Solving the problem of planning a chemical experiment based on genetic algorithms

E V Antipina¹, S A Mustafina¹, A F Antipin¹, S I Mustafina¹

¹Bashkir State University, Zaki Validi street 32, Ufa, Russia, 450076

Abstract. The article is devoted to the task of planning an experiment in chemistry and chemical technology. The problem statement of finding the optimal ratio of the initial substances of reaction is defined. A genetic algorithm for solving this problem is proposed. The algorithm is tested on the example of the reaction of aminomethylation of thiols using tetramethylmethanediamine. The optimal initial concentrations of reagents are calculated to obtain the maximum yield of the target reaction product.

1. Introduction

Solving the problems of chemistry and chemical technology requires expensive and complex experiments. It is worthwhile to use mathematical modeling methods to reduce the material and time costs of a natural experiment. Mathematical modeling methods allow us to identify the main patterns of the chemical process at the stage of the computational experiment.

One of such tasks is planning an experiment in chemistry. The use of mathematical modeling methods makes it possible to determine the optimal initial concentrations of reagents to obtain the highest yield of the target reaction product. This enables to significantly reduce the cost of the synthesis of substances.

The task of finding optimal initial concentrations of reagents is the optimal control problem. The control parameter is the vector of initial concentrations of initial substances. However, the solutions to optimal control problems, found by using most numerical methods, depending on the choice of the initial approximation [1, 2, 3]. To overcome this difficulty is possible by using metaheuristic methods, genetic algorithms in particular. Their significant advantage is the independence of the found solution from the initial approximation [4, 5]. Genetic algorithms exclude reaching the points of the local extremum of the target function. Therefore, they provide high-quality solutions in comparison with classical optimization methods.

We formulate the problem of finding the optimal initial concentrations of the initial substances of a chemical reaction in general terms.

Let the dynamics of concentrations of chemical reaction substances be described by the system of ordinary differential equations [6, 7]

\[
\frac{dx}{dt} = f(t, x(t), T),
\]

with initial conditions
\[ x_i(0) = x_i^0, \quad i = 1, n, \]  

where \( x(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \) – vector of reactant concentrations, \( t \in [0, t_{\text{end}}] \) – reaction time, \( T \) – reaction temperature.

Initial concentrations of substances are given by some ratio:

\[ x_1(0) : x_2(0) : \ldots : x_n(0) = \alpha_1 : \alpha_2 : \ldots : \alpha_n. \]  

It is necessary to determine the ratio of initial concentrations of substances (3), at which the extremum of the optimality criterion is achieved

\[ Q(x) = \sum_{i=1}^{n} \lambda_i x_i(t_{\text{end}}) \rightarrow \text{extr}. \]

The optimality criterion expresses the maximum yield of reaction products (for positive values \( \lambda_i \)) or the minimum content of impurities (for negative values \( \lambda_i \)) at the final point in time during the reaction.

2. Materials and methods
Genetic algorithms for solving optimization problems simulate the process of evolution of living organisms. The evolutionary process consists of the successive change of generations of populations. At the same time, the fittest individuals move to the next generation [8, 9, 10].

We will consider sets of initial concentrations of substances as a mathematical analog of the population:

\[ x^j(0) = (x_1^j(0), x_2^j(0), \ldots, x_n^j(0)), \quad j = 1, P, \]  

where \( P \) is the population size.

Each vector is an analog of an individual, and the element of the vector is a gene. The optimality criterion (4) is a fitness function. To calculate it, the direct kinetic problem must be solved. For this, it is necessary to solve the system of differential equations (1) with the initial conditions (2).

The main stages of the genetic algorithm are:

1) Creation of an initial population of concentrations of initial substances. A set of vectors (5) is generated randomly. The value of the fitness function is calculated for each individual.

2) Selection. Two individuals are selected for subsequent crossing from the current population using the selection operator.

3) Crossover (crossbreeding). A new descendant individual is generated by applying one of the crossover operators.

4) Mutation. A descendant individual is exposed to one of the mutation operators to prevent the population from reaching a point in a local extremum.

5) Population improvement. The least adapted individual from the current population is selected. It is replaced by a descendant individual. Then go to step 2 until the end of the search condition is reached. A descendant individual with the best fitness function value from the latest population will be an optimal set of initial concentrations of chemical reaction reagents.

3. Computing experiment
We calculate the optimal ratio of initial concentrations for the reaction of aminomethylation of thiols by using tetramethylmethanediame to obtain the maximum yield of the reaction product. Nitrogen- and sulfur-containing organic compounds are widely used as effective plant protection products, antioxidant, anticorrosive, antiwear additives for fuels and oils. In [11] experimental studies of the aminomethylation of thiols with tetramethylmethanediame are carried out. The scheme of this reaction is represented by the sequence of stages:
\[ X_1 + X_2 \rightarrow X_3, \]
\[ X_3 + X_4 \rightarrow X_2 + X_5 + X_6, \]

(6)

where \( X_1 = \text{N}_2(\text{CH}_3)_4, X_2 = \text{Sm}, X_3 = \text{N}_2(\text{CH}_3)_4[\text{Sm}], X_4 = \text{HSC}_5\text{H}_11, X_5 = (\text{CH}_3)_2\text{NSC}_5\text{H}_11, X_6 = (\text{CH}_3)_2\text{NH}. \)

The kinetic equations of the rates of the reaction stages are written out according to the law of the acting masses:

\[ \omega_1 = k_1 x_1 x_2, \]
\[ \omega_2 = k_2 x_3 x_4, \]

(7)

where \( x = (x_1, x_2, \ldots, x_6)^T \) – vector of substance concentrations (mole fraction), \( k = (k_1, k_2) \) – vector of kinetic reaction constants (l/(mol·h)), which is calculated according to the Arrhenius equation:

\[ k_j = k_{0j} \exp \left( -\frac{E_j}{RT} \right), \quad j = 1, 2, \]

where \( k_{0j} \) – preexponential multiplier (l/(mol·h)), \( E_j \) – activation energy of the j-th stage (kcal/mol), \( T \) – reaction temperature (K), \( R \) – universal gas constant (cal/(K·mol)).

The matrix of stoichiometric coefficients of substances is given in Table 1.

| \( \omega_1 \) | \( \omega_2 \) |
|---------------|--|
| \( X_1 \)     | -1 | 0   |
| \( X_2 \)     | -1 | 1   |
| \( X_3 \)     |  1 | -1  |
| \( X_4 \)     |  0 | -1  |
| \( X_5 \)     |  0 |  1  |
| \( X_6 \)     |  0 |  1  |

Table 1. Matrix of stoichiometric coefficients of the reaction (7)

The kinetic model of the thiol aminomethylation reaction is represented by a system of differential equations:

\[
\begin{align*}
\frac{dx_1}{dt} &= -\omega_1, \\
\frac{dx_2}{dt} &= -\omega_1 + \omega_2, \\
\frac{dx_3}{dt} &= \omega_1 - \omega_2, \\
\frac{dx_4}{dt} &= -\omega_2, \\
\frac{dx_5}{dt} &= \omega_2, \\
\frac{dx_6}{dt} &= \omega_2.
\end{align*}
\]

(8)

with initial conditions

\[ x_i(0) = x_i^0, \quad i = 1, 6. \]

(9)

The initial substances of the reaction are \( X_1, X_2, X_4 \). The concentrations of these substances at the initial time are related by the ratio

\[ x_1(0) + x_2(0) + x_4(0) = 1. \]
The initial concentrations of the remaining substances are equal to zero. It is necessary to find the ratio of the initial concentrations $x_1(0) : x_2(0) : x_4(0)$ of the reagents $X_1$, $X_2$, $X_4$, which provides the maximum yield of the reaction product $X_5$ at the final moment of the reaction time:

$$G(x, T) = x_5(t_{\text{end}}) \rightarrow \text{max}.$$  

The problem is solved with the following parameters of the genetic algorithm: selection operator-tournament selection, crossover operator – arithmetic crossover, mutation operator-random mutation, maximum population size – 3000, population size – 60, number of individuals in the population – 450. The reaction temperature is 333 K, the reaction time is 1 hour. The solution of the system of differential equations (8) with initial conditions (9) is obtained by the fourth-order Runge-Kutta method.

As a result of the calculations, it is found out that the maximum yield of the reaction product (6) is achieved at the following set of initial concentrations of the starting substances:

$$x_1(0) : x_2(0) : x_4(0) = 0.457 : 0.071 : 0.472$$  

The maximum yield of the reaction product $X_5$ is 0.435 mole fractions.

The change in concentration over time of the reaction product $X_5$ at optimal values of the initial concentrations is shown in Figure 1. The dynamics of concentrations of the initial substances at the ratio (10) is shown in Figure 2.

Let us consider the solution of the direct kinetic problem for the reaction scheme (6) with arbitrary sets of initial concentrations of substances. As can be seen from Table 2, the maximum concentration of the reaction product $X_5$ is found at the ratio of the initial substances (10), which confirms the correct operation of the constructed algorithm.

### Table 2. Dependence of the concentration of the reaction product $X_5$ from the ratio of the initial substances $X_1$, $X_2$, $X_4$

| $X_1$ (mole fraction) | $X_2$ (mole fraction) | $X_4$ (mole fraction) | $X_5$ (mole fraction) |
|-----------------------|-----------------------|-----------------------|-----------------------|
| 0.1                   | 0.2                   | 0.7                   | 0.075                 |
| 0.1                   | 0.5                   | 0.4                   | 0.089                 |
| 0.1                   | 0.7                   | 0.2                   | 0.072                 |
| 0.2                   | 0.1                   | 0.7                   | 0.094                 |
| 0.2                   | 0.6                   | 0.2                   | 0.122                 |
| 0.2                   | 0.2                   | 0.5                   | 0.094                 |
| 0.3                   | 0.3                   | 0.4                   | 0.122                 |
| 0.3                   | 0.4                   | 0.3                   | 0.190                 |
| 0.3                   | 0.5                   | 0.2                   | 0.150                 |
| 0.4                   | 0.2                   | 0.4                   | 0.203                 |
| 0.4                   | 0.3                   | 0.3                   | 0.204                 |
| 0.4                   | 0.4                   | 0.2                   | 0.163                 |
| 0.5                   | 0.1                   | 0.4                   | 0.147                 |
| 0.5                   | 0.2                   | 0.3                   | 0.189                 |
| 0.5                   | 0.3                   | 0.2                   | 0.161                 |
| 0.6                   | 0.1                   | 0.3                   | 0.132                 |
| 0.6                   | 0.2                   | 0.2                   | 0.145                 |
| 0.6                   | 0.3                   | 0.1                   | 0.088                 |
| 0.7                   | 0.1                   | 0.2                   | 0.101                 |
| 0.7                   | 0.2                   | 0.1                   | 0.079                 |
Figure 1. Dynamics of the optimal concentration of the target substance $X_5$.

Figure 2. Dynamics of the optimal concentrations of the initial substances $X_1$, $X_2$, $X_4$.

4. Conclusion
Thus, the developed genetic algorithm for finding the ratio of initial concentrations of the initial substances makes it possible to solve the problem of planning an experiment in chemistry at the stage of the computational experiment. In this case, the solution of the optimization problem will be found for any set of values of the initial concentrations of substances. This is because the work of genetic algorithms does not depend on the starting point of the search for a solution.

The algorithm can be applied to problems of finding optimal control for other processes that are described by nonlinear systems of differential equations. The program based on this algorithm has been developed in the Delphi environment. The program enables us to determine the optimal initial concentrations of substances, at which the maximum criterion of optimality is achieved. The computational experiment is performed for the reaction of aminomethylation of thiols. The optimal initial concentrations of reagents that provide the maximum yield of the target reaction product at the final time are determined.

5. Acknowledgments
The reported study was funded by Ministry of Science and Higher Education of the Russian Federation according to the research project № FZWU-2020-0027.

6. References
[1] Ostrovsky G, Zyatdinov N and Lapteva T 2011 One stage optimization problem with chance constraints Chemical Engineering Transactions 25 225-230.
Biegler L T 2017 Integrated Optimization Strategies for Dynamic Process Operations *Theor. Found. Chem. Eng.* 51(6) 910-927.

Grossmann I, Aguirre P and Bartfeld M 2005 Optimal synthesis of complex distillation columns using rigorous models *Comput. Chem. Eng.* 29 1203.

Holland J N 1975 *Adaptation in Natural and Artificial Systems* (Ann Arbor, Michigan: Univ. Michigan Press) 232.

Wright A 1991 Genetic algorithms for real parameter optimization *Foundat. Genet. Algorithm.* 1 205-218.

Sedov V A and Sedova N A 2014 Modelling collision avoidance actions in closest approach zones by means of neural networks *Asia-Pacific Journal of Marine Science & Education* 4(2) 104-111.

Mustafina S A, Vaytiev V A and Stepashina E V 2014 Search the kinetic parameters of the reduced scheme of α-methylstyrene dimerization reaction *ARPN Journal of Engineering and Applied Sciences* 9(7) 1118-1120.

Mustafina S, Antipin A, Antipina E, Odinokova E, Tuchkina L, Kolyazov K and Mustafina S 2020 Numerical algorithm for finding optimal initial concentrations of chemical reactions *IIUM Engineering Journal* 21(1) 167-174.

Panteleev A V and Metlitskaya D V 2011 An application of genetic algorithms with binary and real coding for approximate synthesis of suboptimal control in deterministic systems *Automation and Remote Control* 72(11) 2328-2338.

Herrera F, Lozano M and Verdegay J L 1998 Tackling real-coded genetic algorithms: operators and tools for the behavior analysis *Artific. Intelligence Rev.* 12(4) 265-319.

Khairullina R R, Tyumkina T V, Ibragimov A G, Akmanov B F and Kunakova R V 2012 N,N,N',N'-tetramethylmethanediamine, efficient reagent for thioles aminomethylation *Russian Journal of Organic Chemistry* 48(2) 175-179.