

Санкт-Петербургский государственный университет
Прикладная математика и информатика

ОТЧЕТ

по учебной практике 1 (научно-исследовательской работе)
(семестр 1)

Градиентный спуск при больших шагах: хаос и фрактальная область сходимости

Выполнил:

Сайфутдинов Анвар Ирекович,
группа 25.Б21-мм

Научный руководитель:

д.ф.-м.н., профессор
Мокаев Тимур Назирович
Кафедра прикладной кибернетики

Санкт-Петербург
2025 г.

Содержание

1	Введение	2
2	Цель работы и постановка задачи	3
2.1	Скалярная факторизация	3
2.2	Итерация градиентного спуска	3
2.3	Критические точки	3
2.4	Матричная факторизация	4
2.5	Критический шаг сходимости	4
3	Обзор литературы	4
3.1	Основной источник	4
3.2	Фрактальные границы бассейнов притяжения	4
3.3	Edge of Stability	5
3.4	Catapult Mechanism	5
3.5	Хаос в квадратичной регрессии	5
4	Методы исследования	6
4.1	Построение карты бассейнов	6
4.2	Метод box-counting	6
4.3	Сравнение с теоретическим η^*	6
5	Результаты экспериментов	7
5.1	Бассейны сходимости: скалярный случай	7
5.2	Влияние регуляризации	8
5.3	Фрактальная размерность границы	8
5.4	Чувствительность к начальным условиям	9
5.5	Сравнение с теоретическим η^*	9
5.6	Матричная факторизация	10
5.7	Траектории и динамика loss	10
5.8	Связь с Edge of Stability и Catapult	11
6	Выводы	12
7	Заключение	12
A	Приложение. Исходный код	15
A.1	Градиент и критический шаг	15
A.2	Box-counting	15

1 Введение

Градиентный спуск — основной метод оптимизации в машинном обучении. Он используется для обучения нейронных сетей, решения задач регрессии, классификации и многих других. Суть метода: чтобы найти минимум функции потерь, мы итеративно двигаемся в направлении антиградиента.

Ключевой параметр метода — *размер шага* (learning rate) η . При малом η сходимость гарантирована, но медленная. При большом η можно сойтись быстрее, но возникает риск неустойчивости: траектория может «прыгать» через минимум или вовсе расходиться.

Недавние исследования [5] показали неожиданный результат: при больших η динамика градиентного спуска становится *хаотической*. Это проявляется в двух феноменах:

1. **Чувствительность к начальным условиям:** малейшие изменения начальной точки приводят к совершенно разным результатам оптимизации.
2. **Фрактальные границы бассейнов:** области сходимости к разным минимумам разделены границей со сложной, самоподобной структурой.

Эти явления важны для практики машинного обучения, поскольку объясняют, почему результаты обучения нейронных сетей могут сильно зависеть от случайной инициализации весов.

Цель работы — изучить хаотическую динамику градиентного спуска на примере задачи матричной факторизации, реализовать численные эксперименты и сопоставить результаты с теоретическими предсказаниями.

Задачи:

1. Изучить основные результаты статьи Liang и Montúfar о критическом шаге и фрактальных границах.
2. Реализовать градиентный спуск для скалярной и матричной факторизации.
3. Визуализировать бассейны сходимости и оценить фрактальную размерность границы.
4. Сравнить экспериментальные результаты с теоретической формулой критического шага.
5. Рассмотреть связь с феноменами «edge of stability» и «catapult mechanism».

2 Цель работы и постановка задачи

2.1 Скалярная факторизация

Начнём с простейшего случая — *скалярной факторизации*. Задача: найти числа u и v , произведение которых близко к заданному y :

$$L(u, v) = \frac{1}{2}(uv - y)^2 + \frac{\lambda}{2}(u^2 + v^2), \quad (1)$$

где $\lambda \geq 0$ — коэффициент регуляризации.

Первое слагаемое штрафует за отклонение произведения uv от цели y . Второе (при $\lambda > 0$) — за слишком большие значения параметров.

Градиент функции (1):

$$\frac{\partial L}{\partial u} = (uv - y)v + \lambda u, \quad \frac{\partial L}{\partial v} = (uv - y)u + \lambda v. \quad (2)$$

2.2 Итерация градиентного спуска

Метод градиентного спуска обновляет параметры по правилу:

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t), \quad (3)$$

где $\theta = (u, v)^\top$, а $\eta > 0$ — размер шага.

Для нашей задачи это даёт:

$$u_{t+1} = u_t - \eta [(u_t v_t - y)v_t + \lambda u_t], \quad (4)$$

$$v_{t+1} = v_t - \eta [(u_t v_t - y)u_t + \lambda v_t]. \quad (5)$$

2.3 Критические точки

При $\lambda = 0$ функция (1) имеет следующие критические точки:

- **Глобальные минимумы:** все точки, где $uv = y$. При $y > 0$ это две ветви гиперболы, например, точки $(1, 1)$, $(-1, -1)$, $(2, 0.5)$ и т.д.
- **Седловая точка:** $(0, 0)$ — это точка, где градиент равен нулю, но это не минимум.

При $\lambda > 0$ минимумы изолированы: остаются только точки $(\pm\sqrt{y - \lambda}, \pm\sqrt{y - \lambda})$ (при $y > \lambda$).

2.4 Матричная факторизация

Обобщение на матричный случай:

$$L(U, V) = \frac{1}{2} \|U^\top V - Y\|_F^2 + \frac{\lambda}{2} (\|U\|_F^2 + \|V\|_F^2), \quad (6)$$

где $U, V \in \mathbb{R}^{d \times r}$ — матрицы параметров, $Y \in \mathbb{R}^{r \times r}$ — целевая матрица, $\|\cdot\|_F$ — норма Фробениуса.

Эта задача важна для рекомендательных систем, сжатия данных и обучения нейронных сетей (где подобная структура возникает в линейных слоях).

2.5 Критический шаг сходимости

Согласно [5], существует *критический шаг* η^* , выше которого глобальные минимумы становятся неустойчивыми. Для скалярного случая:

$$\eta^*(u, v) = \min \left\{ \frac{1}{|y|}, \frac{8}{\|\theta\|^2 + \sqrt{\|\theta\|^4 - 16y(uv - y)}} \right\}, \quad (7)$$

где $\|\theta\|^2 = u^2 + v^2$.

Первый член $1/|y|$ — порог, при превышении которого *все* минимумы нестабильны. Второй член зависит от начальной точки.

3 Обзор литературы

3.1 Основной источник

Работа Liang и Montúfar [5] содержит строгие доказательства следующих результатов:

Теорема 3.1 (О чувствительности). Для любой точки на границе области сходимости и любого $\varepsilon > 0$ существуют точки в ε -окрестности, которые сходятся к разным минимумам с произвольно различающимися нормами.

Теорема 3.2 (О топологической энтропии). При $\eta > \eta^*$ топологическая энтропия отображения градиентного спуска не меньше $\log 3$, что свидетельствует о хаотической динамике.

Также показано, что граница области сходимости имеет *фрактальную* размерность $D_B \approx 1.249$.

3.2 Фрактальные границы бассейнов притяжения

Понятие фрактальных границ бассейнов введено в работе McDonald и др. [1]. Авторы показали, что для многих нелинейных динамических систем граница между

бассейнами притяжения имеет нецелую размерность.

Определение 3.1. Бассейн притяжения аттрактора A — множество начальных условий, из которых траектория сходится к A .

Определение 3.2. Box-counting размерность множества F :

$$D_B(F) = \lim_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \quad (8)$$

где $N(\varepsilon)$ — минимальное число квадратов со стороной ε , покрывающих F .

Если D_B нецелое (например, 1.3), то множество называется *фрактальным*.

3.3 Edge of Stability

Cohen и др. [3] экспериментально обнаружили феномен «*edge of stability*»: при обучении нейронных сетей градиентным спуском максимальное собственное значение гессиана (мера «остроты» минимума) стремится к значению $2/\eta$ и колеблется вокруг него.

Это означает, что градиентный спуск «автоматически» выбирает режим работы на границе устойчивости, даже если изначально learning rate был выбран слишком большим.

3.4 Catapult Mechanism

Lewkowycz и др. [4] объяснили, почему большие learning rates иногда работают лучше ожидаемого. Механизм «катапульти»: если начальная точка находится в «остром» минимуме (с большой кривизной), большой шаг «выбрасывает» траекторию в более «плоскую» область. После нескольких осцилляций оптимизация продолжается в минимуме с меньшей кривизной, что часто даёт лучшую генерализацию.

3.5 Хаос в квадратичной регрессии

Chen и др. [2] провели детальный анализ градиентного спуска на квадратичной регрессии и выделили пять фаз:

1. **Монотонная**: loss убывает монотонно.
2. **Catapult**: немонотонные осцилляции, но сходимость к нулю.
3. **Периодическая**: траектория колеблется между несколькими точками.
4. **Хаотическая**: непредсказуемая динамика.
5. **Расходящаяся**: loss уходит в бесконечность.

4 Методы исследования

4.1 Построение карты бассейнов

Для визуализации бассейнов сходимости:

1. Создаём равномерную сетку точек (u_0, v_0) на плоскости.
2. Для каждой точки запускаем градиентный спуск.
3. Определяем, к какому аттрактору сошлась траектория:
 - расходимость (норма превысила 10^6);
 - минимум с $uv > 0$;
 - минимум с $uv < 0$;
 - седловая точка $(0, 0)$.
4. Раскрашиваем точку в соответствующий цвет.

4.2 Метод box-counting

Для оценки фрактальной размерности границы:

1. Выделяем границу: точки, имеющие соседа другого цвета.
2. Для последовательности размеров ε_k считаем $N(\varepsilon_k)$ — число квадратов, пересекающих границу.
3. Строим график $\log N$ от $\log(1/\varepsilon)$.
4. Размерность D_B — наклон линейной регрессии.

4.3 Сравнение с теоретическим η^*

Для проверки формулы (7):

1. Выбираем набор начальных точек (u_0, v_0) .
2. Для каждой вычисляем теоретический η^* по формуле.
3. Экспериментально находим максимальный η , при котором ещё есть сходимость.
4. Сравниваем теоретические и экспериментальные значения.

5 Результаты экспериментов

5.1 Бассейны сходимости: скалярный случай

На рис. 1 представлены карты бассейнов при $y = 1$, $\lambda = 0$ для разных η .

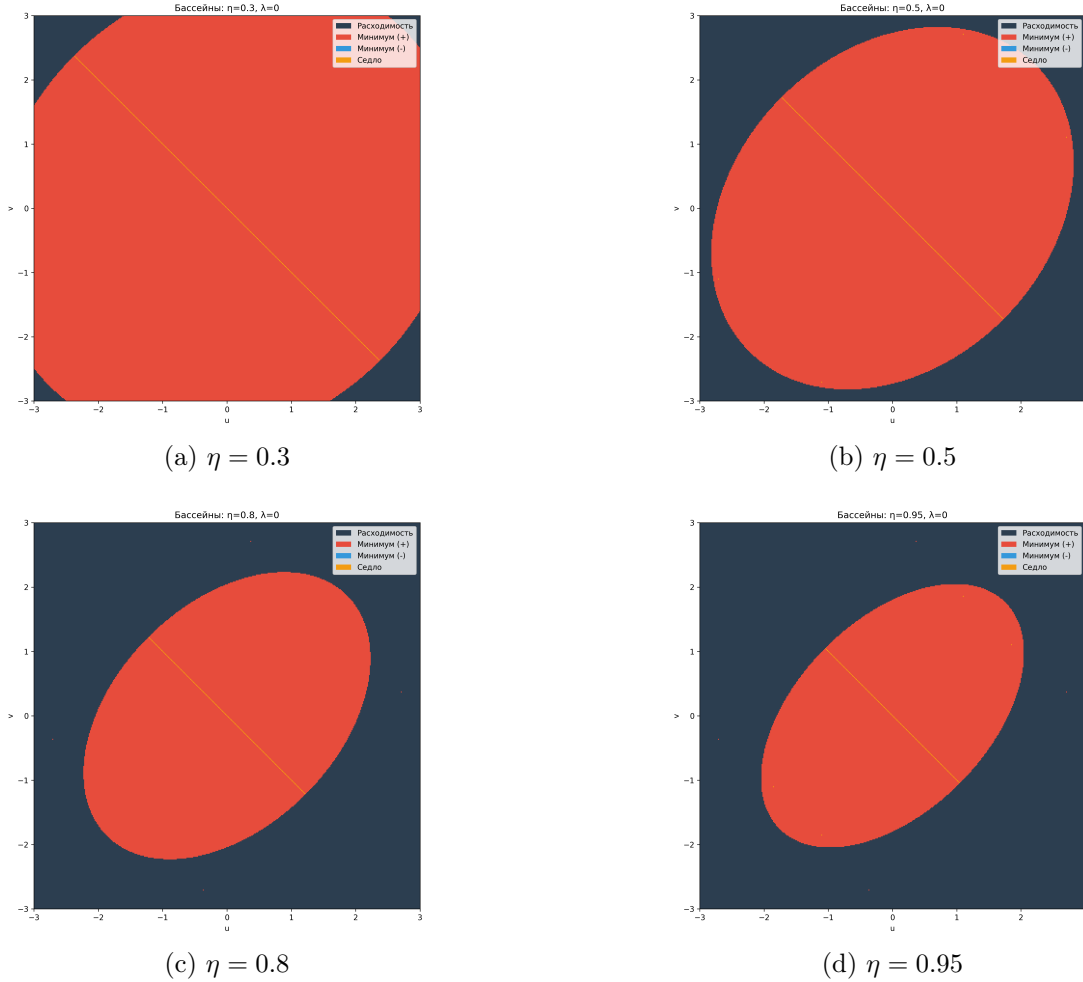


Рис. 1: Бассейны сходимости при разных η , $\lambda = 0$. Красный — минимум с $uv > 0$, синий — минимум с $uv < 0$, тёмный — расходимость.

Наблюдения:

- При $\eta = 0.3$ (рис. а): граница проходит по осям координат, структура простая.
- При $\eta = 0.5$ (рис. б): граница остаётся гладкой, но бассейны немного деформируются.
- При $\eta = 0.8$ (рис. с): появляются области расходимости (тёмные), граница усложняется.
- При $\eta = 0.95$ (рис. д): структура становится сложной, видны «щупальца» бассейнов.

5.2 Влияние регуляризации

На рис. 2 показаны бассейны при $\lambda = 0.1$.

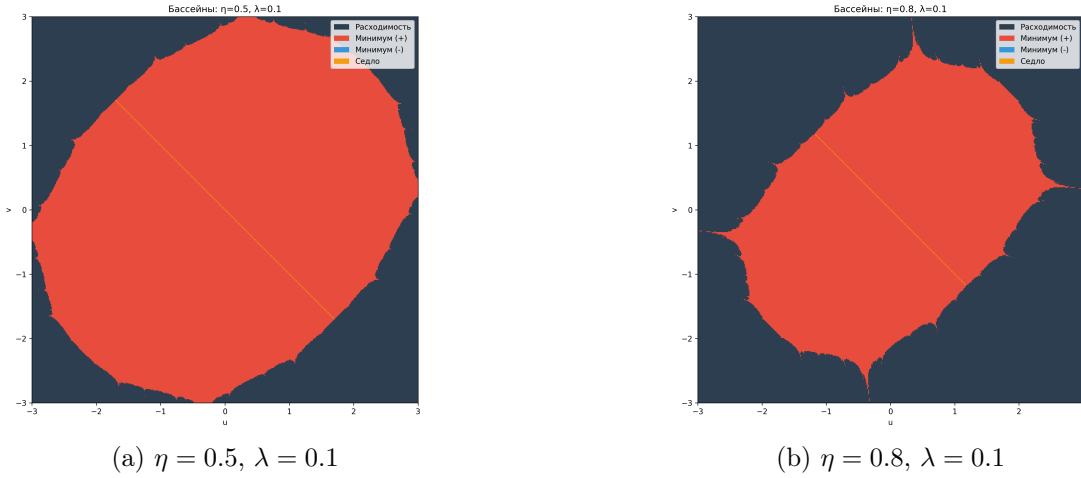


Рис. 2: Бассейны с регуляризацией.

Регуляризация добавляет новый тип поведения — сходимость к седловой точке (оранжевый цвет). Граница приобретает более сложную структуру.

5.3 Фрактальная размерность границы

Для случая $\eta = 0.8, \lambda = 0.1$ выделена граница и оценена её размерность (рис. 3).

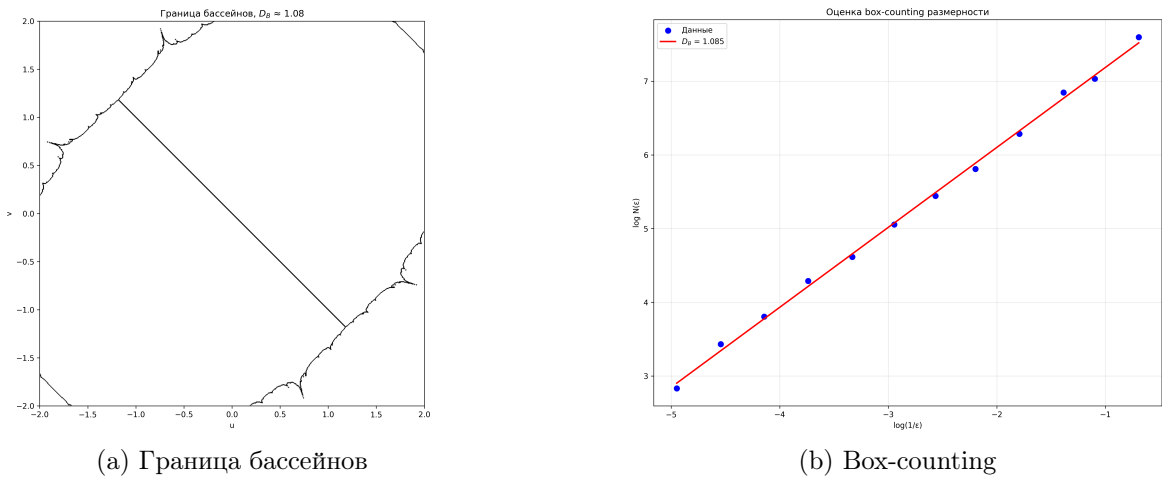


Рис. 3: Оценка фрактальной размерности.

Полученная оценка: $D_B \approx 1.08$.

Размерность близка к единице, что говорит о том, что граница близка к обычной кривой, но имеет некоторую «изрезанность». Теоретическое значение из [5] составляет $D_B \approx 1.249$; расхождение связано с конечным разрешением сетки и выбранными параметрами.

5.4 Чувствительность к начальным условиям

На рис. 4 показаны результаты теста чувствительности: из 300 точек в окрестности $(0.5, 0.5)$ с радиусом 10^{-4} .

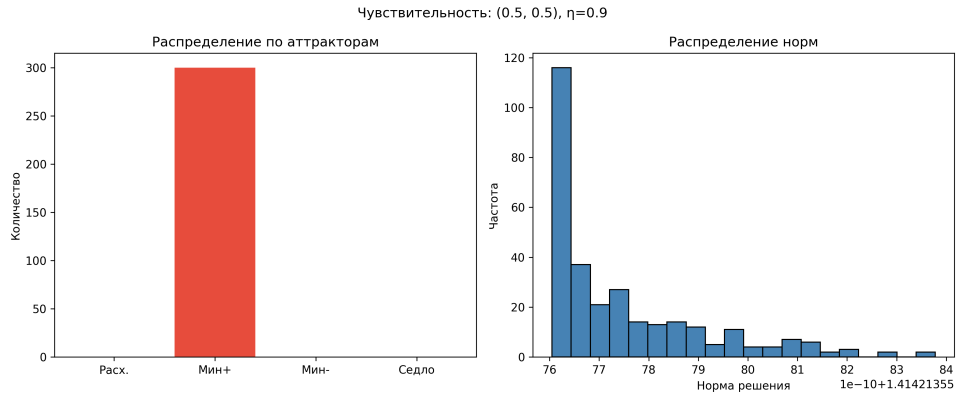


Рис. 4: Чувствительность к начальным условиям при $\eta = 0.9$.

Даже при микроскопических возмущениях траектории сходятся к разным аттракторам — это прямое проявление хаотической динамики.

5.5 Сравнение с теоретическим η^*

На рис. 5 сравниваются теоретические и экспериментальные значения критического шага для различных начальных точек.

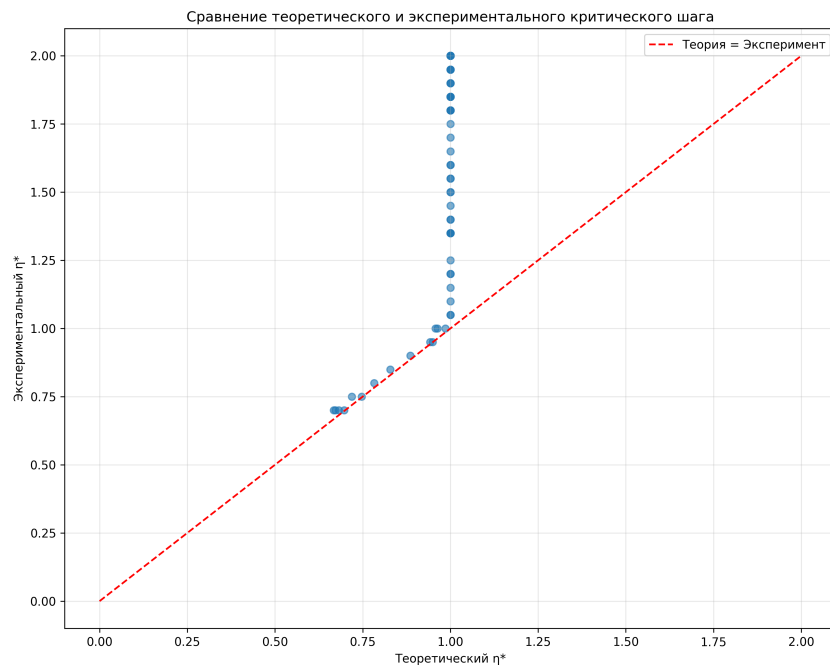
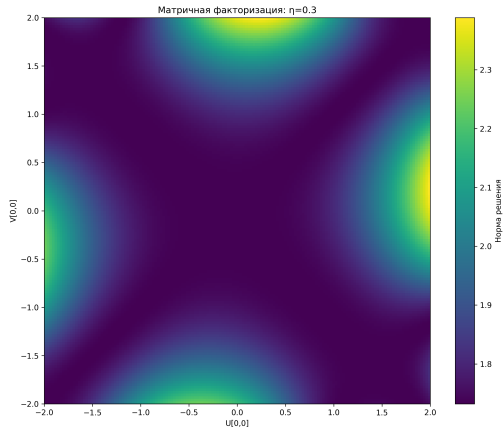


Рис. 5: Теоретический vs экспериментальный критический шаг.

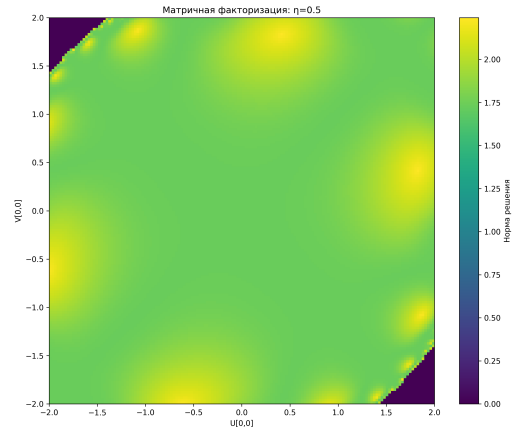
Точки группируются вблизи диагонали, что подтверждает согласие теории и эксперимента. Разброс связан с конечным шагом при экспериментальном определении границы сходимости.

5.6 Матричная факторизация

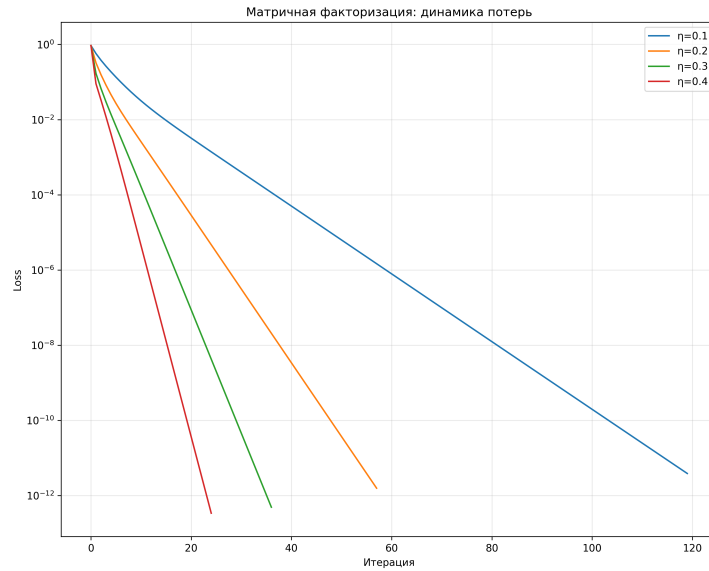
На рис. 6 показаны результаты для матричной факторизации с $Y = \text{diag}(1, 0.5)$.



(a) $\eta = 0.3$



(b) $\eta = 0.5$



(c) Кривые потерь

Рис. 6: Матричная факторизация: (a,b) — срезы бассейнов; (c) — динамика loss.

Наблюдения:

- Структура бассейнов качественно похожа на скалярный случай.
- При увеличении η появляются области расходимости.
- Фрактальная структура сохраняется при переходе к матрицам.

5.7 Траектории и динамика loss

На рис. 7 показаны траектории градиентного спуска.

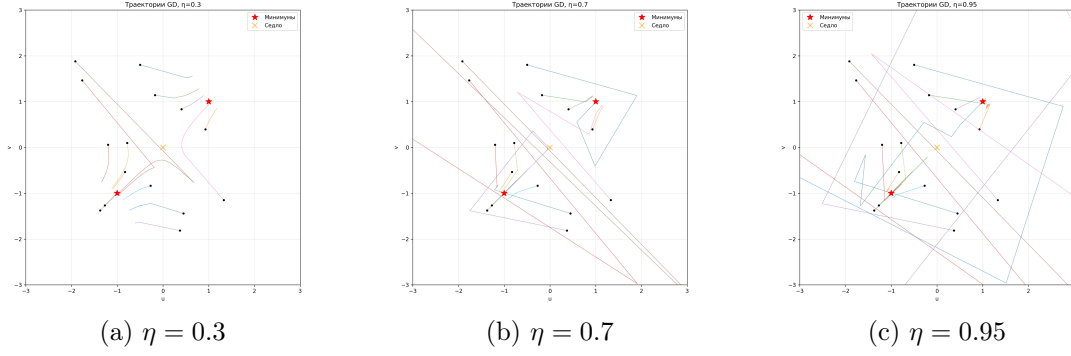


Рис. 7: Траектории при разных η .

При малом η траектории плавные. При большом — появляются резкие повороты и осцилляции.

На рис. 8 показана динамика функции потерь.

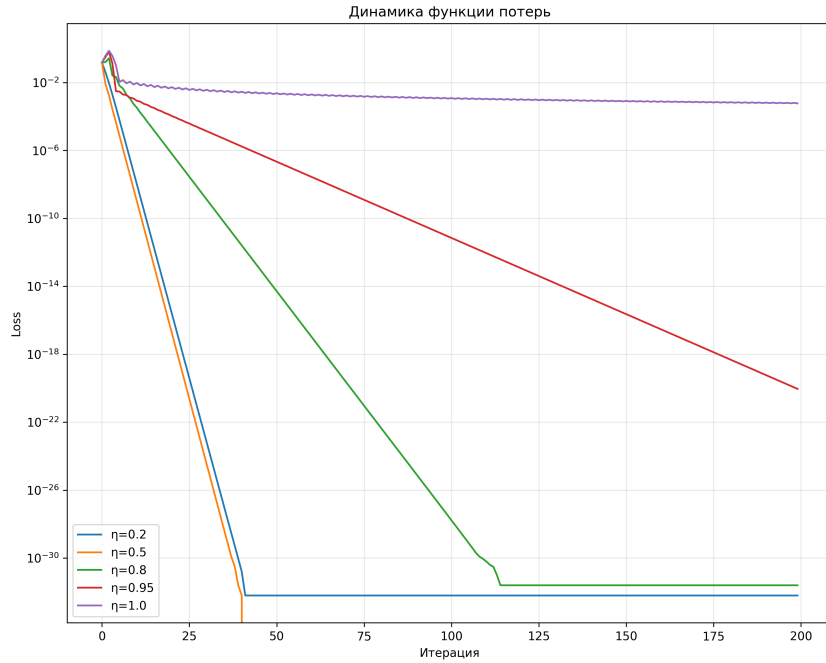


Рис. 8: Динамика loss при разных η .

При $\eta = 1.0$ происходит расходимость — loss растёт экспоненциально.

5.8 Связь с Edge of Stability и Catapult

Наши эксперименты демонстрируют явления, описанные в [3; 4]:

- **Edge of Stability:** при $\eta = 0.8$ – 0.95 траектории не расходятся сразу, а колеблются вблизи границы устойчивости. Loss сначала растёт, потом убывает.
- **Catapult:** при большом начальном η траектория «выбрасывается» из острого минимума в более плоский. Это видно на рис. 8: при $\eta = 0.95$ loss сначала возрастает (spike), затем начинает убывать.

Работа Liang и Montúfar [5] дополняет эту картину: при η выше критического возникает не просто неустойчивость, а полноценный *хаос* — непредсказуемая чувствительность к начальным условиям.

6 Выводы

1. При малых learning rate ($\eta < 0.5$ для $y = 1$) границы бассейнов гладкие, динамика предсказуема.
2. При приближении к критическому значению η^* граница усложняется и приобретает фрактальную структуру.
3. Экспериментальная оценка фрактальной размерности $D_B \approx 1.08$ подтверждает нецелую природу границы.
4. Теоретическая формула критического шага (7) хорошо согласуется с экспериментом.
5. Фрактальные структуры сохраняются при переходе от скалярной к матричной факторизации.
6. Наблюдаемые явления связаны с феноменами «edge of stability» и «catapult mechanism», описанными в современной литературе.

7 Заключение

В работе исследовано поведение градиентного спуска при больших learning rates на задачах скалярной и матричной факторизации. Реализован программный комплекс на Python для:

- визуализации бассейнов сходимости;
- оценки фрактальной размерности границы методом box-counting;
- анализа чувствительности к начальным условиям;
- сравнения с теоретическими предсказаниями.

Экспериментально подтверждено, что при приближении к критическому шагу η^* граница между бассейнами становится фрактальной, а динамика — хаотической. Это объясняет, почему результаты обучения нейронных сетей могут сильно зависеть от случайной инициализации.

Направления дальнейшей работы:

- Исследование матричной факторизации большей размерности.

- Анализ влияния стохастичности (SGD вместо GD).
- Применение результатов для выбора оптимального learning rate.

Список литературы

1. Fractal basin boundaries / S. W. McDonald [и др.] // Physica D. — 1985. — Т. 17, № 2. — С. 125—153.
2. From Stability to Chaos: Analyzing Gradient Descent Dynamics in Quadratic Regression / X. Chen [и др.] // TMLR. — 2024.
3. Gradient Descent on Neural Networks Typically Occurs at the Edge of Stability / J. Cohen [и др.] // ICLR. — 2021.
4. *Lewkowycz A., Gur-Ari G., Dyer E.* The large learning rate phase of deep learning: the catapult mechanism // arXiv preprint arXiv:2003.02218. — 2020.
5. *Liang S., Montúfar G.* Gradient Descent with Large Step Sizes: Chaos and Fractal Convergence Region // arXiv preprint arXiv:2509.25351. — 2025.

А Приложение. Исходный код

ИСХОДНЫЙ КОД: <https://github.com/thesaifutdinovanvar-jpg/gd-chaos>

А.1 Градиент и критический шаг

Листинг 1: Основные функции

```
1 def grad_scalar(u, v, y, lam=0.0):
2     """
3         r = u * v - y
4         du = r * v + lam * u
5         dv = r * u + lam * v
6         return du, dv
7
8 def critical_eta(u, v, y):
9     """
10         eta1 = 1.0 / abs(y) if y != 0 else 1e10
11         norm_sq = u*u + v*v
12         disc = norm_sq**2 - 16 * y * (u*v - y)
13         if disc < 0:
14             return eta1
15         eta2 = 8.0 / (norm_sq + np.sqrt(disc))
16         return min(eta1, eta2)
```

Liang & Montufar

А.2 Box-counting

Листинг 2: Оценка фрактальной размерности

```
1 def box_counting(boundary):
2     sizes, counts = [], []
3     box_size = 2
4     while box_size <= max_size:
5         count = 0
6         for i in range(0, h, box_size):
7             for j in range(0, w, box_size):
8                 if boundary[i:i+box_size, j:j+box_size].any():
9                     count += 1
10            sizes.append(box_size)
11            counts.append(count)
12            box_size = int(box_size * 1.5)
13
14     # log-log
15
16     log_x = np.log(1.0 / np.array(sizes))
17     log_y = np.log(np.array(counts))
```



```
17     D = slope(log_x, log_y)
18     return D
```