ОБ ЭФФЕКТИВНОСТИ ЭВОЛЮЦИОННЫХ АЛГОРИТМОВ БЕЗ ЭЛИТЫ В СЛУЧАЕ РАЗРЕЖЕННОСТИ ОБЛАСТЕЙ УРОВНЯ, НЕСОГЛАСОВАННЫХ ПО ЦЕЛЕВОЙ ФУНКЦИИ[1]

А. В. Еремеев

Во многих известных эволюционных алгоритмах для задач оптимизации используются элитные особи, которые гарантированно сохраняются в популяции эволюционного алгоритма в силу своего преимущества по целевой функции по сравнению с другими имеющимися особями. Не смотря на то, что в живой природе элитных особей не существует, в эволюционных алгоритмах элита обеспечивает постоянное присутствие рекордных решений в популяции и позволяет интенсивно исследовать пространство поиска вблизи таких решений. Тем не менее, известны семейства задач, на которых наличие элитных особей затрудняет исследование новых областей пространства решений, препятствует выходу из локальных оптимумов и увеличивает математическое ожидание времени получения глобального оптимума. Эволюционные алгоритмы без элиты, в частности, при использовании турнирной и линейной ранжирующей селекции, оказываются эффективны на этих задачах, однако требуют подходящей настройки параметров селекции и мутации.

Один из стандартных подходов к анализу эффективности эволюционных алгоритмов основывается на разбиении пространства решений на подмножества (области уровня), пронумерованные в предполагаемом порядке посещения их популяцией ЭА. В настоящей работе рассматривается класс $\mathrm{SparseLocalOpt}_{\alpha,\varepsilon}$ задач псевдобулевой оптимизации, в которых объединение семейства областей уровня, в некотором смысле несогласованного по целевой функции, является ε-разреженным множеством, а множества решений, где целевая функция больше чем в несогласованных областях уровня, имеют плотность не менее α. Основным результатом является новая полиномиальная верхняя оценка математического ожидания времени первого достижения глобального оптимума эволюционными алгоритмами без элиты, справедливая для задач из $\mathrm{SparseLocalOpt}_{\alpha,\varepsilon}$, где элитные ЭА не эффективны, т.е. требуют экспоненциального в среднем времени для получения оптимума. Кроме того, показана эффективность эволюционных алгоритмов без элиты на более широком классе задач. Найдены значения настраиваемых параметров для эволюционных алгоритмов с турнирной и линейной ранжирующей селекцией, при которых гарантируется полиномиальная ограниченность времени оптимизации для некоторых α и ε. Приводится пример использования полученных результатов для семейства задач вершинного покрытия на графах «звезда», а также демонстрируется преимущество эволюционных алгоритмов без элиты по сравнению с простейшим алгоритмом, использующим одну элитную особь.

Ключевые слова: Эволюционный алгоритм, локальный оптимум, время достижения оптимума, плотность, разреженность.

Название статьи на английском языке: On efficiency of non-elitist evolutionary algorithms in case of sparsity of the level sets inconsistent in objective function

Keywords: Evolutionary algorithm, local optimum, optimization time, density, sparsity.

MSC: 68W20, 68W40, 68Q87, 68T20 .

Введение

Отличительной особенностью эволюционных алгоритмов (ЭА) является имитация процесса эволюционной адаптации биологической популяции к условиям окружающей среды. При этом *особи* соответствуют пробным точкам в пространстве решений задачи оптимизации (в данной статье это $\{0,1\}^n$), а *приспособленность особей определяется значениями целевой функции.* Построение новых пробных точек в ЭА осуществляется посредством операторов *мутации* (операторы *кроссинговера*, используемые в генетических алгоритмах [20], здесь не рассматриваются). В данной работе исследуется эффективность ЭА при решении безусловных задач псевдобулевой максимизации $\max\{f(x) \mid x \in \{0,1\}^n\}$, или минимизации $\min\{f(x) \mid x \in \{0,1\}^n\}$, где целевая функция $f : \{0,1\}^n \to \mathbb{R}$. Одним из наиболее часто используемых операторов мутации в случае таких задач является *стандартная мутация* [20], при которой каждый бит имеющейся

---

[1] Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации №075-15-2022-282.

Печатная версия статьи доступна по адресу: https://doi.org/10.21538/0134-4889-2024-30-4-84-105

строки $x \in \{0,1\}^n$ независимо от других меняет свое значение с заданной вероятностью p_{m}. По умолчанию далее будет подразумеваться именно этот оператор мутации.

Одной из основных характеристик работоспособности эволюционных алгоритмов при решении задач оптимизации является *среднее время оптимизации*, т. е. математическое ожидание времени первого достижения оптимума, исчисляемое в количестве вычислений целевой функции [1; 17]. Во многих известных ЭА используются, так называемые, элитные особи, т. е. особи, которые гарантированно сохраняются в популяции в силу своего преимущества по целевой функции по сравнению с другими имеющимися особями [16].

С одной стороны, как показано в [2; 6; 11], на известных семействах одноэкстремальных задач эволюционные алгоритмы без элитных особей в асимптотике не уступают ЭА с элитой. Даже при наличии шума [10; 24] и динамически изменяющейся целевой функции [9; 23] алгоритмы без элиты достаточно успешно справляются с этими задачами. В [3] построено семейство примеров FUNNEL и показано, что элитные алгоритмы $(\mu+\lambda)$ EA на этом семействе имеют экспоненциальное математическое ожидание времени первого достижения оптимума, тогда как ЭА без элиты с использованием турнирной селекцией и подходящей интенсивности мутации затрачивают в среднем полиномиально ограниченное время.

С другой стороны, при решении практических задач [26], а также в теории эволюционных вычислений [12; 27], зачастую рекомендуется использовать элитные особи, т.к. они обеспечивают постоянное присутствие рекордных решений в популяции и позволяют интенсивно исследовать пространство поиска вблизи таких решений. Кроме того, ЭА с элитными особями менее чувствительны к настройке параметров, определяющих интенсивность мутации и селекции [3]. Более того, детальные исследования в [12] показывают, что при использовании (μ,λ)-селекции на известном семестве задач псевдобулевой оптимизации JUMP эволюционные алгоритмы без элиты не могут иметь каких-либо существенных преимуществ перед ЭА с элитой.

Настоящая работа, как и предшествующая ей публикация [4], продолжает исследования свойств ЭА без элитных особей и показывают, что пессимистические выводы из [12] не распространяются на все ЭА без элитных особей и все многоэкстремальные задачи. В частности, в [4] были впервые сформулированы достаточные условия для задач псевдобулевой оптимизации, при выполнении которых ЭА без элиты имеют преимущества перед элитными ЭА. С этой целью был введен класс задач псевдобулевой оптимизации SPARSELOCALOPT$_{\alpha,\varepsilon}$, в которых при подходящем разбиении пространства решений на области уровня, объединение семейства областей уровня, в некотором смысле *не согласованного* по целевой функции, является ε-разреженным. При этом, параметр $\alpha \in [0,1]$ характеризует плотность множества решений, где целевая функция больше, чем в несогласованных областях уровня (подробне см. определения 2–4 ниже).

В [4] получены полиномиальные относительно размерности n пространства решений оценки времени оптимизации в терминах α и ε, показывающие эффективность алгоритмов без элиты на классе задач SPARSELOCALOPT$_{\alpha,\varepsilon}$ при подходящих параметрах мутации и селекции (турнирная селекция с размером турнира $k = 3$ и линейная ранжирующая селекция). С другой стороны, доказаны экспоненциальные в худшем случае нижние оценки на среднее время оптимизации для элитных ЭА, а именно, показано что класс задач SPARSELOCALOPT$_{\alpha,\varepsilon}$ имеет экспоненциальную *элитную стожность оптимизации черного ящика* (elitist black box complexity) в смысле [16] при любых константах $\alpha > 0$ и $\varepsilon > 0$. Этот отрицательный результат означает, в частности, что многие известные элитные эволюционные алгоритмы, такие как $(\mu+\lambda)$ EA со стандартной мутацией [16], недавно предложенные генетические алгоритмы со степенной мутацией [15], генетические алгоритмы со стационарной схемой [26] оказываются неэффективными на классах задач $\mathrm{S\small PARSE}\mathrm{L\small OCAL}\mathrm{O\small PT}_{\alpha,\varepsilon}$ даже при малых (константных) значениях ε.

Основным результатом настоящей работы является новая полиномиальная верхняя оценка математического ожидания времени первого достижения глобального оптимума эволюционными алгоритмами без элиты, справедливая для задач из $\mathrm{SparseLocalOpt}_{\alpha,\varepsilon}$, где элитные ЭА не эффективны. В этой оценке, по сравнению с оценкой из [4], на порядок снижена зависимость от параметра, характеризующего интенсивность селекции в ЭА. Кроме того, показана эффективность эволюционных алгоритмов без элиты на более широком классе задач. Для эволюционных алгоритмов с турнирной и линейной ранжирующей селекцией приводятся доказательства полиномиальной ограниченности среднего времени оптимизации для некоторых α и ε при подходящих значениях настраиваемых параметров. Обсуждается взаимосвязь локальных оптимумов в задачах псевдобулевой оптимизации с областями, где имеется несогласованность областей уровня по целевой функции. Устранены пропуски доказательств и некоторые неточности в формулировках, имеющиеся в работе [4], представленной в трудах конференции GECCO'21.

1. Обозначения и определения

Пусть $\mathbb{N}$ – множество натуральных чисел. Для любого $n \in \mathbb{N}$ полагаем $[n] := \{1, \ldots, n\}$. Скобка Айверсона имеет вид $[\cdot]$. Расстояние Хэмминга обозначается $\mathrm{H}(\cdot, \cdot)$. *Сфера Хэмминга* радиуса $r \in [n]$ с центром в точке $x \in \{0,1\}^n$ определяется как $S_r(x) := \{y \in \{0,1\}^n \mid \mathrm{H}(x, y) = r\}$. Очевидно, $|S_r(x)| = C_n^r$. Для описания асимптотического поведения величин, связанного с неограниченным ростом размерности n, будут использоваться стандартные обозначения $O(\cdot)$ и $\Omega(\cdot)$. Кроме того, будем говорить, что функция от n является полиномиально ограниченной, если существует полином от n, ограничивающий ее сверху. При вводе новых обозначений в формулах используется символ $:=$, и двоеточие находится со стороны определяемой величины.

Решение $x \in \{0,1\}^n$ является локальным оптимумом в окрестности, порожденной метрикой Хэмминга с радиусом d, если $f(x) \geq \max\{f(y) : y \in \{0,1\}^n, \mathrm{H}(x, y) \leq d\}$ в случае задач максимизации или $f(x) \leq \min\{f(y) \mid y \in \{0,1\}^n, \mathrm{H}(x, y) \leq d\}$ в случае задач минимизации.

В данной работе рассматриваются массовые задачи псевдобулевой оптимизации: такая задача представляет собой некоторое бесконечное множество целевых функций, заданных на семействе пространств поиска $\{0,1\}^n$ размерности n посредством оракула, сообщающего значение целевой функции $f(x)$ для любого выбранного решения $x \in \{0,1\}^n$. Для краткости далее слово «массовая» будет опускаться. Цель исследования состоит в оценке асимптотики среднего времени оптимизации ЭА для таких задач оптимизации, когда n неограниченно возрастает.

Для оценки эффективности ЭА здесь используется следующая *модель оптимизации «черного ящика»* (black-box optimization scenario), предложенная в [17]: предполагается, что ЭА применяется в ситуации, когда априори известно только то, к какой задаче псевдобулевой оптимизации относится целевая функция f т. е. известно, что f принадлежит некоторому известному множеству функций $\mathcal{F}$ вида $f : \{0,1\}^n \to \mathbb{R}$ (например, всех одноэкстремальных псевдобулевых функций или всех псевдобулевых функций или всех функций, равных константе всюду, кроме единственной точки глобального оптимума) и известна размерность n пространства решений. Далее ЭА получает информацию о решаемой задаче исключительно посредством обращений к оракулу, вычисляющему значения $f(x)$ для пробных точек x, выбираемых в данном алгоритме в процессе его работы. Время работы алгоритма в рамках данной модели полагается равным количеству обращений к оракулу вычисления $f(x)$ до первого попадания в точку, на которой будет получено оптимальное значение $f(x) = f^*$. Такая модель оптимизации «черного ящика» соответствует практическим приложениям эволюционных алгоритмов, алгоритмов локального поиска, имитации отжига, поиска с запретами и других метаэвристик, когда наиболее трудоемким является вычисление целевой функции, а не выбор очередной пробной точки.

В настоящей работе, как и во многих других исследованиях ЭА (см., например, [4; 14; 19; 22; 25]), будет использоваться *метод уровней*, который основан на разбиении всего пространства поиска на подмножества (области уровня), пронумерованные в предполагаемом порядке их посещения популяцией ЭА. Как правило, области уровня упорядочиваются с учетом значений целевой функции содержащихся в них решений. Например, при *каноническом* разбиении на уровни [22], каждый уровень содержит только решения x с равными значениями $f(x)$ и все уровни пронумерованы в порядке увеличения «качества» решений. Если пространство поиска $\{0,1\}^n$ разбито на m областей уровня $A_1, \ldots, A_m$, тогда для любого $j \in [m]$ обозначим $H_j := \cup_{i=j}^{m} A_i$. Здесь и далее разбиение $A_1, \ldots, A_m$ зависит от размерности n и целевой функции $f \in \mathcal{F}$, однако, для краткости записи в обозначениях областей уровня это не отражается. В работе ЭА информация о разбиении на области уровня никак не учитывается, они используются только для получения априорных оценок времени оптимизации, поэтому области уровня задаются с учетом конкретной целевой функции f.

Популяцией будем называть кортеж $P \in (\{0,1\}^n)^\lambda$, где λ – численность популяции, а i-тая особь из P обозначается как $P(i)$. Параметр λ зависит от размерности n. Для любого $A \subseteq \{0,1\}^n$ обозначим через $|P \cap A|$ число особей P, принадлежащих A, т. е. $|P \cap A| := |\{i : P(i) \in A\}|$.

1.1. Эволюционные алгоритмы без элитных особей

Эволюционные алгоритмы без оператора кроссинговера и без элитных особей могут быть представлены общей схемой алгоритма 1 (см., например, [11]). При построении очередной популяции P_{t+1} на основе имеющейся популяции P_t оператором селекции $\mathrm{Sel}(P_t)$ независимо в совокупности выбираются родительские решения, и копии выбранных родительских решений подвергаются действию рандомизированного оператора мутации $\mathrm{Mut}(x)$. Указанные операторы могут иметь некоторые настраиваемые параметры (например, размер турнира, вероятность мутации), которые подаются на вход алгоритма и не меняются в процессе его выполнения. Распределение вероятностей для номера особи в результате применения оператора $\mathrm{Sel}(P_t)$ обозначим через p_{sel}, а распределение вероятностей для решений, получаемых оператором мутации $\mathrm{Mut}(x)$, обозначим p_{mut}.

Алгоритм 1 эволюционный алгоритм без элитных особей [11]

```
Вход: начальная популяция $P_0$ , настраиваемые параметры для операторов селекции и мутации, а также, условие остановки.
1. Положить $t := 0$.
Итерация $t$:
2. Для каждого $i$ от 1 до $\lambda$ независимо выполнять шаги 2.1,2.2:
    2.1. Селекция: выбрать $I_t(i) := \mathrm{Sel}(P_t)$, положить $x := P_t(I_t(i))$.
    2.2. Мутация: построить $x' := \mathrm{Mut}(x)$, положить $P_{t+1}(i) := x'$.
3. Положить $t := t + 1$.
4. Если условие остановки не выполнено, то перейти на шаг 2, иначе – на шаг 5.
5. Результат: лучшая из найденных особей за время работы ЭА.
```

По умолчанию в алгоритме 1 используется оператор стандартной мутации, в котором вероятность мутации каждого конкретного бита определяется настраиваемым параметром $\chi \in (0, n]$, так что $p_{\mathrm{m}} = \chi/n$. Тогда для любых $x, x' \in \{0,1\}^n$, вероятность получения x' из x есть $(\chi/n)^{\mathrm{H}(x,x')}\,(1 - \chi/n)^{n-\mathrm{H}(x,x')}$. Параметр χ может зависеть от n или быть константой.

Алгоритм 1 представлет собой частный случай более общего алгоритма 2. Для анализа этой схемы применяются, так называемые, *теоремы уровней* (level-based theorems) [2; 14].

Алгоритм 2 популяционного поиска [2]

```
Вход: начальная популяция $P_0$, условие остановки и отображение $D$ из $(\{0,1\}^n)^\lambda$ в пространство распределений над $\{0,1\}^n$.
1. Положить $t := 0$.
Итерация $t$:
2. Для $i$ от 1 до $\lambda$ выполнять шаг 2.1:
    2.1. Независимо в совокупности выбрать $P_{t+1}(i)$ согласно распределению $D(P_t)$.
3. Положить $t := t + 1$.
4. Если условие остановки не выполнено, то перейти на шаг 2, иначе – на шаг 5.
5. Результат: лучшая из найденных особей за время работы ЭА.
```

При исследовании времени оптимизации в теории эволюционных вычислений принято считать, что условие остановки никогда не выполняется. Предположим это и в данной работе.

1.2. Операторы селекции

Для характеризации операторов селекции в данной работе будет использоваться следующий подход. Предположим, что особи популяции P упорядочены по убыванию значений целевой функции в случае задачи на максимум, т. е. $f(P(1)) \geq f(P(2)) \geq \dots \geq f(P(\lambda))$ или по возрастанию целевой функции в случае задачи на минимум. При таком упорядочении номер особи будем называть, также ее *рангом*. Тогда для любых $0 \leq \psi \leq \gamma \leq 1$ обозначим через $\beta(\psi, \gamma, P)$ вероятность выбора особи с рангом от $\lceil \psi\lambda \rceil$ до $\lceil \gamma\lambda \rceil$ в популяции P. В дальнейшем будем опускать P в обозначении $\beta(\cdot,\cdot)$, когда из контекста ясно о какой популяции идет речь. Заметим, что функция β, введенная в [22] и используемая в [2; 11], получается из определяемой здесь функции фиксацией первого аргумента $\psi = 0$.

В эволюционных алгоритмах широко используется оператор *турнирной селекции* [21]. При действии данного оператора, из популяции независимо извлекаются k особей с равномерным распределением, и оператор выдает номер особи с наибольшим рангом среди k выбранных особей. Здесь k является настраиваемым параметром и называется *размером турнира.*

Другой известный оператор селекции – *ранжирующая селекция* (ranking selection) [21]. Функция $\alpha : \mathbb{R} \to \mathbb{R}$ называется *ранжирующей*, если $\alpha(x) \geq 0$ для всех $x \in [0, 1]$, и при этом $\int_0^1 \alpha(x)dx = 1$. В операторе ранжирующей селекции с ранжирующей функцией α вероятность выбора особи среди $\gamma\lambda$ «лучших» особей (т. е. имеющих ранг не более $\gamma\lambda$) полагается равной $\int_0^\gamma \alpha(x)dx$. В случае *линейной ранжирующей селекции* ее интенсивность настраивается параметром $\eta \in (1, 2]$, и ранжирующая функция определяется как $\alpha(x) := \eta(1 - 2x) + 2x$.

В [6] предложен новый оператор *степенной селекции* (power-law selection), который может быть описан в терминах $\beta(0, \gamma)$. В данном случае вероятность выбора особи среди i «лучших» особей полагается равной $\beta(0, i/\lambda) = (i/\lambda)^c$, где параметр $c \in (0, 1)$. Программная реализация ЭА со степенной селекцией и их экспериментальные исследования обсуждаются в [8].

Три описанных выше оператора селекции удовлетворяют следующему естественному свойству *f-монотонности*: если x, x' – особи популяции P и $f(x) \geq f(x')$, то $p_{\text{sel}}(x) \geq p_{\text{sel}}(x')$ в случае задачи на максимум. Неравенство $f(x) \geq f(x')$ заменяется на $f(x) \leq f(x')$ в случае задачи на минимум.

Легко видеть, что для f-монотонного оператора селекции функция β удовлетворяет неравенству $\beta(\psi_1, \psi_1 + \gamma) \geq \beta(\psi_2, \psi_2 + \gamma)$ при любых $\psi_1 \leq \psi_2$.

1.3. Несогласованные области уровня и класс $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$

Для описания областей пространства решений, составляющих сложность для многих элитных ЭА и алгоритмов локального поиска, введем понятие *областей уровня, несогласованных по целевой функцией.*

Определение 1. *При заданной функции $f : \{0, 1\}^n \to \mathbb{R}$ и разбиения множества $\{0, 1\}^n$ на области уровня $A_1, \ldots, A_m$ будем называть область уровня A_i* несогласованной *по целевой функции f относительно области уровня A_j, если $1 \leq i < j \leq m$ и существуют решения $x \in A_i$, $y \in A_j$, такие что $f(x) \geq f(y)$.*

В случае задач минимизации, в определении 1 неравенство $f(x) \geq f(y)$ заменяется на $f(x) \leq f(y)$. Для каждого несогласованного уровня A_i имеется одна или несколько *рассогласованных пар* уровней (f-deceptive pairs [4]) (A_i, A_j), $i < j$, удовлетворяющих определению 1.

Очевидно, что несогласованные уровни отсутствуют в случае канонического разбиения $A_1, \ldots, A_m$, где по определению $A_j = \{x : f(x) = f_j\}$ для всех $j \in [m]$, $f_1 < \cdots < f_m$ в случае задачи на максимум (или $f_1 > \cdots > f_m$ в случае задачи на минимум) и это все значения, принимаемые функцией f на $\{0, 1\}^n$. Однако, если некоторая точка $x' \in A_j$ является локальным оптимумом по окрестности Хэмминга радиуса d, то расстояние до ближайшего «улучшающего решения» (т. е. точки множества H_{j+1}) будет не менее $d + 1$. Вероятность перехода из A_j в H_{j+1} посредством стандартной мутации зависит от расположения мутируемого решения и от свойств множества H_{j+1}. Например, в случае применения оператора Mut к точке локального оптимума $x = x'$, если при этом $|H_{j+1}| = 1$, то для перехода из точки x в H_{j+1} потребуется в среднем $O(n^{d+1})$ попыток. В связи с этим, если задача может иметь локальные (не глобальные) оптимумы по окрестности Хэмминга радиуса d, то при использовании канонического разбиения на уровни получение верхней оценки на среднее время оптимизации, лучшей чем $O(n^{d+1})$, не представляется возможным.

С другой стороны, разбиение на $m = n+1$ областей уровня может быть основано на мере близости решений к некоторому глобальному оптимуму x^* функции f, а именно, $A_j = \{x : \mathrm{H}(x, x^*) = j - 1\}$ для всех $j \in [m]$. При использовании стандартной мутации, где $\chi - const$, переход посредством мутации из любой точки множества A_j в H_{j+1} будет возможен в среднем за $O(n)$ попыток при любом $j \in [m-1]$. Однако, если имеется некоторый локальный оптимум $x' \in A_j$, $j < m - d$, в окрестности, порожденной метрикой Хэмминга с радиусом d, то уровень A_j неизбежно оказывается несогласованным относительно всех уровней $A_{j+1}, \ldots, A_{j+d}$. (Действительно, предположим что решается задача на максимум и рассмотрим последовательность точек $x', x^{(1)}, \ldots, x^{(d)}$, где каждая последующая отличается от предыдущей заменой одного «неверного» бита из локального максимимума x' на «верный» бит из x^*. Тогда в каждом из уровней A_{j+k}, $k = 1, \ldots, d$, имеется по крайней мере одно решение $x^{(k)}$ такое что $f(x^{(k)}) < f(x')$.) Переход посредством стандартной мутации из A_j в H_{j+d+1} в худшем случае потребует в среднем $\Omega(n^{d+1})$ попыток. При большом значении d такая несогласованность может создать сложности в получении верхних оценок меньшего порядка на время оптимизации.

Следующие два определения из [4] (см., также, замечания из [5]) описывают плотные и разреженные множества решений.

Определение 2. *Множество решений $C \subseteq \{0,1\}^n$ называется α-*плотным, *где $\alpha \in [0,1]$ если $\forall x \in C$ выполнено неравенство $|S_1(x) \cap C| \geq \alpha n$.*

Определение 3. *Пусть для каждой целевой функции f задачи псевдобулевой оптимизации $\mathcal{F}$ с пространством поиска $\{0,1\}^n$ задано множество решений $B \subseteq \{0,1\}^n$. Множество B называется ε-разреженным, если*

(SP1) $\forall x \in B, \forall r \in [n-1]$, $|S_r(x) \cap B| \leq \varepsilon \cdot C_n^r$ *при некотором* $\varepsilon \in [0,1]$, *и*
(SP2) $\forall x \in \{0,1\}^n \setminus B, \forall r \in [n-1]$, $|S_r(x) \cap B| = O\big(\frac{1}{n} C_n^r\big)$.

Здесь и далее, параметры α и ε, вообще говоря, могут быть функциями от n. Однако, как правило, достаточно рассматривать ситуации, где α и ε – константы.

В определении класса задач псевдобулевой оптимизации $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$, предложенном в [4] (см. определение 4 ниже),[2] допускается наличие несогласованных областей уровня, но при этом исключаются ситуации когда очередная область уровня оказывается слишком далека от некоторых решений из предыдущей области уровня или когда несогласованные области уровня недостаточно разрежены, или области уровня, следующие за рассогласованными парами, образуют плотное множество.

Определение 4. *Задача псевдобулевой оптимизации $\mathcal{F}$ принадлежит классу $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$, если существует константа d, такая что при любых $n \in \mathbb{N}$ и $f \in \mathcal{F}$ с пространством поиска $\{0,1\}^n$ найдется разбиение множества $\{0,1\}^n$ на*

---

[2] Указанная выше связь между локальными оптимумами и несогласованными уровнями послужила мотивацией для выбора названия класса $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$.

$m = m(n)$ уровней $A_1, \ldots, A_m$, где m ограничено сверху некоторым полиномом от n и выполнены условия:

I. A_m состоит из глобально оптимальных решений для f,

II. $\forall j \in [m-1], \forall x \in A_j, \exists y \in H_{j+1}$, такой что $\mathrm{H}(x, y) \le d$,

а для множества всех рассогласованных пар $(A_{i_1}, A_{j_1}), \ldots, (A_{i_u}, A_{j_u})$ относительно f, где u – число всех таких пар, выполнены условия:

III. $\cup_{v=1}^{u} A_{i_v}$ – ε-разреженное множество,

IV. H_{j_v} – α-плотное множество для каждого $v \in [u]$.

В качестве параметров α и ε, как правило, рассматриваются константы, но, вообще говоря, они могут зависеть от n. В соответствии с определением 4, класс $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ состоит из таких задач псевдобулевой оптимизации, что при «хороших» разбиениях пространства решений на уровни (условия I,II), с одной стороны, объединение всех несогласованных уровней, является ε-разреженным множеством точек (условие III); с другой стороны, для каждой рассогласованной пары уровней (A_i, A_j) объединение уровней с номерами от j до m образует α-плотное множество (условие IV). При этом, «хорошими» считаются разбиения, состоящие из полиномиально ограниченного по n числа уровней, где последний уровень состоит из глобально оптимальных решений, и для любой точки x пространства решений найдется точка из уровня с бо́льшим номером, до которой расстояние Хэмминга не превышает константы.

Заметим, что само по себе условие ε-разреженности множества точек из несогласованных уровней, также как и условие α-плотности множеств H_{j_v} из определения 4, могут быть выполнены тривиально, даже со значениями $\varepsilon = 0$ и $\alpha = 1$ для любого множества функций $\mathcal{F}$. Достаточно выбрать, например, разбиение $\{0,1\}^n$ на 2 уровня, где A_2 – множество глобально оптимальных решений и $A_1 = \{0,1\}^n \backslash A_2$. В таком случае уровни не образуют ни одной рассогласованной пары и, согласно определению 4, условия разреженности и плотности проверять не требуется. Однако, например, в случае единственного глобального оптимума x^* у каждой функции $f \in \mathcal{F}$, такое разбиение с ростом размерности задачи n приведет к линейному по n увеличению расстояния Хэмминга от x^* до наиболее удаленного от x^* элемента в A_1, что противоречит условию II, т.е. такие разбиения не будут «хорошими».

С одной стороны, по мере увеличения параметра ε ослабляется требование к разреженности областей уровня, несогласованных по целевой функции. С другой стороны, снижение параметра α ослабляет требования к плотности областей H_{j_v}. Таким образом, имеет место вложенность классов $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ в том смысле, что при любых $\alpha' \le \alpha$ и $\varepsilon' \ge \varepsilon$ справедливо включение

$$\textsc{SparseLocalOpt}_{\alpha,\varepsilon} \subseteq \textsc{SparseLocalOpt}_{\alpha',\varepsilon'}.$$

В крайней ситуации, когда $\alpha = 0$ и $\varepsilon = 1$, класс $\mathrm{SparseLocalOpt}_{0,1}$ очень широк, т.к. состоит из всех задач псевдобулевой оптимизации $\mathcal{F}$, для которых существует семейство разбиений, такое что несогласованные области уровня удовлетворяют условию разреженности с $\varepsilon = 1$, т.е. достаточно выполнить условие (SP2). В противоположной ситуации, когда $\alpha = 1$ и $\varepsilon = 0$, класс $\mathrm{SparseLocalOpt}_{1,0}$ содержит только такие задачи псевдобулевой оптимизации, для которых пространство решений можно разбить на подмножества без несогласованных областей уровня, и при этом выполнить требование близости соседних областей уровня по метрике Хэмминга.

2. Верхние оценки времени оптимизации

В данном разделе будут сформулированы достаточные условия, при которых эволюционные алгоритмы без элитных особей, использующие стандартный оператор мутации, оказываются эффективны на классе $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ или его расширении $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$ и будут в явном виде указаны верхние оценки на время, которое требуется для достижения глобального оптимума в среднем (теорема 2).

2.1. Условия эффективности эволюционных алгоритмов без элиты

С целью получения верхних оценок на среднее время оптимизации для алгоритма 2 (а значит и для ЭА из алгоритма 1, как его частного случая) в настоящей работе используется следующая теорема уровней из [13].

Теорема 1. *Пусть для каждой $f \in \mathcal{F}$, $f : \{0,1\}^n \to \mathbb{R}$ задано разбиение множества $\{0,1\}^n$ на области уровня $A_1, \ldots, A_m$, пусть P_t – популяция алгоритма 2 на поколении t, $t \in \mathbb{N}$, $T := \min\{t\lambda \mid 0 < |P_t \cap A_m|\}$, и существуют параметры $z_1, \ldots, z_{m-1}, \delta \in (0,1]$, $\gamma_0 \in (0,1)$, такие что при любой популяции P справедливы условия 1)–3):*

1) *для любого $j \in [m-1]$, если $|P \cap H_j| \geq \gamma_0\lambda$, то $\Pr\limits_{y\sim D(P)} (y \in H_{j+1}) \geq z_j$,*

2) *для любого $j \in [m-2]$ и всех $\gamma \in (0, \gamma_0]$,*
если $|P \cap H_j| \geq \gamma_0\lambda$ и $|P \cap H_{j+1}| \geq \gamma\lambda$, то $\Pr_{y\sim D(P)} (y \in H_{j+1}) \geq (1+\delta)\gamma$,

3) *и размер популяции $\lambda \in \mathbb{N}$ удовлетворяет неравенству*

$$\lambda \geq \frac{8}{\gamma_0\delta^2} \ln\left(\frac{cm}{\delta}\left(\ln\lambda + \frac{1}{z^*\lambda}\right)\right),$$

где $z^* := \min_{j\in[m-1]}\{z_j\}$ и c – константа,

тогда при достаточно большой константе c имеем $E[T] \leq \frac{c}{\delta} \cdot \left(\lambda m \ln(\lambda) + \sum_{j=1}^{m-1} \frac{1}{z_j}\right)$.

Если $\delta \to 0$ при $n \to \infty$, то теорема 1 дает асимптотически более точную оценку на $E[T]$ по сравнению с теоремой уровней из [2], т.к. параметр δ здесь входит в знаменатель оценки линейно, тогда как в оценке из [2] зависимость от δ квадратичная. Как следствие теоремы 1, для ЭА с пропорциональной селекцией в [13; 14] удалось асимптотически улучшить верхнюю оценку на $E[T]$, известную из [11]. С другой стороны, из доказательств теорем 2.2 и 3.2 в [2] вытекает, что константа c имеет значение порядка 10^3, поэтому при малых значениях n преимущество имеет теорема уровней из [2].

С целью получения оценки на $E[T]$ для ЭА без элитных особей на классе $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ нам потребуется обобщение теоремы 1 с учетом несогласованных областей уровня. Далее объединение несогласованных областей (deceptive region [4]) будем обозначать через B, т.е. $B := \cup_{v=1}^{u} A_{i_v}$ в обозначениях из определения 4. В новом варианте теоремы уровней условия (G1) и (G2) представляют собой ослабленные версии условий 1) и 2) из теоремы 1, а именно, неравенства условий 1) и 2) теперь накладываются только когда в области B содержится достаточно мало особей текущей популяции. Новое условие (G0) по сути означает, что к следующему поколению число представителей области B в среднем сокращается не менее, чем на $\Delta \cdot 100\%$, если их численность в текущем поколении равна $\psi\lambda$ и $\psi \geq \psi_0$ для некоторого порогового значения $\psi_0 \in (0,1)$.

Теорема 2. *Пусть для каждой $f \in \mathcal{F}$, $f : \{0,1\}^n \to \mathbb{R}$, задано подмножество $B \subset \{0,1\}^n$ и разбиение множества $\{0,1\}^n$ на области уровня $A_1, \dots, A_m$, кроме того, пусть P_t – популяция алгоритма 2 на поколении t, $t \in \mathbb{N}$, $T := \min\{t\lambda \mid 0 < |P_t \cap A_m|\}$, существуют параметры $z_1, \dots, z_{m-1}, \delta \in (0,1]$, $\Delta \in (0,\delta]$, $\psi_0, \gamma_0 \in (0,1)$, такие что при любой популяции P, $y \sim D(P)$, любых $j \le m-1$, $\gamma \le \gamma_0$ и $\psi \ge \psi_0$ справедливы условия (G0)–(G3):*

(G0) *Если $|P \cap B| \le \psi\lambda$, то $\Pr(y \in B) \le (1-\Delta)\psi$,*
(G1) *Если $|P \cap B| \le \psi_0\lambda$ и $|P \cap H_j| \ge \gamma_0\lambda$, то $\Pr(y \in H_{j+1}) \ge z_j$,*
(G2) *Если $|P \cap B| \le \psi_0\lambda$, $|P \cap H_j| \ge \gamma_0\lambda$ и $|P \cap H_{j+1}| \ge \gamma\lambda$, то $\Pr(y \in H_{j+1}) \ge (1+\delta)\gamma$,*
(G3) $\lambda \ge \left(\frac{12}{\gamma_0\Delta^2\psi_0}\right) \ln\left(\frac{8c'm(\ln(\lambda)z_* + \lambda^{-1})}{z_*\Delta}\right)$, *где $z_* := \min_j z_j$ и c' – константа,*

тогда при достаточно большой константе c' имеем $E[T] \le \frac{c'}{\Delta} + \frac{c'}{\delta} \cdot \left(\lambda m \ln(\lambda) + \sum_{j=1}^{m-1} \frac{1}{z_j}\right)$.

Доказательство. Для краткости будем называть особи из множества B *несогласованными*. Разобьем процесс работы ЭА на *этапы* длиной $\tau_1 + 2\tau_2$ поколений каждый, где τ_1 и τ_2 будут указаны позже. Первые τ_1 поколений этапа будем называть *первой фазой* этапа, а оставшиеся $2\tau_2$ поколений – *второй фазой* этапа. Для любого $t_0 \in \mathbb{N}$ определим $Y_t := |P_{t_0+t} \cap B|$, т. е. Y_t – число несогласованных особей на поколении t_0+t. Будем говорить, что произошло событие «неудача» на некоторой фазе этапа, если на одной из итераций в этой фазе выполнилось неравенство $Y_{t+1} \ge \max\{\psi_0\lambda, (1-\Delta/2)Y_t\}$.

Доказательство проводится по следующей схеме. Сначала докажем, что если событие «неудача» (которое имеет достаточно малую вероятность) не происходит на протяжении первой фазы этапа, то число несогласованных особей к концу этой фазы оказывается не более чем $\psi_0\lambda$. Затем докажем, что если не произойдет «неудачи» и на второй фазе этапа, то число несогласованных особей остается не более чем $\psi_0\lambda$ на протяжении всей второй фазы. Это позволит применить теорему 1 и неравенство Маркова для оценки сверху вероятности отсутствия оптимума на одном этапе. Окончательный результат получим рассмотрением последовательности этапов, принимая во внимание вероятность «неудач».

Воспользуемся терминологией из следующего определения (см., например, [28]).

Определение 5. *Случайная величина X «стохастически больше» случайной величины X', если при любом $\phi \in \mathbb{R}$*

$$P\{X' > \phi\} \le P\{X > \phi\}. \tag{2.1}$$

Согласно (G0), биномиально распределенная случайная величина $Z \sim \mathrm{Bin}(\lambda, p_s)$ при $p_s := \max\{\psi_0, Y_t/\lambda\}(1-\Delta)$ будет «стохастически больше» чем Y_{t+1}. Поэтому вероятность неудачи на любом выбранном поколении из первой фазы оценивается сверху:

$$\begin{aligned}\Pr(Y_{t+1} \ge \max\{\psi_0\lambda, (1-\Delta/2)Y_t\}) \\ \le \Pr(Y_{t+1} \ge (1-\Delta/2)\max\{\psi_0\lambda, Y_t\}) \le \\ \le \Pr(Z \ge (1-\Delta/2)\max\{\psi_0\lambda, Y_t\}) = \\ = \Pr(Z \ge E[Z](1+\Delta/(2(1-\Delta)))).\end{aligned} \tag{2.2}$$

Следовательно, по неравенству Чернова (см., например, [18]) получаем:

$$\Pr\left(Y_{t+1} \ge \max\{\psi_0\lambda, (1-\Delta/2)Y_t\}\right) \le \exp\left(-\frac{\Delta^2 \max\{\psi_0\lambda, Y_t\}}{12(1-\Delta)}\right) \le e^{-\frac{\Delta^2\psi_0\lambda}{12(1-\Delta)}}.$$

Выберем наименьшее τ_1 такое, что выполняется неравенство $\lambda\,(1-\Delta/2)^{\tau_1} \le \psi_0\lambda$, т. е.

$$\tau_1 = \left\lceil \frac{\ln(\psi_0)}{\ln(1-\Delta/2)} \right\rceil < \frac{(1-\psi_0)(1-\Delta/2)}{\Delta/2} + 1 < \frac{2}{\Delta} + 1. \tag{2.3}$$

т.к. для всех $x > 0$ имеют место неравенства $\frac{x-1}{x} \le \ln(x) \le x-1$. Если «неудачи» не произойдет на первой фазе, то число несогласованных особей к ее концу будет не более $\lambda(1-\Delta/2)^{\tau_1} \le \psi_0\lambda$.

Для любого этапа i обозначим через E_i случайное событие, состоящее в отсутствии «неудач» на рассматриваемом этапе i, до первого попадания в множество A_m. Заметим, что при реализации события E_i, на любой итерации t во второй фазе этапа i (по крайней мере пока не было попадания в множество A_m) неравенства в условиях 1)-3) теоремы 1 выполнены с теми же параметрами $z_1, \dots, z_{m-1}, \delta \in (0,1]$, $\gamma_0 \in (0,1)$, что и в формулировке настоящей теоремы, если выбрать константу $c' \ge c$, где c – константа из теоремы 1. Действительно, с учетом выбора величины τ_1, при реализации события E_i имеем $|P_t \cap B| \le \psi_0\lambda$ на протяжении второй фазы этапа i. Тогда из предположений (G1) и (G2) следует, что условия 1) и 2) теоремы 1 применимы к условным вероятностям $\Pr_{y\sim D(P_t)}(y \in H_{j+1}|E_i)$ и $\Pr_{y\sim D(P_t)}(y \in H_{j+1}|E_i)$ в условиях 1) и 2) с указанными выше числовыми параметрами. Условие 3) не зависит от состава популяции P_t и следует из предположения (G3), ввиду того что $\Delta \le \delta, \psi_0 \le 1, c' \ge c$.

Величину τ_2 выберем как среднее число поколений до достижения множества A_m в соответствии с оценкой, предоставляемой теоремой 1:

$$\tau_2 := \left\lceil \frac{c'm\ln(\lambda)}{\delta} + \frac{c'}{\delta\lambda}\sum_{j=1}^{m-1}\frac{1}{z_j} \right\rceil < \frac{c'm}{\delta z_*}\cdot\left(\ln(\lambda)z_* + \lambda^{-1}\right) + 1, \tag{2.4}$$

Далее, ввиду неравенств (2.1),(2.3),(2.4) и условия (G3), считая что c' – достаточно большая константа, с вероятностью не более

$$(\tau_1 + 2\tau_2)e^{-\frac{(\Delta)^2\psi_0\lambda}{12(1-\Delta)}} \le \left(\frac{2}{\Delta} + \frac{2c'm(\ln(\lambda)z_* + \lambda^{-1})}{\delta z_*} + 2\right)\left(\frac{z_*\Delta}{8c'm(\ln(\lambda)z_* + \lambda^{-1})}\right) < \frac{1}{3}, \tag{2.5}$$

число несогласованных особей не превышает $\psi_0\lambda$ на протяжении второй фазы.

Применением теоремы 1, а также неравенства Маркова для номера поколения, когда впервые достигается уровень A_m, заключаем что с вероятностью не более $1/2$ популяция ЭА остается вне области A_m на протяжении $2\tau_2$ итераций второй фазы. Следовательно, с учетом неравенства (2.5), вероятность достижения области A_m на протяжении этапа будет не менее $1-1/2-1/3 = 1/6$. Эти аргументы могут быть применены к любому этапу работы ЭА, независимо от предыстории работы алгоритма, т.к. никаких предположений о начальной популяции выше сделано не было.

Таким образом, математическое ожидание для номера этапа, когда происходит первое достижение области A_m, не превышает 6. Требуемая оценка следует из того, что этап состоит из $(\tau_1 + 2\tau_2)$ поколений, а каждое поколение состоит в генерации λ пробных решений. $\square$

2.2. Верхняя оценка времени оптимизации на классе $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$

Следующие две леммы описывают свойства оператора стандартной мутации на плотных и разреженных подмножествах пространства решений.

Лемма 1. *Если множество решений C является α-плотным, то*

$$\Pr\left(\mathrm{Mut}(x) \in C\right) > (1 - \chi/n)^n\,(1 + \alpha\chi) \quad \text{при любом} \quad x \in C. \tag{2.6}$$

Доказательство. Для получения элемента из C в результате мутации решения x достаточно либо сохранить все биты из x без изменений, либо преобразовать x в одно из не менее αn решений во множестве C, лежащих на единичном расстоянии Хэмминга от x. Это происходит с вероятностью не менее $\left(1 - \frac{\chi}{n}\right)^n + \alpha n\left(\frac{\chi}{n}\right)\left(1 - \frac{\chi}{n}\right)^{n-1} > \left(1 - \frac{\chi}{n}\right)^n(1 + \alpha\chi)$. □

Лемма 2. *Пусть для каждой целевой функции $f : \{0,1\}^n \to \mathbb{R}$ задачи $\mathcal{F}$ в пространстве поиска $\{0,1\}^n$ задано подмножество $B = B_f$, являющееся ε-разреженным, тогда*

- $\Pr\left(\mathrm{Mut}(x) \in B\right) \le (1 - \varepsilon)\left(1 - \frac{\chi}{n}\right)^n + \varepsilon + O(n^{-n})$ *при любом* $x \in B$,
- $\Pr\left(\mathrm{Mut}(x) \in B\right) = O(1/n)$ *при любом* $x \notin B$.

Доказательство. Пусть $B_r(x) := \{y \in B \mid H(x,y) = r\}$. При $x \in B$ и $y \sim p_{\mathrm{mut}}(x)$

$$\Pr\left(y \in B\right) = \left(1 - \frac{\chi}{n}\right)^n + \sum_{r=1}^{n}\sum_{z \in B_r(x)} \Pr\left(y = z\right) \le$$

$$\le (1 - \varepsilon)\left(1 - \frac{\chi}{n}\right)^n + \varepsilon\sum_{r=0}^{n} C_n^r\left(\frac{\chi}{n}\right)^r\left(1 - \frac{\chi}{n}\right)^{n-r} =$$

$$= (1 - \varepsilon)\left(1 - \frac{\chi}{n}\right)^n + \varepsilon + \mathrm{O}\big(n^{-n}\big).$$

Аналогично, если $x \notin B$ и $y \sim p_{\mathrm{mut}}(x)$, то

$$\Pr\left(y \in B\right) = \mathrm{O}\left(\frac{1}{n}\sum_{r=0}^{n} C_n^r\left(\frac{\chi}{n}\right)^r\left(1 - \frac{\chi}{n}\right)^{n-r}\right) = \mathrm{O}\left(\frac{1}{n}\right). \;\square$$

Как было отмечено в разделе 1, параметры ε и α класса $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$, а также параметры ЭА λ и χ не обязательно являются константами и могут быть функциями от n. Аналогичное предположение сделано для параметра δ в разделе 2.1. Таким образом, вообще говоря, $\varepsilon = \varepsilon(n), \alpha = \alpha(n)$, $\chi = \chi(n), \lambda = \lambda(n), \delta = \delta(n)$.

В формулировке следующей теоремы допускается, хотя и «не слишком быстрое», приближение $\varepsilon(n)$ и $\chi(n)$ к нулю с ростом n.

Теорема 3. *Пусть в алгоритме 1 используется стандартный оператор мутации с вероятностью мутации $p_{\mathrm{m}} = \chi/n$ и f-монотонный оператор селекции, который описывается функцией $\beta(\psi,\gamma)$, удовлетворяющей неравенствам*

(SM0) $\beta(0,\psi) \le \frac{\psi}{\frac{\varepsilon}{1-\varepsilon} + \left(1 - \frac{\chi}{n}\right)^n}$ *для всех* $\psi \in [\psi_0, 1]$,

(SM2a) $\beta(0,\gamma) \geq \frac{\gamma(1+\delta)}{\left(1-\frac{\chi}{n}\right)^n}$ *для всех* $\gamma \in (0,\gamma_0]$,

(SM2b) $\beta(\psi_0,\psi_0+\gamma) \geq \frac{\gamma(1+\delta)}{\left(1-\frac{\chi}{n}\right)^n(1+\alpha\chi)}$ *для любого* $\gamma \in (0,\gamma_0]$

при некоторых константах $\psi_0, \gamma_0, \alpha \in (0,1)$, $\psi_0 \geq \gamma_0$. *Кроме того, предположим что функции* $\lambda = \lambda(n), 1/\varepsilon = 1/\varepsilon(n)$, $1/\delta = 1/\delta(n)$ *и* $1/\chi = 1/\chi(n)$ *полиномиально ограничены и при достаточно большой константе* k *выполняется неравенство*

(SM3) $k\ln(n)/\varepsilon^2 \leq \lambda$.

Тогда среднее время оптимизации для такого ЭА на классе SparseLocalOpt$_{\alpha,\varepsilon}$ *полиномиально ограничено и составляет* $O(m(\lambda\ln(\lambda)+(n/\chi)^d)/\delta+1/\varepsilon)$, *где* d – *константа из определения класса* SparseLocalOpt$_{\alpha,\varepsilon}$.

Доказательство. Пусть $\mathcal{F}$ – задача псевдобулевой оптимизации из класса SparseLocalOpt$_{\alpha,\varepsilon}$ и для каждой функции $f \in \mathcal{F}$ в соответствии с определением SparseLocalOpt$_{\alpha,\varepsilon}$ задано разбиение пространства $\{0,1\}^n$ на уровни $A_1,\ldots,A_m$, при этом множество всех рассогласованных пар уроней относительно f состоит из (A_{i_1},A_{j_1}), ..., (A_{i_u},A_{j_u}).

Доказательство основано на применении теоремы 2, где $B := \cup_{v=1}^{u} A_{j_v}$. Согласно определению SparseLocalOpt$_{\alpha,\varepsilon}$, семейство множеств $B = B_f$, параметризованных функциями $f \in \mathcal{F}$, является ε-разреженным. Положим $x = P(i)$, где $i = \mathrm{Sel}(P)$, и пусть $y = \mathrm{Mut}(x)$. Для любого $\psi \geq \psi_0$ по лемме 2 выполняется условие (SM0), и по формуле полной вероятности

$$\begin{aligned}\Pr(y\in B) &\leq \Pr(x\in B)\Pr(y\in B \mid x\in B) + \Pr(x\notin B)\Pr(y\in B \mid x\notin B) \leq \\ &\leq \beta(0,\psi)\cdot\left(\varepsilon+(1-\varepsilon)\left(1-\frac{\chi}{n}\right)^n\right) + \mathrm{O}\left(\frac{1}{n}\right) \leq \\ &\leq \psi(1-\varepsilon) + \mathrm{O}\left(\frac{1}{n}\right) \leq \psi(1-\varepsilon) + \frac{\psi}{\psi_0}\mathrm{O}\left(\frac{1}{n}\right).\end{aligned}$$

Таким образом, для достаточно большого n условие (G0) выполняется при $\Delta := \varepsilon$.

Пусть $|P\cap B| \leq \psi_0\lambda$ и $|P\cap H_j| \geq \gamma_0\lambda$. Тогда для получения особи y, принадлежащей H_{j+1}, достаточно выбрать особь $x \in H_j$ оператором селекции, и изменить содержимое в подходящих $r \leq d$ битах процессе мутации. Заметим, что особи, принадлежащие множеству H_j, имеют преимущество по целевой функции перед прочими особями популяции, за исключением не более чем $\psi_0\lambda$ особей, принадлежащих множеству B. Следовательно, по условию (SM2b),

$$\begin{aligned}\Pr(y\in H_{j+1}) &\geq \Pr(x\in H_j)\Pr(y\in H_{j+1}\mid x\in H_j) \geq \\ &\geq \beta(\psi_0,\psi_0+\gamma_0)\,(\chi/n)^r\,(1-\chi/n)^{n-r} \geq \\ &\geq \frac{\gamma_0(1+\delta)}{(1-\chi/n)^n\,(1+\alpha\chi)}\,(\chi/n)^d\,(1-\chi/n)^{n-d}.\end{aligned}$$

Таким образом, условие (G1) выполняется при $z_j = \frac{\gamma_0(1+\delta)\chi^d}{(1+\alpha\chi)(n-\chi)^d}$, $j = 1,\ldots,m$.

Предположим, что $|P\cap B| \leq \psi_0\lambda$, $|P\cap H_j| \geq \gamma_0\lambda$ и $|P\cap H_{j+1}| \geq \gamma\lambda$. Тогда особи из множества H_{j+1} имеют преимущество перед другими особями популяции P, возможно, за исключением особей из B, число которых не превышает $\psi_0\lambda$.

Рассмотрим два случая. Пусть область уровня A_{j+1} входит в некоторую рассогласованную пару уровней (A_i, A_{j+1}). По условию теоремы, H_{j+1} является α-плотным множеством, поэтому из леммы 1 и условия (SM2b) следует, что

$$\begin{aligned}\Pr(y \in H_{j+1}) &\geq \Pr(x \in H_{j+1}) \Pr(y \in H_{j+1} \mid x \in H_{j+1}) \geq \\ &\geq \beta(\psi_0, \psi_0 + \gamma)\,(1 - \chi/n)^n\,(1 + \alpha\chi) \geq \gamma(1 + \delta).\end{aligned}$$

Если A_{j+1} не входит в рассогласованную пару, то особи из H_{j+1} имеют преимущество перед другими особями популяции, поэтому для получения особи из H_{j+1} достаточно выбрать «родительское» решение из H_{j+1} и не изменить ни одного бита в процессе мутации, что согласно (SM2a) происходит с вероятностью

$$\begin{aligned}\Pr(y \in H_{j+1}) &\geq \Pr(x \in H_{j+1}) \Pr(y = x) \geq \\ &\geq \beta(0, \gamma)\,(1 - \chi/n)^n \geq \gamma(1 + \delta).\end{aligned}$$

Таким образом, условие (G2) теоремы 2 выполняется.

Наконец, условие (G3) вытекает из (SM3), если константа k достаточно велика, т.к. γ_0 – константа, а величины λ, $1/\chi$, и $1/\Delta = 1/\varepsilon$ полиномиально ограничены. Из теоремы 2 вытекает оценка

$$E[T] = \mathrm{O}\left(\frac{1}{\varepsilon} + \frac{m\lambda\ln(\lambda)}{\delta} + \frac{1}{\delta}\sum_{j=1}^{m-1}\frac{1}{z_j}\right) = \mathrm{O}\left(\frac{1}{\varepsilon} + \frac{m}{\delta}\lambda\ln(\lambda) + \frac{m}{\delta}\left(\frac{n}{\chi}\right)^d\right),$$

где m, λ, $1/\chi$, $1/\delta$ и $1/\varepsilon$ полиномиально ограничены. □

Заметим, что при условии константных δ и ε полиномиальная ограниченность $E[T]$ была доказана в теореме 15 [4], но в явном виде оценка не была приведена. Как показано в [13; 14], в некоторых случаях более точные оценки времени оптимизации могут быть получены с использованием $\delta = O(1/n)$. Из доказательства теоремы 15 [4] можно извлечь оценку и для случая, когда δ и ε могут быть сколь угодно близки к 0, а именно, $E[T] = O(m(\lambda\ln(\lambda) + (n/\chi)^d)/\delta^2 + 1/\varepsilon)$, однако при этом зависимость от δ оказывается на порядок выше, чем в оценке из теоремы 3.

2.3. Расширение класса SparseLocalOpt$_{\alpha,\varepsilon}$

Рассмотрим расширение класса SparseLocalOpt$_{\alpha,\varepsilon}$, которое не обсуждалось в [4]. Заметим, что в пункте IV в определении 4 требуется, чтобы для каждой рассогласованной пары (A_{i_v}, A_{j_v}), $v = 1, \ldots, u$, множество H_{j_v} являлось α-плотным. Однако, каждое множество H_{j_v}, кроме областей $A_{j_v}, \ldots, A_{j+u}$, может содержать и области, не входящие в рассогласованные пары. Как будет видно из доказательства теоремы 4, нам не потребуетя α-плотности всего множества H_{j_v}. С целью ослабления ограничений из [4] на класс задач, где алгоритм 1 является эффективным, введем сначала определение *относительной* плотности.

Определение 6. *Пусть задано разбиение множества* $\{0,1\}^n$ *на* $m = m(n)$ *уровней* $A_1, \ldots, A_m$. *Уровень* A_j *при* $j \in [m]$ *называется* α-плотным относительно множества H_j, *если для любого* $x \in A_j$ *выполняется неравенство* $|S_1(x) \cap H_j| \geq \alpha n$.

Класс SparseLocalOptB$_{\alpha,\varepsilon}$ определим аналогично SparseLocalOpt$_{\alpha,\varepsilon}$ с той разницей, что вместо пункта IV из определения 4 для класса SparseLocalOptB$_{\alpha,\varepsilon}$ полагаем:

IV'. A_{j_v} – α-плотное множество относительно H_{j_v} для каждого $v \in [u]$.

Очевидно, SparseLocalOpt$_{\alpha,\varepsilon} \subseteq$ SparseLocalOptB$_{\alpha,\varepsilon}$, а значит, в худшем случае класс задач SparseLocalOptB$_{\alpha,\varepsilon}$ не проще чем SparseLocalOpt$_{\alpha,\varepsilon}$. В частности, из результата [4] о сложности SparseLocalOpt$_{\alpha,\varepsilon}$ для ЭА с элитой следует, что в худшем случае и на SparseLocalOptB$_{\alpha,\varepsilon}$ эти алгоритмы неэффективны.

Следующая лемма доказывается аналогично лемме 1.

Лемма 3. *Если при заданном разбиении пространства поиска на области уровня $A_1, \dots, A_m$ множество A_j является α-плотным относительно H_j, то*

$$\Pr\left(\mathrm{Mut}(x) \in H_j\right) > (1 - \chi/n)^n \, (1 + \alpha\chi) \quad \textit{при любом} \quad x \in A_j. \tag{2.7}$$

Теорема 4. *Пусть в алгоритме 1 используется стандартный оператор мутации с вероятностью мутации $p_{\mathrm{m}} = \chi/n$ и f-монотонный оператор селекции, который описывается функцией $\beta(\psi, \gamma)$, удовлетворяющей неравенствам*

(SM0) $\beta(0,\gamma) \le \frac{\gamma}{\frac{\varepsilon}{1-\varepsilon} + \left(1-\frac{\chi}{n}\right)^n}$ *для всех* $\gamma \in [\psi_0, 1]$,

(SM2a) $\beta(0,\gamma) \ge \frac{\gamma(1+\delta)}{\left(1-\frac{\chi}{n}\right)^n}$ *для всех* $\gamma \in (0, \gamma_0]$,

(SM2b') $\beta(\psi_0 + \gamma_0, \psi_0 + \gamma_0 + \gamma) \ge \frac{\gamma(1+\delta)}{\left(1-\frac{\chi}{n}\right)^n (1+\alpha\chi)}$ *для всех* $\gamma \in (0, \gamma_0]$

при некоторых константах $\psi_0, \gamma_0, \alpha \in (0,1)$, $\psi_0 \ge \gamma_0$. Кроме того, пусть функции $\lambda = \lambda(n)$, $1/\varepsilon = 1/\varepsilon(n), 1/\delta = 1/\delta(n)$ и $1/\chi = 1/\chi(n)$ полиномиально ограничены и при достаточно большой константе k выполняется неравенство

(SM3) $k \ln(n)/\varepsilon^2 \le \lambda$.

Тогда математическое ожидание времени оптимизации данного ЭА на классе SparseLocalOptB$_{\alpha,\varepsilon}$ *полиномиально ограничено и составляет $O(m(\lambda \ln(\lambda) + (n/\chi)^d)/\delta + 1/\varepsilon)$, где d – константа из условия II.*

Доказательство. Рассуждения повторяют доказательство теоремы 3, кроме проверки условия (G2). Проверим выполнение условия (G2). Предположим, что $|P \cap B| = \psi\lambda \le \psi_0\lambda$, $|P \cap H_j| \ge \gamma_0\lambda$ и $|P \cap H_{j+1}| \ge \gamma\lambda$. Вероятность селекции особей из некоторых уровней, содержащихся в множестве H_{j+1}, может снижаться из-за наличия в популяции особей из множества B, имеющих преимущество по целевой функции. Для учета этого снижения вероятностей селекции определим множество

$$B_{j+1} := \bigcup \left\{A_i \mid (A_i, A_k) \text{ – рассогласованная пара при } i \le j < k\right\}.$$

Это множество состоит из уровней, предшествующих уровню $j+1$, и несогласованных относительно областей уровня из H_{j+1}. Очевидно, $B_{j+1} \subseteq B$. Определим

$$A^-_{j+1} := \bigcup \left\{A_k \mid \exists i, k, \text{ такие что } i \le j < k \le m-1 \text{ и } (A_i, A_k) \text{ – рассогласованная пара}\right\},$$

состоящее из уровней, входящих в H_{j+1}, в которых вероятность селекции может быть снижена из-за присутствия в P особей из B. Кроме того, определим $H^+_{j+1} := H_{j+1} \setminus A^-_{j+1}$, а также

$$\gamma^- := \frac{1}{\lambda}|P \cap A^-_{j+1}| \text{ и } \gamma^+ := \frac{1}{\lambda}|P \cap H^+_{j+1}|.$$

Очевидно, $\gamma = \gamma^- + \gamma^+$. В популяции $\gamma^+\lambda$ «лучших» особей по целевой функции принадлежат H^+_{j+1}, следом по целевой функции идут особи из B_{j+1}, число которых $|P \cap B_{j+1}| \le |P \cap B| \le \psi_0\lambda$, далее – $\gamma^-\lambda$ особей из A^-_{j+1}.

Как и ранее, положим $x = P(i)$, где $i = \mathrm{Sel}(P)$, $y = \mathrm{Mut}(x)$. По лемме 3 и условию (SM2b'), ввиду f-монотонности селекции,

$$\begin{aligned}\Pr\left(x \in A^-_{j+1}\right)\Pr\left(y \in H_{j+1} \mid x \in A^-_{j+1}\right) \ge \\ \ge \beta(\psi_0+\gamma^+, \psi_0+\gamma^++\gamma^-)\,(1-\chi/n)^n\,(1+\alpha\chi) \ge \\ \ge \beta(\psi_0+\gamma_0, \psi_0+\gamma_0+\gamma^-)\,(1-\chi/n)^n\,(1+\alpha\chi) \ge \gamma^-(1+\delta).\end{aligned}$$

Если же решение x выбрано из множества H^+_{j+1}, то достаточно избежать мутации в каком-либо бите. По условию (SM2a),

$$\Pr\left(x \in H^+_{j+1}\right)\Pr\left(y \in H_{j+1} \mid x \in H^+_{j+1}\right) \ge \beta(0,\gamma^+)\,(1-\chi/n)^n \ge \gamma^+(1+\delta).$$

По формуле полной вероятности,

$$\begin{aligned}\Pr\left(y \in H_{j+1}\right) \ge \Pr\left(x \in A^-_{j+1}\right)\Pr\left(y \in H_{j+1} \mid x \in A^-_{j+1}\right) + \\ + \Pr\left(x \in H^+_{j+1}\right)\Pr\left(y \in H_{j+1} \mid x \in H^+_{j+1}\right) \ge \\ \ge \gamma^-(1+\delta) + \gamma^+(1+\delta) = \gamma(1+\delta).\end{aligned}$$

Таким образом, условие (G2) теоремы 2 выполняется. □

Заметим, что для выполнения условия (G0) в доказательстве теоремы 3 вместо оценки $\Pr(\mathrm{Mut}(x) \in B) = \mathrm{O}(1/n)$ при $x \notin B$ было бы достаточно того, что $\Pr(\mathrm{Mut}(x) \in B) = o(1)$. Это обеспечивается, если в определении ε-разреженного множества условие (SP2) заменить более слабым: $\forall x \in \{0,1\}^n \backslash B, \forall r \in [n-1]$, $|S_r(x) \cap B| = \mathrm{o}(C^r_n)$. Таким образом, класс задач, к которым применима теорема 4 может быть еще расширен.

3. Турнирная и линейная ранжирующая селекция на классе SPARSELOCALOPTB$_{\alpha,\varepsilon}$

В данном разделе приводятся два следствия из теоремы 3, демонстрирующие наборы параметров, при которых ЭА без элитных особей являются эффективными на классе SPARSELOCALOPTB$_{\alpha,\varepsilon}$. Рассматриваются ЭА, основанные на общей схеме алгоритма 1 и использующие турнирную или линейную ранжирующую селекцию. При этом вероятность мутации выбирается достаточно большой, но не более известного порогового значения из [22].

Следствие 1. *Алгоритм 1 с турнирной селекцией при размере турнира $k = 3$, полиномиально ограниченной численностью популяции $\lambda \ge c\ln(n)$, где c – достаточно большая константа, и стандартной мутацией, где $\chi = 1.098612$, имеет полиномиально ограниченное среднее время оптимизации на задачах из класса SPARSELOCALOPTB$_{\alpha,\varepsilon}$, где $\alpha = 1/4$, $\varepsilon = 3/10^5$.*

Доказательство. При размере турнира $k = 3$ имеем

$$\beta(\psi, \psi+\gamma) = (1-\psi)^3\left(1-\left(1-\frac{\gamma}{1-\psi}\right)^3\right).$$

Проверим выполнение условий теоремы 3 с набором параметров

$$\psi_0 := \frac{1}{40}, \gamma_0 := \frac{1}{6250}, \delta := \frac{1}{30000.}$$

Заметим, что вероятность отсутствия изменений при мутации оценивается сверху

$$\left(1 - \frac{\chi}{n}\right)^n < e^{-\chi} < \frac{3335}{10000},$$

а в случае $n \geq 10^4$ справедлива оценка снизу

$$\left(1 - \frac{\chi}{n}\right)^n = \left(1 - \frac{\chi}{n}\right)^{\left(\frac{n}{\chi}-1\right)\chi} \left(1 - \frac{\chi}{n}\right)^{\chi} \geq e^{-\chi} \left(1 - \frac{\chi}{10^4}\right)^{\chi} > \frac{3334}{10000}.$$

Введем вспомогательную функцию

$$h(\psi, \gamma) := \frac{\beta(\psi, \psi + \gamma)}{\gamma} = \gamma^2 + 3(1 - \psi)(1 - \gamma - \psi),$$

и заметим, что она не возрастает по ψ и по γ при $(\psi, \gamma) \in [0, 1/3] \times (0, 1]$, т.к. в этой области

$$\begin{aligned} \frac{\partial h(\psi, \gamma)}{\partial \gamma} &= -3 + 2\gamma + 3\psi \leq 0, \\ \frac{\partial h(\psi, \gamma)}{\partial \psi} &= -6 + 3\gamma + 6\psi \leq -1. \end{aligned}$$

Следовательно, для всех $\gamma \in [\psi_0, 1]$

$$\frac{\beta(0, \gamma)}{\gamma} = h(0, \gamma) \leq h(0, \psi_0) < \frac{29256}{10000} < \frac{1}{\frac{\varepsilon}{1-\varepsilon} + \left(1 - \frac{\chi}{n}\right)^n},$$

откуда вытекает условие (SM0). Аналогично, для всех $\gamma \in (0, \gamma_0]$

$$\frac{\beta(0, \gamma)}{\gamma} = h(0, \gamma) \geq h(0, \gamma_0) = \frac{\beta(0, \gamma_0)}{\gamma_0} > \frac{299952}{100000} > \frac{1 + \delta}{\left(1 - \frac{\chi}{n}\right)^n},$$

поэтому выполнено условие (SM2a). Далее, для всех $\gamma \in (0, \gamma_0]$ имеем

$$\begin{aligned} \frac{\beta(\psi_0 + \gamma_0, \psi_0 + \gamma_0 + \gamma)}{\gamma} &= h(\psi_0 + \gamma_0, \gamma) \geq h(\psi_0 + \gamma_0, \gamma_0) > \frac{57}{20} \\ &> \frac{1 + 30000^{-1}}{\frac{3334}{10000} \cdot \left(1 + \frac{109812}{10^5} \cdot \frac{1}{4}\right)} > \frac{1 + \delta}{(1 - \chi/n)^n (1 + \alpha\chi)}, \end{aligned}$$

и (SM2b') выполнено. Условие (SM3) имеет место при достаточно большой константе c.
□

Следствие 2. *Алгоритм 1 с использованием турнирной селекции при размере турнира $k = 2$ или линейной ранжирующей селекции при $\eta = 2$, полиномиально ограниченной численностью популяции $\lambda \geq c\ln(n)$ при достаточно большой константе c, и параметре мутации $\chi = 0.693146$ имеет полиномиально ограниченное среднее время оптимизации на задачах из класса* SparseLocalOptB$_{\alpha,\varepsilon}$ *при $\alpha = 4/9$ и $\varepsilon = 1/100$.*

Доказательство. При линейной ранжирующей селекции

$$\beta(\gamma_1, \gamma_2) = \int_{\gamma_1}^{\gamma_2} \alpha_{\text{linear}}(x)dx$$

где $\alpha_{\text{linear}}(\gamma) = \eta(1-2\gamma)+2\gamma$. Поэтому при $\eta = 2$ имеем: $\beta(\psi, \psi+\gamma) = \gamma(2-\gamma-2\psi)$. Такая же функция β характеризует турнирную селекцию с размером турнира 2. Выберем параметры

$$\psi_0 := \frac{1}{5}, \quad \gamma_0 := 10^{-7}, \text{ и } \quad \delta := 10^{-6}.$$

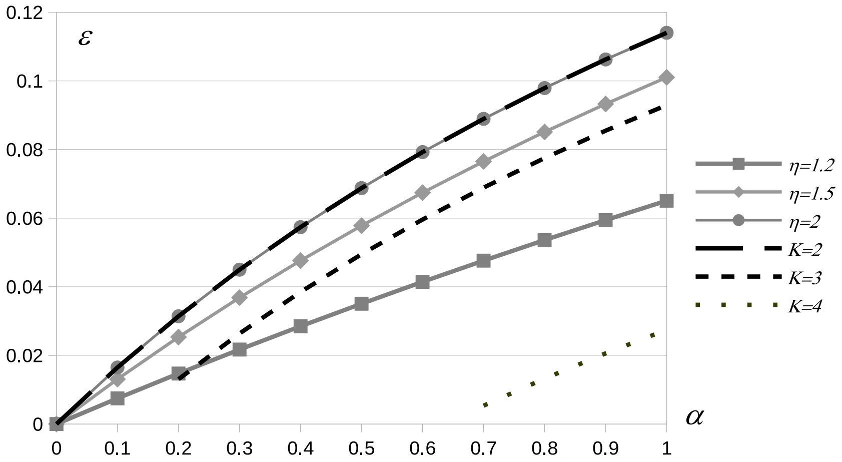
Заметим, что имеет место оценка сверху

$$\left(1 - \frac{\chi}{n}\right)^n < e^{-\chi} < \frac{50000060}{10^8}.$$

С другой стороны, для $n \geq 10^{10}$ имеем нижнюю оценку

$$\left(1 - \frac{\chi}{n}\right)^n = \left(1 - \frac{\chi}{n}\right)^{\left(\frac{n}{\chi}-1\right)\chi} \left(1 - \frac{\chi}{n}\right)^{\chi}$$
$$\geq e^{-\chi}\left(1 - \frac{\chi}{10^4}\right)^{\chi} > \frac{5000005902}{10000000000}.$$

Окончание доказательства аналогично доказательству следствия 1. □



Максимальное ε, удовлетворяющее условиям теоремы 3, как функция от α

Гарантии эффективной работы ЭА распространяются не только на значения параметров α, ε, указанные в следствиях 1 и 2. На рисунке 1 приведены графики максимальных значений ε, удовлетворяющих условиям теоремы 3 для турнирной селекции с размером турнира $k = 2, 3$ и 4, а также, для линейной ранжирующей селекции с параметром $\eta = 1.2, 1.5$ и 2. Графики построены интерполяцией максимальных значений ε, найденных при значениях параметра $\alpha = 0.1, 0.2, \ldots, 1$. Максимальные значения ε были получены с использованием решателя BARON в режиме глобальной оптимизации (максимизация ε), когда переменными являлись $\varepsilon, \psi_0, \chi, \delta, \gamma_0$, а ограничениями – условия (SM0), (SM2a) и (SM2b).

Рисунок 1 показывает, что турнираня селекция при $k = 2$ и линейная ранжирующая селекция при $\eta = 2$ имеют одинаковые графики (что следует из совпадения соответствующих функций $\beta(\psi, \gamma)$) и именно эти значения параметров позволили достичь наиболее высокие значения ε при всех рассмотренных α.

4. Иллюстративные примеры задачи о вершинном покрытии

Взвешенная задача вершинного покрытия (ЗВП) может быть сформулирована следующим образом. Пусть $G = (V, E)$ – граф с множеством вершин $V = \{v_1, v_2, \dots, v_n\}$ и множеством ребер E. Для каждой $v \in V$ определен вес $w(v) > 0$. Подмножество $C \subseteq V$ называется *вершинным покрытием*, если каждое ребро из E инцидентно хотя бы одной вершине из C (такие ребра называются *покрытыми*). Требуется найти вершинное покрытие минимального веса.

Пусть для кодировки решений используются битовые строки длины n. Для битовой строки x длины n определим множество $V(x) \subseteq V$, состоящее из всех вершин v_i, таких что $x_i = 1$, и только их. Как было предложено в [25], гл. 12, приспособленность особей определяется лексикографическим упорядочением двух критериев: 1) количество ребер, не покрытых вершинами из $V(x)$, 2) общий вес вершин в $V(x)$. Аналитически соответствующий критерий минимизации может быть определен как $f(x) := |E'(x)| \cdot S + \sum_{i=1}^{n} w(v_i)x_i$, где $E'(x) := \{e = u_i v_j \in E \ : \ x_i = x_j = 0\}$ – множество непокрытых ребер и $S := \sum_{i=1}^{n} w(v_i)$.

Рассмотрим ЗВП на графе-звезде $K_{1,n-1}$, а именно, предположим что G – полный двудольный граф с $\ell = n - 1$ вершинами $v_1, \dots, v_\ell$ в одной доле, обозначаемой V_2, и одной вершиной v_n в другой доле.

При заданном параметре $\alpha \in (0, 1)$ будем рассматривать только те примеры ЗВП, в которых $w(v_n) \le (1 - \alpha)\ell$ и $w(v_1) = \dots = w(v_\ell) = 1$. Глобально оптимальное покрытие $\{v_n\}$ имеет вес $(1-\alpha)\ell$. Локальный (не глобальный) оптимум относительно окрестности Хэмминга радиуса 1 — это множество V_2.

Утверждение 1. *Пусть ЗВП задана двудольным графом $K_{1,n-1}$, где одна из долей состоит из единственной вершины v_n, а другая доля образована остальными вершинами и им приписаны единичные веса, причем $w(v_{n+1}) \le (1 - \alpha)(n - 1)$ при некотором $0 < \alpha < 1$. Тогда критерий минимизации $f(x)$ принадлежит классу $SparseLocalOptB_{\alpha,1/n}$.*

Доказательство. Разобъем множество всех строк $\{0, 1\}^n$ на области A, B, C и D с аналогичными свойствами, как при задании функции Funnel в [3]. Определим эти области и подмножества уровней следующим образом.

- $A = \cup_{j=1}^{\ell} A_j$, где множество уровня A_j, $j = 1, \dots, \ell$, состоит из таких строк x, для которых $V(x)$ покрывает ровно $j - 1$ ребро, $j \le \ell - 1$. Таким образом, множества уровней $A_1, \dots, A_\ell$ упорядочены по убыванию штрафа $|E'(x)| \cdot S$.

- $B = A_{\ell+1}$, где множество уровня $A_{\ell+1}$ образовано одной битовой строкой x, кодирующей покрытие V_2.

- Области C и D состоят из битовых строк, кодирующих покрытия, в которых содержится вершина v_n. Они образованы следующими множестaвми уровня:

$$A_{\ell+1+j} = \{x \ : \ x_n = 1, \ \sum_{i=1}^{\ell}(1 - x_i) = j\},$$

  $j = 1, \dots, \ell$ и $A_m := \{(0, \dots, 0, 1)\}$, где $m = 2\ell + 1$. Положим, что в область C входят решения из $\bigcup\limits_{1 \le j < \ell - \alpha\ell} A_{\ell+1+j}$ с весом больше, чем $w(v_n) + \alpha\ell$ (таким образом, каждое решение в области C содержит более $\alpha\ell$ вершин из V_2), а область D состоит из решений из $\bigcup\limits_{\ell-\alpha\ell\le j\le\ell} A_{\ell+1+j}$ с весом не более $w(v_n)+\alpha\ell$ (здесь же находится оптимальное решение, т. е. единственный элемент множества уровня A_m).

Каждый уровень в подмножествах A, C и D характеризуется меньшим значением целевой функции, чем предыдущий уровень в том же подмножестве. Для любого $x_A \in A$, для локального оптимума $x_B \in B$, для любого $x_C \in C$, и любого $x_D \in D$ выполняются следующие неравенства:

$$f(x_A) < f(x_B),\ f(x_B) > f_{vcp}(x_C),$$

$$f(x_C) < f_{vcp}(x_D),\ f(x_B) \le f_{vcp}(x_D).$$

Также заметим, что для каждого $x \in A_j$ при $j \le \ell$ существует решение $x' \in H_{j+1}$, такое что $H(x,x')=1$ (например, достаточно инвертировать ноль в позиции n). Для $j=\ell+1,\dots,m-1$ аналогичное утверждение верно, потому что среди первых n битов решения x должна быть по крайней мере одна единица при $j<m$. Таким образом, условие II из определения 4 выполнено при $d=1$ и все предшествующие ему условия также выполнены. Осталось проверить условия III и IV' для рассогласованных пар, где первый уровень пары совпадает с множеством B, а второй – содержится в множестве C.

Разреженность $\varepsilon=1/n$ для области B следует из того факта, что B по определению является одноэлементным множеством. Действительно, для любого $x\in\{0,1\}^n$ и всех $r\in[n-1]$ имеем

$$|S_r(x)\cap B| \le 1 = \frac{1}{\binom{n}{r}}\binom{n}{r} \le \frac{1}{\binom{n}{1}}\binom{n}{r},$$

Следовательно, B является $1/n$-разреженным множеством по определению 3.

Наконец покажем плотность любого уровня A_j из области C относительно множества H_j. Если $x\in A_j\subseteq C$, то $V(x)$ содержит не менее $\alpha\ell+1\ge\alpha n$ вершин из подмножества V_2, и удаление одной из этих вершин дает элемент области C или D). Таким образом, A_j из области C является α-плотным относительно H_j. $\square$

Для сравнения алгоритмов с элитными особями и без них рассмотрим алгоритм рандомизированного локального поиска (RLS) как пример простого EA с элитой. Пусть x^t обозначает текущее решение на итерации t алгоритма RLS. Изначально x^0 генерируется с равномерным распределением вероятностей на $\{0,1\}^n$. Итерация t состоит в построении потомка y путем инверсии одного случайно выбранного бита в x^t. Если $f(y)$ по приспособленности лучше, чем $f(x^t)$, то x^t заменяется на y, в противном случае $x^{t+1}=x^t$.

Чтобы достичь локального оптимума задачи f, начиная с решения где $x_n=0$, алгоритму RLS необходимо сделать не более $n-1$ шагов улучшения, когда добавляются только вершины из V_2, а не вершина v_n. (Шаги, направленные на ухудшение функции приспособленности, не учитываются.) По лемме 12.3 из [25] заключаем, что вероятность достижения локального оптимума с помощью RLS составляет не менее $1/n$. И как только локальный оптимум достигнут, RLS остается в нем навсегда. Следовательно, справедливо следующее

Утверждение 2. *На рассматриваемом семействе задач вершинного покрытия с параметром $0<\alpha<1$ алгоритм RLS с критерием оптимизации f имеет бесконечное математическое ожидание времени оптимизации.*

С другой стороны, из следствия 1 и утверждения 1 вытекает

Следствие 3. *На рассматриваемом семействе задач вершинного покрытия с параметром* $0 < \alpha \leq 1/4$*, когда n достаточно велико, алгоритм 1 с турнирной селекцией с размером турнира* $k = 3$*, полиномиально ограниченной численностью популяции* λ*,* $c\ln(n) \leq \lambda$*, для достаточно большой константы c, при параметре мутации* $\chi = 1.098612$ *имеет полиномиально ограниченное математическое ожидание времени оптимизации.*

Аналогичный результат вытекает из следствия 2 и утверждения 1 для турнирнира размером $k = 2$ и линейной ранговой селекции.

Благодарности

Автор признателен Д.-К. Дангу и П.К. Лере за их участие в работе на начальном этапе.

СПИСОК ЛИТЕРАТУРЫ

1. **Auger A., Doerr B.** Theory of Randomized Search Heuristics: Foundations and Recent Developments. Singapore. World Scientific, 2011, Ser. Theoretical Computer Science, vol. 1, 359 p. doi: 10.1142/7438 .
2. **Corus D., Dang D.-C., Eremeev A.V., and Lehre P.K.** Level-based analysis of genetic algorithms and other search processes // IEEE Trans. Evolutionary Computation, 2018. Vol. 22, no. 5, P. 707–719. doi: 10.1109/TEVC.2017.2753538 .
3. **Dang D.-C., Eremeev A.V., and Lehre P.K.** Escaping local optima with non-elitist evolutionary algorithms // Proc. of AAAI Conference on Artificial Intelligence (AAAI'2021). 2021. P. 12275-12283.
4. **Dang D.-C., Eremeev A.V., and Lehre P.K.** Non-elitist evolutionary algorithms excel in fitness landscapes with sparse deceptive regions and dense valleys // Proc. of the Genetic and Evolutionary Computation Conference (GECCO'2021). 2021. P. 1133–1141. doi: 10.1145/3449639.3459398 .
5. **Dang D.-C., Eremeev A.V., and Lehre P.K.** Corrigendum to "Non-elitist evolutionary algorithms excel in fitness landscapes with sparse deceptive regions and dense valleys" (GECCO 2021), University of Birmingham, 2022.
6. **Dang D.-C., Eremeev A.V., Lehre P.K., and Qin X.** Fast non-elitist evolutionary algorithms with power-law ranking selection // Proc. of the Genetic and Evolutionary Computation Conference (GECCO'2022). 2022. P. 1372–1380. doi: 10.1145/3512290.3528873 .
7. **Dang D.-C., Lehre P.K., and Qin X.** Self-adaptation can help evolutionary algorithms track dynamic optima // Proc. of the Genetic and Evolutionary Computation Conference (GECCO'2023). 2023. P. 1619-–1627. doi: 10.1145/3512290.3528873 .
8. **Dang D.-C., Eremeev A.V., and Qin X.** Empirical evaluation of evolutionary algorithms with power-law ranking selection // Proc. of the 13th IFIP International Conference on Intelligent Information Processing. 2024. P. 217–232,
9. **Dang D.-C., Jansen T., Lehre P.K.** Populations can be essential in tracking dynamic optima // Algorithmica. 2017. Vol. 78, no. 2. P. 660–680. doi: 10.1007/s00453-016-0187-y .
10. **Dang D.-C., Lehre P.K.** Efficient optimisation of noisy fitness functions with population-based evolutionary algorithms // Proc. of the 2015 Conference on Foundations of Genetic Algorithms (FOGA'2015). 2015. P 62–68. doi: 10.1145/2725494.2725508 .
11. **Dang D.-C., Lehre P.K.** Runtime analysis of non-elitist populations: From classical optimisation to partial information // Algorithmica. 2016. Vol. 75. P. 428–461. doi: 10.1007/s00453-015-0103-x .
12. **Doerr B.** Does comma selection help to cope with local optima? // Proc. of the 2020 Genetic and Evolutionary Computation Conference (GECCO 2020). 2020. P. 1304–1313. doi: 10.1145/3377930.3389823 .
13. **Doerr B., Kötzing T.** Multiplicative up-drift // Proc. of the 2019 Genetic and Evolutionary Computation Conference (GECCO 2019). 2019. P. 1470—1478. doi: 10.1145/3321707.3321819 .

14. **Doerr B., Kötzing T.** Multiplicative up-drift // Algorithmica. 2021. Vol. 83, no. 10. P. 3017–3058 doi: 10.1007/s00453-020-00775-7 .
15. **Doerr B., Le H.P., Makhmara R., and Nguyen T.D.** Fast genetic algorithms // Proc. of the 2017 Genetic and Evolutionary Computation Conference (GECCO 2017). 2017. P. 777–784. doi: 10.1145/3071178.3071301 .
16. **Doerr C., Lengler J.** Introducing elitist black-box models: When does elitist behavior weaken the performance of evolutionary algorithms? // Evolutionary Computation. 2017. Vol. 25, no. 4. P. 587–606.
17. **Droste S., Jansen T. and Wegener I.** Upper and lower bounds for randomized search heuristics in black-box optimization // Theory of Computing Systems. 2006. Vol. 39, no. 4. P. 525–544. doi: 10.1007/s00224-004-1177-z .
18. **Dubhashi D., Panconesi A.** Concentration of Measure for the Analysis of Randomized Algorithms. New York: Cambridge University Press, 2009. 195 p. doi: 10.1017/CBO9780511581274 .
19. **Eremeev A.V.** Modeling and analysis of genetic algorithm with tournament selection // Proc. of Artificial Evolution (AE 1999). 2000. Lecture Notes in Computer Science. Vol. 1829. P. 84-95.
20. **Goldberg D.E.** Genetic Algorithms in search, optimization and machine learning. Reading, MA: Addison-Wesley, 1989. 412 p.
21. **Goldberg D.E., Deb K.**, A comparative analysis of selection schemes used in genetic algorithms // Foundations of Genetic Algorithms. 1991. P. 69–93.
22. **Lehre P.K.**, Fitness-levels for non-elitist populations // Proc. of the 2011 Genetic and Evolutionary Computation Conference (GECCO 2011). 2011. P. 2075–2082. doi: 10.1145/2001576.200185 .
23. **Lehre P.K., Qin X.** Self-adaptation can help evolutionary algorithms track dynamic optima // Proc. of the Genetic and Evolutionary Computation Conference (GECCO 2023). 2023. P. 1619–1627. doi: 10.1145/3594805.3607128 .
24. **Lehre P.K., Qin X.** Self-adaptation can improve the noise-tolerance of evolutionary algorithms // Proc. of the 17th ACM/SIGEVO Conference on Foundations of Genetic Algorithms. (FOGA 23). 2023. P. 105–116. doi: 10.1145/3594805.3607128 .
25. **Neumann F., Witt C.** Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity. Natural Computing Series. Berlin, Heidelberg, Springer, 2010. p. 216. doi: 10.1007/978-3-642-16544-3 .
26. **Whitley D.** The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best // Proc. of the Third International Conference on Genetic Algorithms. 1989. P. 116–121.
27. **Борисовский П.А., Еремеев А.В.** О сравнении некоторых эволюционных алгоритмов // Автомат. и телемех. № 3. 2004. С. 3–9.
28. **Зубков А.М., Попов Н.Н.** Отношение частичного порядка, порожденное распределениями числа занятых ячеек // Матем. заметки. 1982. Т. 32, №1. С. 97–102.

Еремеев Антон Валентинович
д-р физ.-мат. наук, доцент
Новосибирский государственный университет
г. Новосибирск,
Омский государственный университет им. Ф.М. Достоевского
г. Омск,
e-mail: eremeev@ofim.oscsbras.ru

Поступила 24.02.2024

Обязательна информация об авторе на английском языке:

Anton Valentinovich Eremeev, Dr. Phys.-Math. Sci., Docent, Novosibirsk State University, Novosibirsk, 630090 Russia; Dostoevsky Omsk State University, Omsk, 644077 Russia
e-mail: eremeev@ofim.oscsbras.ru .