

HYBRID BAT ALGORITHM FOR PARAMETER IDENTIFICATION OF AN *E. COLI* CULTIVATION PROCESS MODEL

Olympia Nikolaeva Roeva¹ and Stefka Stoyanova Fidanova²

¹Bulgarian Academy of Sciences, Institute of Biophysics and Biomedical Engineering, Sofia, Bulgaria

²Bulgarian Academy of Sciences, Institute of Information and Communication Technologies, Sofia, Bulgaria

Correspondence to: Stefka Fidanova

E-mail: stefka@parallel.bas.bg

ABSTRACT

In this paper, a hybrid scheme using Bat Algorithm (BA) and Sequential Quadratic Programming (SQP) method is introduced. In the hybrid BA–SQP, the role of BA is to generate feasible solutions to a problem. The role of SQP is to exploit the information gathered by BA. This process obtains a solution which is at least as good as – but usually better than – the best solution devised by BA. To demonstrate the usefulness of the presented approach, the hybrid scheme was applied to parameter identification of an E. coli MC4110 fed-batch cultivation process model. A comparison with both the conventional BA and SQP method is presented. The results showed that the hybrid BA–SQP has the advantages of both BA’s global search ability and SQP’s local search ability, thus enhancing the overall search ability and computational efficiency. For comparison, the results obtained by applying Ant colony optimization algorithm in conditions similar to those of BA are further shown.

Biotechnol. & Biotechnol. Eq. 2013, 27(6), 4323–4326

Keywords: metaheuristics, Bat Algorithm, SQP, Ant colony optimization, local search, parameter identification, cultivation process, *E. coli*

Introduction

Identification of the real parameters of cellular dynamics models is a research field of growing interest. Robust and efficient methods for parameter identification are of key importance. Today, researchers employ three different approaches to solve optimization problems: exact algorithms, heuristics, and metaheuristics (4). Metaheuristics are generic methods which offer good solutions, even global optima, within reasonable computing time (15). Thus, the use of metaheuristics has received more and more attention (4). An even more efficient behavior and higher flexibility when dealing with real-world and large-scale problems, can be achieved through a combination of a meta-heuristic with other optimization techniques, the so-called hybrid meta-heuristic (4, 12). It has been demonstrated that exact optimization techniques and metaheuristics both have specific advantages, which complement each other (10). Suitable combinations of exact algorithms and metaheuristics can benefit much from synergy and often exhibit significantly higher performance with respect to solution quality and time (11).

In this paper, a hybrid approach between the meta-heuristic Bat Algorithm (BA) and the numerical method Sequential Quadratic Programming (SQP) (BA–SQP) is introduced for the first time, and applied to parameter identification of an *E. coli* cultivation process model.

BA is a new meta-heuristics based on the echolocation behaviour of bats, i.e. on their ability to find their prey and discriminate between different types of insects even in complete darkness (1). This algorithm was proposed by Xin-She Yang

(16, 17). The resulting algorithm is simple in concept and at the same time powerful in implementation (8, 16, 17). SQP is one of the leading numerical methods for solving constrained non-linear optimization problems (5, 9). The algorithms in this class guarantee global convergence but need huge amounts of computational resources, like time and memory, or stop in some local minima.

In order to keep the advantages of both the meta-heuristic method and the numerical one, and to avoid their disadvantages, we propose a hybrid BA–SQP method. The results for *E. coli* MC4110 cultivation process model parameter identification were compared to those obtained by our previous algorithm (7) based on Ant colony optimization (ACO) methodology (6).

Materials and Methods

Hybrid Bat Algorithm – Sequential Quadratic Programming Algorithm

According to the classification of existing approaches combining exact and meta-heuristic algorithms (11), a collaborative combination of BA and SQP methods is proposed here. In the BA–SQP hybrid, in the first step, BA explores the search space in order to isolate the most promising region. In the second step, SQP is introduced to explore the search space and find new better solutions.

Bat Algorithm

BA, as proposed by Yang and Gandomi (16) and Yang (17), is based on the following idealized rules:

1. All bats use echolocation to sense distance, and they also “know” the difference between food/prey and background barriers.
2. Bats fly randomly, with velocity v_i at position x_i with a fixed frequency f_{\min} , varying wavelength λ and loudness

L_0 , to search for prey. They can automatically adjust the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission r in the range of $[0, 1]$, depending on the proximity of their target.

- Although the loudness can vary in many ways, it could vary, for example, from a positive L_0 to a minimum constant value L_{\min} .

The new solutions $x_i(t)$ and velocities $v_i(t)$ at time step t are given by (17):

$$v_i(t) = v_i(t-1) + (x_i(t) - x_*)f_i, \quad (\text{Eq. 1})$$

$$x_i(t) = x_i(t-1) + v_i(t), \quad (\text{Eq. 2})$$

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta, \quad (\text{Eq. 3})$$

where $\beta \in [0, 1]$ is a random vector drawn from a uniform distribution, x_* is the current global best solution which is located after comparing all the solutions among all the n bats, f_i is used to adjust the velocity change.

For the local search part, once a solution is selected among the current best solutions, a new solution for each bat is generated locally using random walk (17):

$$x_{\text{new}} = x_{\text{old}} + \eta L_i(t), \quad (\text{Eq. 4})$$

where $\eta \in [-1, 1]$ is a random number, $L_i(t)$ is the average loudness of all the bats at this time step.

The loudness $L_i(t)$ and the rate $r_i(t)$ of pulse emission have to be updated accordingly as the iterations proceed:

$$\begin{aligned} L_i(t+1) &= \alpha L_i(t), \\ r_i(t+1) &= r_i(0)[1 - \exp(-\gamma t)], \end{aligned} \quad (\text{Eq. 5})$$

where α and γ are constants, whose choice requires some experimenting.

For any $0 < \alpha < 1$, $0 < \gamma$, we have (16):

$$L_i(t) \rightarrow 0, r_i(t) \rightarrow r_i(0), \text{ as } t \rightarrow \infty \quad (\text{Eq. 6})$$

The loudness $L_i(t)$ and emission rates $r_i(t)$ will be updated only if the new solutions are improved, which means that these bats are moving towards the optimal solution.

Sequential Quadratic Programming

SQP is one of the most successful methods for numerical solution of constrained nonlinear optimization problems. It provides powerful algorithmic tools for the solution of large-scale technologically relevant problems.

The general nonlinear optimization problem (NLP) to minimize an objective function f under nonlinear equality and inequality constraints is (9):

$$\begin{aligned} \min_{x \in \mathcal{R}^n} f(x) \\ \text{subject to } c(x) = 0, b(x) \geq 0, x_l \leq x \leq x_u \end{aligned} \quad (\text{Eq. 7})$$

where x is an n -dimensional parameter vector. It is assumed that all problem functions $f(x)$, $c(x)$ and $b(x)$ are continuously differentiable on the whole \mathcal{R}^n .

SQP is an iterative procedure which models the NLP for a given iterate x_k , $k \in \mathbb{N}_0$, by a Quadratic Programming (QP) subproblem, solves that QP subproblem, and then uses the solution to construct a new iterate x_{k+1} . This construction is done in such a way that the sequence $(x_k)_{k \in \mathbb{N}_0}$ converges to a local minimum x_* of the NLP (Eq. (7)) as $k \rightarrow \infty$.

At iteration x_k (for the equality constraints), a basic SQP algorithm defines an appropriate search direction d_k as a solution to the QP subproblem (9):

$$\min_{d \in \mathcal{R}^n} f(x_k) + g(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) + \sum_{i=1}^l \lambda^i \nabla^2 c^i(x_k) \quad (\text{Eq. 8})$$

$$\text{subject to } c(x_k) + A(x_k)d = 0$$

is equal to, or is a symmetric approximation for, the Hessian of the Lagrangian.

The structure of the hybrid BA–SQP can be described by the following pseudo-code:

begin BA

Objective function $f(x)$, $x = (x_1, \dots, x_d)^T$

Initialize the bat population x_i ($i = 1, 2, \dots, n$) and v_i

Define pulse frequency f_i at x_i , pulse rates r_i and the loudness L_i
while ($t < \text{max number of iterations}$)

Generate new solutions by adjusting frequency and velocities

if ($\text{rand} > r_i$)

Select a solution among the best solutions

Generate a local solution around the selected best solutions

end if

Generate new solutions by flying randomly

if ($\text{rand} < L_i \ \& \ f(x_i) < f(x_*)$)

Accept the new solution

Increase r_i and L_i

end if

Rank the bats and find the current best x_*

Check the termination condition

If yes, save final best solution

If no, go to next iteration

end while

Final best solution x

begin SQP

Initialize the initial point $x = x_0$

Initialize the Hessian matrix ($H_0 = I$)

Evaluate f_0 , g_0 , c_0 and A_0

Solve the QP subproblem to find search direction d_k

Update $x_{k+1} = x_k + \alpha_k d_k$

Evaluate f_{k+1} , g_{k+1} , c_{k+1} and A_{k+1}

Convergence check

If yes, go to exit

If no, obtain H_{k+1} by updating H_k

and go back to solve the QP subproblem

Final best solution

Post-process results and visualization

Problem formulation

Application of the general state space dynamical model (3) to the *E. coli* cultivation fed-batch process leads to the following nonlinear differential equation system (14):

$$\frac{dX}{dt} = \mu_{max} \frac{S}{k_s + S} X - \frac{F_{in}}{V} X \quad (\text{Eq. 9})$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}} \mu_{max} \frac{S}{k_s + S} X + \frac{F_{in}}{V} (S_{in} - S) \quad (\text{Eq. 10})$$

$$\frac{dV}{dt} = F_{in} \quad (\text{Eq. 11})$$

where X is biomass concentration, [g/L]; S is substrate concentration, [g/L]; F_{in} is feeding rate, [l/h]; V is bioreactor volume, [L]; S_{in} is substrate concentration in the feeding solution, [g/L]; μ_{max} is the maximum value of the specific growth rate, [1/h]; k_s is saturation constant, [g/L]; $Y_{S/X}$ is yield coefficient, [-].

For the parameter estimation problem, real experimental data of the *E. coli* MC4110 fed-batch cultivation process are used. Measurements of biomass and glucose concentration are used in the identification procedure. The fed-batch cultivation starts at $t = 6.68$ h and proceeds up to $t = 11.54$ h. The cultivation conditions and experimental data have previously been described in detail (2, 13). The initial process conditions are (2):

$$t_0 = 6.68 \text{ h}, X(t_0) = 1.25 \text{ g/L}, S(t_0) = 0.8 \text{ g/L}, S_{in} = 100 \text{ g/L}.$$

The objective function is presented as a minimization of a distance measure J between experimental and model predicted values:

$$J = \sum_{i=1}^n (X_{mod}(i) - X_{exp}(i))^2 + \sum_{i=1}^n (S_{mod}(i) - S_{exp}(i))^2 \rightarrow \min \quad (\text{Eq. 12})$$

where n is the number of data for each state variable (X , S); X_{exp} and S_{exp} are the experimental data; X_{mod} and S_{mod} are model predictions with a given set of the parameters. Model parameters that will be estimated are μ_{max} , k_s and $Y_{S/X}$.

System configuration

All computations were performed using a PC/Intel Core i5-2320 CPU @ 3.00 GHz, 8 GB Memory (RAM), Windows 7 (64 bit) operating system and Matlab 7.5 environment.

Results and Discussion

The proposed hybrid BA-SQP, conventional BA and SQP algorithms were applied for parameter identification of the *E. coli* MC4110 fed-batch cultivation process model. Real experimental data without filtration or any processing were used. The idea was to test the algorithms in such hard real conditions.

BA has its own algorithm parameters that affect its performance in terms of solution quality and computational time. In order to increase the performance of the BA, it is necessary to provide the adjustments of the parameters depending on the problem domain. With an appropriate choice of the algorithm settings, the accuracy of the achieved solutions and the execution time can be optimized. The parameters of the BA (number of bats (NB), number of iterations (NI), α ,

γ , f_{min} , f_{max} , L_0 , L_{min}) were tuned on the basis of a large number of pre-tests according to the parameter identification problem considered here. After the tuning procedures, the main BA parameters were set as follows:

$$f_{min} = 0, f_{max} = 1, \alpha = 0.8, \gamma = 0.7,$$

$$L_0 = 1, L_{min} = 1, \text{NB} = 20, \text{NI} = 100.$$

The way in which the time in the hybrid algorithm is divided between the two methods influences the efficiency and effectiveness of the search process. Numerous tests were performed to find the optimal division of the algorithm time. The BA was run for 5, 10, 15, 20, 25, 30, 35 and 40 iterations before the SQP algorithm was started. The obtained results showed that the optimal number of iterations is 20. For 20 iterations, the BA reached near optimum solution, which is a good initial point for the SQP algorithm. The use of 30 or 40 iterations mainly affected the computational cost and had negligible improvement on the initial point. In the hybrid, a population of 20 bats was used.

Another meta-heuristic which was previously applied to this problem is ACO (6). We ran the ant algorithm proposed in Fidanova and Roeva (7) for the same conditions as BA, in order to compare the performance of the algorithms. The number of ants was 20 and the number of iterations was 100.

In all simulation procedures, a fixed tolerance of $\leq 10^{-4}$ was used. Each algorithm was run 50 times to achieve meaningful statistical analysis.

The numerical results from the parameter identification procedures are summarized in **Table 1**. Considering the application of BA, SQP, BA-SQP and ACO, the obtained average values of the three model parameters (μ_{max} , k_s and $1/Y_{S/X}$) of all 50 runs are presented.

TABLE 1

Obtained model parameters estimations

Parameter	BA	SQP	BA-SQP	ACO
μ_{max}	0.5056	0.4740	0.4992	0.4805
k_s	0.0214	0.0147	0.0205	0.0200
$1/Y_{S/X}$	2.0187	2.0137	2.0157	2.0100
J	6.1051	6.4610 ¹ / 6.7314 ²	5.9182	6.2614
CPU time, s	15.1429	5.4391 ¹ / 5.5021 ²	8.0232	63.6

¹"Good" initial solution, ²"bad" initial solution.

The obtained estimations of the model parameters values of the algorithms were within the admissible ranges (10, 18). The best average result ($J = 5.9182$) was obtained using the hybrid BA-SQP after a total computation time of 8.0232 s. The BA-SQP hybrid technique combines a lot of the features of BA with the SQP optimality criterion. The combination of the meta-heuristic algorithm and the local search method speeds up the search (twice, in comparison to conventional BA) to locate

the better solution. It exhibits the global search capability of the BA, while preserving the efficient local search capability afforded by SQP. A close result was obtained by conventional BA: $J = 6.1051$, but for an almost double computational time.

The SQP solutions strongly depend on the values of the initial solution, as this method usually seeks the neighborhood of the start point. If the initial solution is a “bad” one, the algorithm falls in some local extremum (see **Table 1**).

The ACO algorithm achieved an average solution similar to that obtained by BA after over 50 runs (slightly worse). The execution time was larger due to its constructive nature.

In **Fig. 1**, a comparison between the BA, SQP and BA–SQP model predictions for the process variables and the real experimental data for biomass and substrate concentrations are presented.

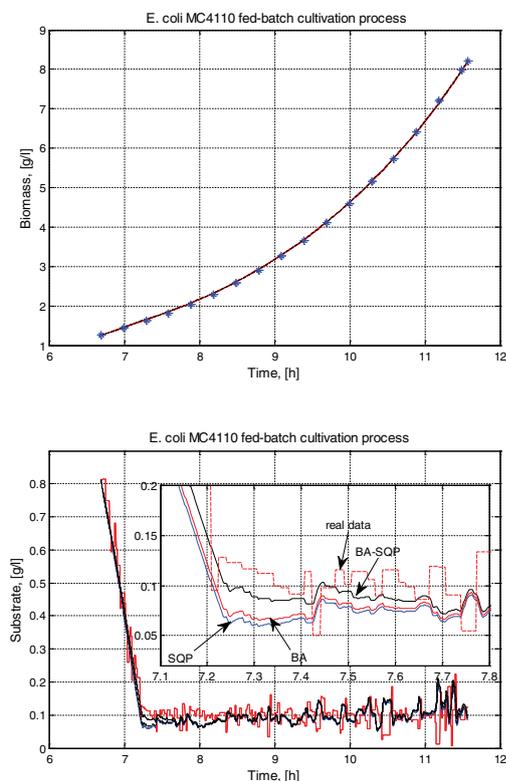


Fig. 1. Comparison between the model predictions and the real process variables.

All the three algorithms successfully model the dynamics of biomass concentration. Concerning the prediction of glucose concentration, the competing methods give different results. The difference is most notable in the first 1 or 2 hours of cultivation. As it can be seen from **Fig. 1**, the hybrid BA–SQP fits the real experimental data more precisely than both conventional BA and SQP.

Conclusions

In this paper, parameter identification of an *E. coli* MC4110 fed-batch cultivation process nonlinear model is proposed, using hybrid BA. To improve the performance of the conventional BA, a combination scheme using BA and SQP

method is introduced. Thus, optimizers work jointly to locate efficiently quality design points better than what either of these could alone. A comparison of proposed BA–SQP hybrid with conventional BA and SQP is done. The mathematical model is considered as a system of three ordinary differential equations describing the two main process variables, namely biomass and substrate. The statistical measures for the meta-heuristic approach were obtained, based on 50 independent runs. The numerical and graphical results showed that the best performance was obtained by the hybrid BA–SQP. It was shown that the proposed hybrid between meta-heuristics BA and exact algorithm SQP is significantly more effective in terms of running time (two times faster) and solution quality, since it benefits from synergy. The results showed that the hybrid BA–SQP algorithm outperforms the meta-heuristics (ACO), as well as the conventional SQP algorithm.

Acknowledgements

This work was partially supported by the National Science Fund of Bulgaria under the Grants DID 02-29 and DMU 02-4.

REFERENCES

1. **Altringham J.D.** (1996) *Bats: Biology and Behaviour*, Oxford University Press, New York.
2. **Arndt M., Hitzmann B.** (2001) In: 8th IFAC Int. Conf. on Comp. Appl. in Biotechn., Canada, 425-429.
3. **Bastin G., Dochain D.** (1991) *On-line Estimation and Adaptive Control of Bioreactors*, Elsevier Science.
4. **Blum C., Roli A.** (2008) *Stud. Comp. Intell.*, **114**, 1-30.
5. **Byrd R.H., Curtis F.E., Nocedal J.** (2010) *SIAM J. Optimiz.*, **20**(5), 2281-2299.
6. **Dorigo M., Stutzle T.** (2004) *Ant Colony Optimization*, MIT Press, Cambridge, MA.
7. **Fidanova S., Roeva O.** (2013) *Biotechnol. Biotech. Eq.*, **27**(3), 3870-3876.
8. **Gandomi A.H., Yang X.-S., Alavi A.H., Talatahari S.** (2013) *Neural Comput. Appl.*, **22**(6), 1239-1255.
9. **Gill Ph.E., Wong E.** (2010) *Sequential Quadratic Programming Methods*, UCSD Department of Mathematics, Technical Report NA-10-03.
10. **Levisauskas D., Galvanauskas V., Henrich S., Wilhelm K., et al.** (2003) *Bioproc. Biosyst. Eng.*, **25**, 255-262.
11. **Puchinger J., Raidl G. R.** (2005) *Lect. Notes Comput. Sc.*, **3562**, 41-53.
12. **Raidl G. R.** (2006) *Lect. Notes Comput. Sc.*, **4030**, 1-12.
13. **Roeva O., Pencheva T., Hitzmann B., Tzonkov St.** (2004) *International Journal of Bioautomation*, **1**, 30-41.
14. **Roeva O.** (2008) *Lect. Notes Comput. Sc.*, **4818**, 601-608.
15. **Toutouh J.** (2010) *Metaheuristics for Optimal Transfer of P2P Information in VANETs*, MSc Thesis, University of Luxembourg <<http://neo.lcc.uma.es/staff/jamal/downloads/vanet-project.pdf>>
16. **Yang X.-S., Gandomi A. H.** (2012) *Eng. Computation.*, **29**(5), 464-483.
17. **Yang X.-S.** (2010) *Stud. Comp. Intell.*, **284**, 65-74.
18. **Zelic B., Vasic-Racki D., Wandrey C., Takors R.** (2004) *Bioproc. Biosyst. Eng.*, **26**, 249-258.