ОБ АЛГОРИТМЕ КОЗИНЦА. II*

В. Н. Малоземов Н. А. Соловьева Г. Ш. Тамасян
v.malozemov@spbu.ru 4vinyo@gmail.com grigoriytamasjan@mail.ru

21 марта 2024 г.

Данный доклад является непосредственным продолжением доклада [1]. Используется нумерация предыдущих разделов, формул, лемм и теорем.

5°. Неусеченные и усеченные итерации. В дальнейшем будем считать, что выполнено условие $\Delta_k > 0$ при всех $k = 0, 1, \ldots$ Это условие гарантирует, в частности, что при всех $k = 0, 1, \ldots$ вектор $w_k = x_k - y_k$ отличен от нулевого.

При $\Delta_k > 0$ план (x_k, y_k) задачи (1) порождает k-ю итерацию алгоритма Козинца. Прежде всего, вычисляется $\widetilde{\lambda}_k$. При $\widetilde{\lambda}_k \in (0, 1]$ k-ю итерацию назовем *неусеченной*, при $\widetilde{\lambda}_k > 1$ — *усеченной*. На неусеченной итерации полагаем $\lambda_k = \widetilde{\lambda}_k$, на усеченной — $\lambda_k = 1$.

Обозначим через D_1 диаметр множества C_1 и через D_2 — диаметр множества C_2. Положим $D = \max\{D_1, D_2\}$.

ЛЕММА 2. *Допустим, что k-я итерация алгоритма Козинца является неусеченной. Тогда справедливо неравенство*

$$\|w_k\|^2 - \|w_{k+1}\|^2 \geqslant \left(\frac{\Delta_k}{D}\right)^2. \tag{21}$$

Доказательство. Имеем $\Delta_k > 0$ и $\lambda_k = \widetilde{\lambda}_k$. Возможны два случая.

1) $\Delta_k = \Delta_k^{(1)} = \langle p_{i_k} - x_k, y_k - x_k \rangle$, где $i_k \in 1 : s$. По описанию алгоритма

$$\widetilde{\lambda}_k = \frac{\Delta_k}{\|p_{i_k} - x_k\|^2}$$

и $x_{k+1} = x_k + \widetilde{\lambda}_k (p_{i_k} - x_k)$, $y_{k+1} = y_k$. Вычислим

$$\|x_{k+1} - y_{k+1}\|^2 = \left\| x_k - y_k + \widetilde{\lambda}_k (p_{i_k} - x_k) \right\|^2 =$$
$$= \|x_k - y_k\|^2 - 2\widetilde{\lambda}_k \langle p_{i_k} - x_k, y_k - x_k \rangle + \widetilde{\lambda}_k^2 \|p_{i_k} - x_k\|^2 =$$
$$= \|x_k - y_k\|^2 - 2\widetilde{\lambda}_k \Delta_k + \frac{\Delta_k^2}{\|p_{i_k} - x_k\|^2} = \|x_k - y_k\|^2 - \frac{\Delta_k^2}{\|p_{i_k} - x_k\|^2}\,.$$

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» http://oml.cmlaboratory.com/

Отсюда непосредственно следует неравенство (21).

2) $\Delta_k = \Delta_k^{(2)} = \langle p_{j_k} - y_k, x_k - y_k\rangle$, где $j_k \in (s+1) : m$. По описанию алгоритма

$$\widetilde{\lambda}_k = \frac{\Delta_k}{\|p_{j_k} - y_k\|^2}$$

и $x_{k+1} = x_k$, $y_{k+1} = y_k + \widetilde{\lambda}_k(p_{j_k} - y_k)$. Вычислим

$$\begin{aligned} \|x_{k+1} - y_{k+1}\|^2 &= \left\|x_k - y_k - \widetilde{\lambda}_k(p_{j_k} - y_k)\right\|^2 = \\ &= \|x_k - y_k\|^2 - 2\widetilde{\lambda}_k\langle p_{j_k} - y_k, x_k - y_k\rangle + \widetilde{\lambda}_k^2\,\|p_{j_k} - y_k\|^2 = \\ &= \|x_k - y_k\|^2 - 2\widetilde{\lambda}_k\Delta_k + \frac{\Delta_k^2}{\|p_{j_k} - y_k\|^2} = \|x_k - y_k\|^2 - \frac{\Delta_k^2}{\|p_{j_k} - y_k\|^2}\,. \end{aligned}$$

Отсюда непосредственно следует неравенство (21).

Лемма доказана. □

Лемма 2 гарантирует, в частности, что на неусеченных итерациях выполняется неравенство $\|w_k\| > \|w_{k+1}\|$. Нетрудно проверить, что оно выполняется и на усеченных итерациях.

Действительно, пусть $\Delta_k = \Delta_k^{(1)} = \langle p_{i_k} - x_k, y_k - x_k\rangle > 0$, где $i_k \in 1 : s$, и

$$\widetilde{\lambda}_k := \frac{\langle p_{i_k} - x_k, y_k - x_k\rangle}{\|p_{i_k} - x_k\|^2} > 1.$$

Последнее неравенство перепишем в виде

$$\langle p_{i_k} - x_k, y_k - x_k\rangle > \|p_{i_k} - x_k\|^2\,. \tag{22}$$

По описанию алгоритма Козинца имеем $\lambda_k = 1$,

$$x_{k+1} = x_k + (p_{i_k} - x_k) = p_{i_k}, \quad y_{k+1} = y_k.$$

Вычислим

$$\begin{aligned} \|x_{k+1} - y_{k+1}\|^2 &= \|x_k - y_k + (p_{i_k} - x_k)\|^2 = \\ &= \|x_k - y_k\|^2 - 2\langle p_{i_k} - x_k, y_k - x_k\rangle + \|p_{i_k} - x_k\|^2\,. \end{aligned}$$

На основании (22) и условия $x_k \neq p_{i_k}$ получаем

$$\|x_{k+1} - y_{k+1}\|^2 < \|x_k - y_k\|^2 - \|p_{i_k} - x_k\|^2 < \|x_k - y_k\|^2\,.$$

В случае $\Delta_k = \Delta_k^{(2)} = \langle p_{j_k} - y_k, x_k - y_k\rangle$, где $j_k \in (s+1) : m$, и

$$\widetilde{\lambda}_k := \frac{\langle p_{j_k} - y_k, x_k - y_k\rangle}{\|p_{j_k} - y_k\|^2} > 1$$

аналогично предыдущему придем к неравенству

$$\|x_{k+1} - y_{k+1}\|^2 < \|x_k - y_k\|^2 - \|p_{j_k} - y_k\|^2 < \|x_k - y_k\|^2 .$$

Таким образом, во всех случаях выполняется неравенство

$$\|w_k\| > \|w_{k+1}\| , \quad k = 0, 1, \ldots \tag{23}$$

Это означает, что числовая последовательность $\{\|w_k\|\}$ строго убывает. Кроме того, она ограничена снизу нулем. По теореме из математического анализа приходим к следующему заключению.

ЛЕММА 3. *Числовая последовательность* $\{\|w_k\|\}$ *имеет конечный предел.*

Нам потребуется одно специфическое свойство усеченных итераций.

ЛЕММА 4. *Количество идущих подряд усеченных итераций конечно.*

Доказательство. Допустим, что k-я итерация алгоритма Козинца является усеченной и следом за ней идут только усеченные итерации. Если $\Delta_k = \Delta_k^{(1)}$, то по описанию алгоритма имеем $x_{k+1} = p_{i_k}$, $y_{k+1} = y_k$. При этом $\|w_{k+1}\| < \|w_k\|$. В случае, когда и далее $\Delta_{k+l} = \Delta_{k+l}^{(1)}$, $l = 1, 2, \ldots$, получим

$$x_{k+l+1} = p_{i_{k+l}}, \quad y_{k+l+1} = y_k, \qquad \|w_{k+l+1}\| < \|w_{k+l}\| .$$

Такой процесс не может быть бесконечным, поскольку точек p_i конечное число ($i \in 1 : s$). Значит, при некотором l выполнится равенство $\Delta_{k+l} = \Delta_{k+l}^{(2)}$. По описанию алгоритма получим

$$x_{k+l+1} = x_{k+l} = p_{i_{k+l-1}}, \quad y_{k+l+1} = p_{j_{k+l}}, \qquad \|w_{k+l+1}\| < \|w_{k+l}\| .$$

Следующие пары (x_r, y_r) будут иметь вид (p_i, p_j), где $i \in 1 : s$, $j \in (s+1) : m$, при строгом убывании величины $\|w_r\|$. И такой процесс не может быть бесконечным ввиду конечности множества пар (p_i, p_j).

Аналогично рассматривается случай $\Delta_k = \Delta_k^{(2)}$.

Лемма доказана. □

Согласно лемме 4, последовательность (19) может содержать конечные блоки планов (x_k, y_k), порождающих усеченные итерации. Если эти блоки исключить из последовательности (19), то останется бесконечная подпоследовательность планов (x_{k_j}, y_{k_j}), порождающих неусеченные итерации.

6°. Доказательство сходимости алгоритма Козинца. Имеется в виду следующее утверждение.

ТЕОРЕМА 4. *Если* $\Delta_k > 0$ *при всех* $k = 0, 1, \ldots,$ *то*

$$\lim_{k\to\infty} w_k = w^*.$$

Доказательство. Рассмотрим бесконечную подпоследовательность (x_{k_j}, y_{k_j}) последовательности (19), состоящую из планов, порождающих неусеченные итерации. По лемме 2 имеем

$$\left\|w_{k_j}\right\|^2 - \left\|w_{k_j+1}\right\|^2 \geqslant \left(\frac{\Delta_{k_j}}{D}\right)^2. \tag{24}$$

Согласно лемме 3 последовательность $\{\|w_k\|^2\}$ сходится. Значит, левая часть неравенства (24) стремится к нулю при $j \to \infty$. Как следствие, получаем

$$\lim_{j\to\infty} \Delta_{k_j} = 0.$$

По теореме 3,

$$\lim_{j\to\infty} w_{k_j} = w^*.$$

Очевидно, что и

$$\lim_{j\to\infty} \left\|w_{k_j}\right\| = \|w^*\|. \tag{25}$$

Вместе с тем, известно, что вся последовательность $\{\|w_k\|\}$ строго убывает и имеет конечный предел (см. (23) и лемму 3). Учитывая (25), приходим к предельному соотношению

$$\lim_{k\to\infty} \|w_k\| = \|w^*\|. \tag{26}$$

Теперь отметим, что в силу (20) справедливо неравенство

$$\|w_k - w^*\|^2 = \|w_k\|^2 - 2\langle w_k, w^*\rangle + \|w^*\|^2 \leqslant \|w_k\|^2 - \|w^*\|^2.$$

Согласно (26), правая часть этого неравенства стремится к нулю. Значит,

$$\lim_{k\to\infty} \|w_k - w^*\|^2 = 0,$$

что равносильно заключению теоремы.

Теорема доказана. □

7°. Алгоритм Козинца (рабочая схема). Рассмотрим возможную вариацию алгоритма Козинца.

Пусть уже имеется k-е приближение (x_k, y_k). Вычислим оценку

$$\Delta^{(1)}(x_k, y_k) := \max_{i\in 1:s}\langle p_i - x_k, y_k - x_k\rangle = \langle p_{i_k} - x_k, y_k - x_k\rangle.$$

- Если $\Delta^{(1)}(x_k, y_k) > 0$, то находим $\lambda_k = \min\left\{1, \frac{\Delta^{(1)}(x_k,y_k)}{\left\|p_{i_k}-x_k\right\|^2}\right\}$ и

$$x_{k+1} = x_k + \lambda_k(p_{i_k} - x_k).$$

- Если $\Delta^{(1)}(x_k, y_k) \leqslant 0$, то полагаем $x_{k+1} = x_k$.

Вычислим вторую оценку

$$\Delta^{(2)}(x_{k+1}, y_k) := \max_{j\in(s+1):m}\langle p_j - y_k, x_{k+1} - y_k\rangle = \langle p_{j_k} - y_k, x_{k+1} - y_k\rangle.$$

- Если $\Delta^{(2)}(x_{k+1}, y_k) > 0$, то находим $\lambda_k = \min\left\{1, \frac{\Delta^{(2)}(x_{k+1},y_k)}{\left\|p_{j_k}-y_k\right\|^2}\right\}$ и

$$y_{k+1} = y_k + \lambda_k(p_{j_k} - y_k).$$

- Если $\Delta^{(2)}(x_{k+1}, y_k) \leqslant 0$, то полагаем $y_{k+1} = y_k$.

При выполнении условий $\Delta^{(1)}(x_k, y_k) \leqslant 0$ и $\Delta^{(2)}(x_{k+1}, y_k) \leqslant 0$ план (x_k, y_k) будет оптимальным, т. к. в этих случаях $x_{k+1} = x_k$ и $y_{k+1} = y_k$. Иначе (x_{k+1}, y_{k+1}) — очередное приближение.

Описание рабочей схемы завершено.

З а м е ч а н и е 1. Основное отличие рабочей схемы от принципиальный состоит в том, что очередное приближение в рабочей схеме строится сразу по обеим переменным, при условии положительности соответствующих оценок.

8°. Примеры. В этом разделе приведены три примера, иллюстрирующие работу предложенных вычислительных схем алгоритма Козинца. Вычисления заканчивались, когда на очередном приближении (x_k, y_k) соответствующие им оценки удовлетворяли условиям:

$$\begin{cases} \Delta^{(1)}(x_k, y_k) < \varepsilon, \\ \Delta^{(2)}(x_{k+1}, y_k) < \varepsilon, \end{cases}$$

где $\varepsilon > 0$ — параметр метода.

В качестве начального приближения (x_0, y_0) брались центры масс множеств P_1 и P_2. Результаты расчетов представлены на графиках.

Используются следующие обозначения: точки множеств P_1 и P_2 изображаются красным и синим цветом; последовательность x_k изображается знаком $+$ красного цвета, y_k — знаком $\times$ синего цвета; предельная точка тем же знаком, но более крупного размера; квадратиком выделены центры масс множеств P_1 и P_2.

ПРИМЕР 1. На рис. 4–7 проиллюстрированы результаты расчетов по принципиальной и рабочей схемам при различных значениях параметра ε. Оптимальное решение выделено зеленым цветом.

При $\varepsilon = 0.1$ оба алгоритма закончили вычисления за 4 итерации. На рисунках 4 и 5 видно, что предельная точка x_4 находится в окрестности оптимальной, а y_4 не совпала с оптимальной только для рабочей схемы.

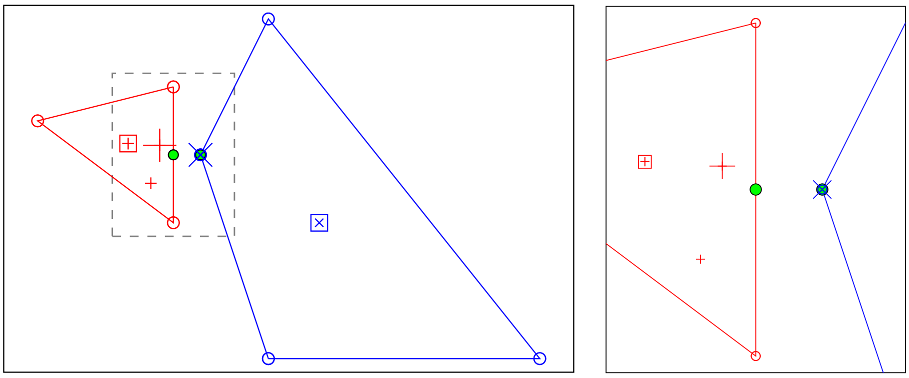

Рис. 4. Принципиальная схема ($\varepsilon = 0.1$)

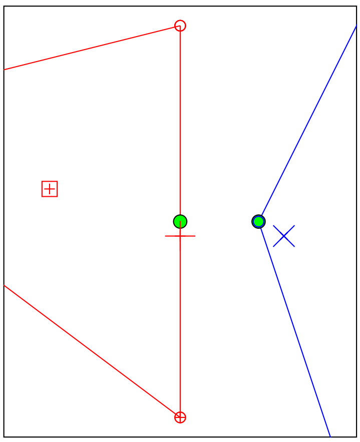

Рис. 5. Рабочая схема ($\varepsilon = 0.1$)

Результаты расчетов обеих схем при $\varepsilon = 0.01$ изображены на рис. 6. В принципиальной схема вычисления закончились за 22 итерации, в рабочей — за 6 итераций. Обе предельные точки совпали с оптимальными только для рабочей схемы.

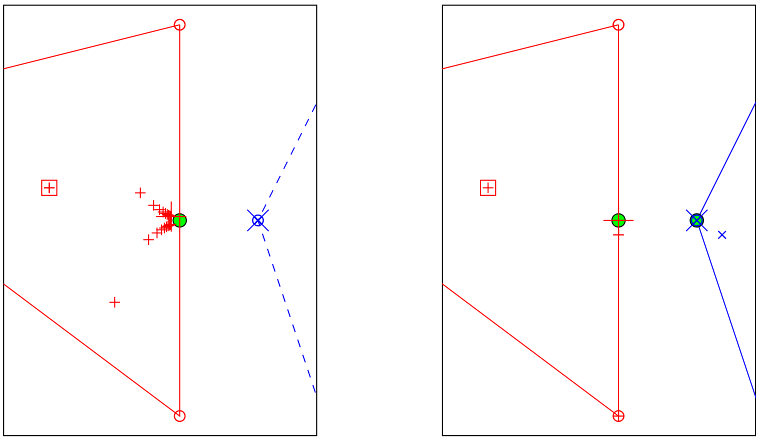

Рис. 6. Расчеты при $\varepsilon = 0.01$

Аналогичная картина сохранилась для рабочей схемы и при $\varepsilon = 0.001$ (см. рис. 7). Как и ранее, вычисления завершились в оптимальной точке за 6 итераций. Намного увеличилось количество итераций в принципиальной схеме — до 240 итераций. Это показывает, что в принципиальной схеме скорость сходимости замедляется при приближении к решению.

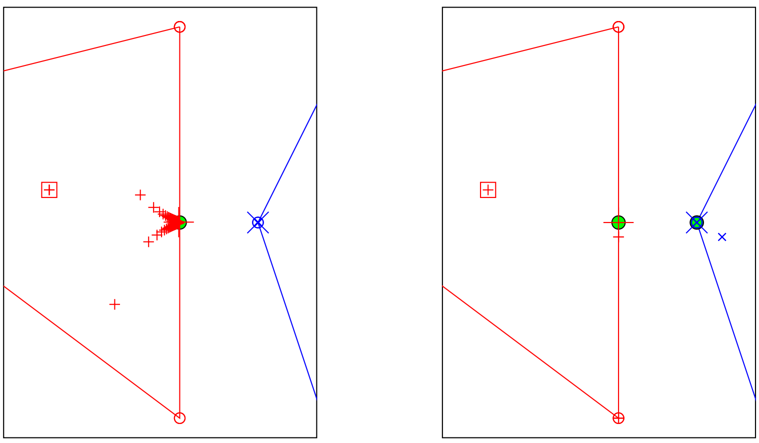

Рис. 7. Расчеты при $\varepsilon = 0.001$

ПРИМЕР 2. На рис. 8–9 приведены результаты расчетов при $\varepsilon = 0.01$ для множеств, у которых $|P_1| = 30$ и $|P_2| = 50$. К единственному решению задачи, последовательность (x_k, y_k), построенная по рабочей схеме, сошлась на порядок быстрее.

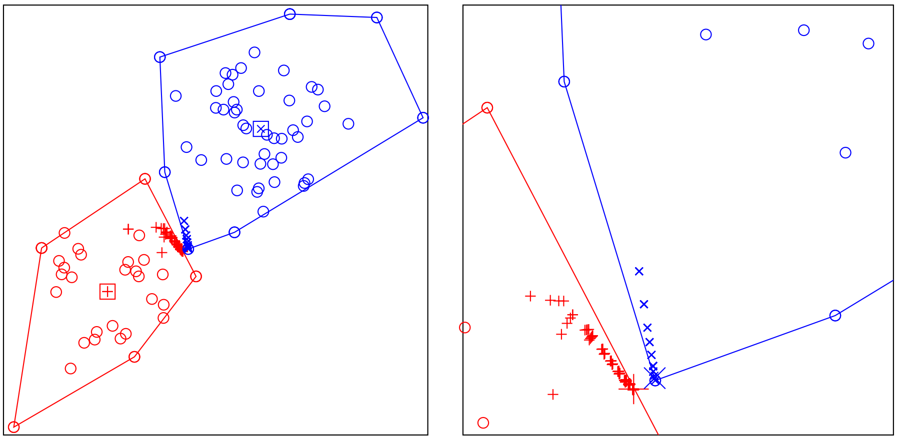

Рис. 8. Принципиальная схема (694 итер.)

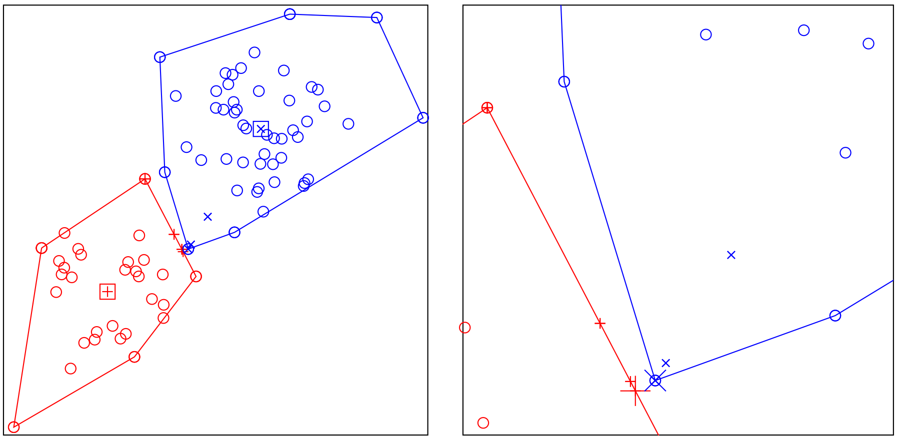

Рис. 9. Рабочая схема (8 итер.)

ПРИМЕР 3. На рис. 10–11 приведены результаты расчетов при $\varepsilon = 0.01$, когда выпуклые оболочки множеств P_1 и P_2 пересекаются. На рисунках видно, что принципиальная и рабочая схемы привели к различным решениям.

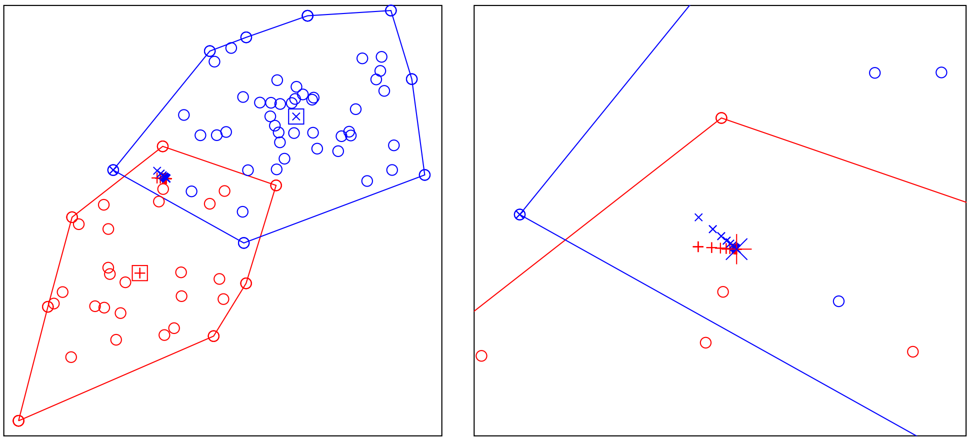

Рис. 10. Принципиальная схема (29 итер.)

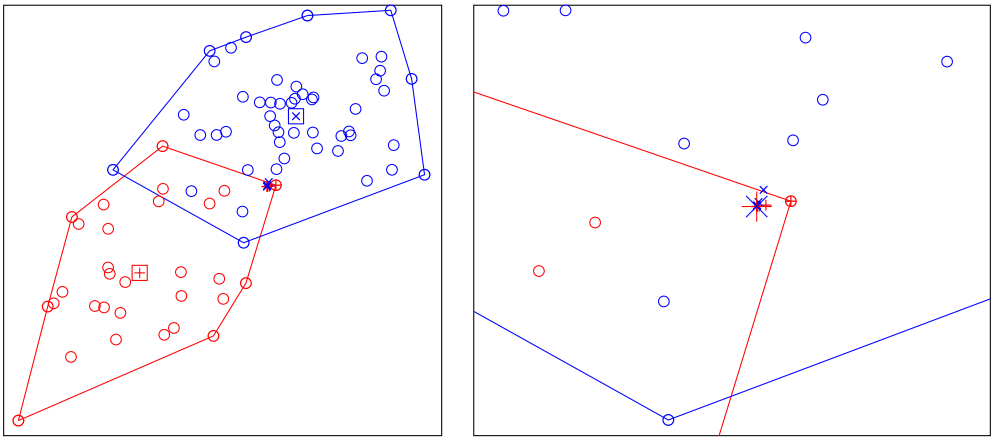

Рис. 11. Рабочая схема (12 итер.)

Замечание 2. Трудоемкость шага в принципиальной схеме соответствует двум шагам в рабочей схеме.

Замечание 3. Были также проведены массовые вычислительные эксперименты по сравнению эффективности рассматриваемых схем алгоритма Косинца. Замерялось время и вычислялось количество итераций каждой схемы на решение одинаковых задач. В множествах P_1 и P_2 размерность векторов доходила до 200, а их количество — до 3000. В среднем, принципиальной схемой задача решается в два раза дольше, чем рабочей схемой. Количество же итераций не намного отличаются друг от друга.

ЛИТЕРАТУРА

1. Малозёмов В. Н., Соловьёва Н. А., Тамасян Г. Ш. *Об алгоритме Козинца. I* // Семинар «OML». Избранные доклады. 14 марта 2024 г. (`http://oml.cmlaboratory.com/reps24.shtml#0314`)

2. Козинец Б. Н. *Об одном алгоритме обучения линейного персептрона.* В сб. «Вычислительная техника и вопросы программирования». Изд. ВЦ ЛГУ, 1964, вып. 3. С. 80–83.

3. Козинец Б. Н. *Рекуррентный алгоритм разделения выпуклых оболочек двух множеств* // Алгоритмы обучения распознаванию образов / Под. ред. В. Н. Вапника. М.: Сов. Радио. 1973. С. 43–50.