

УДК 519.71

О.Г. Руденко, Р.В. Бобнев

ГИБРИДНЫЙ ГЕНЕТИЧЕСКИЙ АЛГОРИТМ НА ОСНОВЕ БИОЛОГИЧЕСКОГО АПОПТОЗА

Введение. Генетические алгоритмы (ГА), предложенные Дж. Холландом [1] и первоначально предназначенные для решения задач оптимизации в качестве достаточно эффективного механизма комбинаторного перебора вариантов решений, в настоящее время широко используются при решении многих проблем, которые могут быть сформулированы как задачи оптимизации, в частности, к таким задачам относится нахождение экстремумов мультимодальных функций. Сам по себе ГА представляет алгоритм, основанный на биологической модели эволюции, а именно естественного отбора. В рамках терминологии ГА все действия производятся над генотипами, которые по аналогии с биологическим термином, представляют собой набор генов, где геномом может выступать практически любое числовое значение, над которым может быть определена операция мутации (понятие мутации кратко описано ниже). В самом простом виде ГА можно разбить на следующие составляющие:

- создание начальной популяции;
- селекция особей и размножение (кроссовер или кроссинговер);
- мутация особей;
- определение уровня приспособленности каждой особи (фитнесс-функции);
- естественный отбор на основе приспособленности каждой особи.

Традиционный ГА показан на рис. 1 сплошными линиями.

Создание популяции подразумевает создание случайной группы особей, каждая из которых является генотипом, состоящим из набора генов.

Выбираемые каким-либо образом две пары, от которых будут произведены потомки, участвуют в репродукции, представляющей собой скрещивание (кроссинговер). Для выбора таких пар применяют ряд методов, среди которых наиболее распространенными являются метод «рулетки», ранжирование, случайный выбор, локальный отбор, отбор на основе усечения, турнирный отбор и др. [1–3]. Сама операция репродукции представляет собой скрещивание (кроссинговер) двух генотипов.

Мутация особей или случайное изменение гена осуществляется путем изменения последовательности знаков (при бинарном представлении это инверсия случайного бита). Так как мутация гена порождает новый ген с другими свойствами, в ГА ее можно заменить на случайный выбор одного гена из заданного множества всех возможных генов.

Определение фитнес-функции и вычисление ее для каждого индивида популяции является сугубо проблемно-зависимой частью, и выбор ее зависит не только от поставленной задачи, но и от способа формализации задачи в рамках ГА. Следует отметить, что если операции кроссинговера и мутации более или менее формализованы, и по ним можно сделать приблизительную оценку производительности и затрат времени вычислений, то определение фитнес-функции осуществляется субъективно и ее вычисление для каждой особи может существенно влиять на время работы алгоритма.

© О.Г. РУДЕНКО, Р.В. БОБНЕВ, 2013

*Международный научно-технический журнал
«Проблемы управления и информатики», 2013, № 1*

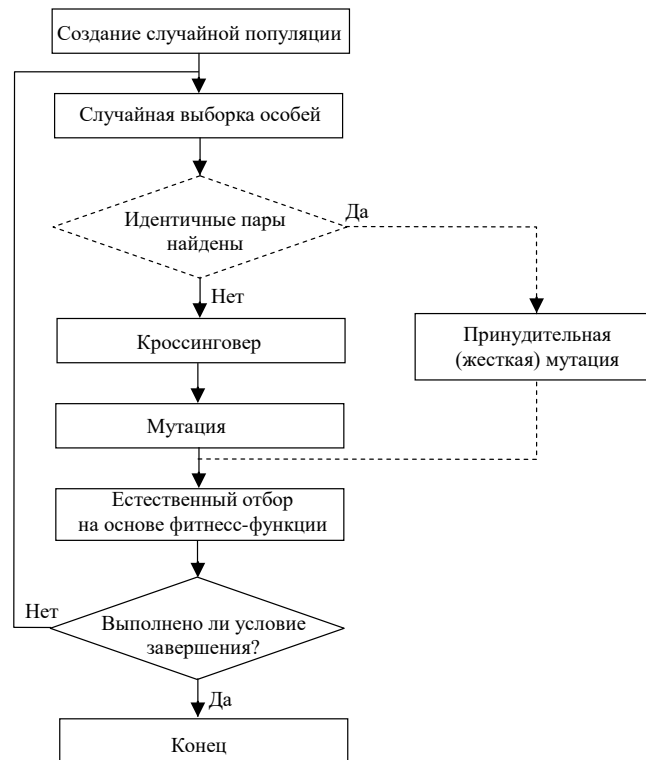


Рис. 1

Естественный отбор является, по сути, отсеиванием неприспособленных особей из всей популяции. Однако простое отсеивание всех слабых особей нецелесообразно, так как из индивидуумов с наивысшим показателем фитнес-функции можно быстро попасть в локальный экстремум, а слабые особи, которые потенциально могут быть новыми решениями, попросту будут отброшены. Данный недостаток должен устраняться мутацией, но не устраняется в силу ее обычно низкой вероятности. Вследствие этого при решении мультимодальных задач зачастую наблюдается преждевременная сходимость в локальные экстремумы и не гарантируется нахождение глобального экстремума [2, 3].

Для устранения данного недостатка были предложены методы инбридинг и аутбридинг, использующие некоторое условие идентичности и отличающиеся лишь способом выбора пар для репродукции путем формирования таких пар на основе близкого или дальнего родства соответственно (под родством обычно понимается расстояние между особями популяции в пространстве параметров). Комбинация этих подходов используется при решении многоэкстремальных задач. Существуют также работы по самоорганизации вероятности мутации, меняющейся в процессе работы ГА, которая корректируется в зависимости от степени сходимости к локальному минимуму. Одна из таких оценок степени сходимости — наблюдение разности среднего и максимального значения целевой функции по популяции.

В данной работе предлагается модификация ГА, отличающаяся от методов инбридинг и аутбридинг тем, что она не оказывает влияния на способ селекции особей. Хотя она имеет некоторое сходство с адаптивным изменением вероятности мутации, существенная разница состоит в том, что предлагаемая модификация видоизменяет способ формирования новых пар потомков. И данная адаптация происходит в процессе работы алгоритма, независимо от того, вся ли популяция уже «забита» локальными экстремумами или только небольшая ее часть.

Апоптоз. В биологии существует понятие апоптоза — явления программируемой клеточной смерти, сопровождаемой набором характерных цитологических признаков (маркеров апоптоза) и молекулярных процессов, имеющих различия у одноклеточных и многоклеточных организмов [4]. Роль апоптоза в организме

крайне важна, так как он позволяет убить «ненужные» идентичные клетки. Основной особенностью апоптоза является то, что этот процесс программирует клетку на смерть путем мутации. Логично предположить, что если для кроссинговера были выбраны две идентичные особи, то каков бы ни был способ обмена генами, потомки будут также идентичными и полностью унаследуют приспособленность родителей. Поэтому единственной возможностью введения разнообразия в генофонд является мутация. Но так как мутация имеет, как правило, низкую вероятность, и учитывая, что она обычно происходит для одного гена, максимум что можно получить, это более точное решение, а при наличии нескольких решений найдено будет только одно. По аналогии с биологическим процессом можно ввести некий механизм апоптоза и в ГА.

Для этого необходимо модифицировать метод селекции и кроссинговера. Если после селекции две особи оказались идентичными, то осуществлять кроссинговер не имеет смысла, так как предки будут полностью соответствовать родителям. Заметим, что подобный механизм регуляции существует и в природе. Для введения данного механизма в ГА необходимо лишь добавить определенную логику: если после селекции оба родителя являются идентичными, то вместо кроссинговера следует провести репликацию и мутацию для двух потомков и одного из родителей. Репликация необходима для поддержания численного соотношения вероятности кроссинговера, а мутация одного из родителей — для уменьшения числа селекций с идентичными родителями. Следует отметить, что одну из мутаций можно произвести стандартным способом (случайный ген), а две — более глобально. Предложенная дополнительная стандартная мутация необходима для повышения точности уже имеющегося решения (если два родителя идентичны, то их приспособленность высока, а это означает, что они либо являются оптимальным решением, либо достаточно близки к нему). Под глобальной мутацией подразумевается мутирование хотя бы половины генов. Это необходимо для достаточного удаления новой особи от своих родителей, что не позволит привести в процессе эволюции опять к ним же. Достичь лучших результатов можно генерированием полностью нового генотипа, добавив, таким образом, целый новый вид. Модификация ГА с использованием механизма апоптоза показана пунктиром на рис. 1.

Особенности работы ГА при введении механизма апоптоза. В классическом ГА вся популяция постепенно забивается клонами одной и той же особи с максимальным значением фитнес-функции (назовем эту особь лидирующей). При этом вероятность двух идентичных потомков при скрещивании стремится к единице до определенного количества эпох. При введении механизма апоптоза каждая операция скрещивания, в которой участвуют идентичные родители, заменяется, по сути, добавлением новых особей. Таким образом, вся популяция проходит один из трех типов скрещивания: скрещивание с лидирующей особью, с другой какой-либо особью (если такие еще остались в популяции) и добавление кардинально новой особи (когда оба предка являются клонами лидера). Как правило, после скрещивания случайной особи с лидирующей потомок получается с лучшим значением фитнес-функции, и далее этот потомок, постепенно перебирая разнообразные гены, эволюционирует в ту же лидирующую особь вследствие влияния генов лидеров на последующие получаемые образы. Если же в скрещивающиеся пары не попала лидирующая особь, происходит обычная работа ГА на области случайной популяции, где нет одного из решений. Это приводит, по аналогии с традиционным ГА, к нахождению еще одного решения в популяции, которая осталась не «забитой» клонами лидера. В результате новая преобладающая особь заполняет всю оставшуюся часть популяции уже своими клонами, и вся исходная популяция разделяется, как правило, на два множества идентичных особей, первое из которых состоит из особи, найденной первой (от которой пытался отойти механизм апоптоза), а второе — из особи, найденной обычным ГА на оставшейся части популяции. Схематически описанный процесс формирования конечной популяции при использовании апоптоза показан на рис. 2.

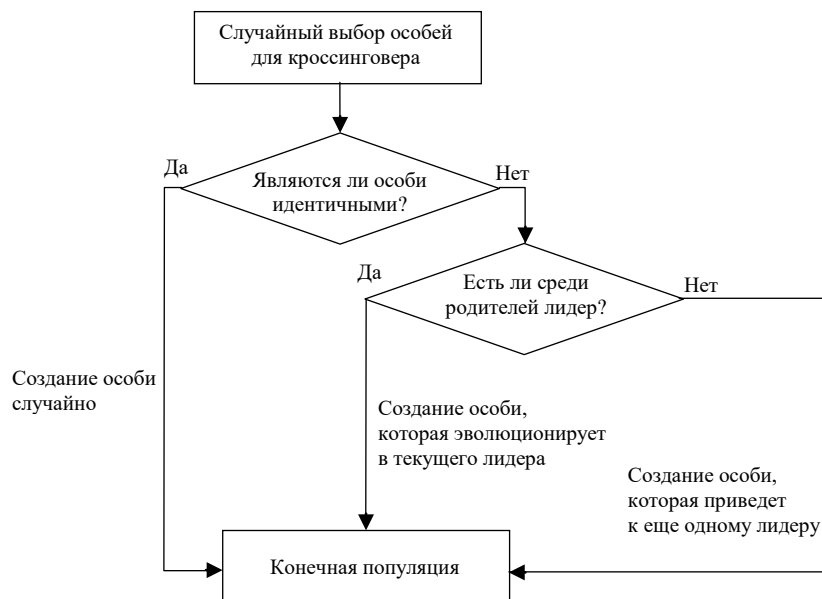


Рис. 2

Как видно из рисунка, конечная популяция после ряда эпох будет состоять из случайных особей (их доля крайне мала), из особей клонов первого найденного лидера и особей клонов лидера, который получился в результате неявного отделения популяции от первого найденного.

Хотя на первый взгляд может показаться, что механизм апоптоза должен вносить достаточное количество случайных особей и по аналогии со вторым найденным лидером будут найдены и остальные решения, на самом деле этого не происходит вследствие того, что вероятность выбора двух идентичных особей, когда популяция заполнена двумя лидирующими особями, существенно меньше вероятности выбора из популяции, содержащей лишь одну лидирующую особь. Так если вначале она стремилась к единице (лишь потому, что был один единственный лидер), то сейчас она стремится к 0,25. Это получается, вследствие того, что выбор двух особей необходимо произвести из двух подмножеств идентичных особей, а не из одного, как было ранее. Поскольку теперь уже есть два вида особей (два лидера), то вероятность выбора каждого из них составляет 0,5, а вероятность двойного такого выбора (когда две особи идентичны между собой и являются клоном одного из лидеров), будет равна 0,25.

Увеличение размера популяции не решает эту проблему, так как в результате лишь увеличиваются размеры подмножеств лидеров: при попадании в кроссингвер лидера и обычной особи операция приводит по пути эволюции к одному из подмножеств, что и влечет за собой разрастание каждого из них. Однако данную особенность работы алгоритма можно использовать как показатель «насыщенности» популяции (под «насыщенностью» понимается такое состояние популяции, при котором практически вся она состоит из клонов одной или более особей, имеющих наивысшие значения функции приспособленности). Покажем, как это можно использовать на практике. Зададим, например, счетчик, состояние которого будет увеличиваться на единицу каждый раз, когда срабатывает механизм апоптоза, при каждом выборе идентичных родителей. Отношение значения этого счетчика к общему числу особей, участвовавших в кроссингвере, даст нам вероятность апоптоза. Как только она приблизится к значению 0,25, можно делать вывод, что ГА достиг своего «насыщения» и популяция в основном состоит из двух подмножеств решений. Далее, если число эпох не превысило заданный порог, можно предложить ряд вариантов нахождения остальных экстремумов функции на основе того факта, что популяция «насыщена».

Опишем некоторые из них.

Вариант 1. Самый простой вариант состоит в заполнении популяции совершенно новыми особями, сохраняя при этом два (или более, если таковые имеются) найденных решения отдельно от новой популяции. Далее происходит работа ГА, аналогичная предыдущей, но найденные ранее решения в алгоритме не участвуют. При этом, однако, нет гарантии того, что после очередного прохода алгоритма ГА на вновь созданной популяции не будут получены дублирующие решения, которые были сохранены ранее.

Вариант 2. Во втором варианте после достижения порога «насыщения» популяция заполняется новыми особями псевдослучайно. В этом случае каждому из найденных решений сопоставляется определенный ареал с заданным радиусом (подразумевается, что для данного ареала, который сопоставляется с решением, это решение оптимально, с максимальным значением фитнес-функции). Генерирование новых особей происходит с использованием правила, по которому вновь получающиеся особи не должны попадать в этот ареал. При этом возникает вопрос адекватности определения радиуса каждого из ареалов. Данный подход (как и предыдущий) не дает стопроцентной гарантии того, что после работы ГА не будет дублирующих решений, он всего лишь существенно уменьшает вероятность их появления.

Вариант 3. Наиболее разумным кажется способ, при котором найденные решения участвуют в процессе работы апоптоза ГА после «насыщения» популяции, но не участвуют в кроссинговере. Для этого, когда для кроссинговера выбраны две новые особи из популяции, следует сравнивать их не только между собой, но и со всеми найденными ранее решениями, которые были сохранены после достижения «насыщения». Если одна из вновь выбранных особей идентична какому-либо из решений или отстоит от него на заданном минимальном расстоянии (в случае если известен ареал обитания, как в предыдущем варианте), то данную особь следует заменить полностью новой. Это необходимо для того, чтобы с высокой вероятностью особь, являющаяся либо уже одним из найденных решений, либо очень близкой к нему, не превратилась в одного из «лидеров» и не заполнила популяцию своими клонами. Следует заметить, что данная процедура замедлит работу ГА после прохождения первого этапа обнаружения порога «насыщения», так как теперь после выбора двух родителей придется сравнивать их не только между собой, но и со всеми найденными ранее решениями. А чем больше будет найдено решений, тем медленнее будет происходить процесс кроссинговера.

Вариант 4. Еще одним вариантом использования найденной зависимости может быть определение приблизительного числа эпох, за которое ГА доходит до «насыщения». Определив это число, можно оценить количество эпох (например, среднее значение количества эпох, за которые популяция доходила до «насыщения»), при котором ГА не достигает «насыщения», но при этом получаются достаточно точные результаты, схожие с показанными для экспериментов с высокой вероятностью мутации. Грубо говоря, задача сводится к определению числа эпох в зависимости от параметров мутации и вероятности кроссинговера, за которое популяция придет в «насыщенное» состояние. Например, обозначим T найденное число эпох. После этого выбирается такое число эпох, при котором популяция не успеет дойти до «насыщения», но особи при этом уже будут располагаться на местах искомым решений. Естественно, выбранное число эпох будет меньше, чем T , но находиться оно должно вблизи него. Варьируя расстояние конечного числа эпох до «точки насыщения», можно регулировать точность найденных решений.

Описанные выше варианты предполагают, что популяция насыщена в случае отношения размера популяции и количества дубликатов особей равного или пре-

вышающего 0,25. Однако можно принять данный порог значительно меньшим, тем самым внося разнообразие популяции намного чаще, но и с меньшей точностью решений.

Моделирование. Для сравнения эффективности работы предложенной модификации с классическим ГА был проведен ряд экспериментов по нахождению экстремумов поверхностей мультимодальных функций.

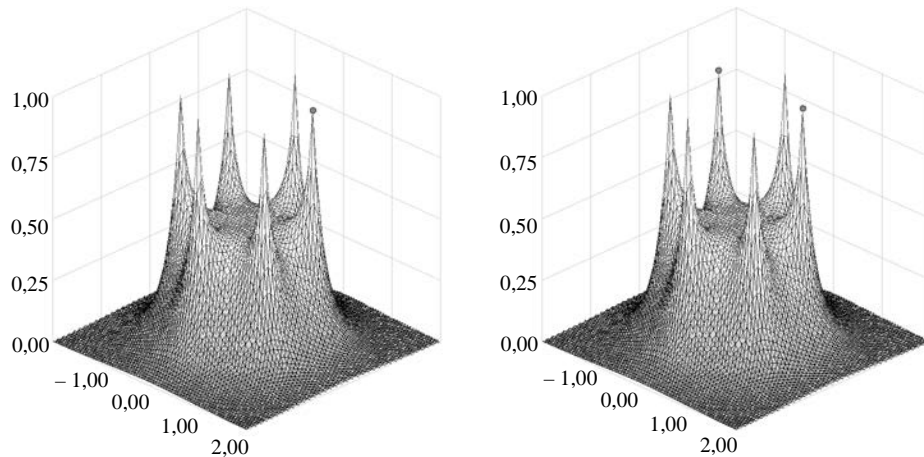
Для моделирования использовались следующие функции:

$$g(z) = \frac{1}{1 + |z^6 - 1|}, \quad z \in C, \quad z = x + iy; \quad x, y \in [-2, 2] \quad (\text{рис. 3, а}); \quad (1)$$

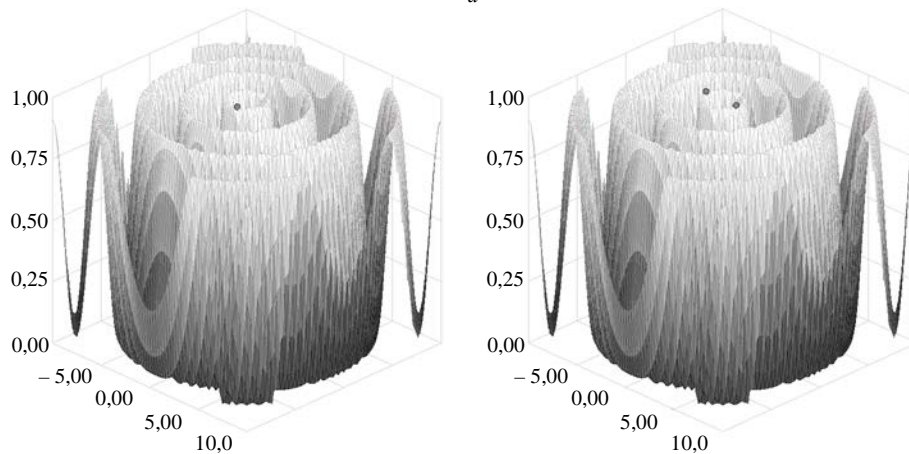
$$g(z) = 0,5 + \frac{\sin^2(\sqrt{x^2 + y^2}) - 0,5}{(1 + 0,001(x^2 + y^2))}, \quad x, y \in [-10, 10] \quad (\text{рис. 3, б}); \quad (2)$$

$$g(x, y) = x \sin(4\pi x) - y \sin(4\pi y + \pi) + 1, \quad x, y \in [-2, 2] \quad (\text{рис. 3, в}). \quad (3)$$

Эксперимент 1. В этом эксперименте сравнивалась работа классического алгоритма и предложенной модификации ГА с апоптозом. При моделировании были выбраны следующие параметры: вероятность мутации 0,003, вероятность кроссинговера 0,7, число эпох 80, начальный размер популяции 100, максимальный 200. Результаты эксперимента приведены на рис. 3. Здесь рисунки слева соответствуют классическому ГА, а справа — предложенной модификации.



а



б

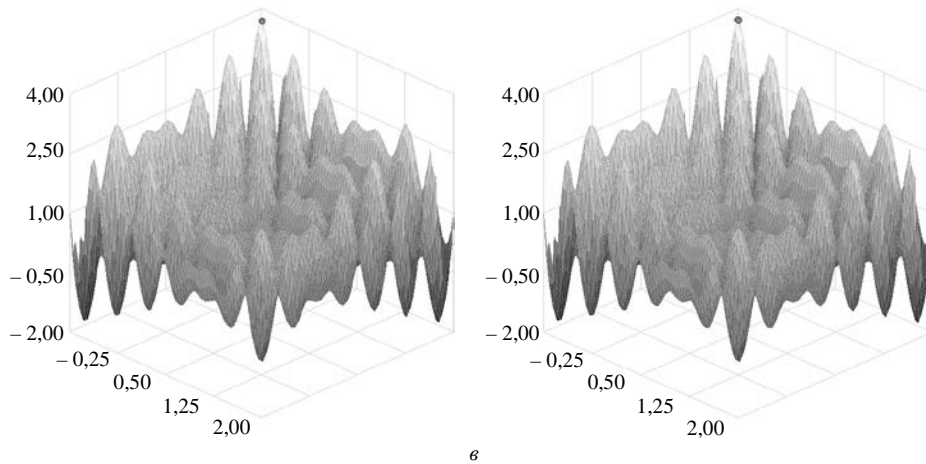


Рис. 3

Из экспериментов следует, что при стандартных параметрах ГА предложенная модификация хотя и дает лучшие результаты, она лишь частично справляется с поставленной задачей, так как находит не все решения, а как правило, два из них. Следует добавить, что для поверхностей, у которых есть только один ярко выраженный глобальный экстремум, оба решения иногда могут быть практически видны как одно (например, рис. 3, в), поскольку оба могут лежать крайне близко один к другому.

Эксперимент 2. Был проведен эксперимент с одной из известных модификаций ГА, описанной в [5], в которой используются масштабирование фитнес-функции и распределение ее в определенном ареале, зависящем от разрозненности популяции. В этом случае функция приспособленности выглядит следующим образом:

$$f_s(x_i) = \frac{f(x_i)}{\sum_{j=1}^N sh(d(x_i, x_j))}, \quad (4)$$

где $f(x_i)$ — стандартная фитнес-функция для i -й особи; $sh(d)$ — функция соседства

$$sh(d) = \begin{cases} 1 - \left(\frac{d}{\sigma_{sh}}\right)^\alpha, & d < \sigma_{sh}, \\ 0, & d \geq \sigma_{sh} \end{cases} \quad (5)$$

α — константа, регулирующая форму соседства (наиболее часто $\alpha = 1$); σ_{sh} — заданный порог разрозненности особей (в данной работе взято значение 0,003). Функция $d = d(x_i, x_j)$ определяет расстояние i -й особи до j -го индивидуума. Выбор наиболее эффективной функции расстояния между особями является отдельной задачей (в данных экспериментах выбрано евклидово расстояние). Очевидно, что данная модификация существенно увеличивает время работы ГА, так как количество вычислений в одной эпохе ГА значительно больше по сравнению с классическим ГА.

На рис. 4, а–в приведены результаты работы описанного выше модифицированного ГА без апоптоза (рисунки слева) для указанных ранее поверхностей и с применением модификации с апоптозом (рисунки справа).

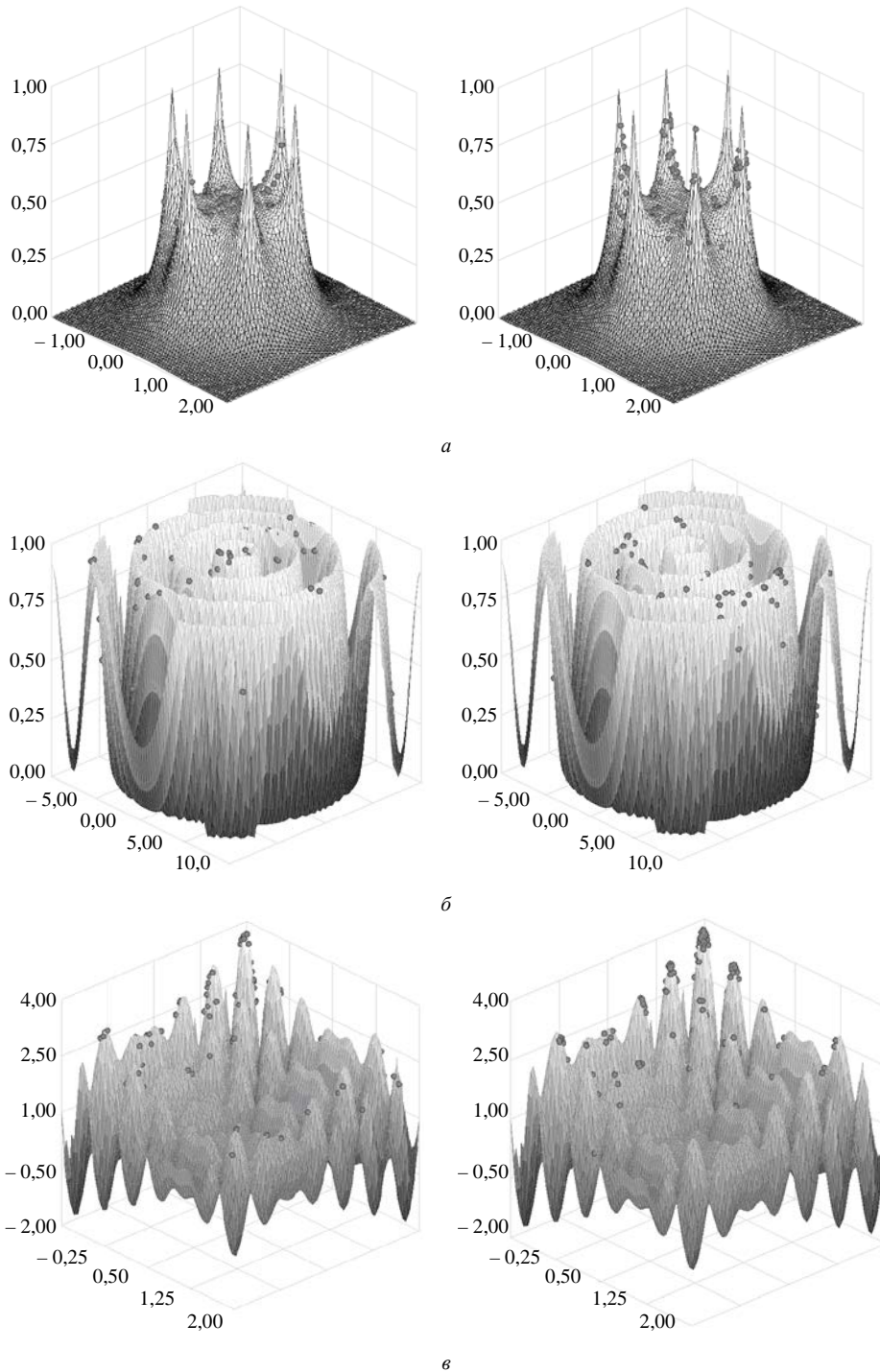


Рис. 4

Эксперимент 3. Исследовалась работа различных рассмотренных выше вариантов использования порога «насыщения». На рис. 5 отображены результаты моделирования первого и четвертого предложенных вариантов. Для обоих вариантов использовались одинаковые параметры мутации и кроссинговера (0,003 и 0,7 соответственно), начальная популяция состояла из 100 индивидуумов, конечная — из 200, число эпох равно 80. При реализации четвертого варианта было принято следующее правило для определения нужного количества эпох, для остановки ГА: число эпох для остановки равно 70 % от числа, которое потребовалось для достижения «насыщения», т.е. если задано максимальное число эпох, равное 80, а реально ГА дошел до насыще-

ния за 25 (это приблизительное число, за которое доходит ГА до «насыщения» в описанных экспериментах), то всего будет пройдено 43 эпохи, что практически в два раза меньше ожидаемого числа. Отметим, что отношение в 70 % взято произвольно (две третьих от числа, необходимого для «насыщения») и должно быть вычислено по ходу работы ГА на основе отношения числа дубликатов особей к общему числу особей, участвовавших в кроссинговере, для каждой из эпох.

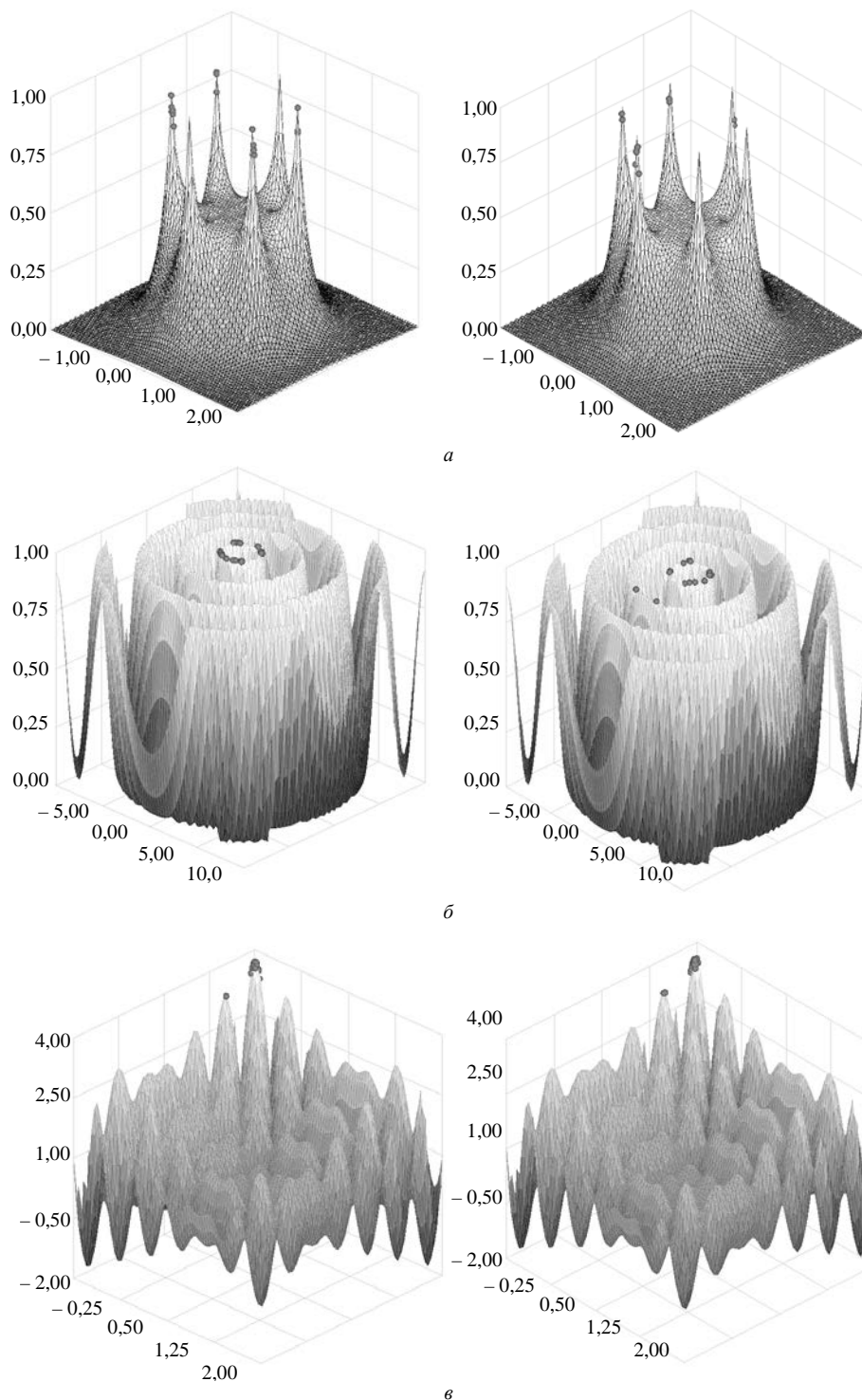


Рис. 5

Эксперимент 4. В данном эксперименте показан сравнительный анализ предложенной модификации ГА и известной реализации иммунного алгоритма — алгоритма клонального отбора (CLONALG) [6]. Исследовалось также влияние величины порога «насыщенности» популяции на свойства получаемых решений. Как было сказано ранее, порогом, при котором следует выполнить один из предложенных вариантов, может быть число, гораздо меньшее 0,25. При этом естественно разброс решений будет значительно шире, но и количество их будет существенно больше. На рис. 6 приведены результаты для случая выбора порога «насыщения», равного 0,04 (для лучшего сравнения результаты отображены проекциями на плоскость входных данных).

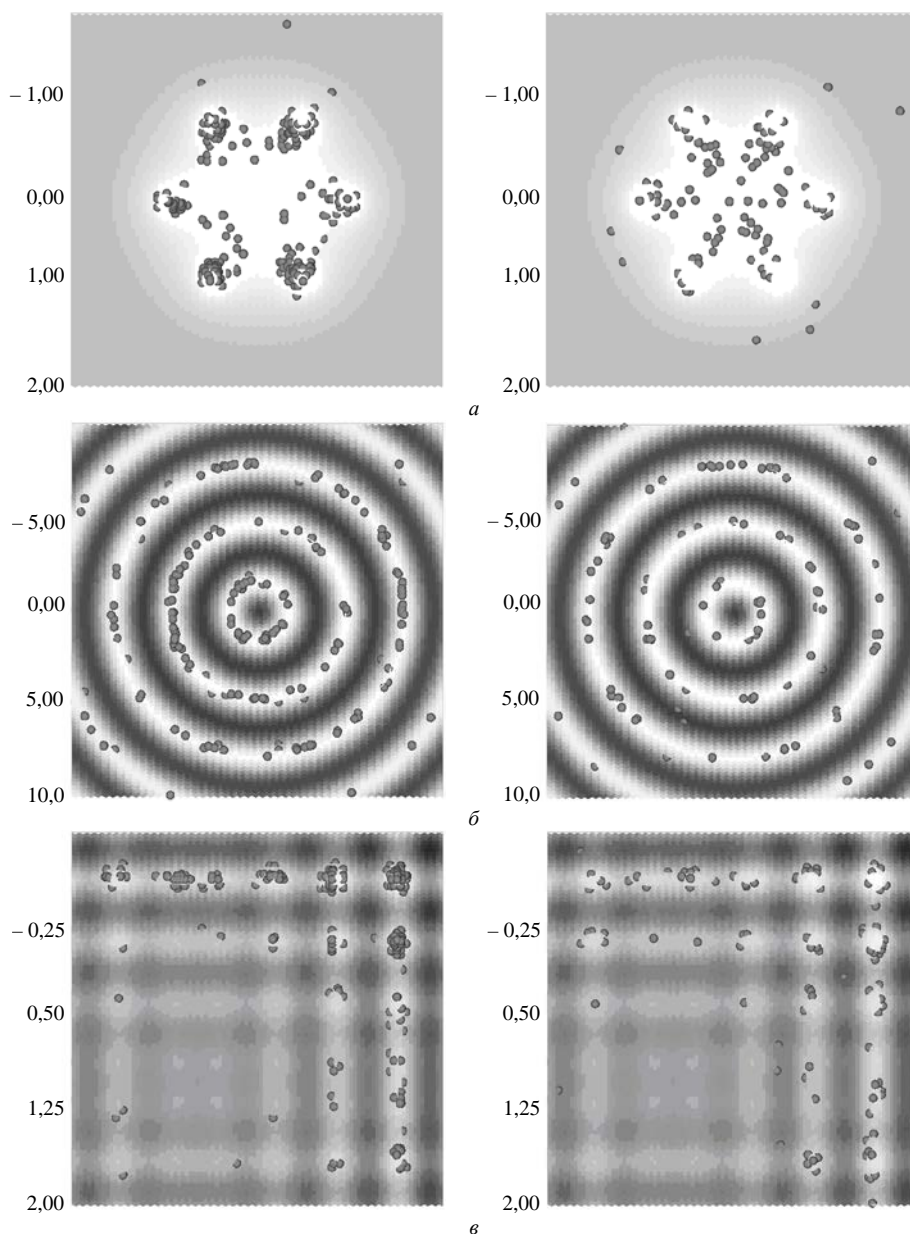


Рис. 6

Заключение. Экспериментальное исследование предложенных модификаций ГА подтвердило их работоспособность и целесообразность использования для нахождения экстремумов мультимодальных функций. Данные модификации, как

и традиционный ГА, требуют корректного определения числа эпох, однако в отличие от классического ГА, при незаданном или неизвестном числе эпох механизм апоптоза может достаточно просто определить их количество без просмотра всей популяции, а лишь используя данные, полученные в ходе работы алгоритма. Более того, предложенный четвертый вариант обеспечивает определение необходимого количества эпох для поиска решений, что существенно упрощает вычислительную реализацию данных модифицированных алгоритмов.

Следует отметить, что параметр порога насыщения может выступать универсальным коэффициентом для регуляции точности и количества находимых решений, что существенно упрощает задачу применения данного алгоритма.

О.Г. Руденко, Р.В. Бобнев

ГІБРИДНИЙ ГЕНЕТИЧНИЙ АЛГОРИТМ НА ОСНОВІ БІОЛОГІЧНОГО АПОПТОЗУ

Розглянуто класичний генетичний алгоритм та основні проблеми, що виникають при його реалізації. Запропоновано модифікацію даного алгоритму, яка змінює спосіб формування нових пар нащадків на основі механізму біологічного апоптозу. Введено поняття насиченості популяції та запропоновано використовувати поріг насиченості для покращення роботи алгоритму. Наведено результати експериментальних досліджень, які підтверджують ефективність запропонованих модифікацій при знаходженні екстремумів мультимодальних функцій.

O.G. Rudenko, R.V. Bobniev

HYBRID GENETIC ALGORITHM BASED ON THE BIOLOGICAL APOPTOSIS

A classic genetic algorithm and the key issues associated with its implementation are considered. A modification of the classic genetic algorithm which changes the way of creating new offspring's pairs is presented. The modification is based on the biological apoptosis theory. The saturation population conception and the suggestion about usage of this conception as termination condition with the purpose of algorithm improving have been introduced. The experimental research results proving efficiency of suggested modifications in case of multimodal functions have been shown.

1. *Holland J.H.* Adaptation in natural and artificial systems. — Ann Arbor, MI : Univ. Michigan Press, 1975. — 183 p.
2. *Рутковская Д., Пилинский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы. — М. : Горячая линия – Телеком, 2006. — 452 с.
3. *Скобцов Ю.А.* Основы эволюционных вычислений. — Донецк : ДонНТУ, 2008. — 326 с.
4. *Гордеева А.В., Лабас Ю.А., Звягильская Р.А.* Апоптоз одноклеточных организмов: механизмы и эволюция // Биохимия. — 2004. — **69**, вып. 10. — С. 1301–1313.
5. *Goldberg D.E., Richardson J.* Genetic algorithms with sharing for multimodal function optimization // Proc. of the Second Int. Conf. on Genetic Algorithms, Mahwah, NJ, USA, Lawrence Erlbaum Associates, Inc. — 1987. — P. 41–49.
6. *De Castro L.N., Von Zuben F.J.* Learning and optimization using the clonal selection principle // IEEE Transact. on Evolut. Comput., Special Issue on Artificial Immune Systems. — 2002. — **6**, N 3. — P. 239–251.

Получено 19.01.2012

Статья представлена к публикации акад. НАН Украины А.В. Палагиным.