Глава 4. Задача коммивояжера

В задаче коммивояжера рассматривается n городов и матрица попарных расстояний между ними. Требуется найти такой порядок посещения городов, чтобы суммарное пройденное расстояние было минимальным, каждый город посещался ровно один раз и коммивояжер вернулся в тот город, с которого начал свой маршрут. Другими словами, во взвешенном полном графе требуется найти гамильтонов цикл минимального веса.

Задача коммивояжера занимает особое место в комбинаторной оптимизации и исследовании операций. Исторически она была одной из тех задач, которые послужили толчком для развития этих направлений. Простота формулировки, конечность множества допустимых решений, наглядность и, в тоже время, колоссальные затраты на полный перебор до сих пор подталкивают математиков к разработке все новых и новых численных методов. Фактически, все свежие идеи сначала тестируются на этой задаче. С точки зрения приложений, она не представляет интерес. Куда важнее её обобщения для транспорта и логистики, когда несколько транспортных средств ограниченной грузоподъемности должны обслуживать клиентов, посещая их в заданные временные окна. В задаче коммивояжера выброшены все детали приложений. Оставлена только комбинаторная суть, чисто математическая проблема, с которой не удается справиться уже полвека. Именно этой задаче посвящена данная глава.

4.1 Вычислительная сложность

В теории вычислительной сложности принято рассматривать задачи распознавания, то есть задачи, в которых ответом может

быть только *«Да»* или *«Нет»*. Например, правда ли, что заданный граф является деревом? Среди задач распознавания принято выделять классы P и NP. Напомним, что класс P состоит из задач распознавания, разрешимых за полиномиальное время. Другими словами, число элементарных операций для решения таких задач ограничено сверху полиномом от длины записи исходных данных. Класс NP является более широким. Он включает в себя все задачи распознавания, в которых ответ *«Да»* может быть проверен за полиномиальное время. Задача принадлежит этому классу, если, даже не умея ее решать, можно “легко” проверить ответ, подглядев его или найдя в интернете. При этом достаточно уметь проверять только ответ *«Да»*. Иногда проверка ответа *«Да»* может быть легче или сложнее проверки ответа *«Нет».* Рассмотрим, например, задачу о гамильтоновости графа: задан простой неориентированный граф, требуется узнать, есть ли в нем гамильтонов цикл? Покажем, что эта задача принадлежит классу NP. Предположим, что граф действительно гамильтонов, и кто-то подсказал нам ответ, указав один из таких циклов. Спрашивается, можно ли проверить эту подсказку за полиномиальное время? Для этого нужно проверить, что указанный набор ребер образует цикл, и он покрывает все вершины. Очевидно, что это “легко” можно сделать и, следовательно, задача принадлежит классу NP. Заметим, что ответ «Нет» проверить здесь куда труднее. Говорят, что задача распознавания принадлежит классу co-NP, если ответ «Нет» можно проверить за полиномиальное время. Легко показать, что следующая задача принадлежит классу co-NP. Задан простой неориентированный граф, правда ли, что он не является гамильтоновым?

В классе P можно проверить любой ответ и, значит, $P \subseteq NP$. Доказать или опровергнуть обратное включение пока никому не удается. На сегодняшний день, это одна из центральных проблем

математики (http://www.claymath.org/millennium/) За ее решение Американское математическое общество предлагает приз — миллион долларов.

Многолетние интенсивные исследования подсказывают, что P ≠ NP. Косвенным доказательством этой гипотезы служит тот факт, что в классе NP обнаружены так называемые NP-полные задачи. Задачу из класса NP называют *NP-полной*, если существование полиномиального алгоритма для ее решения влечет существование полиномиальных алгоритмов для всех задач из класса NP. К настоящему времени известно огромное число NP-полных задач [4, 14]. Однако ни для одной из них так и не удалось разработать точный полиномиальный алгоритм. В частности, задача о гамильтоновости графа является NP-полной. В задаче коммивояжера требуется найти гамильтонов цикл минимальной длины. Это не задача распознавания. Она не лежит в классе NP, но она не проще проверки гамильтоновости графа. Действительно, если существует точный полиномиальный алгоритм для задачи коммивояжера, то можно легко построить точный полиномиальный алгоритм и для проверки гамильтоновости графа. Для этого достаточно по графу $G = (V, E)$, чью гамильтоновость мы исследуем, построить матрицу расстояний (c_{ij}) задачи коммивояжера по следующему правилу:

$$c_{ij} = \begin{cases} 1, \text{если } (i,j) \in E \\ 2, \text{если } (i,j) \notin E \end{cases}, \quad i, j \in V.$$

Если решение задачи коммивояжера имеет ответ n, то граф G является гамильтоновым. Обратное тоже верно. Задачи вне класса NP, которые не проще NP-полных задач, называют *NP-трудными*. Задача коммивояжера принадлежит этому классу. Фактически мы получили доказательство даже более сильного утверждения.

Теорема 4.1. Задача коммивояжера является NP-трудной даже, когда матрица (c_{ij}) является симметричной и состоит из 1 и 2.

Легко проверить, что неравенство треугольника $c_{ij} \leq c_{ik} + c_{kj},\ 1 \leq i, j, k \leq n$, в этом случае также выполняется. Говорят, что задача является NP-трудной в сильном смысле, если она остается NP-трудной даже при унарной кодировке исходных данных. При двоичной кодировке целое число B требует $\lceil \log_2 B \rceil$ ячеек памяти. При унарной кодировке требуется B ячеек памяти. Переход к унарной кодировке влечет рост требуемой памяти и, как следствие, рост длины записи исходных данных. Грубо говоря, задача будет NP-трудной в сильном смысле, если она остается NP-трудной, если все числа в исходных данных будут ограничены некоторой константой. Из нашего доказательства следует, что задача коммивояжера является NP-трудной в сильном смысле. Задача о рюкзаке, например, является NP-трудной, но не NP-трудной в сильном смысле.

Полученное утверждение говорит о том, что точный полиномиальный алгоритм для задачи коммивояжера построить, скорее всего, не удастся. По-видимому, таких просто не существует. Следовательно, надо либо выйти за рамки полиномиальных алгоритмов, либо искать приближенные решения задачи. Следующая теорема говорит о том, что второй подход может оказаться не проще точного решения задачи. Пусть для примера I задачи коммивояжера величина $Opt(I)$ задает длину кратчайшего гамильтонова цикла, а некоторый алгоритм A на этом примере дает ответ $A(I)$.

Теорема 4.2. Если существует приближенный полиномиальный алгоритм A и такая положительная константа r, что для любого примера I задачи коммивояжера справедливо неравенство

$$A(I) \leq r \cdot Opt(I),$$

то классы P и NP совпадают.

Доказательство. Рассмотрим снова задачу о гамильтоновом цикле и построим матрицу (c_{ij}) по следующему правилу:

$$c_{ij} = \begin{cases} 1, & \text{если } (i,j) \in E \\ nr, & \text{если } (i,j) \notin E \end{cases}$$

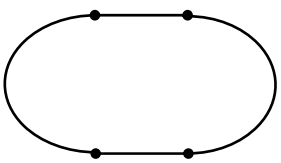
Применим приближенный алгоритм A. Если $A(I) = n$, то граф содержит гамильтонов цикл.

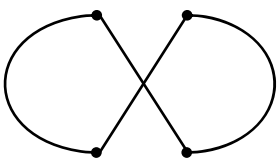
Предположим, что $A(I) > n$. Тогда $A(I) \geq nr + (n-1)$, так как путь коммивояжера проходит как минимум через одно "тяжелое" ребро. Но в этом случае граф не может иметь гамильтонов цикл, так как алгоритм ошибается не более чем в r раз. Следовательно, полиномиальный алгоритм A дает правильный ответ для NP-полной задачи и P = NP. ■

Из доказательства теоремы следует, что константу r можно заменить на любую, например, экспоненциально растущую функцию от n. Повторяя рассуждения, получаем, что относительная точность полиномиальных алгоритмов в худшем случае не может быть ограничена, в частности, величиной $2^{p(n)}$, где $p(n)$ — произвольный полином от размерности задачи, если P ≠ NP.

Заметим, что в отличие от доказательства теоремы 4.1, новая конструкция уже не гарантирует выполнение неравенства треугольника. Это наводит на мысль, что при дополнительных ограничениях на матрицу (c_{ij}) можно получить полиномиальные алгоритмы с гарантированной точностью. Это действительно так, и ниже будут представлены такие алгоритмы для матриц, удовлетворяющих неравенству треугольника. Подробный обзор таких алгоритмов можно найти, например, в [13].

Рассмотрим теперь другой класс алгоритмов — итерационные алгоритмы локального улучшения. Каждая итерация будет





иметь полиномиальную трудоемкость, но число итераций может оказаться в худшем случае экспоненциальной величиной. Пусть C — гамильтонов цикл, а $N(C)$ — окрестность цикла C, то есть все циклы, которые можно получить из цикла C некоторой локальной перестройкой. Например, $N(C)$ — окрестность 2–замена: выбираем в C два несмежных ребра и меняем их на два других, чтобы снова получить гамильтонов цикл (см. рис. 4.1).

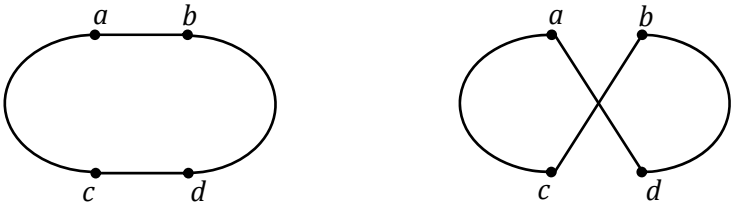


Рис.4.1 Окрестность 2–замена

Мощность такой окрестности — $n(n-3)/2$. Аналогично можно получить окрестность 3–замена, 4–замена и т.д.

Алгоритм локального улучшения

1. Выбрать начальный цикл C и вычислить его длину $f(C)$.
2. Найти наилучшего соседа C' для цикла C:
$$f(C') = \min\{F(T) \mid T \in N(C)\}.$$
3. Если $F(C') < F(C)$, то положить $C := C'$ и вернуться на 2 иначе STOP, C — локальный минимум.

Теорема 4.3. Пусть A — алгоритм локального улучшения и окрестность $N(C)$ имеет полиномиальную мощность. Если существует такая положительная константа r, что для любого примера I задачи коммивояжера справедливо неравенство $A(I) \leq r \cdot Opt(I)$, то классы P и NP совпадают.

Доказательство. (Аналогично предыдущему) Предположим, что такая константа существует. Рассмотрим задачу о гамильтоновом цикле и получим точный полиномиальный алгоритм ее решения. Снова по графу $G = (V, E)$, построим матрицу:

$$c_{ij} = \begin{cases} 1, \text{если } (i,j) \in E \\ nr, \text{если } (i,j) \notin E \end{cases}, \quad i, j \in V.$$

Выбираем произвольный начальный цикл. Его длина в худшем случае равна $n^2 r$. На каждом шаге локального улучшения часть тяжелых ребер заменяется легкими ребрами. Значит, число итераций не превышает n. Каждая итерация имеет полиномиальную трудоемкость, так как число соседних решений полиномиально. Следовательно, алгоритм A является полиномиальным для заданного класса примеров и, как следует из доказательства предыдущей теоремы, A — точный алгоритм. Значит, P = NP. ■

4.2 Конструктивные алгоритмы

Конструктивными алгоритмами принято называть такие алгоритмы, которые постепенно, шаг за шагом, строят допустимое решение задачи. В этом смысле алгоритмы локального улучшения не являются конструктивными. Рассмотрим один из конструктивных алгоритмов для задачи коммивояжера, известный под названием "*Иди в ближайший из непройденных городов*". Далее будем предполагать, что матрица (c_{ij}) удовлетворяет неравенству треугольника.

Алгоритм $A_{\text{Б}}$

1. Выбираем произвольный город и обозначаем его i_1.
2. Цикл $k := 1, \dots, n-1$

Находим ближайший город к i_k из непомеченных городов, обозначаем его i_{k+1}:

$$c_{i_k i_{k+1}} = \min_{j \neq i_1,\ldots,i_k} c_{i_k j}$$

и помечаем город i_k.

Теорема 4.4. Для любого $r > 1$ найдется такой пример I задачи коммивояжера, что даже при выполнении неравенства треугольника справедливо неравенство $A_{\text{Б}}(I) > r \cdot Opt(I)$.

Рассмотрим пример задачи коммивояжера, в котором города расположены на окружности на одинаковом расстоянии друг от друга. Длина окружности здесь является оптимальным значением, но алгоритм $A_{\text{Б}}$ может сильно ошибаться. На рис. 4.2 показан пример для $n = 15$. Величина c_{ij} определяется как длина кратчайшего пути между i и j, получаемого с использованием ребер, изображенных на рисунке. Нетрудно проверить, что эти расстояния удовлетворяют неравенству треугольника. Периметр дает оптимальный маршрут 15, алгоритм $A_{\text{Б}}$ выдает зачерненный маршрут длины 27.

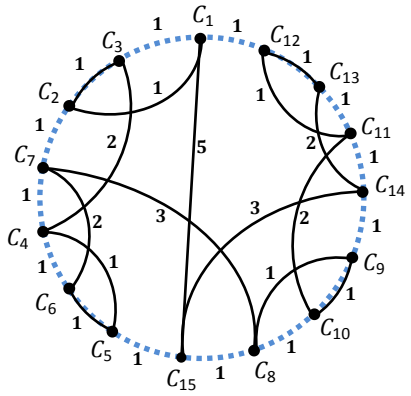


Рис. 4.2 Результат работы алгоритма $A_{\text{Б}}$

Рассмотрим теперь алгоритм, который имеет гарантированную точность $r = 2$. Он основан на следующей идее [14]. Для пол-

ного взвешенного графа задачи коммивояжера построим алгоритмом Краскала A_K остовное дерево минимального веса. Заметим, что $A_k(I) \leq Opt(I)$ для любого примера I, так как удаление ребра из оптимального решения задачи коммивояжера дает остовное дерево. Заменим каждое ребро на два ребра. Получим эйлеров граф, каждая вершина имеет четную степень. Построим произвольный эйлеров цикл (рис. 4.3).

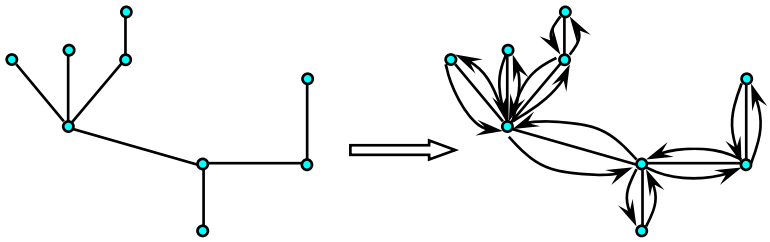


Рис. 4.3. Двойной обход остовного дерева

Листья посещаются 1 раз, остальные вершины по нескольку раз. Перестроим его так, чтобы получился гамильтонов цикл. Начиная с произвольной вершины, двигаемся вдоль эйлерова цикла и помечаем вершины. Если очередная вершина уже помечена, то пропускаем ее и двигаемся дальше, пока не найдем непомеченную вершину или не вернемся в первую вершину. Цепочку дуг для помеченных вершин заменяем прямой дугой в непомеченную или первую вершину (см. рис. 4.4). Обозначим этот алгоритм A_{ST}.

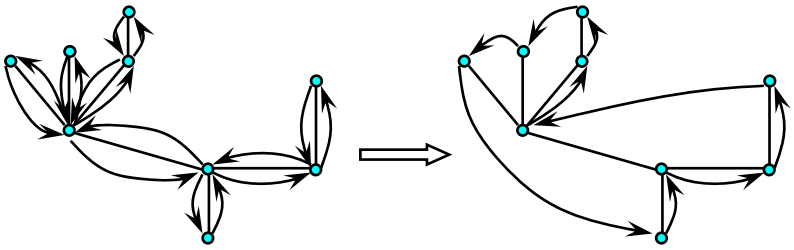


Рис. 4.4. Перестройка остовного дерева

Теорема 4.5. Для любого примера I задачи коммивояжера с симметричной матрицей (c_{ij}), удовлетворяющей неравенству треугольника, алгоритм A_{ST} получает гамильтонов цикл не более чем в два раза длиннее оптимального, то есть $A_{ST}(I) \leq 2Opt(I)$.

Доказательство. Для двойного обхода остовного дерева имеем $2A_K(I) < 2Opt(I)$. Пусть новое ребро (i,j), не содержащееся в двойном обходе, заменяет цепочку ребер $(i,k_1),(k_1,k_2),\dots,(k_m,j)$. Из неравенства треугольника следует, что $c_{ij} \leq c_{ik_1} + \cdots + c_{k_m j}$. Следовательно, $A_{ST}(I) \leq 2A_K(I) \leq 2Opt(I)$.

Теорема 4.6. Оценка точности $r = 2$ является неулучшаемой.

Доказательство. Приведем семейство исходных данных задачи коммивояжера, на котором оценка 2 достигается асимптотически. Рассмотрим следующий пример на евклидовой плоскости с $3(n+1)$ вершинами (см. рис. 4.5) .

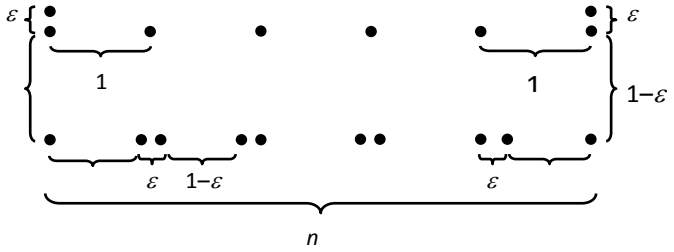


Рис. 4.5. Пример на евклидовой плоскости

Для этого примера минимальное остовное дерево имеет вид (рис. 4.6.):

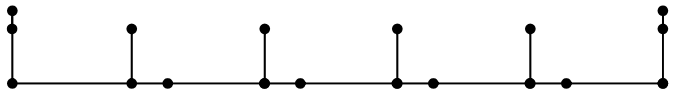


Рис. 4.6. Минимальное остовное дерево

Длина остовного дерева $A_K = n + (n + 1)(1 - \varepsilon) + 2\varepsilon$. Алгоритм перестройки двойного обхода получит решение (рис. 4.7)

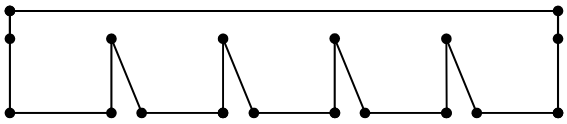


Рис. 4.7. Результат алгоритма A_{ST}

Длина этого гамильтонова цикла $A_{ST} \approx 2n + 2n\,(1 - \varepsilon)$ (рис. 4.8). Оптимальное решение задачи $Opt(I) \approx 2_n + 2$.

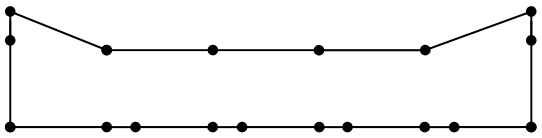


Рис. 4.8. Оптимальное решение

Таким образом, при $n \to \infty$ и $\varepsilon \to 0$ получаем $A_{ST}\,(I) \,/\, Opt(I) \to 2$. ■

Теоремы 4.5 и 4.6 говорят не только о точности алгоритма A_{ST} в худшем случае. Они утверждают также, что оценка справедлива для любого эйлерова цикла, который будет перестраиваться в гамильтонов цикл. Но эйлеровых циклов может быть экспоненциально много. Выбирая из всех вариантов наилучший, можно существенно понизить ошибку. Владимир Дейнеко и Александр Тискин [15] нашли точный полиномиальный алгоритм для выбора наилучшего эйлерова цикла. Оценка точности в худшем случае не изменилась, но поведение алгоритма в среднем показывает, что отклонение от нижней оценки оптимума не превышает 10%. На сегодняшний день это один их лучших полиномиальных алгоритмов с точки зрения средней погрешности.

В заключение этого раздела приведем еще один полиномиальный алгоритм, предложенный независимо Николаем Кристофидесом и Анатолием Сердюковым. Эта улучшенная версия предыдущего алгоритма имеет оценку точности 3/2. Как и раньше сначала строится остовное дерево минимального веса, затем к нему добавляются ребра так, чтобы получился эйлеров граф, а потом строится эйлеров цикл и перестраивается в гамильтонов цикл. Новшество состоит в добавлении ребер для получения эйлерова графа. Вместо дублирования ребер остова, в дереве выделяется множество вершин V' нечетной степени. Их число четно, $|V'| = 2k$, так как сумма всех степеней вершин графа должна быть четной. На множестве V' находится совершенное паросочетание минимального веса: k ребер, имеющие минимальный суммарный вес и покрывающие все вершины. Эта задача полиномиально разрешима. Более того, вес такого паросочетания не превосходит половины длины любого гамильтонова цикла. Действительно, гамильтонов цикл в исходном графе легко переделать в гамильтонов цикл для подграфа на вершинах из V'. Этого можно добиться, исключив вершины не принадлежащие V'. В силу неравенства треугольника получим гамильтонов цикл не большей длины, чем исходный цикл. Он задает на множестве V' два паросочетания. Они получаются, если брать ребра через одно. Наименьший из весов этих паросочетаний не превосходит половины длины гамильтонова цикла. Добавим это паросочетание к остовному дереву. Получим эйлеров граф. Его вес не превосходит 3/2 длины минимального гамильтонова цикла. Перестраивая этот граф в гамильтонов цикл старым способом, получаем требуемое. Можно показать, что эту оценку нельзя улучшить, т.е. есть примеры, на которых этот алгоритм действительно ошибается в 3/2 раза.

Заметим, что новый подход снова порождает эйлеров граф, в котором может оказаться экспоненциально много эйлеровых циклов. Выбирая наилучший из них, можно существенно понизить погрешность получаемых приближенных решений. К сожалению, выбор наилучшего эйлерова цикла в данном случае оказывается NP-трудной задачей. Тем не менее, реализация этой идеи, пусть даже в виде приближенного решения, приводит, как и в предшествующем случае к прекрасным результатам.

4.3 Нижние оценки

Получение нижних оценок оптимума является важной задачей. Во-первых, они позволяют оценить погрешность конструктивных алгоритмов и вообще любых приближенных алгоритмов. Во-вторых, эти оценки используются в точных методах, например, методе ветвей и границ [12, 16]. Их качество часто оказывается критическим фактором при получении глобального оптимума. Ниже будут представлены четыре нижних оценки, каждая из которых имеет свои сильные и слабые стороны.

4.3.1 Примитивная нижняя оценка

Пусть величины $a_i = \min_{j \neq i} c_{ij},\ i = 1, \dots, n,$ задают минимальную плату за выезд из города. Так как коммивояжер должен выехать их каждого города, то $\sum_{i=1}^{n} a_i$ является, очевидно, нижней оценкой оптимума. Рассмотрим теперь матрицу $c'_{ij} = c_{ij} - a_i$ и для каждого города посчитаем минимальную плату за въезд: $b_j = \min_{j \neq i} c'_{ij},\ j = 1, \dots, n.$ Тогда величина $H(c_{ij}) = \sum_{i=1}^{n} a_i + \sum_{j=1}^{n} b_j$ будет также нижней оценкой.

Теорема 4.7. Величина $H(c_{ij})$ является нижней оценкой оптимума.

Доказательство. Оптимальные решения для матриц (c_{ij}) и (c'_{ij}) связаны соотношением $Opt(c_{ij}) = Opt(c'_{ij}) + \sum_{i=1}^{n} a_i$. Положим $c''_{ij} = c'_{ij} - b_j,\ 1 \le i,j \le n$. Тогда $Opt(c_{ij}) = Opt(c''_{ij}) + \sum_{i=1}^{n} a_i + \sum_{j=1}^{n} b_j \ge \sum_{i=1}^{n} a_i + \sum_{j=1}^{n} b_j$. ■

Основным достоинством этой нижней оценки является ее простота и наглядность. Она требует $O(n^2)$ операций, то есть линейной трудоемкости от числа элементов матрицы расстояний.

4.3.2. Оценка линейного программирования

Представим задачу коммивояжера в терминах целочисленного линейного программирования. Введем переменные

$$x_{ij} = \begin{cases} 1, & \text{если из города } i \text{ едем в город } j \\ 0 & \text{в противном случае} \end{cases}.$$

Задача коммивояжера может быть записана следующим образом:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

при ограничениях

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad 1 \le j \le n,$$

$$\sum_{j=1}^{n} x_{ij} = 1, \qquad 1 \le i \le n,$$

$$\sum_{i \in S} \sum_{j \notin S} x_{ij} \ge 1, \qquad S \ne \emptyset, S \subset \{1, \dots, n\},$$

$$x_{ij} \in \{0,1\}, \qquad 1 \le i,j \le n.$$

Целевая функция задает длину гамильтонова цикла. Первое и второе равенство требует въехать и выехать из каждого города. Однако этого недостаточно. Набор подциклов, покрывающий все города, будет удовлетворять этим равенствам, но не будет гамильтоновым циклом. Поэтому нужно дополнительное неравенство для исключения подциклов. Это неравенство требует для любого подмножества городов S существование хотя бы одной дуги, ведущей в остальные города.

Заменим условие целочисленности переменных на условие $0 \leq x_{ij} \leq 1,\ 1 \leq i, j \leq n$. Получим задачу линейного программирования. Оптимальное значение в этой задачи будет нижней оценкой, потому что такая замена расширяет область допустимых решений. Эта оценка будет не хуже примитивной оценки.

Задача линейного программирования полиномиально разрешима [17]. Однако не стоит торопиться с выводами. Для задачи линейного программирования действительно есть точный алгоритм, трудоемкость которого полиномиально зависит от длины записи исходных данных. Но в задаче коммивояжера только n^2 чисел. Если они ограничены константой, то длина входа $O(n^2)$. В нашей задаче линейного программирования вход значительно длиннее. Только ограничений для исключения подциклов $O(2^n)$. Фактически, мы не можем даже выписать эту задачу из-за огромного числа ограничений. Ситуация кажется тупиковой. Тем не менее, из нее есть как минимум два выхода. Первый состоит в замене условий на подциклы другим условием, имеющим полиномиальное число ограничений. Это можно сделать несколькими способами (см., например, [18]). Тогда нижняя оценка становится полиномиально вычислимой, но ее качество заметно падает. Второй способ состоит в последовательном наращивании ограничений на подциклы по ме-

ре необходимости. Удалим сначала все такие ограничения и решим задачу. Найдем по заданному решению подмножество S, для которого нарушаются условие на подциклы и добавим его в систему ограничений. Действуя таким способом, можно последовательно улучшать нижнюю оценку до тех пор, пока либо не получим точное решение исходной задачи линейного программирования, либо задача не станет слишком большой для расчетов. В любом случае получаем нижнюю оценку оптимума.

4.3.3 1-деревья для симметричных матриц

Вес минимального остовного дерева, как мы уже знаем, дает оценку снизу, но ее можно улучшить. Остовное дерево состоит из $(n-1)$ ребра, а гамильтонов цикл содержит n ребер. Добавление одного ребра к остовному дереву порождает так называемые *1-деревья*. Они тоже дают нижние оценки. Итак, мы хотим найти гамильтонов цикл минимального веса, то есть:

- ровно n ребер, которые
- покрывают все вершины и образуют связный граф,
- имеют минимальный суммарный вес,
- каждая вершина инцидентна ровно двум ребрам.

Ослабим последнее условие, заменив его следующим:

- одна заданная вершина инцидентна ровно двум ребрам.

Так как мы ослабили одно из условий, то получим нижнюю оценку. Алгоритм построения минимального по весу 1-дерева для выделенной вершины состоит в следующем. Удаляем эту вершину и смежные с ней ребра. На получившемся графе строим остовное дерево минимального веса. Получаем $n-2$ ребер. Добавляем два ребра, инцидентных выделенной вершине, имеющих минималь-

ный суммарный вес. Полученное 1-дерево дает нижнюю оценку. Она зависит от выбора вершины. Построив 1-деревья для всех вершин и выбрав из нижних оценок наибольшую, получим требуемое.

4.3.4 Задача о назначениях

Рассмотрим следующую задачу о назначении рабочих на станки [12, 17]. Дано: n рабочих, n станков и матрица c_{ij}, задающая время выполнения работы на j-м станке i-м рабочим. Требуется найти расстановку рабочих по станкам, которая дает наименьшее суммарное рабочее время. Введем переменные задачи

$$x_{ij} = \begin{cases} 1, & \text{если рабочий } i \text{ работает на станке } j \\ 0 & \text{в противном случае} \end{cases}.$$

Задача о назначениях может быть записана в следующем виде:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

при ограничениях

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad 1 \le j \le n,$$

$$\sum_{j=1}^{n} x_{ij} = 1, \qquad 1 \le i \le n,$$

$$x_{ij} \in \{0,1\}, \qquad 1 \le i, j \le n.$$

Сравнивая эту запись с аналогичной записью для задачи коммивояжера, легко заметить, что они отличаются только одной группой ограничений для исключения подциклов. Следовательно, оптимальное решение задачи о назначениях дает нижнюю оценку для задачи коммивояжера.

Приведем точный полиномиальный алгоритм решения задачи о назначениях. Пусть $\Delta = (\Delta_1, \dots, \Delta_n)$ — некоторый вектор. Элемент c_{ij} называется Δ-минимальным для строки i, если $c_{ij} - \Delta_j \le c_{ik} - \Delta_k$ для всех $k = 1, \dots, n$.

Теорема 4.8. Пусть для некоторого вектора Δ существует набор Δ-минимальных элементов $(c_{1j(1)}, \dots, c_{nj(n)})$ по одному в каждой строке и каждом столбце. Тогда этот набор является оптимальным решением задачи о назначениях.

Доказательство. Решение $(c_{1j(1)}, \dots, c_{nj(n)})$ является допустимым и

$$\sum_{i=1}^{n} c_{ij(i)} = \sum_{i=1}^{n} (c_{ij(i)} - \Delta_{j(i)}) + \sum_{j=1}^{n} \Delta_j.$$

В правой части равенства первая сумма является минимальной среди всех допустимых назначений. Вторая сумма является константой. Значит, полученное решение является оптимальным. ■

Определение 4.1. Для вектора Δ выделим в каждой строке по одному Δ-минимальному элементу и назовем его *Δ-основой*. Другие Δ–минимальные элементы будем называть *альтернативными Δ-основами.* Число столбцов матрицы (c_{ij}) без Δ-*основ* назовем *дефектом.*

Идея алгоритма состоит в следующем. Начинаем с нулевого вектора Δ и считаем дефект. На каждом этапе алгоритма дефект уменьшается на 1, то есть не более чем за n этапов получим оптимальное решение.

Описание одного этапа

1. Выберем столбец без Δ-основы и обозначим его S_1.

2. Увеличим Δ_{S_1} на максимальное δ так, чтобы все Δ-минимальны элементы остались Δ-минимальными (возможно

$\delta = 0$). Получим для некоторой строки i_1 новый Δ-минимальный элемент $c_{i_1 S_1}$, назовем его альтернативной основой для строки i_1 (рис. 4.9).

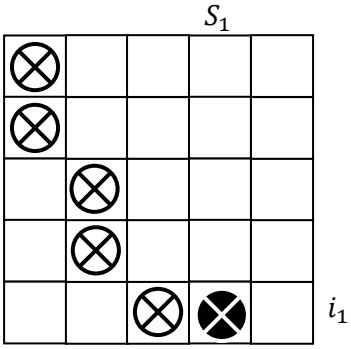


Рис. 4.9. Появление альтернативной основы

3. Для строки i_1 столбец $j(i_1)$ с Δ-основой пометим меткой S_2.

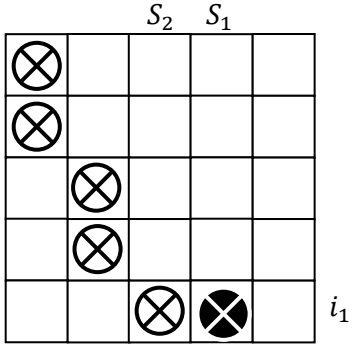


Рис. 4.10. Выбор второго столбца

4. Увеличим Δ_{s_1} и Δ_{s_2} на максимальное δ так, чтобы все Δ-основы остались Δ-минимальными элементами. Найдем новую альтернативную основу в одном из столбцов S_1 или S_2 . Пусть она оказалась в строке i_2 . Пометим столбец $j(i_2)$ меткой S_3 (Рис 4.11)

и будем продолжать этот процесс до тех пор, пока не встретим столбец с двумя или более основами.

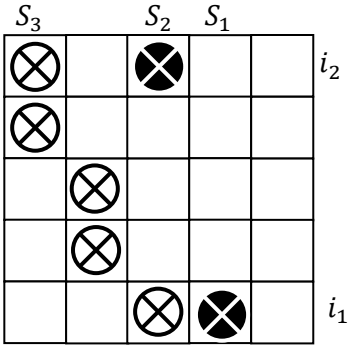


Рис. 4.11. Получение новой основы

5. Строим новый набор Δ-основ. Заменой основы в строке назовем следующую операцию: альтернативная основа становится основой, а старая перестает быть основой. Произведем замену основ в строке, где лежит последняя альтернативная основа (строка i_k). Тогда в столбце $j(i_k)$ число основ уменьшится на 1, но останется положительным (Рис. 4.12).

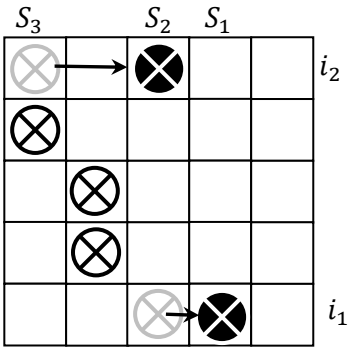


Рис. 4.12. Замена основы

6. В столбце, где появилась новая основа, возьмем старую основу и в этой строке тоже проведем замену основ и т.д. до тех пор,

пока не доберемся до столбца S_1. В итоге, столбец S_1 получит основу, а число основ в столбце $j(i_k)$ уменьшится на 1.

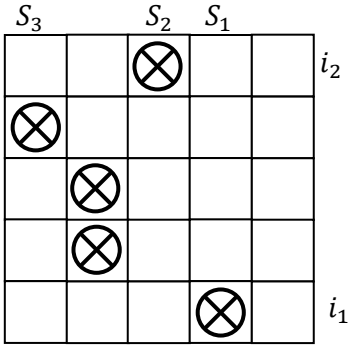


Рис. 4.13. Новое распределение основ

Посчитаем трудоемкость одного этапа. Мы последовательно рассматриваем сначала один столбец S_1 , затем два столбца S_1, S_2 и т.д. Всего таких шагов может быть не больше *n.* На каждом шаге надо выбрать δ, это требует $O(n^2)$ операций. Значит, всего требуется $O(n^4)$ операций.

4.4. Локальный поиск

Идеи локального поиска при решении задач дискретной оптимизации являются, по-видимому, наиболее естественными и наглядными. Первые шаги их реализации относятся к началу 50-х, началу 60-х годов 20 столетия. Они связаны, в основном, с задачей коммивояжера. Позднее эти идеи использовались для задач размещения, построения сетей, расписаний и др. Однако, довольно быстро выяснилось, что методы локального улучшения не гарантируют нахождения глобального оптимума, и отсутствие концептуального прогресса ослабило интерес к данному направлению. В последние 20 лет наблюдается возрождение этого подхода. Оно связано как с новыми алгоритмическими схемами, построенными на аналогиях с живой и неживой природой, так и с новыми теоретическими результатами в области локального поиска. Изменился общий взгляд на построение локальных алгоритмов. Требование монотонного улучшения по целевой функции больше не является доминирующим. Наиболее мощные алгоритмы допускают произвольное ухудшение и многие из них могут рассматриваться как способ порождения конечных неразложимых цепей Маркова на подходящем множестве состояний. В данном разделе приводится краткое введение в эту бурно развивающуюся область дискретной оптимизации.

4.4.1. Основные определения

Пусть тройка $(\mathfrak{I}, Sol, f)$задает задачу дискретной оптимизации с множеством входов $\mathfrak{I}$ (исходных данных), конечным множеством допустимых решений $Sol(I), I \in \mathfrak{I}$, и целевой функцией $f: s \to R, s \in Sol(I)$. Для определенности будем считать, что требуется найти минимум функции f на множестве $Sol(I)$ и само реше-

ние $s^* \in Sol(I)$, на котором этот минимум достигается. Решение s^* будем называть *глобальным минимумом,* и множество глобальных минимумов будем обозначать через S^*. Для каждого $s \in Sol(I)$ определим функцию окрестности $N: Sol(I) \to 2^{Sol(I)}$, которая для каждого допустимого решения задает множество соседних решений, в некотором смысле близких к данному. Множество $N(s)$ будем называть окрестностью решения s в множестве $Sol(I)$. Функция окрестности может быть достаточно сложной и отношение соседства не всегда симметрично.

Определение 4.2. Решение $s \in Sol(I)$ называется *локальным минимумом* по отношению к функции N, если$f(s) \leq f(s')$ для всех $s' \in N(s)$.

Множество локальных минимумов обозначим через $\hat{S}$. Это множество, очевидно, зависит от выбора функции N.

Определение 4.3. Функция окрестности N называется *точной*, если $\hat{S} \subseteq S^*$.

Стандартный алгоритм локального улучшения начинает работу с некоторого начального решения, выбранного случайным образом или с помощью какого-либо вспомогательного алгоритма. На каждом шаге локального улучшения происходит переход от текущего решения к соседнему решению с меньшим значением целевой функции до тех пор, пока не будет достигнут локальный минимум.

Алгоритмы локального улучшения широко применяются для решения NP-трудных задач. Многие полиномиально разрешимые задачи могут рассматриваться как задачи, легко решаемые таким способом. При подходящем выборе полиномиальной окрестности соответствующая теорема может быть сформулирована в следую-

щем виде: допустимое решение не является глобальным оптимумом, если и только если оно может быть улучшено некоторым локальным образом. Ниже приводится несколько примеров таких задач. Они указывают на важность локального поиска при построении оптимизационных алгоритмов и достаточно общий характер этого подхода.

1. **Линейное программирование**. Геометрически алгоритм симплекс метода можно интерпретировать как движение по вершинам многогранника допустимой области. Вершина не является оптимальной, если и только если существует смежная с ней вершина с меньшим значением целевой функции. Алгебраически, предполагая невырожденность задачи, базисное допустимое решение не является оптимальным, если и только если оно может быть улучшено локальным изменением базиса, т.е. заменой одной базисной переменной на небазисную. Получающаяся таким образом окрестность является точной и имеет полиномиальную мощность.

2. **Минимальное остовное дерево**. Остовное дерево не является оптимальным, если и только если локальной перестройкой, добавляя одно ребро и удаляя из образовавшегося цикла другое ребро, можно получить новое остовное дерево с меньшим суммарным весом. Операция локальной перестройки задает отношение соседства на множестве остовных деревьев. Окрестность любого дерева имеет полиномиальную мощность, а функция окрестности является точной.

3. **Максимальное паросочетание**. Паросочетание не является максимальным, если и только если существует увеличивающий путь. Два паросочетания называют соседними, если их симметрическая разность образует путь. Определенная таким образом окрестность является точной и имеет полиномиальную мощность. Ана-

логичные утверждения справедливы для взвешенных паросочетаний, совершенных паросочетаний минимального веса, задач о максимальном потоке и потоке минимальной стоимости.

На каждом шаге локального улучшения функция окрестности задает множество возможных направлений движения. Часто это множество состоит из нескольких элементов и имеется определенная свобода в выборе следующего решения. Правило выбора может оказать существенное влияние на трудоемкость алгоритма и результат его работы. Например, в задаче о максимальном потоке алгоритм Форда-Фалкерсона (который тоже можно рассматривать как вариант локального улучшения) имеет полиномиальную временную сложность при выборе кратчайшего пути для увеличения потока и экспоненциальную временную сложность без гарантии получить глобальный оптимум при произвольном выборе пути. Таким образом, при разработке алгоритмов локального поиска важно не только правильно определить окрестность, но и верно задать правило выбора направления спуска.

Интуитивно кажется, что в окрестности надо брать элемент с наименьшим значением целевой функции. Однако, как мы увидим ниже, разумным оказывается не только такой выбор, но и движение в "абсурдном" направлении, когда несколько шагов с ухудшением могут привести (и часто приводят) к лучшему локальному минимуму. При выборе окрестности хочется иметь как можно меньше соседей, чтобы сократить трудоемкость одного шага. С другой стороны, более широкая окрестность, вообще говоря, приводит к лучшему локальному оптимуму. Поэтому при создании алгоритмов каждый раз приходится искать оптимальный баланс между этими противоречивыми факторами. Ясных принципов разрешения этих

противоречий на сегодняшний день неизвестно, и для каждой задачи этот вопрос решается индивидуально.

Определение 4.4. *Графом соседства* $G_N = (Sol(I), E)$ будем называть взвешенный ориентированный граф, вершинами которого являются допустимые решения задачи, а дугами — упорядоченные пары (s, s'), если $s' \in N(s)$. Веса в этом графе приписаны вершинам и равны соответствующим значениям целевой функции.

Граф соседства G_N(*neighborhood graph*) иногда называют ландшафтом целевой функции или просто ландшафтом (*landscape, fitness landscape*). При определении функции окрестности важно следить за тем, чтобы получающийся граф был строго связен, то есть для каждой пары вершин s, s' существовал путь из s в s'. Это свойство является важным при анализе асимптотического поведения алгоритмов, например, вероятностных метаэвристик, о которых пойдет речь ниже. Если же это свойство не выполняется, то стараются получить хотя бы свойство вполне связности, когда из любой вершины существует путь в вершину $s^* \in S^*$. Если же и этого свойства нет, то теряется уверенность в достижении глобального оптимума локальными методами. Приходится ограничиваться локальными оптимумами, либо переопределять функцию окрестности.

Анализ эффективности локального поиска условно можно разделить на два направления: эмпирические и теоретические исследования. Как это ни странно, но они дают разные оценки возможностям этого подхода [19, 20].

Эмпирические результаты. Для многих NP-трудных задач локальный поиск позволяет находить приближенные решения, близкие по целевой функции к глобальному оптимуму. Трудоемкость алгоритмов часто оказывается полиномиальной, причем степень полинома достаточно мала. Так для задачи о разбиении множества вершин

графа на две равные части разработаны алгоритмы локального поиска минимизации разреза со средней трудоемкостью $O(n \log n)$, которые дают всего несколько процентов погрешности.

Для задачи коммивояжера алгоритмы локального поиска являются наилучшими с практической точки зрения. Один из таких алгоритмов с окрестностью Лина-Кернигана в среднем имеет погрешность около 2% и максимальная размерность решаемых задач достигает 1 000 000 городов. На случайно сгенерированных задачах такой колоссальной размерности итерационная процедура Джонсона позволяет находить решения с отклонением около 0,5 % за несколько минут на современных компьютерах.

Для задач теории расписаний, размещения, покрытия, раскраски графов и многих других NP-трудных задач алгоритмы локального поиска показывают превосходные результаты. Более того, их гибкость при изменении математической модели, простота реализации и наглядность превращают локальный поиск в мощное средство для решения практических задач.

Теоретические результаты. Исследование локального поиска с точки зрения гарантированных оценок качества показывают границы его возможностей. Построены трудные для локального поиска примеры, из которых следует, что

1) минимальная точная окрестность может иметь экспоненциальную мощность;

2) число шагов для достижения локального оптимума может оказаться экспоненциальным;

3) значение локального оптимума может сколь угодно сильно отличаться от глобального оптимума.

Эти результаты подталкивают к более пристальному рассмотрению задачи нахождения локального оптимума, ее трудоемкости в среднем и худшем случаях. Абстрактная (массовая) задача локального поиска L задается множеством ее индивидуальных примеров, каждый из которых однозначно определяет целевую функцию, функцию окрестности и множество допустимых решений.

Определение 4.5. Задача L принадлежит классу *PLS (polynomial time local search)*, если за полиномиальное время от длины записи исходных данных можно выполнить следующие три операции:

1) Для произвольного примера I проверить его корректность и, если $I \in \mathfrak{I}$, то найти некоторое допустимое решение.

2) Для любого решения проверить, является ли оно допустимым и, если да, то вычислить значение целевой функции для него.

3) Узнать является ли данное решение локальным оптимумом и, если нет, то найти в его окрестности решение с лучшим значением целевой функции.

Грубо говоря, в классе PLS содержатся все задачи локального поиска, для которых проверка локальной оптимальности может быть осуществлена за полиномиальное время. Этот класс не пуст, так как задача коммивояжера с окрестностью 2-замена содержится в нем. Следующая теорема говорит о том, что, скорее всего, в этом классе нет NP-трудных задач.

Теорема 4.9. Если задача L из класса PLS является NP-трудной, то NP = co-NP.

В классе PLS содержатся полиномиально разрешимые задачи, то есть задачи, для которых можно найти локальный оптимум за полиномиальное время. Многие NP-трудные задачи без весовых

функций с любой полиномиально проверяемой окрестностью относятся к таковым: задачи о клике, о покрытии, о независимом множестве и др. Задача коммивояжера с матрицей (c_{ij}), состоящей из 1 и 2 будет таковой при любой полиномиально проверяемой окрестности. Однако, в общем случае вопрос о вычислительной сложности нахождения локального оптимума пока остается открытым. Так же как и для задач распознавания, в классе PLS можно определить сведение одной задачи к другой и доказать теорему о существовании PLS-полных задач. Это наиболее трудные задачи в этом классе. Существование полиномиального алгоритма хотя бы для одной из них влечет полиномиальную разрешимость для всех задач из класса PLS. Задача коммивояжера с окрестностью -замена является PLS-полной при $k \geq 8$. Более того, удается показать, что стандартный алгоритм локального улучшения для этой задачи требует в худшем случае экспоненциального числа итераций независимо от того, какое правило замещения будет применяться. Более подробно с теорией PLS-полных задач можно познакомиться в [19, 21].

4.4.2. Окрестности на перестановках

Как уже отмечалось выше, выбор окрестности играет важную роль при построении алгоритмов локального поиска. От него существенно зависит трудоемкость одного шага алгоритма, общее число шагов и, в конечном счете, качество получаемого локального оптимума. На сегодняшний день нет и, возможно никогда не будет, единого правила выбора окрестности. Для каждой задачи функцию N приходится определять заново, учитывая специфику данной задачи. Более того, по-видимому для каждой задачи можно предложить несколько функций окрестности с разными по мощности множествами соседей и, как следствие, разными множествами локальных оптимумов. Ниже будут приведены три примера выбора окре-

стностей для задачи коммивояжера, которые иллюстрируют возможные пути построения окрестностей и их свойства.

Будем по-прежнему рассматривать задачу коммивояжера с симметричной матрицей расстояний и гамильтонов цикл представлять в виде перестановки $\pi = \{i_1, \dots, i_n\}$. Определим окрестность $N(\pi)$ как множество всех перестановок, отличающихся от π только в двух позициях ($city\text{-}swap$). Множество $N(\pi)$ содержит ровно $n(n-1)/2$ элементов и вычислительная сложность одного шага локального поиска с учетом вычисления целевой функции не превосходит $O(n^2)$ операций.

Обозначим через $f(\pi)$ длину гамильтонова цикла и определим разностный оператор ∇^2 для $f(\pi)$ следующим образом [22]:

$$\nabla^2 f(\pi) = \frac{1}{|N(\pi)|} \sum_{\pi' \in N(\pi)} f(\pi') - f(\pi).$$

Этот оператор задает среднее отклонение целевой функции в окрестности данной перестановки. Для локального оптимума π справедливо неравенство $\nabla^2 f(\pi) > 0$. Пусть f_{av} — средняя длина цикла на множестве всех допустимых решений задачи. Тогда справедливы следующие утверждения.

Теорема 4.10. Функция $f' = f - f_{av}$ удовлетворяет уравнению

$$\nabla^2 f' = -\frac{4}{n} f'.$$

Следствие 4.1. Любой локальный минимум π имеет длину $f(\pi) \leq f_{av}$.

Следствие 4.2. Алгоритм локального поиска, начиная с произвольной перестановки, достигнет локального оптимума за $O(nk)$ шагов, если длина максимального тура превосходит среднее значение f_{av} не более чем в 2^k раз.

Аналогичные утверждения справедливы и для следующих задач:

1) о разбиении $2n$-вершиного графа на две части по n вершин с минимальным суммарным весом ребер, соединяющих эти части;

2) о раскраске вершин графа в n цветов так, чтобы смежные вершины имели разные цвета;

3) о разбиении -элементного множества на два подмножества так, чтобы суммарный вес одного подмножества совпал с суммарным весом другого подмножества;

4) о 3-выполнимости в следующей постановке: для n булевых переменных задан набор троек; каждая переменная может входить в набор с отрицанием или без него; набор считается выполненным, если он содержит разные значения, то есть хотя бы одно *истинное* и хотя бы одно *ложное*; требуется узнать, существует ли назначение переменных, при котором все наборы будут выполнены. Развитие этих идей можно найти в [23].

4.4.3. Окрестности Лина-Кернигана

Рассмотрим гамильтонов цикл как последовательность ребер. Напомним, что окрестностью 2-замена называют множество всех гамильтоновых циклов, получающихся из заданного заменой двух несмежных ребер. Такая окрестность содержит $n(n-3)/2$ элементов, что несколько меньше, чем в окрестности city-swap. На рис. 4.14 показано различие между этими окрестностями. Для окрестности 2-замена удаление ребер (a, b) и (e, f) приводит к появлению двух новых ребер (a, e) и (b, f) и обратному обходу дуг от d до c. Если в окрестности city-swap в перестановке поменять местами b и e, то появятся четыре новых ребра $(a, e), (e, c), (d, b)$ и (b, f), но обход дуг от d до c останется прежним.

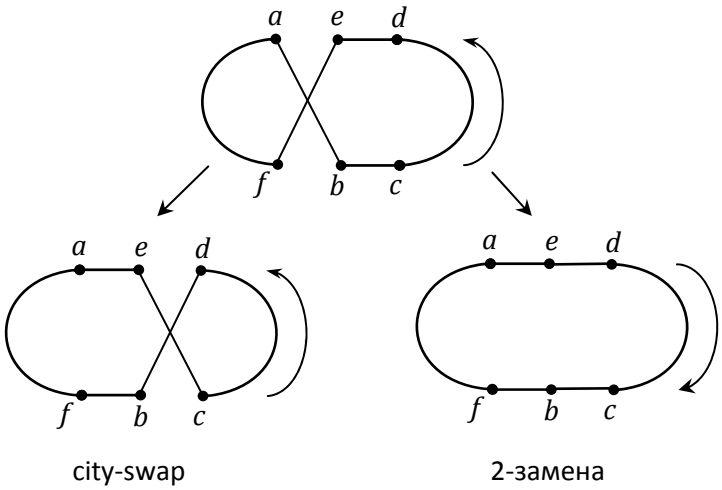


Рис. 4.14. Различия между окрестностями

Если вершины расположены на плоскости, а веса ребер вычисляются как евклидовы расстояния, то пересечение ребер (a, b) и (e, f) свидетельствует о возможности улучшения данного тура в силу неравенства треугольника. Однако это условие не является необходимым и улучшение возможно даже без пересечения ребер.

На основе окрестности 2-замена Лином и Керниганом предложена оригинальная эвристика. Она позволяет заменять произвольное число ребер и переходит от одного тура к другому, используя принципы жадных алгоритмов. Основная идея эвристики заключается в следующем. Удалим из гамильтонова цикла произвольное ребро, скажем (a, b). В полученном пути один конец (вершину a) будем считать фиксированной, а другой конец будем менять, перестраивая гамильтонов путь. Добавим ребро из вершины b, например (b, c), и разорвем образовавшийся единственный цикл так, чтобы снова получить гамильтонов путь. Для этого придется удалить ребро, инцидентное вершине c. Обозначим его (c, d). Но-

вый гамильтонов путь имеет концевые вершины a и d (см. рис. 4.15). Эту процедуру будем называть ротацией. Для получения нового гамильтонова цикла достаточно добавить ребро (a, d). Согласно алгоритму Лина-Кернигана, переход от одного тура к другому состоит из удаления некоторого ребра, выполнении серии последовательных ротаций и, наконец, замыкания концевых вершин полученного гамильтонова пути. Существуют различные варианты этой основной схемы, которые отличаются правилами выбора ротаций и ограничениями на множества удаляемых и добавляемых ребер. В алгоритме Лина-Кернигана ротации выбираются так, чтобы минимизировать разность $c_{bc} - c_{cd}$. При этом множества удаляемых и добавляемых ребер в серии ротаций не должны пересекаться. Последнее ограничение гарантирует, что число ротаций не превысит n^2 и трудоемкость одного шага (перехода от одного цикла к другому) останется полиномиальной. Общее число шагов алгоритма, по-видимому, не может быть ограничено полиномом и известен вариант окрестности Лина-Кернигана, для которого задача коммивояжера становится PLS-полной [19].

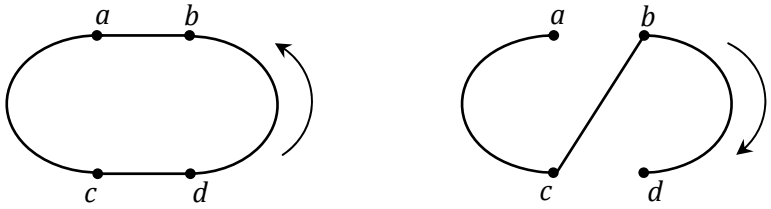


Рис. 4.15. Процедура ротации

4.4.4. Экспоненциальные окрестности

Одним из новых перспективных направлений в области локального поиска является исследование свойств окрестностей экспоненциальной мощности. Если лучшее решение в такой окрестно-

сти можно найти за полиномиальное время, то алгоритм локального поиска с такой окрестностью получает определенное преимущество перед остальными. Для задачи коммивояжера первые примеры таких окрестностей были предложены в работе [24]. Изложим, в качестве иллюстрации, идею одного из таких подходов. Этот пример наглядно показывает тесную связь между экспоненциальными окрестностями и полиномиально разрешимыми случаями задачи коммивояжера. Наличие такой связи, по-видимому, будет стимулировать поиск новых полиномиально разрешимых случаев для трудных комбинаторных задач, что в свою очередь может привести к дальнейшему прогрессу в области локального поиска.

Основная идея может быть представлена следующим образом. Предположим, что множество из n городов разбито на две части $x_1, \dots, x_m$ и $y_1, \dots, y_k$, так, что $k + m = n$ и $m \leq k$. Для удобства и простоты изложения добавим в первую группу $k - m$ фиктивных городов $x_{m+1}, \dots, x_k$ и рассмотрим окрестность $N(y_1, \dots, y_{k,} x_1, \dots, x_m)$, состоящую из всех циклов вида

$$y_1 x_{\tau_1} y_2 x_{\tau_2}, \dots, y_k x_{\tau_k} y_1$$

где $(\tau_1, \dots, \tau_k)$ — произвольная перестановка k элементов. Содержательно, каждый такой цикл получается вставкой городов x_i между парами городов y_j, y_{j+1}. Множество $N(y_1, \dots, y_{k,} x_1, \dots, x_m)$ имеет экспоненциальную мощность $\left[\frac{n+1}{2}\right]!$ при $2m \leq n \leq 2m + 1$. Оптимальный тур в этом множестве может быть найден с помощью решения задачи о назначениях следующим образом. При $j = 1, \dots, k$ положим

$$d_{ij} = \begin{cases} c(y_j x_i) + c(x_i y_{j+1}), & i = 1, \dots, m \\ c(y_j y_{j+1}), & i = m+1, \dots, k. \end{cases}$$

Введем переменные

$$z_{ij} = \begin{cases} 1, & \text{если } x_i \text{ располагается между } y_j, y_{j+1} \\ 0 & \text{в противном случае} \end{cases}$$

и рассмотрим задачу

$$\min \sum_{i=1}^{k} \sum_{j=1}^{k} d_{ij} z_{ij}$$

$$\sum_{i=1}^{k} z_{ij} = 1, \ \ j = 1, \dots, k,$$

$$\sum_{j=1}^{k} z_{ij} = 1, \ \ i = 1, \dots, k,$$

$$z_{ij} \in \{0,1\}, \quad i, j = 1, \dots, k.$$

Оптимальное решение задачи определяет наилучшее размещение городов $x_i, i = 1, \dots, m$, между городами $y_j, j = 1, \dots, k$, то есть оптимальный тур в множестве $N(y_1, \dots, y_k, x_1, \dots, x_m)$.

Заметим, что максимальная мощность окрестности $N(y_1, \dots, y_k, x_1, \dots, x_m)$ достигается при $m < n/2$. Точнее, пусть $m = n/2 - p$, $p \geq 0$, целое, $n \geq 5$ и $\sigma(n) \equiv n (\mathrm{mod}\ 2)$. Для действительного числа r обозначим через $[r]_0$ (соответственно $[r]_1$) максимальное целое (полуцелое, то есть число вида $q/2$ для нечетных q), не превосходящее r. Тогда максимальная мощность $N_{\max}(n)$ окрестности превосходит $\left[\frac{n+1}{2}\right]!$ и равна [25]:

$$N_{\max}(n) = (n/2 + p_0)!/2p_0!,$$

где $p_0 = \left[\sqrt{\frac{1}{8}n + \frac{9}{8}} + \frac{3}{8}\right]_{\sigma(n)}$.

Применение этой формулы позволяет получить асимптотическое выражение для мощности окрестности.

Теорема 4.12 . $N_{\max} = \Theta\left(\frac{e^{\sqrt{n/2}}\left[\frac{n+1}{2}\right]!}{n^{1/4+[1/2]_{\sigma(n)}}}\right)$.

Перейдем теперь к наиболее интригующему свойству данной окрестности, связанному к расстоянием между турами в графе окрестностей. Оказывается, что с помощью множества $N(y_1, \dots, y_k, x_1, \dots, x_m)$ можно так определить новую окрестность, что расстояние между любыми двумя турами не будет превосходить 4 [26]. Сам факт, что максимальное расстояние между турами не зависит от размерности задачи, представляется удивительным. Более того, это константа очень мала; всего 4 шага достаточно сделать, чтобы добраться из заданного решения до любого другого и, в частности, до оптимального. К сожалению, мы не знаем в какую именно сторону нужно сделать эти четыре шага, иначе задача стала бы полиномиально разрешимой. Кроме того, мы не знаем, как много локальных оптимумов типа 2-замена или city-swap содержится в такой экспоненциальной окрестности. Тем не менее, факт ограниченности диаметра свидетельствует о больших потенциальных возможностях такой окрестности.

Пусть T — произвольный гамильтонов цикл. Обозначим через $\mathcal{F}_{nm}$ семейство всех подмножеств мощности m, состоящих из несмежных вершин цикла T. Каждый элемент семейства $(x_1, \dots, x_m) \in \mathcal{F}_{nm}$ однозначно определяет множество циклов $N(y_1, \dots, y_{n-m}, x_1, \dots, x_m)$. Объединение таких множеств обозначим через $N_m(T)$. Можно показать, что при фиксированном m мощность семейства $\mathcal{F}_{nm}$ равна

$$|\mathcal{F}_{nm}| = \binom{n-m}{m} + \binom{n-m-1}{m-1},$$

то есть выражается полиномом от n и, следовательно, оптимальный цикл в множестве $N_m(T)$ можно найти за полиномиальное время. Будем говорить, что цикл R является соседним к циклу T, если в цикле T найдутся m несмежных вершин $(x_1, \dots, x_m)$ таких, что $R \in N(y_1, \dots, y_{n-m}, x_1, \dots, x_m)$.

Теорема 4.13. При $m = \lfloor (n-1)/2 \rfloor$ для любых циклов T_1 и T_5 существуют такие циклы T_2, T_3 и T_4, что T_i является соседним к T_{i-1}, $i = 2, 3, 4, 5$.

В общем случае при произвольных m длина пути в графе окрестностей G_N между двумя произвольными циклами ограничена сверху величиной $4\lceil \lfloor (n-1)/2 \rfloor / m \rceil$.

4.5. Метаэвристики

Идеи локального поиска получили свое дальнейшее развитие в так называемых *метаэвристиках*, то есть в общих схемах построения алгоритмов, которые могут быть применены практически к любой задаче дискретной оптимизации. Все метаэвристики являются итерационными процедурами и для многих из них установлена асимптотическая сходимость наилучшего найденного решения к глобальному оптимуму. В отличие от алгоритмов с оценками, метаэвристики не привязываются к специфике решаемой задачи. Это достаточно общие итерационные процедуры, использующие рандомизацию и элементы самообучения, интенсификацию и диверсификацию поиска, адаптивные механизмы управления, конструктивные эвристики и методы локального поиска. К метаэвристикам принято относить методы имитации отжига (*Simulated Annealing* (SA)), поиск с запретами (*Tabu Search* (TS)), генетические алгоритмы (*Genetic Algorithms* (GA)) и эволюционные методы (*Evolutionary Computation* (EC)), а также поиск с чередующимися окрестностями (*Variable Neighborhood Search* (VNS)), муравьиные колонии (*Ant Colony Optimization* (ACO)), вероятностные жадные алгоритмы (*Greedy Randomized Adaptive Search Procedure* (GRASP)) и др. [27].

Идея этих методов основана на предположении, что целевая функция имеет много локальных экстремумов, а просмотр всех допустимых решений невозможен, несмотря на конечность их числа. В такой ситуации нужно сосредоточить поиск в наиболее перспективных частях допустимой области. Таким образом, задача сводится к выявлению таких областей и быстрому их просмотру. Каждая из метаэвристик решает эту проблему по-своему.

Метаэвристики принято делить на траекторные методы, когда на каждой итерации имеется одно допустимое решение и осуществляется переход к следующему, и на методы, работающие с семейством (популяцией) решений. К первой группе относятся TS, SA, VNS. Траекторные методы оставляют в пространстве поиска траекторию, последовательность решений, где каждое решение является соседним для предыдущего относительно некоторой окрестности. В методах TS, SA окрестность определяется заранее и не меняется в ходе работы. Целевая функция вдоль траектории меняется немонотонно, что позволяет "выбираться" из локальных экстремумов и находить всё лучшие и лучшие приближенные решения. Элементы самоадаптации позволяют менять управляющие параметры алгоритмов, используя предысторию поиска. Более сложные методы, например, VNS, используют несколько окрестностей и меняют их систематически в целях диверсификации. Фактически, при смене окрестности происходит смена ландшафта. Сознательное изменение ландшафта, как, например, это происходит в методе шума (*Noising method*, [28]), благотворно сказывается на результатах поиска. Изучение ландшафтов, их свойств, например, изрезанности (ruggedness) позволяет давать рекомендации по выбору окрестностей.

Ко второй группе методов относятся GA, EC, ACO и др. На каждой итерации этих методов строится новое решение задачи, которое основывается уже не на одном, а на нескольких решениях из популяции. В генетических и эволюционных алгоритмах для этих целей используются процедуры скрещивания (crossover) и целенаправленных мутаций. В методах ACO применяется другая идея, основанная на сборе статистической информации о наиболее удачных найденных решениях. Эта информа-

ция учитывается в вероятностных жадных алгоритмах и подсказывает, какие именно компоненты решений (ребра графа, предприятия, элементы системы технических средств) чаще всего приводили ранее к малой погрешности. Методы этой группы, как правило, основаны на аналогиях в живой природе. Идея АСО является попыткой имитации поведения муравьев, которые почти не имеют зрения и ориентируются по запаху, оставленному предшественниками. Сильно пахнущее вещество, феромон, является для них индикатором деятельности предшественников. Оно аккумулирует в себе предысторию поиска и подсказывает дорогу к муравейнику. Попытки подглядеть у природы способы решения трудных комбинаторных задач находят все новые и новые воплощения в численных методах. Например, при инфекции организм старается подобрать (породить, сконструировать) наиболее эффективную защиту. Наблюдение за этим процессом привело к рождению нового метода – искусственным иммунным системам. Исследование жизнедеятельности пчелиного улья, где только одна матка оставляет потомство, послужило основой для новых генетических методов. Однако наибольший прогресс наблюдается на пути гибридизации, например, построения гиперэвристик, автоматически подбирающих наиболее эффективные эвристики для данного примера и симбиоз с классическими методами математического программирования.

Остановимся подробнее на последнем направлении. Здесь наблюдается достаточно широкий спектр исследований [20]:

- построение экспоненциальных по мощности окрестностей, в которых поиск наилучшего решения осуществляется за

полиномиальное время путем решения вспомогательной оптимизационной задачи;

- исследование операторов скрещивания, базирующихся на точном или приближенном решении исходной задачи на подмножестве, заданном родительскими решениями;
- гибридизация метаэвристик с точными методами, например, с методами ветвей и границ и методом динамического программирования.
- разработка новых точных методов, использующих идеи локального поиска;
- применение разных математических формулировок решаемой задачи и использование различных кодировок решений и др.

Обзор достижений в данном направлении можно найти в [27]. Ниже будут представлены пять метаэвристик, четыре из которых являются траекторными алгоритмами. Каждая из них имеет свою идею и механизм ее реализации. Несмотря на то, что эта глава посвящена задаче коммивояжера, читатель легко увидит способы адаптации данных алгоритмов к другим задачам.

4.5.1. Алгоритм имитации отжига

Экзотическое название данного алгоритма связано с методами имитационного моделирования в статистической физике, основанными на технике Монте-Карло. Исследование кристаллической решетки и поведения атомов при медленном остывании тела привело к появлению на свет стохастических алгоритмов, которые оказались чрезвычайно эффективными в комбинаторной оптимизации. Впервые это было замечено в 1983

году [29]. В настоящее время этот подход является популярным как среди практиков, благодаря своей простоте, гибкости и эффективности, так и среди теоретиков, поскольку для него удается доказать асимптотическую сходимость к глобальному оптимуму.

Алгоритм имитации отжига относится к классу пороговых алгоритмов локального поиска. На каждом шаге в окрестности текущего решения выбирается новое решение. Если разность по целевой функции между ними не превосходит заданного га t_k, то новое решение заменяет текущее. В противном случае выбирается другое соседнее решение. Общая схема пороговых алгоритмов может быть представлена следующим образом.

Пороговый алгоритм

1. Выбрать начальное решение s и положить $f^* := f(s), k := 0$.

2. Пока не выполнен критерий остановки, делать следующее:

2.1. Выбрать $s' \in N(s)$ случайным образом.

2.2. Если $f(s') - f(s) < t_k$ то $s := s'$.

2.3. Если $f^* > f(s)$, то $f^* := f(s)$.

2.4. Положить $k := k + 1$.

3. Предъявить наилучшее найденное решение.

В зависимости от способа задания последовательности $\{t_k\}$ различают три типа алгоритмов

1. *Последовательное улучшение*: $t_k = 0$ для всех k —вариант классического локального спуска с монотонным улучшением по целевой функции.

2. *Пороговое улучшение*: $t_k = c_k,\ k = 0,1,2,\ldots,\ c_k \geq c_{k+1} \geq 0$ $\lim_{k\to\infty} c_k \to 0$ — вариант локального поиска, когда допускается ухудшение по целевой функции до некоторого порога, и этот порог последовательно снижается до нуля.

3. *Имитация отжига*: t_k— неотрицательная случайная величина с математическим ожиданием $\mathbb{E}(t_k) = c_k$ — вариант локального поиска, когда допускается произвольное ухудшение по целевой функции, но вероятность такого перехода обратно пропорциональна величине ухудшения, точнее для любого $s' \in N(s)$

$$P_{ss'} = \begin{cases} 1, & \text{если } f(s') \leq f(s) \\ \exp\left(\dfrac{f(s) - f(s')}{c_k}\right), & \text{если } f(s') > f(s) \end{cases}$$

Последовательность $\{c_k\}$ играет важную роль при анализе сходимости и выбирается так, чтобы $c_k \to 0$ при $k \to \infty$. Иногда параметр c_k называют температурой, имея ввиду указанные выше истоки. Для анализа данного и последующих алгоритмов потребуются некоторые определения из теории конечных цепей Маркова [19, 30].

Определение 4.6. Пусть $\mathfrak{S}$ обозначает множество возможных исходов некоторого случайного процесса. *Цепь Маркова* есть последовательность испытаний, когда вероятность исхода в каждом испытании зависит только от результата предшествующего испытания.

Пусть $x(k)$— случайная переменная, обозначающая результат *k*-го испытания. Тогда для каждой пары $i, j \in \mathfrak{S}$ вероятность перехода от i к j при *k*-ом испытании задается выражением

$$P_{ij}(k) = \mathbb{P}\{x(k) = j \mid x(k-1) = i\}.$$

Матрица (P_{ij}) называется *переходной матрицей*. Цепь Маркова называется *конечной*, если множество исходов конечно, и *однородной*, если переходные вероятности не зависят от номера шага k.

Для алгоритма имитации отжига испытание состоит в выборе очередного решения на *k*-м шаге. Вероятность перехода зависит от текущего решения s_k, так как переход возможен только в соседнее решение. Она не зависит от того, каким путем добрались до этого решения. Таким образом, последовательность решений $\{s_k\}$ образует цепь Маркова. Эта цепь конечна, так как множество допустимых решений конечно. Для каждой пары $s, s' \in Sol$ вероятность перехода от s к s' задается выражением

$$P_{ss'}(k) = \begin{cases} D_{ss'}A_{ss'}(c_k), & \text{если } s \neq s', \\ 1 - \sum_{l \neq s} D_{sl}A_{sl}(c_k), & \text{если } s = s'. \end{cases}$$

где $D_{ss'}$ есть вероятность выбора решения s' из окрестности $N(s)$, $A_{ss'}(c_k)$— вероятность принятия решения s'. Если в окрестности решения выбираются равновероятно, то

$$D_{ss'} = \begin{cases} \dfrac{1}{|N(s)|}, & \text{если } s' \in N(s), \\ 0 & \text{в противном случае} \end{cases}$$

а вероятность принятия s' определяется как

$$A_{ss'}(c_k) = \exp\frac{\max\{0,\ f(s') - f(s)\}}{c_k}.$$

Если $c_k = c$, то цепь Маркова является однородной. В общем случае алгоритм имитации отжига порождает конечную неод-

нородную цепь Маркова. При исследовании асимптотического поведения алгоритмов важную роль играет так называемое стационарное распределение, вектор q_i размерности $|\mathfrak{I}|$, компоненты которого определяются как

$$q_i = \lim_{k\to\infty} \mathbb{P}\{x(k) = i \mid x(0) = j\}, i \in \mathfrak{I},$$

если такой предел существует и не зависит от j. Величина q_i равна вероятности оказаться в состоянии i после достаточно большого числа шагов.

Теорема. Пусть $c_k = c$ для всех k и для любой пары $s, s' \in Sol$ найдется положительное целое число p и такие решения $l_0, l_1, \dots, l_p \in Sol$, что $l_0 = s,\ l_p = s'$ и $D_{l_k l_{k+1}} > 0$ для $0 \le k \le p-1$. Тогда для цепи Маркова, порожденной алгоритмом имитации отжига, существует единственное стационарное распределение, компоненты которого задаются формулой

$$q_s(c) = \frac{\exp(-f(s)/c)}{\sum_{l\in Sol}\exp(-f(l)/c)}, \qquad s \in Sol,$$

и

$$\lim_{c\downarrow 0} q_s(c) = \begin{cases} 1/|S^*|, & \text{если } s \in S^*, \\ 0 & \text{в противном случае.} \end{cases}$$

Последнее равенство, по сути, означает, что с ростом числа итераций вероятность оказаться в точке глобального оптимума стремится к 1, если значение порога c_k стремится к нулю, то есть

$$\lim_{c\downarrow 0}\lim_{k\to\infty} \mathbb{P}_c\{s_k \in S^*\} = 1.$$

На практике, конечно, нет возможности выполнить бесконечное число итераций. Поэтому приведенная теорема является только источником вдохновения при разработке алгоритмов. Существует много эвристических способов выбора конечной последова-

тельности $\{c_k\}$ с целью повышения вероятности обнаружения глобального оптимума.

В большинстве работ следуют рекомендациям первопроходцев и используют схему геометрической прогрессии:

1. Начальное значение: $c_0 = \Delta f_{\max}$— максимальная разность между соседними решениями.

2. Понижение порога: $c_{k+1} = \alpha c_k,\ k = 0, 1, \dots, K-1$, где α— положительная константа, достаточно близкая к 1.

3. Финальное значение: $c_K > 0$ определяется либо по числу сделанных изменений, либо как максимальное c_k, при котором алгоритм не меняет текущее решение в течение заданного числа шагов. При каждом значении c_k алгоритм выполняет порядка $|N|$ шагов, не меняя значение порога. Общая схема имитации отжига может быть представлена следующим образом.

Алгоритм SA

1. Выбрать начальное решение s и положить $f^* \coloneqq f(s),\ \mathrm{s}^* \coloneqq \mathrm{s}$, определить начальную температуру t и коэффициент α.

2. Повторять, пока не выполнен критерий остановки.

2.1. Цикл $k \coloneqq 1, \dots, |N|$

Выбрать решение $s' \in N(s)$ случайным образом.

Вычислить $\Delta = f(s) - f(s')$.

Если $\Delta \geq 0$, то $s \coloneqq s'$ иначе $s \coloneqq s'$ с вероятностью $\exp(\Delta/t)$.

Если $f^* > f(s)$, то $f^* \coloneqq f(s),\ \ s^* \coloneqq s$.

2.2 Понизить температуру: $t \coloneqq \alpha t$.

3. Предъявить наилучшее найденное решение s^*.

На рис. 4.16 показано типичное поведение алгоритма. На первых итерациях при высокой температуре наблюдается большой разброс по целевой функции. Процесс поиска чем-то напоминает броуновское движение.

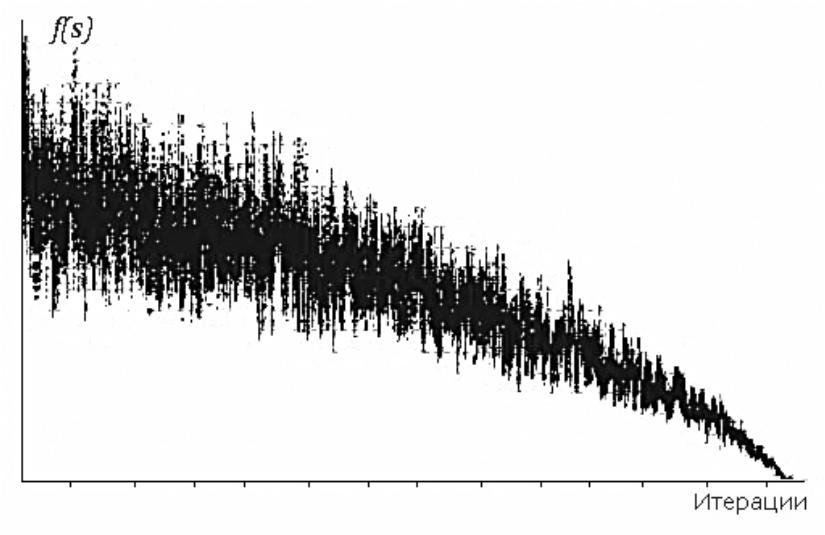


Рис. 4.16. Поведение алгоритма SA

По мере понижения температуры разброс уменьшается, и наблюдается явная тенденция к падению среднего значения целевой функции. При низкой температуре наблюдается стагнация процесса. Все чаще и чаще не удается перейти в соседнее решение, и алгоритм останавливается в одном из локальных оптимумов.

Среди пороговых алгоритмов следует выделить еще один с забавным названием *великий потоп* (Great Deluge). Правда для задач на минимум его следует называть скорее *великая засуха*. На каждой итерации этого алгоритма имеется некоторое допустимое решение задачи s и в его окрестности, как и

раньше, выбирается решение s' случайным образом. Если $f(s') < f(s)$, то выполняется переход в новое решение. В противном случае значение $f(s')$ сравнивается с *уровнем воды W.* Если $f(s') \leq W$, то все равно выполняется переход в новое решение. Если же переход не выполнен, то выбирается новое соседнее решение. Порог W сначала устанавливается достаточно большим, чтобы были возможны любые переходы. Затем этот порог уменьшается на каждой итерации на заданную величину ΔW. Процесс останавливается в локальном минимуме со значением выше уровня воды. Общая схема алгоритма не сильно отличается от имитации отжига, но приводит к качественно другой картине.

Алгоритм GD

1. Выбрать начальное решение s и положить $f^* := f(s)$, $s^* := s$, определить начальный порог W и скорость его падения ΔW.
2. Повторять, пока не выполнен критерий остановки.
   2.1. Выбрать решение $s' \in N(s)$ случайным образом.
   2.2. Вычислить $\Delta = f(s) - f(s')$.
   2.3. Если $\Delta \geq 0$, то $s := s'$ иначе если $f(s') \leq W$, то $s := s'$.
   2.4. Если $f^* > f(s)$, то $f^* := f(s)$, $s^* := s$.
   2.5. Понизить порог: $W := W - \Delta W$.
3. Предъявить наилучшее найденное решение s^*.

На рис. 4.17 показано типичное поведение алгоритма. Падение порога на каждой итерации вынуждает искать направления спуска. Как рыба, которая не хочет оказаться на суше, мы ищем все более и более глубокие места. Заметим, что для реализации алгоритма достаточно уметь вычислять значение целевой функции и определять окрестность. Параметры W и ΔW

легко настраиваются под специфику задачи, что открывает широкий простор для приложений.

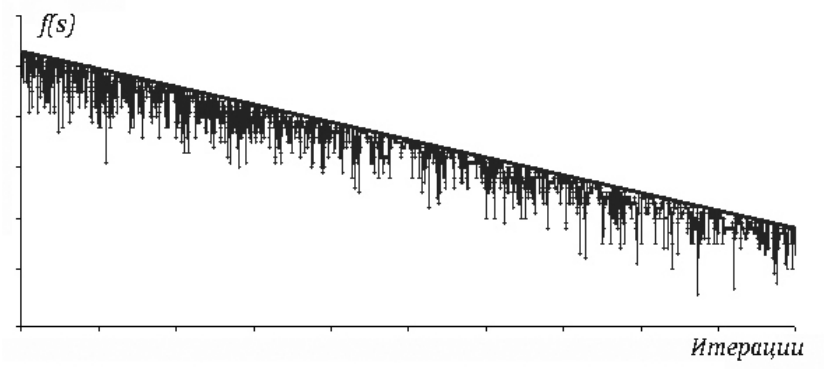


Рис. 4.17. Поведение алгоритма GD

4.5.2. Поиск с запретами

Основоположником алгоритма поиска с запретами является Ф. Гловер, который в 1986 году предложил принципиально новую схему локального поиска [31]. Она позволяет алгоритму не останавливаться в точке локального оптимума, как это предписано в стандартном алгоритме локального улучшения, а путешествовать от одного оптимума к другому, в надежде найти среди них глобальный оптимум. Основным механизмом, позволяющим алгоритму выбираться из локального оптимума, является список запретов. Он строится по предыстории поиска, то есть по нескольким последним решениям, $s_k, s_{k-1}, \dots, s_{k-l+1}$, и запрещает часть окрестности текущего решения $N(s_k)$. Точнее, на каждом шаге алгоритма новое решение s_{k+1} является оптимальным решением подзадачи

$$f(s_{k+1}) = min\{f(s) \mid s \in N(s_k) \setminus Tabu_l(s_k)\}.$$

Список запретов учитывает специфику задачи и, как правило, запрещает использование тех "фрагментов" решения (ребер графа, координат вектора, цвета вершин), которые менялись на последних l шагах алгоритма. Константа $l \geq 0$ определяет его память. При "короткой памяти" $(l = 0)$ получаем стандартный алгоритм локального улучшения.

Существует много вариантов реализации этой идеи. Приведем один из них, для которого удается установить асимптотические свойства. Рассмотрим рандомизированную окрестность $N_p(s)$. Каждый элемент окрестности $N(s)$ включается в множество $N_p(s)$ с вероятностью p независимо от других элементов. С ненулевой вероятностью множество $N_p(s)$ может совпадать с $N(s)$, может оказаться пустым или содержать ровно один элемент. Общая схема алгоритма поиска с запретами может быть представлена следующим образом.

Алгоритм TS

1. Выбрать начальное решение s и положить $f^* := f(s)$, $s^* := s,\ Tabu_l(s) := \emptyset$.

2. Повторять, пока не выполнен критерий остановки.

   2.1. Сформировать окрестность $N_p(s)$.

   2.2. Если $N_p(s) \neq \emptyset$, то найти такое решение s', что

   $$f(s') = \min\{f(j) \mid j \in N_p(s) \setminus Tabu_l(s)\}.$$

   2.3. Если $f^* > f(s'),$ то $f^* := f(s')$ и $s^* := s'$.

   2.4. Положить $s := s'$ и обновить список запретов.

3. Предъявить наилучшее найденное решение s^*.

Параметры p и l являются управляющими для данного алгоритма. Выбор их значений зависит от размерности задачи и мощности окрестности. Примеры адаптации этой схемы к разным задачам комбинаторной оптимизации можно найти, например, в [20].

Пусть $l = 0$. Тогда алгоритм *TS* порождает конечную однородную цепь Маркова. Можно показать, что в этом случае переходные вероятности $P_{ij},\ i,j \in \Im,$ определяются равенством

$$P_{ij} = \begin{cases} 0, & \text{если } j \notin N(i), \\ p(1-p)^{k-1}, & \text{если } j \in N(i) \text{ и } f(j) = \min\limits_{\tau \in N(i)}{}^{k}(f(\tau)), \end{cases}$$

где $\min^k$ означает *k*-й минимальный элемент множества. Если граф окрестностей G_N строго связен, то для такой цепи Маркова существует единственное стационарное распределение, и вероятность не найти глобальный оптимум стремится к нулю со скоростью геометрической прогрессии. Заметим, что для стационарного распределения не получено точной формулы, как это удалось сделать для алгоритма имитации отжига. Утверждается только существование и единственность такого распределения. При малых размерностях его можно получить численно из соотношений

$$q_i = \sum_{j \in \Im} q_j P_{ij},\ \ i \in \Im, \quad \sum_{j \in \Im} q_j = 1.$$

Аналитического выражения для решения такой системы, по-видимому, не существует. Тем не менее, исследование этого распределения и, в частности, его компонент, соответствующих глобальному оптимуму представляет несомненный интерес.

Пусть $l > 0$. В этом случае последовательность решений уже не является цепью Маркова, так как выбор следующего ре-

шения зависит от предыстории. Однако, если рассмотреть множество кортежей $\langle s_k, s_{k-1}, \ldots, s_{k-l+1} \rangle$ длины l, то такая последовательность, порождаемая алгоритмом, уже будет конечной однородной цепью Маркова. Матрица переходных вероятностей имеет теперь значительно большую размерность, но ее элементы будут определяться по сути теми же формулами, что и раньше.

Список запретов оказывает существенное влияние на поведение алгоритма. В частности, если на некотором шаге все соседние решения оказываются под запретом, то процесс поиска прекращается. Таким образом, даже свойство строгой связности графа G_N не гарантирует желаемого асимптотического поведения. Тем не менее, при определенных условиях на длину этого списка удается обеспечить сходимость по вероятности наилучшего найденного решения к глобальному оптимуму.

На рис. 4.18 показано типичное поведение алгоритма *TS*. На первых итерациях он ведет себя как стандартный алгоритм локального улучшения. Получив первый локальный минимум, алгоритм вынужден выполнить несколько шагов с ухудшением, после чего снова открывается путь к новым локальным оптимумам. Рандомизация окрестности привносит определенное разнообразие в процесс поиска и предотвращает зацикливание. Более того, она позволяет сократить трудоемкость одного шага алгоритма и часто повышает вероятность нахождения глобального оптимума за предписанное число итераций. Чем сложнее и запутаннее пример, тем меньшую долю окрестности следует просматривать. Как правило, значения параметра p в интервале от 0.2 до 0.3 приводят к хорошим результатам даже при небольшом числе итераций.

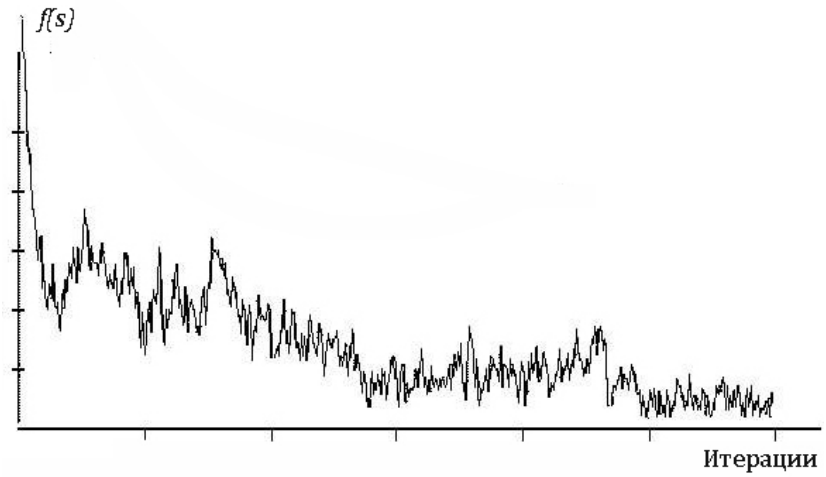


Рис. 4.18. Поведение алгоритма *TS*

4.5.3.Локальный поиск с чередующимися окрестностями

Идея этого подхода предложена Н. Младеновичем и П. Хансеном в 1997 г. [32]. Она состоит в систематическом изменении функции окрестности и соответствующем изменении ландшафта в ходе локального поиска. Существует много вариантов ее реализации, особенно для задач большой размерности. Легкость адаптации к различным моделям и высокая эффективность, особенно в сочетании с другими метаэвристиками, объясняет ее конкурентоспособность при решении NP-трудных задач.

Обозначим через $N_k, k = 1, \dots, k_{\max}$, конечное множество функций окрестностей, предварительно выбранных для локального поиска. Предлагаемый алгоритм опирается на следующие три тезиса.

– Локальный минимум для одной окрестности не обязательно является локальным минимумом для другой.

– Глобальный минимум является локальным минимумом относительно любой окрестности.

– Для многих задач дискретной оптимизации локальные минимумы относительно одной или нескольких окрестностей расположены достаточно близко друг к другу.

Последнее тезис является эмпирическим. Он позволяет сузить область поиска глобального оптимума, используя информацию об уже обнаруженных локальных оптимумах. Именно эта идея лежит в основе алгоритмов связывающих путей, различных операторов скрещивания для генетических алгоритмов и многих других подходов. Алгоритм чередующихся окрестностей неявно использует этот тезис. Общая схема алгоритма может быть представлена следующим образом.

Алгоритм VNS

1. Выбрать окрестности $N_k,\ k = 1, \dots, k_{\max}$, и начальное решение s.

2. Повторять, пока не выполнен критерий остановки.

2.1. Положить $k \coloneqq 1$.

2.2. Повторять, пока $k \leq k_{\max}$:

(a) Выбрать $s' \in N_k(s)$ случайным образом.

(b) Применить локальный поиск с начального решения s'. Полученный локальный оптимум обозначим s''.

(c) Если $f(s'') < f(s),$ то $s \coloneqq s''$ иначе $k \coloneqq k + 1$.

3. Предъявить наилучшее найденное решение s.

В качестве критерия остановки может использоваться максимальное время счета или число итераций без смены лучшего найденного решения. Наряду с последовательным порядком смены окрестностей могут использоваться различные групповые стратегии. Случайный выбор на шаге 2.2.(*a*) применяется во избежание зацикливания. На шаге 2.2.(*b*) локальный поиск может использовать различные алгоритмы SA, TS и др.

Существует много вариантов этой схемы. Например, на шаге 2.2.(*c*) переход осуществляется только при уменьшении значения целевой функции. Это условие можно заменить по аналогии с алгоритмом имитации отжига и разрешать движение в точку с большим значением целевой функции с некоторой вероятностью, зависящей от этого ухудшения. Идею наискорейшего спуска можно осуществить путем следующих изменений. Вместо шага 2.2, на котором выполняется переход согласно *k*-й окрестности, следует выбрать наилучший переход по всем рассматриваемым окрестностям. Другой способ состоит в модификации шага 2.2.(*a*) и выборе не случайного решения, а наилучшего в -й окрестности. Порядок смены окрестностей тоже может варьироваться.

На рис. 4.19 показано типичное поведение алгоритма VNS. Высокие пики на последних итерациях соответствуют переходам на шаге 2.2.(*a*), где выбирается случайным образом соседнее решение. Несмотря на заметное ухудшение по целевой функции, локальный поиск на шаге 2.2.(*b*) быстро возвращается к хорошим решениям. Тенденция к последовательному улучшению рекорда хорошо видна на этом рисунке.

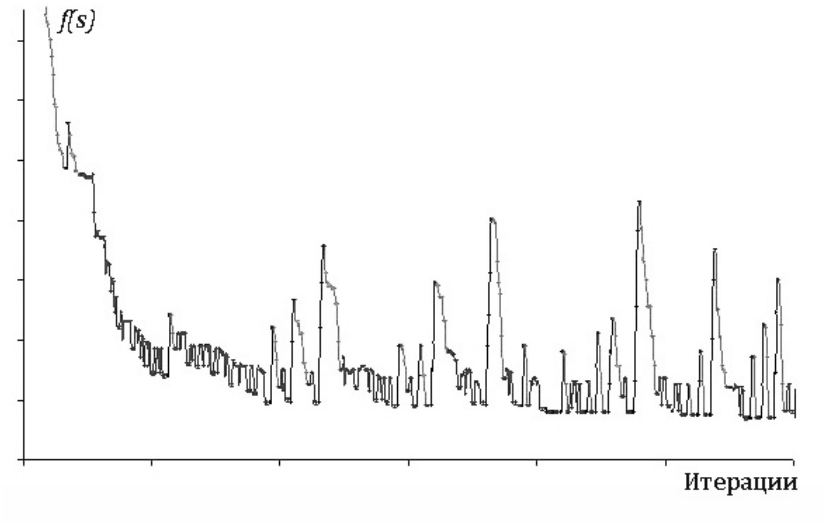


Рис. 4.19. Поведение алгоритма VNS

4.5.4. Генетические алгоритмы

Идея генетических алгоритмов заимствована у живой природы и состоит в организации эволюционного процесса, конечной целью которого является получение оптимального решения сложной комбинаторной задачи. Разработчик генетических алгоритмов выступает в данном случае как "создатель", который должен правильно установить законы эволюции, чтобы достичь этой цели. Впервые эти нестандартные идеи были применены к решению оптимизационных задач в середине 70-х годов прошлого столетия. Примерно через десять лет появились первые теоретические обоснования этого подхода. На сегодняшний день генетические алгоритмы доказали свою конкурентоспособность при решении многих NP-трудных задач, особенно в приложениях, где математические модели имеют сложную структуру и применение стандартных методов типа

ветвей и границ, динамического или линейного программирования крайне затруднено.

Стандартный генетический алгоритм начинает свою работу с формирования начальной *популяции* — конечного набора допустимых решений задачи. Эти решения могут быть выбраны случайным образом или получены с помощью конструктивных алгоритмов. Выбор начальной популяции не имеет значения для сходимости процесса в асимптотике, однако формирование "хорошей" начальной популяции (например, из множества локальных оптимумов) может заметно сократить время достижения глобального оптимума.

На каждом шаге эволюции с помощью вероятностного оператора *селекции* выбираются два решения, *родители* s_1, s_2. Оператор *скрещивания* по этим решениям строит новое решение s', которое затем подвергается небольшим случайным модификациям. Их принято называть мутациями. Затем решение добавляется в популяцию, а решение с наименьшим значением целевой функции удаляется из популяции. Общая схема такого алгоритма может быть записана следующим образом.

Алгоритм GA

1. Выбрать начальную популяцию.
2. Повторять, пока не выполнен критерий остановки.
    2.1. Выбрать родителей s_1, s_2 из популяции.
    2.2. Построить новое решение s' по решениям s_1, s_2.
    2.3. Модифицировать s'.
    2.4. Обновить популяцию.

3. Предъявить наилучшее решение в популяции.

Остановимся подробнее на основных операторах этого алгоритма: селекции, скрещивании и мутации. Среди операторов селекции наиболее распространенными являются два вероятностных оператора: *пропорциональной* и *турнирной* селекции. При пропорциональной селекции вероятность выбрать решение в качестве одного из родителей зависит от качества этого решения: чем меньше значение целевой функции, тем выше вероятность, а сумма вероятностей равна 1. При турнирной селекции формируется случайное подмножество из элементов популяции и среди них выбирается один элемент с наименьшим значением целевой функции. Турнирная селекция имеет определенные преимущества. Она не теряет своей избирательности, когда в ходе эволюции все элементы популяции становятся примерно равными по значению целевой функции. Операторы селекции строятся таким образом, чтобы с ненулевой вероятностью любой элемент популяции мог бы быть выбран в качестве одного из родителей. Более того, допускается ситуация, когда оба родителя представлены одним и тем же элементом популяции. Возвращаясь к алгоритму *VNS*, представляется разумным брать в качестве одного из родителей наилучший элемент популяции.

Как только два решения выбраны, к ним применяется вероятностный оператор скрещивания (*crossover*). Существует много различных версий этого оператора, среди которых простейшим, по-видимому, является однородный оператор. По родительским решениям он строит новое решение, присваивая каждой координате этого вектора с вероятностью 0,5 соответствующее значение одного из родителей. Если родительские век-

тора совпадали, скажем, по первой координате, то новый вектор «унаследует» это значение. Геометрически, для булевых векторов, оператор скрещивания случайным образом выбирает в гиперкубе вершину s', которая принадлежит минимальной грани, содержащей вершины s_1, s_2. Можно сказать, что оператор скрещивания старается выбрать новое решение где-то между родительскими, полагаясь на удачу. Более аккуратная процедура могла бы выглядеть таким образом. Новое решение выбирается как оптимальное решение исходной задачи на соответствующей грани гиперкуба. Если расстояние между s_1, s_2 достаточно велико, то задача оптимального скрещивания может совпадать с исходной. Тем не менее, даже приближенное решение этой задачи вместо случайного выбора заметно улучшает работу генетического алгоритма.

Оператор мутации с заданной вероятностью p_m меняет значение каждой координаты на противоположное. Например, вероятность того, что вектор $(0,0,0,0,0)$ в ходе мутации перейдет в вектор $(1,1,1,0,0)$ равна $p_m\, p_m\, p_m\, (1-p_m)\, (1-p_m) > 0$. Таким образом, с ненулевой вероятностью новое решение s' может перейти в любое другое решение, что и объясняет асимптотические свойства алгоритма. Отметим, что модификация решения s' может состоять не только в случайной мутации, но и в частичной перестройке решения алгоритмами локального поиска. Применение локального улучшения позволяет генетическому алгоритму сосредоточиться только на локальных оптимумах. Множество локальных оптимумов может оказаться экспоненциально большим, и на первый взгляд такой вариант алгоритма не будет иметь преимуществ. Однако экспериментальные исследования распределения локальных оптимумов свидетельствуют о высокой концентрации их в непосредственной

близости от глобального оптимума. Это наблюдение отчасти объясняет работоспособность генетических алгоритмов. Если в популяции собираются локальные оптимумы, которые сконцентрированы в одном месте, и очередное решение s' выбирается где-то между двумя произвольными локальными оптимумами, то такой процесс имеет много шансов найти глобальный оптимум.

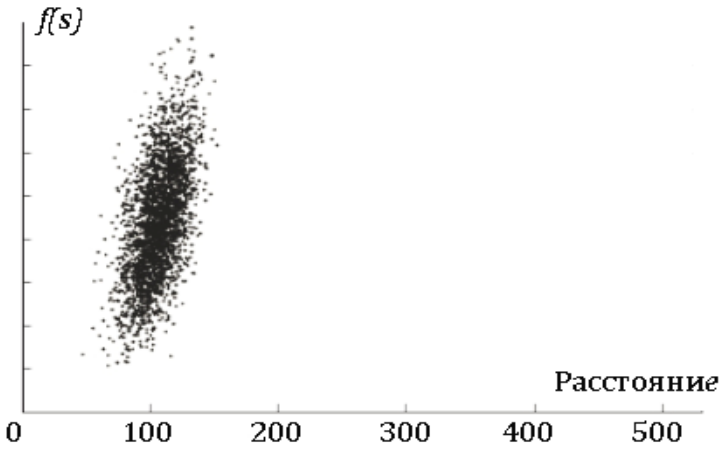


Рис. 4.20. Распределение локальных оптимумов

На рис. 4.20 показаны результаты численных экспериментов для задачи коммивояжера при $n = 532$. Каждая точка на рисунке соответствует локальному оптимуму по окрестности 2-замена. Ось y показывает значение целевой функции для локальных оптимумов. По оси x откладывается расстояние от локального оптимума до глобального. Максимально возможное расстояние равно 532. Однако среднее расстояние не превышает 200. Это означает, что все локальные оптимумы имеют более 300 общих ребер с глобальным оптимумом. Они все достаточно близки друг к другу [33].

В заключение этой главы хочется еще раз подчеркнуть тот факт, что разработка численных методов для решения NP-трудных задач дискретной оптимизации дает широкий простор для творчества. Учет специфики решаемой задачи может подсказать структуру окрестностей, выбор схемы и идею самого алгоритма, но единого наилучшего для всех задач метода построить не удастся, и теоремы о *небесплатных завтраках* (No Free Lunch Theorems) ясное тому подтверждение.

4.6 Упражнения

Упражнение 4.1. При построении примитивной нижней оценки для задачи коммивояжера можно сначала вычислять плату за въезд, в потом плату за выезд, но можно и наоборот, сначала плату за выезд, а потом плату за въезд. Покажите, что эти два варианта могут давать разные нижние оценки.

Упражнение 4.2. Покажите, что нижняя оценка линейного программирования из раздела 4.3.2. не хуже оптимального значения в целочисленной задаче о назначениях.

Упражнение 4.3. Приведите пример неотрицательной матрицы расстояний, для которой разница между оптимумом в задаче коммивояжера и оптимумом в задаче о назначениях была бы сколь угодно велика.

Упражнение 4.4. Усовершенствуйте полиномиальный алгоритм для задачи о назначениях, понизив его трудоемкость до $O(n^3)$.

Упражнение 4.5. Приведите пример неотрицательной матрицы расстояний, для которой разница между оптимумом в задаче

коммивояжера и длиной минимального 1-дерева была бы сколь угодно велика.

Упражнение 4.6. Покажите, что минимальные 1-деревья дают нижнюю оценку для задачи коммивояжера, которая не больше оптимума задачи о назначениях.

Упражнение 4.7. Докажите, что оценка 3/2 для алгоритма Кристофидеса-Сердюкова является неулучшаемой.

Упражнение 4.8. Известно, что выбор наилучшего эйлерова цикла при построении приближенного алгоритма с оценкой 3/2 по Кристофидесу и Сердюкову является NP-трудной задачей. Постройте для ее решения генетический алгоритм, алгоритм поиска с запретами и алгоритм имитации отжига.

Упражнение 4.9. Для задачи коммивояжера разработайте гибридную схему генетического алгоритма и поиска с запретами, генетического алгоритма и имитации отжига, локального поиска с чередующимися окрестностями и поиска с запретами.

Упражнение 4.10. Постройте генетический алгоритм и алгоритм поиска с чередующимися окрестностями для задачи Штейнера.

Упражнение 4.11. В открытой задаче коммивояжера требуется посетить все города, но не возвращаться в исходный город. Как решить эту задачу с помощью алгоритма для классической задачи коммивояжера.

Упражнение 4.12. Покажите, что открытая задача коммивояжера является NP–трудной.

Упражнение 4.13. Сформулируйте задачу обхода конем всех полей шахматной доски как открытую задачу коммивояжера.

Упражнение 4.14. В задаче коммивояжера на узкое место требуется минимизировать длину максимального ребра вместо суммы длин ребер. Запишите эту задачу в терминах целочисленного линейного программирования.

Упражнение 4.15. Покажите, что задача коммивояжера на узкое место является NP–трудной.

Упражнение 4.6. Предположим, что коммивояжер должен посетить город *i* сразу (не обязательно сразу) после города *j*. Запишите эту задачу в терминах целочисленного линейного программирования.

Упражнение 4.17. Предположим, что коммивояжер должен посещать город *i* не раньше момента времени a_i и не позже b_i, $1 \le i \le n$. Если он приезжает раньше срока, то вынужден ждать. Запишите эту задачу в терминах целочисленного линейного программирования.

Упражнение 4.18. Предположим, что коммивояжер посещает города с целью доставки грузов. В городе i=1 находится склад. В каждый город j требуется доставить q_j единиц груза, j>1. У коммивояжера имеется транспортное средство грузоподъемностью Q и $q_j \le Q$ для всех j. Коммивояжер может возвращаться на склад, брать груз и снова направляться в непосещенные города. Его цель – минимизировать пройденное расстояние, посетить все города и вернуться на склад. Запишите эту задачу в терминах целочисленного линейного программирования.

Упражнение 4.19. В условиях предшествующей задачи будем предполагать, что коммивояжер может посещать любой город много раз и условие $q_j \le Q$ может нарушаться. Запишите эту за-

дачу в терминах частично-целочисленного линейного программирования.

Упражнение 4.20. Покажите, что задача коммивояжера с экспоненциальной окрестностью из раздела 4.4.4. принадлежит классу PLS.

Упражнение 4.21. Приведите пример задачи распознавания, которая не принадлежит классу NP (в предположении, что P≠NP).

Упражнение 4.22. Приведите пример задачи распознавания, которая не принадлежит классу co-NP (в предположении, что P≠NP).

Упражнение 4.23. Приведите пример задачи распознавания, которая не принадлежит классам NP и co-NP (в предположении, что P≠NP).

Упражнение 4.24. Покажите, что если оптимизационная задача является NP-трудной, то существование для нее точной полиномиально проверяемой окрестности влечет совпадение классов NP и co-NP.

Упражнение 4.25. Покажите, что если оптимизационная задача является NP-трудной в сильном смысле, то существование для нее точной полиномиально проверяемой окрестности влечет совпадение классов P и NP.

Библиографический список

1. Кристофидес Н. Теория графов. Алгоритмический подход. М: Мир, 1978.
2. Wichmann B.A. Ackermann's function: a study in the efficiency of calling procedures // BIT, 1976, 16, 103-110.
3. Monma C., Suri S. Transitions in geometric minimum spanning trees // Discrete ans Computational Geometry, 1992, 8(3), 265-293.
4. Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи. М.: Мир, 1982.
5. Дементьев В. Т., Ерзин А. И., Ларин Р. М., Шамардин Ю. В. Задачи оптимизации иерархических структур. Новосибирск: Изд-во Новосибирского ун-та, 1996.
6. Robins G., Zelikovsky A. Improved Steiner tree approximation in graphs // Proc. 10th Ann. ACM-SIAM Symp. on Discrete Algorithms, ACM-SIAM, 2000, 770-779.
7. Hanan M. On Steiner's problem with rectilinear distance // SIAM J. Applied Math., 1966, 14, 255–265.
8. Hwang F. K. On steiner minimal trees with rectilinear distance // SIAM J. Applied Math., 1976, 30 (1), 104–114.
9. Zelikovsky A. Z. An 11/8-approximation algorithm for the Steiner problem on networks with rectilinear distance // Janos Bolyai Mathematica Societatis Conf.: Sets, Graphs, and Numbers, 1992, 733–745.
10. Erzin A. I., Cho J. D. A Deep-Submicron Steiner Tree // Mathematical and Computer Modelling, 2000, 31, 215-226.

11. Ерзин А. И., Чо Д. Д. Задача построения синхронизирующего сигнального дерева // Автоматика и телемеханика, 2003, № 3, 163-176.

12. Ерзин А. И. Введение в исследование операций: Учеб. Пособие. Новосибирск: НГУ, 2006.

13. Korte B, Vugen J. Combinatorial optimization. Berlin: Springer, 2007.

14. Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К. Алгоритмы. Построение и анализ. 2 издание. М.: Вильямс, 2009.

15. Deineko V., Tiskin A. Fast minimum-weight double-tree shortcutting for metric TSP: is the best one good enough? // ACM Journal on Experimental Algorithmics. 2009. Vol. 14. N 4.6.

16. Сигал И.Х., Иванова А.П. Введение в прикладное дискретное программирование: модели и вычислительные алгоритмы. М.:ФИЗМАТЛИТ, 2007.

17. Пападимитриу Х., Стайглиц К. Комбинаторная оптимизация. Алгоритмы и сложность. М.: Мир, 1985.

18. Алексеева Е.В. Построение математических моделей целочисленного линейного программирования. Примеры и задачи. Новосибирск: НГУ, 2012.

19. Yannakakis M. Computational complexity. In: Aarts E., Lenstra J. (Eds.) Local search in combinatorial optimization. NY:Wiley, 1997.

20. Кочетов Ю.А. Методы локального поиска для дискретных задач размещения Модели и алгоритмы. Saarbrucken: Lambert Academic Publishing, 2011.

21. Ю.А. Кочетов. Вычислительные возможности локального поиска в комбинаторной оптимизации // Журнал вычислительной

математики и математической физики 2008 т.48, № 5. С. 788-807.

22. Grover L.K. Local search and the local structure of NP-complete problems // Operations Research Letters. 1992. Vol. 12, N 4. P. 235–244.

23. Angel E., Zissimopoulos V. On the classification of NP-complete problems in terms of their correlation coefficient // Discrete Appl. Math. 2. V. 9. P. 261–277.

24. Burkard R.E., Deineko V.G., Woeginger G.J. The traveling salesman problem and the PQ-tree // Lecture Notes in Computer Science, Vol. 1084. Berlin: Springer, 1996. P. 490–504.

25. Gutin G. Exponential neighborhood local search for the traveling salesman problem // Comput. Oper. Res. 1999. Vol. 26. P. 313–320.

26. Gutin G., Yeo A. Small diameter neighborhood graphs for the traveling salesman problem: four moves from tour to tour // Comput. Oper. Res. 1999. Vol. 26. P. 321–327.

27. Talbi El-G., Brotcorne L. Metaheuristics for Bi-level Optimization (Studies in Computational Intelligence). Berlin: Springer, 2013

28. Charon I. Hudry H. The noising methods. A survey. In: Ribeiro C.C. Hansen P. (eds.) Meta-heuristics: Advances and Trends in Local Search Paradigms for Optimization. Boston: Kluwer. Acad. Publ., 1998. P. 245–262.

29. KirkpatrickS., Gelatt C.D., Vecchi M.P. Optimization by simulated annealing // Science. 1983. Vol. 220, P.671–680.

30. Кемени Дж., Снелл Дж. Конечные цепи Маркова. М.: Наука, 1970.

31. Glover F., Laguna M. Tabu Search. Dordrecht: Kluwer Acad. Publ., 1997.

32. Mladenović N and Hansen P. Variable neighbourhood search, Computers and Operations Research, 1997. Vol. 24. P. 1097–1100.

33. Boese K.D., Kahng A.B., Muddu S. A new adaptive multi–start technique for combinatorial global optimizations // Oper. Res. Lett. 1994. Vol. 16, N 2. P.101–114.

34. Wolsey L. A. Integer Programming. N. Y.: John Wiley & Sons, Inc, 1998.