

А.А. Фефелов, В.И. Литвиненко, М.А. Таиф, И.А. Лурье

## Параметрическая идентификация S-системы с применением модифицированного алгоритма клонального отбора

Предложен метод реконструкции генных регуляторных сетей, который в сочетании с алгоритмом оптимизации (эволюционным или иммунным) позволяет повысить скорость и точность решения задачи параметрической идентификации S-системы. Изложена суть метода последовательной трансформации пространства решений, управляемой результатами работы алгоритма оптимизации.

**Ключевые слова:** генные регуляторные сети, параметрическая идентификация, экспрессия генов, обыкновенные дифференциальные уравнения, S-система, алгоритм клонального отбора, оптимизация пространства поиска, скорость сходимости алгоритма.

Запропоновано метод реконструкції генних регуляторних мереж, який в поєднанні з алгоритмом оптимізації (еволюційним або імунним) дозволяє підвищити швидкість і точність розв'язання задачі параметричної ідентифікації S-системи. Викладено суть методу послідовної трансформації простору розв'язань, керованої результатами роботи алгоритму оптимізації.

**Ключові слова:** генні регуляторні мережі, параметричну ідентифікацію, експресія генів, звичайні диференціальні рівняння, S-система, алгоритм клонального відбору, оптимізація простору пошуку, швидкість збіжності алгоритму.

**Введение.** Современный мир характеризуется колоссальной сложностью систем, наполняющих его. Одним из примеров подобных систем есть биологические организмы. Моделирование биологических систем и их взаимодействий имеет ключевое значение для понимания механизмов их функционирования, что подтверждается результатами многочисленных исследований, которые особенно интенсивно проводятся в последние несколько десятилетий. Разработано множество инструментов и технологий, позволяющих строить и исследовать модели биологических систем и процессов. Одним из фундаментальных направлений современной науки, специализация которой заключается в расшифровке биологической информации на генетическом уровне, – молекулярная биология. Используя профили экспрессии генов, получаемые при помощи ДНК-микрочипов [1], выявляются совокупности генов, участвующих в том или ином биологическом процессе. Взаимодействия между различными генами или их группами в рамках одного процесса моделируется при помощи генной регуляторной сети (ГРС). Формально ГРС представляют в виде взвешенного ориентированного графа, вершины которого соответствуют генам сети, а дуги определяют связи между этими генами. ГРС находят многочисленные применения в медицине и биологии, среди которых можно выделить: обнаружение генети-

ческих заболеваний, разработку новых лекарственных препаратов, уменьшение побочных эффектов от лечения, формирование гипотез о функциональном назначении неизвестных генов [2, 3] и т.д. Однако на данном этапе нерешенной остается проблема создания новых быстрых и точных методов идентификации моделей, позволяющих использовать данные экспрессии для реконструкции архитектуры и поведения ГРС.

### Постановка проблемы

Проблема реконструкции ГРС включает в себя два основных аспекта: выбор модели генной сети и разработка метода идентификации выбранной модели. На данный момент для описания структуры и поведения ГРС предложен ряд математических моделей, среди которых наибольшее распространение получили: булевы сети [4], линейные модели [5], дифференциальные уравнения [6], ассоциативные сети [7], статические и динамические байесовские сети [8–10], нейронные сети [11], модели пространства состояний [12]. Системы обыкновенных дифференциальных уравнений (ОДУ) – наиболее точные модели, позволяющие максимально близко к реальности воспроизводить динамику ГРС. Именно по этой причине одна из таких моделей выбрана в данной статье. Чаще всего систему ОДУ представляют в форме S-системы [13], которая, с одной стороны, есть нелинейной моделью и поэтому достаточ-

но точно описывает поведение генной сети, а с другой, обладая характерной структурой, дает возможность легко реконструировать топологию ГРС. В общем виде  $S$ -система выглядит следующим образом:

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^N x_j^{g_{ij}} - \beta_i \prod_{j=1}^N x_j^{h_{ij}}, \quad (1)$$

где  $x_i(t)$  – переменная состояния, выражающая концентрацию продукта экспрессии  $i$ -го гена в момент времени  $t$ ;  $N$  – число компонентов (генов) в сети; параметры  $g_{ij}$  и  $h_{ij}$  определяют характер и степень воздействия гена  $x_j$  на ген  $x_i$ ;  $\alpha_i, \beta_i$  – неотрицательные коэффициенты.

Уменьшаемое  $\alpha_i \prod_{j=1}^N x_j^{g_{ij}}$  в правой части равенства (1) соответствует совокупному воздействию элементов ГРС, приводящему к увеличению концентрации  $x_i$ , а вычитаемое  $\beta_i \prod_{j=1}^N x_j^{h_{ij}}$

отражает все воздействия, приводящие к уменьшению концентрации  $x_i$ . При увеличении концентрации положительные значения  $g_{ij}$  подразумевают активацию геном  $j$  гена  $i$  (усиливает продукцию гена), в то время как отрицательные значения  $g_{ij}$  свидетельствуют о подавляющем воздействии гена  $j$  на ген  $i$  (ослабляет продукцию гена). В фазе уменьшения концентрации эффекты усиления и ослабления деградации соответствуют положительным и отрицательным значениям параметра  $h_{ij}$ . Значения  $g_{ij} = 0$  ( $h_{ij} = 0$ ) говорят об отсутствии связи между генами  $i$  и  $j$ .

Основная трудность идентификации моделей в форме  $S$ -системы – высокая размерность задачи. Количество параметров ( $\alpha_i, \beta_i, g_{ij}, h_{ij}$ ), необходимых найти, определяется выражением  $2N(N+1)$ . Так как вычислительная сложность данной задачи составляет  $O(n^2)$ , ее практически невозможно решить аналитическими методами, особенно при большом количестве участвующих в ГРС генов. В связи с этим раз-

работка эффективного метода поиска оптимальных параметров модели актуальна.

### Эволюционные алгоритмы и искусственные иммунные системы

Эволюционные алгоритмы (ЭА), как и искусственные иммунные системы (ИИС), – мощные инструменты оптимизации, часто применяемые для решения задач реконструкции ГРС. ЭА – это вычислительная парадигма, в основе которой лежат методы, моделирующие процессы естественной эволюции живых существ. Модели ИИС основаны на результатах исследований теоретической иммунологии. Согласно этим исследованиям естественные иммунные механизмы высших существ обладают признаками, свойственными системам распознавания образов. Иммунные механизмы адаптивны и устойчивы к небольшим изменениям характеристик распознаваемых антигенов (вирусов и бактерий). Они способны развивать свои навыки распознавания чужеродных агентов. Несмотря на различие парадигм, ЭА и ИИС имеют много общего. И те, и другие используют популяцию индивидуумов (вариантов решений), которая из поколения в поколение подвергается воздействию эволюционных или иммунных операторов. В общем случае этими операторами есть мутация и селекция. Популяция решений, первоначально генерируемая случайным образом, под воздействием мутации и селекции, а также других специфических операторов в каждом новом поколении улучшает свои характеристики. Оценка популяции в количественном выражении проводится посредством присвоения каждому индивидууму значения пригодности для ЭА или аффинности для ИИС. На основании оценок индивидуумы могут покидать популяцию или переходить в новое поколение.

На данный момент для реконструкции ГРС предложено несколько разновидностей ЭА и ИИС. В работе [14] предложен комбинированный алгоритм, использующий генетическое программирование (ГП) и метод наименьших квадратов (МНК). В качестве модели ГРС выбрана общая форма системы ОДУ, а оценка индивидуумов проводится посредством средне-

квадратической невязки между наблюдаемыми и экспериментально полученными данными временных рядов экспрессии генов. В [15] предложен двухэтапный метод реконструкции ГРС, где на первом этапе с применением ГП проводится аппроксимация данных экспрессии с получением дифференцируемых функций, а на втором этапе значения производных, вычисленные после дифференцирования функций, используются гибридным эволюционным алгоритмом для оценивания индивидуумов и поиска структуры и параметров модели ГРС, представленной в форме  $S$ -системы. Такой подход позволяет избавиться от сложных вычислений этапа интегрирования и повысить общую скорость расчетов. В работе [16] для реконструкции  $S$ -системы применен ЭА, называемый тригонометрической дифференциальной эволюцией (ТДЭ) [17]. Процесс реконструкции включает три фазы: структурной и параметрической идентификации и синхронизации. При этом каждая следующая фаза использует результаты, полученные в предыдущей фазе. В работе [18] дифференцированием временных рядов данных экспрессии  $S$ -система переводится в развязанную форму. В развязанной системе ОДУ каждое уравнение, входящее в ее состав, можно идентифицировать независимо от других, что сокращает размерность пространства поиска. Затем к каждому уравнению применяется метод ТДЭ для поиска его структуры и параметров. В работе [19] предложен метод реконструкции  $S$ -системы, основанный на многоцелевой оптимизации. Индивидуумы многоцелевого ЭА состоят из двух частей: бинарной и вещественной. Бинарная часть кодирует структуру ГРС, а именно: наличие или отсутствие связи между генами, а вещественная часть содержит значения параметров  $S$ -системы. Для каждой из частей сформулирована своя целевая функция. Результат работы алгоритма – множество Парето-оптимальных альтернатив, из которых с использованием процедуры автоматической селекции формируются окончательные решения. В работе [20] для оптимизации параметров  $S$ -системы использована ИИС в форме классического кло-

нального алгоритма, которая была улучшена в [21] путем гибридизации с одним из квазиньютоновских методов. В полученном гибриде ИИС выполняет функции глобального оптимизатора, в то время как квазиньютоновский метод занимается локальным поиском. В [22] клональный алгоритм применяется к линейризованной модели  $S$ -системы, где показатели степени заменены двумя весовыми матрицами: матрицей активации и подавления. Это позволяет уменьшить количество оптимизируемых параметров и увеличить скорость сходимости алгоритма путем исключения показательных зависимостей. В работе [23] авторы использовали гибридный метод клонального отбора и вейвлет-нейронной сети для реконструкции ГРС, представленной общей формой системы ОДУ. В данном случае нейронная сеть, как универсальный аппроксиматор, выбрана для описания функциональных зависимостей правой части системы ОДУ, а клональный алгоритм применяется для идентификации структуры и параметров нейронной сети.

В статье предлагается метод идентификации ГРС, отличающийся от рассмотренных ранее наличием итерационного процесса, в котором проводится оптимизация пространства поиска значений параметров вычислительной модели.

### Предлагаемое решение

Обозначим через  $v$  любой из параметров  $S$ -системы  $\alpha, \beta, g$  или  $h$ . Для поиска оптимального значения  $v$  эволюционным или иммунным алгоритмом необходимо, чтобы  $v$  был задан в промежутке  $\{v \in \mathfrak{R} : a \leq v \leq b\}$ . Концы промежутка  $a$  и  $b$  зависят от типа параметра и устанавливаются благодаря предварительным знаниям о предметной области перед началом процесса оптимизации. Так как рассматриваемые здесь типы поисковых алгоритмов принадлежат к инструментам комбинаторной оптимизации, множество значений  $V, v \in V$  конечно. При использовании бинарного кодирования индивидуумов мощность  $V$  определяется как  $|V| = 2^{m_v}$ , где  $m_v$  – длина сегмента строки индивидуума (количество бит), выделенного для

кодирования параметра  $v$ . Общее количество вариантов решений, среди которых алгоритм должен найти оптимальный, равно  $2^{m_v^{2N(N+1)}}$ . Очевидно, что уменьшение  $m_v$  позволяет существенно сократить вычислительную нагрузку и повысить производительность поискового алгоритма. Однако здесь следует учитывать тот факт, что все варианты решений размещаются в пространстве решений не как попало, а строго в узлах многомерной сетки, образованной значениями из множеств  $V_i, i = 1, \dots, 2N(N+1)$  параметров  $S$ -системы (рис. 1, *a*). Если глобальный оптимум  $P_{opt}$  не совпадает ни с одним из узлов этой сетки, то при данном  $m_v$  алгоритм поиска не в состоянии обнаружить оптимальное решение. Более того, если  $m_v$  достаточно мал, то возможные лучшие решения, такие как  $P_1$ , будут находиться слишком далеко от глобального оптимума и тем самым не будут удовлетворять требованиям к точности результата.

Суть этого предложения заключается в том, чтобы в процессе оптимизации, периодически, при достижении определенных условий трансформировать пространство поиска посредством пересчета промежутков возможных значений  $v$  с их одновременным центрированием относительно обнаруженного текущего лучшего решения (рис. 1, *б*). Как видно на представленном рисунке, после получения решения  $P_1$  концы промежутков возможного изменения параметров  $v_1$  и  $v_2$  были пересчитаны, что привело к трансформации пространства поиска. В результате, не изменяя значения  $m_v$ , авторы статьи создали более частую сетку, которая позволила на следующей итерации получить лучшее решение  $P_2$ , находящееся ближе к  $P_{opt}$ , чем решение  $P_1$ . Затем пространство трансформируется относительно решения  $P_2$ , и процесс повторяется до достижения требуемой точности результата.

Положительная особенность данного подхода – возможность работать с малыми значениями  $m_v$ , существенно уменьшая количество вариантов решений и ускоряя сходимость алгоритма поиска.

Трансформация предполагает не только сжатие пространства, как показано на рисунке 1, *б*, но и возможное расширение в том случае, если лучшее решение в очередной итерации оказалось на концах промежутка одного или нескольких параметров модели. При этом трансформация может включать одновременное сжатие пространства по одним измерениям и расширение по другим, что делает предложенный подход более гибким.

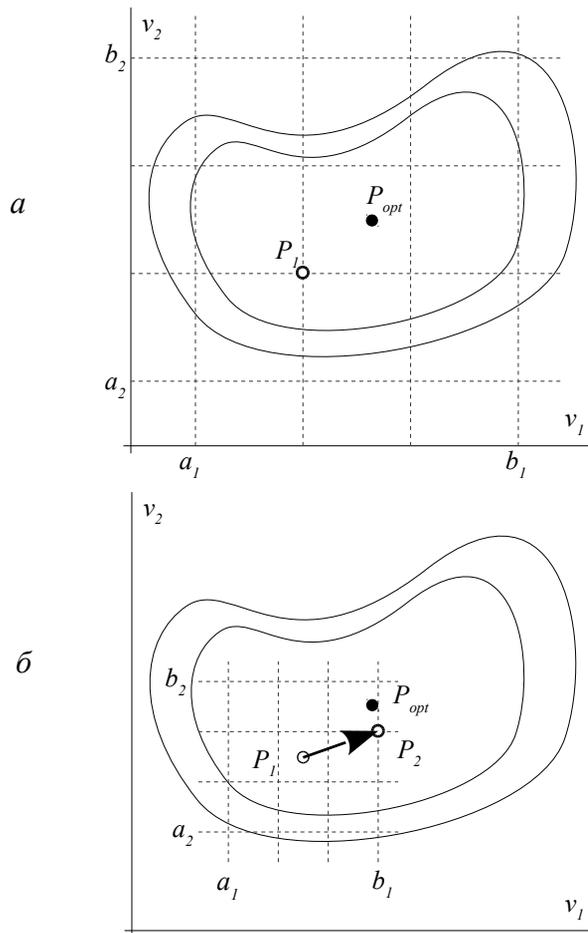


Рис. 1. *a* – первоначальное размещение вариантов решений в двумерном пространстве поиска; *б* – результат трансформации пространства поиска относительно решения  $P_1$

Положительная особенность данного подхода – возможность работать с малыми значениями  $m_v$ , существенно уменьшая количество вариантов решений и ускоряя сходимость алгоритма поиска. Трансформация предполагает не только сжатие пространства, как показано на рисунке 1, *б*, но и возможное расширение в том случае, если лучшее решение в очередной

итерации оказалось на концах промежутка одного или нескольких параметров модели. При этом трансформация может включать одновременное сжатие пространства по одним измерениям и расширение по другим, что делает предложенный подход более гибким.

Формально процесс трансформации пространства решений можно представить следующим образом:

$$\begin{aligned} a^{j+1} &= v_p^j - \frac{s^{j+1}(b^0 - a^0)}{2}; \\ b^{j+1} &= v_p^j + \frac{s^{j+1}(b^0 - a^0)}{2}, \end{aligned} \quad (2)$$

где  $a^{j+1}, b^{j+1}$  – концы промежутка значений параметра  $v$  для следующей трансформации;  $a^0, b^0$  – начальный промежуток значений параметра  $v$   $S$ -системы;  $v_p^j$  – лучшее значение параметра  $v$   $S$ -системы, полученное посредством алгоритма оптимизации в текущей ( $j$ -й) итерации;  $s^{j+1}$  – коэффициент масштабирования промежутка значений параметра  $v$  для следующей трансформации, который рассчитывается по формуле:

$$s^{j+1} = \begin{cases} s^j k_g, & \text{если } (v_p^j - a^j) \leq \varepsilon \text{ или } (b^j - v_p^j) \leq \varepsilon \\ s^j k_s, & \text{в противном случае} \end{cases}, \quad (3)$$

где  $s^j$  – коэффициент масштабирования промежутка значений параметра  $v$  в текущей ( $j$ -й) итерации;  $k_g$  ( $k_g > 1$ ) – коэффициент расширения промежутка;  $k_s$  ( $0 < k_s < 1$ ) – коэффициент сжатия промежутка;  $a^j, b^j$  – концы промежутка значений параметра  $v$  в текущей ( $j$ -й) итерации;  $\varepsilon$  – пороговое значение, которое фиксирует факт совпадения параметра  $v$  с левым или правым концом промежутка.

Следует отметить, что выражения (2) и (3) справедливы как для бинарного, так и для вещественного способов кодирования индивидуумов поискового алгоритма.

На рис. 2 представлена блок-схема, демонстрирующая работу предложенного метода. В данной схеме  $v_p$  подразумевает лучшее зна-

чение параметра  $v$   $S$ -системы, полученное за все время работы алгоритма.

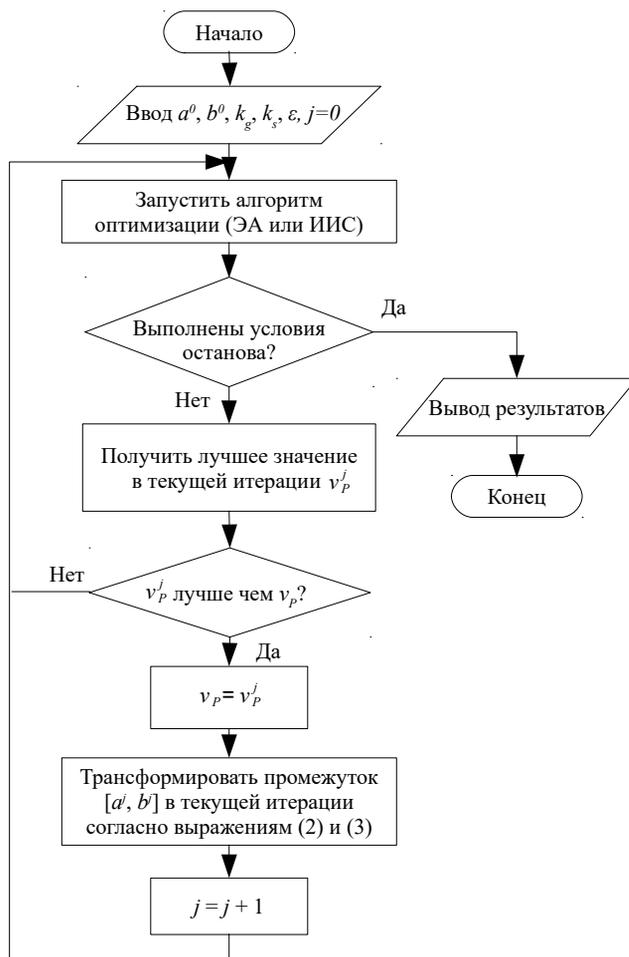


Рис. 2. Блок-схема предложенного алгоритма реконструкции ГРС

Работу предложенного метода и алгоритма можно описать следующим образом. После однократного пуска эволюционного или иммунного алгоритма оптимизации с установленным количеством поколений, лучшее решение, полученное в результате этого пуска, сравнивается с лучшим решением, полученным за все предыдущие пуски. Сравнение проводится на основании значений ошибки аппроксимации наблюдаемых данных экспрессии генов. Если новое решение оказывается лучше всех предыдущих, проводится трансформация пространства решений по формулам (2) и (3). Далее происходит повторный пуск алгоритма оптимизации, и процесс повторяется до тех пор, пока не будут выполнены условия останова (не будет достигнуто необходимое количество итераций  $j$  или некото-

рое минимальное значение ошибки). Таким образом в каждой следующей итерации  $j$  ЭА или ИИС работают с новым пространством решений, которое постепенно сжимается в окрестности глобального оптимума.

В качестве метода оптимизации в статье использована ИИС, представленная в форме алгоритма клонального отбора [24]. Блок-схема работы данного алгоритма показана на рис. 3.

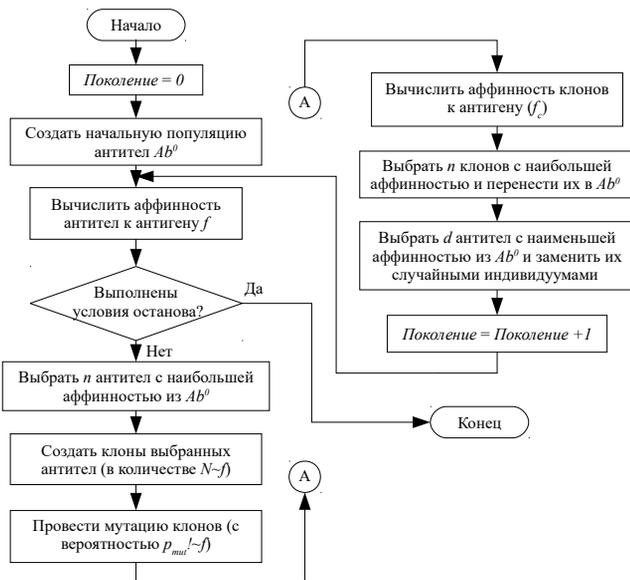


Рис. 3. Блок-схема алгоритма клональной селекции

В алгоритме клонального отбора поиск оптимального индивидуума заключается в последовательной адаптации популяции антител – вариантов решения задачи к входящему антигену – целевой функции. Во время адаптации популяция антител подвергается ряду циклических воздействий, таких как селекция, клонирование, мутация и повторная селекция. Одним из параметров работы каждой из фаз служит мера аффинности, выражающая степень близости индивидуума к оптимальному решению. Для поддержки разнообразия популяции в конце каждого поколения в ней создается некоторое количество новых случайных антител.

При решении задачи реконструкции  $S$ -системы кодирование индивидуумов ИИС проводится по схеме, изображенной на рис. 4. В дан-

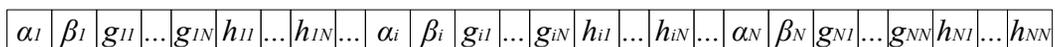


Рис. 4. Структура антитела, кодирующего  $S$ -систему

ной схеме индексирование параметров совпадает с тем, которое используется в выражении (1).

Для получения меры аффинности антител к антигену вычисляется ошибка модели на временных рядах данных экспрессии генов. В данной статье для расчета ошибки применяется следующее выражение [25]:

$$f = \text{MIN} \sum_{i=1}^N \sum_{j=1}^T \left( \frac{x_i^M(t_0 + j\Delta t) - x_i(t_0 + j\Delta t)}{x_i(t_0 + j\Delta t)} \right)^2, \quad (4)$$

где  $t_0$  – начальное время;  $\Delta t$  – временной шаг;  $T$  – количество данных временного ряда экспрессии;  $x_i^M(t_0 + j\Delta t)$  – значения концентрации, полученные решением системы ОДУ (1);  $x_i(t_0 + j\Delta t)$  – наблюдаемые значения концентрации из временного ряда экспрессии.

Так как система (1) не имеет аналитического решения, ее решают одним из методов численного интегрирования, в нашем случае методом Рунге–Кутты четвертого порядка. Значения  $x_i^M$  для каждого индивидуума популяции ИИС получаются решением задачи Коши с начальными условиями  $x_i(t_0)$ .

## Исследования

Для проверки эффективности предлагаемого подхода авторы выбрали две задачи на идентификацию параметров  $S$ -системы. В обеих используются искусственные ГРС с фиксированной структурой. Это означает, что нам известны связи между генами ГРС, но неизвестны веса этих связей. Первая задача представлена в [25] и содержит ГРС, состоящую из двух генов. Во второй задаче [15] регуляторная сеть содержит четыре гена. В обоих случаях сравнили результаты реконструкции ГРС с работой классического клонального алгоритма, описанного в [24].

## Реконструкция ГРС, состоящей из двух генов

Параметры  $S$ -системы для двухгенной ГРС ( $N = 2$ ) представлены в табл. 1. Динамика дан-

ной сети, рассчитанная с начальными условиями  $(x_1(t_0), x_2(t_0)) = (1,0; 1,5)$ , показана на рис. 5.

**Таблица 1.** Параметры  $S$ -системы для ГРС, состоящей из двух генов

$i$	$\alpha_i$	$\beta_i$	$g_{i1}$	$g_{i2}$	$h_{i1}$	$h_{i2}$
1	3,0	3,0	0,0	-2,5	-1,0	0,0
2	3,0	3,0	2,5	0,0	0,0	2,0

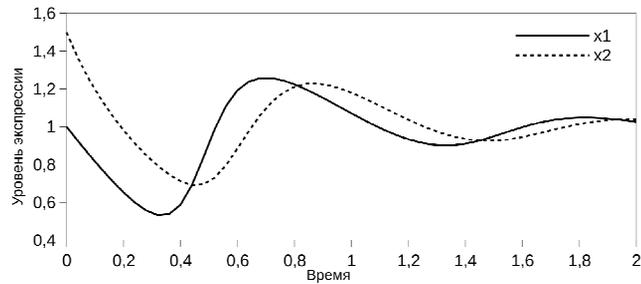


Рис. 5. Временной ряд данных экспрессии для двухгеновой ГРС

Для построения таблицы наблюдаемых данных экспрессии, необходимых для проведения экспериментов, система (1) решена с параметрами, взятыми из табл. 1. Решение  $S$ -системы осуществлялось методом Рунге–Кутты четвертого порядка с количеством временных отсчетов  $T = 50$ .

Параметры алгоритма реконструкции, используемые в данном эксперименте, представлены в табл. 2.

**Таблица 2.** Параметры алгоритма реконструкции для первого эксперимента

Название параметра	Значение параметра
Начальные промежутки значений для параметров $\alpha$ и $\beta$	[0,0; 5,0]
Начальные промежутки значений для параметров $g$ и $h$	[-3,0; 3,0]
Способ кодирования индивидуумов	бинарный
Точность представления параметров ( $m_v$ )	8 бит
Размер основной популяции	300
Размер популяции клонов	900
Количество поколений	20
Коэффициент отбора	0,7 ( $n = 210$ )
Тип отбора	турнирный
Размер турнира	4
Уровень мутации	0,01
Коэффициент расширения промежутка ( $k_g$ )	1,5
Коэффициент сжатия промежутка ( $k_s$ )	0,8
Пороговое значение $\varepsilon$	0,0
Условие останова (количество трансформаций пространства)	50

В результате эксперимента получена  $S$ -система с параметрами, представленными в табл. 3. При этом ошибка модели, рассчитанная по выражению (4), составила  $4,275 \cdot 10^{-7}$ .

**Таблица 3.** Полученные параметры  $S$ -системы для двухгеновой ГРС

$i$	$\alpha_i$	$\beta_i$	$g_{i1}$	$g_{i2}$	$h_{i1}$	$h_{i2}$
1	2,994	2,994	0,0	-2,505	-1,002	0,0
2	2,998	2,999	2,501	0,0	0,0	2,001

Как видно из таблицы, вычисленные значения параметров практически идентичны параметрам оригинальной системы (табл. 1). Это лучший результат, полученный в течение серии испытаний. С целью проверки устойчивости предложенного подхода авторы провели десять испытаний алгоритма с последующим усреднением результатов, а также сравнили средний результат с аналогичными расчетами, выполненными при помощи классического алгоритма клонального отбора, настройки которого отличаются от представленных в табл. 2 двумя значениями: точностью представления параметров (в данном случае  $m_v = 16$  бит) и количеством поколений (6200 поколений). Данные отличия связаны с тем, что, во-первых, классический клональный отбор не предполагает трансформации пространства решений, и, следовательно, требует более высокой точности представления значений. Во-вторых, из блок-схемы, изображенной на рис. 2, следует, что трансформация пространства происходит только после тех пусков алгоритма оптимизации, в результате которых появляется новое лучшее решение. Выбрали максимальное количество пусков из числа наблюдаемых в серии из десяти испытаний, и умножили это значение на 20 поколений, получив число 6200 поколений.

Для сравнения средних показателей вычисленных параметров  $S$ -системы рассчитаны их стандартные отклонения. Результаты расчетов представлены в табл. 4, где видно, что предлагаемый подход позволил значительно увеличить точность идентификации параметров  $S$ -системы данной конфигурации в сравнении с классическим клональным алгоритмом.

**Таблица 4.** Сравнительная таблица стандартных отклонений параметров  $S$ -системы, полученных для серии из десяти испытаний в первом эксперименте

$i$	$S(\alpha_i)$	$S(\beta_i)$	$S(g_{i1})$	$S(g_{i2})$	$S(h_{i1})$	$S(h_{i2})$
Предлагаемый метод						
1	0,27	0,25	–	0,18	0,06	–
2	0,09	0,07	0,05	–	–	0,04
Классический клональный алгоритм						
1	0,99	1,01	–	0,42	0,49	–
2	0,67	0,59	0,53	–	–	0,57

В табл. 5 представлены сравнительные данные ошибок модели для серии из десяти испытаний. Для расчета средней ошибки использовалась мера среднего арифметического.

График сходимости предложенного алгоритма в испытании, продемонстрировавшем наилучший результат, показан на рис. 6.

Испытания проводились с использованием аппаратного обеспечения *Intel Core i3-3110M* (2,4 ГГц). При этом было задействовано только одно ядро процессора. Среднее время работы предложенного алгоритма в одном испытании составило 24,5 минуты. Классический клональный алгоритм работал в среднем 56 минут.

**Таблица 5.** Сравнительная таблица ошибок модели при аппроксимации временных рядов данных экспрессии генов в первом эксперименте

Ошибка модели ( $f$ )	Предлагаемый метод	Классический клональный алгоритм
Минимальная	$4,275 \cdot 10^{-7}$	0,0027
Максимальная	0,0035	0,541
Средняя	$6,0 \cdot 10^{-4}$	0,154

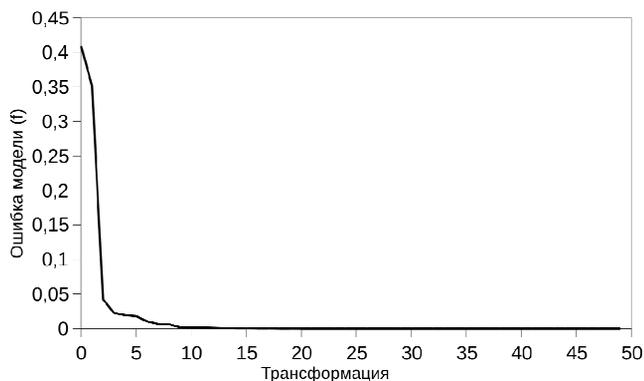


Рис. 6. График изменения ошибки модели в зависимости от трансформации пространства решений в первом эксперименте

### Реконструкция ГРС, состоящей из четырех генов

Параметры  $S$ -системы для регуляторной сети, состоящей из четырех генов ( $N = 4$ ), пред-

ставлены в табл. 6. Здесь, как и в предыдущем эксперименте, данные экспрессии генерировались решением системы (1) с параметрами, взятыми из табл. 6. Во втором эксперименте выбраны следующие начальные условия  $(x_1(t_0), x_2(t_0), x_3(t_0), x_4(t_0)) = (1,4; 2,7; 1,2; 0,4)$ . Количество данных временного ряда  $T = 26$ .

**Таблица 6.** Параметры  $S$ -системы для ГРС, состоящей из четырех генов

$i$	$\alpha_i$	$\beta_i$	$g_{i1}$	$g_{i2}$	$g_{i3}$	$g_{i4}$	$h_{i1}$	$h_{i2}$	$h_{i3}$	$h_{i4}$
1	12,0	10,0	0,0	0,0	-0,8	0,0	0,5	0,0	0,0	0,0
2	8,0	3,0	0,5	0,0	0,0	0,0	0,0	0,75	0,0	0,0
3	3,0	5,0	0,0	0,75	0,0	0,0	0,0	0,0	0,5	0,2
4	2,0	6,0	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,8

В данном случае большинство настроек алгоритма реконструкции сходны с указанными в табл. 2. Поэтому далее (табл. 7) приводятся настройки, чьи значения отличаются от установленных ранее.

**Таблица 7.** Изменившиеся параметры алгоритма реконструкции во втором эксперименте

Название параметра	Значение параметра
Начальные промежутки значений для параметров $\alpha$ и $\beta$	[0,0; 15,0]
Начальные промежутки значений для параметров $g$ и $h$	[-1,0; 1,0]
Точность представления параметров ( $m_i$ )	4 бит
Количество поколений	50

Проведена серия из десяти испытаний и выбран лучший результат ( $f = 0,00112$ ). График временного ряда экспрессии, полученный для лучшей модели, в сравнении с оригинальными данными, показан на рис. 7. Линии на графике находятся очень близко друг к другу, что делает их трудноразличимыми.

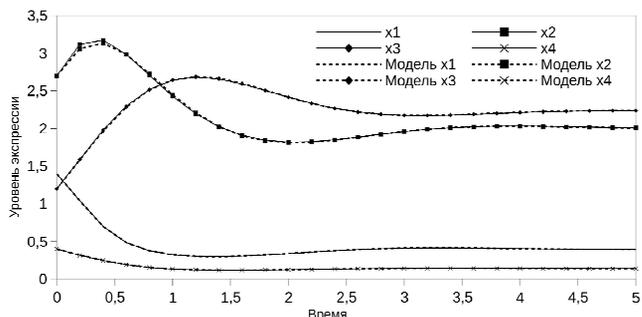


Рис. 7. График лучшей модели в серии из десяти испытаний во втором эксперименте

Параметры  $S$ -системы, соответствующие динамике, изображенной на рис. 7, представлены в табл. 8. Классический клональный алгоритм, с которым проводится сравнение результатов, настроен аналогично предыдущему эксперименту ( $m_v = 16$  бит, количество поколений – 6200).

**Таблица 8.** Параметры лучшей модели в серии из десяти испытаний во втором эксперименте

$I$	$\alpha_i$	$\beta_i$	$g_{i1}$	$g_{i2}$	$g_{i3}$	$g_{i4}$	$h_{i1}$	$h_{i2}$	$h_{i3}$	$h_{i4}$
1	13,8	12,1	0,0	0,0	-0,62	0,0	0,4	0,0	0,0	0,0
2	14,4	9,4	0,22	0,0	0,0	0,0	0,0	0,32	0,0	0,0
3	3,9	5,8	0,0	0,62	0,0	0,0	0,0	0,0	0,39	0,13
4	1,8	6,7	0,63	0,0	0,0	0,0	0,0	0,0	0,0	0,98

Для сравнения средних показателей вычисленных параметров  $S$ -системы рассчитаны их стандартные отклонения. Результаты расчетов представлены в табл. 9. В данном случае предлагаемый метод также демонстрирует преимущества перед классическим клональным алгоритмом, хотя они не столь существенны, как в первом эксперименте. Есть два параметра ( $\alpha_4$  и  $h_{34}$ ), у которых стандартное отклонение в испытаниях с классическим клональным отбором оказалось меньше, чем в испытаниях с разработанным методом.

**Таблица 9.** Сравнительная таблица стандартных отклонений параметров  $S$ -системы, полученных для серии из десяти испытаний во втором эксперименте

$I$	$S(\alpha_i)$	$S(\beta_i)$	$S(g_{i1})$	$S(g_{i2})$	$S(g_{i3})$	$S(g_{i4})$	$S(h_{i1})$	$S(h_{i2})$	$S(h_{i3})$	$S(h_{i4})$
Предлагаемый метод										
1	2,91	2,48	–	–	0,18	–	0,14	–	–	–
2	3,26	3,05	0,15	–	–	–	–	0,24	–	–
3	2,6	3,07	–	0,25	–	–	–	–	0,33	0,38
4	2,79	2,21	0,19	–	–	–	–	–	–	0,29
Классический клональный алгоритм										
1	6,05	3,4	–	–	0,55	–	0,21	–	–	–
2	3,35	4,73	0,43	–	–	–	–	0,35	–	–
3	5,1	5,64	–	0,64	–	–	–	–	0,36	0,3
4	2,18	3,1	0,4	–	–	–	–	–	–	0,4

В таблице 10 представлены сравнительные данные ошибок модели для серии из десяти испытаний. Для расчета средней ошибки использовалась мера среднего арифметического.

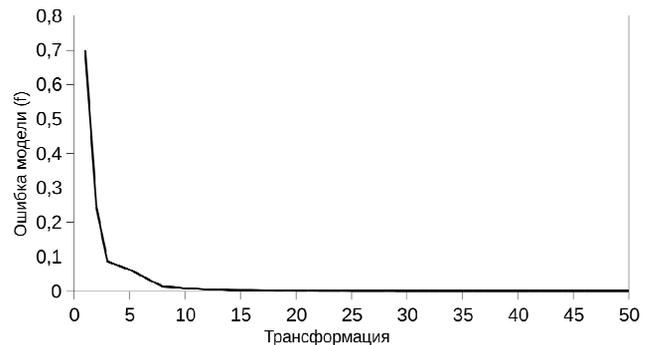
Судя по таблице видно, что во всех случаях ошибки модели демонстрируют преимущества

предложенного подхода. График сходимости предложенного алгоритма в испытании, продемонстрировавшем наилучший результат, показан на рис. 8.

**Таблица 10.** Сравнительная таблица ошибок модели при аппроксимации временных рядов данных экспрессии генов во втором эксперименте

Ошибка модели ( $f$ )	Предлагаемый метод	Классический клональный алгоритм
Минимальная	0,00112	0,03
Максимальная	0,0089	1,021
Средняя	0,00466	0,6

Как и в первом эксперименте, испытания проводились с использованием аппаратного обеспечения *Intel Core i3-3110M* (2.4 ГГц). При этом было задействовано только одно ядро процессора. Среднее время работы предложенного алгоритма в одном испытании составило 45 минут. Классический клональный алгоритм работал в среднем 1,5 часа.



**Рис. 8.** График изменения ошибки модели в зависимости от трансформации пространства решений во втором эксперименте

Для проведения экспериментов использовалась информационная система реконструкции ГРС, объектно-ориентированная архитектура которой представлена в [26].

**Заключение.** Метод реконструкции ГРС основан на многократной трансформации пространства решений, при которой пространство сжимается в окрестности глобального оптимума. Данный подход может быть использован для модификации ЭА или ИИС. В статье трансформация пространства решений применяется в сочетании с искусственной иммунной системой, представленной в форме алгоритма клональной селекции. Для проверки эффективно-

сти предложенного подхода проведены эксперименты по реконструкции искусственных ГРС. Цель экспериментов – нахождение значенных параметров S-систем при известной структуре. Результаты экспериментов подтвердили преимущества разработанного метода и алгоритма перед классическим алгоритмом кластерного отбора.

Предложенный метод не избавляет алгоритм поиска от возможности попадания в локальные оптимумы, поэтому развитие способностей выбирать из них есть одно из приоритетных направлений дальнейших исследований. В дальнейшем планируется проверить эффективность метода на реальных биологических данных.

1. *DeRisi J.L., Lyer V.R., Brown P.O.* Exploring the metabolic and genetic control of gene expression on a genomic scale // *Science*. – 1997. – **278**. – P. 680–686.
2. *Rawool S.B., Venkatesh K.V.* Steady state approach to model gene regulatory networks – simulation of microarray experiments // *Biosystems*. – 2007. – **90**(3). – P. 636–655.
3. *Chen P.C.Y., Chen J.W.* A Markovian approach to the control of genetic regulatory networks // *Biosystems*. – 2007. – **90**(2). – P. 535–545.
4. *Liang S., Fuhrman S., Somogyi R.* Reveal, a general reverse engineering algorithm for inference of genetic network architectures // *Pacific Symp. on Biocomputing (PSB'98)*, Jan. 1998: proc. – Maui, Hawaii, USA, 1998. – P. 18–29.
5. *Linear modeling of mRNA expression levels during CNS development and injury / P. D'haeseleer, X. Wen, S. Fuhrman et al.* // *4th Pacific Symp. on Biocomputing (PSB'99)*, Jan. 1999: proc. – Maui, Hawaii, USA, 1999. – P. 41–52.
6. *Chen T., He H.L., Church G.M.* Modeling gene expression with differential equations // *4th Pacific Symp. on Biocomputing (PSB'99)*, Jan. 1999: proc. – Big Island of Hawaii, Hawaii, USA, 1999. – P. 29–40.
7. *Schafer J., Strimmer K.* An empirical Bayes approach to inferring large-scale gene association networks // *Bioinformatics*. – 2005. – **21**(6). – P. 754–764.
8. *Using Bayesian networks to analyze expression data / N. Friedman, M. Lital, I. Nachman et al.* // *J. of Computational Biology*. – 2000. – **7**(3–4). – P. 601–620.
9. *Advances to Bayesian network inference for generating causal networks from observational biological data / J. Yu, V.A. Smith, P.P. Wang et al.* // *Bioinformatics*. – 2004. – **20**(18). – P. 3594–3603.
10. *Werhli A.V., Grzegorzczak M., Husmeier D.* Comparative evaluation of reverse engineering gene regulatory

networks with relevance networks, graphical Gaussian models and Bayesian networks // *Bioinformatics*. – 2006. – **22**(20). – P. 2523–2531.

11. *Reconstruction and validation of gene regulatory networks with neural networks / H. Hache, C. Wierling, H. Lehrach et al.* // *2nd Foundations of Systems Biology in Engineering Conf. (FOSBE '07)*, Sept. 2007: proc. – Stuttgart, Germany, 2007. – P. 319–324.
12. *Modeling T-cell activation using gene expression profiling and state-space models / C. Rangel, J. Angus, Z. Ghahramani et al.* // *Bioinformatics*. – 2004. – **20**(9). – P. 1361–1372.
13. *Savageau M.A.* Introduction to S-systems and the underlying power-law formalism // *Mathematical and Computer Modelling*. – 1988. – **11**. – P. 546–551.
14. *Sakamoto E., Iba H.* Inferring a system of differential equations for a gene regulatory network by using genetic programming // *Congress on Evolutionary Computation, 27–30 May 2001: proc.* – COEX, Seoul, Korea, 2001. – P. 720–726.
15. *Kim K.-Y., Cho D.-Y., Zhang B.-T.* Multi-stage evolutionary algorithms for efficient identification of gene regulatory networks // *Applications of evolutionary computing: EvoWorkshops 2006*, Apr. 2006: proc. – Budapest, Hungary, 2006. – P. 45–56.
16. *Noman N., Iba H.* Inference of Gene Regulatory Networks Using S-system and Differential Evolution // *Genetic and Evolutionary Computation Conference (GECCO'05)*, June 2005: proc. – Washington DC, USA, 2005. – P. 439–446.
17. *Fan H.-Y., Lampinen J.* A trigonometric mutation operation to differential evolution // *J. of Global Optimization*. – 2003. – **27**(1). – P. 105–129.
18. *Noman N., Iba H.* Reverse engineering genetic networks using evolutionary computation // *Genome Informatics*. – 2005. – **16**(2). – P. 205–214.
19. *Chen Y., Chen D., Zou X.* Inference of biochemical S-systems via mixed-variable multiobjective evolutionary optimization // *Computational and mathematical methods in medicine*. – 2017. – **2017**. – ID 3020326. – 9 p.
20. *Jereesh A.S., Govindan V.K.* A clonal based algorithm for the reconstruction of genetic network using S-system // *Int. J. of Research in Engineering and Technology (IJRET)*. – 2013. – **2**. – P. 44–50.
21. *Jereesh A.S., Govindan V.K.* Clono-hybrid algorithm for the reconstruction of gene regulatory network using S-system // *Int. J. of Pure and Applied Bioscience*. – 2014. – **2**(6). – P. 241–248.
22. *Jereesh A.S., Govindan V.K.* Two weight matrix model based GNET reconstruction using clonal algorithm // *J. of Advanced Bioinformatics Applications and Research*. – 2014. – **5**(3). – P. 121–133.
23. *Hybrid Approach for Gene Regulatory Networks Reconstruction by a System of Ordinary Differential Equations / A.A. Fefelov, V.I. Lytvynenko, M.A. Taif et al.* – *Upr. sist. maš.* – 2017. – **3**. – P. 63–72. (In Russian).

24. *De Castro L.N., Zuben F.J. Von Learning and optimization using the clonal selection principle // IEEE Transactions on Evolutionary Computation.* – 2002. – 6(3). – P. 239–251.
25. *Tominaga D., Koga N., Okamoto M. Efficient numerical optimization algorithm based on genetic algorithm for inverse problem // Genetic and Evolutionary Computation Conf. (GECCO'00), July 2000: proc.* – Las Vegas, Nevada, USA, 2000. – 251. – P. 251–258.
26. *Object-Oriented Architecture of the Information System for the Reconstruction of the Gene Regulatory Networks / A.A. Fefelov, V.I. Lytvynenko, M.A. Taif et al.* – Upr. sist. maš. – 2017. – N 4. – P. 67–75, 82. (In Russian).

Поступила 03.10.2017

E-mail: fao1976@ukr.net, immun56@gmail.com,  
taifmohamedali@gmail.com, lurieira@gmail.com  
© А.А. Фефелов, В.И. Литвиненко, М.А. Таиф,  
И.А. Лурье, 2017

UDC 574:004.2

A.A. Fefelov<sup>1</sup>, V.I. Lytvynenko<sup>2</sup>, M.A. Taif<sup>3</sup>, I.A. Lurie<sup>4</sup>

<sup>1</sup> PhD in Techn. Sciences, Associate Professor, Department of Design of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, fao1976@ukr.net

<sup>2</sup> Doctor of Technical Sciences, Professor, Head of the Department of Informatics and Computer Science of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, immun56@gmail.com

<sup>3</sup> Graduate student of the Department of Informatics and Computer Science, Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, taifmohamedali@gmail.com

<sup>4</sup> PhD in Techn. Sciences, Associate Professor, Department of Informatics and Computer Science of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, lurieira@gmail.com

### Parametric Identification of the S-System by the Modified Clonal Selection Algorithm

**Keywords:** gene regulatory networks, reverse engineering, gene expression, ordinary differential equations, clonal selection algorithm, wavelet-neural network.

**Introduction.** Modeling the biological systems and their interactions has the key value for the mechanisms of the functioning understanding. Recently, a lot of tools and technologies have been developed. They allow us to build and study the models of the biological systems and processes. One of such tools is the reconstruction or reverse engineering of gene regulatory networks, which helps us identify the structural and dynamic properties of the system based on the observations of its behavior and certain knowledge in the relevant subject area. Using the gene expression profiles obtained by DNA microarray, gene sets involved in a certain biological process are revealed. However, at this stage of development, the use of expression for reconstructing architecture and behavior of regulatory networks remains unsolved.

**Purpose.** The aim of this work is to create an effective method of the optimal parameters of the mathematical model of a gene regulatory network searching based on the ordinary differential equations system represented in the form of S-system.

**Method.** A method is based on the successive transformation of decision space, guided by the results of the separate starting of clonal selection algorithm, hereupon space compresses in the vicinity of the global optimum.

**Results.** A method for reconstructing the gene regulatory networks based on a modified clonal selection algorithm is developed. The method uses time series data of the gene expression profiles for searching interconnections between GRN components. The efficiency of the proposed method is confirmed by the experimental studies.

**Conclusion.** The developed method and the algorithm increase the speed of the convergence of the optimization algorithms, and at the same time improve their accuracy in solving the problem of parametric identification of S-System. The proposed method can be used for modification of the evolutionary algorithms or artificial immune systems. Besides, in our future research we plan to test the method effectiveness on the real biological data.

