



## Определения и формулировки

- 1. Положительно определённая матрица.
- 2. Евклидова норма вектора.
- 3. *p*-норма вектора.
- 4. Как выглядит единичный шар в p норме на плоскости для  $p=1,2,\infty$ ?
- 5. Норма Фробениуса для матрицы.
- 6. Спектральная норма матрицы.
- 7. Скалярное произведение двух векторов.
- 8. Скалярное произведение двух матриц, согласованное с нормой Фробениуса.
- 9. Собственные значения матрицы. Спектр матрицы.
- 10. Связь спектра матрицы и её определенности.
- 11. Спектральное разложение матрицы.
- 12. Сингулярное разложение матрицы.
- 13. Связь определителя и собственных чисел для квадратной матрицы.
- 14. Связь следа и собственных чисел для квадратной матрицы.
- 15.  $\Lambda$ инейная сходимость последовательности.
- 16. Сублинейная сходимость последовательности.
- 17. Сверхлинейная сходимость последовательности.
- 18. Квадратичная сходимость последовательности.
- 19. Рассмотрим итеративный процесс  $x_k \in \mathbb{R}$ , сходящийся к решению  $x^* \in \mathbb{R}$  квадратично. Как изменяется количество верных значащих цифр в решении после одной итерации метода?
- 20. Формулировка теста корней для определения скорости сходимости последовательности.
- 21. Формулировка теста отношений для определения скорости сходимости последовательности.
- 22. Унимодальная функция.
- 23. Метод дихотомии.
- 24. Метод золотого сечения.
- 25. Метод параболической интерполяции.
- 26. Условие достаточного убывания для неточного линейного поиска.
- 27. Условия Гольдштейна для неточного линейного поиска.
- 28. Условие ограничения на кривизну для неточного линейного поиска.
- 29. Градиент функции  $f(x): \mathbb{R}^n \to \mathbb{R}$ .
- 30. Гессиан функции  $f(x): \mathbb{R}^n \to \mathbb{R}$ .
- 31. Якобиан функции  $f(x): \mathbb{R}^n \to \mathbb{R}^m$ .
- 32. Формула для аппроксимации Тейлора первого порядка  $f_{x_0}^I(x)$  функции  $f(x):\mathbb{R}^n o \mathbb{R}$  в точке  $x_0$ .
- 33. Формула для аппроксимации Тейлора второго порядка  $f_{x_0}^{II}(x)$  функции  $f(x):\mathbb{R}^n o \mathbb{R}$  в точке  $x_0$ .
- 34. Связь дифференциала функции df и градиента  $\nabla f$  для функции  $f(x):\mathbb{R}^n \to \mathbb{R}.$
- 35. Связь второго дифференциала функции  $d^2f$  и гессиана  $\nabla^2f$  для функции  $f(x):\mathbb{R}^n \to \mathbb{R}.$
- 36. Формула для приближенного вычисления производной функции f(x) :  $\mathbb{R}^n$  ightarrow  $\mathbb{R}$  по k-ой координате с помощью метода конечных разностей.
- 37. Пусть  $f=f(x_1(t),\dots,x_n(t))$ . Формула для вычисления  $\frac{\partial f}{\partial t}$  через  $\frac{\partial x_i}{\partial t}$  (Forward chain rule).
- 38. Пусть L функция, возвращающая скаляр, а  $v_k$  функция, возвращающая вектор  $x \in \mathbb{R}^t$ . Формула для вычисления  $\frac{\partial L}{\partial v_k}$  через  $\frac{\partial L}{\partial x_i}$  (Backward chain rule).
- 39. Идея Хатчинсона для оценки следа матрицы с помощью matvec операций.
- 40. Афинное множество. Афинная комбинация. Афинная оболочка.
- 41. Выпуклое множество. Выпуклая комбинация. Выпуклая оболочка.
- 42. Конус. Выпуклый конус. Коническая комбинация. Коническая оболочка.
- 43. Внутренность множества.
- 44. Относительная внутренность множества.





- 45. Сумма Минковского.
- 46. Любые 2 операции с множествами, сохраняющие выпуклость.
- 47. Выпуклая функция.
- 48. Строго выпуклая функция.
- 49. Надграфик функции  $f(x): \mathbb{R}^n \to \mathbb{R}$ .
- 50. Множество подуровней функции  $f(x): \mathbb{R}^n \to \mathbb{R}$ .
- 51. Дифференциальный критерий выпуклости первого порядка.
- 52. Дифференциальный критерий выпуклости второго порядка.
- 53. Связь выпуклости функции и её надграфика.
- 54.  $\mu$ -сильно выпуклая функция.
- 55. Дифференциальный критерий сильной выпуклости первого порядка.
- 56. Дифференциальный критерий сильной выпуклости второго порядка.
- 57. Любые 2 операции с функциями, сохраняющие выпуклость.
- 58. Теорема Тейлора.
- 59. Необходимые условия локального экстремума.
- 60. Достаточные условия локального экстремума.
- $61.\,$  Общая задача математического программирования. Функция  $\varLambda$ агранжа.
- 62. Теорема Каруша Куна Таккера в форме необходимых условий решения задачи математического программирования.
- 63. Условие Слейтера.
- 64. Задача выпуклого программирования.
- 65. Двойственная функция в задаче математического программирования.
- 66. Двойственная задача для задачи математического программирования.
- 67. Сильная двойственность. Зазор двойственности.
- 68. Локальный анализ чувствительности с помощью множителей Лагранжа.
- 69. Задача линейного программирования. Задача линейного программирования в стандартной форме.
- 70. Возможные случаи двойственности в задаче линейного программирования.
- 71. Симплекс метод.
- 72. Нахождение первоначальной угловой точки с помощью двухфазного симплекс метода.
- 73. Сходимость симплекс метода.
- 74. Показать, что направление антиградиента направление наискорейшего локального убывания функции.
- 75. Дифференциальное уравнение градиентного потока.
- 76. Метод градиентного спуска.
- 77. Наискорейший спуск.
- 78. Как направлены две соседние итерации метода наискорейшего спуска по отношению друг к другу?
- 79. Липшицева парабола для гладкой функции.
- 80. Размер шага наискорейшего спуска для квадратичной функции.
- 81. Характер сходимости градиентного спуска к локальному экстремуму для гладких невыпуклых функций в терминах  $\mathcal O$  от числа итераций метода.
- 82. Характер сходимости градиентного спуска для гладких выпуклых функций в терминах  $\mathcal O$  от числа итераций метода.
- 83. Характер сходимости градиентного спуска для гладких и сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
- 84. Связь спектра гессиана с константами сильной выпуклости и гладкости функции.
- 85. Условие Поляка-Лоясиевича (градиентного доминирования) для функций.
- 86. Сходимость градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.



- 87. Связь PL-функций и сильно выпуклых функций.
- 88. Привести пример выпуклой, но не сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
- 89. Привести пример сильно выпуклой задачи линейных наименьших квадратов (возможно, с регуляризацией).
- 90. Нижние оценки для гладкой выпуклой оптимизации с помощью методов первого порядка в терминах  $\mathcal{O}$  от числа итераций метода.
- 91. Нижние оценки для гладкой сильно выпуклой оптимизации с помощью методов первого порядка в терминах  $\mathcal{O}$  от числа итераций метода.
- 92. Отличие ускоренной и неускоренной линейной сходимости для методов первого порядка.
- 93. Метод тяжелого шарика (Поляка).
- 94. Понятие локальной и глобальной сходимости численного метода оптимизации.
- 95. Ускоренный градиентный метод Нестерова для выпуклых гладких функций.
- 96. Ускоренный градиентный метод Нестерова для сильно выпуклых гладких функций.
- 97.  $\,A$ -сопряженность двух векторов.  $\,A$ -ортогональность. Скалярное произведение  $\langle\cdot,\cdot
  angle_{A}.$
- 98. Процедура ортогонализации Грама-Шмидта.
- 99. Метод сопряженных направлений.
- 100. Метод сопряженных градиентов.
- 101. Зависимость сходимости метода сопряженных градиентнов от спектра матрицы.
- 102. Характер сходимости метода сопряженных градиентов в терминах  $\mathcal O$  от числа итераций метода.
- 103. Метод Полака-Рибьера.
- 104. Метод Ньютона.
- 105. Сходимость метода Ньютона для квадратичной функции.
- 106. Характер сходимости метода Ньютона для сильно выпуклых гладких функций куда и как сходится.
- 107. Демпфированный метод Ньютона.
- 108. Идея квазиньютоновских методов. Метод SR-1.
- 109. Афинная инвариантность метода Ньютона.
- 110. Проекция.
- 111. Достаточное условие существования проекции точки на множество.
- 112. Достаточное условие единственности проекции точки на множество.
- 113. Метод проекции градиента.
- 114. Критерий проекции точки на выпуклое множество (Неравенство Бурбаки-Чейни-Гольдштейна).
- 115. Проекция как нерастягивающий оператор.
- 116. Метод Франк-Вульфа.
- 117. Характер сходимости метода проекции градиента для гладких выпуклых функций в терминах  ${\mathcal O}$ от числа итераций метода.
- 118. Характер сходимости метода проекции градиента для гладких сильно выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
- 119. Характер сходимости метода Франк-Вульфа для гладких выпуклых функций в терминах  $\mathcal O$  от числа итераций метода.
- 120. Характер сходимости метода Франк-Вульфа для гладких сильно выпуклых функций в терминах  ${\mathcal O}$ от числа итераций метода.
- 121. Нижние оценки для условной оптимизации с помощью оракула линейной минимизации для сильно выпуклых гладких функций в терминах  $\mathcal O$  от числа итераций метода.
- 122. Привести пример выпуклой негладкой задачи линейных наименьших квадратов (возможно, с регуляризацией).
- 123. Субградиент. Субдифференциал.
- 124. Как считать субдифференциал поточечного максимума выпуклых функций.
- 125. Субградиентный метод.





- 126. Характер сходимости субградиентного метода для негладких выпуклых Липшицевых функций в терминах  $\mathcal{O}$  от числа итераций метода.
- 127. Какому условию должна удовлетворять стратегия выбора шага, чтобы субградиентный метод сходился для выпуклых Липшицевых функций?
- 128. Характер сходимости субградиентного метода для негладких сильно выпуклых Липшицевых функций в терминах  $\mathcal O$  от числа итераций метода. Стратегия выбора шага.
- 129. Нижние оценки для негладкой выпуклой оптимизации с помощью методов первого порядка в терминах  $\mathcal{O}$  от числа итераций метода.
- 130. Нижние оценки для негладкой сильно выпуклой оптимизации с помощью методов первого порядка в терминах  $\mathcal O$  от числа итераций метода.
- 131. Проксимальный оператор.
- 132. Оператор проекции как частный случай проксимального оператора.
- 133. Характер сходимости проксимального градиентного метода для гладких выпуклых функций f в терминах  $\mathcal{O}$  от числа итераций метода.
- 134. Характер сходимости проксимального градиентного метода для гладких сильно выпуклых функций f в терминах  $\mathcal{O}$  от числа итераций метода.
- 135. Аналитическое выражение для  $\operatorname{prox}_{\lambda \|x\|_1}$ .
- 136. Аналитическое выражение для  $\operatorname{prox}_{\frac{\mu}{2}\|x\|_2^2}$ .
- 137. Проксимальный оператор как нерастягивающий оператор.
- 138. Характер сходимости ускоренного проксимального градиентного метода для гладких выпуклых функций f в терминах  $\mathcal O$  от числа итераций метода.
- 139. Метод стохастического градиентного спуска.
- 140. Идея мини-батча для метода стохастического градиентного спуска. Эпоха.
- 141. Характер сходимости стохастического градиентного спуска для гладких выпуклых функций в терминах  $\mathcal{O}$  от числа итераций метода.
- 142. Характер сходимости стохастического градиентного спуска для гладких PL-функций в терминах  ${\mathcal O}$ от числа итераций метода.
- 143. Характер работы стохастического градиентного спуска с постоянным шагом для гладких PLфункций.
- 144. Основная идея методов уменьшения дисперсии.
- 145. Meтод SVRG.
- 146. Метод SAG.
- 147. Метод Adagrad.
- 148. Метод AdamW.
- 149. Метод Shampoo.
- 150. Метод Muon.
- 151. Как сравниваются методы в AlgoPerf Benchmark.
- 152. Идея проекции функции потерь нейронной сети на прямую, плоскость.
- 153. Grokking.
- 154. Double Descent.
- 155. Взрыв/Затухание градиентов при обучении глубоких нейронных сетей.
- 156. Идея gradient checkpointing.
- 157. Идея аккумуляции градиентов.
- 158. Зачем увеличивать батч при обучении больших нейросетевых моделей. Warmup.
- 159. Идея cooldown фазы для построения расписания learning rate. В чём преимущество по сравнению c cosine scheduler?
- 160. Data Parallel обучение на нескольких видеокартах.
- 161. GPipe Pipeline параллелизм.
- 162. PipeDream Pipeline параллелизм.
- 163. Дообучение больших моделей с помощью LoRA адаптеров.





- 164. Сколько токенов в обучающей выборке приходилось на модель Chinchilla на один обучаемый параметр модели?
- 165. Метод двойственного градиентного подъема.
- 166. Связь константы сильной выпуклости f и гладкости  $f^*$ .
- 167. Идея dual decomposition.
- 168. Метод двойственного градиентного подъема для линейных ограничений-неравенств.
- 169. Метод модифицированной функции Лагранжа.
- 170. Метод ADMM.

## Теоремы с доказательствами

- 1. Тест корней для определения скорости сходимости последовательности.
- 2. Метод дихотомии и золотого сечения для унимодальных функций. Скорость сходимости.
- 3. Автоматическое дифференцирование. Вычислительный граф. Forward/ Backward mode (в этом вопросе нет доказательств, но необходимо подробно описать алгоритмы).
- 4. Выпуклость надграфика как критерий выпуклости функции.
- 5. Дифференциальный критерий сильной выпуклости первого порядка.
- 6. Дифференциальный критерий сильной выпуклости второго порядка.
- 7. Необходимые условия безусловного экстремума.
- 8. Достаточные условия безусловного экстремума.
- 9. Симплекс метод для задачи линейного программирования с ограничениями в форме неравенств. Двухфазный симплекс метод.
- 10. Теорема о проверке оптимальности решения в линейном программировании. Теорема о сильной двойственности в задаче линейного программирования.
- 11. Теорема сходимости градиентного спуска для гладких выпуклых функций.
- 12. Теорема сходимости градиентного спуска для гладких PL функций.
- 13. Теорема сходимости градиентного спуска для сильно выпуклых квадратичных функций. Оптимальные гиперпараметры.
- 14. Вывод ускоренного метода для квадратичной функции с помощью полиномов Чебышёва.
- 15. 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 А-ортогонализации достаточно хранить лишь предыдущий шаг метода, а не все предыдущие).
- 16. Теорема сходимости метода Ньютона для сильно выпуклых функций с Липшицевым гессианом.
- 17. Теорема о сходимости метода проекции градиента для выпуклых гладких функций.
- 18. Теорема о сходимости метода проекции градиента для сильно выпуклых гладких функций.
- 19. Теорема о сходимости метода Франк-Вульфа для выпуклых гладких функций.
- 20. 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?
- $21.\,$  Теорема о сходимости субградиентного метода для сильно выпуклых Липшицевых функций.
- 22. Вывод формул обновления оценок обратного гессиана и гессиана квазиньютоновских методов SR-1, DFP, BFGS.
- 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 стохастического градиентного спуска в гладком PL-случае.