метод	требования	Т. о сходимости	сходимость
градиентный метод (здесь линейная аппроксимация) метод 1 порядка безусловная min	безусловная минимизация $f(x)$ - дифф на $\mathbb{R}^n$ $h_k = -\nabla f(x)$ (направление - антиградиент) α_k - фикс / Армихо / Гольдштейн $f_k(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle$	• если $f(x)$ - дифф на $\mathbb{R}^n$ $\nabla f(x)$ удовл. усл. Липш. $f(x)$ - ограничена снизу $0<\alpha<\frac{2}{L}$ для $\alpha_k=\alpha$ тогда $\lim \nabla f(x_k)=0$ и $f(x_{k+1}) \le f(x_k)$ • если $f(x)$ - дифф на $\mathbb{R}^n$ $\nabla f(x)$ - липшицев с const=L $f(x)$ - сильно выпукл с const μ $0<\alpha<\frac{2}{L}$, $\alpha=\alpha_k$ тогда $\lVert x_k - x^* \rVert \le Cq^k$, $0\le q\le 1$ • если $f(x)$ - дважды дифф на $\mathbb{R}^n$ $\mu I \le \nabla^2 f(x) \le LI$, $\mu>0$ тогда $\lVert x_k - x^* \rVert \le \lVert x_0 - x^* \rVert q^k$ $q^* = \frac{L-\mu}{L+\mu}$ $\alpha = \alpha^* = \frac{2}{L+\mu}$	$f(x^N) - f(x^*) \le \frac{2LR^2}{N}$ для f- выпуклой $O\left(\frac{LR^2}{\varepsilon}\right)$ шагов $\Omega\left(\sqrt{\frac{LR^2}{\varepsilon}}\right)$ - - - - - для f - сильно выпуклой $O\left(\varkappa \ln \frac{LR^2}{\varepsilon}\right)$ $\Omega\left(\sqrt{\varkappa} \ln \frac{\mu R^2}{\varepsilon}\right)$ $\varkappa = \frac{L}{\mu}$ - число обусл.
наискорейший спуск 1 порядок безусловная min	градиентный метод + $\alpha_k = \arg\min_{\alpha\ge 0} \varphi_k(\alpha)$, $\varphi_k(\alpha) = f(x_k - \alpha\nabla f(x_k))$	если $f(x)$ - непр. дифф и $\{x: f(x) \le f(x_0)\}$ то $\exists x_{k_i} \to x^*$ и $\nabla f(x^*) = 0$ (условие Липшица на ∇ заменим непрерывностью ∇)	
метод Ньютона (квадратичная аппроксимация) 2 порядок безусловная min	f - дважды дифф на $\mathbb{R}^n$ $f_k(x) = f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \langle \frac{1}{2} \nabla^2 f(x_k) \cdot (x - x_k), x - x_k \rangle$ $h_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ он локальный $\nabla^2 f(x) \succeq \mu I$, $\mu>0$ $\nabla^2 f(x)$ - липшицев если нет условия Липшица, то геометрическая скорость сход. если $\nabla^2 f \preceq \mu I$, $\mu>0$, то $\nabla^2 f(x) + \Delta E$ где $\Delta E = \arg\min \lVert \Delta E \rVert$	если • $\nabla^2 f(x)$ удовл. усл. Липш. с константой L • $f(x)$ - сильно выпукл с const=μ • $q = \left(\frac{L}{2\mu^2}\right) \lVert \nabla f(x_0) \rVert < 1$ то $\lVert x_k - x^* \rVert \le \frac{2\mu}{L} q^{2^k}$ если • f - дважды непр дифф и её гессиан липшицев с L • $\exists x^*$ - лок. min: $\nabla^2 f(x) \succeq \mu I_n$ • $\lVert x_0 - x^* \rVert_2 \le \frac{2\mu}{3L}$ то $\lVert x_k - x^* \rVert_2 \le \frac{L \lVert x_k - x^* \rVert_2^2}{2(\mu - L\lVert x_k - x^* \rVert_2)}$	$\lVert x_k - x^* \rVert \le \frac{2\mu}{L} q^{2^k}$ (квадратичная скорость) $O(n^3)$ - вычислений
демпфирован. метод Ньютона 2 порядок безусловная min	придаём глобальную сходимость $\alpha_k = \arg\min f(x_k - \alpha[\nabla^2 f(x_k)]^{-1} \nabla f(x_k))$ либо Армихо, наискорейший спуск, одномерная min	если • $f(x)$ - сильно выпуклая • 2 непр. дифф • $\mu \lVert s \rVert^2 \le \langle s, \nabla^2 f(x) s \rangle \le L \lVert s \rVert^2$ то глобальная сходимость	сильно выпуклая ф-ция → сверхлинейная скорость липшицевая 2-ая производная - квадратичная скорость
квазиньютоновский безусловная min	$x_{k+1} = x_k - \alpha_k H_k \nabla f(x_k)$, $H_k \ne \nabla^2 f(x_k)$ H_k - симметричная, положительно-опред. $H_k - [\nabla^2 f(x_k)]^{-1} \to 0$ - в пределе МН $H_{k+1}(\nabla f(x_{k+1}) - \nabla f(x_k)) = x_{k+1} - x_k$ $\Delta H_k = H_{k+1} - H_k$ $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ $\delta_k = x_{k+1} - x_k$	$\Delta H_k = \frac{(\delta_k - H_k y_k)(\delta_k - H_k y_k)^T}{\langle \delta_k - H_k y_k, y_k \rangle}$ - одноранговая коррекция $\Delta H_k = \frac{\delta_k \delta_k^T}{\langle y_k, \delta_k \rangle} - \frac{H_k y_k y_k^T H_k}{\langle H_k y_k, y_k \rangle}$ - Дэвидона - Флетчера - Пауэлла	

МСГ 1 порядок безусловный min	$f(x)$ - дифф. на $\mathbb{R}^n$ $\langle \nabla f(x_i), \nabla f(x_j)\rangle = 0$, $\langle A h_i, h_j\rangle = 0$ $x_{k+1} = x_k + \alpha_k p_k$, $p_0 = \nabla f(x_0)$ $\alpha_k = \arg\min f(x_k + \alpha p_k) = -\frac{\langle \nabla f(x_k), p_k\rangle}{\langle p_k, A p_k\rangle}$ $\beta_k = \frac{\langle \nabla f(x_{k+1}), A p_k\rangle}{\langle p_k, A p_k\rangle}$, $p_{k+1} = \nabla f(x_{k+1}) - \beta_k p_k$	если выпуклая квадратичная ф-ция достигает своего min на $\mathbb{R}^n$, то МСГ находит точный min не более чем за n шагов	в окрестности точки строгого min - локальная квадратичная сходимость в общем случае - глобальная сходимость как у градиентного
метод Нестерова 1 порядок безусловный min	$\begin{cases} x_{k+1} = y_k - \frac{1}{L}\nabla f(y_k) \\ y_{k+1} = x_{k+1} + \beta_k (x_{k+1} - x_k)\end{cases}$ $\alpha_k = \frac{1}{L}$ $\beta_k = \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}}$ f - дифф. на $\mathbb{R}^n$ откуда L? А вот: $f(y) \le f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|x-y\|_2^2$ нужно подбирать шаг!	если f - выпуклая с липшицевым градиентом и шаг $\alpha_k = \frac{1}{L}$, то $f(x_k) - f^* \le \frac{2L\|x_0 - x^*\|_2^2}{(k+1)^2}$ если f - сильно выпуклая с липшицевым градиентом и $\alpha_k = \frac{1}{L}$, то $f(x_k) - f^* \le L\|x_k - x_0\|_2^2 \left(1 - \frac{1}{\sqrt{k}}\right)^k$	оценка сверху совпадает с оценкой снизу нелинейные комбинации градиентов могут дать более быстрые на практике методы
метод тяжёлого шарика 1 порядок безусловный min	$x_{k+1} = x_k - \alpha\nabla f(x_k) + \beta(x_k - x_{k-1})$ α_k - шаг вдоль антиградиента f - сильно выпуклая с const μ и липшицевым градиентом с const L не обязательно монотонный решает проблему осцилляций для плохо обусловленных задач	если f - сильно выпукл. с липш. градиентом, $\alpha_k = \frac{4}{(\sqrt{L} + \sqrt{\mu})^2}$ и $\beta_k = \max(\lvert 1 - \sqrt{\alpha_k L}\rvert^2, \lvert 1 - \sqrt{\alpha_k \mu}\rvert^2)$, то $\left\|\begin{pmatrix}x_{k+1}-x^*\\ x_k - x^*\end{pmatrix}\right\|_2 \le \left(\frac{\sqrt{k}-1}{\sqrt{k}+1}\right)^k \left\|\begin{pmatrix}x_1-x^*\\ x_0-x^*\end{pmatrix}\right\|_2$	оценка совпадает с оценкой снизу для методов 1 порядка для сильно-выпуклых ф-ций
стохастический градиентный спуск 1 порядок безусловный min	$\mathbb{E}[g(x)\mid x] = \nabla f(x)$ - вдруг не всегда знаем ∇ в методе градиентного спуска: $\nabla f(x^k)$ заменен на $g(x^k)$ $f(x): \mathbb{R}^n \to \mathbb{R}$ имеет липшицев градиент с const L и μ-сильно выпукла но проблемы с точностью - их решаем с помощью мини-батчинга или уменьшающегося шага	• $\mathbb{E}(f(x^N)) - f(x^*) \le (1-\mu\gamma)^N(\mathbb{E}(f(x^0)) - f(x^*)) + \sigma^2/2\gamma$ • $f(x^N) - f(x^*) \le (1-\frac{\mu}{L})^N(f(x^0) - f(x^*))$ - если f - сильно выпуклая с μ и имеет липшицев градиент с const L в $\|\cdot\|_2$ • $\mathbb{E}[\|g^L(x) - \nabla f(x)\|_2^2] = \frac{1}{L^2}\mathbb{E}\left[\left\|\sum_{i=1}^{L} g^i - \nabla f(x)\right\|_2^2\right] \le \frac{\sigma^2}{L}$ - mini batch • $\mathbb{E}[f(x^N)] - f(x^*) \le \frac{2L}{\mu k + 2L}\max\{f(x^0) - f(x^*), \frac{\sigma^2}{\mu}\}$ - уменьш. шаг	$O(\frac{L}{\mu}\ln\frac{1}{\varepsilon})$ шагов если σ = 0 - совпадает с градиентным спуском
МПГ условный min	$\min_{x\in Q} f(x)$ Q - замкнутое и выпуклое f(x) - дифф. на Q если x^* - точка min, то x^* - решение нерав. $\langle \nabla f(x^*), x - x^*\rangle \ge 0$, если f(x) - выпукла, то это условие и достаточное f - L-гладкая ф-ция $\pi_X(a)$ - проекция $\Leftrightarrow \langle \pi_X(a) - a, x - \pi_X(a)\rangle \ge 0$ если f - L-гладкая, $x_k \in Q$, то $x^{k+1} = \pi_Q(x^k - \frac{1}{L}\nabla f(x^k))$ для L-гладкой ф-ции на каждом шаге метода с длиной шага $\gamma = \frac{1}{L}$ мы min квадратичную аппроксимацию на Q f(x) убывает от итерации к итерации $x^{k+1} = \pi_Q(x^k - \gamma_k \nabla f(x^k)) = \arg\min_{x\in Q}\{f(x^k) + \langle \nabla f(x^k), x - x^k\rangle + \frac{1}{2\gamma_k}\|x - x^k\|_2^2\}$	если f - выпуклая, L-гладкая, $Q \subset \mathbb{R}^n$ - выпукло и замкнуто, $\bar{x}^N = \frac{1}{N}\sum_{k=1}^{N} x^k$, тогда $f(\bar{x}_N) - f(x^*) \le \frac{LR^2}{2N}$, где $R = \|x_0 - x^*\|_2$ если f - μ-сильно-выпуклая, L-гладкая, $Q \subset \mathbb{R}^n$ - выпукло и замкнуто, то $\|x^N - x^*\|_2^2 \le (1 - \frac{\mu}{L})^N \|x_0 - x^*\|_2^2$	$O(\frac{LR^2}{\varepsilon})$ сублинейная скорость сход. в случае выпуклой f в случае f-сильно выпуклой - линейная скорость $O(\frac{L}{\mu}\ln\frac{\|x_0 - x^*\|_2^2}{\varepsilon})$

Метод зеркального спуска условная min	μ-сильно выпуклая относительно $\|\cdot\|$, если $f(y) \ge f(x) + \langle \nabla f(x), y-x\rangle + \frac{\mu}{2}\|y-x\|_2^2$ L-гладкая относительно $\|\cdot\|$, если $f(y) \le f(x) + \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|_2^2$ $d(y) \ge d(x) + \langle \nabla d(x), y-x\rangle + \frac{1}{2}\|y-x\|^2$ - прокс ф. $V_d(x,y) = d(x) - d(y) - \langle \nabla d(y), x-y\rangle$ - div метод: $\begin{cases} \nabla d(y^k) = \nabla d(x^k) - \gamma_k \nabla f(x^k) \\ x^{k+1} = \arg\min_{x\in Q} V_d(x, y^k) \end{cases}$ d - сильно выпуклая с $\mu=1$ V_d - сильно выпуклая по x f - дифф $x^{k+1} = \arg\min \left[\underbrace{\langle \nabla f(x^k), x-x^k\rangle}_{\text{линейное приближение}} + \underbrace{\frac{1}{\gamma_k} V_d(x,x^k)}_{\text{отвечает за проекцию на } Q} \right]$ почему зеркальный? 1) $\nabla d(x) : (Q_0, \|\cdot\|) \to (\mathbb{R}^n, \|\cdot\|_*)$ 2) $x^k \to \nabla d(x^k)$ из $(\mathbb{R}^n, \|\cdot\|_*)$ 3) в $(\mathbb{R}^n, \|\cdot\|_*)$ шаг и получим $\nabla d(y^k)$ 4) $\nabla d(y^k) \to (Q_0, \|\cdot\|)$ и получим x^{k+1} градиентный спуск происходит в двойственном пр-ве	если · $\|\nabla f(x)\|_* \le M$ · $R>0: R^2 \ge 2\inf_{x\in X^*} V_d(x,x_0)$ · $h_k = \frac{\varepsilon}{M\|\nabla f(x^k)\|}$ - шаг то $\forall k \ge K = \frac{M^2R^2}{\varepsilon^2} \hookrightarrow$ $f(\bar{x}^k) - f^* \le \varepsilon$ здесь X^* - мн-во решений $\min_{x\in Q} f$ M, R завися от $\|\cdot\|$	в пространствах большой размерности можно получить выигрыш по кол-ву итераций