САНКТ–ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

УДК 519.7

Вздыхалкина Екатерина Константиновна

Наилучшее отделение двух множеств с помощью нескольких гиперплоскостей

Диссертация на соискание ученой степени
кандидата физико–математических наук
по специальности 01.01.09 — дискретная математика
и математическая кибернетика

Научный руководитель
доктор физ.–мат. наук, профессор
Л. Н. Полякова

Санкт–Петербург 2014

СОДЕРЖАНИЕ

Введение

1. Область прикладной математики, которая теперь называется *математической диагностикой* [35], активно развивается с 60-х годов прошлого столетия. Она включает широкий круг задач: распознавание образов, классификацию и кластеризацию данных, машинное обучение, медицинскую и техническую диагностику. Простейшей задачей такого рода является отделение двух множеств в конечномерном или бесконечномерном пространствах.

При решении задач математической диагностики используются статистические методы [2,53] , методы математического программирования (линейного [12,38,41,51], билинейного [30], выпуклого [52]) и методы глобальной оптимизации [28]. Значительный вклад в развитие этого направления внесли М. А. Айзерман, Э. М. Браверман, Л. И. Розоноэр [1], В. Н. Вапник [2, 53], А. Л. Горелик [4], Ю. И. Журавлёв [6], Ю. И. Неймарк [11], В. Н. Фомин [24], Я. З. Цыпкин [25], В. А. Якубович [26]. Особо следует отметить О. Л. Мангасаряна. По его работам 1965–2014 годов [в частности, 41-47, 29, 30, 37] можно проследить за всеми этапами развития математической диагностики.

При решении конкретных задач математической диагностики возникает необходимость в построении правила, в соответствии с которым судят о принадлежности данной точки тому или другому множеству. Это правило называют *диагностическим правилом*. Чаще всего его реализуют в виде *дискриминантной функции (функционала)*. С начала 90-х годов прошлого столетия в качестве дискриминантной активно используются недифференцируемые функции. С их помощью удаётся, в частности, улучшить результаты в следующем направлении: в случае, когда выпуклые оболочки двух отделяемых множеств пересекаются, более точно выделить «смешанную полосу», содержащую точки как одного, так и другого множества.

Для решения негладких задач математической диагностики привлекаются методы недифференцируемой оптимизации [35, 36, 27, 5, 7]. Примечательно, что при их реализации важную роль играет линейное программирование. Отметим также работы [31,33,34,39,48–50], которые обогатили арсенал методов математической диагностики.

Данное исследование проводится в русле работ Беннет-Мангасаряна [29] и Асторино-Гаудиозо [27]. Изучается задача наилучшего отделения выпуклой оболочки одного конечного множества от другого конечного множества с помощью нескольких гиперплоскостей. Используется недифференциремая дискриминантная функция с параметром. Параметр вводится для преодоления вычислительных трудностей, связанных с принципиальной неединственностью решения соответствующих экстремальных задач.

2. Диссертация состоит из шести параграфов и четырёх дополнений.

В §1 рассматривается задача наилучшего линейного отделения двух конечных множеств A и B из $\mathbb{R}^n$. Пусть

$$A = \{a_i\}_{i=1}^m \quad \text{и} \quad B = \{b_j\}_{j=1}^k. \tag{1}$$

Введём функцию [29]

$$f(g) = \frac{1}{m}\sum_{i=1}^m \big[\langle w, a_i\rangle - \gamma + c\big]_+ + \frac{1}{k}\sum_{j=1}^k \big[-\langle w, b_j\rangle + \gamma + c\big]_+,$$

где $g = (w, \gamma)$, $c > 0$ — параметр (в [29] вместо c стоит единица), $[u]_+ = \max\{0, u\}$ — плюсиковая функция. Справедливо следующее утверждение.

Теорема 1.1. *Множества A и B строго линейно отделимы тогда и только тогда, когда существует вектор g_*, на котором $f(g_*) = 0$.*

Учитывая, что $f(g) \geq 0$ при всех g, заключаем, что задача строгого линейного отделения множества A от множества B сводится к экстремальной задаче

$$f(g) \to \min_{g \in \mathbb{R}^{n+1}} . \tag{2}$$

В свою очередь задача (2) эквивалентна задаче линейного программирования

$$\frac{1}{m}\sum_{i=1}^{m} y_i + \frac{1}{k}\sum_{j=1}^{k} z_j \to \min, \tag{3}$$

$$-\langle w, a_i \rangle + \gamma + y_i \geq c, \quad i \in 1:m;$$

$$\langle w, b_j \rangle - \gamma + z_j \geq c, \quad j \in 1:k;$$

$$y_i \geq 0,\ i \in 1:m; \quad z_j \geq 0,\ j \in 1:k.$$

Задача (3) всегда имеет решение. Обозначим его $\big(w_*, \gamma_*, \{y_i^*\}, \{z_j^*\}\big)$ и пусть μ — минимальное значение целевой функции. Если $\mu = 0$, то гиперплоскость H_*, определяемая уравнением $\langle w_*, x \rangle = \gamma_*$, строго отделяет множество A от множества B. При этом можно вычислить ширину отделяющей полосы.

При $\mu > 0$ множества A и B не допускают строгого линейного отделения. В этом случае будем говорить, что указанная ранее гиперплоскость H_*, является наилучшей гиперплоскостью, приближённо отделяющей множество A от множества B. Можно вычислить ширину «смешанной полосы», содержащей точки обоих множеств.

При $\mu > 0$ имеется тонкость: может оказаться, что $w_* = \mathbb{O}$. Доказывается, что в этом случае у задачи (3) существует другое решение с $w_* \neq \mathbb{O}$.

Оно строится с помощью вспомогательной задачи линейного программирования.

Приводятся примеры строгого и наилучшего приближённого линейного отделения. Выясняется, что аддитивный параметр c играет роль нормирующего множителя.

§1 написан на основе работ [8, 10, 40].

В §2 переходим к основной задаче — отделению выпуклой оболочки одного конечного множества от другого конечного множества с помощью h гиперплоскостей (задача h-отделения).

Пусть множества A и B имеют вид (1). Обозначим $\operatorname{co}(A)$ выпуклую оболочку множества A.

Введём функцию от матрицы [27]

$$F(G) = \frac{1}{m}\sum_{i=1}^{m} \max_{s\in 1:h}\big[\langle w^s,\ a_i\rangle - \gamma_s + c\big]_+ + \frac{1}{k}\sum_{j=1}^{k} \min_{s\in 1:h}\big[-\langle w^s,\ b_j\rangle + \gamma_s + c\big]_+.$$

Здесь G — матрица размера $h \times (n+1)$ со строками

$$g^s = (w^s,\ \gamma_s), \quad s \in 1 : h,$$

$c > 0$ — параметр (в [27] вместо c стоит единица). Матрицу G указанных размеров будем называть *подходящей*, если у неё все w^s ненулевые.

Справедливо следующее утверждение.

Теорема 2.1. *Выпуклая оболочка множества A и множество B строго h-отделимы тогда и только тогда, когда существует подходящая матрица G_*, на которой $F(G_*) = 0$.*

Учитывая, что $F(G) \geq 0$ при всех G, заключаем, что задача строгого h-отделения сводится к экстремальной задаче

$$F(G) \to \min_{G\in\mathbb{R}^{h,n+1}} . \tag{4}$$

При минимизации функции $F(G)$ наибольшую трудность доставляет второе слагаемое (сумма минимумов). Оно делает функцию $F(G)$ невыпуклой (к тому, что она негладкая).

В §3 показывается, что задача (4) с помощью «леммы о сумме минимумов» сводится к конечному числу задач линейного программирования. Для этого вводятся индексные цепочки $S = (s_1,\ s_2, \ldots,\ s_k)$, где $s_j \in 1:h$ при каждом $j \in 1:k$. Каждой такой цепочке S сопоставляется экстремальная задача

$$\frac{1}{m}\sum_{i=1}^{m}\max_{s\in 1:h}\big[\langle w^s,\ a_i\rangle - \gamma_s + c\big]_+ + \frac{1}{k}\sum_{j=1}^{k}\big[-\langle w^{s_j},\ b_j\rangle + \gamma_{s_j} + c\big]_+ \to \min_G,$$

эквивалентная задаче линейного программирования

$$\frac{1}{m}\sum_{i=1}^{m} p_i + \frac{1}{k}\sum_{j=1}^{k} q_j \to \min, \tag{5}$$

$$-\langle a_i,\ w^s\rangle + \gamma_s + p_i \ge c, \quad i \in 1:m,\ s \in 1:h;$$

$$\langle b_j,\ w^{s_j}\rangle - \gamma_{s_j} + q_j \ge c, \quad j \in 1:k;$$

$$p_i \ge 0, \quad i \in 1:m; \quad q_j \ge 0, \quad j \in 1:k.$$

Задача (5) имеет решение при всех S. Выделим цепочку S_*, на которой минимальное значение целевой функции в задаче (5) принимает наименьшее значение. Обозначим $\big(\{w_*^s\},\ \{\gamma_s^*\},\ \{p_i^*\},\ \{q_j^*\}\big)$ соответствующее решение задачи (5). В §3 доказывается (теорема 3.1), что матрица G_*, составленная из строк (w_*^s, γ_s^*), $s \in 1:h$, является решением задачи (4).

Возможны такие случаи.

1) $F(G_*) = 0$ и G_* — подходящая матрица. Тогда система гиперплоскостей H_*^s, определяемых уравнениями $\langle w_*^s, x\rangle = \gamma_s^*$, $s \in 1:h$, строго отделяет $\mathrm{co}(A)$ от B.

2) $F(G_*) = 0$, но матрица G_* — неподходящая. В этом случае справедливо

утверждение (теорема 2.2): не все w_*^s нулевые; если обозначить через J множество индексов ненулевых w_*^s, то множества $\mathrm{co}(A)$ и B строго $\bigl(h-|J|\bigr)$-отделимы.

3) $F(G_*) > 0$ и G_* — подходящая матрица. По аналогии со случаем $h = 1$ будем говорить, что система гиперплоскостей H_*^s, $s \in 1:h$, осуществляет наилучшее приближённое h-отделение множества $\mathrm{co}(A)$ от B.

В §3 приведён пример строгого 2-отделения.

§2 и §3 написаны на основе работ [19, 21, 22].

3. Итак, в §3 установлен принципиальный факт: задача h-отделения сводится к конечному числу задач линейного программирования. Однако количество таких задач линейного программирования может быть большим. Это побуждает обратиться к итерационным методам, которые позволяют получить *приближённое* решение задачи h-отделения с требуемой точностью.

Из общих соображений следует, что функция от матрицы $F(G)$ дифференцируема по направлениям (в качестве которых также выступают матрицы). На этой основе в §4 строится метод «градиентного типа» для приближённого решения задачи h-отделения (4). Опишем его принципальную схему.

Начнём с производной по направлению. По определению

$$F'(G,\ V) = \lim_{t\to+0} \frac{F(G+tV)-F(G)}{t}.$$

Для того чтобы записать для $F'(G,\ V)$ явную формулу, введём обозначения

$$f(v,\ u) = \langle v,\ u\rangle + c,$$

$$\hat{a}_i = \begin{pmatrix} a_i \\ -1 \end{pmatrix}, \quad \check{b}_j = \begin{pmatrix} -b_j \\ 1 \end{pmatrix},$$

$$\hat{\varphi}_i(G) = \max_{s\in 1:h} f(g^s,\ \hat{a}_i), \quad \check{\psi}_j(G) = \min_{s\in 1:h} f(g^s,\ \check{b}_j),$$

$$\varphi_i(G) = \big[\hat{\varphi}_i(G)\big]_+, \quad \psi_j(G) = \big[\check{\psi}_j(G)\big]_+.$$

Тогда

$$F(G) = \frac{1}{m}\sum_{i=1}^{m} \varphi_i(G) + \frac{1}{k}\sum_{j=1}^{k} \psi_j(G).$$

Положим далее

$$\hat{R}_i(G) = \big\{s \in 1:h \mid f(g^s,\ \hat{a}_i) = \hat{\varphi}_i(G)\big\},$$

$$\check{R}_j(G) = \big\{s \in 1:h \mid f(g^s,\ \check{b}_j) = \check{\psi}_j(G)\big\}.$$

Теорема 4.1. *Справедлива формула*

$$F'(G,\ V) = \frac{1}{m}\sum_{i=1}^{m} \varphi_i'(G,\ V) + \frac{1}{k}\sum_{j=1}^{k} \psi_j'(G,\ V),$$

где

$$\varphi_i'(G,\ V) = \begin{cases} \max\limits_{s\in \hat{R}_i(G)} \langle v^s,\ \hat{a}_i\rangle, & \text{если } \hat{\varphi}_i(G) > 0, \\ \max\limits_{s\in \hat{R}_i(G)} \big[\langle v^s,\ \hat{a}_i\rangle\big]_+, & \text{если } \hat{\varphi}_i(G) = 0, \\ 0, & \text{если } \hat{\varphi}_i(G) < 0; \end{cases}$$

$$\psi_j'(G,\ V) = \begin{cases} \min\limits_{s\in \check{R}_j(G)} \langle v^s,\ \check{b}_j\rangle, & \text{если } \check{\psi}_j(G) > 0, \\ \min\limits_{s\in \check{R}_j(G)} \big[\langle v^s,\ \check{b}_j\rangle\big]_+, & \text{если } \check{\psi}_j(G) = 0, \\ 0, & \text{если } \check{\psi}_j(G) < 0. \end{cases}$$

В Дополнении C приводится доказательство этой теоремы.

Переходим к описанию одного шага численного метода минимизации функции $F(G)$. Пусть имеется ν-е приближение G_ν. Для определения направления спуска из точки G_ν решаем вспомогательную экстремальную задачу

$$F'(G,\ V) \to \min, \tag{6}$$

где минимум берётся по матрицам V размера $h \times (n+1)$, все элементы которых ограничены по модулю некоторой константой $K > 0$. Задача (6) сводится к небольшому числу задач линейного программирования. Она имеет решение V_ν. Если $F'(G_\nu,\ V_\nu) \geq 0$, то G_ν — стационарная точка функции $F(G)$. Вычисления прекращаются. В противном случае матрица V_ν является направлением убывания функции $F(G)$ из точки G_ν. Находим шаг $t_\nu > 0$ в направлении V_ν, обеспечивающий гарантированное уменьшение функции $F(G)$. Полагаем $G_{\nu+1} = G_\nu + t_\nu V_\nu$, после чего вычисления повторяются. Описание принципиальной схемы метода «градиентного типа» для решения задачи (4) завершено.

В §4 приводится пример на применение данного метода к решению задачи 3-отделения. Основное внимание уделяется организации вычислений.

§4 написан на основе работ [16, 23].

4. В §5 предлагается ещё один численный метод решения задачи (4), максимально усиливающий её специфику. Опишем общий шаг этого метода.

Фиксируем положительные параметры точности ε_A, ε_B и σ.

Пусть имеется ν-е приближение — матрица G_ν со строками g^s_ν, $s \in 1 : h$. Для получения очередного приближения $G_{\nu+1}$ выполняем следующие операции:

1) Вычисляем $F(G_\nu)$.

2) Формируем индексные множества

$$
\begin{aligned}
I_\nu &= \bigl\{ i \in 1:m \mid \max_{s \in 1:h} f(g_\nu^s,\ \hat{a}_i) > -\varepsilon_A \bigr\},\\
J_\nu &= \bigl\{ j \in 1:k \mid \min_{s \in 1:h} f(g_\nu^s,\ \check{b}_j) > -\varepsilon_B \bigr\},\\
L_j^\nu &= \bigl\{ s \in 1:h \mid f(g_\nu^s,\ \check{b}_j) = \min_{p \in 1:h} f(g_\nu^p,\ \check{b}_j) \bigr\}, \quad j \in J_\nu.
\end{aligned}
$$

3) Перебираем индексные цепочки $S = \{s_j\}_{j \in J_\nu}$, $s_j \in L_j^\nu$, пока не найдётся цепочка $\hat{S} = \{\hat{s}_j\}_{j \in J_\nu}$ со следующим свойством: решение $\hat{V}_\nu$ экстремальной задачи

$$
\hat{Q}(G_\nu + V) := \frac{1}{m} \sum_{i \in I_\nu} \varphi_i(G_\nu + V) + \frac{1}{k} \sum_{j \in J_\nu} \bigl[f(g_\nu^{\hat{s}_j} + v^{\hat{s}_j},\ \check{b}_j) \bigr]_+ \to \min,
$$

где минимум берётся по матрицам V, все элементы которых ограничены по модулю некоторой константой $K > 0$, удовлетворяет неравенству

$$
\hat{Q}(G_\nu + \hat{V}_\nu) < F(G_\nu) - \sigma. \tag{7}
$$

4) Полагаем $G_{\nu+1} = G_\nu + \hat{t}_\nu \hat{V}_\nu$, где $\hat{t}_\nu$ — точка минимума функции $F(G_\nu + t\hat{V}_\nu)$ на отрезке $[0,\ 1]$.

Если цепочка $\hat{S}$, обеспечивающая неравенство (7), отсутствует, то G_ν — почти локально оптимальная матрица. В этом случае вычисления прекращаются.

Доказывается, что описанный процесс конечен.

§5 написан на основе работы [9].

5. В §6 на примере задачи 4-отделения проводятся широкие эксперименты по анализу численных методов, предложенных в §4 и §5. Обсуждается общая для обоих методов проблема выбора начального приближения.

6. В диссертации имеются четыре Дополнения.

Дополнение A посвящено линейному программированию. Приводится теорема существования решения и две теоремы двойственности. Делается замечание о практическом решении задач линейного программирования в среде MATLAB.

В Дополнении B собраны многочисленные свойства плюсиковой функции.

В Дополнении C выводится формула для производной по направлению от функции матрицы $F(G)$.

В Дополнении D доказывается лемма о сумме минимумов.

Все результаты из Дополнений активно используются в основном тексте.

7. На защиту выносятся следующие результаты:

- анализ задачи наилучшего линейного отделения как негладкой параметрической задачи;
- анализ задачи наилучшего h-отделения как негладкой и невыпуклой параметрической задачи;
- разработка метода «градиентного типа» для решения задачи h-отделения;
- разработка безградиентного метода h-отделения, максимально учитывающего специфику данной задачи.

8. Отдельные результаты диссертации докладывались на приводимых ниже конференциях и семинарах:

- на 40-й, 41-й, 42-й и 44-й Международных научных конференциях аспирантов и студентов «Процессы управления и устойчивость», Санкт-Петербург [17–20];

- на Международной конференции «Конструктивный негладкий анализ и смежные вопросы», Санкт-Петербург, 2012 [32];

- на Всероссийской молодёжной научной школе-семинаре «Дискретные модели и методы принятия решений», Новосибирск, 2013 (диплом за лучший доклад);

- на Санкт-Петербургском городском семинаре «Дискретный гармонический анализ и геометрическое моделирование» [8–10, 13, 14, 21–23];

- на семинарах кафедры математической теории моделирования систем управления факультета ПМ-ПУ (зав. кафедрой проф. В. Ф. Демьянов).

Основные результаты опубликованы в работах [15, 16, 40], в изданиях из списка ВАК.

Автор приносит глубокую благодарность своему научному руководителю профессору Людмиле Николаевне Поляковой и профессору Василию Николаевичу Малозёмову за постоянное внимание к моей работе и неоценимую помощь.

§ 1. Наилучшее линейное отделение двух множеств

На линейном уровне исследуется задача наилучшего приближённого отделения двух конечных множеств. Эта задача сводится к задаче негладкой оптимизации, при анализе которой используется вся мощь теории линейного программирования.

В идейном плане мы следуем работе [29].

1.1. Пусть в пространстве $\mathbb{R}^n$ заданы два конечных множества

$$A = \{a_i\}_{i=1}^m \quad \text{и} \quad B = \{b_j\}_{j=1}^k.$$

Множества A и B называются *строго отделимыми*, если существуют ненулевой вектор $w \in \mathbb{R}^n$ и вещественное число γ, такие, что

$$\langle w, a_i \rangle < \gamma \quad \text{при всех } i \in 1:m, \tag{1.1}$$

$$\langle w, b_j \rangle > \gamma \quad \text{при всех } j \in 1:k. \tag{1.2}$$

При выполнении условий (1.1) и (1.2) говорят также, что гиперплоскость H, определяемая уравнением $\langle w, x \rangle = \gamma$, *строго отделяет* множество A от множества B.

1.2. Введём функцию

$$f(g) = \frac{1}{m} \sum_{i=1}^m \big[\langle w, a_i \rangle - \gamma + c\big]_+ + \frac{1}{k} \sum_{j=1}^k \big[-\langle w, b_j \rangle + \gamma + c\big]_+, \tag{1.3}$$

где $g = (w, \gamma)$, $g \in \mathbb{R}^{n+1}$, $c > 0$ — параметр и $[u]_+ = \max\{0, u\}$. Ясно, что $f(g) \geq 0$ при всех g.

Теорема 1.1. *Множества A и B строго отделимы тогда и только тогда, когда существует вектор g_*, на котором $f(g_*) = 0$.*

Доказательство. Пусть $f(g_*) = 0$ на некотором векторе $g_* = (w_*, \gamma_*)$.

Покажем прежде всего, что $w_* \neq \mathbb{O}$. В противном случае

$$f(g_*) = (-\gamma_* + c)_+ + (\gamma_* + c)_+ = \begin{cases} -\gamma_* + c & \text{при } \gamma_* \leq -c, \\ 2c & \text{при } \gamma_* \in [-c, c], \\ \gamma_* + c & \text{при } \gamma_* \geq c. \end{cases}$$

Отсюда следует, что $f(g_*) \geq 2c$. Это противоречит условию $f(g_*) = 0$.

Далее, условие $f(g_*) = 0$ гарантирует, что все слагаемые

$$\big[\langle w_*, a_i\rangle - \gamma_* + c\big]_+ \quad \text{и} \quad \big[-\langle w_*, b_j\rangle + \gamma_* + c\big]_+$$

равны нулю. Это возможно лишь тогда, когда

$$\langle w_*, a_i\rangle - \gamma_* + c \leq 0 \quad \text{при всех } i \in 1:m, \tag{1.4}$$

$$-\langle w_*, b_j\rangle + \gamma_* + c \leq 0 \quad \text{при всех } j \in 1:k. \tag{1.5}$$

Остаётся отметить, что неравенства (1.4) и (1.5) обеспечивают выполнение условий строгой отделимости (1.1) и (1.2) с $w = w_*$, $\gamma = \gamma_*$.

Докажем обратное утверждение. Пусть выполнены условия (1.1) и (1.2). Обозначим

$$d := \min_{j\in 1:k}\langle w, b_j\rangle - \max_{i\in 1:m}\langle w, a_i\rangle > 0, \tag{1.6}$$

$$w_* = (\tfrac{2c}{d})w, \quad \gamma_* = \tfrac{1}{2}\big[\min_{j\in 1:k}\langle w_*, b_j\rangle + \max_{i\in 1:m}\langle w_*, a_i\rangle\big].$$

Согласно (1.6) и определению w_*

$$\min_{j\in 1:k}\langle w_*, b_j\rangle - \max_{i\in 1:m}\langle w_*, a_i\rangle = 2c.$$

Имеем

$$\max_{i\in 1:m}\langle w_*, a_i\rangle = 2\gamma_* - \min_{j\in 1:k}\langle w_*, b_j\rangle = 2\gamma_* - 2c - \max_{i\in 1:m}\langle w_*, a_i\rangle,$$

так что

$$\max_{i\in 1:m}\langle w_*, a_i\rangle = \gamma_* - c. \tag{1.7}$$

Аналогично

$$\min_{j\in 1:k}\langle w_*, b_j\rangle = 2\gamma_* - \max_{i\in 1:m}\langle w_*, a_i\rangle = 2\gamma_* + 2c - \min_{j\in 1:k}\langle w_*, b_j\rangle,$$

так что

$$\min_{j\in 1:k}\langle w_*, b_j\rangle = \gamma_* + c. \tag{1.8}$$

Положим $g_* = (w_*, \gamma_*)$. На основании (1.7) и (1.8) получим $f(g_*) = 0$.

Теорема доказана. $\square$

1.3. При доказательстве теоремы 1 описано преобразование вектора $g = (w, \gamma)$, порождающего строго отделяющую гиперплоскость $H = \{x \mid \langle w, x\rangle = \gamma\}$, в вектор $g_* = (w_*, \gamma_*)$, на котором $f(g_*) = 0$. Дело в том, что на само́м векторе g значение $f(g)$ может быть положительным (это зависит от параметра c).

Пример 1.1. В качестве A и B возьмём одноточечные множества на плоскости $\mathbb{R}^2$: $A = \{a\}$, $B = \{b\}$, где $a = (0,0)$ и $b = (0,2)$. Вектор $g = (w, \gamma)$ с компонентами $w = (0,1)$ и γ из интервала $(0,2)$ порождает прямую $x_2 = \gamma$, строго отделяющую точку a от точки b (см. рис. 1.1).

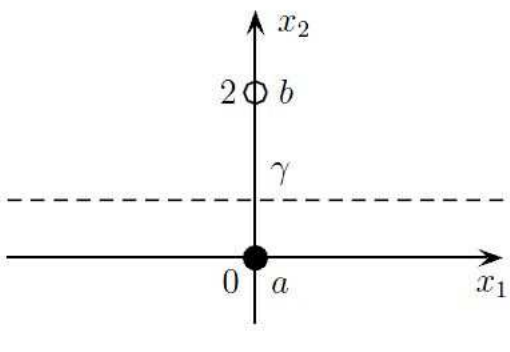

Рис. 1.1. Строгое отделение двух точек

Вместе с тем,

$$f(g) = [-\gamma + c]_+ + [-2 + \gamma + c]_+.$$

На рис. 1.2 представлен график $f(g)$ как функции от γ при $c \in (0, 1]$.

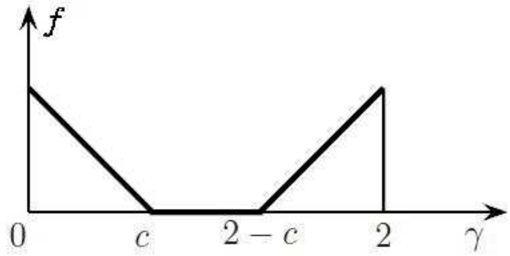

Рис. 1.2. График функции $f(g)$

Видим, что $f(g) = 0$ при $\gamma \in [c, 2-c]$. При $\gamma \in (0, c) \cup (2-c, 2)$ прямая $x_2 = \gamma$ по-прежнему строго отделяет точку a от точки b, но $f(g) > 0$.

1.4. Рассмотрим экстремальную задачу

$$f(g) \to \min, \tag{1.9}$$

где $f(g)$ — функция вида (1.3). Эта задача эквивалентна задаче линейного программирования

$$\frac{1}{m}\sum_{i=1}^{m} y_i + \frac{1}{k}\sum_{j=1}^{k} z_j \to \min, \tag{1.10}$$

$$-\langle w, a_i\rangle + \gamma + y_i \geq c, \quad i \in 1:m;$$

$$\langle w, b_j\rangle - \gamma + z_j \geq c, \quad j \in 1:k;$$

$$y_i \geq 0,\ i \in 1:m; \quad z_j \geq 0,\ j \in 1:k.$$

Множество планов задачи (1.10) непусто (например, планом является вектор с компонентами $w = \mathbb{O}$, $\gamma = 0$, $y_i \equiv c$, $z_j \equiv c$) и целевая функция ограничена снизу нулём. Значит, задача (1.10) имеет решение. По эквивалентности существует решение и у задачи (1.9), причём минимальные значения целевых функций у этих задач равны между собой. Это общее значение обозначим через μ. Отметим также, что если $\big(w_*, \gamma_*, \{y_i^*\}, \{z_j^*\}\big)$ — решение задачи (1.10), то $g_* = (w_*, \gamma_*)$ — решение задачи (1.9).

1.5. При $\mu = 0$ получим $f(g_*) = 0$. По теореме 1 вектор $g_* = (w_*, \gamma_*)$ порождает гиперплоскость $H = \bigl\{x \in \mathbb{R}^n \mid \langle w_*, x\rangle = \gamma_*\bigr\}$, строго отделяющую множество A от множества B.

Вектор g_* можно привести к каноническому виду. Положим

$$w_0 = w_*/\|w_*\|,$$

$$\gamma_0 = \tfrac{1}{2}\bigl[\min_{j\in 1:k}\langle w_0, b_j\rangle + \max_{i\in 1:m}\langle w_0, a_i\rangle\bigr],$$

$$c_0 = \tfrac{1}{2}\bigl[\min_{j\in 1:k}\langle w_0, b_j\rangle - \max_{i\in 1:m}\langle w_0, a_i\rangle\bigr],$$

$$g_0 = (w_0, \gamma_0).$$

Тогда при всех $i \in 1:m$

$$\langle w_0, a_i\rangle - \gamma_0 + c_0 = \langle w_0, a_i\rangle - \max_{i\in 1:m}\langle w_0, a_i\rangle \leq 0,$$

и при всех $j \in 1:k$

$$-\langle w_0, b_j\rangle + \gamma_0 + c_0 = -\langle w_0, b_j\rangle + \min_{j\in 1:k}\langle w_0, b_j\rangle \leq 0.$$

Значит, $f(g_0) = 0$ при $c = c_0$. Гиперплоскость $H_0 = \bigl\{x \mid \langle w_0, x\rangle = \gamma_0\bigr\}$ строго отделяет множество A от множества B, причём ширина отделяющей полосы равна $2c_0$.

На рис. 1.3 приведён пример строгого отделения.

1.6. Как отмечалось в п. 1.4, задача (1.9) всегда имеет решение. При $\mu > 0$ по теореме 1.1 множества A и B не допускают строгого линейного отделения. В этом случае будем говорить, что гиперплоскость $H_* = \bigl\{x \mid \langle w_*, x\rangle = \gamma_*\bigr\}$, порождаемая решением $g_* = (w_*, \gamma_*)$ задачи (1.9), является *наилучшей гиперплоскостью, приближённо отделяющей множество A от множества B* (при данном значении параметра c).

Здесь, однако, имеется тонкость: нет гарантии, что компонента w_* вектора g_* отлична от нулевой. Разберёмся в этой ситуации.

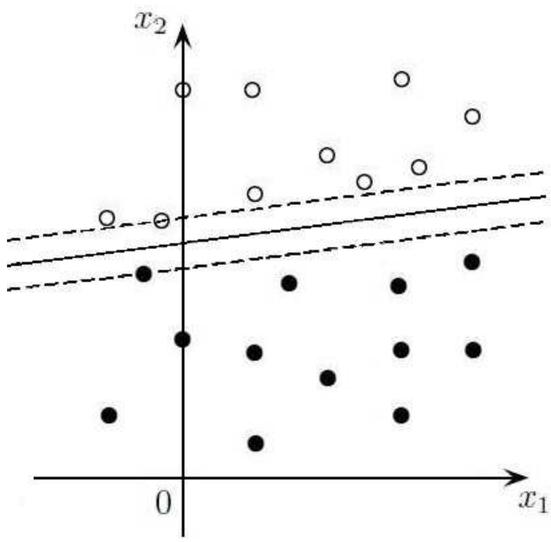

Рис. 1.3. Пример строгого отделения двух множеств

Теорема 1.2. *Для того чтобы задача* (1.9) *имела решение* $g_* = (w_*, \gamma_*)$ *с* $w_* = \mathbb{O}$, *необходимо и достаточно, чтобы выполнялось условие*

$$\frac{1}{m}\sum_{i=1}^{m} a_i = \frac{1}{k}\sum_{j=1}^{k} b_j. \tag{1.11}$$

Д о к а з а т е л ь с т в о. Н е о б х о д и м о с т ь. При $w_* = \mathbb{O}$ легко вычисляется экстремальное значение целевой функции у задачи линейного программирования (1.10). Действительно,

$$\mu = f(g_*) = \min_{\gamma}\big\{[-\gamma + c]_+ + [\gamma + c]_+\big\} = 2c.$$

Такое же экстремальное значение имеет задача линейного программирования, двойственная к задаче (1.10). В силу разрешимости двойственной задачи совместна система

$$c\Bigl(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j\Bigr) = 2c, \tag{1.12}$$

$$-\sum_{i=1}^{m} u_i a_i + \sum_{j=1}^{k} v_j b_j = \mathbb{O}, \tag{1.13}$$

$$\sum_{i=1}^{m} u_i - \sum_{j=1}^{k} v_j = 0, \tag{1.14}$$

$$0 \le u_i \le \tfrac{1}{m},\ i \in 1:m; \quad 0 \le v_j \le \tfrac{1}{k},\ j \in 1:k. \tag{1.15}$$

Из (1.12) и (1.14) следует, что

$$\sum_{i=1}^{m} u_i = 1, \quad \sum_{j=1}^{k} v_j = 1.$$

Принимая во внимание (1.15), заключаем, что все u_i равны $\frac{1}{m}$ и все v_j равны $\frac{1}{k}$. Теперь формула (1.13) становится эквивалентной равенству (1.11).

Достаточность. Запишем задачу, двойственную к (1.10):

$$c\Bigl(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j\Bigr) \to \max$$

при ограничениях (1.13)–(1.15). В силу (1.11) набор $u_i \equiv \frac{1}{m}$, $v_j \equiv \frac{1}{k}$ является планом этой задачи. Значение целевой функции на нём равно $2c$.

Вместе с тем, на плане

$$w = \mathbb{O}, \quad \gamma = 0, \quad y_i \equiv c, \quad z_j \equiv c \tag{1.16}$$

задачи (1.10) значение целевой функции также равно $2c$. Отсюда следует, что план (1.16) задачи (1.10) с $w = \mathbb{O}$ является оптимальным.

Теорема доказана. □

Пример 1.2. Рассмотрим на плоскости $\mathbb{R}^2$ два двухточечных множества

$$A = \bigl\{(0,0), (1,1)\bigr\}, \quad B = \bigl\{(1,0), (0,1)\bigr\}$$

(см. рис. 1.4). В данном случае выполняется условие (1.11). По теореме 2 задача (1.9) имеет решение $g_* = (w_*, \gamma_*)$ с $w_* = \mathbb{O}$. При этом $\mu = 2c$.

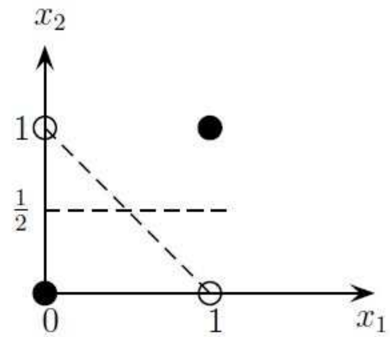

Рис. 1.4. Множества A и B из примера 1.2

Покажем, что у задачи (1.9) существует другое решение $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$.

Согласно (1.3)

$$f(g) = \tfrac{1}{2}\big\{[-\gamma+c]_+ + [w^1+w^2-\gamma+c]_+\big\} + \tfrac{1}{2}\big\{[-w^1+\gamma+c]_+ + [-w^2+\gamma+c]_+\big\}.$$

Здесь $w = (w^1, w^2)$. Положим

$$w_0 = (c, c), \quad \gamma_0 = c, \quad g_0 = (w_0, \gamma_0).$$

Тогда $f(g_0) = 2c$. Значит, на векторе g_0 достигается минимум функции $f(g)$. Гиперплоскость $H_0 = \{x \mid x_1 + x_2 = 1\}$ является наилучшей гиперплоскостью, приближённо отделяющей множество A от множества B.

Таким же свойством обладают вектор $g_1 = (w_1, \gamma_1)$ с $w_1 = (0, c)$, $\gamma_1 = \frac{c}{2}$ и гиперплоскость $H_1 = \{x \mid x_2 = \frac{1}{2}\}$ (см. рис. 1.4).

1.7. Особенность,отмеченная в примере 1.2, имеет общий характер.

Теорема 1.3. *При* $\mu > 0$ *у задачи* (1.9) *существует решение* $g_0 = (w_0, \gamma_0)$ *с* $w_0 \neq \mathbb{O}$.

Д о к а з а т е л ь с т в о. Допустим, что у решения $g_* = (w_*, \gamma_*)$ задачи (1.9) компонента w_* оказалась нулевой. Построим другое решение $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$.

По теореме 1.2 выполняется соотношение (1.11) и $\mu = 2c$. Возьмём произвольный ненулевой вектор $p \in \mathbb{R}^n$ и рассмотрим задачу линейного программирования

$$\langle p, w\rangle \to \min, \tag{1.17}$$
$$-\frac{1}{m}\sum_{i=1}^{m} y_i - \frac{1}{k}\sum_{j=1}^{k} z_j = -2c;$$
$$-\langle w, a_i\rangle + \gamma + y_i \geq c, \quad i \in 1:m;$$
$$\langle w, b_j\rangle - \gamma + z_j \geq c, \quad j \in 1:k;$$
$$y_i \geq 0,\ i \in 1:m; \quad z_j \geq 0,\ j \in 1:k.$$

Вектор (1.16) удовлетворяет ограничениям задачи (1.17), то есть является её планом. Покажем, что этот план не может быть оптимальным.

В случае оптимальности плана (1.16) у задачи, двойственной к (1.17), должен существовать план с таким же (нулевым) значением целевой функции. Таким образом, должна быть совместной система

$$c\left(\sum_{i=1}^{m} u_i + \sum_{j=1}^{k} v_j - 2\zeta\right) = 0, \tag{1.18}$$
$$-\sum_{i=1}^{m} u_i a_i + \sum_{j=1}^{k} v_j b_j = p, \tag{1.19}$$
$$\sum_{i=1}^{m} u_i - \sum_{j=1}^{k} v_j = 0, \tag{1.20}$$
$$0 \leq u_i \leq \tfrac{1}{m}\zeta,\ i \in 1:m; \quad 0 \leq v_j \leq \tfrac{1}{k}\zeta,\ j \in 1:k. \tag{1.21}$$

Покажем, однако, что эта система несовместна.

Из (1.18) и (1.20) следует, что

$$\sum_{i=1}^{m} u_i = \zeta, \quad \sum_{j=1}^{k} v_j = \zeta.$$

В силу (1.21) получаем $u_i \equiv \frac{1}{m}\zeta$, $v_j \equiv \frac{1}{k}\zeta$. Равенство (1.19) принимает вид

$$\zeta\biggl(-\frac{1}{m}\sum_{i=1}^{m} a_i + \frac{1}{k}\sum_{j=1}^{k} b_j\biggr) = p.$$

Но это противоречит условию (1.11) (напомним, что $p \neq \mathbb{O}$).

Установлено, что план (1.16) задачи (1.17) с нулевым значением целевой функции не является оптимальным. Значит, существует план

$$\bigl(w_0,\ \gamma_0,\ \{u_i^0\},\ \{v_j^0\}\bigr) \tag{1.22}$$

с отрицательным значением целевой функции. У такого плана должно быть $w_0 \neq \mathbb{O}$.

Теперь отметим, что план (1.22) задачи (1.17) удовлетворяет ограничениям задачи (1.10) и на нём целевая функция задачи (1.10) принимает наименьшее возможное значение, равное $2c$ (напомним, что $\mu = 2c$). В силу эквивалентности задач (1.9) и (1.10) вектор $g_0 = (w_0, \gamma_0)$ с $w_0 \neq \mathbb{O}$ будет решением задачи (1.9).

Теорема доказана. □

З а м е ч а н и е. В качестве ненулевого вектора p можно взять, например, любую ненулевую разность $b_{j_0} - a_{i_0}$. В этом случае множество планов задачи, двойственной к (1.17), которое определяется условиями (1.19)– (1.21), будет непустым. Вместе с непустотой множества планов само́й задачи (1.17) это гарантирует наличие у задачи (1.17) оптимального плана.

1.8. При $\mu > 0$ решение $g_0 = (w_0, \gamma_0)$ задачи (1.9) с $w_0 \neq \mathbb{O}$ можно привести к каноническому виду. Как и в п. 1.5 положим

$$w_1 = w_0/\|w_0\|,$$

$$\gamma_1 = \tfrac{1}{2}\Big[\min_{j\in 1:k}\langle w_1, b_j\rangle + \max_{i\in 1:m}\langle w_1, a_i\rangle\Big],$$

$$c_1 = \tfrac{1}{2}\Big[\min_{j\in 1:k}\langle w_1, b_j\rangle - \max_{i\in 1:m}\langle w_1, a_i\rangle\Big],$$

$$g_1 = (w_1, \gamma_1).$$

В данном случае $c_1 \le 0$. При $c_1 = 0$ гиперплоскость $H_1 = \big\{x \mid \langle w_1, x\rangle = \gamma_1\big\}$ не строго отделяет множество A от множества B. При $c_1 < 0$ та же гиперплоскость H_1 является наилучшей, приближённо отделяющей множество A от множества B.

Согласно определению w_1, γ_1, c_1 имеем

$$\begin{aligned} \langle w_1, a_i\rangle - \gamma_1 + c_1 \le 0, &\quad i \in 1:m \\ -\langle w_1, b_j\rangle + \gamma_1 + c_1 \le 0, &\quad j \in 1:k. \end{aligned}$$

При $c_1 < 0$ эти неравенства определяют «смешанную полосу»

$$c_1 \le \langle w_1, x\rangle - \gamma_1 \le -c_1,$$

которая содержит как точки множества A, так и точки множества B. Ширина смешанной полосы равна $2|c_1|$.

1.9. На рис. 1.5 приведён пример наилучшего приближённого отделения двух множеств.

1.10. Чтобы подчеркнуть зависимость от параметра c, будем писать $f(g,\ c)$, $\mu(c)$ вместо $f(g)$ и μ. Очевидно, что при всех $c > 0$ справедлива формула

$$f(cg,\ c) = cf(g,\ 1).$$

Поэтому

$$\mu(c) = \min_g f(g,\ c) = \min_g f(cg,\ c) = c\min_g f(g,\ 1) = c\mu(1).$$

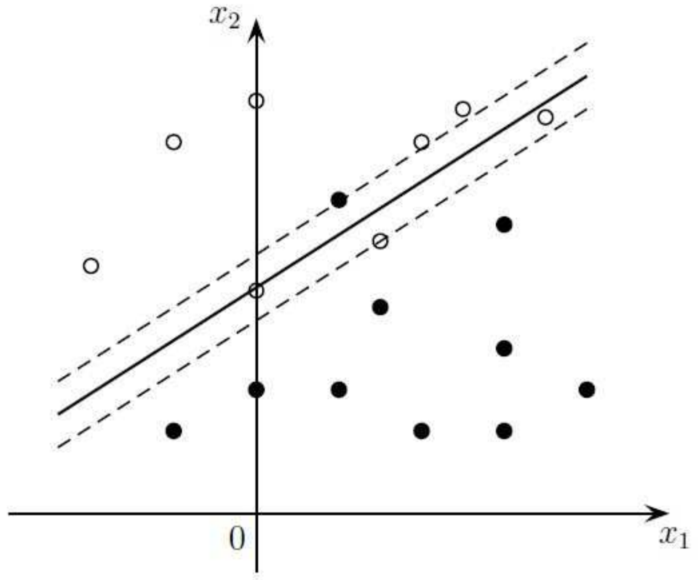

Рис. 1.5. Наилучшее приближённое отделение двух множеств

Более того, если g_1 — решение задачи (1.9) при $c = 1$, то вектор $g_c = cg_1$ будет решением задачи (1.9) при произвольном $c > 0$. Таким образом, аддитивный параметр $c > 0$ играет роль нормирующего множителя.

§ 2. Постановка задачи строгого h-отделения

2.1. Пусть в $\mathbb{R}^n$ заданы два конечных множества

$$A = \{a_i\}_{i=1}^m \quad \text{и} \quad B = \{b_j\}_{j=1}^k.$$

Следуя [27], назовём выпуклую оболочку множества A и множество B *строго h-отделимыми*, если существует h гиперплоскостей вида

$$H_s = \big\{x \in \mathbb{R}^n \mid \langle w^s,\ x\rangle = \gamma_s\big\}, \quad w^s \neq \mathbb{O}, \quad s \in 1:h,$$

таких, что выполняются неравенства

$$\begin{aligned} &\langle w^s,\ a_i\rangle < \gamma_s \quad \text{при всех} \quad i \in 1:m \quad \text{и всех} \quad s \in 1:h, \\ &\langle w^s,\ b_j\rangle > \gamma_s \quad \text{при каждом} \quad j \in 1:k \quad \text{и некотором} \quad s \in 1:h. \end{aligned} \tag{2.1}$$

На рис. 2.1 приведён пример строгого 2-отделения.

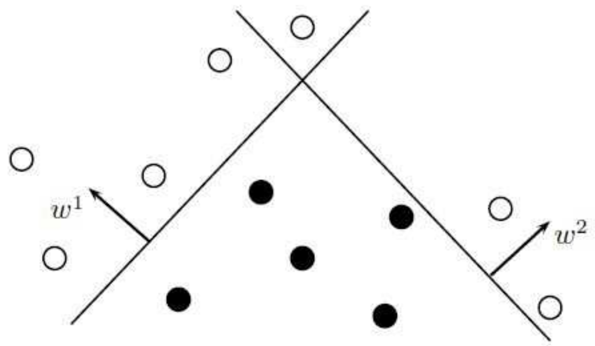

Рис. 2.1. Пример строгого 2-отделения

Введём функцию

$$F(G) = \frac{1}{m}\sum_{i=1}^{m} \max_{s\in 1:h}\big[\langle w^s,\ a_i\rangle - \gamma_s + c\big]_+ + \frac{1}{k}\sum_{j=1}^{k} \min_{s\in 1:h}\big[-\langle w^s,\ b_j\rangle + \gamma_s + c\big]_+.$$

Здесь G — матрица размера $h \times (n+1)$ со строками

$$g^s = (w^s,\ \gamma_s), \quad s \in 1:h;$$

$c > 0$ — параметр. Матрицу G указанного вида будем называть *подходящей*, если у неё все элементы w^s ненулевые ($w^s \neq \mathbb{O}$, $s \in 1:h$). Ясно, что $F(G) \geq 0$ при всех G.

Теорема 2.1. *Выпуклая оболочка множества A и множество B строго h-отделимы тогда и только тогда, когда существует подходящая матрица G_*, такая, что $F(G_*) = 0$.*

Д о к а з а т е л ь с т в о. Н е о б х о д и м о с т ь. Пусть выполнены соотношения (2.1) и $w^s \neq \mathbb{O}$ при всех $s \in 1:h$. Обозначим

$$\delta := \min_{i\in 1:m,\, s\in 1:h}\big[-\langle w^s,\ a_i\rangle + \gamma_s\big] > 0.$$

Каждому $j \in 1:k$ соответствует индекс $s_j \in 1:h$, такой, что

$$\delta_j := \langle w^{s_j},\ b_j\rangle - \gamma_{s_j} > 0.$$

При $\delta_* = \min\{\delta,\ \delta_1, \ldots,\ \delta_k\}$, $\delta_* > 0$, выполняются неравенства

$$\delta_* \le -\langle w^s,\ a_i\rangle + \gamma_s, \quad i \in 1:m,\ s \in 1:h,$$

$$\delta_* \le \langle w^{s_j},\ b_j\rangle - \gamma_{s_j}, \quad j \in 1:k.$$

Положим $w_*^s = \frac{c}{\delta_*} w^s$, $\gamma_s^* = \frac{c}{\delta_*}\gamma_s$. Получим

$$\langle w_*^s,\ a_i\rangle - \gamma_s^* + c \le 0, \quad i \in 1:m,\ s \in 1:h;$$

$$-\langle w_*^{s_j},\ b_{s_j}\rangle + \gamma_{s_j}^* + c \le 0, \quad j \in 1:k.$$

Отсюда по определению плюсиковой функции следует, что на подходящей матрице G_* со строками $(w_*^s,\ \gamma_s^*)$, $s \in 1:h$, выполняется равенство $F(G_*) = 0$.

Достаточность. Если $F(G_*) = 0$ на некоторой подходящей матрице G_*, то

$$\max_{s\in 1:h}\big[\langle w_*^s,\ a_i\rangle - \gamma_s^* + c\big]_+ = 0 \quad \text{при всех} \quad i \in 1:m;$$

$$\min_{s\in 1:h}\big[-\langle w_*^s,\ b_j\rangle + \gamma_s^* + c\big]_+ = 0 \quad \text{при всех} \quad j \in 1:k.$$

Это значит, что

$$\big[\langle w_*^s,\ a_i\rangle - \gamma_s^* + c\big]_+ = 0 \quad \text{при всех} \quad i \in 1:m \quad \text{и всех} \quad s \in 1:h;$$

$$\big[-\langle w_*^s,\ b_j\rangle + \gamma_s^* + c\big]_+ = 0 \quad \text{при каждом} \quad j \in 1:k \quad \text{и некотором} \quad s \in 1:h.$$

По определению плюсиковой функции данные соотношения гарантируют выполнение условий (2.1) с $w^s = w_*^s$, $\gamma_s = \gamma_s^*$.

Теорема доказана. □

Замечание. В случае $\mathrm{co}(A) \cap B = \varnothing$ всегда существует строгое $|B|$-отделение. Для этого при каждом $j \in 1:k$ нужно построить гиперплоскость $H_j = \big\{x \in \mathbb{R}^n \mid \langle w^j,\ x\rangle = \gamma_j\big\}$, строго отделяющую точку b_j от замкнутого выпуклого множества $\operatorname{co}(A)$. На самом деле, некоторые гиперплоскости H_j могут отделять от $\operatorname{co}(A)$ сразу несколько точек множества B, поэтому $h \leq |B|$.

На рис. 2.2 приведён пример строгого $|B|$-отделения.

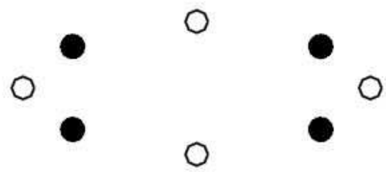

Рис. 2.2. Строго $|B|$-отделимые множества

2.2. Равенство $F(G_*) = 0$ может выполняться на матрице G_*, которая не является подходящей. В этой связи представляет интерес следующее утверждение [3].

Теорема 2.2. *Пусть* $F(G_*) = 0$. *Тогда*

1) *у матрицы* G_* *не все компоненты* w_*^s *равны нулю;*
2) *если* $w_*^s = \mathbb{O}$ *на множестве* $J \subset 1:h$, *то* $\operatorname{co}(A)$ *и* B *строго* $\big(h - |J|\big)$*-отделимы.*

Д о к а з а т е л ь с т в о. Если $F(G_*) = 0$ и все компоненты w_*^s матрицы G_* ненулевые, то

$$\max_{s\in 1:h}[c - \gamma_s^*]_+ + \min_{s\in 1:h}[c + \gamma_s^*]_+ = 0.$$

Но это противоречит свойству 11) плюсиковой функции (см. Дополнение B).

Обозначим через J множество индексов $s \in 1:h$, на которых $w_*^s = \mathbb{O}$. Условие $F(G_*) = 0$ перепишем в виде

$$\begin{gathered}\frac{1}{m}\sum_{i=1}^{m} \max\Big\{\max_{s\in 1:h\setminus J}\big[\langle w_*^s,\ a_i\rangle - \gamma_s^* + c\big]_+,\ \max_{s\in J}[c - \gamma_s^*]_+\Big\} + \\ + \frac{1}{k}\sum_{j=1}^{k} \min\Big\{\min_{s\in 1:h\setminus J}\big[-\langle w_*^s,\ b_j\rangle + \gamma_s^* + c\big]_+,\ \min_{s\in J}[c + \gamma_s^*]_+\Big\} = 0.\end{gathered} \tag{2.2}$$

Из равенства нулю первой суммы и свойства 5) плюсиковой функции следует, что

$$\max_{s\in 1:h\setminus J}\big[\langle w_*^s,\ a_i\rangle - \gamma_s^* + c\big] \le 0, \quad i\in 1:m; \tag{2.3}$$

$$c - \gamma_s^* \le 0, \quad s \in J. \tag{2.4}$$

В силу (2.4), $c+\gamma_s^* \ge 2c$ при всех $s\in J$, так что

$$\min_{s\in J}[c+\gamma_s^*]_+ \ge 2c.$$

По свойству 6) плюсиковой функции равенство нулю второй суммы из (2.2) возможно только тогда, когда

$$\min_{s\in 1:h\setminus J}\big[-\langle w_*^s,\ b_j\rangle + \gamma_s^* + c\big] \le 0, \quad j\in 1:k. \tag{2.5}$$

На основании (2.3) и (2.5) получаем

$$\langle w_*^s,\ a_i\rangle < \gamma_s^* \quad \text{при всех} \quad i\in 1:m \quad \text{и всех} \quad s\in 1:h\setminus J;$$
$$\langle w_*^s,\ b_j\rangle > \gamma_s^* \quad \text{при каждом} \quad j\in 1:k \quad \text{и некотором} \quad s\in 1:h\setminus J.$$

Это и означает, что множества $\mathrm{co}(A)$ и B строго $\big(h-|J|\big)$-отделимы.

Теорема доказана. $\square$

2.3. В дальнейшем мы будем исследовать экстремальную задачу

$$F(G) \to \min, \tag{2.6}$$

где минимум берётся по всем матрицам G размера $h\times(n+1)$. Будет показано, что задача (2.6) всегда имеет решение.

Пусть G_* — какое-нибудь решение задачи (2.6) и $\mu = F(G_*)$. Если $\mu=0$, то по теореме 2.2 множества $\mathrm{co}(A)$ и B строго $\big(h-|J|\big)$-отделимы, где J — множество индексов $s\in 1:h$, на которых $w_*^s = \mathbb{O}$. В частности, если G_* — подходящая матрица, то множества $\mathrm{co}(A)$ и B строго h-отделимы.

При $\mu > 0$ по теореме 2.1 строгое h-отделение невозможно. Если соответствующая матрица G_* подходящая, то будем говорить, что она обеспечивает *наилучшее приближённое h-отделение* множеств $\mathrm{co}(A)$ и B.

§ 3. Строгая h-отделимость и линейное программирование

3.1. Как показано в предыдущем параграфе, задачу строгого h-отделения выпуклой оболочки множества $A = \{a_i\}_{i=1}^m$ от множества $B = \{b_j\}_{j=1}^k$ можно формализовать так:

$$F(G) := \frac{1}{m}\sum_{i=1}^{m}\varphi_i(G) + \frac{1}{k}\sum_{j=1}^{k}\psi_j(G) \to \min, \tag{3.1}$$

где

$$\varphi_i(G) = \max_{s\in 1:h}\big[\langle w^s,\ a_i\rangle - \gamma_s + c\big]_+,$$

$$\psi_j(G) = \min_{s\in 1:h}\big[-\langle w^s,\ b_j\rangle + \gamma_s + c\big]_+.$$

Неизвестной является $\big(h \times (n+1)\big)$-матрица G со строками $(w^s,\ \gamma_s)$, $s \in 1:h$; $c > 0$ — параметр. Принципиальным является тот факт, что задача (3.1) сводится к конечному числу задач линейного программирования [3].

3.2. Обозначим $\Pi = \big\{S = (s_1,\ldots,\ s_k) \mid s_j \in 1:h \text{ при всех } j \in 1:k\big\}$.

Лемма 3.1. *Справедливо равенство*

$$\inf_G F(G) = \min_{S\in\Pi}\inf_G\Big\{\frac{1}{m}\sum_{i=1}^{m}\varphi_i(G) + \frac{1}{k}\sum_{j=1}^{k}\big[-\langle w^{s_j},\ b_j\rangle + \gamma_{s_j} + c\big]_+\Big\}. \tag{3.2}$$

Доказательство. По лемме о сумме минимумов (см. Дополнение D)

$$\sum_{j=1}^{k}\psi_j(G) = \min_{S\in\Pi}\sum_{j=1}^{k}\big[-\langle w^{s_j},\ b_j\rangle + \gamma_{s_j} + c\big]_+.$$

Отсюда и из определения функции $F(G)$ следует, что

$$F(G) = \min_{S\in\Pi} \Big\{ \frac{1}{m} \sum_{i=1}^{m} \varphi_i(G) + \frac{1}{k} \sum_{j=1}^{k} \big[-\langle w^{s_j},\ b_j \rangle + \gamma_{s_j} + c \big]_+ \Big\}.$$

Значит,

$$\inf_G F(G) = \inf_G \min_{S\in\Pi} \Big\{ \frac{1}{m} \sum_{i=1}^{m} \varphi_i(G) + \frac{1}{k} \sum_{j=1}^{k} \big[-\langle w^{s_j},\ b_j \rangle + \gamma_{s_j} + c \big]_+ \Big\}. \tag{3.3}$$

Остаётся в правой части (3.3) поменять местами инфимум по G и минимум по $S \in \Pi$. $\square$

3.3. Лемма показывает, что задача (3.1) эквивалентна конечному числу экстремальных задач вида

$$\frac{1}{m} \sum_{i=1}^{m} \max_{s\in 1:h} \big[\langle w^s,\ a_i \rangle - \gamma_s + c \big]_+ + \frac{1}{k} \sum_{j=1}^{k} \big[-\langle w^{s_j},\ b_j \rangle + \gamma_{s_j} + c \big]_+ \to \inf_G, \tag{3.4}$$

соответствующих различным $S \in \Pi$. В свою очередь, задача (3.4) эквивалентна задаче линейного программирования

$$\frac{1}{m} \sum_{i=1}^{m} p_i + \frac{1}{k} \sum_{j=1}^{k} q_j \to \inf, \tag{3.5}$$

$$-\langle a_i,\ w^s \rangle + \gamma_s + p_i \geq c, \quad i \in 1:m,\ s \in 1:h;$$

$$\langle b_j,\ w^{s_j} \rangle - \gamma_{s_j} + q_j \geq c, \quad j \in 1:k;$$

$$p_i \geq 0, \quad i \in 1:m; \quad q_j \geq 0, \quad j \in 1:k.$$

Задача (3.5) имеет решение при всех $S \in \Pi$. Выделим цепочку S_*, на которой минимальное значение целевой функции в задаче (3.5) принимает наименьшее значение. Обозначим $\big(\{w_*^s\},\ \{\gamma_s^*\},\ \{p_i^*\},\ \{q_j^*\}\big)$ соответствующее решение задачи (3.5). В силу леммы 3.1 матрица G_*, составленная из строк (w_*^s, γ_s^*), $s \in 1:h$, является решением задачи (3.1).

Приходим к следующему заключению.

Теорема 3.1. *Задача* (3.1) *эквивалентна конечному числу задач линейного программирования вида* (3.5) *в том смысле, что решение задачи* (3.5) *при* $S_* \in \Pi$, *которому соответствует наименьшее значение целевой функции, порождает решение задачи* (3.1).

3.4. Рассмотрим пример. Пусть на плоскости заданы множества A и B, состоящие соответственно из точек

$$a_1 = (-2, 0),\ a_2 = (2, 0),\ a_3 = (0, 2),\ a_4 = (0, 1);$$

$$b_1 = (0, 3),\ b_2 = (3, 0),\ b_3 = (-3, 0).$$

Очевидно, что $\mathrm{co}(A) \cap B = \varnothing$ (см. рис. 3.1).

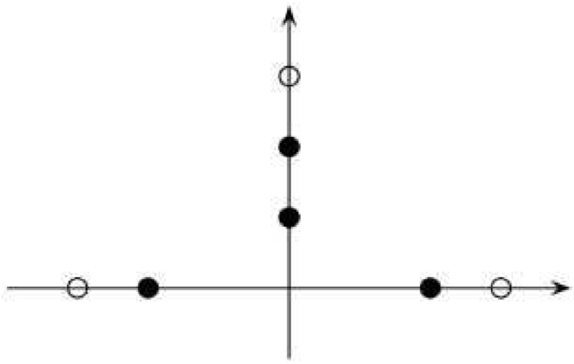

Рис. 3.1. Множества A и B

Решим задачу строгой 2-отделимости. В данном случае

$$n = 2,\ m = 4,\ k = 3,\ h = 2.$$

Выясним, как выглядит задача (3.5) при $S = (1,\ 1,\ 2)$. Выпишем вектор неизвестных

$$z = (w_1^1,\ w_2^1,\ \gamma_1,\ w_1^2,\ w_2^2,\ \gamma_2,\ p_1,\ p_2,\ p_3,\ p_4,\ q_1,\ q_2,\ q_3)$$

и матрицу ограничений

$$
D = \left[\begin{array}{rrrrrrrrrrrrr}
2 & 0 & 1 & & & & 1 & & & & & & \\
-2 & 0 & 1 & & & & & 1 & & & & & \\
0 & -2 & 1 & & & & & & 1 & & & & \\
0 & -1 & 1 & & & & & & & 1 & & & \\
 & & & 2 & 0 & 1 & 1 & & & & & & \\
 & & & -2 & 0 & 1 & & 1 & & & & & \\
 & & & 0 & -2 & 1 & & & 1 & & & & \\
 & & & 0 & -1 & 1 & & & & 1 & & & \\
0 & 3 & -1 & & & & & & & & 1 & & \\
3 & 0 & -1 & & & & & & & & & 1 & \\
 & & & -3 & 0 & -1 & & & & & & & 1
\end{array}\right].
$$

Задача (3.5) примет вид

$$
\frac{1}{4}\sum_{i=1}^{4} p_i + \frac{1}{3}\sum_{j=1}^{3} q_j \to \inf, \tag{3.6}
$$

$$
Dz \geq c\, e,
$$

$$
p_i \geq 0, \quad i \in 1:4; \quad q_j \geq 0, \quad j \in 1:3,
$$

где e — вектор, все компоненты которого равны единице. При $c = 1$ решением данной задачи является вектор

$$
z_* = (111.2210,\ 112.0012,\ 272.9097,\ -78.1474,\ 27.3511,\ 192.1601,
$$

$$
0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000).
$$

Минимальное значение целевой функции равно нулю. Строгое 2-отделение множества $\mathrm{co}(A)$ от B при $S = (1,\, 1,\, 2)$ выглядит так, как показано на рис. 3.2.

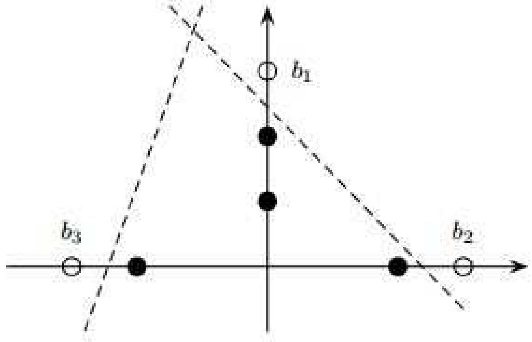

Рис. 3.2. Строгое 2-отделение при $S = (1,\ 1,\ 2)$

Из общих соображений (см. пункт 1.10) следует, что при $c = 0.1$ решением задачи (3.6) является вектор $0.1z_*$. Вместе с тем, непосредственное решение задачи линейного программирования (3.6) при $c = 0.1$ по программе из MATLAB даёт такой результат:

$$z = (108.1756,\ 108.8484,\ 256.1022,\ -76.0149,\ 25.6934,\ 186.8208,$$
$$0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000).$$

Так проявляется неединственность решения задачи строгого h-отделения.

Задание вектора индексов $S = (1,\ 1,\ 2)$ соответсвует разбиению множества B на два подмножества $\{b_1,\ b_2\} \cup \{b_3\}$. Эти два подмножества согласовано отделяются от $\mathrm{co}(A)$ с помощью двух прямых

$$\langle w^1,\ x\rangle = \gamma_1 \quad \text{и} \quad \langle w^2,\ x\rangle = \gamma_2.$$

Существуют ещё два разбиения множества B на два подмножества:

$$\{b_1,\ b_3\} \cup \{b_2\} \quad \text{и} \quad \{b_2,\ b_3\} \cup \{b_1\}.$$

Им соответствуют векторы $S = (1,\ 2,\ 1)$ и $S = (2,\ 1,\ 1)$.

Результат строгого 2-отделения при $S = (1,\ 2,\ 1)$ показан на рис. 3.3.

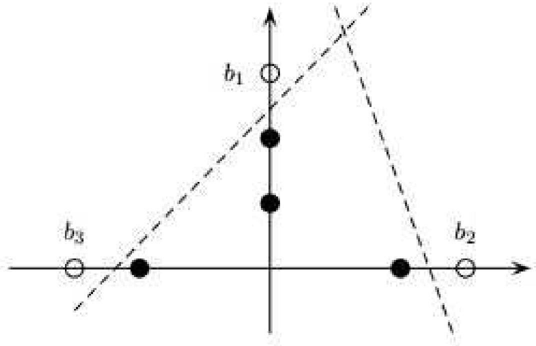

Рис. 3.3. Строгое 2-отделение при $S = (1,\ 2,\ 1)$

Этот случай симметричен случаю $S = (1,\ 1,\ 2)$.

При $S = (2,\ 1,\ 1)$ строгой 2-отделимости нет (см. рис. 3.4).

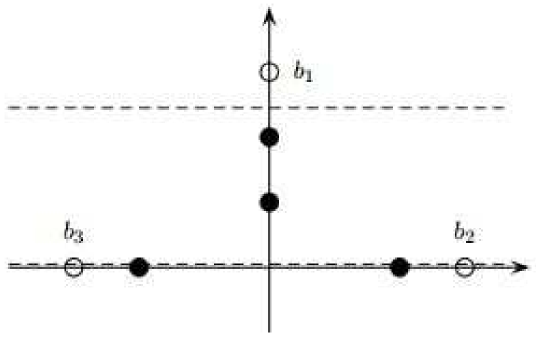

Рис. 3.4. Случай $S = (2,\ 1,\ 1)$

Решением задачи, аналогичной (3.6), при $c = 1$ является вектор

$$z = (0.0000,\ -112.0230,\ -1.0000,\ 0.0000,\ 111.8673,\ 273.7824,$$
$$2.0000,\ 2.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000,\ 0.0000).$$

Минимальное значение целевой функции равно единице.

3.5. В общем случае задание вектора $S \in \Pi$ соответствует разбиению множества B, состоящего из k векторов, на h подмножеств. Число таких разбиений и определяет количество задач линейного программирования вида (3.5), к решению которых сводится решение задачи (3.1).

Если заранее известно, что множества $\mathrm{co}(A)$ и B строго h-отделимы, то решение задачи (3.1) можно упростить. После разбиения множества B

на h подмножеств следует *независимо* решать задачи линейного отделения каждого из этих подмножеств от множества A. В случае успешного отделения совокупность разделяющих гиперплоскостей образует решение задачи (3.1).

В рассмотренном выше примере при $S = (1,\ 1,\ 2)$ будем независимо решать задачи линейного отделения множеств $\{b_1,\ b_2\}$ и $\{b_3\}$ от A. Запишем соответствующие задачи линейного программирования (см. (1.10))

$$\begin{gathered}
\frac{1}{4}\sum_{i=1}^{4} p_i + \frac{1}{2}\sum_{j=1}^{2} q_j \to \inf, \\
-\langle a_i,\ w^1\rangle + \gamma_1 + p_i \geq c, \quad i \in 1:4; \\
\langle b_j,\ w^1\rangle - \gamma_1 + q_j \geq c, \quad j \in 1:2; \\
p_i \geq 0, \quad i \in 1:4; \quad q_j \geq 0, \quad j \in 1:2,
\end{gathered} \tag{3.7}$$

и

$$\begin{gathered}
\frac{1}{4}\sum_{i=1}^{4} p_i + q_3 \to \inf, \\
-\langle a_i,\ w^2\rangle + \gamma_2 + p_i \geq c, \quad i \in 1:4; \\
\langle b_3,\ w^2\rangle - \gamma_2 + q_3 \geq c; \\
p_i \geq 0, \quad i \in 1:4; \quad q_3 \geq 0.
\end{gathered} \tag{3.8}$$

Их решения $\{w^1,\ \gamma_1\}$ и $\{w^2,\ \gamma_2\}$ определяют две прямые, строго отделяющие $\mathrm{co}(A)$ от B (см. рис. 3.5).

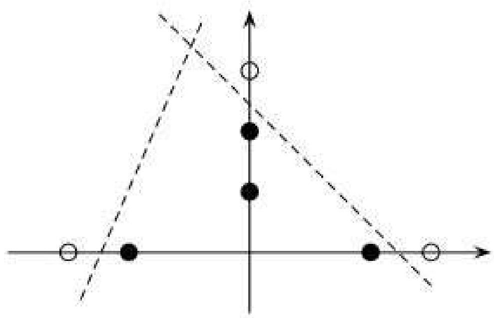

Рис. 3.5. Решения задач (3.7) и (3.8)

§ 4. Метод «градиентного типа» строгого h-отделения

4.1. Итак, задача (3.1) строгого h-отделения сводится к конечному числу задач линейного программирования. Это приниципиальный факт. Однако с практической точки зрения здесь возникают трудности, поскольку количество соответствующих задач линейного программирования может быть достаточно большим. В такой ситуации приобретают интерес приближённые методы.

Ниже будет показано, что функция от матрицы $F(G)$ дифференцируема по направлениям (в качестве которых также выступают матрицы). Это позволит построить метод градиентного типа для решения задачи (3.1).

4.2. Введём обозначения

$$f(v,\ u) = \langle v,\ u\rangle + c, \quad v,\ u \in \mathbb{R}^{n+1},$$

$$\hat{a}_i = \begin{pmatrix} a_i \\ -1 \end{pmatrix},\ i \in 1:m, \quad \check{b}_j = \begin{pmatrix} -b_j \\ 1 \end{pmatrix},\ j \in 1:k.$$

С их помощью функции $\varphi_i(G)$ и $\psi_j(G)$ можно представить в виде

$$\varphi_i(G) = \max_{s\in 1:h}\big[f(g^s,\ \hat{a}_i)\big]_+, \quad \psi_j(G) = \min_{s\in 1:h}\big[f(g^s,\ \check{b}_j)\big]_+.$$

По-прежнему рассматривается задача

$$F(G) := \frac{1}{m}\sum_{i=1}^{m}\varphi_i(G) + \frac{1}{k}\sum_{j=1}^{k}\psi_j(G) \to \min_G. \tag{4.1}$$

Введём дополнительные обозначения

$$\hat{\varphi}_i(G) = \max_{s\in 1:h} f(g^s,\ \hat{a}_i), \quad \check{\psi}_j(G) = \min_{s\in 1:h} f(g^s,\ \check{b}_j),$$

так что (см. Дополнение B, свойство 5) и 6) плюсиковой функции)

$$\varphi_i(G) = \big[\hat{\varphi}_i(G)\big]_+, \quad \psi_j(G) = \big[\check{\psi}_j(G)\big]_+.$$

Положим

$$\hat{R}_i(G) = \big\{s \in 1:h \mid f(g^s,\ \hat{a}_i) = \hat{\varphi}_i(G)\big\},$$
$$\check{R}_j(G) = \big\{s \in 1:h \mid f(g^s,\ \check{b}_j) = \check{\psi}_j(G)\big\}.$$

Производные функций $\varphi_i(G)$ и $\psi_j(G)$ в точке G по направлению V определяются следующим образом:

$$\varphi_i'(G,\ V) = \lim_{t\to+0} \frac{\varphi_i(G+tV) - \varphi_i(G)}{t},$$
$$\psi_j'(G,\ V) = \lim_{t\to+0} \frac{\psi_j(G+tV) - \psi_j(G)}{t}.$$

Справедлива

Теорема 4.1. *Производные* $\varphi_i'(G,\ V)$, $\psi_j'(G,\ V)$ *существуют для всех* $\big(h\times(n+1)\big)$*-матриц* G *и* V*. При этом*

$$\varphi_i'(G,\ V) = \begin{cases} \max\limits_{s\in\hat{R}_i(G)} \langle v^s,\ \hat{a}_i\rangle, & \text{если } \hat{\varphi}_i(G) > 0, \\ \max\limits_{s\in\hat{R}_i(G)} \big[\langle v^s,\ \hat{a}_i\rangle\big]_+, & \text{если } \hat{\varphi}_i(G) = 0, \\ 0, & \text{если } \hat{\varphi}_i(G) < 0; \end{cases}$$

$$\psi_j'(G,\ V) = \begin{cases} \min\limits_{s\in\check{R}_j(G)} \langle v^s,\ \check{b}_j\rangle, & \text{если } \check{\psi}_j(G) > 0, \\ \min\limits_{s\in\check{R}_j(G)} \big[\langle v^s,\ \check{b}_j\rangle\big]_+, & \text{если } \check{\psi}_j(G) = 0, \\ 0, & \text{если } \check{\psi}_j(G) < 0. \end{cases}$$

Доказательство этой теоремы приведено в Дополнении C.

Отметим, что согласно (4.1)

$$F'(G,\ V) = \frac{1}{m}\sum_{i=1}^{m} \varphi_i'(G,\ V) + \frac{1}{k}\sum_{j=1}^{k} \psi_j'(G,\ V). \tag{4.2}$$

4.3. Переходим к описанию «градиентного» метода. Возьмём начальное приближение G_0. Пусть уже имеется ν-е приближение G_ν.

Решаем вспомогательную задачу

$$F'(G_\nu, V) \to \min_{V \in \Omega}, \tag{4.3}$$

где Ω — множество матриц $V = \{v^s(\alpha)\}$, элементы которых удовлетворяют неравенствам

$$|v^s(\alpha)| \leq K, \quad s \in 1:h,\ \alpha \in 1:n+1.$$

Задача (4.3) сводится к небольшому числу задач линейного программирования. Она имеет решение (в силу ограниченности множества Ω). Обозначим его V_ν. Если

$$F'(G_\nu, V_\nu) \geq 0,$$

то G_ν — стационарная точка функции $F(G)$. Вычисления прекращаются. В противном случае матрица V_ν является направлением убывания функции $F(G)$ из точки G_ν. Находим точку минимума функции $F(G_\nu + tV_\nu)$ при $t > 0$. Обозначим её t_ν. Полагаем $G_{\nu+1} = G_\nu + t_\nu V_\nu$, после чего вычисления повторяются. Описание принципиальной схемы градиентного метода для решения задачи (4.1) завершено.

4.4. Приведём пример строгого 3-отделения двух конечных множеств на плоскости с помощью градиентного метода. Основное внимание будем уделять организации вычислений.

Пусть на плоскости заданы множества A и B, состоящие соответственно из точек

$$a_1 = (-3, 0),\ a_2 = (1, 3),\ a_3 = (2, -1),\ a_4 = (0, 1);$$

$$b_1 = (-2, 2),\ b_2 = (2, 3),\ b_3 = (4, 1),\ b_4 = (-1, -2).$$

Очевидно, что $\mathrm{co}(A) \cap B = \varnothing$ (см. рис. 4.1).

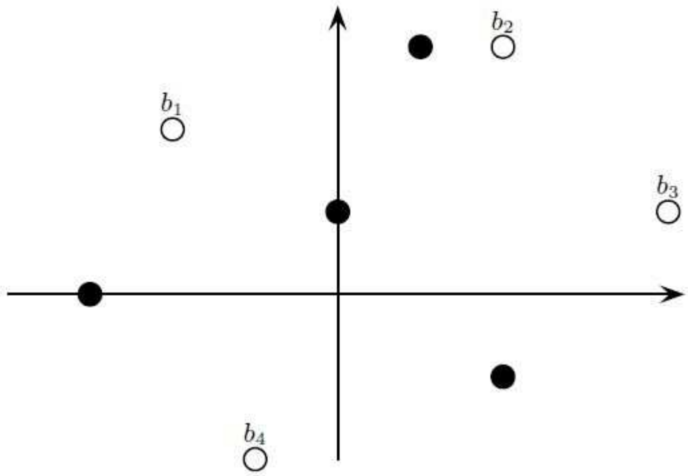

Рис. 4.1. Задача строгого 3-отделения

Сначала будем решать задачу строгого 3-отделения при $c = 0$. Возьмём начальное приближение (см. рис. 4.2)

$$G_0 = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & -1 & 4 \end{pmatrix}.$$

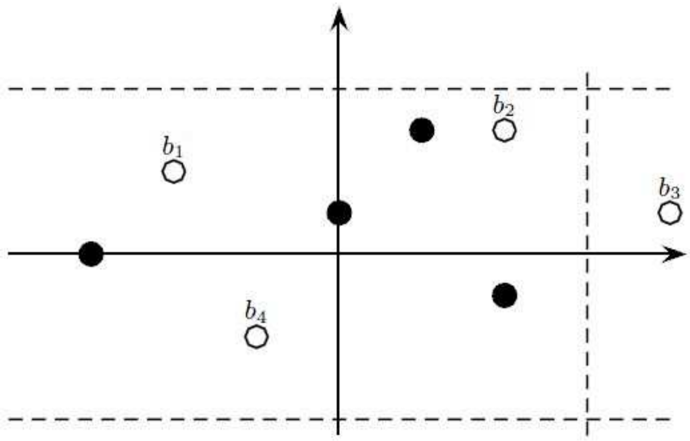

Рис. 4.2. Начальное приближение G_0

Заполним таблицу 1 и таблицу 2.

Последние строки таблицы 1 и таблицы 2 позволяют вычислить значение целевой функции: $F(G_0) = \frac{5}{4}$.

Таблица 1: Вычисление функций $\varphi_i(G_0)$

i	1	2	3	4
$f(g_0^1, \hat{a}_i)$	-6	-2	-1	-3
$f(g_0^2, \hat{a}_i)$	-4	-1	-5	-3
$f(g_0^3, \hat{a}_i)$	-4	-7	-3	-5
$\hat{\varphi}_i(G_0)$	-4	-1	-1	-3
$\hat{R}_i(G_0)$	$\{2,3\}$	$\{2\}$	$\{1\}$	$\{1,2\}$
$\varphi_i(G_0)$	0	0	0	0

Таблица 2: Вычисление функций $\psi_j(G_0)$

j	1	2	3	4
$f(g_0^1, \check{b}_j)$	5	1	-1	4
$f(g_0^2, \check{b}_j)$	2	1	3	6
$f(g_0^3, \check{b}_j)$	6	7	5	2
$\check{\psi}_j(G_0)$	2	1	-1	2
$\check{R}_j(G_0)$	$\{2\}$	$\{1,\ 2\}$	$\{1\}$	$\{3\}$
$\psi_j(G_0)$	2	1	0	2

Переходим к вычислению производных по направлению. Имеем

$$\varphi_i'(G_0,\ V) \equiv 0 \quad \text{при всех} \quad i \in 1:4;$$

$$\psi_1'(G_0,\ V) = \langle v^2, \check{b}_1 \rangle = 2v^2(1) - 2v^2(2) + v^2(3);$$

$$\psi_2'(G_0,\ V) = \min_{s \in \{1,\ 2\}} \langle v^s, \check{b}_2 \rangle =$$

$$= \min\{-2v^1(1) - 3v^1(2) + v^1(3),\ -2v^2(1) - 3v^2(2) + v^2(3)\};$$

$$\psi_3'(G_0,\ V) \equiv 0;$$

$$\psi_4'(G_0,\ V) = \langle v^3, \check{b}_4 \rangle = v^3(1) + 2v^3(2) + v^3(3).$$

Производная по направлению V функции $F(G)$ при $G = G_0$ в случае $\psi_2'(G_0,\ V) = -2v^1(1) - 3v^1(2) + v^1(3)$ принимает вид

$$F'(G_0,\ V) = -\tfrac{1}{2}v^1(1) - \tfrac{3}{4}v^1(2) + \tfrac{1}{4}v^1(3) + \tfrac{1}{2}v^2(1) - \tfrac{1}{2}v^2(2) + \tfrac{1}{4}v^2(3) +$$
$$+ \tfrac{1}{4}v^3(1) + \tfrac{1}{2}v^3(2) + \tfrac{1}{4}v^3(3).$$

Решаем задачу линейного программирования

$$F'(G_0,\ V) \to \min_{V \in \Omega}, \tag{4.4}$$

где в качестве Ω возьмём множество (3×3)-матриц V, все элементы которых по модулю не превосходят 10 ($K = 10$). Очевидно, что решением задачи (4.4) является матрица

$$V_0 = \begin{pmatrix} 10 & 10 & -10 \\ -10 & 10 & -10 \\ -10 & -10 & -10 \end{pmatrix};$$

при этом $F'(G_0,\ V_0) = -37.5$.

Таким образом, получили, что в случае $\psi_2'(G_0,\ V) = -2v^1(1) - 3v^1(2) + v^1(3)$ решение V_0 — направление наискорейшего убывания функции $F(G)$ при $G = G_0$. Минимум функции $F(G_0 + tV_0)$ на луче $t > 0$ достигается при $t_0 = 0.05$, поэтому

$$G_1 := G_0 + t_0 V_0 = \begin{pmatrix} 1.5 & 0.5 & 2.5 \\ -0.5 & 1.5 & 3.5 \\ -0.5 & -1.5 & 3.5 \end{pmatrix}$$

(см. рис. 4.3).

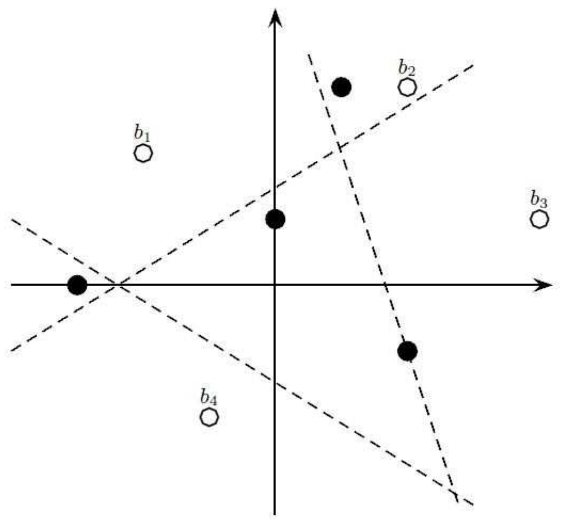

Рис. 4.3. Матрица G_1

Повторяем процесс.

Таблица 3: Вычисление функций $\varphi_i(G_1)$

i	1	2	3	4
$f(g_1^1, \hat{a}_i)$	-7	0.5	0	-2
$f(g_1^2, \hat{a}_i)$	-2	0.5	-6	-2
$f(g_1^3, \hat{a}_i)$	-2	-8.5	-3	-5
$\hat{\varphi}_i(G_1)$	-2	0.5	0	-2
$\hat{R}_i(G_1)$	$\{2,\ 3\}$	$\{1,\ 2\}$	$\{1\}$	$\{1,\ 2\}$
$\varphi_i(G_1)$	0	0.5	0	0

Таблица 4: Вычисление функций $\psi_j(G_1)$

j	1	2	3	4
$f(g_1^1, \check{b}_j)$	4.5	-2	-4	5
$f(g_1^2, \check{b}_j)$	-0.5	0	4	6
$f(g_1^3, \check{b}_j)$	5.5	9	7	0
$\check{\psi}_j(G_1)$	-0.5	-2	-4	0
$\check{R}_j(G_1)$	$\{2\}$	$\{1\}$	$\{1\}$	$\{3\}$
$\psi_j(G_1)$	0	0	0	0

Очевидно, что $F(G_1) = 0.125$. Переходим к вычислению производных по направлению. Имеем

$$\varphi_i'(G_1,\ V) \equiv 0 \quad \text{при} \quad i = 1,\ 4;$$

$$\varphi_2'(G_1,\ V) = \max_{s \in 1:2} \langle v^s,\ \hat{a}_2 \rangle;$$

$$\varphi_3'(G_1,\ V) = \big[\langle v^1,\ \hat{a}_3 \rangle\big]_+;$$

$$\psi_j'(G_1,\ V) \equiv 0 \quad \text{при} \quad j = 1,\ 2,\ 3;$$

$$\psi_4'(G_1,\ V) = \big[\langle v^3,\ \check{b}_4 \rangle\big]_+.$$

Производная по направлению V функции $F(G)$ при $G = G_1$ принимает вид

$$F'(G_1,\ V) = \tfrac{1}{4}\Big\{\max_{s \in 1:2} \langle v^s,\ \hat{a}_2 \rangle + \big[\langle v^1,\ \hat{a}_3 \rangle\big]_+ + \big[\langle v^3,\ \check{b}_4 \rangle\big]_+\Big\}.$$

Задача (4.4) при $K = 10$ сводится к задаче линейного программирования

$$\tfrac{1}{4}\big\{t_1 + t_2 + t_3\big\} \to \min$$

при ограничениях

$$\big|v^s(\alpha)\big| \le 10, \quad \alpha \in 1:3,\ s \in 1:3;$$

$$t_1 \ge \langle v^1,\ \hat{a}_2\rangle, \quad t_1 \ge \langle v^2,\ \hat{a}_2\rangle,$$

$$t_2 \ge \langle v^1,\ \hat{a}_3\rangle, \quad t_3 \ge \langle v^3,\ \check{b}_4\rangle,$$

$$t_2 \ge 0, \quad t_3 \ge 0.$$

Для решения последней задачи использовалась программа из MATLAB [13]. Получили матрицу

$$V_1 = \begin{pmatrix} -10 & -10 & 10 \\ -10 & -10 & 10 \\ -5.43 & -7.1 & -5.43 \end{pmatrix};$$

при этом $F'(G_1,\ V_1) = -12.5$. Минимум функции $F(G_1 + tV_1)$ на луче $t > 0$ достигается при $t_1 = 0.01$, поэтому

$$G_2 := G_1 + t_1 V_1 = \begin{pmatrix} 1.4 & 0.4 & 2.6 \\ -0.6 & 1.4 & 3.6 \\ -0.55 & -1.57 & 3.45 \end{pmatrix},$$

и $F(G_2) = 0$. Учитывая, что G_2 — подходящая матрица, заключаем, что G_2 — решение рассматриваемой задачи (см. рис. 4.4).

4.5. Выбор параметра $c = 0$ не соответствует теории (по теории нужно брать $c > 0$). При $c = 0$ очевидным решением задачи (3.1) является матрица $G = \mathbb{O}$. Однако решение неединственно. Мы начинали вычисления с подходящей матрицы G_0 и пришли к подходящей матрице G_2, которая реализует h-отделение, правда, нестрогое.

Теперь вместо $c = 0$ положим $c = \frac{1}{2}$. В качестве начального приближения возьмём ту же матрицу G_0. Заполним таблицу 5 и таблицу 6.

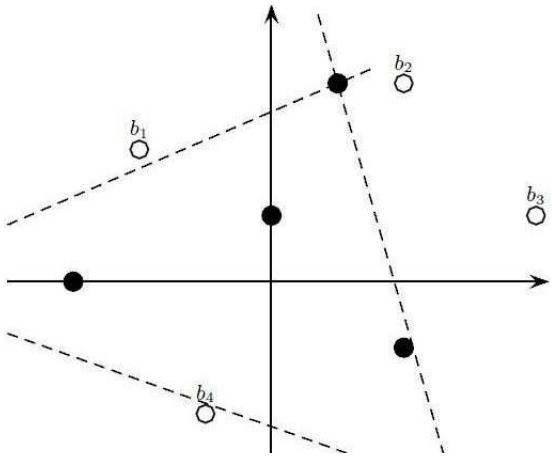

Рис. 4.4. Матрица G_2 — решение задачи при $c = 0$

Таблица 5: Вычисление функций $\varphi_i(G_0)$

i	1	2	3	4
$f(g_0^1, \hat{a}_i)$	-5.5	-1.5	-0.5	-2.5
$f(g_0^2, \hat{a}_i)$	-3.5	-0.5	-4.5	-2.5
$f(g_0^3, \hat{a}_i)$	-3.5	-6.5	-2.5	-4.5
$\hat{\varphi}_i(G_0)$	-3.5	-0.5	-0.5	-2.5
$\hat{R}_i(G_0)$	$\{2,3\}$	$\{2\}$	$\{1\}$	$\{1,2\}$
$\varphi_i(G_0)$	0	0	0	0

Таблица 6: Вычисление функций $\psi_j(G_0)$

j	1	2	3	4
$f(g_0^1, \check{b}_j)$	5.5	1.5	-0.5	4.5
$f(g_0^2, \check{b}_j)$	2.5	1.5	3.5	6.5
$f(g_0^3, \check{b}_j)$	6.5	7.5	5.5	2.5
$\check{\psi}_j(G_0)$	2.5	1.5	-0.5	2.5
$\check{R}_j(G_0)$	$\{2\}$	$\{1,\ 2\}$	$\{1\}$	$\{3\}$
$\psi_j(G_0)$	2.5	1.5	0	2.5

По последним строкам таблицы 5 и таблицы 6 находим значение целевой функции: $F(G_0) = 1.625$. Далее имеем

$$G_1 = \begin{pmatrix} 1.5 & 0.5 & 2.5 \\ -0.5 & 1.5 & 3.5 \\ -0.5 & -1.5 & 3.5 \end{pmatrix}.$$

Непосредственно проверяется, что $F(G_1) = 0.5$. (см. рис. 4.5).

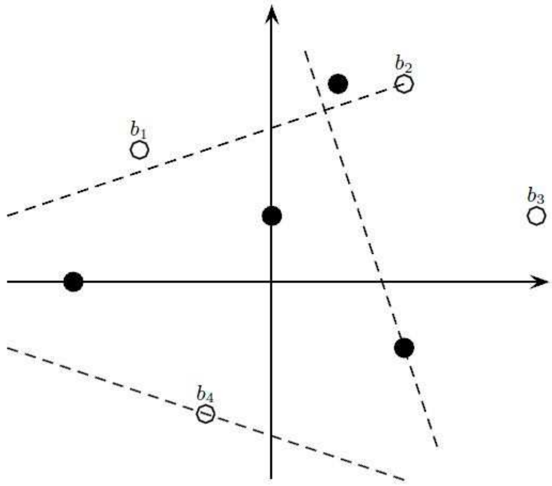

Рис. 4.5. Подходящая матрица G_1

Повторяем процесс, и на следующем шаге получаем матрицу

$$G_2 = \begin{pmatrix} 1.2 & 0.3 & 2.8 \\ -0.8 & 1.2 & 3.5 \\ -0.8 & -1.8 & 3.2 \end{pmatrix};$$

при этом $F(G_2) = 0$ (см. рис. 4.6). Значит, подходящая матрица G_2 осуществляет строгое 3-отделение.

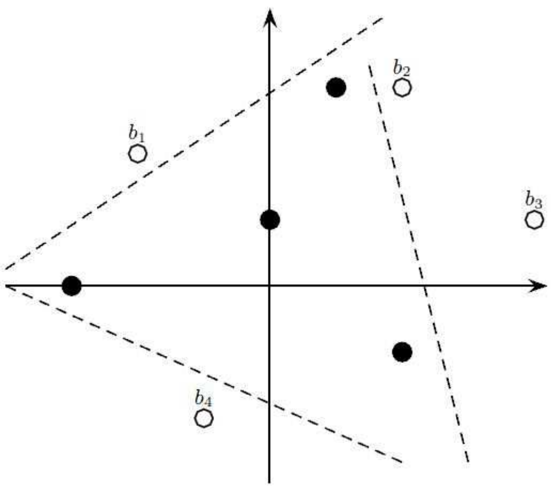

Рис. 4.6. Матрица G_2 — решение задачи при $c = \frac{1}{2}$

§ 5. Безградиентный метод локального поиска

5.1. Рассмотрим ещё один, безградиентный, метод решения задачи (4.1), максимально учитывающий её специфику. Описание метода потребует некоторой подготовки.

В пространстве $\mathbb{R}^{n+1}$ будем использовать равномерную норму векторов,

$$\|v\| = \max_{\alpha \in 1:n+1} \bigl| v(\alpha) \bigr|,$$

и согласованную с ней норму матриц G,

$$\|G\| = \max_{s \in 1:h} \sum_{\alpha=1}^{n+1} \bigl| g^s(\alpha) \bigr|.$$

Очевидно, что для любой строки g^s матрицы G и любого вектора $v \in \mathbb{R}^{n+1}$ выполняется неравенство

$$\bigl| \langle g^s,\ v \rangle \bigr| \leq \|G\| \cdot \|v\|, \quad s \in 1:h. \tag{5.1}$$

Возьмём два параметра точности $\varepsilon_A > 0$ и $\varepsilon_B > 0$. С каждой матрицей G свяжем индексные множества

$$I = \Bigl\{ i \in 1:m \mid \max_{s \in 1:h} f(g^s,\ \hat{a}_i) > -\varepsilon_A \Bigr\},$$

$$J = \Bigl\{ j \in 1:k \mid \min_{s \in 1:h} f(g^s,\ \check{b}_j) > -\varepsilon_B \Bigr\}.$$

Положим

$$\varepsilon = \min\{\varepsilon_A / C_A,\ \varepsilon_B / C_B\},$$

где

$$C_A = \max_{i \in 1:m} \|\hat{a}_i\|, \quad C_B = \max_{j \in 1:k} \|\check{b}_j\|.$$

Лемма 5.1. *Для любой матрицы G выполняется равенство*

$$F(G+V) = \frac{1}{m} \sum_{i \in I} \varphi_i(G+V) + \frac{1}{k} \sum_{j \in J} \psi_j(G+V) \tag{5.2}$$

при условии, что $\|V\| \leq \varepsilon$.

Доказательство. Нужно проверить, что при указанных V

$$\varphi_i(G+V)=0, \text{ если } i\notin I; \quad \psi_j(G+V)=0, \text{ если } j\notin J. \tag{5.3}$$

По определению функции f имеем

$$f(g^s+v^s,\ \hat{a}_i)=f(g^s,\ \hat{a}_i)+\langle v^s,\ \hat{a}_i\rangle.$$

Далее (см. Дополнение B, свойство 5))

$$\varphi_i(G+V)=\Big[\max_{s\in 1:h}\big(f(g^s,\ \hat{a}_i)+\langle v^s,\ \hat{a}_i\rangle\big)\Big]_+. \tag{5.4}$$

По определению множества I при $i\notin I$ и $s\in 1:h$ выполняется неравенство

$$f(g^s,\ \hat{a}_i)\le -\varepsilon_A,$$

поэтому согласно (5.1) при всех $s\in 1:h$

$$f(g^s,\ \hat{a}_i)+\langle v^s,\ \hat{a}_i\rangle\le -\varepsilon_A+\|V\|C_A\le 0.$$

На основании (5.4) и определения плюсиковой функции получаем первое равенство из (5.3).

Аналогично (см. Дополнение B, свойство 6))

$$\psi_j(G+V)=\Big[\min_{s\in 1:h}\big(f(g^s,\ \check{b}_j)+\langle v^s,\ \check{b}_j\rangle\big)\Big]_+. \tag{5.5}$$

При $j\notin J$ существует индекс $s\in 1:h$, на котором

$$f(g^s,\ \check{b}_j)\le -\varepsilon_B.$$

На том же индексе s имеем

$$f(g^s,\ \check{b}_j)+\langle v^s,\ \check{b}_j\rangle\le -\varepsilon_B+\|V\|C_B\le 0.$$

Отсюда и из (5.5) следует второе равенство из (5.3).

Лемма доказана. □

Отметим, что в формулировке леммы 5.1 окрестность $\|V\|\le\varepsilon$ одинакова для всех матриц G.

5.2. Обозначим

$$l_j = \min_{s \in 1:h} f(g^s,\ \check{b}_j).$$

С матрицей G свяжем индексные множества

$$L_j = \bigl\{ s \in 1:h \mid f(g^s,\ \check{b}_j) = l_j \bigr\}, \quad j \in J.$$

Они соответствуют множествам $\check{R}_j(G)$, введённым в п. 4.2, $L_j = \check{R}_j(G)$.

Имеем

$$\min_{s \in L_j} \bigl[f(g^s,\ \check{b}_j) \bigr]_+ = \bigl[\min_{s \in L_j} f(g^s,\ \check{b}_j) \bigr]_+ = [l_j]_+ =$$
$$= \bigl[\min_{s \in 1:h} f(g^s,\ \check{b}_j) \bigr]_+ = \min_{s \in 1:h} \bigl[f(g^s,\ \check{b}_j) \bigr]_+ = \psi_j(G).$$

Положив в (5.2) $V = \mathbb{O}$, получим

$$F(G) = \frac{1}{m} \sum_{i \in I} \varphi_i(G) + \frac{1}{k} \sum_{i \in J} \min_{s \in \check{R}_j(G)} \bigl[f(g^s,\ \check{b}_j) \bigr]_+.$$

Лемма 5.2. *Для каждой матрицы G найдётся такое число $\varepsilon_G \in (0, \varepsilon]$, что*

$$F(G+V) = \frac{1}{m} \sum_{i \in I} \varphi_i(G+V) + \frac{1}{k} \sum_{j \in J} \min_{s \in \check{R}_j(G)} \bigl[f(g^s + v^s,\ \check{b}_j) \bigr]_+$$

при условии, что $\|V\| \le \varepsilon_G$.

Доказательство. В силу леммы 5.1 достаточно проверить, что при $j \in J$ и малых $\|V\|$

$$\min_{s \in L_j} f(g^s + v^s,\ \check{b}_j) = \min_{s \in 1:h} f(g^s + v^s,\ \check{b}_j). \tag{5.6}$$

Пусть

$$\min_{s \notin L_j} f(g^s,\ \check{b}_j) = l_j + \Delta_j, \quad j \in J,$$

где $\Delta_j > 0$. Положим

$$\varepsilon_G = \min\bigl\{ \varepsilon,\ \min_{j \in J} \tfrac{\Delta_j}{3C_B} \bigr\}.$$

При $\|V\| \le \varepsilon_G$ и $s \notin L_j$ имеем

$$f(g^s + v^s,\ \check{b}_j) = f(g^s,\ \check{b}_j) + \langle v^s,\ \check{b}_j\rangle \ge l_j + \Delta_j - \|V\| C_B \ge l_j + \tfrac{2}{3}\Delta_j. \quad (5.7)$$

В то же время при тех же V и $s \in L_j$

$$f(g^s + v^s,\ \check{b}_j) = l_j + \langle v^s,\ \check{b}_j\rangle \le l_j + \tfrac{1}{3}\Delta_j. \quad (5.8)$$

На основании (5.7) и (5.8) заключаем, что при $\|V\| \le \varepsilon_G$ выполняется равенство (5.6).

Лемма доказана. □

5.3. Зафиксируем матрицу G_0 со строками g_0^s, $s \in 1:h$, константу $K > \varepsilon$ и рассмотрим экстремальную задачу

$$Q(G_0 + V) := \frac{1}{m}\sum_{i\in I_0} \varphi_i(G_0 + V) + \frac{1}{k}\sum_{j\in J_0} \min_{s\in L_j^0}\bigl[f(g_0^s + v^s,\ \check{b}_j)\bigr]_+ \to \min_{V\in\Omega}, \quad (5.9)$$

где индексные множества I_0, J_0 и L_j^0 соответсвуют матрице G_0. Множество планов Ω определим так:

$$\bigl|v^s(\alpha)\bigr| \le K, \quad \alpha \in 1:n+1,\ s \in 1:h.$$

В частности, матрица V, у которой $\|V\| \le K$, является планом задачи (5.9). Действительно,

$$\|v^s(\alpha)\| \le \sum_{p=1}^{n+1}\bigl|v^s(p)\bigr| \le \max_{s\in 1:h}\sum_{p=1}^{n+1}\bigl|v^s(p)\bigr| = \|V\| \le K.$$

Отметим, что по лемме 5.2

$$Q(G_0 + V) = F(G_0 + V),\ \text{если}\ \|V\| \le \varepsilon_{G_0}. \quad (5.10)$$

Вместе с тем, $L_j^0 \subset 1:h$, поэтому

$$\psi_j(G_0 + V) \le \min_{s\in L_j^0}\bigl[f(g_0^s + v^s,\ \check{b}_j)\bigr]_+.$$

Отсюда и из леммы 5.1 следует, что

$$F(G_0 + V) \le Q(G_0 + V),\ \text{если}\ \|V\| \le \varepsilon. \quad (5.11)$$

Обозначим через Q_0 минимальное значение целевой функции в задаче (5.9). Согласно (5.10)

$$Q_0 \le Q(G_0) = F(G_0). \tag{5.12}$$

Теорема 5.1. *При выполнении условия*

$$Q_0 = F(G_0) \tag{5.13}$$

матрица G_0 *является точкой локального минимума функции* $F(G)$ *вида* (4.1).

Доказательство. На основании (5.10), условия $\varepsilon < K$ и равенства (5.13) при $\|V\| \le \varepsilon_{G_0}$ имеем

$$F(G_0 + V) = Q(G_0 + V) \ge Q_0 = F(G_0).$$

Это и означает, что G_0 — точка локального минимума функции $F(G)$. $\square$

5.4. Если матрица G_0 не является точкой локального минимума функции $F(G)$, то согласно (5.12) выполняется неравенство $F(G_0) > Q_0$. На самом деле, мы ориентируемся на приближённые вычисления, поэтому будем считать, что

$$F(G_0) - Q_0 > \sigma,$$

где σ — наперёд заданный положительный параметр точности. Последнее неравенство перепишем в виде

$$Q_0 < F(G_0) - \sigma. \tag{5.14}$$

Матрице G_0 сопоставим совокупность индексных множеств $S = \{s_j\}_{j \in J_0}$, где $s_j \in L_j^0$. Эту совокупность обозначим Π_0. По лемме о сумме минимумов (см. Дополнение D и п. 3.2) величина Q_0 допускает представление

$$Q_0 = \min_{S \in \Pi_0} \min_{V \in \Omega} \Big\{ \frac{1}{m} \sum_{i \in I_0} \varphi_i(G_0 + V) + \frac{1}{k} \sum_{j \in J_0} \big[f(g_0^{s_j} + v^{s_j},\ \check{b}_j) \big]_+ \Big\}. \tag{5.15}$$

На основании (5.14) и (5.15) можно утверждать, что существует индексное множество $\hat{S} = \{\hat{s}_j\}_{j \in J_0}$ из Π_0, такое, что

$$\min_{V \in \Omega} \Big\{ \frac{1}{m} \sum_{i \in I_0} \varphi_i(G_0 + V) + \frac{1}{k} \sum_{j \in J_0} \big[f(g_0^{\hat{s}_j} + v^{\hat{s}_j},\ \check{b}_j) \big]_+ \Big\} < F(G_0) - \sigma. \tag{5.16}$$

Рассмотрим вспомогательную задачу

$$\hat{Q}(G_0 + V) := \frac{1}{m} \sum_{i \in I_0} \varphi_i(G_0 + V) + \frac{1}{k} \sum_{j \in J_0} \big[f(g_0^{\hat{s}_j} + v^{\hat{s}_j},\ \check{b}_j) \big]_+ \to \min_{V \in \Omega}. \tag{5.17}$$

Ясно, что $\hat{Q}(G_0 + V) \geq Q(G_0 + V)$ при любых V. Важным моментом является то, что в силу определения множеств L_j^0 справедливы равенства

$$\hat{Q}(G_0) = Q(G_0) = F(G_0). \tag{5.18}$$

Задача (5.17) эквивалентна задаче линейного программирования

$$\frac{1}{m} \sum_{i \in I_0} \xi_i + \frac{1}{k} \sum_{j \in J_0} \eta_j \to \min, \tag{5.19}$$

$$\xi_i - \langle v^s,\ \hat{a}_i \rangle \geq f(g_0^s,\ \hat{a}_i), \quad i \in I_0,\ s \in 1:h;$$

$$\eta_j - \langle v^{\hat{s}_j},\ \check{b}_j \rangle \geq f(g_0^{\hat{s}_j},\ \check{b}_j), \quad j \in J_0;$$

$$\xi_i \geq 0, \quad i \in I_0; \quad \eta_j \geq 0, \quad j \in J_0;$$

$$-K \leq v^s(\alpha) \leq K, \quad \alpha \in 1:n+1,\ s \in 1:h.$$

Эта задача имеет решение, поскольку множество её планов непусто и целевая функция ограничена снизу нулём. По эквивалентности задача (5.17) также имеет решение. Обозначим его $\hat{V}$. Положим $\hat{G} = G_0 + \hat{V}$. Согласно (5.16), $\hat{Q}(\hat{G}) < F(G_0) - \sigma$. С учётом (5.18) получаем

$$\hat{Q}(\hat{G}) - \hat{Q}(G_0) < -\sigma. \tag{5.20}$$

Матрицу $\hat{V}$ мы воспринимаем как направление спуска. Разберёмся с шагом спуска. Положим

$$\hat t = \begin{cases} 1, & \text{если } \|\hat V\| \le \varepsilon, \\ \varepsilon/\|\hat V\|, & \text{если } \|\hat V\| > \varepsilon, \end{cases}$$

и введём матрицу $G_1 = G_0 + \hat t\,\hat V$. Очевидно, что $\hat t \in (0, 1]$. Более того,

$$\hat t \ge \frac{\varepsilon}{(n+1)K}. \tag{5.21}$$

Действительно, при $\|\hat V\| \le \varepsilon$ неравенство (5.21) очевидно (поскольку $\varepsilon < K$). Пусть $\|\hat V\| > \varepsilon$. По условию, $\hat V \in \Omega$, поэтому

$$\|\hat V\| = \max_{s\in 1:h} \sum_{\alpha=1}^{n+1} \bigl|\hat v^s(\alpha)\bigr| \le (n+1)K.$$

Отсюда и из определения $\hat t$ следует (5.21).

Имеем $\|\hat t\,\hat V\| \le \varepsilon$. Согласно (5.11)

$$F(G_1) \le Q(G_1). \tag{5.22}$$

Далее

$$Q(G_1) \le \hat Q(G_1) = \hat Q\bigl(G_0 + \hat t(\hat G - G_0)\bigr). \tag{5.23}$$

Покажем, что

$$\hat Q\bigl(G_0 + \hat t(\hat G - G_0)\bigr) \le \hat Q(G_0) + \hat t\bigl[\hat Q(\hat G) - \hat Q(G_0)\bigr]. \tag{5.24}$$

Воспользуемся соотношениями

$$f\bigl(g_0^s + \hat t(\hat g^s - g_0^s),\ u\bigr) = f\bigl(\hat t\,\hat g^s + (1-\hat t\,)g_0^s,\ u\bigr) =$$
$$= \hat t\,f(\hat g^s,\ u) + (1-\hat t\,)f(g_0^s,\ u),$$

из которых следует, что (см. Дополнение В, свойства 2) и 4))

$$\Bigl[f\bigl(g_0^s + \hat t(\hat g^s - g_0^s),\ u\bigr)\Bigr]_+ \le \hat t\bigl[f(\hat g^s,\ u)\bigr]_+ + (1-\hat t\,)\bigl[f(g_0^s,\ u)\bigr]_+. \tag{5.25}$$

Подставив в (5.25) $\hat a_i$ и $\check b_j$ вместо u и применив операции взятия максимума и суммирования, придём к неравенству

$$\hat Q\bigl(G_0 + \hat t(\hat G - G_0)\bigr) \le \hat t\,\hat Q(\hat G) + (1-\hat t\,)\hat Q(G_0),$$

равносильному (5.24). По существу, установлена выпуклость функции $\hat{Q}(G)$ вида

$$\hat{Q}(G) = \frac{1}{m}\sum_{i\in I_0}\varphi_i(G) + \frac{1}{k}\sum_{j\in J_0}\big[f(g^{\hat{s}_j},\ \check{b}_j)\big]_+.$$

На основании (5.22), (5.23), (5.24), (5.18), (5.20) и (5.21) получаем

$$F(G_1) \le Q(G_1) \le \hat{Q}(G_0) + \hat{t}\big[\hat{Q}(\hat{G}) - \hat{Q}(G_0)\big] \le$$
$$\le F(G_0) - \hat{t}\sigma \le F(G_0) - \frac{\varepsilon\sigma}{(n+1)K}.$$

Значит,

$$F(G_0) - F(G_1) \ge \frac{\varepsilon\sigma}{(n+1)K}. \tag{5.26}$$

Отметим, что неравенство (5.26) сохранится, если вместо $\hat{t}$ взять точку минимума функции $F(G_0 + t\hat{V})$ на отрезке $[0,1]$.

К матрице G_1 можно применить те же рассуждения, что и к матрице G_0. Ввести индексные множества I_1, J_1, L_j^1, функцию $Q(G_1 + V)$ с минимальным на Ω значением Q_1. Если

$$F(G_1) - Q_1 \le \sigma,$$

то G_1 — по определению почти локально оптимальная матрица. Вычисления прекращаются. В противном случае с помощью функции $\hat{Q}(G_1 + V)$ строим новую матрицу G_2, такую, что

$$F(G_1) - F(G_2) \ge \frac{\varepsilon\sigma}{(n+1)K}.$$

Далее процесс повторяется. Строится последовательность матриц $\{G_\nu\}$, для которой

$$F(G_\nu) - F(G_{\nu+1}) \ge \frac{\varepsilon\sigma}{(n+1)K}.$$

В силу неотрицательности $F(G)$ описанный процесс конечен. Через конечное число шагов получим почти локально оптимальную матрицу G_*.

5.5. Опишем общий шаг вычислительного процесса. В нём используются глобальные параметры ε_A, ε_B и σ.

Пусть имеется ν-е приближение — матрица G_ν со строками g_ν^s, $s \in 1:h$. Для получения очередного приближения $G_{\nu+1}$ выполняются следующие операции:

1) Вычисляется значение $F(G_\nu)$.

2) Формируются индексные множества

$$I_\nu = \bigl\{ i \in 1:m \mid \max_{s \in 1:h} f(g_\nu^s,\ \hat{a}_i) > -\varepsilon_A \bigr\},$$
$$J_\nu = \bigl\{ j \in 1:k \mid \min_{s \in 1:h} f(g_\nu^s,\ \check{b}_j) > -\varepsilon_B \bigr\},$$
$$L_j^\nu = \bigl\{ s \in 1:h \mid f(g_\nu^s,\ \check{b}_j) = \min_{p \in 1:h} f(g_\nu^p,\ \check{b}_j) \bigr\}, \quad j \in J_\nu.$$

3) Перебираются индексные цепочки $S = \{s_j\}_{j \in J_\nu}$, $s_j \in L_j^\nu$ (см. Дополнение D), пока не найдётся цепочка $\hat{S} = \{\hat{s}_j\}_{j \in J_\nu}$, на которой решение $\hat{V}_\nu$ экстремальной задачи

$$\hat{Q}(G_\nu + V) := \frac{1}{m} \sum_{i \in I_\nu} \varphi_i(G_\nu + V) + \frac{1}{k} \sum_{j \in J_\nu} \bigl[f(g_\nu^{\hat{s}_j} + v^{\hat{s}_j},\ \check{b}_j) \bigr]_+ \to \min_{V \in \Omega}$$

удовлетворяет неравенству

$$\hat{Q}(G_\nu + \hat{V}_\nu) < F(G_\nu) - \sigma. \tag{5.27}$$

4) В качестве очередного приближения берётся матрица

$$G_{\nu+1} = G_\nu + \hat{t}_\nu \hat{V}_\nu,$$

где $\hat{t}_\nu$ — точка минимума функции $F(G_\nu + t\hat{V}_\nu)$ на отрезке $[0,\ 1]$.

Замечание. Если цепочка $\hat{S}$, обеспечивающая неравенство (5.27), отсутствует, то G_ν — почти локально оптимальная матрица.

§ 6. Численные эксперименты

6.1. В этом параграфе приводятся результаты численных экспериментов по применению метода градиентного типа (§4) и безградиентного метода (§5) к решению задачи h-отделения двух множеств $\mathrm{co}(A)$ и B из $\mathbb{R}^n$, где

$$A = \{a_i\}_{i=1}^{m}, \quad B = \{b_j\}_{j=1}^{k}.$$

Общей проблемой обоих методов является выбор начального приближения. Представляется естественным такой подход.

В множестве $B_1 := B$ берём произвольную точку b_{j_1} и с помощью гиперплоскости $H_1 = \big\{x \mid \langle w^1, x\rangle = \gamma_1\big\}$ линейно отделяем её от $\mathrm{co}(A)$. Далее, в множестве $B_2 = \big\{b_j \in B_1 \mid \langle w^1, b_j\rangle < \gamma_1\big\}$ берём произвольную точку b_{j_2} и с помощью гиперплоскости $H_2 = \big\{x \mid \langle w^2, x\rangle = \gamma_2\big\}$ линейно отделяем её от $\mathrm{co}(A)$. Процесс продолжается до тех пор, пока не будет построена гиперплоскость $H_h = \big\{x \mid \langle w^h, x\rangle = \gamma_h\big\}$, линейно отделяющая точку $b_{j_h} \in B_h$ от $\mathrm{co}(A)$. Здесь

$$B_h = \big\{b_j \in B_{h-1} \mid \langle w^{h-1}, b_j\rangle < \gamma_{h-1}\big\}.$$

Матрицу G_0 со строками (w^s, γ_s), $s = 1, \ldots, h$, можно взять в качестве начального приближения.

На самом деле, в описанном процессе на p-м шаге вместо одной точки $b_{j_p} \in B_p$ можно брать несколько точек из B_p. Предельный случай выглядит так: множество B разбивается на h непересекающихся подмножеств и каждое подмножество, независимо, линейно отделяется от $\mathrm{co}(A)$.

6.2. Рассмотрим конкретный пример. Пусть на плоскости заданы множества A и B, состоящие соответственно из точек (см. рис. 6.1.)

$$a_1 = (-3, 0),\ a_2 = (-2, 0),\ a_3 = (-2, -3),\ a_4 = (0, 2),\ a_5 = (0, 1),$$

$$a_6 = (1, 1),\ a_7 = (2, 0),\ a_8 = (2, 4),\ a_9 = (3, 1),\ a_{10} = (4, -2);$$

$$b_1 = (-6, 0),\ b_2 = (-3.5, -1),\ b_3 = (-3, 2),\ b_4 = (-3, -3.5),$$

$$b_5 = (-2, -5),\ b_6 = (-1, 5),\ b_7 = (1, 5),\ b_8 = (2, -5),$$

$$b_9 = (3, 8),\ b_{10} = (4, -5),\ b_{11} = (5, 2),\ b_{12} = (6, -2).$$

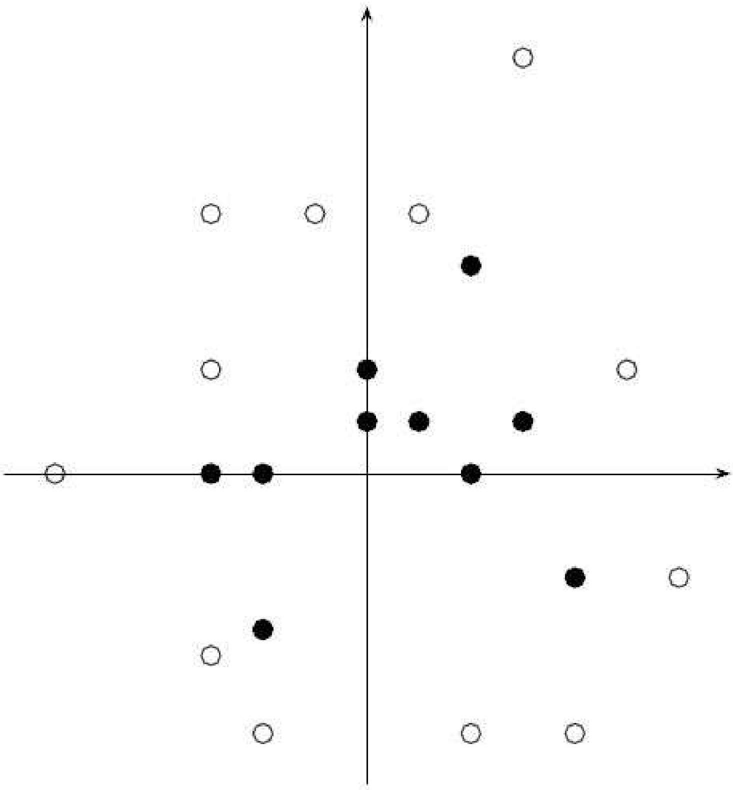

Рис. 6.1. Множества A и B

Для h-отделения этих множеств при $h = 4$ воспользуемся сначала методом градиентного типа.

Построим начальное приближение. Возьмём точку b_6 и линейно отделим её от множества A (см. § 1). Получим вектор

$$g_1 = (w_1, \gamma_1) = (-0.53,\ 0.85,\ 3.27),$$

определяющий гиперплоскость, линейно отделяющую точку b_6 от множества A (см. рис. 6.2).

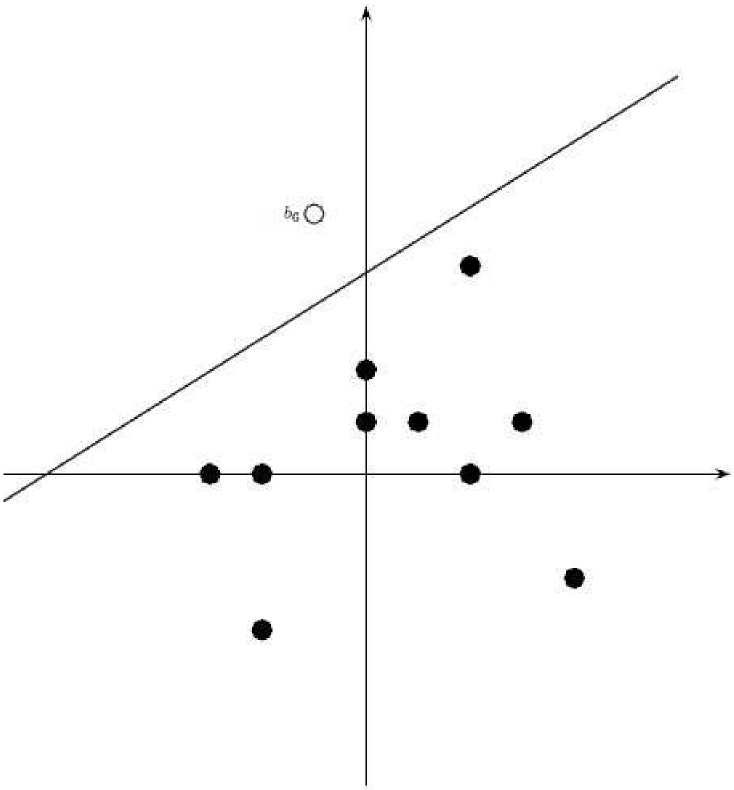

Рис. 6.2. Линейное отделение одной точки

Аналогично отделим точки b_2, b_8 и b_{12}. Получим соответственно векторы (см. рис. 6.3)

$$g_2 = (w_2, \gamma_2) = (-0.98,\ -018,\ 3.29),$$

$$g_3 = (w_3, \gamma_3) = (0.18,\ -0.98,\ 3.99),$$

$$g_4 = (w_4, \gamma_4) = (0.99,\ -0.11,\ 5.18).$$

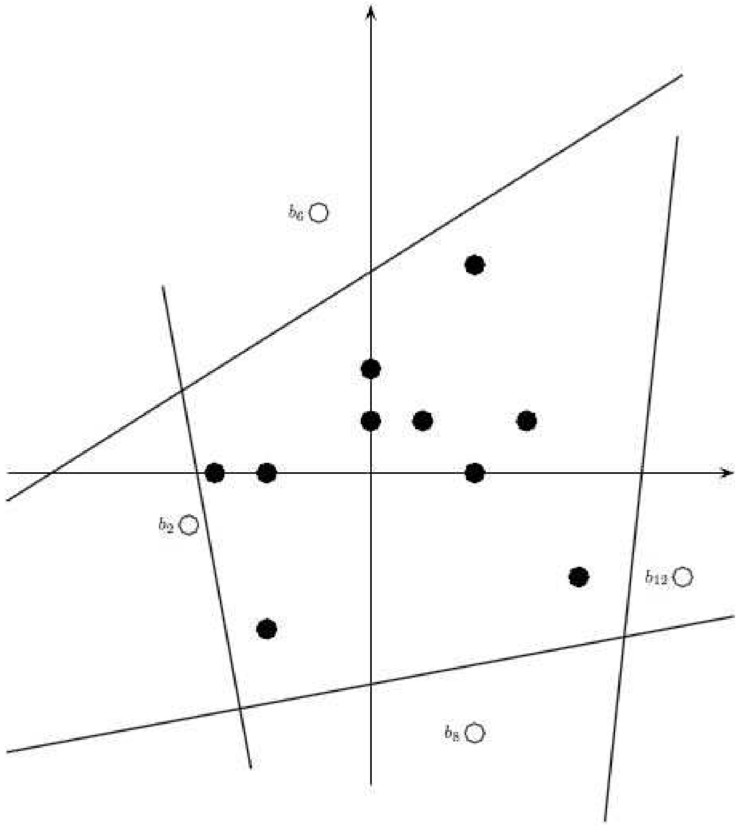

Рис. 6.3. Линейное отделение четырёх точек

Теперь в качестве начального приближения G_0 возьмем матрицу со строками g_1, g_2, g_3, g_4, так что (см. рис. 6.4)

$$G_0 = \begin{pmatrix} -0.53 & 0.85 & 3.27 \\ -0.98 & -0.18 & 3.29 \\ 0.18 & -0.98 & 3.99 \\ 0.99 & -0.11 & 5.18 \end{pmatrix}.$$

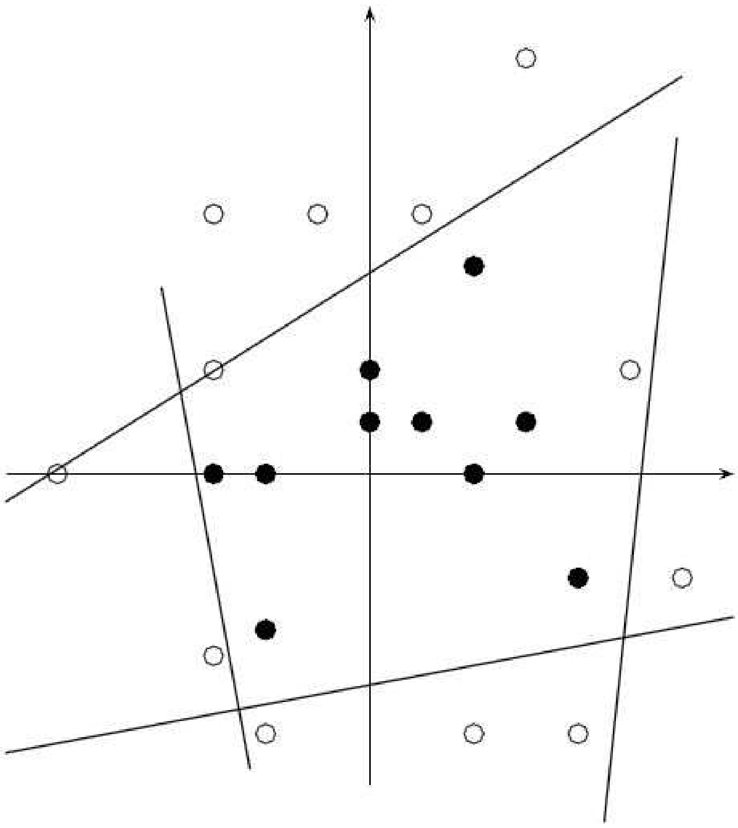

Рис. 6.4. Начальное приближение G_0

Будем решать задачу строгого 4-отделения методом градиентного типа при $c = 0.5$ (см. п. 4.3).

Заполним таблицу 7 и таблицу 8.

Таблица 7: Вычисление функций $\varphi_i(G_0)$

i	1	2	3	4	5	6	7	8	9	10
$f(g_0^1, \hat{a}_i)$	-1.17	-1.71	-4.24	-1.08	-1.93	-2.46	-3.84	-0.99	-3.53	-6.60
$f(g_0^2, \hat{a}_i)$	0.16	-0.82	-0.28	-3.15	-2.97	-3.95	-4.75	-6.46	-5.92	-6.36
$f(g_0^3, \hat{a}_i)$	-4.04	-3.86	-0.91	-5.46	-4.47	-4.29	-3.12	-6.87	-3.92	-0.79
$f(g_0^4, \hat{a}_i)$	-7.67	-6.67	-6.35	-4.89	-4.79	-3.79	-2.69	-2.12	-1.80	-0.49
$\hat{\varphi}_i(G_0)$	0.16	-0.82	-0.28	-1.08	-1.93	-2.46	-2.69	-0.99	-1.80	-0.49
$\hat{R}_i(G_0)$	$\{2\}$	$\{2\}$	$\{2\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{4\}$	$\{1\}$	$\{4\}$	$\{4\}$
$\varphi_i(G_0)$	0.16	0	0	0	0	0	0	0	0	0

Таблица 8: Вычисление функций $\psi_j(G_0)$

j	1	2	3	4	5	6	7	8	9	10	11	12
$f(g_0^1, \check{b}_j)$	0.57	2.75	0.48	5.13	6.94	−0.99	0.08	9.07	−1.39	10.14	4.75	8.66
$f(g_0^2, \check{b}_j)$	−2.11	0.16	1.19	0.20	0.92	3.70	5.67	4.85	8.18	6.82	9.06	9.33
$f(g_0^3, \check{b}_j)$	5.59	4.15	7.01	1.60	−0.05	9.59	9.22	−0.79	11.80	−1.16	5.54	1.42
$f(g_0^4, \check{b}_j)$	11.65	9.06	8.88	8.29	7.14	7.20	5.22	3.17	3.54	1.18	0.92	−0.49
$\check{\psi}_j(G_0)$	−2.11	0.16	0.48	0.20	−0.05	−0.99	0.08	−0.79	−1.39	−1.16	0.92	−0.49
$\check{R}_j(G_0)$	{2}	{2}	{1}	{2}	{3}	{1}	{1}	{3}	{1}	{3}	{4}	{4}
$\psi_j(G_0)$	0	0.16	0.48	0.20	0	0	0.08	0	0	0	0.92	0

Значение целевой функции $F(G_0)$ равно 0.17. Производная по направлению V функции $F(G)$ при $G = G_0$ принимает вид:

$$F'(G_0,\ V) = \frac{1}{10}\langle v^2, \hat{a}_1\rangle + \frac{1}{12}\Big(\langle v^2, \check{b}_2\rangle + \langle v^1, \check{b}_3\rangle + \langle v^2, \check{b}_4\rangle + \langle v^1, \check{b}_7\rangle + \langle v^4, \check{b}_{11}\rangle\Big).$$

Решаем задачу линейного программирования

$$F'(G_0,\ V) \to \min_{V\in\Omega},$$

где в качестве Ω возьмём множество (4×3)-матриц V, все элементы которых по модулю не превосходят 10 ($K = 10$). Очевидно, что решением этой задачи является матрица

$$V_0 = \begin{pmatrix} -10 & 10 & -10 \\ -10 & -10 & -10 \\ 0 & 0 & 0 \\ 10 & 10 & -10 \end{pmatrix};$$

при этом $F'(G_0,\ V_0) = -22.67$.

Получили направление наискорейшего убывания V_0 функции $F(G)$ при $G = G_0$. Минимум функции $F(G_0 + tV_0)$ на луче $t > 0$ достигается при $t_0 = 0.008$, поэтому (см. рис. 6.5)

$$G_1 := G_0 + t_0V_0 = \begin{pmatrix} -0.61 & 0.93 & 3.19 \\ -1.06 & -0.26 & 3.21 \\ 0.18 & -0.98 & 3.99 \\ 1.07 & -0.02 & 5.10 \end{pmatrix}.$$

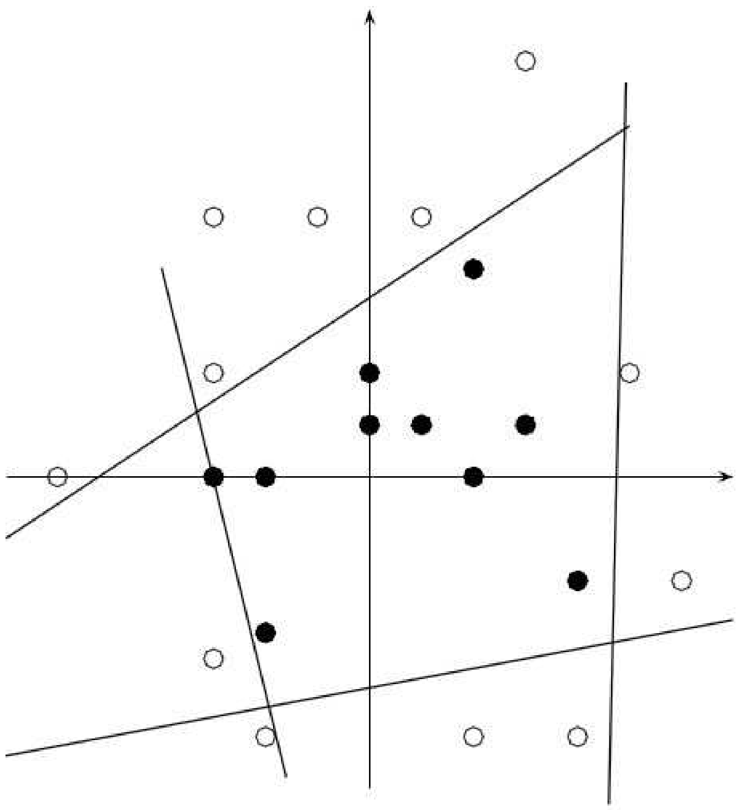

Рис. 6.5. Первое приближение G_1

Повторяем процесс. Заполним таблицу 9 и таблицу 10.

Таблица 9: Вычисление функций $\varphi_i(G_1)$

i	1	2	3	4	5	6	7	8	9	10
$f(g_0^1, \hat{a}_i)$	-0.85	-1.47	-4.24	-0.84	-1.77	-2.38	-3.92	-0.83	-3.61	-6.99
$f(g_0^2, \hat{a}_i)$	0.48	-0.58	0.20	-3.23	-2.97	-4.03	-4.83	-6.94	-6.16	-6.44
$f(g_0^3, \hat{a}_i)$	-4.04	-3.86	-0.91	-5.46	-4.47	-4.29	-3.12	-6.87	-3.92	-0.79
$f(g_0^4, \hat{a}_i)$	-7.82	-6.75	-6.67	-4.65	-4.62	-3.55	-2.45	-1.48	-1.40	-0.25
$\hat{\varphi}_i(G_0)$	0.48	-0.58	0.20	-0.84	-1.77	-2.38	-2.45	-0.83	-1.40	-0.25
$\hat{R}_i(G_0)$	$\{2\}$	$\{2\}$	$\{2\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{4\}$	$\{1\}$	$\{4\}$	$\{4\}$
$\varphi_i(G_0)$	0.48	0	0.20	0	0	0	0	0	0	0

Таблица 10: Вычисление функций $\psi_j(G_1)$

j	1	2	3	4	5	6	7	8	9	10	11	12
$f(g_0^1, \check{b}_j)$	0.01	2.47	0.001	5.09	7.09	−1.54	−0.32	9.55	−1.87	10.77	4.90	9.22
$f(g_0^2, \check{b}_j)$	−2.67	−0.27	1.03	−0.39	0.27	3.94	6.07	4.53	8.97	6.66	9.54	9.56
$f(g_0^3, \check{b}_j)$	5.59	4.15	7.01	1.60	−0.05	9.59	9.22	−0.79	11.80	−1.16	5.54	1.42
$f(g_0^4, \check{b}_j)$	12.04	9.33	8.87	8.73	7.62	6.80	4.65	3.32	2.58	1.17	0.28	−0.89
$\check{\psi}_j(G_0)$	−2.67	−0.27	0.001	−0.39	−0.05	−1.54	−0.32	−0.79	−1.87	−1.16	0.28	−0.89
$\check{R}_j(G_0)$	{2}	{2}	{1}	{2}	{3}	{1}	{1}	{3}	{1}	{3}	{4}	{4}
$\psi_j(G_0)$	0	0	0.001	0	0	0	0	0	0	0	0.28	0

Вычисляем $F(G_1) = 0.09$. Производная по направлению V функции $F(G)$ при $G = G_1$ принимает вид

$$F'(G_1,\ V) = \frac{1}{10}\Big(\langle v^2, \hat{a}_1\rangle + \langle v^2, \hat{a}_3\rangle\Big) + \frac{1}{12}\Big(\langle v^2, \check{b}_2\rangle + \langle v^1, \check{b}_3\rangle + \langle v^4, \check{b}_{11}\rangle\Big).$$

Решаем задачу линейного программирования

$$F'(G_1,\ V) \to \min_{V\in\Omega},$$

где в качестве Ω возьмём множество (4×3)-матриц V, все элементы которых по модулю не превосходят 10 ($K = 10$). Очевидно, что решением этой задачи является матрица

$$V_1 = \begin{pmatrix} -10 & 10 & -10 \\ 10 & 10 & 10 \\ 0 & 0 & 0 \\ 10 & 10 & -10 \end{pmatrix};$$

при этом $F'(G_1,\ V_1) = -21.67$.

Получили направление наискорейшего убывания V_1 функции $F(G)$ при $G = G_1$. Минимум функции $F(G_1 + tV_1)$ на луче $t > 0$ достигается при $t_1 = 0.005$, поэтому

$$G_2 := G_1 + t_1V_1 = \begin{pmatrix} -0.66 & 0.98 & 3.14 \\ -1.01 & -0.21 & 3.26 \\ 0.18 & -0.98 & 3.99 \\ 1.12 & 0.02 & 5.05 \end{pmatrix}$$

и $F(G_2) = 0.02$. Учитывая, что G_2 — подходящая матрица, заключаем, что G_2 — решение рассматриваемой задачи (см. рис. 6.6).

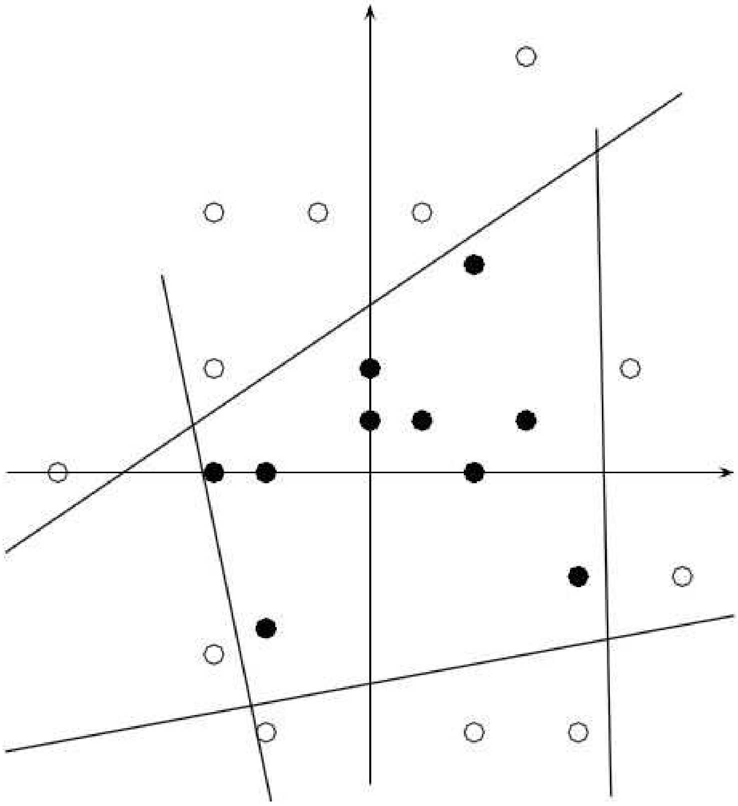

Рис. 6.6. Решение G_2

6.3. Теперь рассмотрим решение данного примера при тех же начальных условиях с помощью безградиентного метода (см. §5).

Положим $c = 0.5$, $\varepsilon_A = 0.1$, $\varepsilon_B = 0.1$ и $\sigma = 0.0001$. Строим начальное приближение как в п. 6.2. Получим

$$G_0 = \begin{pmatrix} -0.5333 & 0.8460 & 3.2735 \\ -0.9837 & -0.1801 & 3.2869 \\ 0.1837 & -0.9830 & 3.9915 \\ 0.9944 & -0.1055 & 5.1832 \end{pmatrix}$$

и $F(G_0) = 0.1706$.

Формируем индексные множества

$$I_0 = \bigl\{ i \in 1:m \mid \max_{s \in 1:h} f(g_\nu^s,\ \hat{a}_i) > -\varepsilon_A \bigr\},$$

$$J_0 = \bigl\{ j \in 1:k \mid \min_{s \in 1:h} f(g_\nu^s,\ \check{b}_j) > -\varepsilon_B \bigr\},$$

$$L_j^0 = \bigl\{ s \in 1:h \mid f(g_\nu^s,\ \check{b}_j) = \min_{p \in 1:h} f(g_\nu^p,\ \check{b}_j) \bigr\}, \quad j \in J_\nu.$$

Имеем (см. таблицу 7 и таблицу 8)

$$I_0 = \{1\},$$

$$J_0 = \{2,\ 3,\ 4,\ 5,\ 7,\ 11\},$$

$$L_2^0 = \{2\},\ L_3^0 = \{1\},\ L_4^0 = \{2\},\ L_5^0 = \{3\},\ L_7^0 = \{1\},\ L_{11}^0 = \{4\}.$$

Решаем задачу линейного программирования вида (5.19). Запишем вектор неизвестных

$$z = \bigl(\xi_1,\ \eta_2,\ \eta_3,\ \eta_4,\ \eta_5,\ \eta_7,\ \eta_{11},\ v^1(1),\ v^1(2),\ v^1(3),$$
$$v^2(1),\ v^2(2),\ v^2(3),\ v^3(1),\ v^3(2),\ v^3(3),\ v^4(1),\ v^4(2),\ v^4(3)\bigr),$$

матрицу ограничений

$$D = \begin{pmatrix}
1 & & & & & & & 3 & 0 & 1 & & & & & & & & & \\
1 & & & & & & & & & & 3 & 0 & 1 & & & & & & \\
1 & & & & & & & & & & & & & 3 & 0 & 1 & & & \\
1 & & & & & & & & & & & & & & & & 3 & 0 & 1 \\
 & 1 & & & & & & & & & -3.5 & -1 & -1 & & & & & & \\
 & & 1 & & & & & -3 & 2 & -1 & & & & & & & & & \\
 & & & 1 & & & & & & & -3 & -3.5 & -1 & & & & & & \\
 & & & & 1 & & & & & & & & & -2 & -5 & -1 & & & \\
 & & & & & 1 & & 1 & 5 & -1 & & & & & & & & & \\
 & & & & & & 1 & & & & & & & & & & 5 & 2 & -1
\end{pmatrix}$$

и вектор правой части ограничений

$$b = \bigl(-1.1736,\ 0.1642,\ -4.0426,\ -7.6664,$$
$$0.1638,\ 0.4816,\ 0.2054,\ -0.0561,\ 0.0768,\ 0.9222\bigr).$$

Тогда задача линейного программирования (5.19) примет вид

$$\tfrac{1}{10}\xi_1 + \tfrac{1}{12}(\eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_7 + \eta_{11}) \to \min,$$

$$Dz \geq b,$$

$$\xi_i \geq 0, \quad i \in I_0; \quad \eta_j \geq 0, \quad j \in J_0;$$

$$-10 \leq v^s(\alpha) \leq 10, \quad \alpha \in 1:3,\ s \in 1:4.$$

Решением этой задачи является матрица

$$V_0 = \begin{pmatrix} 0.0012 & 0.0070 & -0.0031 \\ 0.0025 & -0.0071 & 0.0006 \\ 0.0053 & -0.0066 & 0.0009 \\ 0.0050 & 0.0037 & -0.0040 \end{pmatrix},$$

при этом выполняется неравенство (см. п. 5.5)

$$\hat{Q}(G_0 + V_0) < F(G_0) - \sigma,$$

где $\hat{Q}(G_0 + V_0) = 0.1698$ и $\sigma = 0.0001$.

В качестве очередного приближения берём матрицу (см. рис. 6.7)

$$G_1 := G_0 + t_0 V_0 = \begin{pmatrix} -0.5101 & 0.9859 & 3.2113 \\ -0.9340 & -0.3225 & 3.2988 \\ 0.2892 & -1.1150 & 4.0099 \\ 1.0952 & -0.0314 & 5.1041 \end{pmatrix},$$

где $t_0 = 30$ — точка минимума функции $F(G_0 + tV_0)$ при $t_0 > 0$. Значение целевой функции $F(G_1)$ равно 0.0546.

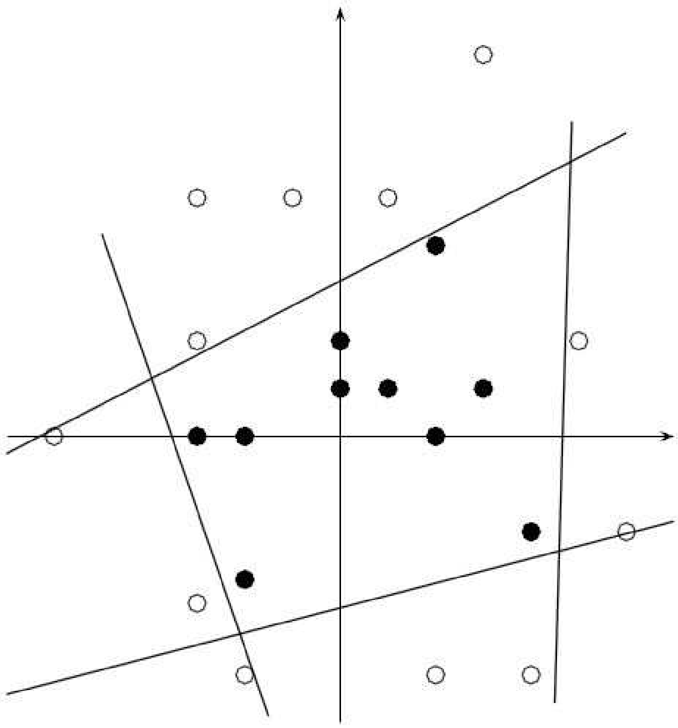

Рис. 6.7. Первое приближение G_1

Повторяем процесс. Формируем индексные множества

$$I_1 = \{1,\ 3\},$$

$$J_1 = \{2,\ 3,\ 11\},$$

$$L_2^1 = \{2\},\ L_3^1 = \{1\},\ L_{11}^1 = \{4\}.$$

Решаем задачу линейного программирования вида (5.19). Решением является матрица

$$V_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

так что $Q_1 = Q(G_1 + V_1) = Q(G_1) = F(G_1)$. По теореме 5.1 это гарантирует локальную оптимальность матрицы G_1.

6.4. Рассмотрим другой способ выбора начального приближения. Разобъем множество B на четыре подгруппы по три точки. Далее будем линейно отделять каждую подгруппу от множества A.

Построим начальное приближение. Возьмём точки $\{b_1,\ b_3,\ b_4\}$ и линейно отделим их от множества A (см. § 1). Получим вектор

$$g_1 = (w_1, \gamma_1) = (-1,\ 0,\ 3),$$

определяющий гиперплоскость, линейно отделяющую эти точки от множества A (см. рис. 6.8).

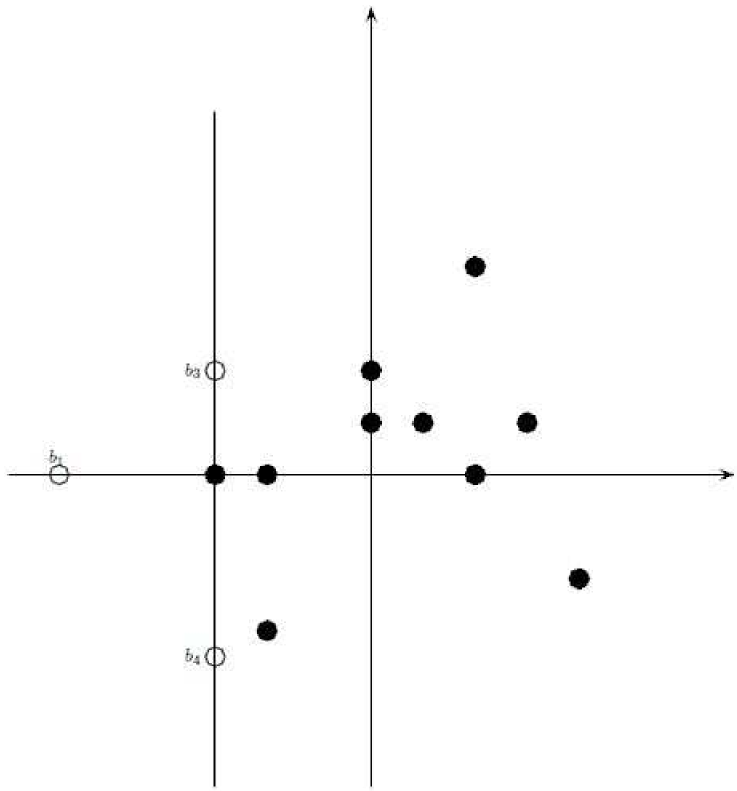

Рис. 6.8. Наилучшее линейное отделение первой группы точек

Далее возьмём точки $\{b_2,\ b_5,\ b_8\}$, получим вектор

$$g_2 = (w_2, \gamma_2) = (-0.59,\ -0.81,\ 3.23),$$

(см. рис. 6.9).

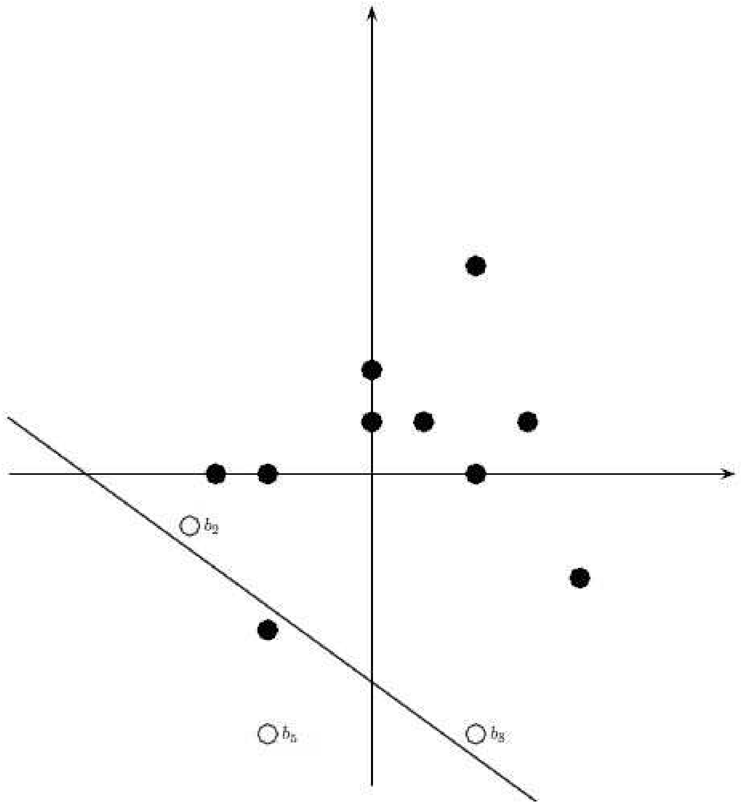

Рис. 6.9. Наилучшее линейное отделение второй группы точек

Аналогично отделяем точки $\{b_9,\ b_{10},\ b_{12}\}$, получаем вектор

$$g_3 = (w_3, \gamma_3) = (0.99,\ 0.08,\ 3.72),$$

и точки $\{b_6,\ b_7,\ b_{11}\}$, получаем вектор

$$g_4 = (w_4, \gamma_4) = (0.45,\ 0.89,\ 4.47)$$

(см. рис. 6.10 и рис. 6.11).

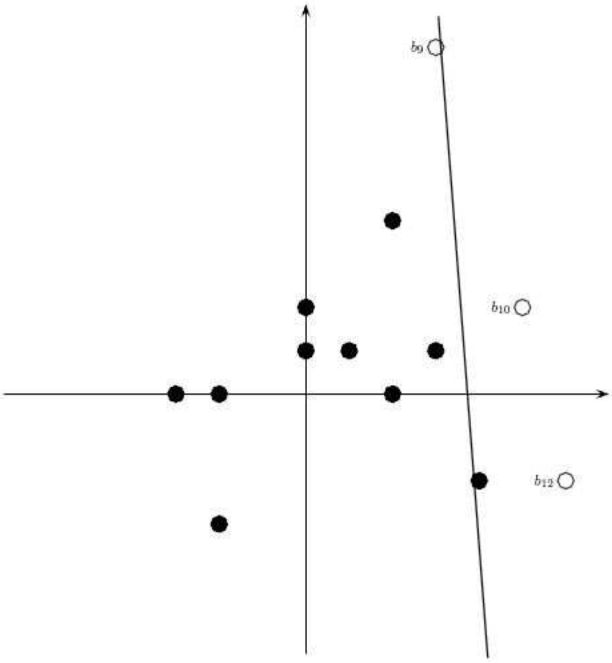

Рис. 6.10. Наилучшее линейное отделение третей группы точек

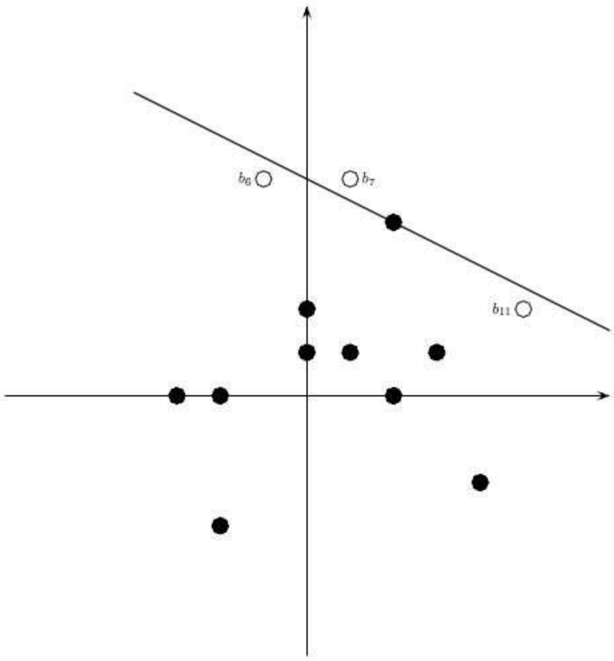

Рис. 6.11. Наилучшее линейное отделение четвертой группы точек

Теперь в качестве начального приближения возьмем матрицу G_0 со строками g_1, g_2, g_3, g_4, так что (см. рис. 6.12)

$$G_0 = \begin{pmatrix} -1 & 0 & 3 \\ -0.59 & -0.81 & 3.23 \\ 0.99 & 0.08 & 3.72 \\ 0.45 & 0.89 & 4.47 \end{pmatrix}.$$

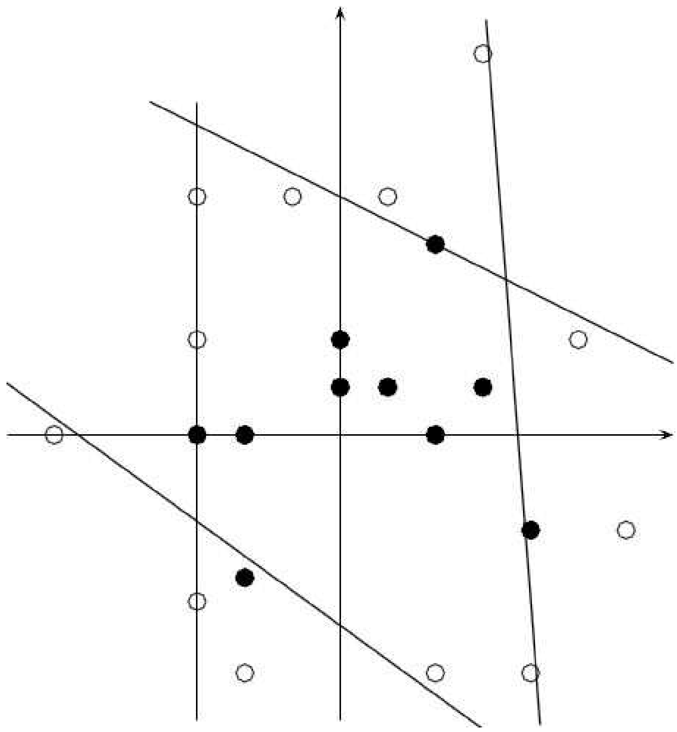

Рис. 6.12. Начальное приближение G_0

Будем решать задачу строгого 4-отделения при $c = 0.5$ методом градиентного типа (см. п. 4.3).

Заполним таблицу 11 и таблицу 12.

Таблица 11: Вычисление функций $\varphi_i(G_0)$

i	1	2	3	4	5	6	7	8	9	10
$f(g_0^1, \hat{a}_i)$	0.50	-0.50	-0.50	-2.50	-2.50	-3.50	-4.50	-5.50	-5.50	-6.50
$f(g_0^2, \hat{a}_i)$	-0.97	-1.56	0.87	-4.35	-3.54	-4.13	-3.91	-7.73	-5.31	-3.47
$f(g_0^3, \hat{a}_i)$	-6.21	-5.21	-5.44	-3.07	-3.14	-2.15	-1.22	0.08	-0.15	0.61
$f(g_0^4, \hat{a}_i)$	-5.31	-4.87	-7.55	-2.18	-3.08	-2.63	-3.08	0.95	-1.74	-3.97
$\hat{\varphi}_i(G_0)$	0.50	-0.50	0.87	-2.18	-2.50	-2.15	-1.22	0.95	-0.15	0.61
$\hat{R}_i(G_0)$	$\{1\}$	$\{1\}$	$\{2\}$	$\{4\}$	$\{1\}$	$\{3\}$	$\{3\}$	$\{4\}$	$\{3\}$	$\{3\}$
$\varphi_i(G_0)$	0.50	0	0.87	0	0	0	0	0.95	0	0.61

Таблица 12: Вычисление функций $\psi_j(G_0)$

j	1	2	3	4	5	6	7	8	9	10	11	12
$f(g_0^1, \check{b}_j)$	-2.50	0	0.50	0.50	1.50	2.50	4.50	5.50	6.50	7.50	8.50	9.50
$f(g_0^2, \check{b}_j)$	0.21	0.87	3.59	-0.86	-1.48	7.19	8.36	0.87	11.97	2.04	8.29	5.65
$f(g_0^3, \check{b}_j)$	10.20	7.78	7.06	7.48	6.60	4.83	2.84	2.61	0.61	0.61	-0.92	-1.61
$f(g_0^4, \check{b}_j)$	7.65	7.43	4.52	9.44	10.33	0.95	0.05	8.55	-3.52	7.65	0.95	4.08
$\check{\psi}_j(G_0)$	-2.50	0	0.50	-0.86	-1.48	0.95	0.05	0.87	-3.52	0.61	-0.92	-1.61
$\check{R}_j(G_0)$	$\{1\}$	$\{1\}$	$\{1\}$	$\{2\}$	$\{2\}$	$\{4\}$	$\{4\}$	$\{2\}$	$\{4\}$	$\{3\}$	$\{3\}$	$\{3\}$
$\psi_j(G_0)$	0	0	0.50	0	0	0.95	0.05	0.87	0	0.61	0	0

Значение целевой функции $F(G_0)$ равно 0.54. Производная по направлению V функции $F(G)$ при $G = G_0$ принимает вид:

$$F'(G_0,\ V) = \frac{1}{10}\Big(\langle v^1, \hat{a}_1\rangle + \langle v^2, \hat{a}_3\rangle + \langle v^4, \hat{a}_8\rangle + \langle v^3, \hat{a}_{10}\rangle\Big) +$$
$$+ \frac{1}{12}\Big(\big[\langle v^1, \check{b}_2\rangle\big]_+ + \langle v^1, \check{b}_3\rangle + \langle v^4, \check{b}_6\rangle + \langle v^4, \check{b}_7\rangle + \langle v^2, \check{b}_8\rangle + \langle v^3, \check{b}_{10}\rangle\Big).$$

Решаем задачу линейного программирования

$$F'(G_0,\ V) \to \min_{V \in \Omega},$$

где в качестве Ω возьмём множество (4×3)-матриц V, все элементы которых по модулю не превосходят 10 ($K = 10$). Решением этой задачи является матрица

$$V_0 = \begin{pmatrix} -5.71 & 10 & 10 \\ 10 & -10 & 10 \\ 10 & -10 & 10 \\ -10 & 10 & -10 \end{pmatrix};$$

при этом $F'(G_0,\ V_0) = -20.22$.

Получили направление наискорейшего убывания V_0 функции $F(G)$ при $G = G_0$. Минимум функции $F(G_0 + tV_0)$ на луче $t > 0$ достигается при $t_0 = 0.01$, поэтому

$$G_1 := G_0 + t_0 V_0 = \begin{pmatrix} -1.06 & 0.1 & 3.1 \\ -0.49 & -0.91 & 3.33 \\ 1.09 & -0.02 & 3.82 \\ 0.35 & 0.99 & 4.37 \end{pmatrix},$$

$F(G_1) = 0.43$ (см. рис. 6.13).

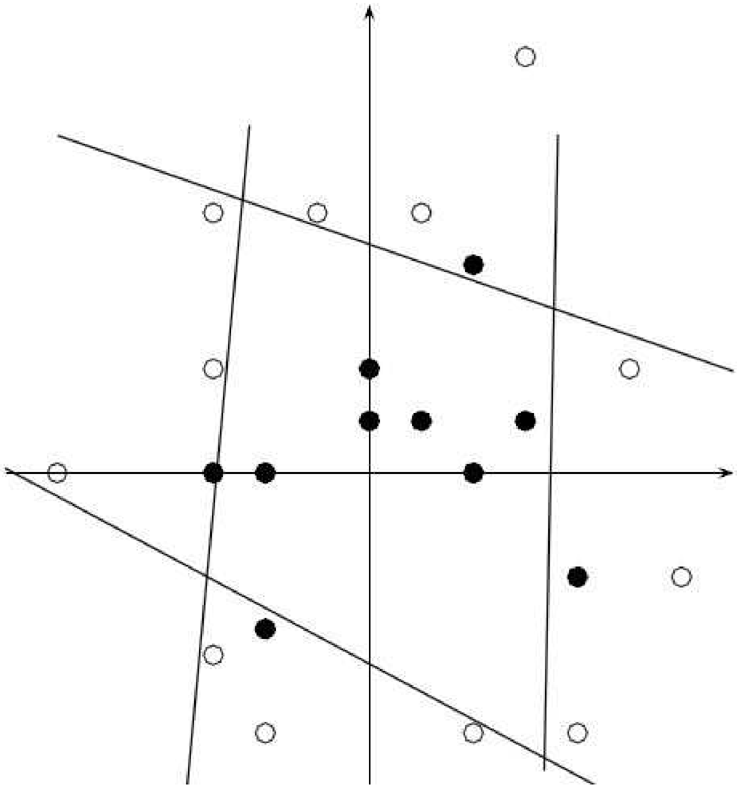

Рис. 6.13. Первое приближение G_1

Повторяем процесс. На следующем шаге получаем матрицу

$$G_2 = \begin{pmatrix} -1.26 & 0.3 & 2.9 \\ -0.29 & -1.11 & 3.53 \\ 0.90 & 0.18 & 4.02 \\ 0.15 & 1.19 & 4.57 \end{pmatrix},$$

$F(G_2) = 0.35$ (см. рис. 6.14).

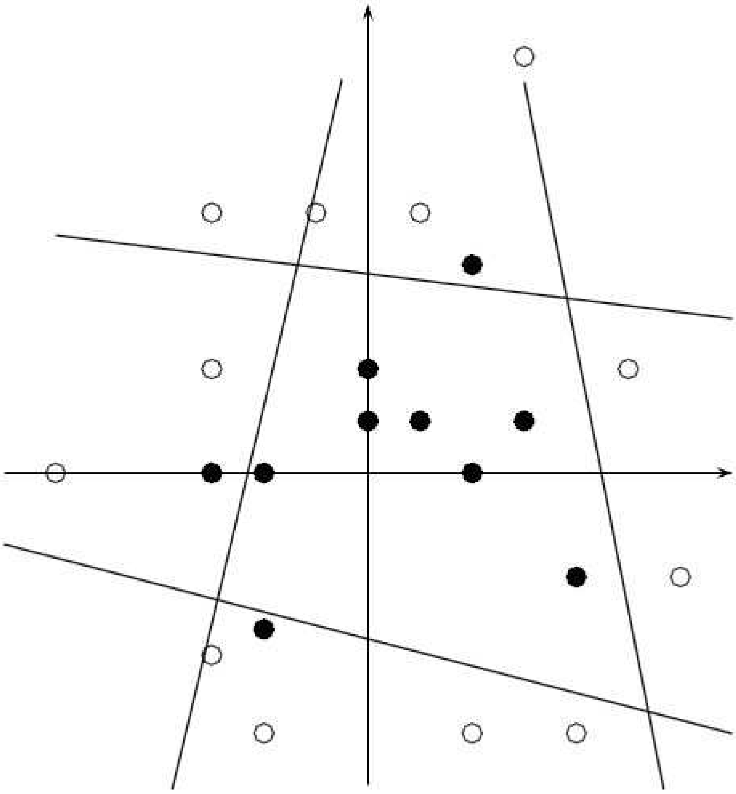

Рис. 6.14. Второе приближение G_2

Повторяем процесс. На следующем шаге получаем матрицу

$$G_3 = \begin{pmatrix} -1.16 & 0.3 & 3 \\ -0.19 & -1.01 & 3.63 \\ 0.90 & 0.18 & 4.02 \\ 0.05 & 1.09 & 4.67 \end{pmatrix},$$

$F(G_3) = 0.16$ (см. рис. 6.15).

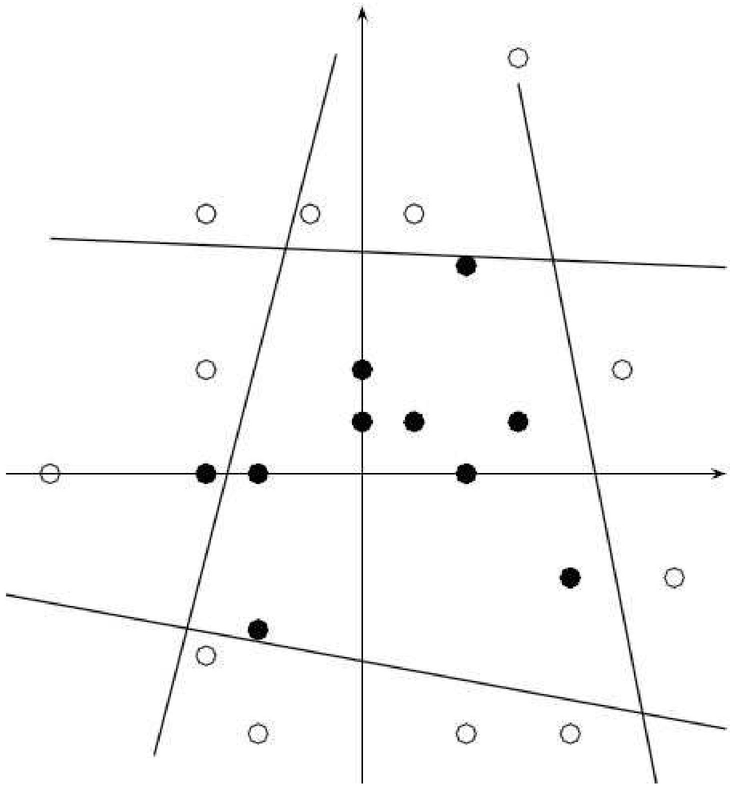

Рис. 6.15. Третье приближение G_3

6.5. Далее рассмотрим решение данного примера при тех же начальных условиях, что и в п. 6.4, с помощью метода из §5.

Положим $c = 0.5$, $\varepsilon_A = 0.1$, $\varepsilon_B = 0.1$ и $\sigma = 0.0001$. Строим начальное приближение как в п. 6.4. Получим

$$G_0 = \begin{pmatrix} -1 & 0 & 3 \\ -0.5882 & -0.8087 & 3.2349 \\ 0.9971 & 0.0767 & 3.7198 \\ 0.4472 & 0.8944 & 4.4721 \end{pmatrix}$$

и $F(G_0) = 0.5416$.

Формируем индексные множества

$$I_0 = \bigl\{i \in 1:m \mid \max_{s \in 1:h} f(g_\nu^s,\ \hat{a}_i) > -\varepsilon_A\bigr\},$$

$$J_0 = \bigl\{j \in 1:k \mid \min_{s \in 1:h} f(g_\nu^s,\ \check{b}_j) > -\varepsilon_B\bigr\},$$

$$L_j^0 = \bigl\{s \in 1:h \mid f(g_\nu^s,\ \check{b}_j) = \min_{p \in 1:h} f(g_\nu^p,\ \check{b}_j)\bigr\}, \quad j \in J_\nu.$$

Имеем (см. таблицу 11 и таблицу 12)

$$I_0 = \{1,\ 3,\ 8,\ 10\},$$

$$J_0 = \{2,\ 3,\ 6,\ 7,\ 8,\ 10\},$$

$$L_2^0 = \{1\},\ L_3^0 = \{1\},\ L_6^0 = \{4\},\ L_7^0 = \{4\},\ L_8^0 = \{2\},\ L_{10}^0 = \{3\}.$$

Решаем задачу линейного программирования вида (5.19). Запишем вектор неизвестных

$$z = \big(\xi_1,\ \xi_3,\ \xi_8,\ \xi_{10},\ \eta_2,\ \eta_3,\ \eta_6,\ \eta_7,\ \eta_8,\ \eta_{10},\ v^1(1),\ v^1(2),\ v^1(3),$$
$$v^2(1),\ v^2(2),\ v^2(3),\ v^3(1),\ v^3(2),\ v^3(3),\ v^4(1),\ v^4(2),\ v^4(3)\big),$$

матрицу ограничений

$$D = \begin{pmatrix}
1 & & & & & & & & & & 3 & 0 & 1 & & & & & & & & & \\
1 & & & & & & & & & & & & & 3 & 0 & 1 & & & & & & \\
1 & & & & & & & & & & & & & & & & 3 & 0 & 1 & & & \\
1 & & & & & & & & & & & & & & & & & & & 3 & 0 & 1 \\
 & 1 & & & & & & & & & 2 & 3 & 1 & & & & & & & & & \\
 & 1 & & & & & & & & & & & & 2 & 3 & 1 & & & & & & \\
 & 1 & & & & & & & & & & & & & & & 2 & 3 & 1 & & & \\
 & 1 & & & & & & & & & & & & & & & & & & 2 & 3 & 1 \\
 & & 1 & & & & & & & & -3 & -4 & 1 & & & & & & & & & \\
 & & 1 & & & & & & & & & & & -3 & -4 & 1 & & & & & & \\
 & & 1 & & & & & & & & & & & & & & -3 & -4 & 1 & & & \\
 & & 1 & & & & & & & & & & & & & & & & & -3 & -4 & 1 \\
 & & & 1 & & & & & & & -4 & 2 & 1 & & & & & & & & & \\
 & & & 1 & & & & & & & & & & -4 & 2 & 1 & & & & & & \\
 & & & 1 & & & & & & & & & & & & & -4 & 2 & 1 & & & \\
 & & & 1 & & & & & & & & & & & & & & & & -4 & 2 & 1 \\
 & & & & 1 & & & & & & -3.5 & -1 & -1 & & & & & & & & & \\
 & & & & & 1 & & & & & -3 & 2 & -1 & & & & & & & & & \\
 & & & & & & 1 & & & & & & & & & & & & & -1 & 5 & -1 \\
 & & & & & & & 1 & & & & & & & & & & & & 1 & 5 & -1 \\
 & & & & & & & & 1 & & & & & 2 & -5 & -1 & & & & & & \\
 & & & & & & & & & 1 & & & & & & & 4 & -5 & -1 & & &
\end{pmatrix}$$

и вектор правой части ограничений

$$b = \big(0.5000,\ -0.9703,\ -6.2111,\ -5.3137,\ -0.5000,\ 0.8676,\ -5.4441,$$
$$-7.5497,\ -5.5000,\ -7.7343,\ 0.0783,\ 0.9471,\ -6.5000,\ -3.4703,\ 0.6152,$$
$$-3.9721,\ 0,\ 0.5000,\ 0.9473,\ 0.0529,\ 0.8678,\ 0.6149\big).$$

Тогда задача линейного программирования (5.19) примет вид

$$\tfrac{1}{10}(\xi_1+\xi_3+\xi_8+\xi_{10}) + \tfrac{1}{12}(\eta_2+\eta_3+\eta_6+\eta_7+\eta_8+\eta_{10}) \to \min,$$
$$Dz \geq b,$$
$$\xi_i \geq 0, \quad i \in I_0; \quad \eta_j \geq 0, \quad j \in J_0;$$
$$-10 \leq v^s(\alpha) \leq 10, \quad \alpha \in 1:3,\ s \in 1:4.$$

Решением этой задачи является матрица

$$V_0 = \begin{pmatrix} -0.0061 & 0.0044 & 0.0007 \\ -0.0001 & -0.0031 & 0.0066 \\ -0.0016 & -0.0046 & 0.0070 \\ -0.0062 & 0.0068 & 0.0022 \end{pmatrix},$$

при этом выполняется неравенство (см. п. 5.5)

$$\hat{Q}(G_0+V_0) < F(G_0) - \sigma,$$

где $\hat{Q}(G_0+V_0) = 0.5382$ и $\sigma = 0.0001$.

В качестве очередного приближения берём матрицу (см. рис. 6.16)

$$G_1 := G_0 + t_0 V_0 = \begin{pmatrix} -1.1826 & 0.1310 & 3.0208 \\ -0.5904 & -0.9010 & 3.4325 \\ 0.9494 & -0.0625 & 3.9284 \\ 0.2624 & 1.0993 & 4.5388 \end{pmatrix},$$

где $t_0 = 30$ — точка минимума функции $F(G_0 + tV_0)$ при $t_0 > 0$. Значение целевой функции $F(G_1)$ равно 0.4391.

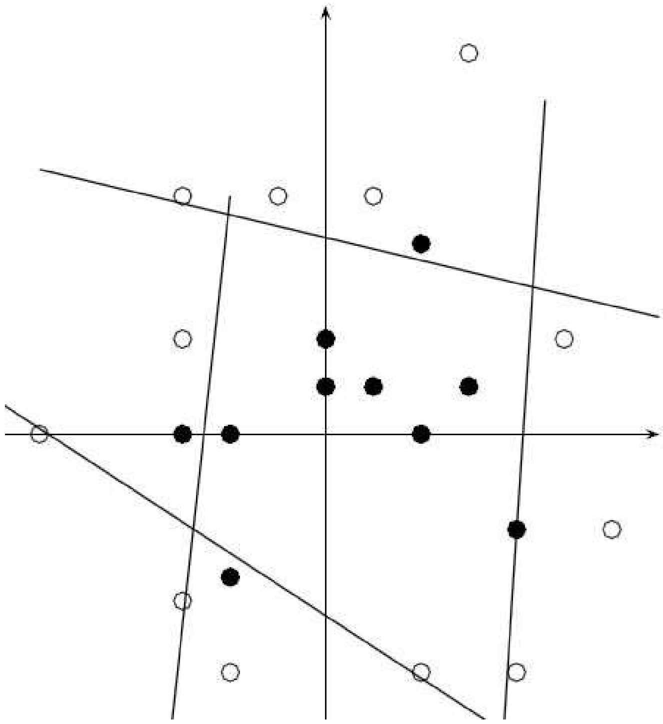

Рис. 6.16. Первое приближение G_1

Повторяем процесс. Формируем индексные множества

$$I_1 = \{1,\ 3,\ 8,\ 10\},$$

$$J_1 = \{8,\ 10\},$$

$$L_8^1 = \{2\},\ L_{10}^1 = \{3\}.$$

Решаем задачу линейного программирования вида (5.19). Решением является матрица

$$V_1 = \begin{pmatrix} 0.0004 & -0.0036 & 0.0034 \\ 0.0009 & -0.0028 & 0.0001 \\ -0.0001 & -0.0035 & 0.0004 \\ -0.0002 & -0.0027 & 0.0015 \end{pmatrix}$$

и $\hat{Q}(G_1 + V_1) = 0.4392$. Таким образом, G_1 — почти локально оптимальная матрица при $\sigma = 0.0001$.

ДОПОЛНЕНИЕ А

Двойственность в линейном программировании

А.1. Рассмотрим задачу линейного программирования

$$\begin{gathered} f(x) := c[N] \times x[N] \to \min \\ A[M_1,\ N] \times x[N] \geq b[M_1], \\ A[M_2,\ N] \times x[N] = b[M_2], \\ x[N_1] \geq \mathbb{O}[N_1], \end{gathered} \tag{А.1}$$

где $N_1 \subset N$. Вектор $x = x[N]$, удовлетворяющий ограничениям задачи (А.1), называется *планом*. Множество планов обозначим Ω. Требуется найти *оптимальный план* — вектор $x_* \in \Omega$, на котором целевая функция $f(x)$ принимает наименьшее на Ω значение (см. например [3, с. 16]).

Теорема А.1. *Оптимальный план существует тогда и только тогда, когда множество планов Ω непусто и целевая функция $f(x)$ ограничена снизу на Ω.*

А.2. Обозначим $M = M_1 \cup M_2$, $N_2 = N \setminus N_1$ и запишем двойственную задачу линейного программирования

$$\begin{gathered} g(u) := b[M] \times u[M] \to \max \\ u[M] \times A[M,\ N_1] \leq c[N_1], \\ u[M] \times A[M,\ N_2] = c[N_2], \\ u[M_1] \geq \mathbb{O}[M_1]. \end{gathered} \tag{А.2}$$

Множество планов задачи (А.2) обозначим Λ.

Первая теорема двойственности. *Из существования оптимального плана у одной из двойственных задач* (A.1), (A.2) *следует существование оптимального плана и у другой задачи. При этом справедливо соотношение двойственности*

$$\min_{x\in\Omega} f(x) = \max_{u\in\Lambda} g(u).$$

Следствие. *Для того чтобы планы $x_0 \in \Omega$, $u_0 \in \Lambda$ двойственных задач были оптимальными, необходимо и достаточно, чтобы выполнялось равенство $f(x_0) = g(u_0)$.*

A.3. Обычно исследуется пара двойственных задач линейного программирования и результаты формируются одновременно для прямой и двойственной задачи.

Теорема A.2. *Для того чтобы обе задачи* (A.1) *и* (A.2) *имели оптимальные планы, необходимо и достаточно, чтобы множества их планов Ω и Λ были непустыми.*

Вторая теорема двойственности. *Планы x_0, u_0 двойственных задач* (A.1) *и* (A.2) *являются оптимальными тогда и только тогда, когда выполняются условия дополнительности*

$$\begin{aligned} \big(A[i,\ N] \times x_0[N] - b[i]\big) \times u_0[i] = 0, &\quad i \in M_1; \\ x_0[j] \times \big(c[j] - u[M] \times A[M,\ j]\big) = 0, &\quad j \in N_1 \end{aligned}$$

A.4. Если x_0 — оптимальный план задачи (A.1), то всё множество оптимальных планов описывается системой линейных соотношений

$$
\begin{aligned}
c[N] \times x[N] &= c[N] \times x_0[N], \\
A[M_1,\ N] \times x[N] &\geq b[M_1], \\
A[M_2,\ N] \times x[N] &= b[M_2], \\
x[N_1] &\geq \mathbb{O}[N_1].
\end{aligned} \tag{A.3}
$$

Иногда требуется найти другой оптимальный план, отличный от x_0. Это можно сделать, введя подходящую целевую функцию

$$h(x) = d[N] \times x[N]$$

и минимизируя её при ограничениях (A.3).

A.5. Доказательство приведенных в этом Дополнении фактов имеется, например, в [3, глава 1].

A.6. Для решения задач линейного программирования разработаны эффективные численные методы. В пакете MATLAB задействованы два метода: метод внутренней точки (Large-Scale Algorithm) и вариант симплекс-метода (Medium-Scale Algorithm). По умолчанию используется метод внутренней точки. Программа реализована в виде функции Linprog. Описанию возможностей этой функции (с большим количеством примеров) посвящена работа [13].

ДОПОЛНЕНИЕ В

Свойства плюсиковой функции

Имеются в виду свойства функции $[x]_+ = \max\{0, x\}$.

1) $x \le [x]_+ \le |x|$

2) $[x+y]_+ \le [x]_+ + [y]_+$

Доказательство. Согласно 1) имеем

$$x + y \le [x]_+ + [y]_+.$$

Вместе с тем, $0 \le [x]_+ + [y]_+$. Значит,

$$[x+y]_+ = \max\{0, x+y\} \le [x]_+ + [y]_+.$$

3) $\big|[x]_+ - [y]_+\big| \le |x-y|$;

Доказательство. Имеем

$$x = y + (x-y) \le [y]_+ + |x-y|,$$

так что $[x]_+ \le [y]_+ + |x-y|$. Аналогично $[y]_+ \le [x]_+ + |y-x|$. Из двух последних неравенств следует требуемое.

4) $[cx]_+ = c[x]_+$ при $c > 0$.

Доказательство. При $x \le 0$ равенство принимает вид $0 = 0$, а при $x > 0$ соответственно $cx = cx$.

5) $\max\limits_{s \in 1:h} [x_s]_+ = [\max\limits_{s \in 1:h} x_s]_+$.

Доказательство. Если $\max\limits_{s \in 1:h} x_s \le 0$, то утверждение очевидно ($0 = 0$). В противном случае и левая и правая части доказываемого соотношения равны $\max\limits_{s \in 1:h} x_s$.

6) $\min\limits_{s\in 1:h}[x_s]_+ = [\min\limits_{s\in 1:h} x_s]_+$.

Доказательство. Если $\min\limits_{s\in 1:h} x_s > 0$, то обе части доказываемого соотношения равны $\min\limits_{s\in 1:h} x_s$. В противном случае утверждение очевидно $(0 = 0)$.

7) $\max\limits_{s\in 1:h}[x_s + y_s]_+ \leq \max\limits_{s\in 1:h}[x_s]_+ + \max\limits_{s\in 1:h}[y_s]_+$.

Доказательство следует из свойства 2).

8) $\left|\max\limits_{s\in 1:h}[x_s]_+ - \max\limits_{s\in 1:h}[y_s]_+\right| \leq \max\limits_{s\in 1:h}|x_s - y_s|$.

Доказательство. Воспользоваться отмеченным ранее неравенством

$$[x_s]_+ \leq [y_s]_+ + |x_s - y_s|.$$

9) $\min\limits_{s\in 1:h}[x_s]_+ \leq \min\limits_{s\in 1:h}[y_s]_+ + \max\limits_{s\in 1:h}|x_s - y_s|$.

Доказательство. В случае $\min\limits_{s\in 1:h} x_s \leq 0$ утверждение в силу свойства 6) очевидно (в левой части стоит ноль). Пусть все x_s положительны. Тогда

$$[x_s]_+ = x_s = y_s + (x_s - y_s) \leq [y_s]_+ + \max_{s\in 1:h}|x_s - y_s|.$$

Отсюда следует, что

$$\min_{s\in 1:h}[x_s]_+ \leq [y_s]_+ + \max_{s\in 1:h}|x_s - y_s|$$

и

$$[y_s]_+ \geq \min_{s\in 1:h}[x_s]_+ - \max_{s\in 1:h}|x_s - y_s|.$$

Взяв в левой части минимум по $s \in 1 : h$, придём к неравенству, которое равносильно требуемому.

10) $\left|\min\limits_{s\in 1:h}[x_s]_+ - \min\limits_{s\in 1:h}[y_s]_+\right| \leq \max\limits_{s\in 1:h}|x_s - y_s|$.

Доказательство основано на свойстве 9).

11) При $c > 0$ справедливо неравенство

$$\max_{s\in 1:h}[c - x_s]_+ + \min_{s\in 1:h}[c + x_s]_+ \geq 2c.$$

Равенство достигается только тогда, когда величина $x_* = \min\limits_{s\in 1:h} x_s$ принадлежит отрезку $[-c,\ c]$.

Доказательство. Очевидно, что функция $[c - x]_+$ монотонно убывает (не строго), поэтому

$$\max_{s\in 1:h}[c - x_s]_+ = [c - x_*]_+.$$

Аналогично, функция $[c + x]_+$ монотонно возрастает (не строго), поэтому

$$\min_{s\in 1:h}[c + x_s]_+ = [c + x_*]_+.$$

Остаётся учесть, что

$$[c - x_*]_+ + [c + x_*]_+ \geq 2c,$$

причём равенство достигается только тогда, когда $x_* \in [-c,\ c]$ (см. рис. B).

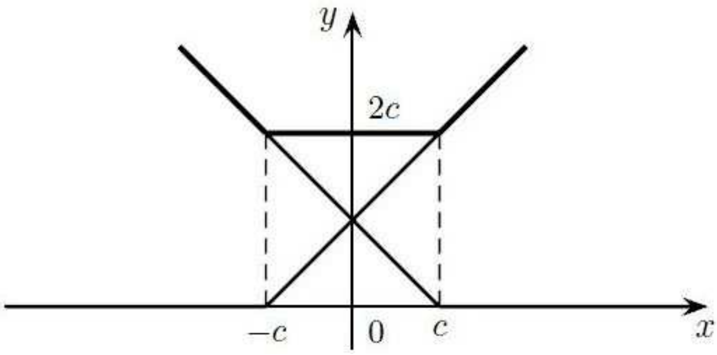

Рис. B. График функции $y = [c - x]_+ + [c + x]_+$

ДОПОЛНЕНИЕ C

Производные по направлению

C.1. Возьмём конкретную функцию

$$f(v,\ u) = \langle v,\ u\rangle + c, \quad u,\ v \in \mathbb{R}^{n+1},$$

и точку $\widehat{a} \in \mathbb{R}^{n+1}$. Составим новые функции

$$\widehat{\varphi}(G) = \max_{s\in 1:h} f(g^s,\ \widehat{a}) \quad \text{и} \quad \varphi(G) = \big[\widehat{\varphi}(G)\big]_+.$$

Здесь G — матрица со строками $g^1, \dots,\ g^h$. Нас интересует производная функции $\varphi(G)$ по направлению V, которая обозначается $\varphi'(G,\ V)$ и определяется следующим образом:

$$\varphi'(G,\ V) = \lim_{t\to+0} \frac{\varphi(G+tV) - \varphi(G)}{t}.$$

Обозначим

$$\widehat{R}(G) = \big\{s \in 1:h \mid f(g^s,\ \widehat{a}) = \widehat{\varphi}(G)\big\}.$$

Теорема C.1. *Для производной по направлению $\varphi'(G,\ V)$ справедлива формула*

$$\varphi'(G,\ V) = \begin{cases} \max\limits_{s\in\widehat{R}(G)} \langle v^s,\ \widehat{a}\rangle, & \text{если } \widehat{\varphi}(G) > 0, \qquad \text{(C.1)} \\ \max\limits_{s\in\widehat{R}(G)} \big[\langle v^s,\ \widehat{a}\rangle\big]_+, & \text{если } \widehat{\varphi}(G) = 0, \qquad \text{(C.2)} \\ 0, & \text{если } \widehat{\varphi}(G) < 0. \qquad \text{(C.3)} \end{cases}$$

C.2. Для доказательства этой теоремы нам потребуется одно вспомогательное утверждение.

Лемма. *При фиксированных G и V и малых $t > 0$ выполняется равенство*

$$\max_{s \in 1:h} f(g^s + tv^s,\ \widehat{a}) = \max_{s \in \widehat{R}(G)} f(g^s + tv^s,\ \widehat{a}).$$

Доказательство. Достаточно проверить, что в условиях леммы

$$\min_{s \in \widehat{R}(G)} f(g^s + tv^s,\ \widehat{a}) > \max_{s \notin \widehat{R}(G)} f(g^s + tv^s,\ \widehat{a}). \tag{C.4}$$

Обозначим

$$\widehat{\varepsilon} = \widehat{\varphi}(G) - \max_{s \notin \widehat{R}(G)} f(g^s,\ \widehat{a}).$$

Ясно, что $\widehat{\varepsilon} > 0$. Имеем

$$f(g^s + tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a}) = t\langle v^s,\ \widehat{a}\rangle,$$

поэтому при малых $t > 0$ будет выполнятся неравенство

$$\bigl|f(g^s + tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a})\bigr| \le \frac{\widehat{\varepsilon}}{4} \quad \forall s \in 1:h. \tag{C.5}$$

При $s \notin \widehat{R}(G)$ согласно определению $\widehat{\varepsilon}$ и (C.5) получаем

$$\begin{gathered} f(g^s + tv^s,\ \widehat{a}) = f(g^s,\ \widehat{a}) + \bigl[f(g^s + tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a})\bigr] \le \\ \le \bigl(\widehat{\varphi}(G) - \widehat{\varepsilon}\bigr) + \frac{\widehat{\varepsilon}}{4} = \widehat{\varphi}(G) - \frac{3\widehat{\varepsilon}}{4}, \end{gathered}$$

в то время как при $i \in \widehat{R}(G)$

$$f(g^s + tv^s,\ \widehat{a}) \ge f(g^s,\ \widehat{a}) + \bigl|f(g^s + tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a})\bigr| \ge \widehat{\varphi}(G) - \frac{\widehat{\varepsilon}}{4}.$$

На основании двух последних неравенств приходим к (C.4).

Лемма доказана. □

C.3. Обратимся к доказательству теоремы. Рассмотрим три случая.

1) $\widehat{\varphi}(G) > 0$. По непрерывности функции $\widehat{\varphi}(G)$ и определению плюсиковой функции при малых $t > 0$ имеем

$$\varphi(G + tV) = \widehat{\varphi}(G + tV).$$

В силу леммы

$$\frac{\varphi(G+tV)-\varphi(G)}{t} = \frac{1}{t}\left[\max_{s\in\widehat{R}(G)} f(g^s+tv^s,\ \widehat{a}) - \max_{s\in\widehat{R}(G)} f(g^s,\ \widehat{a})\right] =$$

$$= \frac{1}{t}\max_{s\in\widehat{R}(G)}\bigl[f(g^s+tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a})\bigr] = \max_{s\in\widehat{R}(G)}\langle v^s,\ \widehat{a}\rangle. \tag{C.6}$$

Мы воспользовались тем, что при $s \in \widehat{R}(G)$ значения $f(g^s,\ \widehat{a})$ равны между собой. Из (C.6) следует (C.1) (даже без предельного перехода).

2) $\widehat{\varphi}(G) = 0$. По лемме и свойству 5) плюсиковой функции (см. Дополнение B) при малых $t > 0$ имеем

$$\max_{s\in 1:h}\bigl[f(g^s+tv^s,\ \widehat{a})\bigr]_+ = \max_{s\in\widehat{R}(G)}\bigl[f(g^s+tv^s,\ \widehat{a})\bigr]_+,$$

поэтому

$$\varphi(G+tV) = \left[\max_{s\in 1:h} f(g^s+tv^s,\ \widehat{a})\right]_+ = \max_{s\in\widehat{R}(G)}\bigl[f(g^s+tv^s,\ \widehat{a})\bigr]_+.$$

По свойству 4) плюсиковой функции

$$\frac{\varphi(G+tV)-\varphi(G)}{t} = \frac{1}{t}\max_{s\in\widehat{R}(G)}\bigl[f(g^s+tv^s,\ \widehat{a})\bigr]_+ =$$

$$= \max_{s\in\widehat{R}(G)}\left[\frac{f(g^s+tv^s,\ \widehat{a}) - f(g^s,\ \widehat{a})}{t}\right]_+ = \max_{s\in\widehat{R}(G)}\bigl[\langle v^s,\ \widehat{a}\rangle\bigr]_+. \tag{C.7}$$

Мы воспользовались тем, что в данном случае $\varphi(G) = 0$ и $f(g^s,\ \widehat{a}) = 0$ при всех $s \in \widehat{R}(G)$. Из (C.7) следует (C.2) (и опять без предельного перехода).

3) $\widehat{\varphi}(G) < 0$. По непрерывности при малых $t > 0$ будет $\widehat{\varphi}(G+tV) < 0$. Формула (C.3) очевидна, поскольку по определению плюсиковой функции и $\varphi(G)$, и $\varphi(G+tV)$ при малых $t > 0$ равны нулю.

Теорема доказана. $\square$

C.4. Возьмём ещё одну точку $\check{b} \in \mathbb{R}^{n+1}$ и рассмотрим функции

$$\check{\psi}(G) = \min_{s\in 1:h} f(g^s,\ \check{b}), \quad \psi(G) = \bigl[\check{\psi}(G)\bigr]_+.$$

Обозначим

$$\check{R}(G) = \bigl\{s \in 1:h \mid f(g^s,\ \check{b}) = \check{\psi}(G)\bigr\}.$$

Теорема C.2. *Для производной по направлению* $\psi'(G,\ V)$ *справедлива формула*

$$\psi'(G, V) = \begin{cases} \min\limits_{s \in \check{R}(G)} \langle v^s,\ \check{b} \rangle, & \textit{если } \check{\psi}(G) > 0, \\ \min\limits_{s \in \check{R}(G)} \big[\langle v^s,\ \check{b} \rangle \big]_+, & \textit{если } \check{\psi}(G) = 0, \\ 0, & \textit{если } \check{\psi}(G) < 0. \end{cases}$$

Доказательство аналогично доказательству теоремы C.1. Нужно воспользоваться свойствами 4) и 6) плюсиковой функции (см. Дополнение B) и равенством

$$\min_{s \in 1:h} f(g^s + tv^s,\ \check{b}) = \min_{s \in \check{R}(G)} f(g^s + tv^s,\ \check{b}),$$

справедливым при малых $t > 0$.

ДОПОЛНЕНИЕ D

Лемма о сумме минимумов

Пусть

$$M = \sum_{j=1}^{k} \min_{s \in 1:h_j} p(s,\ j).$$

На рис. D схематично представлено индексное множество, на котором определена функция $p(s,\ j)$.

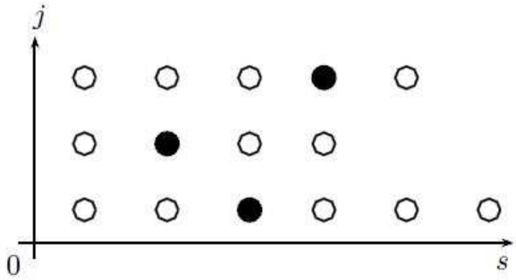

Рис. D. Двухиндексное множество

Обозначим через Π множество всех цепочек $S = (s_1,\ s_2, \ldots,\ s_k)$, где $s_j \in 1 : h_j$ при каждом $j \in 1 : k$. Выделим в Π цепочку $S^* = (s_1^*,\ s_2^*, \ldots,\ s_k^*)$, элементы которой удовлетворяют условиям

$$s_j^* \in 1 : h_j, \quad p(s_j^*,\ j) = \min_{s \in 1:h_j} p(s,\ j).$$

Введём функцию

$$P(S) = \sum_{j=1}^{k} p(s_j,\ j).$$

Ясно, что $M = P(S^*)$.

Лемма о сумме минимумов. *Справедливо равенство*

$$M = \min_{S \in \Pi} P(S). \tag{D.1}$$

Д о к а з а т е л ь с т в о. При каждом $S \in \Pi$ имеем

$$P(S) = \sum_{j=1}^{k} p(s_j,\ j) \geq \sum_{j=1}^{k} \min_{s \in 1:h_j} p(s,\ j) = M,$$

так что $P(S) \geq M$ при всех $S \in \Pi$. Вместе с тем, как отмечалось, $P(S^*) = M$, причём $S^* \in \Pi$. Приходим к равенству (D.1). $\square$

Перепишем формулу (D.1) в развёрнутом виде

$$\sum_{j=1}^{k} \min_{s \in 1:h_j} p(s,\ j) = \min_{S \in \Pi} \sum_{j=1}^{k} p(s_j,\ j).$$

Таким образом, лемма показывает, как правильно переставлять местами знаки «$\sum$» и «min».

СПИСОК ЛИТЕРАТУРЫ

1. Айзерман М. А., Браверман Э. М., Розоноэр Л. И. *Метод потнциальных функций в теории обучения машин.* М.: Наука, 1970.

2. Вапник В. Н., Червоненкис А. Я. *Теория распознавания образов (статистические проблемы обучения).* М.: Наука, 1974.

3. Гаврин М. К., Малозёмов В. Н. *Экстремальные задачи с линейными ограничениями.* Л.: Изд-во ЛГУ, 1984.

4. Горелик А. Л., Скрипкин В. А. *Методы распознования.* 4-е изд. М.: Высшая школа, 2004.

5. Демьянов В. Ф. *Идентификация точек двух выпуклых множеств* // Вестник СПбГУ. Сер. 1. 2001. Вып. 3 (№ 17), с. 14–20.

6. Журавлёв Ю. И. *Об алгебраическом подходе к решению задач распознования и классификации* // Проблемы кибернетики: Вып. 33. М.: Наука, 1978. С. 5–68.

7. Зубова О. А. *Идентификация нескольких множеств в многомерном пространстве* // Вестник СПбГУ. Сер. 10. 2007. № 4, с. 17–22.

8. Малозёмов В. Н., Чернэуцану Е. К. *О математической диагностике (линейная модель)* // Семинар «DHA & CAGD». Избранные доклады. 17 апреля 2010 г. (http://dha.spb.ru/PDF/LinMod.pdf)

9. Малозёмов В. Н., Чернэуцану Е. К. *Численный метод строгого h-отделения* // Семинар «DHA & CAGD». Избранные доклады. 15 октября 2011 г. (http://dha.spb.ru/PDF/hSepNumerical.pdf)

10. Малозёмов В. Н., Чернэуцану Е. К. *Наилучшее линейное отделение двух множеств* // Семинар «DHA & CAGD». Избранные доклады. 6 октября 2012 г. (http://dha.spb.ru/PDF/LinSep.pdf)

11. Неймарк Ю. И., Баталова З. С. и др. *Распознавание образов и медицинская диагностика.* М.: Наука, 1972.

12. Первозванский А. А. *Распознавание абстрактных образов, как задача линейного программирования* // Известия АН СССР, Техническая кибернетика. 1965. № 4, с. 41–44.

13. Сергеев А. Н., Соловьёва Н. А., Чернэуцану Е. К. *Решение задач линейного программирования в среде* MATLAB // Семинар «DHA & CAGD». Программные реализации. 12 февраля 2011 г. (http://dha.spb.ru/PDF/MatLabLP.pdf)

14. Соловьёва Н. А., Чернэуцану Е. К. *Решение задач квадратичного программирования в среде* MATLAB // Семинар «DHA & CAGD». Программные реализации. 12 февраля 2011 г. (http://dha.spb.ru/PDF/MatLabQP.pdf)

15. Чернэуцану Е. К. *Анализ задачи строгого h-отделения двух множеств.* Вестник СПбГУ. Сер. 10. 2012. № 4, с. 85-–91.

16. Чернэуцану Е. К. *Метод градиентного типа для решения задачи строгого h-отделения* // Вестник СПбГУ. Сер. 10. 2013. № 2, с. 67—75.

17. Чернэуцану Е. К., Лебедев Д. М. *Кусочно-линейные функции, допускающие минимальное разложение* / Процессы управления и устойчи-

вость: Труды 40-й международной научной конференции аспирантов и студентов. Санкт-Петербург, 2009. С. 784–789.

18. Чернэуцану Е. К. *Нахождение разделяющей гиперплоскости между двумя политопами* / Процессы управления и устойчивость: Труды 41-й международной научной конференции аспирантов и студентов. Санкт-Петербург, 2010. С. 736–739.

19. Чернэуцану Е. К. *Строгая h-отделимость двух множеств и линейное программирование* / Процессы управления и устойчивость: Труды 42-й международной научной конференции аспирантов и студентов. Санкт-Петербург, 2011. С. 259–265.

20. Чернэуцану Е. К. *Численные эксперименты по строгой h-отделимости* / Процессы управления и устойчивость: Труды 44-й международной научной конференции аспирантов и студентов. Санкт-Петербург, 2013. С. 88–93.

21. Чернэуцану Е. К. *Строгая h-отделимость двух множеств* // Семинар «DHA & CAGD». Избранные доклады. 18 декабря 2010 г. (http://dha.spb.ru/PDF/hSeparability.pdf)

22. Чернэуцану Е. К. *Строгая h-отделимость и линейное программирование* // Семинар «DHA & CAGD». Избранные доклады. 29 января 2011 г. (http://dha.spb.ru/PDF/hSeparabilityLP.pdf)

23. Чернэуцану Е. К. *Численные эксперименты по строгой h-отделимости* // Семинар «DHA & CAGD». Избранные доклады. 17 декабря 2011 г. (http://dha.spb.ru/PDF/hSepEx.pdf)

24. Фомин В. Н. *Математическая теория обучаемых опознающих систем.* Л.: Изд-во ЛГУ, 1976.

25. Цыпкин Я. З. *Основы теории обучающихся систем.* М.: Наука, 1970.

26. Якубович В. А. *Некоторые общие теоретические принципы построения обучаемых опознающих систем* / Сб.: Вычислительная техника и вопросы программирования, 4. Л.: Изд-во ЛГУ, 1965. С. 3–71.

27. Astorino A., Gaudioso M. *Polyhedral separability througt successive LP* // JOTA. 2002. Vol. 112. No. 2, pp. 265—293.

28. Bagirov A. M., Rubinov A. M. and Yearwood J. *A global optimization approach to classification* // Optimization and Engineering. 2002. Vol. 3, pp. 129–155.

29. Bennett K. P., Mangassarian O. L. *Robust linear programming discrimination of two linearly inseparable sets* // Optimization Methods and Software. 1992. Vol. 1, pp. 23–34.

30. Bennett K. P., Mangassarian O. L. *Bilinear Separation of Two Sets in n-Space.* Computational Optimization and Applications. 1993. Vol. 2, pp. 207–227.

31. Burges C. J. C. *A Tutorial on Support Vector Machines for Pattern Recognition* // Data Mining and Knowledge Discovery. 1998. Vol. 2, pp. 121–167.

32. Cherneutsanu E. *On strict h-polyhedral separability of two sets* /International conference «Conatructive nonsmooth analysis and related topics». Abstracts. Saint-Petersburg, June 18-23, 2012. P. 33-34.

33. Cristianini N. and Shawe-Taylor J. *An Introduction to Support Vector Machines and other kernel based methods.* Cambridge University Press, 2000.

34. Cucker F. and Smale S. *On the mathematical foundations of learning* // Bulletin (New Series) of the American Mathematical Society. 2002. 39(1), pp. 1–49.

35. Demyanov V. F. *Mathematical diagnostics via nonsmooth analysis* // Optimisation Methods and Software. 2005. Vol 20, No 2-3, pp. 197–218.

36. Demyanov V. F., Astorino A. and Gaudioso M. *Nonsmooth problems in Mathematical Diagnostics* // Nonconvex Optimization and Its Applications. 2001. Vol. 54, pp. 11–30.

37. Fung M. and Mangasarian O. L. *Privacy-Preserving Linear and Nonlinear Approximation via Linear Programming* // Optimization Methods and Software. 2013. Vol. 28, pp. 207–216.

38. Glover F. *Improved Linear Programming Models for Discriminant Analysis* // Decision Sciences. 1990. Vol. 21, pp. 771—785.

39. Hansen P. and Jaumard B. *Cluster analysis and mathematical programming* // Mathematical Programming. 1997. Vol. 79, pp. 191–215.

40. Malozemov V. N. and Cherneutsanu E. K. *The best linear separation of two sets* // Springer Optimization and its Applications. Springer Sciense+Business Media. New York, 2014. Vol. 87, pp. 175–183.

41. Mangasarian O. L. *Linear and nonlinear separation of patterns by linear programming* // Operations Research. 1965. Vol. 13, pp. 444–452.

42. Mangasarian O. L. *Multisurface Method of Pattern Separation* // IEEE Transactions on Information Theory. 1968. Vol. 14, pp. 801–807.

43. Mangasarian O. L. *Mathematical programming in data mining* // Data Mining and Knowledge Discovery. 1997. Vol. 1, pp. 183–201.

44. Mangasarian O. L. *Arbitrary-Norm Separating Plane* // Operations Research Letters. 1999. Vol. 24, pp. 15–23.

45. Mangasarian O. L. *Unsupervised Classification via Convex Absolute Value Inequalities.* Data Mining Institute Technical Report 14-01, March 2014.

46. Mangasarian O. L., Setiono R. and Wolberg W. H. *Pattern Recognition via Linear Programming: Theory and Application to Medical Diagnosis* // Large-Scale Numerical Optimization, Philadelphia, PA. SIAM. 1990. pp. 22–31.

47. Mangasarian O. L. and Wolberg W. H. *Cancer Diagnosis via Linear Programming* // SIAM News. 1990. Vol. 23, pp. 1–18.

48. Mirkin B. *Mathematical Classification and Clustering.* Kluwer Academic Publishers, 1996.

49. Quinlan J. R. *Programs for Machine Learning.* Morgan Kaufmann, San Mateo, 1993.

50. Scholkopf B., Smola A. *Learning with Kernels.* The MIT Press, 2002.

51. Smith F. W. *Pattern Classifier Design by Linear Programming* // IEEE Transactions on Computers. 1968. Vol. 4, pp. 17.

52. Rosen J. B. *Pattern separation by convex programming* // Journal of Mathematical Analysis and Applications. 1965. Vol. 10, pp. 123–134.

53. Vapnik V. *The Nature of Statistical Learning Theory.* Springer-Verlag, 2000.