

ПРИМЕНЕНИЕ ПЛАНИРУЕМОГО ЭКСПЕРИМЕНТА С ЦЕЛЬЮ ИССЛЕДОВАНИЯ ВЛИЯНИЯ ПАРАМЕТРОВ ГЕНЕТИЧЕСКОГО АЛГОРИТМА НА ПРОЦЕСС ПОИСКА ГЛОБАЛЬНОГО ЭКСТРЕМУМА

А.Н. Савин, И.В. Дружинин, А.А. Ерофтиев

Введение

В настоящее время при разработке новых материалов и устройств часто требуется решать оптимизационные задачи, которые, как известно, сводятся к поиску экстремумов целевой функции (ЦФ) различными методами. Решение подобных задач возможно с помощью генетических алгоритмов [1].

При решении оптимизационных задач с помощью генетических алгоритмов исследователи сталкиваются с проблемой правильного выбора параметров генетического алгоритма. Соответственно, актуальным является исследование влияния и выявление оптимальных параметров, так как это может значительно повысить надежность и уменьшить время поиска.

Целью данной работы является моделирование с помощью средств планируемого численного эксперимента зависимостей скорости и надежности нахождения глобального экстремума посредством генетического алгоритма от его параметров.

Глобальная оптимизация с помощью генетического алгоритма

В задачах условной оптимизации с ограничениями минимум функции

$$f(x) = f(x_1, \dots, x_n) \rightarrow \min, x \in \Omega, \Omega \subset R^n \quad (1)$$

ищется на некотором собственном подмножестве Ω метрического пространства R^n , где x определяется явными ограничениями

$$l_j \leq x_j \leq u_j, j = 1, \dots, n, \quad (2)$$

Пусть имеется некоторое множество X , состоящее из элементов x , принадлежащих подмножеству $\Omega \subset R^n$, и на нем определена скалярная функция $f(x)$. Говорят, что $f(x)$ имеет локальный минимум в точке x^* , если существует некоторая конечная ε -окрестность этой точки, в которой выполняется условие:

$$f(x^*) < f(x), \|x - x^*\| \leq \varepsilon \quad (3)$$

У функции может быть много локальных минимумов. Если выполняется неравенство $f(x^*) < f(x)$, $x \in X$, где $x \neq x^*$ – любая точка множества X , то говорят о глобальном минимуме функции $f(x)$ [2].

Поиск глобального минимума функции (1) при наличии ограничений (2) на варьируемые параметры может быть осуществлён с помощью генетических алгоритмов [3].

ГА представляют собой модель процесса эволюции в природе. В них используются аналоги механизмов генетического наследования и естественного отбора.

Эволюционная теория утверждает, что каждый биологический вид целенаправленно развивается и изменяется для того, чтобы наилучшим образом приспособиться к окружающей среде. Основным механизмом эволюции – это естественный отбор. Его суть состоит в том, что более приспособленные особи имеют больше возможностей для выживания и размножения и, следовательно, приносят больше потомства, чем плохо приспособленные особи. При этом благодаря передаче генетической информации (генетическому наследованию) потомки наследуют от родителей основные их качества. Таким образом, потомки сильных индивидуумов также будут относительно хорошо приспособленными, а их доля в общей массе особей будет возрастать. После смены нескольких десятков или сотен поколений средняя приспособленность особей данного вида заметно возрастает.

В теории ГА используется биологическая терминология в упрощенном виде:

- хромосома – вектор чисел;
- ген – бит хромосомы;
- особь – набор хромосом;
- популяция – совокупность особей.

В качестве критерия жизнеспособности особи выступает функция приспособленности F – отображение совокупности хромосом особи на множество вещественных чисел. Точное выражение для функции приспособленности составляется индивидуально для каждой задачи.

В этом случае элементы x подмножества Ω в (1) можно рассматривать как особи некоторой популяции точек n -мерного пространства, а значения функции $f(x)$ с измененным знаком в этих точках можно рассматривать как значение функции приспособленности $F = -f(x)$.

Основные шаги работы классического ГА можно записать в следующем виде (рис. 1):

1. Генерация начальной популяции из k особей.
2. Вычисление функции приспособленности каждой особи и отброс наименее приспособленных особей.
3. Выбор среди оставшихся особей двух родителей.
4. Применение к родительским хромосомам оператора скрещивания с вероятностью p_c ; при отсутствии скрещивания дочерние хромосомы равны родительским.
5. Применение к полученным дочерним хромосомам оператора мутации с вероятностью p_m ; в отсутствие мутации хромосомы не изменяются.
6. Повторение шагов 3–5, пока популяция вновь не будет содержать k особей.
7. Повторение шагов 1–6 до тех пор, пока не будет достигнут критерий окончания процесса.

Скрещивание (кроссовер) – это операция, при которой две хромосомы обмениваются своими частями. В модели, описываемой ГА, кроссовер выполняет оператор скрещивания C . Существует несколько алгоритмов оператора скрещивания [3], самые простые из них – одно- и многоточечный.

В одноточечном варианте происходит разрыв двух родительских хромосом в случайной позиции, после чего они обмениваются полученными участками. При многоточечном кроссовере хромосомы обмениваются участками, полученными в результате нескольких разрывов.

Мутацией называется случайное изменение одного или нескольких генов хромосомы. Мутация производится оператором мутации M .

В качестве критерия окончания алгоритма могут выступать 2 условия:

- достижение заданного максимального количества итераций (поколений);
- схождение популяции.

Схождением популяции называется такое ее состояние, когда все особи популяции практически одинаковы и находятся в области некоторого экстремума.

ГА контролируется следующими параметрами:

- объемом популяции k ;
- вероятностью мутации p_m ;
- вероятностью скрещивания p_c ;
- числом новых особей на каждом этапе развития популяции m ;
- максимальным числом итераций алгоритма N ;
- при использовании критерия сходимости популяции – точностью схождения ε .

В данной работе представлены результаты анализа влияния первых четырех из вышеперечисленных параметров генетического алгоритма на время и точность нахождения глобального экстремума при использовании варианта алгоритма с одноточечным разрывом.

Построение модели с помощью планируемого эксперимента

Одним из путей проведения требуемого анализа генетического алгоритма является использование полиномиальных регрессионных моделей, построенных по данным численного или физического планируемых экспериментов, и, соответственно, учитывающих влияние вышеуказанных факторов.

Математическая модель в этом случае может быть представлена следующим выражением [4]:

$$\eta = f(x_1, x_2, \dots, x_k), \quad (4)$$

где x_i , $1 \leq i \leq k$ – независимые переменные (факторы),
 η – моделируемая выходная характеристика алгоритма,
 $f(x_1, x_2, \dots, x_k)$ – полином требуемого порядка.

Построение адекватных регрессионных моделей генетического алгоритма по экспериментальным данным, т.е. решение задачи нахождения коэффициентов уравнения регрессии (4) возможно лишь при соблюдении ряда жестких требований: план эксперимента должен быть симметричным, ортогональным, ротатабельным, D-оптимальным и т.д. [4].

Теория планирования эксперимента [4] предлагает различные методы построения оптимальных по разным критериям планов. Часто используются, например, полный факторный эксперимент, в котором реализуются все возможные сочетания уровней факторов, и его дробные реплики. Однако высокая

трудоемкость, а зачастую даже непригодность метода обусловлена катастрофическим возрастанием числа необходимых опытов с ростом числа учитываемых факторов. В связи с этим возникает задача повышения эффективности планируемого эксперимента путём предельного сокращения числа проводимых опытов при сохранении адекватности модели.

Одним из вариантов решения этой задачи является метод построения композиционных планов экспериментов высоких порядков на основе симплекса, обладающих свойствами симметричности, ортогональности, ротатабельности, D-оптимальности и содержащих при этом небольшое количество опытов. Алгоритм построения симплекс-суммируемого плана n-го порядка приведен в [4].

План эксперимента для моделирования генетического алгоритма

В данной работе для проведения численного эксперимента по моделированию работы генетического алгоритма был использован симплекс-суммируемый план 4 порядка, обладающий свойствами симметричности, ортогональности, ротатабельности и D-оптимальности.

В качестве независимых параметров (факторов) эксперимента рассматривались объем популяции k , вероятность мутации p_m , вероятность скрещивания p_c и число новых особей на каждом этапе развития популяции m .

Границы изменения варьируемых факторов в плане эксперимента подбирались на основе проведенного ранее исследования возможности распараллеливания генетического алгоритма [5] (табл. 1).

Таблица 1. Диапазон изменения факторов в плане эксперимента.

	k	m/2	pc	pm
min	33	2	0.274	0.059
max	78	7	0.826	0.091

В качестве анализируемых выходных характеристик генетического алгоритма было решено рассмотреть время работы алгоритма, выраженное в необходимом для получения решения количестве вычислений целевой функции, а также вероятность попадания результата в область глобального минимума. Для указанных характеристик было решено получить модели 2 порядка.

Полученный на основе алгоритма [4] симплекс-суммируемый план четвертого порядка приведён в табл. 2.

Таблица 2. Симплекс-суммируемый план эксперимента в натуральных округленных величинах параметров ГА.

№	k	m/2	pc	pm
1	33	3	0.458	0.071
2	78	3	0.458	0.071
3	55	7	0.458	0.071
4	55	5	0.826	0.071
5	55	5	0.55	0.091
6	55	5	0.55	0.075
7	39	4	0.328	0.072
8	71	3	0.588	0.072
9	70	6	0.512	0.072
10	39	5	0.772	0.072
11	55	5	0.55	0.072
12	55	5	0.274	0.079
13	55	2	0.642	0.079
14	78	6	0.642	0.079
15	33	6	0.642	0.079
16	55	5	0.55	0.059
17	71	4	0.328	0.078
18	39	3	0.588	0.078
19	71	5	0.772	0.078
20	39	6	0.512	0.078
21	55	5	0.55	0.078

План удовлетворяет условиям оптимальности [4] и содержит 21 опыт, что позволяет определить коэффициенты неполных полиномов 2 порядка, включающих параметры генетического алгоритма и их квадраты:

$$\tau = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m a_{i+4} x_i^2, \quad (5)$$

$$\rho = b_0 + \sum_{i=1}^m b_i x_i + \sum_{i=1}^m b_{i+4} x_i^2, \quad (6)$$

где τ , ρ – выходные характеристики генетического алгоритма, время работы алгоритма и вероятность попадания результата в область глобального минимума соответственно,

x_i – его параметры,

a_i , b_i – коэффициенты регрессии для выходных характеристик генетического алгоритма.

При этом обеспечивается достаточная избыточность.

Для нахождения коэффициентов уравнений (5) и (6) использовался шаговый регрессионный метод [6], позволяющий оценивать влияние указанных факторов на характеристики генетического алгоритма и включать в уравнения регрессии в первую очередь те факторы, которые оказывают наиболее существенное влияние на зависимые переменные, т.е. факторы с наибольшей частной корреляцией с зависимой переменной.

Поиск глобального минимума осуществлялся у многоэкстремальных ЦФ (7), (8) и (9) при количестве параметров оптимизации n , равном соответственно 4, 6 и 8.

- число параметров оптимизации $n=4$:

$$f(x_1, x_2, x_3, x_4) = \sum_{i=1}^4 \{0.1x_j^2 - 0.3\cos(9x_j)\}, \quad (7)$$

где $-0.5 < x_j < 1$, $j = 1, \dots, 4$;

$f_{\min} = -1.2$ при $x_j = 0$, $j = 1, \dots, 4$.

- число параметров оптимизации $n=6$:

$$f(x_1, x_2, \dots, x_5, x_6) = \sum_{i=1}^6 \{0.1x_j^2 - 0.3\cos(9x_j)\}, \quad (8)$$

где $-0.5 < x_j < 1$, $j = 1, \dots, 6$;

$f_{\min} = -1.8$ при $x_j = 0$, $j = 1, \dots, 6$.

- число параметров оптимизации $n=8$:

$$f(x_1, x_2, \dots, x_7, x_8) = \sum_{i=1}^8 \{0.1x_j^2 - 0.3\cos(9x_j)\}, \quad (9)$$

где $-0.5 < x_j < 1$, $j = 1, \dots, 8$;

$f_{\min} = -2.4$ при $x_j = 0$, $j = 1, \dots, 8$.

Процесс поиска глобального минимума завершился при выполнении следующих условий: при сходимости популяции, которое оценивалось по максимальному расстоянию между значениями функции приспособленности $\max_k(|\Delta f|)$, или по достижении максимального количества итераций алгоритма $N = 10000$.

Полученные при этом корреляции параметров с временем работы алгоритма, выраженным в необходимом для получения решения количестве вычислений целевой функции, а также с вероятностью попадания результата в область глобального минимума представлены в табл. 3.

Таблица 3. Корреляции натуральных характеристик ГА и их квадратов с числом запусков вычисления ЦФ в процессе выполнения ГА и вероятностью попадания результата выполнения ГА в область глобального минимума для разного количества аргументов целевой функции.

n	k	m/2	pc	pm	k2	m2/4	pc2	pm2
Число запусков вычисления ЦФ в процессе выполнения ГА								
4	0.812	-0.379	0.019	0.241	0.815	-0.326	0.001	0.235
6	0.606	0.392	0.060	0.484	0.610	0.459	0.028	0.485
8	0.151	0.739	0.036	0.409	0.157	0.802	-0.003	0.414
Вероятность попадания результата выполнения ГА в область глобального минимума								
4	0.956	-0.050	-0.109	0.063	0.923	-0.047	-0.123	0.071
6	0.967	-0.020	-0.090	0.105	0.947	-0.013	-0.115	0.112
8	0.957	-0.080	-0.102	0.138	0.949	-0.076	-0.135	0.143

Значения коэффициентов полиномов (5) и (6) представлены в табл. 4 и табл. 5 соответственно.

Таблица 4. Коэффициенты регрессии для расчёта числа запусков вычисления ЦФ в процессе выполнения ГА для разного количества аргументов целевой функции.

n	4	6	8
a0	367.73266162	2043.70549139	16178.567775
a1	4.26837526	-5.77197475	-106.05555855
a2	-148.15155182	-386.17808074	-1954.60997784
a3	-81.49595755	-246.08914335	-156.37312384
a4	-4090.36941153	-39871.7540169	-337191.052933
a5	0.00251330	0.12489119	1.06890869
a6	13.98601102	50.03390008	294.21752399
a7	98.49133251	250.4219348	-85.861382
a8	49162.4860944	363667.509614	2725628.35605

Таблица 5. Коэффициенты регрессии для расчёта вероятности попадания результата выполнения ГА в область глобального минимума для разного количества аргументов целевой функции.

n	4	6	8
b0	-0.21767527	-0.89718040	-0.94755971
b1	0.02576801	0.02534969	0.01460169
b2	-0.00193986	0.00763478	0.04152101
b3	0.50602928	1.17568895	1.61199231
b4	-2.18687935	1.66766550	0.71579381
b5	-0.00015213	-0.00011743	-0.00002033
b6	-0.00060958	-0.00129618	-0.00586314
b7	-0.54834324	-1.17156045	-1.57790135
b8	23.98330553	9.91187973	23.22239932

Оценка статистической значимости коэффициентов моделей проводилась проверкой гипотезы о статистической незначимости совокупности коэффициентов регрессии уравнений с помощью F-критерия (критерия Фишера). При этом гипотеза является несостоятельной, если вычисленные значения F-критерия превышают табличные критические значения при выбранных уровнях значимости [6].

Основные параметры, характеризующие статистические свойства моделей, рассчитанные в предположении о нормальном законе распределения ошибок представлены в табл. 6.

Таблица 6. Расчётные статистические характеристики регрессионных моделей количества вычислений ЦФ в процессе выполнения ГА и вероятности попадания

результата выполнения ГА в область глобального минимума для разного количества аргументов целевой функции

n	Коэффициент множественной корреляции	F-критерий (Fтабл(0.01) = 5,66)	95% доверительный интервал (%)
Количество вычислений ЦФ в процессе выполнения ГА			
4	0.9963	31.94	26.0
6	0.9795	55.48	18.7
8	0.9811	60.04	35.3
Вероятность попадания результата выполнения ГА в область глобального минимума			
4	0.9997	138.22	5.7
6	0.9963	308.94	6.9
8	0.9961	295.93	10.0

Как видно из табл. 6, вычисленные по числу степеней свободы и заданному уровню значимости моделей F-критерии значительно превышают критическое значение $F_{табл}(0.01) = 5.66$ [7]. Таким образом, все построенные модели являются статистически достоверными при уровне значимости $\alpha = 0.01$.

Расчитанные относительные погрешности при нормальном законе распределения ошибок и уровне достоверности $p = 0.05$ (табл. 6) находятся в пределах $\pm 18.9\% \div \pm 26\%$ для количества вычисления ЦФ в процессе выполнения ГА и $\pm 5.7\% \div \pm 10\%$ для вероятности попадания результата выполнения ГА в область глобального минимума.

Экспериментальная проверка адекватности регрессионных моделей

С целью проверки адекватности построенных моделей t и p , а также для оценки их точности было проведено экспериментальное тестирование. Значения параметров генетического алгоритма для контрольного эксперимента приведены в табл. 7.

Таблица 7. Коэффициенты регрессии для расчёта вероятности попадания результата выполнения ГА в область глобального минимума для разного количества аргументов целевой функции

№	k	m/2	pc	pm
1	75	5	0.209	0.09
2	59	6	0.224	0.062
3	65	8	0.5	0.075

Таблица 8. Относительные погрешности расчёта количества вычислений ЦФ в процессе выполнения ГА и вероятности попадания результата выполнения ГА в область глобального минимума по регрессионным моделям для разного количества аргументов целевой функции.

n	Расчётное значение	Экспериментальное значение	Погрешность расчёта (%)
Число запусков вычисления ЦФ в процессе выполнения ГА Контрольный эксперимент №1			
4	328.2431	334.564	1.9
6	950.0324	920.091	-3.3
8	3513.3629	3164.004	-11.0
Вероятность попадания результата выполнения ГА в область глобального минимума Контрольный эксперимент №1			
4	0.9135	0.904	-1.1
6	0.7742	0.745	-3.9
8	0.6148	0.62	0.8
Число запусков вычисления ЦФ в процессе выполнения ГА Контрольный эксперимент №2			
4	164,9672	153.677	-7.3
6	505.3875	377.456	-33.9

8	2038.4663	1043.704	-95.3
Вероятность попадания результата выполнения ГА в область глобального минимума Контрольный эксперимент №2			
4	0.7819	0.758	-3.2
6	0.5349	0.517	-3.5
8	0.2968	0.361	17.8
Число запусков вычисления ЦФ в процессе выполнения ГА Контрольный эксперимент №3			
4	240.6019	232.381	-3.5
6	446.8316	453.931	1.6
8	632.8148	897.861	29.5
Вероятность попадания результата выполнения ГА в область глобального минимума Контрольный эксперимент №3			
4	0.8838	0.888	0.5
6	0.74	0.733	-1.0
8	0.5838	0.605	3.5

В табл. 8 приведены относительные погрешности расчета рассматриваемых выходных характеристик генетического алгоритма по моделям в сравнении с их статистически рассчитанными значениями при 95% доверительном интервале.

Таким образом, относительная погрешность расчёта по полученным регрессионным моделям количества вычислений ЦФ в процессе выполнения генетического алгоритма, как видно из табл. 8, не превышает 11%. Относительная погрешность расчета вероятности попадания результата выполнения алгоритма в область глобального минимума при использовании тех же моделей составляет не более 3.9%.

Исследование влияния параметров генетического алгоритма на его выходные характеристики

Первоначальная оценка влияния факторов ГА на оптимизацию осуществлялась по рассчитанным при построении модели значениям коэффициентов корреляции. Как видно из табл. 3, сильное влияние на число запусков вычисления ЦФ в процессе выполнения ГА оказывает размер популяции k , число детей в каждом поколении m и вероятность применения

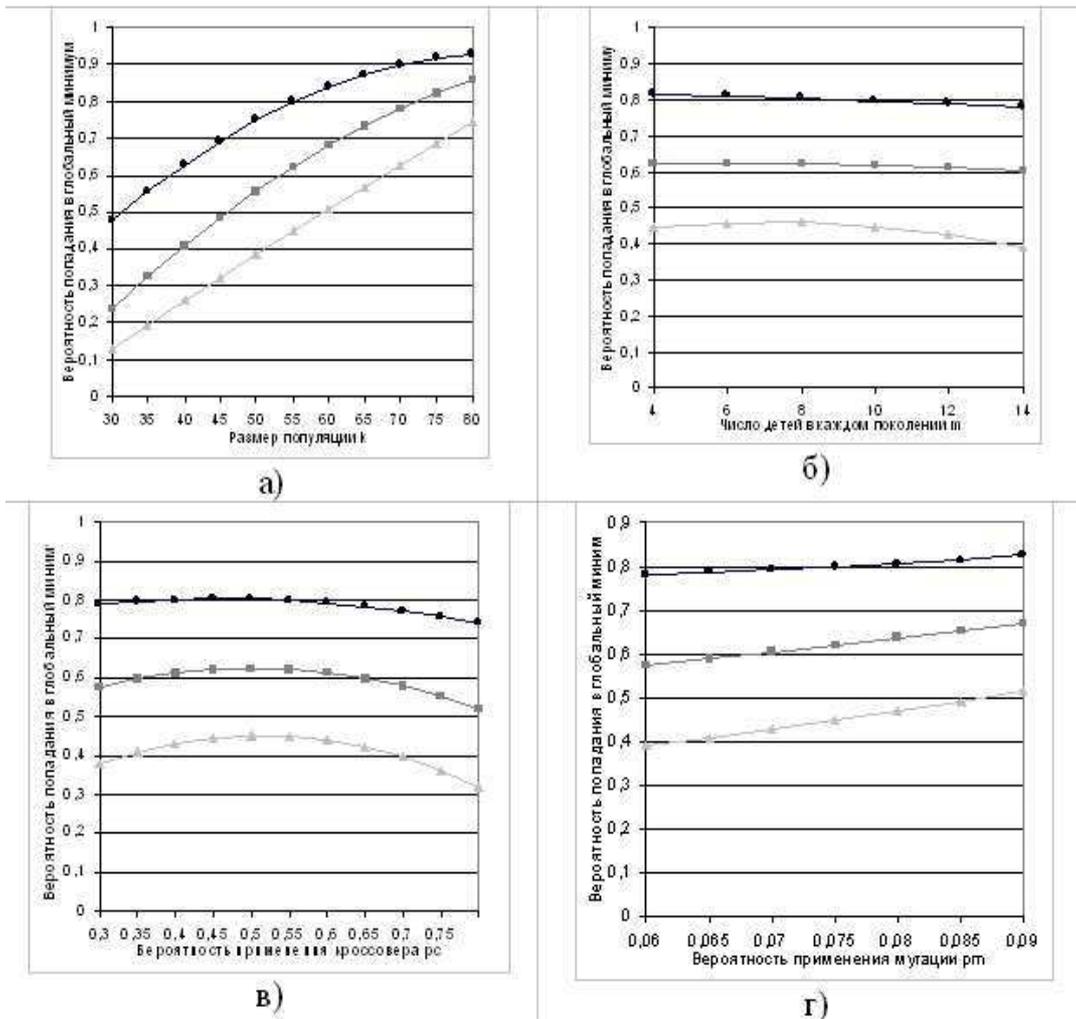


Рис. 1. Графики зависимости вероятности попадания в глобальный минимум от размера популяции k (а), числа детей в каждом поколении m (б), вероятности применения кроссовера p_c (в) и вероятности применения мутации p_m (г)

—●— — $n=4$, —■— — $n=6$, —○— — $n=8$

мутации p_m , а вероятность применения кроссовера p_c влияния почти не оказывает. На вероятность попадания результата выполнения ГА в область глобального минимума в свою очередь существенное влияние оказывает только размер популяции k .

Построенные регрессионные модели выходных характеристик генетического алгоритма (5), (6) позволяют получить их явные зависимости от каждого из рассмотренных факторов, т.к. не содержат перекрестных членов.

Графики зависимостей числа вычислений ЦФ от размера популяции k (а), числа детей в каждом поколении m (б), вероятности применения кроссовера p_c (в) и вероятности применения мутации p_m (г) приведены на рис. 1.

Графики вероятности попадания результата выполнения ГА в область глобального минимума от размера популяции k (а), числа детей в каждом поколении m (б), вероятности применения кроссовера p_c (в) и вероятности применения мутации p_m (г) приведены на рис. 2.

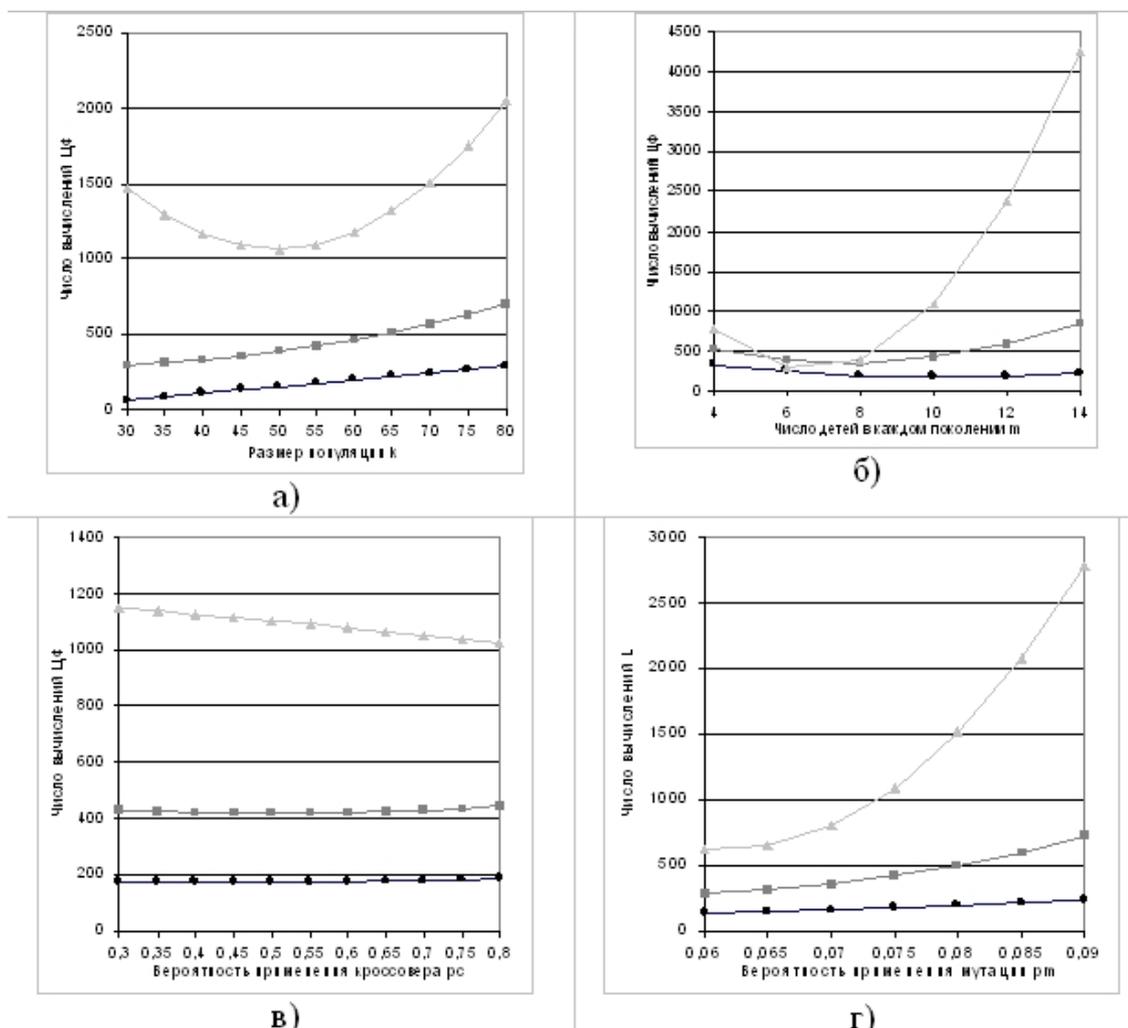


Рис. 2. Графики зависимости числа вычислений ЦФ от размера популяции k (а), числа детей в каждом поколении m (б), вероятности применения кроссовера p_c (в) и вероятности применения мутации p_m (г)

—●— — $n=4$, —■— — $n=6$, —○— — $n=8$

Данные зависимости могут быть использованы для определения оптимальных параметров ГА при решении оптимизационных задач.

Заключение

В ходе данной работы был проведен корреляционный анализ для оценки степени влияния параметров ГА на его выходные характеристики.

Были построены регрессионные модели второго порядка времени нахождения глобального минимума ЦФ и вероятности попадания результата выполнения ГА в область глобального минимума. Была проведена оценка статистической достоверности и экспериментальная проверка адекватности моделей. Было установлено, что погрешность расчетов по сравнению с экспериментом не превышает теоретически предсказанных значений.

Таким образом, результаты проведенных исследований могут быть применены при выборе оптимальных параметров ГА для поиска глобального экстремума.

ЛИТЕРАТУРА:

1. A.A. Eroftiev, N.E. Timofeeva, A.N. Savin. Parallel Computing in Application to Global Optimization Problem Solving // Grid and Visualization Systems: MIPRO, 2011 Proceedings of the 34th International Convention .– Zagreb, Croatia: DENONA, 2011. P. 185-190.
2. R. Myers, E.R. Hancock. Empirical Modelling of Genetic Algorithms // Evolutionary Computation, Winter 2001, Vol. 9, No. 4, P. 461-493.
3. А.Н. Савин, Д.М. Доронин, И.А. Накрап, И.Н. Салий. Метод построения симплекс-суммируемого плана n -го порядка проведения эксперимента для моделирования характеристик замедляющих систем // Физика

волновых процессов и радиотехнические системы. – Самара: Изд-во Самарского гос-о ун-та, 2010. – Том 13. N 2 — с.68-75

4. В.Д. Кошур. Адаптивный алгоритм глобальной оптимизации на основе взвешенного усреднения координат и нечетко-нейронных сетей // Нейроинформатика, 2006, том 1, № 2