Численные методы, Весна 2021 ВШЭ. Задание 10.[a]

1. **Восстановление зашумлённого изображения (30)** Возьмите файл data_5_1.npz, содержащий матрицы A и C (изображение и фильтр) и откройте его, используя numpy:

```
with np.load('data.npz') as data:
    A, C = data['A'], data['C']
```

Для работы с изображением нам будет удобно выстраивать элементы матрицы A в вектор–столбец a:

```
def mat2vec(A):
    h, w = A.shape
    a = np.zeros(h*w, dtype=A.dtype)
    A = np.flipud(A)
    for i, row in enumerate(A):
        a[i*w:i*w+w] = row
    return a
```

так что обратное преобразование – от вектора a в матрицу A осуществляется с помощью функции

```
def vec2mat(a, shape):
    h, w = shape
    A = np.zeros(shape, dtype=a.dtype)
    for i in range(h):
        A[i, :] = a[i*w:i*w+w]
    return np.flipud(A)
```

Изображение, содержащееся в матрице A получено из некоторого оригинала A_0 путем свёртки его с фильтром C и добавлением шума. Фильтр C осуществляет ‘размытие’ изображения, одновременно меняя его размер от 16×51 к 25×60. Используя соответствующие вектора a и a_0, эту операцию можно записать так

$$a_0 \to a = Ca_0 + \epsilon,$$

где ϵ - вектор, состоящий из нескоррелированных случайных величин из нормального распределения. Ваша задача состоит в том, чтобы располагая зашумленным изображением A и фильтром C, восстановить исходное изображение A_0.

- **(5)** Постройте изображение, содержащееся в A (у вас должен получиться Рис. 1).
- **(7)** Исследуйте действие фильтра C на изображения: составьте (на свой выбор) матрицу, и проверьте, что с соответствующим изображением делает фильтр C. Вам понадобятся операции `a = mat2vec(A)` и `A0 = vec2mat(a0, shape)` для перехода от матричного к векторному представлению и обратно.
- **(8)** Наивный способ восстановить изображение A_0 по изображению A состоит в том, чтобы решить систему $a = Ca_0$ относительно вектора a_0. Какой является эта система: недо– или переопределённой? Используйте SVD матрицы C чтобы найти a_0 и постройте соответствующее изображение A_0.
- **(10)** Для того, чтобы улучшить результат, поэкспериментируйте с количеством удержанных собственных значений при решении системы уравнений в предыдущем пункте. Что находится на изображении A_0?

2. **Ранжирование веб–страниц (35)** Одна из самых известных задач о вычислении собственных векторов – задача о ранжировании n веб-страниц. Подход, который вам нужно будет реализовать в этой задаче, был одним из главных в работе Google на раннем этапе. Всё, что мы собираемся использовать – структуру

[a] Дополнительно указаны: (количество баллов за задачу)[имя задачи на nbgrader]

взаимных ссылок между страницами. `PageRank` определяется рекурсивно: важность i–й страницы определяется как среднее значение важностей всех страниц, которые ссылаются на i–ю. Обозначим важность i–й страницы p_i, тогда это определение может быть записано в виде линейной системы:

$$p_i = \sum_j \frac{p_j}{L(j)} l_{ij},$$

где $l_{ij} = 1$ если j–я страница ссылается на i–ю (в противном случае $l_{ij} = 0$), а $L(j)$ – количество исходящих ссылок со страницы j. Система может быть переписана в виде задачи на собственное значение:

$$p = Gp, \quad G_{ij} = \frac{l_{ij}}{L(j)}.$$

Если в графе есть 'подвешенные' узлы (все элементы какого-то столбца равны нулю), то весь столбец заполняется числом $1/n$. Наконец, вводится параметр $0 < \beta < 1$ так что матрица G заменяется на

$$G \to \beta G + \frac{1-\beta}{n} ee^T,$$

где e – вектор, состоящий из единиц. Обратите внимание, что задача свелась к нахождению собственного вектора p матрицы G, отвечающего собственному значению 1. Можно показать [*], что 1 – максимально возможное собственное значение матрицы G.

- **(4)** Придумайте самостоятельно небольшой граф связности (10 узлов), постройте соответствующие матрицы l и G и найдите численно собственный вектор, отвечающий `PageRank`.
- **(1)** Скачайте файл, в котором представлен ориентированный граф, узлы которого составляют страницы stanford.edu, а направленные рёбра – ссылки между ними (граф задан матрицей смежности l). Распакуйте архив и загрузите его:

```
from scipy import sparse
def dataset2csr(filename, nodes, edges):
    rows = []; cols = []
    with open(filename, 'r') as f:
        for line in f.readlines()[4:]:
            o, d = (int(x)-1 for x in line.split())
            rows.append(d)
            cols.append(o)
    return(sparse.csr_matrix(([True]*edges, (rows, cols)), shape=(nodes, nodes)))

l = dataset2csr(filename='web-Stanford.txt', nodes = 281903, edges=2312497)
```

- **(6)** Найдите `PageRank` для матрицы из предыдущего пункта. Для этого реализуйте степенную итерацию для нахождения собственного вектора, отвечающего максимальному собственному значению G. Возьмите $\beta = 0.8$.
- **(9)** Итерируйте до тех пор, пока 1–норма изменения вектора-кандидата не станет меньше 10^{-4}. Сколько итераций понадобилось?
- **(5)** Какому собственному значению отвечает найденный вектор и у какого узла наибольший `PageRank`?
- **(10)** Докажите утверждение (*).

3. **(20)** Рассмотрите матрицу A размера 32×32, задаваемую следующей формулой:

$$A_{ij} = -\delta_{i,j} + \delta_{i,j-1} + \delta_{i,j-2}.$$

- **(5)** Найдите спектр матрицы A.
- **(5)** Используя функцию `scipy.linalg.expm`, постройте $||e^{At}||_2$ как функцию t на интервале $0 \le t \le 50$.
- **(10)** Используя (без доказательства) эквивалентность утверждений i и iii из задачи 26.1 Trefethen, Bau, изобразите в комплексной плоскости множество σ_ϵ [см. определение в Задаче 3 листка 4] для $\epsilon = 10^{-i}, \quad i = 1, .., 5$.

4. **(15)** Постройте интерполяцию функций

$$y_1(x) = \sin(6x) + \sin(60e^x),\ y_2(x) = \frac{1}{1 + 1000(x+0.5)^2} + \frac{1}{\sqrt{1 + 1000(x-0.5)^2}}$$

на отрезке $[-1, 1]$ полиномами Чебышева. Постройте зависимость ошибки приближения от количества узлов. Сколько узлов нужно удержать в каждом из этих случаев для получения достаточно точного приближения?

5. **(30) Изотерма реального газа)** Рассмотрим изотермы реального газа, описываемого уравнением ван-дер-Ваальса

$$p = \frac{RT}{V-b} - \frac{a}{V^2}\ .$$

Здесь p – давление газа, V – объем, T — температура, а a, b и R — постоянные. Видно, что при постоянной температуре T ниже определенной (так называемой критической) температуры, зависимость давления от объема становится немонотонной. Физически эта область — область сосуществования жидкость-газ, а часть изотермы, на которой объем растет при увеличении давления является нефизической. Фактически же давление остается постоянным во всей области сосуществования, причем положение горизонтального участка изотермы определяется правилом Максвелла и геометрически находится из равенства площадей участков изотермы. (см. напр., Д.В. Сивухин *Термодинамика* или R. Swendsen, *An introduction to statistical mechanics and thermodynamics* или любой иной учебник по термодинамике). Построить изотермы газа с учетом конструкции Максвелла, т.е. вычислить положение горизонтального участка изотермы *Указание: Первым действием найти критические параметры (10) — т.е. найти точку, в которой при обращаются в ноль первая и вторая производные давления по объему — и переписать уравнение состояния в безразмерных единицах (20).*

6. **(30) Прямоугольная яма II** Рассмотрим квантовую частицу в прямоугольной яме. В задаче SquareWell (листок 7) были найдены собственные значения и собственные функции гамильтониана. Также были построены — сеточным методом — собственные функции частицы с дополнительным возмущающим потенциалом вида $V(x) = \gamma x(x-a)$. Теперь рассмотрите задачу в базисе собственных функций невозмущенного гамильтониана (просто прямоугольной ямы) и найдите уровни энергии и собственные функции возмущенной системы с потенциалом $V(x)$. Сравните результаты с решением задачи сеточным методом. Сколько собственных значений невозмущенной системы нуsжно взять для того, чтобы получить значение энергий первых трех состояний с точностью 1%? Для малых значений γ сравните результаты с теорией возмущений (см. например Ландау-Лифшиц, Квантовая механика).

7. **(30) Одномерная диффузия** Рассмотрим простейшее одномерное уравнение диффузии функции распределения частиц по импульсам $f(p,t)$:

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial p}\left(D\frac{\partial f}{\partial p}\right) \tag{1}$$

В тривиальном случае, когда $D = const$, оно имеет аналитическое решение.

(5) Получить аналитическое решение уравнения (1) в случае, когда $D = 1$, $f(t = 0, p) = \delta(p)$. *Указание: перейти к Фурье-пространству, рассмотреть уравнение на Фурье-образ функции распределения.*

(10) Численно решить уравнение (1) при тех же условиях, что в предыдущем пункте на временном интервале $t = [0, 100]$.

К решению уравнения диффузии можно подойти и с другой стороны: при диффузии траектория каждой частицы в фазовом пространстве является случайной, т.е. для ее описания можно использовать модель случаного блуждания. Можно показать, что импульс каждой частицы будет являться случайной величиной и его изменение в каждый момент времени можно описать формулой:

$$p(t + dt) = p(t) + \sqrt{2D}W_t \tag{2}$$

где $W_t = \varepsilon\sqrt{dt}$ – изменение Винеровского процесса, $\varepsilon \sim N(0, 1)$ – нормально-распределенная случайная величина.

(10) Напишите программу, которая будет моделировать случайные фазовые траектории $\sim 10^6$ частиц и позволит проследить эволюцию функции распределения во времени. Для этого нужно: 1) задать ансамбль частиц с импульсами, удовоетворяющими начальному условию, 2) задать шаг по времени, 3) для каждой частицы на каждом шаге сгенерерировать нормальное случайное число и рассчитать изменение импульса, 4) собрать функцию распределения частиц, зная импульс каждой частицы (бинировать импульсы и посчитать, какое количество частиц попало в каждый бин).

(5) Сравнить все три полученных решения.

8. **(30) Metropolis-Hastings** В 1986 г. космический шатл Challenger взорвался во время старта, семь астронавтов погибли. Оказалось, что взрыв был вызван разрывом резиновых элементов (O-ring failure), скрепляющих детали обшивки. Авария произошла при относительно низкой температуре (31° F, или 0° C), что наводит на мысль о повышении вероятности аварии с понижением температуры (резина “задубела”). Данные с предыдущих (к счастью, безаварийных) запусков по наличию факта разрыва резиновых элементов (O-ring failures) представлены в файле `Challenger.csv`. Первая колонка – индикатор (1/0 = есть/нет O-ring failures), вторая колонка отвечает температуре окружающей среды T_i при i-ом старте.

   (a) **(5)** Рассмотрите модель логистической регрессии для аппроксимации вероятности O-ring failure, т.е.

   $$\Pr(Y_i = 1|T_i) \stackrel{\text{def}}{=} p(T_i) = \frac{e^{\alpha+\beta T_i}}{1 + e^{\alpha+\beta T_i}} \tag{3}$$

   Найти оценки параметров α и β, минимизируя квадрат ошибки.

   (b) **(5)** Применяя формулу Байеса, записать апостериорное совместное распределение величин α и β, используя равномерные априорные распределения (uniform priors), т.е.

   $$p(\alpha, \beta) = c\mathcal{L}(T_1, ..., T_n; \alpha, \beta)p_0(\alpha)g_0(\beta) \tag{4}$$

   где $p_0(\alpha) \sim \text{Unif}(0, 30)$, $g_0(\beta) \sim \text{Unif}(-1, 1)$ – плотности равномерного закона. Нормировочная константа c не важна.

   Указание: пусть при T_1 индикатор O-ring failure $y_1 = 1$, при T_2 индикатор $y_2 = 0$ и т.д. Тогда функция правдоподобия начинается так: $\mathcal{L}(T_1, ..., T_n; \alpha, \beta) = p(T_1) \times (1 - p(T_2)) \times ...$, т.к. (3) есть вероятность $Y = 1$ при температуре T_1, а $1 - p(T_2)$ есть вероятность $Y = 0$ при температуре T_2.

   (c) **(10)** Реализуйте алгоритм Метрополиса-Гастингса для генерации величин из распределения (4).

   (d) **(5)** Запустите соотвествующую марковскую цепь длины не менее 50000 и получите предсказание (т.е. среднее) вероятности O-ring failure при 60, 50 и 40 °F. Найти также с.к.о. указанных вероятностей и построить гистрограммы данных вероятностей.

   (e) **(3)** Построить маржинальные распределения (гистограммы) для α и β.

   (f) **(2)** Предложите какую-либо диагностику, удостоверяющую, что марковская цепь вышла на стационарный режим.

   Все гистограммы должны иметь по крайней мере 15 столбцов.

9. **(15) Условие излучения** Рассматривается монохроматическая плоская волна $e^{ikx-i\omega t}$, распространяющаяся в положительном направлении оси x. После отделения времени уравнение для пространственной части имеет вид

   $$\frac{d^2u}{dx^2} + k^2u = 0 \tag{5}$$

   (a) Рассмотрите указанное уравнение на отрезке $[0, 1]$. На левом конце поставьте условие падающей волны, $u(0) = e^{ikx}\big|_{x=0} = 1$. На правом конце поставьте т.н. условие излучения,

   $$\left.\left(\frac{du}{dx} - iku\right)\right|_{x=1} = 0 \tag{6}$$

   (b) **(1)** Попытайтесь объяснить логику условия (6)

   (c) **(10)** Дискретизируйте у-ние (5) и, главное, условие излучения (6) со **вторым** порядком точности. *Указание: с уравнением проблем нет, вспомните задание по у-нию Шредингера. С условием (6) наиболее простой подход – использовать равномерную сетку с узлами, сдвинутыми относительно концов интервала на полшага. При таком подходе не забудтье записать условие на левом конце $x = 0$ со вторым порядком точности (это несложно).*

(d) **(2)** Решите соотвествующую трехдиагональную систему в комплексных числах. Сравните полученное решение с точным e^{ikx} на двух последовательных сетках с шагами, отличающимися в два раза. Во сколько раз упадет ошибка на мелкой сетке по сравнению с грубой? Волновое число задайте самостоятельно, например, $k = 1$.

(e) **(2)** Поставьте на правом конце $x = 1$ какое-нибудь другое произвольное условие, например $u(1) = 1$. Сравните полученное решение с точным.