

ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ В ДИСКРЕТНОЙ МАТЕМАТИКЕ

DOI 10.17223/20710410/25/10

УДК 510.52

ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ ПОСТРОЕНИЯ КОМПОЗИЦИОННЫХ МОДЕЛЕЙ ЛИПШИЦ-ОГРАНИЧЕННЫХ ОТОБРАЖЕНИЙ

И. С. Калинин

*Национальный исследовательский университет «МИЭТ», г. Москва, Россия***E-mail:** gaminot@gmail.com

Работа посвящена вопросам численного построения композиционных моделей липшиц-ограниченных сюръективных функций одного аргумента. Композиционные модели являются частным случаем функциональной аппроксимации, получаемым путём композиции функций из заданного множества. Доказывается NP-трудность задачи построения оптимальной композиционной модели при заданном множестве функций, используемых для построения модели, и определённой приближаемой функции. Рассматриваются различные алгоритмы нахождения приближённых композиционных моделей, часть из которых имеет полиномиальную сложность; оцениваются возможности применения данных подходов.

Ключевые слова: композиция функций, композиционные модели, NP-полнота, липшиц-ограниченность, вычислительная сложность.

Введение

Композиционные модели используются в различных областях науки и техники, в том числе при представлении и анализе экспериментальных данных, для оптимизации процессов вычислений, анализа программного обеспечения методом «чёрного ящика», получения эквивалентных функциональных преобразований. Примеры применения композиционных моделей приводятся в работах [1–3]. Существенным недостатком композиционных моделей является отсутствие вычислительно эффективных алгоритмов их получения в общем случае. Известны эффективные алгоритмы для некоторых подклассов отображений, например для полиномов [4].

Точной композиционной моделью длины n для отображения f , построенной по системе функций $F = \{g_1, g_2, \dots, g_m\}$, назовём композицию функций $g_{i_1}(\dots g_{i_n}(x))$, такую, что $g_{i_1}(\dots g_{i_n}(x)) = f$. Оптимальной композиционной моделью длины n будем называть композицию $g_{i_1}(\dots g_{i_n}(x))$, такую, что по заданной метрике μ достигается $\min(\mu(g_{i_1}(\dots g_{i_n}(x)), f))$. Приближённой композиционной моделью будем называть результат процесса минимизации функции $H(i_1, \dots, i_n) = \mu(g_{i_1}(\dots g_{i_n}(x)), f)$ по набору индексов (i_1, \dots, i_n) , $i_j \in \{1, \dots, m\}$, $j = 1, \dots, n$, с использованием методов, не гарантирующих глобальной оптимальности найденного решения в общем случае.

Основной целью работы является доказательство отсутствия эффективных алгоритмов (NP-трудности) построения оптимальных композиционных моделей в случае, если функции из $F \cup \{f\}$ являются липшиц-ограниченными сюръективными отображениями $[0, 1] \rightarrow [0, 1]$. Под липшиц-ограниченными отображениями понимаются такие

функции, что существует константа L , для которой верно

$$\forall x_1, x_2 \in [0, 1] (|f(x_1) - f(x_2)| \leq L|x_1 - x_2|).$$

В работе NP-полнота понимается в смысле сводимости к ней за полиномиальное время остальных задач класса NP [5, 6].

Рассматриваются также вопросы построения алгоритмов поиска приближённой композиционной модели, приводятся имеющиеся методы решения данной задачи, кратко проводится их сравнение. В заключении подводятся итоги и перечисляются некоторые открытые вопросы.

1. NP-трудность построения оптимальной и точной композиционной модели

В настоящее время доказана NP-трудность задачи MGS — поиска минимальной генерирующей последовательности перестановок для заданной перестановки. NP-трудность MGS доказывается редукцией к ней проблемы ЗХС, входящей в список 21 проблемы Карпа [5, 6]. В приводимом далее доказательстве NP-трудности задачи построения композиционной модели используются некоторые идеи работы [7], где проводится доказательство NP-трудности задачи MGS.

Дадим формулировки задач, которые фигурируют в доказательстве:

- ЗХС(S, U) (точное покрытие 3-множествами). Дано множество $S = \{u_1, \dots, u_{3n}\}$ и подмножество $U \subset S^3$, $|U| = m$. Определить, существует ли подмножество $S' \subset U$, $|S'| = n$, такое, что для любого $u \in S$ есть ровно одно 3-множество $s' \in S'$, содержащее u .

Примечание: элементы из U — неупорядоченные множества из трёх различных элементов S , называемые 3-множествами.

- СОМР-L(F, f, μ, ε) (распознавание композиции липшиц-ограниченных функций, приближающей целевую функцию с заданной погрешностью). Дан набор функций $F = \{g_1, \dots, g_m\}$ и целевая функция f , причём функции f, g_i являются сюръективными липшиц-ограниченными отображениями $[0, 1] \rightarrow [0, 1]$. Определить, существует ли набор (i_1, \dots, i_n) , такой, что $\mu(g_{i_1} \dots g_{i_n}, f) \leq \varepsilon$.

В качестве метрики μ может рассматриваться любая метрика, для которой вычисление (или приближение с устанавливаемой погрешностью) значения между определяемыми в доказательстве функциями может быть произведено за полиномиальное время.

Теорема 1. ЗХС(S, U) редуцируется к СОМР-L(F, f, μ, ε).

Доказательство. Каждый элемент $s \in U$ можно задать как бинарный вектор b_s длины $3n$, где $b_s(i) = 1$, если $u_i \in s$, а все прочие позиции в b_s заняты нулями. Очевидно, получение таких векторов может быть выполнено за время, полиномиальное от n и m .

Каждому бинарному вектору b_s поставим в соответствие кусочно-определённую липшиц-ограниченную функцию из $C^1[0, 1]$. В качестве базовой функции выберем $t(x) = \frac{9}{2} \text{sign}(x) ((1 - |x|)^3 - (1 - |x|)^2)$, $x \in [-1, 1]$, введём также функцию окна

$$w(i, N, x) = \begin{cases} 1, & x \in \left[\frac{i-1}{N}, \frac{i}{N} \right], \\ 0, & x \notin \left[\frac{i-1}{N}, \frac{i}{N} \right] \end{cases}$$

и функции

$$g_s(x) = x + \frac{1}{3n} \sum_{i=1}^{3n} b_s(i) w(i, 3n, x) t(6nx - 2i + 1).$$

Систему функций F для задачи COMP-L определим так: $F = \{g_s : s \in U\}$, $|F| = m$; а целевую функцию f как

$$f(x) = x + \frac{1}{3n} \sum_{i=1}^{3n} w(i, 3n, x) t(6nx - 2i + 1).$$

Построение и вычисление функций из F и функции f выполняется за время, линейное от n и m . Константа Липшица всех функций из F и функции f равна 4.

Построенные функции g_{s1} , g_{s2} при выполнении композиции ведут себя следующим образом:

- 1) Если s_1 и s_2 не содержат одинаковых элементов u_i , то $g_{s1}(g_{s2}) = g_{s2}(g_{s1})$ соответствуют функции g , построенной для бинарного вектора $b = b_{s1} \vee b_{s2}$, так как

$$\begin{aligned} g_{s1}(g_{s2}(x)) &= x + \frac{1}{3n} \sum_{i=1}^{3n} b_{s1}(i) w(i, 3n, x) t(6nx - 2i + 1) + \\ &+ \frac{1}{3n} \sum_{i=1}^{3n} b_{s2}(i) w(i, 3n, x) t(6nx - 2i + 1) = \\ &= x + \frac{1}{3n} \sum_{i=1}^{3n} (b_{s1}(i) \vee b_{s2}(i)) w(i, 3n, x) t(6nx - 2i + 1). \end{aligned}$$

- 2) Если s_1 и s_2 содержат одинаковые элементы, то функция в соответствующих частях изменяется следующим образом: если $b_{s1}(i) = b_{s2}(i) = 1$ для некоторого i , то для $x \in \left[\frac{i-1}{3n}, \frac{i}{3n} \right]$ выполняется равенство

$$g_{s1}(g_{s2}(x)) = x + \frac{1}{3n} t(6nx - 2i + 1) + \frac{1}{3n} t \left(6n \left(x + \frac{1}{3n} t(6nx - 2i + 1) \right) - 2i + 1 \right),$$

что отличается от значений g_{s1} и g_{s2} . При дальнейших композициях $g_{s1}(g_{s2}(x))$ с функциями $g_s(x)$, такими, что $b_s(i) = 1$, на отрезке $\left[\frac{i-1}{3n}, \frac{i}{3n} \right]$ будут образовываться многократные подстановки $t(x)$ в саму себя, не совпадающие с исходной функцией.

Подобные свойства позволяют определить отсутствие, однократное или многократное покрытие элемента u_i в результате объединения 3-множеств $s \in S$, представляемых функциями из F . Приведём возможные результаты композиции функций $g_{s1}(x)$ и $g_{s2}(x)$ в таблице:

$b_{s1}(i)$ — бит покрытия элемента 3-множеством s_1	$b_{s2}(i)$ — бит покрытия элемента 3-множеством s_2	Функция результата композиции на отрезке $\left[\frac{i-1}{3n}, \frac{i}{3n} \right]$
0	0	x
1	0	$x + \frac{t(6nx - 2i + 1)}{3n}$
0	1	$x + \frac{t(6nx - 2i + 1)}{3n}$
1	1	$x + \frac{1}{3n} t(6nx - 2i + 1) + \frac{1}{3n} t \left(6n \left(x + \frac{1}{3n} t(6nx - 2i + 1) \right) - 2i + 1 \right)$

Свойства построенной системы функций показывает рис. 1, на котором сверху вниз изображены функция g_{s_1} для $s_1 = \{u_1, u_2, u_4\}$; функция g_{s_2} для $s_2 = \{u_2, u_3, u_5\}$; композиция функций $g_{s_1}(g_{s_2})$ при $m = 2$.

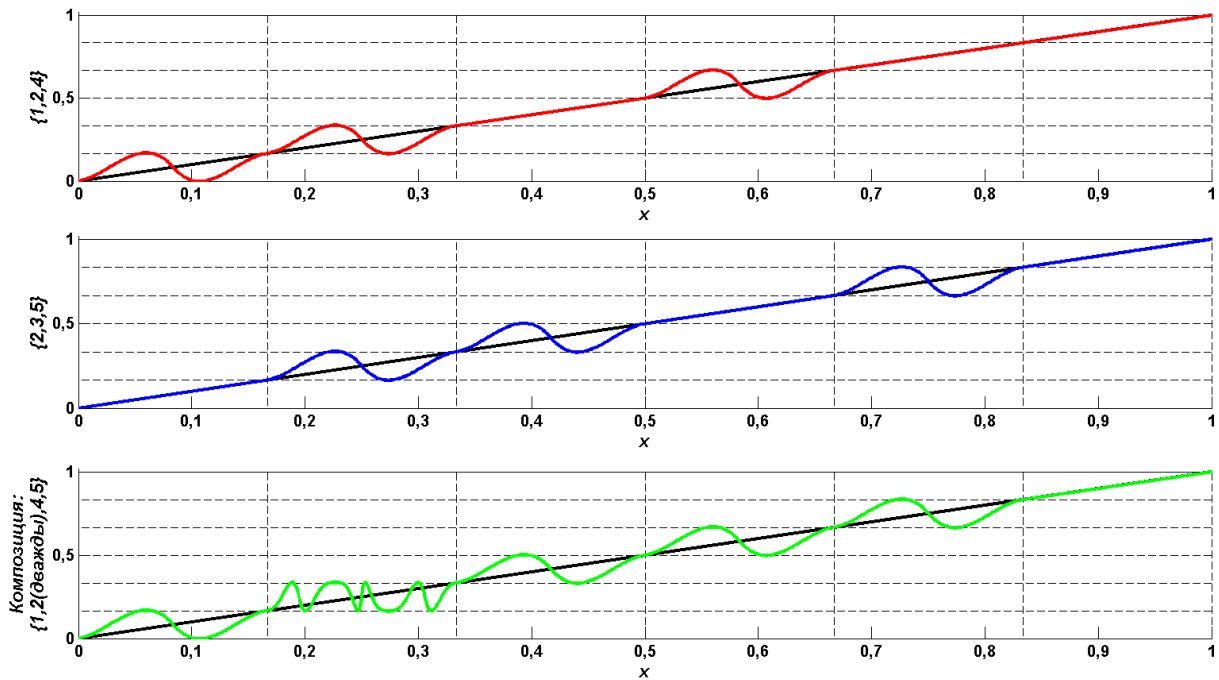


Рис. 1. Функции, представляющие 3-множества, и их композиция

Пусть определено существование решения задачи $\text{COMP-L}(F, f, \mu, \varepsilon)$ при $\varepsilon = 0$, тогда если оно существует, то покрытие для задачи ЗХС также существует, и найденный набор индексов (i_1, \dots, i_m) соответствует S' — решению задачи ЗХС. Отсутствие решения задачи COMP-L ведёт к отсутствию решения задачи ЗХС. Для получения покрытия из набора индексов используется обратное преобразование функций с данными индексами в бинарные векторы, а их — в 3-множества. Сложность данного преобразования, очевидно, является полиномиальной от n и m . ■

Замечание 1. Может показаться, что задача может быть NP-полной в силу необходимости вычисления значений метрики между липшиц-ограниченными функциями, но в данном случае метрика Чебышева для приведенных наборов функций может быть вычислена за полиномиальное от m и n время на основе таблицы расстояний между подстановками функции $t(x)$. Аналогично может рассматриваться метрика Чебышева по любой дискретной системе узлов, позволяющей различать композиции функции $t(x)$.

Замечание 2. Задача построения точной композиционной модели является частным случаем задачи $\text{COMP-L}(F, f, \mu, \varepsilon)$ при $\varepsilon = 0$. Значит, в доказательстве можно совместно рассматривать задачу распознавания композиционной модели, приближающей целевую функцию с заданной погрешностью, и задачу построения точной композиционной модели.

Замечание 3. Задача построения оптимальной композиционной модели NP-трудна, так как задача распознавания композиционной модели $\text{COMP-L}(F, f, \mu, \varepsilon)$ NP-полна. Предположим, решена задача построения оптимальной композиционной модели и известен $\min(\mu(g_{i_1}(\dots g_{i_n}), f))$; тогда задача распознавания решается путём

сравнения ε с найденным минимумом. Следовательно, задача построения оптимальной композиционной модели не может решаться проще задачи распознавания композиционной модели.

Таким образом, задача построения точной композиционной модели липшиц-ограниченной функции является NP-полной, а задача построения оптимальной композиционной модели — NP-трудной. Названные задачи скорее всего не имеют эффективных алгоритмов решения (проблема $P \neq NP$). Однако данный факт не влияет на возможность выполнять поиск приближённой композиционной модели.

2. Методы поиска приближённой композиционной модели

Рассмотрим различные методы поиска приближённых композиционных моделей и их свойства. Методы поиска приближённых композиционных моделей по основному принципу работы можно разделить на использующие:

- аппроксимации специальными видами функций;
- параметрическую оптимизацию;
- теорию поиска в метрических пространствах;
- случайный или генетический поиск.

При аппроксимации специальными функциями необходимо подобрать класс отображений C так, что:

- 1) возможно аппроксимировать f и $g_i \in F$ функциями данного класса;
- 2) в выбранном классе существуют эффективные алгоритмы построения композиционных моделей.

Если такой класс C отображений находится, то функция f заменяется на $\hat{f} \in C$ (входящую в выбранный класс), а функции $g_i \in F$ заменяются на $\hat{g}_i \in \hat{F} \subset C$. Далее решается задача построения точной/оптимальной композиционной модели в классе C . Пусть решение в классе C найдено, погрешность равна $\varepsilon = \mu(\hat{f}, \hat{g}_{i_1}(\dots \hat{g}_{i_n}))$. Рассмотрим погрешность решения (i_1, \dots, i_n) для исходной задачи: $\varepsilon \leq \mu(f, \hat{f}) + \mu(\hat{f}, \hat{g}_{i_1}(\dots \hat{g}_{i_n})) + \mu(\hat{g}_{i_1}(\dots \hat{g}_{i_n}), g_{i_1}(\dots g_{i_n})) = \mu(f, \hat{f}) + \hat{\varepsilon} + \mu(\hat{g}_{i_1}(\dots \hat{g}_{i_n}), g_{i_1}(\dots g_{i_n}))$. Таким образом, погрешность ограничена сверху функцией от расстояний между исходными функциями и их функциями-представителями из класса C , а также погрешностью решения задачи в классе C . Более точный анализ погрешностей можно провести для конкретной метрики, например метрики Чебышева:

$$\varepsilon = \mu(f, g_{i_1}(\dots g_{i_n})) \leq \mu(f, \hat{f}) + \hat{\varepsilon} + \sum_{j=1}^n \left(\prod_{k=1}^{j-1} \left[\min(L_{g_{i_k}}, L_{\hat{g}_{i_k}}) \right] \mu(g_{i_j}, \hat{g}_{i_j}) \right),$$

при этом, если обозначить $\varepsilon_c = \max_j \left(\max(\mu(\hat{g}_j, g_j)), \mu(\hat{f}, f) \right)$, а $L = \max_i (L_{g_i})$, то получаем оценку $\hat{\varepsilon} \leq \varepsilon_c \left(\frac{L^n - 1}{L - 1} + 1 \right) + \varepsilon$.

Так как $L > 1$ (по условию на отображения $f, g_i \in F$), то верхняя оценка погрешности возрастает очень быстро с ростом длины композиционной модели n , при этом на практике L редко бывает меньше 3. Таким образом, аппроксимационный подход может применяться, например, для классов C полиномиальных [4] и рациональных [8] функций, для которых существуют полиномиальные алгоритмы построения композиционных моделей. При этом аппроксимационный подход будет эффективен, когда

функции f и F с малой погрешностью ε_c приближаются функциями из C , а длина n композиционной модели мала.

Методы поиска приближённой композиционной модели, основанные на параметрическом подходе, традиционно включают две стадии:

- параметрическую оптимизацию с использованием традиционных (классических) методов оптимизации;
- прямой перебор вариантов с целью определения подходящего.

Данные стадии могут быть сгруппированы различным образом. Например, в работе [3] автор предлагает, рассматривая бесконечное множество F (с семействами функций, зависящими от параметров), изначально, на основе приближённых значений параметров, выбрать перебором подходящую композиционную модель. При выполнении перебора поиск выполняется с точностью до определения семейства каждой входящей в композицию функции. После этого автор [3] предлагает применить параметрическую оптимизацию для определения конкретных функций в каждом семействе. Полагается, что $F = F_1 \cup F_2 \cup \dots \cup F_m$; на первом этапе перебором ищется

$$\min_{\{(i_1, \dots, i_n) : i_1, \dots, i_n \in \{1, \dots, m\}\}} \left(\mu \left(f, g_1 = \text{select}(f, F_{i_1}) \in F_{i_1} \left(\dots \right. \right. \right. \\ \left. \left. \left. \dots g_n = \text{select}(g_{n-1}^{-1}(\dots g_1^{-1}(f), F_{i_n})) \in F_{i_n} \right) \right) \right),$$

где $\text{select}(f, F_i)$ выбирает некоторым вычислительно простым алгоритмом параметры функции из F_i для того, чтобы она приближала f (например, на основе метода МНК). На втором этапе ищется минимум функции

$$H(\bar{p}_1, \dots, \bar{p}_n) = \mu(f, g_1 = F_{i_1}(\bar{p}_1) (\dots g_n = F_{i_n}(\bar{p}_n))) ,$$

где (i_1, \dots, i_n) — индексы, выбранные в результате поиска на первом этапе. Альтернативой является изначально сведение задачи к задаче параметрической оптимизации, а затем поиск лучшего целочисленного округления полученного решения, подобный подход рассмотрен в работе [9]. Недостатками этих подходов является сложность предсказания качества получаемого решения и быстрое возрастание числа параметров, по которым проводится оптимизация, с ростом длины композиционной модели. Последний недостаток не позволяет использовать многие методы оптимизации, поэтому в большинстве случаев производится поиск только локального экстремума.

Поскольку на множестве $\{f\} \cup \{g_{i_1}(\dots g_{i_n}) : i_1, \dots, i_n \in \{1, \dots, m\}\} = \{f\} \cup F^n$ в задаче построения композиционной модели определена метрика, то могут использоваться методы теории поиска в метрических пространствах, рассматриваемые в работах [10, 11]. Оптимизация поиска в метрических пространствах достигается за счёт применения неравенства треугольника, на основе предвычислений расстояний во множестве F^n , дальнейшая оптимизация с ограничением точности получаемого результата может быть достигнута с использованием методов либо раннего прекращения поиска, либо методом ослабления условий сравнений в ходе поиска. Проблема данного подхода заключается в необходимости хранения предвычислений размера $O(|F|^n)$ для оптимизации процесса поиска. Это объясняется тем, что оценка расстояния $\mu(f, g)$ за счёт неравенства треугольника может быть произведена, только если существует t , такое, что $\mu(f, t)$ и $\mu(t, g)$ могут быть оценены. Таким образом, отношение «оцениваемости» расстояния $U_{\text{est}} \subseteq [\{f\} \cup F^n]^2$ является транзитивным замыканием отношения

«определённости» $U_{\text{def}} \subseteq [\{f\} \cup F^n]^2$ расстояния на множестве $\{f\} \cup F^n$ при условии, что оценки строятся за счёт неравенства треугольника с использованием «определённых» значений расстояния. Следовательно, $U_{\text{est}} \subseteq [\{f\} \cup F^n]^2$, только если отношение «определённости» расстояния U_{def} включает дерево на вершинах $\{f\} \cup F^n$, что требует минимум $O(|F|^n)$ вычисленных расстояний. С учётом необходимости хранения предвычисленных значений расстояний невозможно организовать поиск для больших значений n и множеств F со значительным количеством функций. Также следует иметь в виду, что если f заранее не известна, а предвычисления проводятся исключительно над множеством функций F^n , что чаще всего и необходимо в практических задачах, то, пока не будет найдено $(i_1, \dots, i_n) : \mu(f, g_{i_1}(\dots g_{i_n})) \ll 1$, предвычисленные значения не могут быть использованы для эффективной оценки расстояний $\mu(f, g_{j_1}(\dots g_{j_n}))$, используемых в работе алгоритма. Таким образом, на первых этапах работы алгоритма эффективность предвычислений незначительна.

Методы, основанные на случайном или генетическом поиске, применяются к задачам, аналогичным задаче поиска приближённой композиционной модели, в работах [1, 2]. Поскольку методология самих этих процедур хорошо известна [1, 12], обратим внимание лишь на проблемы их применения к задаче поиска приближённой композиционной модели. Случайный поисковый алгоритм для повышения эффективности работы требует определения перечня параметров распределения вектора модификации текущего решения, которые должны адаптироваться, исходя из истории поиска. Пока объём параметров адаптации незначителен, их выбор, как и определение правил адаптации, составляет значительную проблему, а возрастание эффективности незначительно. Если параметров адаптации много, то адаптация происходит крайне медленно и возрастание эффективности также незначительно. В случае применения генетического алгоритма проблема его использования заключается в определении функции смешивания элементов, такой, что коэффициент корреляции средней метрики элемента, получающегося на выходе функции смешивания, со средней метрикой элементов, попадающих на вход функции смешивания, не равен нулю. Это условие следует из теоремы Прайса [13] об эффективности генетического алгоритма, где под метрикой элемента понимается значение метрики между данным вариантом композиционной модели (названным здесь элементом) и целевой функцией f . На текущий момент для задачи построения приближённой композиционной модели такая процедура смешивания не известна.

Заключение

Задача построения точной композиционной модели липшиц-ограниченной функции является NP-полной, а задача построения оптимальной композиционной модели — NP-трудной. Автору не известны непереборные алгоритмы решения данных задач, например алгоритмы с константами $c, d < 1$ и сложностью $O((c|F|)^{dn})$, хотя их существование вполне возможно.

Алгоритмы поиска приближённой композиционной модели могут иметь полиномиальную сложность (например, алгоритмы, основанные на аппроксимации специальными видами функций и параметрической оптимизации), но применимость этих алгоритмов ограничена. Более универсальные алгоритмы, основанные на теории поиска в метрических пространствах, генетическом и случайном поисках, не гарантируют объёма вычислений (или предвычислений) менее $O(|F|^n)$, поэтому актуальным является построение и математическое обоснование алгоритмов, позволяющих снизить объём вычислений. Отдельную область исследований представляют алгоритмы, использующие

функции распределения расстояний между элементами $\{f\} \cup F^n$ и применяющие эту информацию для оптимизации поиска. Для таких алгоритмов могут быть вычислены оценки на достигаемое повышение эффективности и вероятностные характеристики точности решения, исходя из способов оценки распределения расстояний.

ЛИТЕРАТУРА

1. Koza J. R. Genetic Programming: on the Programming of Computers by Means of Natural Selection. London: A Bradford Book, 1998. 815 p.
2. Luke S. Essentials of metaheuristics. <http://cs.gmu.edu/~sean/book/metaheuristics>. 2009.
3. Лабутин С. А., Пугин М. В. Анализ сигналов и зависимостей: учеб. пособие. Н. Новгород: Нижегород. гос. тех. ун-т, 2001. 158 с.
4. Seong J. K., Elber G., and Kim M. S. Polynomial decomposition and its applications. <http://cana.kaist.ac.kr/seong/decomposition.pdf>. 2003.
5. Karp R. M. Reducibility among combinatorial problems // GJ-474 report. 1971. P. 87–103. http://www.seas.upenn.edu/~bhusnur4/cit596_spring2014/karp-1971.pdf
6. Панадимитриу Х. Х., Стайглиц К. Комбинаторная оптимизация: Алгоритмы и сложность. М.: Мир, 1987. 520 с.
7. Even S. and Goldreich O. The minimum-length generator sequence problem is NP-hard // J. Algorithms. 1981. No. 2. P. 311–313.
8. Alonso C., Gutierrez J., and Recio T. A rational function decomposition algorithm by near-separated polynomials // J. Symbolic Comput. 1995. V. 19. P. 527–544.
9. Калинин И. С. Алгоритм построения декомпозиции непрерывной функции одного аргумента по заданному множеству функций // Инновации в науке, образовании и бизнесе: X Междунар. научн. конф. Калининград: КГТУ, 2012. Ч. 2. С. 160–163.
10. Chavez E., Navarro G., Baeza-Yates R., and Marroquin J. L. Search in metric spaces // ACM Computing Surveys. 2001. V. 33. No. 3. P. 273–321.
11. Zezula P., Amato G., Dohnal V., and Batko M. Similarity Search: The Metric Space Approach. N. Y.: Springer Verlag, 2006. 220 p.
12. Растрюгин Л. А., Рина К. К., Тарасенко Г. С. Адаптация случайного поиска. Рига: Зинатне, 1978. 243 с.
13. Alenber L. The schema theorem and Price's theorem // Foundations of Genetic Algorithms 3. San Francisco: Morgan Kaufmann, 1995. P. 23–49.