2) Метод Ньютона

2.1) Описание процесса

Будем ∢ зад. условной минимизации: $J(u) \to \inf$, $u \in U \subseteq H$, (1) в $\swarrow$ мн-во U допустимых эл-тов и необязательно совпадает со всем пр-вом H, а ф-ция $J(u) \in C^2(U)$

Итерац. посл-ть метода строится след. образом. ∃ уже известно k-е приближение $u_k \in U$. Возьмем квадратичн. часть приращ-я $J(u) - J(u_k) = J_k(u) + \bar{o}(\|u-u_k\|^2)$, где

$J_k(u) = \langle J'(u_k), u-u_k \rangle + \frac{1}{2} \langle J''(u_k)(u-u_k), u-u_k \rangle, u \in U$, (2)

после чего вычислим u_{k+1} по ф-ле $u_{k+1} = \underset{u \in U}{\operatorname{argmin}} J_k(u)$ (3)

Почему данный метод принято называть мет. Ньют.? Чтобы пояснить название, ∢ случай $U = H \Rightarrow$ реш-е зад. $J_k(u) \to \inf$, $u \in U$, необходимо удовл. ур-нию $J_k'(u) = 0$, т.е. $J_k'(u_{k+1}) = 0 = J'(u_k) + J''(u_k)(u_{k+1} - u_k) = 0$, откуда $\left[u_{k+1} = u_k - (J''(u_k))^{-1} J'(u_k) \text{ (H)}\right]$ (типа решаем зад. $J'(u) = 0$)

If мы теперь вспомним обычн. мет. Ньют. д/реш-я ур-я $f(x) = 0$, то есть процесс $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$, $k = 0, 1, \ldots$, то видим, что мет. (3) — это тот же мет. реш-я ур-я $J'(u) = 0$ в случае гильб. пр-в. Отсюда и назв-е мет.

Более общ. схема мет. Ньют.: $u_{k+1} = u_k + \alpha_k(\bar{u}_k - u_k)$, где $\bar{u}_k = \underset{u \in U}{\operatorname{argmin}} J_k(u)$. Величина $\alpha_k \in [0, 1]$ наз-ся шагом мет. Ньют. В завис-ти от способа выбора α_k можно получить различ. вар-ты мет. Ньют. Например, if взять $\alpha_k \equiv 1$, то $u_{k+1} = \bar{u}_k$, и мы получаем так называемый классич. мет. Ньют. Он имеет квадратичн. ск-ть сх-ти, но, как будет показано ниже, сх-ся лишь локально

2.2) Сх-ть метода

Т (о сх-ти мет. Ньют.) ∃ U — вып., замкн. мн-во из H.

$\operatorname{int} U \neq \emptyset$. Ф-ция $J(u)$ определена и явл-ся сил. вып. с const $\mu > 0$, $J(u) \in C^2(U)$, $J''(u)$ удовл. на U усл-ю Липшица с const L, т.е. $\|J''(u) - J''(v)\| \leq L\|u - v\|$, $\forall u, v \in U$. (5)] начал. хорошим выбором

приближ-е $u_0 \in U$ удовл. соотнош-ю $\|u_0 - u_*\| < \frac{2\mu}{L}$, (6) где u_* — реш-е зад. (1) и $\langle J''(u)\xi, \xi \rangle \geq \mu\|\xi\|^2$ (7) Тогда мет. (3) порождает посл-ть $\{u_k\}$: $\|u_k - u_*\| \leq \frac{2\mu}{L} q^{2^k}$, $k = 0, 1, 2, \ldots$,

$q = \frac{L\|u_0 - u_*\|}{2\mu} < 1$ (8)

↓ квадратичн. ск-ть сх-ти

▲ При выполнении усл. Т зад. минимиз. (1) удовл. всем требов-м сил. вып. Т Вейерштр., поэтому ее реш-е u_* $\exists$ и !. $J_k(u) = \langle J'(u_k), u - u_k \rangle + \frac{1}{2}\langle J''(u_k)(u - u_k), u - u_k \rangle \in C^2(U)$, (9) причем $J_k'(u) = J'(u_k) + J''(u_k)(u - u_k)$, $J_k''(u) = J''(u_k)$ (10). Отсюда $\Rightarrow$, что $\langle J_k''(u)\xi, \xi \rangle = \langle J''(u)\xi, \xi \rangle \geq \mu\|\xi\|^2$, $\forall u \in U$, $\xi \in H$. Поэтому ф-ция $J_k(u)$ также явл-ся сил. вып. на U с const μ. Стало быть, и зад. нахожд-я т. u_{k+1} тоже удовл. всем треб.-ям Т Вейерш. д/сил. вып. ф-ций, и u_{k+1} $\exists$ и !. Итак, при выполнении усл-й доказываем. Т классич. метод Ньют. определен корректно.

Теперь восп-ся крит. опт-ти д/исходн. зад. и д/ зад. поиска u_{k+1}: $\langle J'(u_*), u - u_* \rangle \geq 0$, $\forall u \in U$; (11)

$\langle J_k'(u_{k+1}), u - u_{k+1} \rangle = \langle J'(u_k) + J''(u_k)(u_{k+1} - u_k), u - u_{k+1} \rangle \geq 0$, $\forall u \in U$ (12)

Подставляя в 1е из этих нерав-в $u = u_{k+1} \in U$, а во 2ое $u = u_* \in U$, после чего складывая полученные соотнош-я, мы имеем: $\langle J'(u_*) - J'(u_k) - J''(u_k)(u_{k+1} - u_k), u_{k+1} - u_* \rangle \geq 0 \Leftrightarrow \langle J'(u_*) - J'(u_k), u_{k+1} - u_* \rangle - \langle J''(u_k)(u_* - u_k), u_{k+1} - u_* \rangle \geq \langle J''(u_k)(u_{k+1} - u_*), u_{k+1} - u_* \rangle$

Оценим лев. и прав. части получ. нер-ва. Д/этого восп-ся

...ной конечн. приращ-й, нер-вом К-Б, усл. (5) и крит. сил. вып.: $\langle J'(u_*) - J'(u_k), u_{k+1} - u_* \rangle - \langle J''(u_k)(u_* - u_k), u_{k+1} - u_* \rangle =$
$= \int_0^1 \langle J''(u_k + t(u_* - u_k))(u_* - u_k), u_{k+1} - u_* \rangle dt - \langle J''(u_k)(u_* - u_k), u_{k+1} - u_* \rangle = \int_0^1 \langle (J''(u_k + t(u_* - u_k)) - J''(u_k))(u_* - u_k), u_{k+1} - u_* \rangle dt$
$\leqslant \int_0^1 L \| t(u_* - u_k) \| \cdot \| u_* - u_k \| \cdot \| u_{k+1} - u_* \| dt = \frac{L}{2} \| u_* - u_k \|^2 \| u_{k+1} - u_* \|,$
$\langle J''(u_k)(u_{k+1} - u_*), u_{k+1} - u_* \rangle \geqslant \mu \| u_{k+1} - u_* \|^2$

С учетом этих оценок мы получаем: $\frac{L}{2} \| u_* - u_k \| \| u_{k+1} - u_* \| \geqslant$
$\geqslant \mu \| u_{k+1} - u_* \|^2 \iff \| u_{k+1} - u_* \| \leqslant \frac{L}{2\mu} \| u_k - u_* \|^2, \ k = 0, 1, 2, \ldots$

Отсюда уже нетрудно вывести нужный рез-т, используя мет. мат. индукции. Действительно, при $k=0$ с учетом $q = \frac{L}{2\mu} \| u_0 - u_* \| < 1$ мы имеем: $\| u_1 - u_* \| \leq \frac{L}{2\mu} \| u_0 - u_* \|^2 = \frac{2\mu}{L} q^2 =$
$= \frac{2\mu}{L} q^{2^1}$

Наконец, по индукции предполагая, что $\| u_k - u_* \| \leq \frac{2\mu}{L} q^{2^k}$, имеем: $\| u_{k+1} - u_* \| \leqslant \frac{L}{2\mu} \| u_k - u_* \|^2 \leqslant \frac{L}{2\mu} \left(\frac{2\mu}{L} q^{2^k} \right)^2 = \frac{2\mu}{L} q^{2^{k+1}}$ ∎

Достоинства: высокая ск-ть сх-ти.

Недостатки: $J''(u)$ сложно найти

- На каждом шаге необх. решать доп. зад. minимиз.
- картинка Метод требует хорошего начал. приближ-я → мб пробл., чтобы избежать, исп-ть континуир. методы

Зам. $\exists$ квазиньютоновские методы $u_{k+1} = u_k - A_k J'(u_k)$, где оп-р $A_k \in \mathcal{L}(H \to H)$: $\lim\limits_{k \to \infty} \| A_k - (J''(u_k))^{-1} \| = 0$

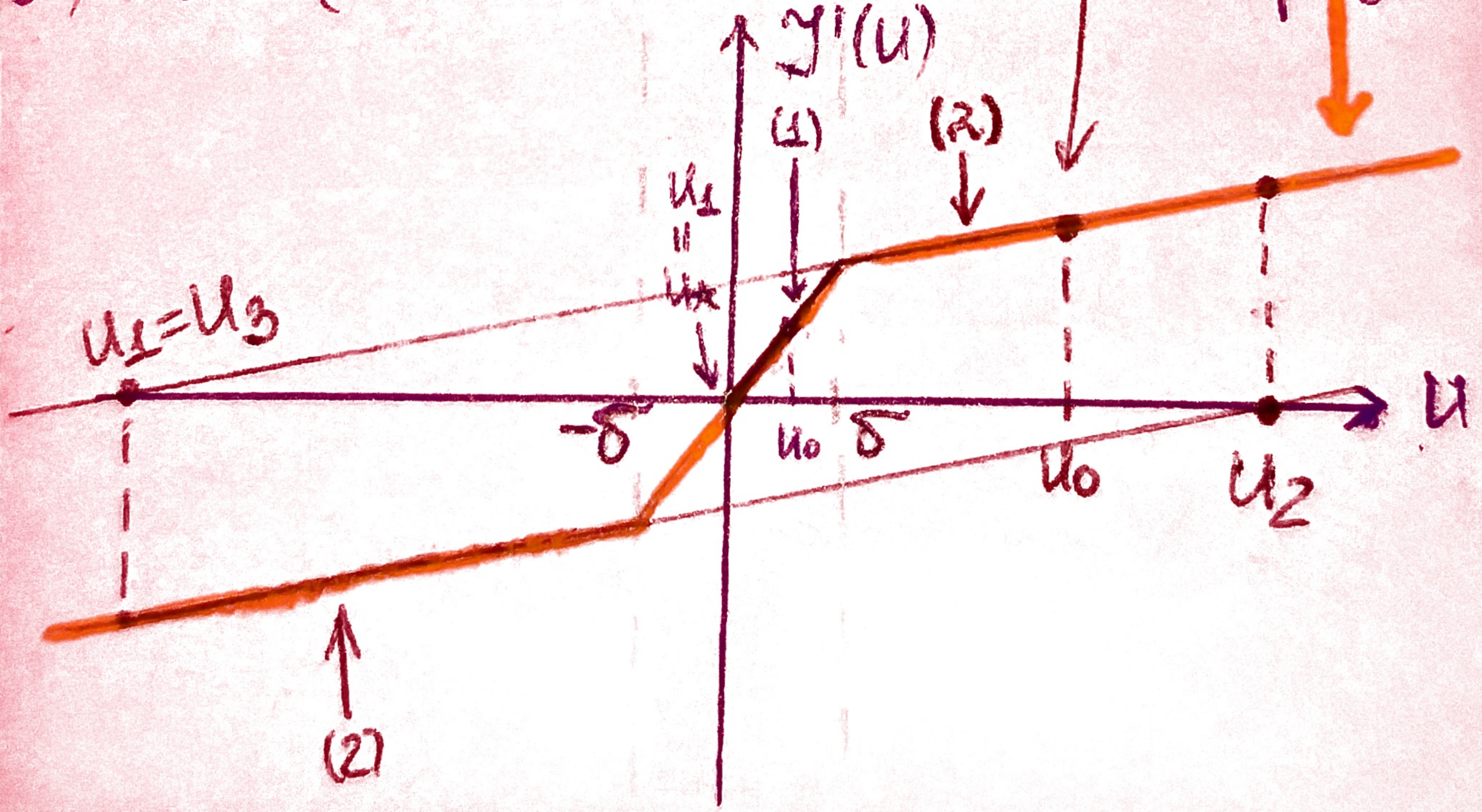
Пример $J(u)\to\inf$, $u\in H=\mathbb{R}^1 \Rightarrow$

$\Rightarrow J'(u)=0$; мет. Н.: $u_{k+1}=u_k - (J''(u_k))^{-1}J'(u_k)$

$u_*=0$

1) $u_0\in(-\delta,\delta)\Rightarrow u_1=u_*$

2) $u_0\notin(-\delta,\delta)\Rightarrow u_k \not\to u_*$



Метод покоординатного спуска

На пред. лекциях, в основном, были Σ итерац. методы минимиз., к-е д/своей реализации требуют вычисл-я 1ых или 2х производ-х целевой ф-ции. Однако в практич. зад. нередко либо минимизируем. ф-ция не обладает нужной гладкостью, либо вычисл-е и производ-х с нужной точностью оказывается слишком трудоемким. Д/подобных ситуаций желательно иметь в запасе мет. мин., при реализации α треб-ся вычислять лишь знач-я ф-ции. Одним из таких мет. явл-ся метод покоординатн. спуска.

Рассм. этот метод применительно к зад. мин. без огранич-й в конечномер. пр-ве: $J(u) \to \inf, u \in \mathbb{R}^n$ (1) Выберем в пр-ве $\mathbb{R}^n$ некот. базис, например, стандартн. ОНБ $\{e_i\}_{i=1}^n$. При реализации мет. будет производ-ся циклич. перебор этих базисн. в-ов, поэтому д/удобства описания итерац. процесса (теоретически бесконечного) выстроим их в единую ∞ посл-ть: $p_0 = e_1, p_1 = e_2, \ldots, p_{n-1} = e_n, p_n = e_1, p_{n+1} = e_2, \ldots, p_{2n-1} = e_n, p_{2n} = e_1, \ldots$ (2)

Перед запуском метода выбир-ся некот. старт. т. $u_0 \in \mathbb{R}^n$, стартов. шаг $\alpha_0 > 0$ и коэфф-т дробления шага $\lambda \in (0,1)$. Допустим, что на k-й итерации найдено k-е приближ-е u_k и текущ. знач.-е шага $= \alpha_k > 0$. Д/опр-ния очередного приближ-я выпол-ся след. действия. Берется базис. в-р p_k и вычисл-ся знач-е ф-ции в т. $u = u_k + \alpha_k p_k$.

If $J(u_k + \alpha_k p_k) < J(u_k)$, то $u_{k+1} = u_k + \alpha_k p_k$, $\alpha_{k+1} = \alpha_k$ (3) и процесс продолж-ся из т. u_{k+1} со след. по пор-ку базисн. направл-м p_{k+1}. If $J(u_k + \alpha_k p_k) \geqslant J(u_k)$, то вычисл-ся знач-е ф-ции в противополож. т. $u = u_k - \alpha_k p_k$.

If $J(u_k - \alpha_k p_k) < J(u_k)$, то $u_{k+1} = u_k - \alpha_k p_k$, $\alpha_{k+1} = \alpha_k$ (4) и переходим к след. итерации с очередным базис. в-ром p_{k+1}. Будем наз-ть $(k+1)$-ую итерацию удачной, if переход от

u_k к u_{k+1} произошел по одному из усл. (3) или (4) и сопровождался строгим убыванием знач-я ф-ции. If же $J(u_k - \alpha_k p_k) \geq J(u_k)$, т.е. нарушены оба усл-я (3) и (4), назовем $(k+1)$-ю итерацию неудачной. Дальнейшие действия завис. от предыстории. В процессе вычисл-й ведется подсчет числа неудачн. итераций, случившихся подряд. If их общ. кол-во вместе с неудачей на текущ. $(k+1)$-м шаге еще не достигло n, то полагают $u_{k+1} = u_k$, $\alpha_{k+1} = \alpha_k$ (5) и переходят к очередн. базисн. направл-ю p_{k+1}. If же все $n-1$ итерации, предшествующие неудаче, зафиксир.-й на $(k+1)$-м шаге, также были неудачн., то производится дробление шага α_k с априорно выбранным коэфф-м $\lambda \in (0,1)$: $u_{k+1} = u_k$, $\alpha_{k+1} = \lambda \alpha_k$ (6) и проверка основных усл-й (3) и (4) продолж-ся на более мелкой сетке. Дост. усл-я сх-ти описанного вар-та мет. покоорд. спуска содержит след. Т.

Т. $\exists$ ф-ция $J(u)$ вып. на $\mathbb{R}^n$ и принадлежит классу $C^1(\mathbb{R}^n)$, а начал. приближ-е u_0: соответствующ. ему мн-во Лебега $M_0 = \{u \in \mathbb{R}^n \mid J(u) \leq J(u_0)\}$ огр. Тогда посл-ть u_k, вырабатываемая мет-м (2)-(6), сх-ся и по ф-ции и по аргументу: $\lim\limits_{k\to\infty} J(u_k) = J_*$, $\lim\limits_{k\to\infty} \rho(u_k, U_*) = 0$ (7)

▲ По усл., мн-во M_0 огр, а из непр-ти ф-ции $J(u) \Rightarrow$ его замкн-ть. Кроме того, это мн-во по построению содержит в себе все точки, перспективные в плане мин. $J(u)$ на всем пр-ве $\mathbb{R}^n$, поэтому в силу классич. (конечномер) Т Вейерштр. будем иметь $J_* > -\infty$ и $U_* \neq \varnothing$. Из описания мет. (2)-(6) $\Rightarrow$ монотонность: $J(u_{k+1}) \leq J(u_k)$, $k = 0, 1, \ldots$, поэтому $u_k \in M_0$ и $\exists \lim\limits_{k\to\infty} J(u_k) \geq J_*$. Покажем, что $\lim\limits_{k\to\infty} \alpha_k =$

$=0$, т.е. моментов дробления шага мет. будет ∞ много.
Допустим противное: $\exists$ последнее дробление состоялось на N-м шаге, после чего процесс вычисл-й (∞) продолж-ся с фикс. шагом $\alpha_k = \alpha_N = \alpha > 0$, $k = N, N+1, \dots$ $*$ в пр-ве $\mathbb{R}^n$ дискрет. сетку (решетку) M_α с одним и тем же равномерн. шагом $\alpha > 0$ по всем координатн. направл-м и поместим т. u_N в один из ее узлов. Из описания мет. покоорд. спуска $\Rightarrow$, что все следующ. приближ-я u_k явл-ся узлами этой самой решетки, т.е. $u_k \in M_\alpha$, $\forall k \geq N$, и остаются в пределах огр. мн-ва M_0, в к. может нах-ся лишь конечн. число узлов сетки M_α. Но при перемещ-ии по конеч. мн-ву узлов невозможно ∞ число раз наблюдать строгое убывание знач-й ф-ции, без к. обязательно произошло бы дробление шага. Полученное противоречие показывает, что процесс дробления α_k бесконечен и $\lim\limits_{k\to\infty} \alpha_k = 0$.

$\exists$ $k_1 < k_2 < \dots < k_m < \dots$ — номера тех итераций, на к. длина шага α_k дробится. В соотв. с описан. проц-й (2)–(6) этим дроблением предшествуют серии ровно из n неудачных итераций по всем базисн. направл-м: $J(u_{k_m} + \alpha_{k_m} e_i) \geq J(u_{k_m})$, $J(u_{k_m} - \alpha_{k_m} e_i) \geq J(x_{k_m})$, $i = 1, 2, \dots$ (8)

Из посл-ти точек u_{k_m}, принадлежащих огр. мн-ву M_0, можно выбрать сх-ся п/посл-ть. Без огранич-я общности можем считать, что сама п-ть u_{k_m} сх-ся к некот. т. u_*. С помощью ф-лы конечн. приращ-й перепишем нер-ва (8) в виде $\langle J'(u_{k_m} + \theta_m^+ \alpha_{k_m} e_i), e_i \rangle \alpha_{k_m} \geq 0$,
$\langle J'(u_{k_m} - \theta_m^- \alpha_{k_m} e_i), e_i \rangle (-\alpha_{k_m}) \geq 0$, (9) при некотор. $\theta_m^+, \theta_m^- \in [0,1]$

$m=1,2,\ldots$. Поделим нер-ва (9) на $\alpha_{k_m}>0$, после чего перейдем в них к $\lim\limits_{m\to\infty}$. С учетом усл-я $J'(x)\in C(\mathbb{R}^n)$ непр-ти grad-а, огр-ти знач-й $\theta_m^{\pm}$ и сх-ти $\alpha_{k_m}\to 0$ получим соотнош-я $\langle J'(u_*), e_i\rangle \geqslant 0$, $\langle J'(u_*), e_i\rangle \leqslant 0$, $i=\overline{1,n}$. Т.к. $\{e_i\}_{i=1}^n$ – базис в $\mathbb{R}^n$, то отсюда $\Rightarrow$, что $J'(u_*)=0$, а поскольку ф-ция $J(u)$ вып., то т. u_* явл-ся одной из т. min: $u_*\in U_*$. Получ-ся, что п/посл-ть u_{k_m} явл-ся мин-ющей: $\lim\limits_{m\to\infty} J(u_{k_m}) = J(u_*) = J_*$. Отсюда и из монот-ти посл-ти $J(u_k)\Rightarrow$, что минимизирующ. явл-ся и вся посл-ть u_k: $\lim\limits_{k\to\infty} J(u_k) = J_*$, т.е. 1ое из утв. (7) доказано. 2е утв. о сх-ти по arg-ту выводится из сх-ти по ф-ции и компактности в $\mathbb{R}^n$ мн-ва M_0 (см. классич. T Вейерштр.) ■

Заметим, что при реализации описанного вар-та мет. покоорд. спуска не исп-ся знач-я grad-ов мин-емой ф-ции, однако в усл. T о сх-ти метода присутствует требование гладк-ти этой ф-ции. Приведем пример, показывающ., что при отсутствии гладк-ти без усиления остальных усл. T сх-ть мет. покоорд. спуска гарантировать нельзя.

Пример $\divideontimes$ двумерную зад. min-ции $J(u)=(x-1)^2+(y-1)^2+2|x-y|\to\inf$, $u=(x,y)\in\mathbb{R}^2$

не дифф-ма

Ф-ция J непр., сил. вып., огр. снизу на всем пр-ве $\mathbb{R}^2$, достигает своей нижней грани $J_*=0$ в единств. т. $u_*=(x_*=1, y_*=1)$ и на прямой $y=x$ не явл-ся дифф-мой. Нетрудно проверить, что при выборе в $\mathbb{R}^2$ стандарт. базис. направлений $e_1=(1,0)$, $e_2=(0,1)$ в случае запуска мет. покоорд. спуска из начала координат $u_0=(0,0)$ независимо от выбора стартов. шага $\alpha_0>0$ процесс зациклится в начал. т. $u_k=u_0=$

$= (0,0)$, $k = 1, 2, \dots$, и сх-ти к реш-ю зад. мин-ции не будет ни по ф-ции, ни по арг-у: $J(u_k) - J_* = J(u_0) - J_* = 2$, $|u_k - u_*| = |u_0 - u_*| = \sqrt{2}$, $k = 1, 2, \dots$

∃ и друг. вар-ты мет. покоорд. спуска. Так, вместо стандарт. базиса из единичн. координатн. в-ов, исп-ся в (2), можно взять в R^n произвол. базис $\{e_i\}_{i=1}^{n}$, не обязательно ортонормир. При этом утв-я T останутся в силе, правда, „покоорд. спуском" соответствующ. итерац. процесс можно будет назвать с некот. натяжкой. Кстати, if в приведенном с негладк. ф-цией $J(u)$ в кач-ве базис. направл-й вместо стандарт. координатн. взять в-ры $e_1 = (1, 1)$, $e_2 = (1, -1)$, то из того же самого начал. приближ-я $u_0 = (0, 0)$ д/∀ старт. шага $\alpha_0 > 0$ посл-ть u_k будет сх-ся к реш-ю зад.

Достоинства: простота, скромные требования к гладкости мин-емой ф-ции

Недостатки: невысокая ск-ть сх-ти, неспособность к распознаванию принадлежности вычисляемых приближ-й u_k оптим. мн-ву U_*

Д/З

12.1 Примените метод проекции градиента с постоянным шагом $\alpha = \frac{1}{2}$ к следующей задаче условной оптимизации в пространстве $\mathbf{E}^2$

$$f(u) = 2x^2 + (x - y)^2 \to \inf, \quad u \in U = \{u = (x, y) \in \mathbf{E}^2 : 1 \leqslant x \leqslant 3, \ -2 \leqslant y \leqslant 2\},$$

взяв в качестве начального приближения точку $u_0 = (3, 0)$

13.1 Сделать два шага метода покоординатного спуска для решения задачи минимизации в пространстве $\mathbf{E}^2$

$$f(u) = (x - 1)^2 + (y - 1)^2 \to \inf, \quad u = (x, y) \in \mathbf{E}^2$$

Замечание. см. лекцию 13