1

Математические основы многомерной оптимизации

Теорема. *Пусть функция $f(x)$ ограничена снизу, дифференцируема на $\mathbb{R}^n$ и ее градиент удовлетворяет условию Липшица с постоянной $L > 0$:*

$$\|\nabla f(x')-\nabla f(x'')\| \le L\|x'-x''\|, \quad \forall x', x'' \in \mathbb{R}^n$$

Тогда для любой начальной точки x^0 в итерационной процедуре:

$$x^{k+1} = x^k + \alpha y^k,$$
$$y^k = -\nabla f(x^k),\ k = 0, 1, 2, \ldots$$

можно выбрать такое число $\alpha > 0$, постоянное для всех k, что

$$\lim_{k\to\infty}\|\nabla f(x^k)\| = 0,$$

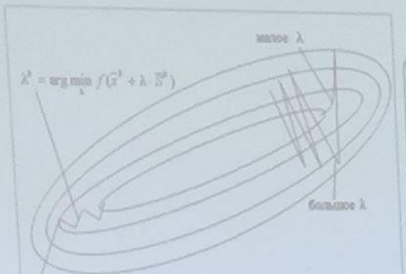
причем последовательность $\{x_k\}$ будет релаксационной.

2

Метод наискорейшего спуска

Steepest Descent method

Идея подхода: *при построении релаксационной последовательности используем антиградиент как направление спуска*



Подготовительный этап.

Зададим начальное приближение x^0 и точность расчета ε:

$$|\nabla f(x^i)| < \varepsilon.$$

Шаг 1:

Определим направление спуска и длину шага на текущей итерации:

$$y^i = -\nabla f(x^{i-1})$$
$$\alpha^i = \min_{\alpha>0} f\left(x^{i-1} + \alpha y^i\right)$$

Вычислим приближение к точке минимума x^i:

$$x^i = x^{i-1} + \alpha^i y^i$$

Шаг 2:

Проверим выполнение условия останова:

Если $|\nabla f(x^i)| \ge \varepsilon$, то $i = i + 1$ и возвращаемся к шагу 1.

Иначе – завершаем поиск со значением $x^* = x^i$.

6 Метод сопряженных градиентов

Conjugate Gradient (Fletcher-Reeves) method

Идея подхода: *при построении релаксационной последовательности используем значение антиградиента на текущем шаге и направление спуска на предыдущем шаге*

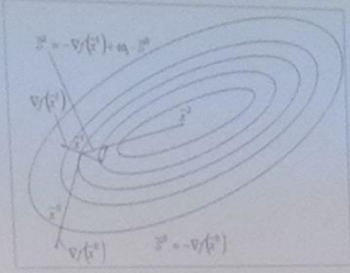
Подготовительный этап:

Зададим начальное приближение x^0 *и точность расчета* ε:

$$|\nabla f(x^i)| < \varepsilon$$

Определим первое приближение к точке минимума с шагом α^0 *и направлением спуска* s^0:

$$s^0 = -\nabla f(x^0), \quad \alpha^0 = \min_{\alpha>0} f\left(x^0 + \alpha s^0\right), \quad x^1 = x^0 + \alpha^0 s^0$$



Шаг 1:

Определим весовой коэффициент w^i, *направление спуска* s^i *и длину шага* α^i *на текущей итерации:*

$$w^i = \frac{\left|\nabla f\left(x^i\right)\right|^2}{\left|\nabla f\left(x^{i-1}\right)\right|^2}, \quad s^i = -\nabla f(x^i) + w^i s^{i-1}, \quad \alpha^i = \min_{\alpha>0} f\left(x^i + \alpha s^i\right)$$

Вычислим приближение к точке минимума x^{i+1}:

$$x^{i+1} = x^i + \alpha^i s^i$$

Шаг 2:

Проверим выполнение условия останова:

Если $|\nabla f(x^{i+1})| \geq \varepsilon$, *то* $i = i + 1$ *и возвращаемся к шагу 1*

Иначе – завершаем поиск со значением $x^* = x^{i+1}$.

При $i = n + 1$ *итераций необходимо принять* $x^0 = x^{n+1}$ *и вернуться к шагу 1*

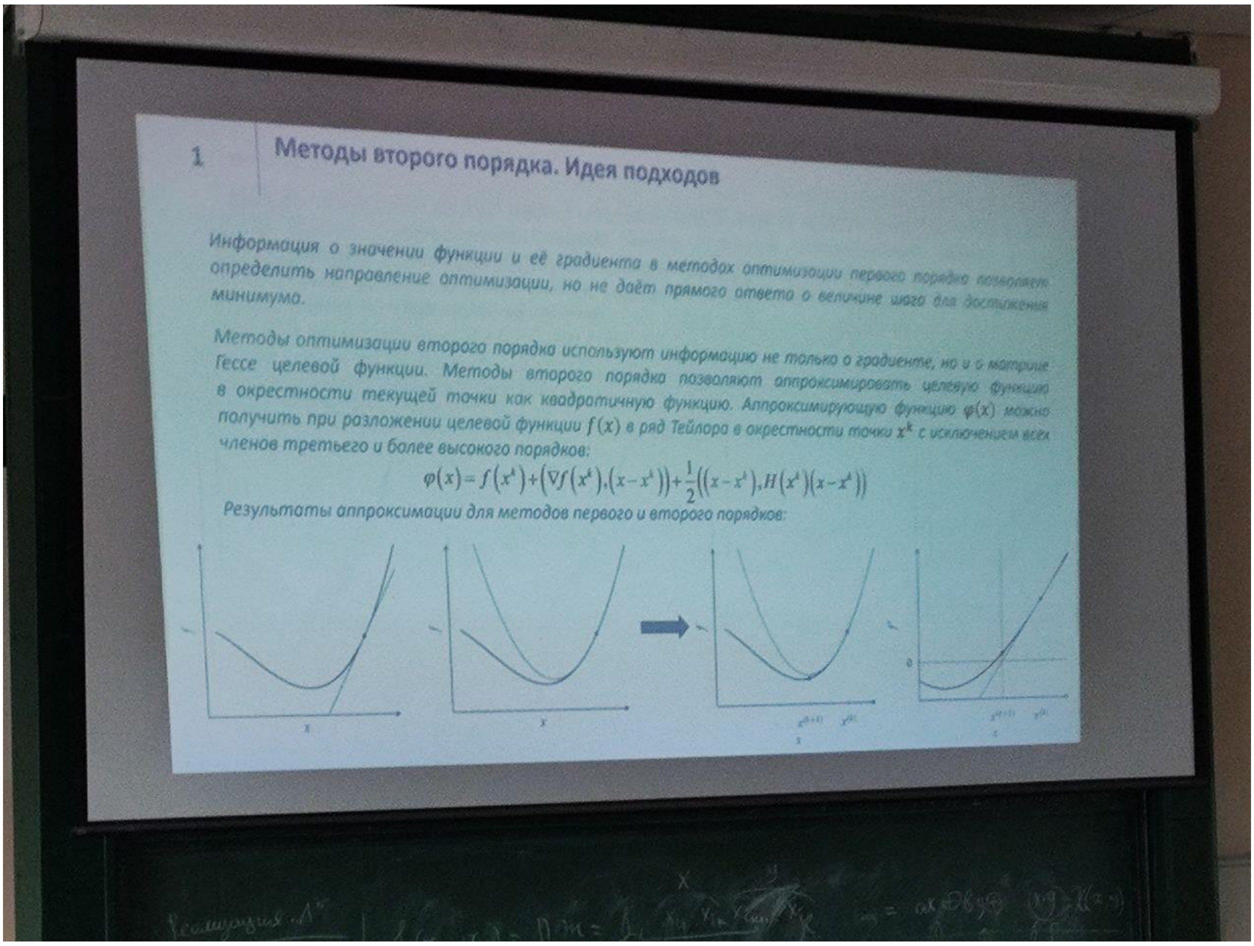
9 Метод Полака-Райбера

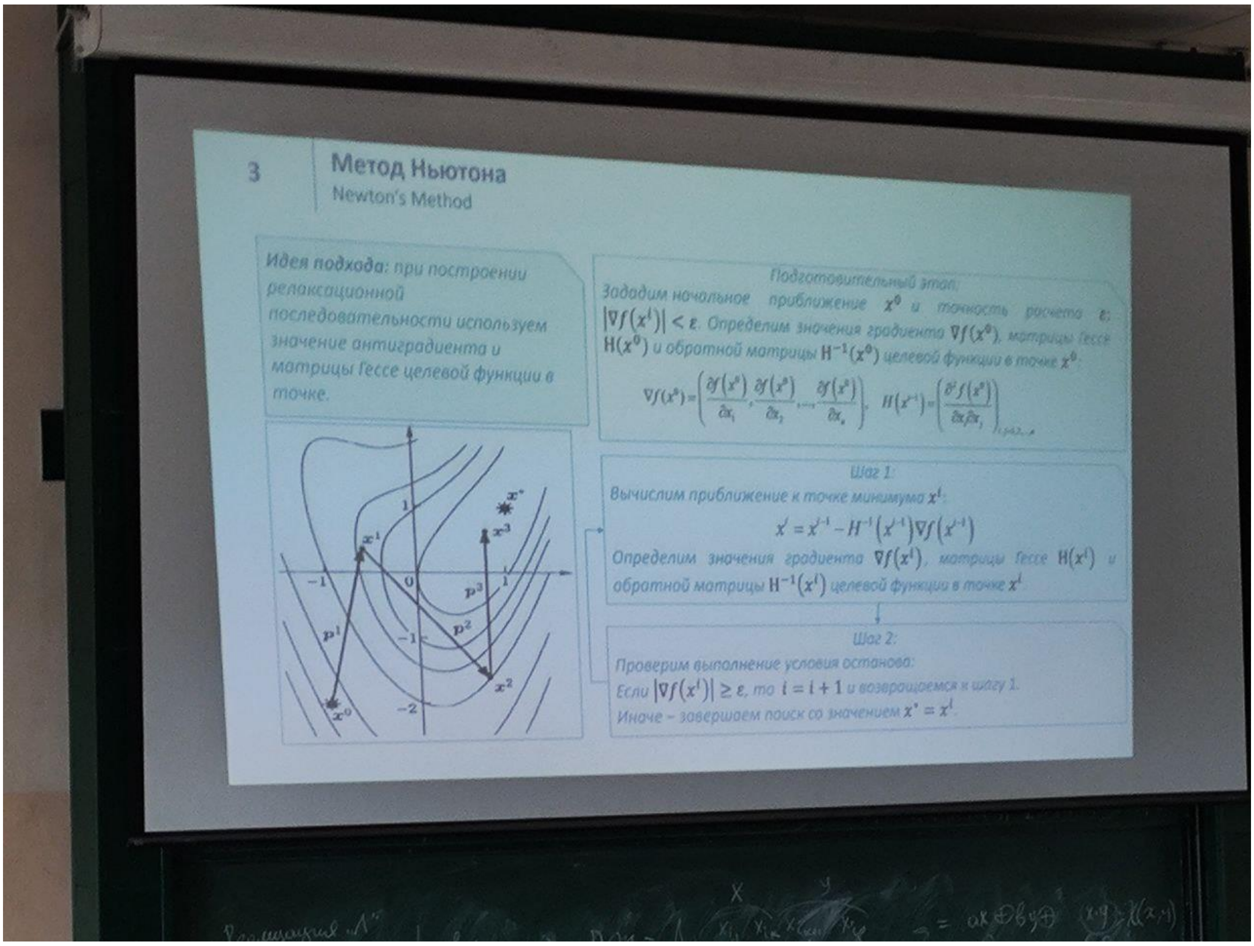
Polak-Ribière method

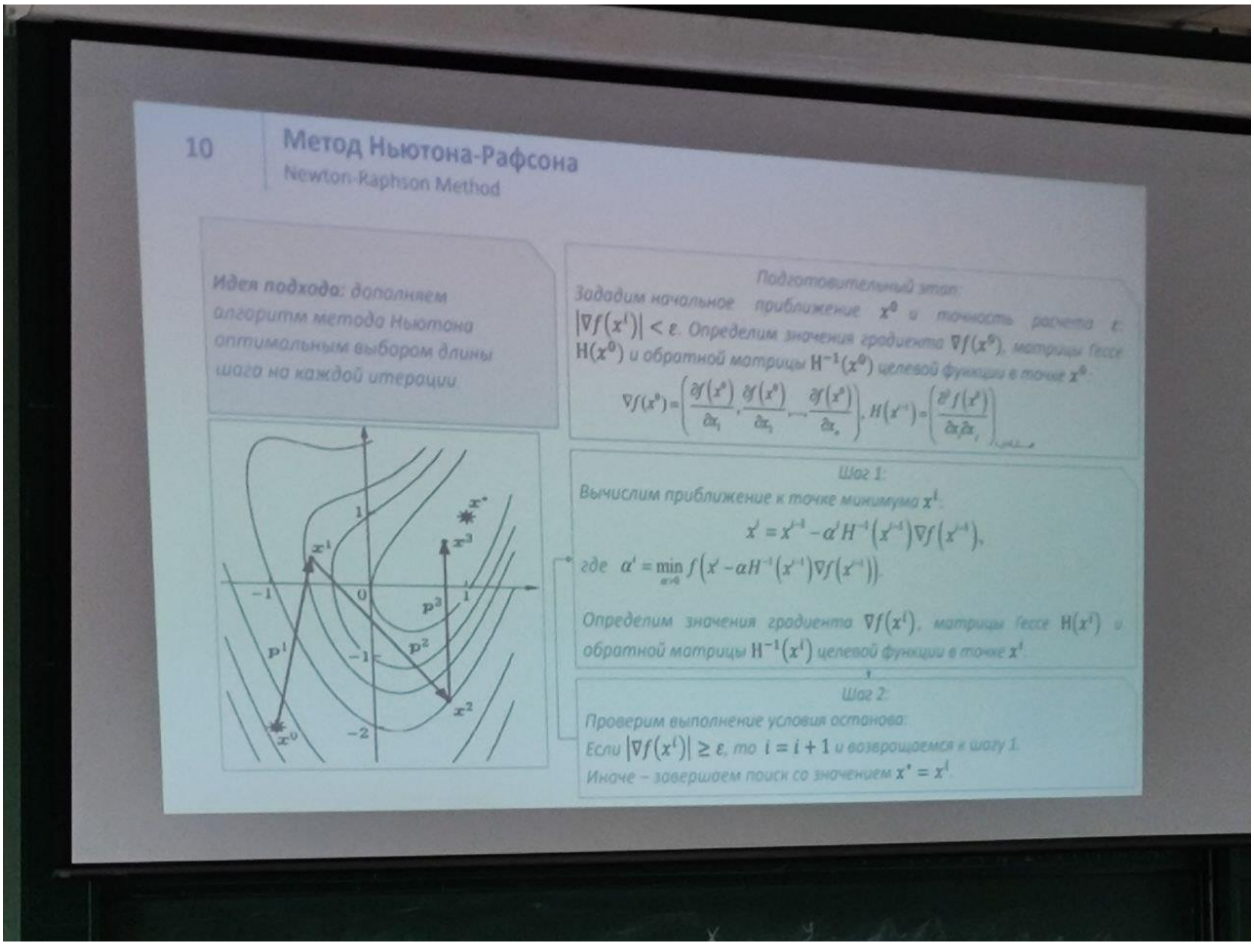
Идея подхода: *вариация метода сопряженных градиентов, предполагающая альтернативный способ вычисления весовых коэффициентов:*

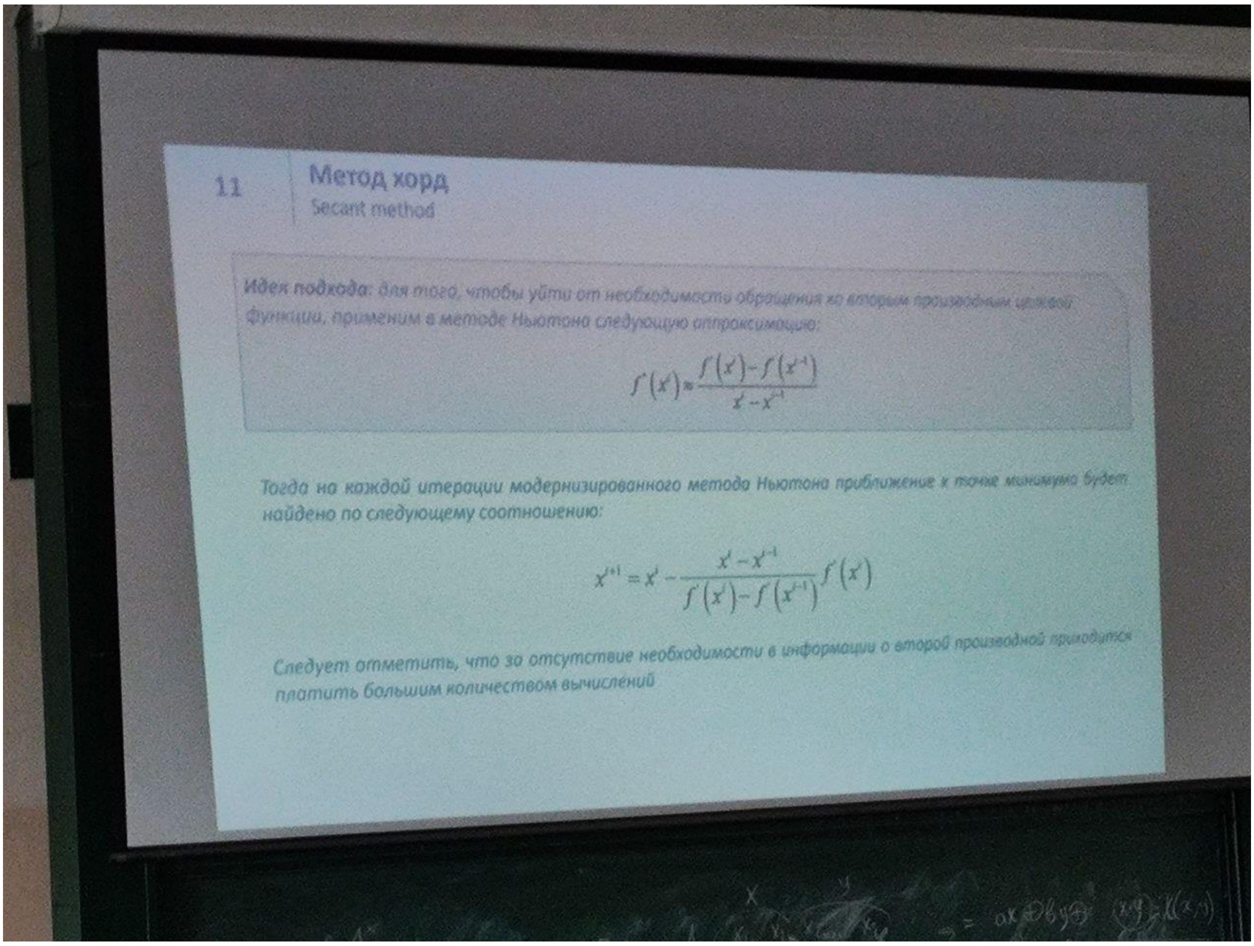
$$w^i = \frac{\nabla^T f\left(x^i\right)\left[\nabla f\left(x^i\right) - \nabla f\left(x^{i-1}\right)\right]}{\nabla^T f\left(x^{i-1}\right) s^{i-1}}$$

Подход Полака-Райбера позволяет избежать «перезапуска» алгоритма при достижении счетчиком итераций i *значения* $n + 1$. *Это свойство обеспечивает большую эффективность метода Полака-Райбера для произвольных целевых функций.*


1
Методы второго порядка. Идея подходов
Информация о значении функции и её градиента в методах оптимизации первого порядка позволяет определить направление оптимизации, но не даёт прямого ответа о величине шага для достижения минимума.
Методы оптимизации второго порядка используют информацию не только о градиенте, но и о матрице Гессе целевой функции. Методы второго порядка позволяют аппроксимировать целевую функцию в окрестности текущей точки как квадратичную функцию. Аппроксимирующую функцию $\varphi(x)$ можно получить при разложении целевой функции $f(x)$ в ряд Тейлора в окрестности точки x^k с исключением всех членов третьего и более высокого порядков:
$\varphi(x) = f(x^k) + (\nabla f(x^k), (x - x^k)) + \frac{1}{2}((x - x^k), H(x^k)(x - x^k))$
Результаты аппроксимации для методов первого и второго порядков:


3
Метод Ньютона
Newton's Method
Идея подхода: при построении релаксационной последовательности используем значение антиградиента и матрицы Гессе целевой функции в точке.
Подготовительный этап:
Зададим начальное приближение x^0 и точность расчета ε: $|\nabla f(x^i)| < \varepsilon$. Определим значения градиента $\nabla f(x^0)$, матрицы Гессе $H(x^0)$ и обратной матрицы $H^{-1}(x^0)$ целевой функции в точке x^0:
$\nabla f(x^0) = \left(\frac{\partial f(x^0)}{\partial x_1}, \frac{\partial f(x^0)}{\partial x_2}, \ldots, \frac{\partial f(x^0)}{\partial x_n}\right)$, $H(x^0) = \left(\frac{\partial^2 f(x^0)}{\partial x_i \partial x_j}\right)$
Шаг 1:
Вычислим приближение к точке минимума x^i:
$x^i = x^{i-1} - H^{-1}(x^{i-1}) \nabla f(x^{i-1})$
Определим значения градиента $\nabla f(x^i)$, матрицы Гессе $H(x^i)$ и обратной матрицы $H^{-1}(x^i)$ целевой функции в точке x^i.
Шаг 2:
Проверим выполнение условия останова:
Если $|\nabla f(x^i)| \geq \varepsilon$, то $i = i + 1$ и возвращаемся к шагу 1.
Иначе – завершаем поиск со значением $x^* = x^i$.
x^0
x^1
x^2
x^3
x^*
p^1
p^2
p^3
−1
0
1
−2


10
Метод Ньютона-Рафсона
Newton-Raphson Method
Идея подхода: дополняем алгоритм метода Ньютона оптимальным выбором длины шага на каждой итерации
Подготовительный этап:
Зададим начальное приближение x^0 и точность расчета ε: $|\nabla f(x^i)| < \varepsilon$. Определим значения градиента $\nabla f(x^0)$, матрицы Гессе $H(x^0)$ и обратной матрицы $H^{-1}(x^0)$ целевой функции в точке x^0.
$\nabla f(x^0) = \left(\frac{\partial f(x^0)}{\partial x_1}, \frac{\partial f(x^0)}{\partial x_2}, \dots, \frac{\partial f(x^0)}{\partial x_n}\right), H(x^{i-1}) = \left(\frac{\partial^2 f(x^0)}{\partial x_i \partial x_j}\right)$
Шаг 1:
Вычислим приближение к точке минимума x^i:
$x^i = x^{i-1} - \alpha^i H^{-1}(x^{i-1}) \nabla f(x^{i-1}),$
где $\alpha^i = \min_{\alpha>0} f\left(x^i - \alpha H^{-1}(x^{i-1}) \nabla f(x^{i-1})\right).$
Определим значения градиента $\nabla f(x^i)$, матрицы Гессе $H(x^i)$ и обратной матрицы $H^{-1}(x^i)$ целевой функции в точке x^i.
Шаг 2:
Проверим выполнение условия останова:
Если $|\nabla f(x^i)| \geq \varepsilon$, то $i = i + 1$ и возвращаемся к шагу 1.
Иначе – завершаем поиск со значением $x^* = x^i$.
x^*
x^1
x^3
x^2
x^0
p^1
p^2
1
−1
0
−2


11
Метод хорд
Secant method
Идея подхода: для того, чтобы уйти от необходимости обращения ко вторым производным целевой функции, применим в методе Ньютона следующую аппроксимацию:
$f'(x^i) \approx \frac{f(x^i) - f(x^{i-1})}{x^i - x^{i-1}}$
Тогда на каждой итерации модернизированного метода Ньютона приближение к точке минимума будет найдено по следующему соотношению:
$x^{i+1} = x^i - \frac{x^i - x^{i-1}}{f'(x^i) - f'(x^{i-1})} f'(x^i)$
Следует отметить, что за отсутствие необходимости в информации о второй производной приходится платить большим количеством вычислений