Лабораторная работа №1
Методы нулевого и первого порядка

https://github.com/GooddiLK/met-opt-lab1

Пластинин Алексей M3237
t.me/plstnn

Малков Александр M3237
t.me/AlexM_37

Цель работы:

Выяснить эффективность работы различных методов градиентного спуска в зависимости от выбора начальной точки, шага, значения гиперпараметров, числа обусловленности на различных функциях.

Используемые методы для планирования шага:

- Постоянный / кусочно-постоянный
- Экспоненциальное и полиномиальное затухание
- Условие Армихо
- Условие Вольфа с бинпоиском

Исследование

Рассматриваемые функции:

- $x^2 + y^2$
- $3x^2 - 4xy + 10y^2$
- $(x^2 + y - 11)^2 + (x + y^2 - 7)^2$ - Функция Химмельблау

Стоит также отметить, что в реализации градиент не нормировался.

Влияние числа обусловленности:

Для $k = 1$ лучше всего работает константный шаг, он сходится за несколько итераций, тем не менее все методы работают хорошо.
Для $k \approx 10$ константный шаг не работает, эффективнее всего себя показал полиномиальный шаг, Армихо и Вольфе работают, но требуют кратно больше рерсов для вычислений.

Для функции Химмельблау работают только Армихо и Вольфе. Их эффективность примерно равна.

Влияние гиперпараметров:

У константного шага значение гиперпараметра очень велико. Если он слишком большой, можно "пролететь" точку минимума, если слишком маленький, медленная сходимость. В зависимости от скорости сходимости функции требуется подбирать подходящее значение шага.

У Армихо α_0 влияет на скорость схождения, чем ниже гиперпараметр, тем быстрее работает метод. Стоит отметить что на F_2 слишком маленькое значение этого параметра снижает эффективность. Для F_3 же меньше - лучше.
У параметра c_1 обратная зависимость с α_0. Он также балансирует между скоростью сходимости и стабильностью.

У Вольфе α_0 и c_1 аналогичены Армихо. Чем выше c_2, тем быстрее сходится метод, но также повышаются риски.

Выбор начальной точки:

Выбор начальной точки заметно влияет на скорость работы Армихо, Вольфа, постоянного шага.
Меньшая значимость на экспоненциальном.
Почти не влияет на геометрический и полиномиальный методы вычисления шага.

Для функции Химмельблау разные начальные точки могут давать различные ответы.

Данные:

Влияние размера постоянного шага на скорость сходимости из (2,2).

Размер шага	Кол-во итераций F_1	Кол-во итераций F_2
0.5	2	-
0.4	6	-
0.3	9	-
0.2	13	-
0.1	25	-
0.05	48	22
0.025	90	42

Влияние параметров у Армихо из (10, 10):

α_0	q	ε	c_1	F_1	$F_{1'}$	F_2	$F_{2'}$	F_3	$F_{3'}$
0.01	0.5	0.001	0.05	282	281	129	128	25	22
0.1	0.5	0.001	0.05	33	32	18	14	67	16
0.5	0.5	0.001	0.05	2	2	56	15	109	19
1	0.5	0.001	0.05	3	2	71	15	128	19
2	0.5	0.001	0.05	4	2	86	15	147	19
3	0.5	0.001	0.05	26	12	54	9	110	12
4	0.5	0.001	0.05	5	2	101	15	166	19
5	0.5	0.001	0.05	23	7	134	19	134	15
10	0.5	0.001	0.05	30	7	153	19	149	15
-	-	-	-	-	-	-	-	-	-
1	0.5	0.001	0.1	3	2	71	15	128	19
1	0.5	0.001	0.2	3	2	71	15	71	10
1	0.5	0.001	0.5	26	12	79	16	95	12
1	0.5	0.001	0.8	77	25	134	25	303	34
-	-	-	-	-	-	-	-	-	-
1	0.05	0.001	0.05	125	62	55	27	132	45
1	0.1	0.001	0.05	65	32	127	45	38	12
1	0.2	0.001	0.05	31	15	26	9	44	11
1	0.3	0.001	0.05	19	9	193	64	59	13
1	0.4	0.001	0.05	13	6	54	14	66	12
1	0.6	0.001	0.05	13	6	109	18	109	12
1	0.7	0.001	0.05	19	9	201	25	434	33
1	0.8	0.001	0.05	31	15	433	36	492	27
-	-	-	-	-	-	-	-	-	-
3	0.1	0.001	0.05	19	9	73	26	121	30

Влияние параметров у Вольфе из (10, 10):

α_0	c_1	c_2	ε	F_1	$F_{1'}$	F_2	$F_{2'}$	F_3	$F_{3'}$
1	0.001	0.1	0.001	14	14	46	46	93	93
1	0.01	0.1	0.001	14	14	55	55	93	93
1	0.05	0.1	0.001	14	14	55	55	78	78
-	-	-	-	-	-	-	-	-	-
1	0.001	0.1	0.001	14	14	46	46	93	93
1	0.001	0.2	0.001	14	14	74	74	163	163
1	0.001	0.3	0.001	14	14	74	74	-	-
1	0.001	0.4	0.001	14	14	1221	1221	-	-

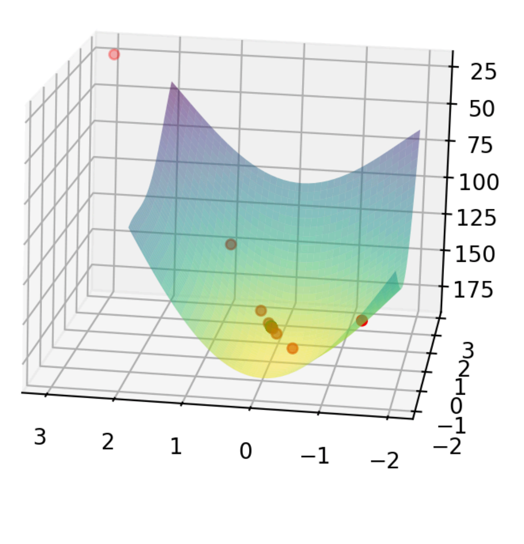

Figure 1: Wolfe F1

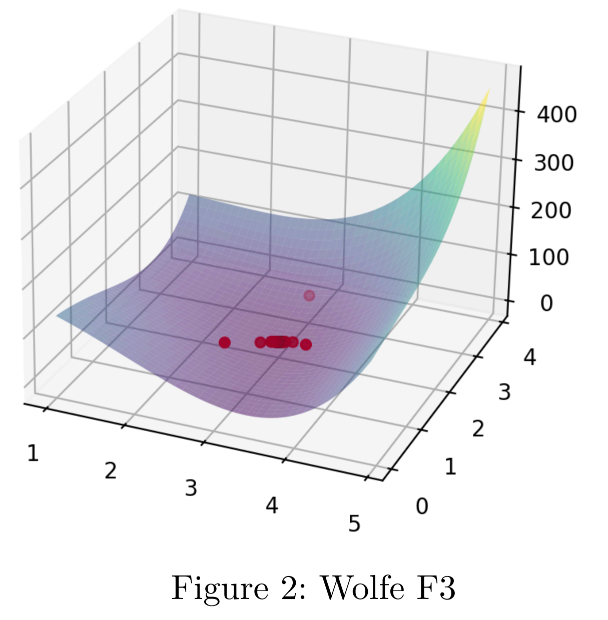

Figure 2: Wolfe F3

Выводы:

Анализ показал, что нет универсальной стратегии выбора шага. Для простых функций лучше всего подходит постоянный шаг, в то время как для сложных, 4 стандартных алгоритмы вычисления шага вообще не работают.

Армихо и Вольфе являются стабильными, они работают почти из любой точки и почти при любых гиперпараметрах, тем не менее эффективность методов на простых функциях иногда низкая.
Также они принимают несколько гиперпараметров, которые влияют на эффективность метода в совокупности, из-за этого очень труно подобрать оптимальные параметры, даже зная вид функции, на которой будет применяться метод.
В sciPy нет реализации стандартного градиентного спуска. Однако там есть подбор шага по условиям Вольфа. Он используется в методах BFGS, L-BFGS, CG, но это квазиньютоновские методы.