейного программирования. Стандартная форма задачи ЛП.
  65. Симплекс-метод (достаточно описать шаги алгоритма).
  66. Двухфазный симплекс-метод.
  67. Сходимость симплекс-метода.
  68. Двойственная задача к задаче ЛП.
  69. Направление антиградиента как направление наискорейшего локального убывания.
  70. Дифференциальное уравнение градиентного потока.
  71. Метод градиентного спуска.
  72. Наискорейший спуск.
  73. Как направлены градиенты на двух соседних итерациях метода наискорейшего спуска по отношению друг к другу?
  74. Характер сходимости градиентного спуска для гладких невыпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  75. Характер сходимости градиентного спуска для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  76. Характер сходимости градиентного спуска для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
  77. Связь PL-функций и сильно выпуклых функций.
  78. Нижние оценки для минимизации гладких выпуклых функций с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
  79. Нижние оценки для минимизации гладких сильно выпуклых функций с помощью методов первого порядка в терминах  $\Omega$  от числа итераций.
  80. Отличие ускоренной и неускоренной линейной сходимости методов первого порядка.
  81. Метод тяжёлого шарика (Поляка) для сильно выпуклой квадратичной функции. Характер сходимости. Оптимальные гиперпараметры.
  82. Локальная и глобальная сходимость численного метода оптимизации.
  83. Ускоренный градиентный метод Нестерова для выпуклых гладких функций. Характер сходимости в терминах  $\mathcal{O}$  от числа итераций.
  84. Ускоренный градиентный метод Нестерова для сильно выпуклых гладких функций. Характер сходимости в терминах  $\mathcal{O}$  от числа итераций.
  85.  $A$ -сопряжённость ( $A$ -ортогональность) векторов. Скалярное произведение  $\langle \cdot, \cdot \rangle_A$ .
  86. Процесс ортогонализации Грама-Шмидта.

87. Метод сопряжённых направлений (CD).
88. Метод сопряжённых градиентов (достаточно выписать формулы одной итерации).
89. Характер сходимости метода сопряжённых градиентов для сильно выпуклой квадратичной функции в терминах  $\mathcal{O}$  от числа итераций метода.
90. Зависимость сходимости метода сопряжённых градиентов от спектра матрицы  $A$ .
91. Липшицева парабола для гладкой функции.
92. Связь спектра гессиана с константами сильной выпуклости и гладкости функции. Число обусловленности.
93. Метод Ньютона.
94. Сходимость метода Ньютона для квадратичной функции.
95. Характер сходимости метода Ньютона для сильно выпуклых гладких функций — куда и как сходится.
96. Демпфированный метод Ньютона.
97. Метод SR-1 (без точной формулы).
98. Проекция точки на множество.
99. Достаточное условие существования проекции точки на множество.
100. Достаточное условие единственности проекции точки на множество.
101. Критерий проекции точки на выпуклое множество (неравенство Бурбаки-Чейни-Гольдштейна).
102. Оператор проекции как нерастягивающий оператор.
103. Метод проекции градиента.
104. Характер сходимости метода проекции градиента для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
105. Характер сходимости метода проекции градиента для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
106. Метод Франк-Вульфа.
107. Характер сходимости метода Франк-Вульфа для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
108. Характер сходимости метода Франк-Вульфа для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
109. Нижние оценки для условной оптимизации с помощью оракула линейной минимизации (LMO) для гладких сильно выпуклых функций в терминах  $\Omega$  от числа итераций метода.
110. Привести пример выпуклой негладкой задачи линейных наименьших квадратов (возможно, с регуляризацией).
111. Субградиент. Субдифференциал.
112. Субдифференциал поточечного максимума выпуклых функций (теорема Дубовицкого-Милютина).
113. Субградиентный метод.
114. Какому условию должна удовлетворять стратегия выбора шага, чтобы субградиентный метод сходился для выпуклых липшицевых функций?
115. Характер сходимости субградиентного метода для негладких выпуклых липшицевых функций в терминах  $\mathcal{O}$  от числа итераций. Стратегия выбора шага.
116. Характер сходимости субградиентного метода для негладких сильно выпуклых липшицевых функций в терминах  $\mathcal{O}$  от числа итераций. Стратегия выбора шага.
117. Нижние оценки для негладкой выпуклой оптимизации методами первого порядка в терминах  $\Omega$  от числа итераций.
118. Нижние оценки для негладкой сильно выпуклой оптимизации методами первого порядка в терминах  $\Omega$  от числа итераций.
119. Проксимальный оператор  $\text{prox}_{f,\alpha}$ .
120. Оператор проекции как частный случай проксимального оператора.
121. Аналитическое выражение для  $\text{prox}_{\lambda\|x\|_1}$ .

122. Аналитическое выражение для  $\text{prox}_{\frac{\lambda}{2}\|\cdot\|_2}$ .
123. Проксимальный оператор как нерастягивающий оператор.
124. Характер сходимости проксимального градиентного метода для гладких выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций метода.
125. Характер сходимости проксимального градиентного метода для гладких сильно выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций метода.
126. Характер сходимости ускоренного проксимального градиентного метода для гладких выпуклых функций  $f$  в терминах  $\mathcal{O}$  от числа итераций метода.
127. Метод стохастического градиентного спуска.
128. Идея мини-батча. Эпоха.
129. Характер сходимости стохастического градиентного спуска для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций.
130. Характер сходимости стохастического градиентного спуска для гладких PL-функций в терминах  $\mathcal{O}$  от числа итераций.
131. Характер работы стохастического градиентного спуска с постоянным шагом для гладких PL-функций.
132. Метод SAG (Stochastic Average Gradient).
133. Метод AdamW.
134. Сопряжённая (двойственная) норма.
135. Linear Minimization Oracle (LMO).
136. Связь LMO с сопряжённой нормой.
137. Неевклидов градиентный спуск.
138. Неевклидов нормированный градиентный спуск.
139. Неевклидов градиентный спуск в случае, когда аргумент — матрица.
140. Неевклидов нормированный спуск в случае, когда аргумент — матрица.
141. Полярный фактор матрицы.
142. LMO для спектральной нормы.
143. Формулировка метода нормированного спуска в норме Фробениуса.
144. Формулировка метода нормированного спуска в поэлементной  $\ell_\infty$ -норме.
145. Формулировка метода нормированного спуска в спектральной норме.
146. Метод Muon.
147. Итерации Ньютона-Шульца для вычисления полярного фактора.
148. Сколько памяти занимают веса крупнейшей открытой модели (DeepSeek V4-Pro,  $\approx 1,6$  трлн параметров) в форматах FP32, FP16, BF16?
149. Сколько токенов в обучающей выборке приходилось на модель Chinchilla на один обучаемый параметр модели?
150. Идея warmup в расписании learning rate.
151. Идея аккумуляции градиентов.
152. Дообучение с помощью LoRA-адаптеров.
153. Идея проекции функции потерь нейронной сети на прямую и на плоскость.
154. Grokking.
155. Double Descent.
156. Сопряжённая функция  $f^*$ .
157. Метод двойственного градиентного подъёма (dual ascent).
158. Связь сильной выпуклости функции  $f$  и липшицевости градиента сопряжённой  $f^*$ .
159. Идея двойственной декомпозиции (dual decomposition).
160. Метод модифицированной функции Лагранжа (ALM).
161. Метод ADMM.

## Теоремы с доказательствами

1. Тест корней для определения скорости сходимости последовательности.
2. Скорость сходимости метода дихотомии для унимодальных функций.
3. Скорость сходимости метода золотого сечения для унимодальных функций.
4. Автоматическое дифференцирование. Вычислительный граф. Алгоритмы прямого (forward) и обратного (reverse) режима (без доказательств; описать алгоритмы подробно).
5. Выпуклость надграфика как критерий выпуклости функции.
6. Дифференциальный критерий сильной выпуклости первого порядка (доказательство).
7. Дифференциальный критерий сильной выпуклости второго порядка (доказательство).
8. Необходимое условие безусловного экстремума первого порядка (доказательство).
9. Достаточные условия безусловного экстремума второго порядка (доказательство).
10. Теорема о проверке оптимальности решения в задаче линейного программирования.
11. Теорема о сильной двойственности в задаче ЛП.
12. Теорема сходимости градиентного спуска для гладких выпуклых функций.
13. Теорема сходимости градиентного спуска для гладких PL-функций.
14. Сходимость градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
15. Доказательство сходимости метода сопряжённых градиентов и вывод формул метода (за какое число шагов сходится метод, как выбираются направления и почему в  $A$ -ортогонализации достаточно хранить лишь предыдущее направление).
16. Теорема о локальной квадратичной сходимости метода Ньютона для сильно выпуклых функций с липшицевым гессианом.
17. Вывод формул обновления оценок гессиана и обратного гессиана в квазиньютоновских методах SR1, DFP, BFGS.
18. Теорема о сходимости метода проекции градиента для гладких выпуклых функций.
19. Теорема о сходимости метода проекции градиента для гладких сильно выпуклых функций.
20. Теорема о сходимости метода Франк-Вульфа для гладких выпуклых функций.
21. Теорема о сходимости субградиентного метода для выпуклых липшицевых функций. Стратегии выбора шага для сходимости.
22. Теорема о сходимости субградиентного метода для сильно выпуклых липшицевых функций.
23. Теорема о сходимости проксимального градиентного метода для выпуклой гладкой функции  $f$ .
24. Теорема о сходимости проксимального градиентного метода для сильно выпуклой гладкой функции  $f$ .
25. Теорема о сходимости SGD для гладких PL-функций с постоянным и убывающим шагом.