Домашнее задание № 2 по предмету "Экономико-математичекие модели оптимизации"

Григорчак К.А.
Ханов Э.Д.
Группа АМ-21-06

Москва, 2025

Содержание

1 Метод Дэвидона-Флетчера-Пауэлла (DFP)

1.1 Общее описание метода

Метод Дэвидона-Флетчера-Пауэлла (DFP) — это один из классических квазиньютоновских методов безусловной оптимизации. Он предназначен для нахождения локального минимума функции $f : R^n \to R$, не требуя вычисления второй производной (матрицы Гессе). Вместо этого метод строит приближение к обратной матрице Гессе и обновляет его на каждой итерации с использованием информации о градиенте.

1.2 Принцип работы метода DFP

На каждой итерации метода вычисляется направление антиградиента, скорректированное приближением обратной матрицы Гессе. Пусть x_k — текущая точка, $g_k = \nabla f(x_k)$ — градиент в ней, H_k — приближение к обратной матрице Гессе. Тогда направление спуска: $d_k = -H_k g_k$.

Далее проводится одномерная минимизация вдоль направления d_k для нахождения шага α_k:

$$x_{k+1} = x_k + \alpha_k d_k$$

После этого пересчитываются величины:

$$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k$$

1.3 Формула обновления в DFP

Обновление приближения обратной матрицы Гессе производится по следующей формуле:

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$$

Эта формула гарантирует симметричность и положительную определённость H_{k+1}, если H_k также симметрична и положительно определена, а $s_k^T y_k > 0$.

1.4 Вывод формулы DFP

Метод Дэвидона–Флетчера–Пауэлла (DFP) основан на обновлении приближения к матрице Гессе, используя информацию о градиентах на двух последовательных итерациях. Метод DFP непосредственно строит приближение к *обратной* матрице Гессе H, минуя вычисление самой матрицы Гессе B.

В методе DFP новая матрица H_{k+1} выбирается как решение задачи минимизации отклонения от предыдущей итерации H_k при условии, что новое приближение должно удовлетворять секущему условию:

$$H_{k+1}, :$$

$$H_{k+1} = \arg\min_H \|H - H_k\|_F^2 \qquad Hy_k = s_k,$$

где $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2}$ — это фробениусова норма матрицы, а

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

— приращения аргумента и градиента соответственно.

Для решения этой задачи минимизации с ограничением воспользуемся методом множителей Лагранжа. Введём матрицу множителей Лагранжа $\Lambda \in R^{n \times n}$, и построим функционал Лагранжа:

$$\mathcal{L}(H, \Lambda) = \|H - H_k\|_F^2 + Tr\left[\Lambda^\top (Hy_k - s_k)\right],$$

где $Tr[\cdot]$ обозначает след матрицы.

Найдём стационарную точку по переменной H:

$$\frac{\partial \mathcal{L}}{\partial H} = 2(H - H_k) + \Lambda y_k^\top = 0,$$

откуда:

$$H = H_k - \frac{1}{2}\Lambda y_k^\top.$$

Подставим в ограничение:

$$H y_k = s_k \quad \Rightarrow \quad \left(H_k - \frac{1}{2}\Lambda y_k^\top\right) y_k = s_k,$$

$$H_k y_k - \frac{1}{2}\Lambda(y_k^\top y_k) = s_k,$$

$$\Lambda = \frac{2}{y_k^\top y_k}(H_k y_k - s_k).$$

Подставляя обратно, получаем:

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^\top}{(s_k - H_k y_k)^\top y_k}.$$

Альтернативно, этот результат можно привести к более удобной форме с использованием симметричного выражения:

$$H_{k+1} = H_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k},$$

которая чаще всего используется на практике и сохраняет симметрию и положительную определённость H_k, если $s_k^\top y_k > 0$.

Таким образом, эта формула обеспечивает обновление приближения обратного гессиана в методе DFP с сохранением сходимости и устойчивости.

Эта формула удовлетворяет следующим важным свойствам:

- Обеспечивает выполнение секущего условия $H_{k+1} y_k = s_k$.
- Сохраняет симметричность и положительную определённость H_{k+1}, если H_k симметрична и положительно определена, и выполняется $s_k^T y_k > 0$.
- Добавляет информацию о кривизне вдоль s_k $\left(\frac{s_k s_k^T}{s_k^T y_k}\right)$ и корректирует поведение вдоль y_k $\left(\frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right)$.

Таким образом, метод DFP строит новое приближение к обратной матрице Гессе, минимизируя изменение относительно предыдущего приближения при выполнении секущего условия. Это делает метод эффективным инструментом для оптимизации гладких функций при отсутствии точной информации о второй производной.

1.5 Выбор начального приближения для метода DFP

Для метода DFP не существует строгого универсального условия выбора начального приближения x_0, однако выполнение следующих условий обеспечивает локальную суперлинейную сходимость:

- $\nabla f(x_0) \neq 0$ — градиент функции в начальной точке не должен быть нулевым, иначе метод не стартует;
- Начальное приближение x_0 должно находиться в окрестности локального минимума, при этом функция f должна быть дважды непрерывно дифференцируемой;
- Предполагается, что гессиан в точке минимума положительно определён — это обеспечивает корректное поведение аппроксимации H_k на каждом шаге;
- Начальное приближение для обратной гессиан-матрицы выбирается как $H_0 = I$, что соответствует единичной матрице. Также допустимы другие положительно определённые симметричные матрицы.

1.6 Условия сходимости метода DFP

Существуют теоретические условия, при которых метод DFP гарантированно сходится к стационарной точке:

- Функция f ограничена снизу, то есть существует $f^* \in R$, такое что $f(x) \geq f^*$ для всех x;
- Градиент $\nabla f(x)$ непрерывен и удовлетворяет условию Липшица: существует $L > 0$, такое что для всех x, y
$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|;$$
- Начальное приближение H_0 симметрично и положительно определено;
- Для всех k выполняется $s_k^T y_k > 0$ (выполняется автоматически при использовании условий Вольфа);
- Шаг α_k выбирается с использованием линейного поиска, удовлетворяющего условиям:
  - условие Армихо: $f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k$;
  - условие кривизны: $\nabla f(x_k + \alpha_k d_k)^T d_k \geq c_2 \nabla f(x_k)^T d_k$,

  где $0 < c_1 < c_2 < 1$.

При выполнении этих условий метод DFP сходится к стационарной точке x^*, причём

$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0.$$

Если дополнительно функция f имеет положительно определённую матрицу Гессе в точке минимума, то сходимость метода будет сверхлинейной.

1.7 Заключение

Метод DFP представляет собой эффективный способ минимизации гладких функций без необходимости вычисления второй производной. Он применяется в задачах, где матрица Гессе либо недоступна, либо её вычисление слишком трудоемко. Несмотря на появление более устойчивых модификаций (например, метода BFGS), DFP сохраняет своё значение как важный шаг в развитии квазиньютоновских методов оптимизации.

2 Метод Бройдена — Флетчера — Гольдфарба — Шанно (BFGS)

2.1 Общее описание метода

Метод BFGS (Broyden–Fletcher–Goldfarb–Shanno) — один из наиболее популярных и широко используемых квазиньютоновских методов безусловной оптимизации. Как и метод DFP, он предназначен для поиска локального минимума гладкой функции $f : R^n \to R$ без необходимости вычисления полной матрицы Гессе. Вместо этого строится и обновляется приближение к обратной матрице Гессе на каждой итерации с использованием информации о градиенте. Метод BFGS считается более устойчивым и надёжным, чем DFP, и на практике часто показывает лучшую сходимость.

2.2 Принцип работы метода BFGS

На k-ой итерации метода вычисляется направление спуска с использованием приближённой обратной матрицы Гессе H_k:

$$d_k = -H_k g_k, \quad g_k = \nabla f(x_k).$$

После этого осуществляется одномерная минимизация вдоль направления d_k для выбора шага α_k и обновления точки:

$$x_{k+1} = x_k + \alpha_k d_k.$$

Далее вычисляются векторы:

$$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k.$$

2.3 Формула обновления в BFGS

Обновление приближения к обратной матрице Гессе в методе BFGS производится по формуле:

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}.$$

Эта формула обеспечивает симметричность и положительную определённость H_{k+1}, при условии, что H_k также симметрична и положительно определена, а $y_k^T s_k > 0$.

2.4 Вывод формулы BFGS

Как и в методе DFP, обновление приближённой матрицы в методе BFGS выводится из условия секущей. Однако в отличие от метода DFP, где обновляется приближение к *обратной* матрице Гессе H, в BFGS сначала строится новое приближение к самой матрице Гессе B, а затем получается $H = B^{-1}$.

Метод BFGS можно вывести из задачи минимизации фробениусовой нормы отклонения новой оценки гессиана B_{k+1} от текущей B_k:

$$B_{k+1} = \arg\min_B \|B - B_k\|_F^2, \qquad B s_k = y_k,$$

где:

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Решение этой задачи, аналогично методу DFP, приводит к симметричному и положительно определённому обновлению, сохраняющему секущие условия. Итоговая формула обновления B_k:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k} + \frac{y_k y_k^\top}{y_k^\top s_k}.$$

Это выражение минимизирует изменение гессиана при сохранении корректности направления градиента.

Обратную матрицу $H_{k+1} = B_{k+1}^{-1}$ можно получить, применив обобщённую формулу обратного матричного обновления (например, обратную формулу Вудбери или Шермана-Моррисона):

$$H_{k+1} = \left(I - \frac{s_k y_k^\top}{y_k^\top s_k}\right) H_k \left(I - \frac{y_k s_k^\top}{y_k^\top s_k}\right) + \frac{s_k s_k^\top}{y_k^\top s_k}.$$

Эта форма используется в практических реализациях BFGS, поскольку работает с H_k напрямую (а не с B_k) и лучше подходит для численной реализации.

Интуитивная трактовка данной формулы:

- Первая часть $\left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right)$ представляет собой коррекцию предыдущего приближения H_k, обеспечивающую симметрию и удовлетворение секущему условию.
- Вторая часть $\frac{s_k s_k^T}{y_k^T s_k}$ добавляет информацию о кривизне функции вдоль направления s_k.
- Вся формула обеспечивает симметричность и (при $y_k^T s_k > 0$) положительную определённость матрицы H_{k+1}.

Таким образом, в методе BFGS обновление матрицы основывается на решении задачи минимизации изменения гессиана B, в то время как в DFP — изменения его обратной матрицы H. Это различие и обеспечивает численную устойчивость метода BFGS.

2.5 Выбор начального приближения для метода BFGS

Метод BFGS, как и DFP, требует подходящего начального приближения x_0, однако имеет лучшую устойчивость при плохом выборе x_0. Тем не менее, для гарантии сходимости желательно выполнение следующих условий:

- $\nabla f(x_0) \neq 0$ — начальная точка не должна быть стационарной;
- Точка x_0 должна находиться в области, где функция f является дважды непрерывно дифференцируемой и её гессиан положительно определён;
- Начальное приближение x_0 желательно выбирать ближе к предполагаемому минимуму, особенно в задачах с несколькими экстремумами;
- Матрица H_0 (начальная аппроксимация обратного гессиана) по умолчанию берётся равной I, однако может быть задана любая положительно определённая симметричная матрица.

2.6 Условия сходимости метода BFGS

Для сходимости метода BFGS требуется выполнение следующих условий:

- Функция f ограничена снизу: существует $f^* \in R$, такое что $f(x) \geq f^*$ для всех x;
- Градиент $\nabla f(x)$ непрерывен и удовлетворяет условию Липшица: существует $L > 0$, такое что для всех x, y
$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|;$$
- Начальное приближение H_0 симметрично и положительно определено;
- Выполняется условие $y_k^T s_k > 0$ для всех k (что обеспечивается при использовании условий Вольфа при линейном поиске);
- Шаг α_k выбирается таким образом, чтобы удовлетворять условиям:
  - условие Армихо: $f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k$;
  - условие кривизны: $\nabla f(x_k + \alpha_k d_k)^T d_k \geq c_2 \nabla f(x_k)^T d_k$,

  где $0 < c_1 < c_2 < 1$.

При выполнении этих условий метод BFGS сходится к стационарной точке x^*, и
$$\lim_{k \to \infty} \|\nabla f(x_k)\| = 0.$$
Если в точке минимума функция f имеет положительно определённую матрицу Гессе, то сходимость метода будет сверхлинейной.

2.7 Заключение

Метод BFGS — это мощный и надёжный инструмент для минимизации гладких функций без необходимости вычисления второй производной. Он превосходит метод DFP по численной устойчивости и чаще используется на практике. Благодаря своим свойствам, метод BFGS стал стандартом дефакто среди квазиньютоновских методов в задачах оптимизации.

3 Сравнительный анализ методов DFP и BFGS

В таблице 1 представлены достигнутые минимумы для трёх классических тестовых функций: Розенброка, Химмельблау и МакКормика. Для оценки точности вычислений указаны численные значения, полученные методами DFP и BFGS.

Известные точные значения глобальных минимумов этих функций:

- **Функция Розенброка:** $f(1, 1) = 0$
- **Функция Химмельблау:** $f(3, 2) = 0$
- **Функция МакКормика:** $f(-0.5472, -1.5472) \approx -1.9132$

Оба метода демонстрируют высокую точность и сходные результаты, приближаясь к точным значениям с разной скоростью сходимости.

Таблица 1: Сравнение достигнутых минимумов функции

Функция	**Метод**	**Минимум** x^*	$f(x^*)$
Розенброка	DFP	$[0.99999896,\ 0.99999791]$	$1.09 \cdot 10^{-12}$
	BFGS	$[1.00000055,\ 1.00000112]$	$3.55 \cdot 10^{-13}$
Химмельблау	DFP	$[2.99999997,\ 1.99999987]$	$4.02 \cdot 10^{-13}$
	BFGS	$[3.00000000,\ 2.00000000]$	$3.18 \cdot 10^{-16}$
МакКормика	DFP	$[-0.54719727,\ -1.54719743]$	-1.9132
	BFGS	$[-0.54719771,\ -1.54719775]$	-1.9132

Также в таблице ниже приведены количества итераций, необходимых для достижения заданной точности.

Таблица 2: Число итераций до сходимости при различных точностях

Функция	**Метод**	$\varepsilon = 10^{-1}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}
Розенброка	DFP	14	15	19	20	20
	BFGS	8	10	11	12	12
Химмельблау	DFP	10	11	11	12	12
	BFGS	11	12	14	14	14
МакКормика	DFP	3	5	5	6	7
	BFGS	3	5	6	7	8

Выводы

Оба метода сходятся к очень близким точкам минимума и демонстрируют высокую точность. Однако можно отметить следующие особенности:

- Метод BFGS работает быстрее на функции Розенброка: для достижения заданной точности требуется значительно меньше итераций по сравнению с DFP.
- На функции Химмельблау методы показывают схожие результаты, однако DFP сходится немного быстрее при высоких точностях.
- Для функции МакКормика оба метода демонстрируют практически одинаковую скорость сходимости, но BFGS требует чуть больше итераций на высокой точности.

Таким образом, метод BFGS в среднем обеспечивает более быструю сходимость, особенно на плохо обусловленных функциях (например, Розенброка), в то время как DFP может показывать лучшую стабильность на более простых задачах.

Графическое представление траекторий

Ниже представлены графики траекторий движения для обеих реализаций:

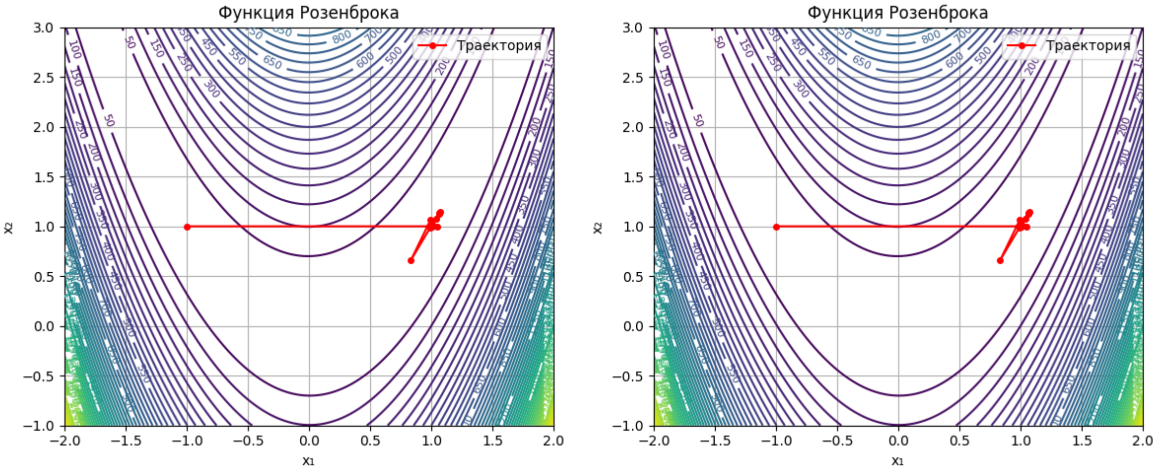

Рис. 1: Траектории на функции Розенброка: DFP (слева), BFGS (справа)

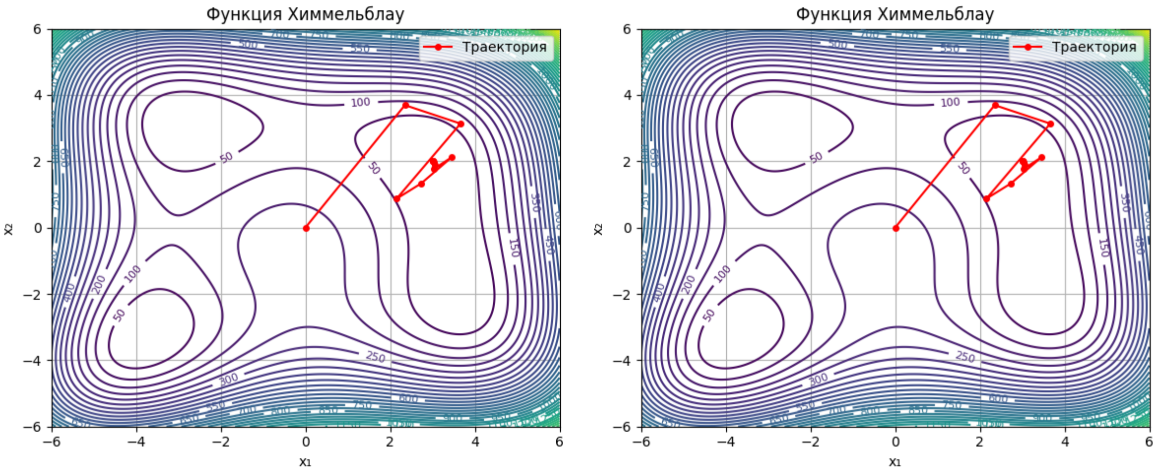

Рис. 2: Траектории на функции Химмельблау: DFP (слева), BFGS (справа)

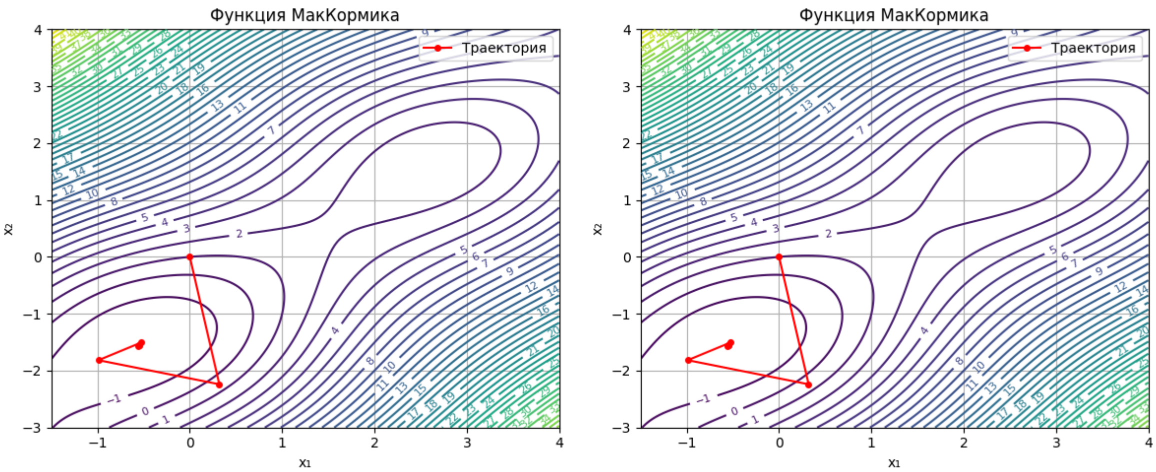

Рис. 3: Траектории на функции МакКормика: DFP (слева), BFGS (справа)