Parameter identification algorithm for convection-diffusion transport model

A N Kuvshinova¹, A V Tsyganov¹, Yu V Tsyganova², H R Tapia Garza¹

¹Ulyanovsk State University of Education, Lenin square 4/5, Ulyanovsk, Russia, 43207
²Ulyanovsk State University, Lev Tolstoy street 42, Ulyanovsk, Russia, 432017

Abstract. This paper considers an application of gradient-type optimization algorithm for solving the problem of convection velocity and diffusion coefficient identification in the convection-diffusion transport model. A solution is obtained by the means of numerical minimization of the parameter identification criterion for a discrete linear stochastic model in state space. The log-likelihood function is used as the identification criterion. Numerical experiments were conducted to confirm the efficiency of the proposed solution.

1. Introduction and problem statement
Convection-diffusion transport models are an indispensable tool for describing various natural and anthropogenic processes [1], [2]. These models contain parameters that must be specified to uniquely determine the solution of boundary value problems. But in practice, situations often arise when some of these parameters are unknown or given approximately and they need to be determined or clarified.

In the simplest one-dimensional case, the convection-diffusion transport model can be described by equation (1) with initial condition (2) and boundary conditions (3):

\[ \frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} = \alpha \frac{\partial^2 c}{\partial x^2}, \quad a < x < b, \quad 0 < t < +\infty, \]

(1)

\[ c(x, 0) = \varphi(x), \quad a \leq x \leq b, \]

(2)

\[ c(a, t) = f_1(t), \quad c(b, t) = f_2(t), \quad 0 < t < +\infty, \]

(3)

where \( c(x, t) \) is the function of interest (for example, the concentration of the pollutant), \( x \) is the spatial variable, \( t \) is the time variable, \( v \) is the convection velocity, \( \alpha \) is the diffusion coefficient, \( \varphi(x) \), \( f_1(t) \), \( f_2(t) \) are given functions, \( a, b \) are the boundaries of the considered segment.

When solving a wide range of problems in ecology, geophysics, seismology, and other fields, the problem of identifying the coefficients of the equation often arises (1) in the convection-diffusion transport model. Depending on the equation under consideration and the boundary conditions, various methods can be used to solve this problem. In [3], [4] to find the coefficients of the equation (1) time series analysis methods based on the method of least squares, the extended Kalman filter, and their combinations are used. In [5], metaheuristic optimization algorithms are used to identify the convection velocity in equation (1).
In this paper, we propose the use of the gradient (iterative) method of numerical optimization to find the optimal estimate of a discrete linear stochastic model parameter according to a given criterion of identification quality.

The choice of a gradient algorithm to minimize the quality criterion of parameter identification is because the use of numerical methods is guaranteed under the conditions of convergence theorems [6].

2. Discrete linear stochastic model
We move from equations (1)–(3) to a discrete linear stochastic model represented in a general form by equations in state space:

\[
\begin{align*}
  c_k &= F(\theta)c_{k-1} + B(\theta)u_{k-1} + G(\theta)w_{k-1}, \\
  z_k &= H(\theta)c_k + \xi_k, \\
  k &= 1, 2, \ldots,
\end{align*}
\]  

(4)

where \( c_k \in \mathbb{R}^n \) is the system state vector, \( u_k \in \mathbb{R}^r \) is the input (control) vector, \( z_k \in \mathbb{R}^m \) is the measurements vector, noises \( w_k \in \mathbb{R}^q \) and \( \xi_k \in \mathbb{R}^m \) form independent normally distributed sequences with zero mean and covariance matrices \( Q(\theta) \geq 0 \) and \( R(\theta) > 0 \), respectively, matrices \( F(\theta) \in \mathbb{R}^{n \times m}, B(\theta) \in \mathbb{R}^{n \times r}, G(\theta) \in \mathbb{R}^{n \times q}, H(\theta) \in \mathbb{R}^{m \times n}, Q(\theta) \in \mathbb{R}^{q \times q}, R(\theta) \in \mathbb{R}^{m \times m} \) depend on the parameter \( \theta \).

In practice, for the numerical solution of non-stationary problems in convection-diffusion equations, two- and three-layer finite-difference schemes are most frequently used. To obtain a discrete linear stochastic model corresponding to the equations (1)–(3) consider in the plane \( Oxt \) a regular grid (5) with spatial step \( \Delta x \) and time step \( \Delta t \):

\[
x_i = a + i\Delta x, t_k = k\Delta t, i = 0, 1, \ldots, N, k = 0, 1, \ldots
\]  

(5)

We denote \( c_i^k = c(x_i, t_k), \varphi_i = \varphi(x_i), f_1^k = f_1(t_k), f_2^k = f_2(t_k) \) and write down the finite-difference scheme for (1)–(3):

\[
\frac{c_i^k - c_{i-1}^{k-1}}{\Delta t} + v \frac{c_{i+1}^{k-1} - c_{i-1}^{k-1}}{2\Delta x} = \alpha \frac{c_{i+1}^{k-1} - 2c_i^{k-1} + c_{i-1}^{k-1}}{\Delta x^2},
\]  

(6)

\[
i = 1, 2, \ldots, N - 1, k = 1, 2, \ldots,
\]

\[
c_0^i = \varphi_i, i = 0, 1, \ldots, N,
\]  

(7)

\[
c_0^k = f_1^k, c_N^k = f_2^k, k = 1, 2, \ldots
\]  

(8)

From equation (6) it follows that the value of the function \( c(x, t) \) at the internal points can be found through its values at the points of the previous time layer as follows:

\[
c_i^k = (r_1 + r_2)c_{i-1}^{k-1} + (1 - 2r_2)c_i^{k-1} + (r_2 - r_1)c_{i+1}^{k-1},
\]  

(9)

where \( r_1 = \frac{v\Delta t}{2\Delta x}, r_2 = \frac{\alpha\Delta t}{\Delta x^2} \). We rewrite (9) as

\[
c_i^k = a_1c_{i-1}^{k-1} + a_2c_i^{k-1} + a_3c_{i+1}^{k-1},
\]  

(10)

where \( a_1 = r_1 + r_2, a_2 = 1 - 2r_2, a_3 = r_2 - r_1. \)
The desirable discrete linear stochastic model can be represented as follows:

\[
\begin{bmatrix}
  c_1^k \\
  c_2^k \\
  \vdots \\
  c_{n-2}^k \\
  c_{n-1}^k \\
  c_n^k
\end{bmatrix} =
\begin{bmatrix}
  a_2 & a_3 & 0 & \cdots & 0 & 0 & 0 \\
  a_1 & a_2 & a_3 & \cdots & 0 & 0 & 0 \\
  0 & a_1 & a_2 & \cdots & 0 & 0 & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
  0 & 0 & 0 & \cdots & a_2 & a_3 & 0 \\
  0 & 0 & 0 & \cdots & a_1 & a_2 & a_3 \\
  0 & 0 & 0 & \cdots & 0 & a_1 & a_2
\end{bmatrix}
\begin{bmatrix}
  c_1^{k-1} \\
  c_2^{k-1} \\
  \vdots \\
  c_{n-2}^{k-1} \\
  c_{n-1}^{k-1} \\
  c_n^{k-1}
\end{bmatrix}
+ \begin{bmatrix}
  a_1 & 0 \\
  0 & 0 \\
  0 & 0 \\
  \vdots & \vdots \\
  0 & 0 \\
  0 & a_3
\end{bmatrix}
\begin{bmatrix}
  f_1^{k-1} \\
  f_2^{k-1} \\
  \vdots \\
  \vdots \\
  \vdots \\
  u_{k-1}
\end{bmatrix},
\]

(11)

Note that the first equation in the model (11) is deterministic, the initial value of the state vector is given by the equations (7) and the boundary conditions (8) act as control parameters. The state vector size is \( n = N - 2 \), i.e., the state vector \( c_k \) consists of all internal points of the grid. A similar model was considered in [7], where it was used for dynamic identification of boundary conditions, however, the input action vector \( u_{k-1} \) was supposed to be unknown.

We assume that the noise characteristics are known, and the steps of the space-time grid \( \Delta x \) and \( \Delta t \) are given. Then the unknown model parameters (11) to be determined are the convection velocity \( v \) and the diffusion coefficient \( \alpha \), which the coefficients \( a_1, a_2 \) and \( a_3 \) of the equation (10) and therefore the matrices \( F \) and \( B \), depend on. In this case, the model (11) can be briefly written as:

\[
\begin{align*}
  c_k &= F_{k-1}(\theta)c_{k-1} + B_{k-1}(\theta)u_{k-1}, \\
  z_k &= H_k c_k + \xi_k, \\
  k &= 1, 2, \ldots
\end{align*}
\]

(12)

where \( \theta = (v, \alpha)^T \in \mathbb{R}^2 \).

3. Gradient parameter identification methods

We consider the problem of parameter identification of the model (12) from noisy observable data to estimate an unknown (in a general case, vector) parameter. The goal of parameter identification is to find an unknown parameter \( \theta \) by known input signals \( U_0^{K-1} = \{u_0, u_1, \ldots, u_{K-1}\} \) and output data from observations \( Z_1^K = \{z_1, \ldots, z_K\} \) in accordance with the selected identification quality criterion \( J(\theta; Z_1^K, U_0^{K-1}) \). In this case, the problem of estimating an unknown parameter requires solving the nonlinear programming problem

\[
\hat{\theta}_{\text{min}} = J(\theta; Z_1^K, U_0^{K-1}),
\]

(13)

where \( \theta \in \mathbb{D}(\theta) \subseteq \mathbb{R}^p \).

Nonlinear programming methods include a large group of numerical methods that can be divided into two groups: direct and indirect [8]. The first group includes methods in which all the studied points belong to an admissible region. Such methods include gradient methods, the steepest-descent method, the Frank-Wolfe method, etc. The second group includes methods in which the points under investigation may not belong to an admissible region. A common of these methods is the penalty function method. All methods of penalty functions differ from each other in the way they determine the “penalty”.

3
We use gradient (iterative) algorithms to solve the problem (13). Gradient methods are approximate (iterative) methods for solving nonlinear programming problems and allow us to solve almost any problem. The basic concept used in all gradient methods is the concept of the gradient of a function as the direction of its quickest change.

We denote \( \nu_k = z_k - H_k\hat{c}_{k|k-1} \), \( c_k - \hat{c}_k \), \( P_{k|k-1} = \mathbb{E}\{c_k - \hat{c}_k|z_1, z_2, \ldots, z_{k-1}\} \), \( P_k = \mathbb{E}\{(c_k - \hat{c}_k)(\hat{c}_k - c_k)^T|z_1, z_2, \ldots, z_k\} \).

Then, as a function for implementing the gradient algorithm, we select the identification criterion (13) in the form of a logarithmic likelihood function [9]

\[
J_{MLF}(\theta; Z^K_1, U^K_0) = \frac{Km}{2} \ln(2\pi) + \frac{1}{2} \sum_{k=1}^{K} \{ \ln[\det(\Sigma_{\nu,k})] + \nu_k^T \Sigma_{\nu,k}^{-1} \nu_k \},
\]

(14)

where \( \Sigma_{\nu,k} = \mathbb{E}\{\nu_k\nu_k^T\} = R_k + H_kP_{k|k-1}H_k^T \).

We calculate the derivative with respect to the parameter \( \theta_i \) (\( i = 1, 2 \)):

\[
\frac{\partial J_{MLF}}{\partial \theta_i} = -\frac{1}{2} \sum_{k=1}^{K} \{ \text{tr}(\Sigma_{\nu,k}^{-1} \partial \Sigma_{\nu,k} / \partial \theta_i) + \nu_k^T \Sigma_{\nu,k}^{-1} \partial \nu_k / \partial \theta_i + \nu_k^T \Sigma_{\nu,k}^{-1} \nu_k \}.
\]

(15)

The required values \( \Sigma_{\nu,k} \) and \( \nu_k \) can be computed with the use of the Kalman filter [10].

I. Kalman filter equations

The classical Kalman filter consists of two stages: the time update stage (prediction) and the measurement update stage (correction). We write the Kalman filter equations for the model (11).

A. Time update (prediction)

For \( k = 1, 2, \ldots \) prediction state estimate \( \hat{c}_{k|k-1} \) and the prediction error covariance matrix \( P_{k|k-1} \) are calculated by the formulas (16) and (17) respectively:

\[
\hat{c}_{k|k-1} = F_{k-1}(\theta)\hat{c}_{k-1} + B_{k-1}(\theta)u_{k-1},
\]

(16)

\[
P_{k|k-1} = F_{k-1}(\theta)P_{k-1}F_{k-1}^T(\theta).
\]

(17)

B. Measurement update (correction)

The Kalman gain coefficients, the correction of the previously obtained prediction state estimate, and the error prediction covariance matrix are calculated by the formulas:

\[
K_k = P_{k|k-1}H_k^T(R_k + H_kP_{k|k-1}H_k^T)^{-1} = P_{k|k-1}H_k^T(\Sigma_{\nu,k})^{-1},
\]

(18)

\[
\hat{c}_k = \hat{c}_{k|k-1} + K_k(z_k - H_k\hat{c}_{k|k-1}) = \hat{c}_{k|k-1} + K_k\nu_k,
\]

(19)

\[
P_k = (I - K_kH_k)P_{k|k-1}.
\]

(20)

II. The sensitivity equations of the Kalman filter by the parameter \( \theta \)

Let \( \theta = (\theta_1, \theta_2)^T = (\nu, \alpha)^T \).

\[
\frac{\partial \hat{c}_{k|k-1}}{\partial \theta_i} = \frac{\partial F_{k-1}(\theta)}{\partial \theta_i} \hat{c}_{k-1} + F_{k-1}(\theta) \frac{\partial \hat{c}_{k-1}}{\partial \theta_i} + \frac{\partial B_{k-1}(\theta)}{\partial \theta_i} u_{k-1}.
\]

(21)

\[
\frac{\partial P_{k|k-1}}{\partial \theta_i} = \frac{\partial F_{k-1}(\theta)}{\partial \theta_i} P_{k-1}F_{k-1}^T(\theta) + F_{k-1}(\theta) \frac{\partial P_{k-1}(\theta)}{\partial \theta_i} F_{k-1}^T(\theta) + F_{k-1}(\theta) P_{k-1} \left( \frac{\partial F_{k-1}(\theta)}{\partial \theta_i} \right)^T,
\]

(22)
\[
\frac{\partial K_k}{\partial \theta_i} = \frac{\partial P_{k|k-1}}{\partial \theta_i} H^T \Sigma_{\nu,k}^{-1} - P_{k|k-1} H^T \Sigma_{\nu,k}^{-1} \frac{\partial \Sigma_{\nu,k}}{\partial \theta_i} \Sigma_{\nu,k}^{-1}, \quad (23)
\]
\[
\frac{\partial \Sigma_{\nu,k}}{\partial \theta_i} = H_k \frac{\partial P_{k|k-1}}{\partial \theta_i} H_k^T, \quad (24)
\]
\[
\frac{\partial c_k}{\partial \theta_i} = \frac{\partial c_{k|k-1}}{\partial \theta_i} + \frac{\partial K_k}{\partial \theta_i} \nu_k + K_k \frac{\partial \nu_k}{\partial \theta_i}, \quad (25)
\]
\[
\frac{\partial \nu_k}{\partial \theta_i} = -H_k \frac{\partial c_{k|k-1}}{\partial \theta_i}, \quad (26)
\]
\[
\frac{\partial P_k}{\partial \theta_i} = -\frac{\partial K_k}{\partial \theta_i} H_k P_{k|k-1} + (I - K_k H_k) \frac{\partial P_{k|k-1}}{\partial \theta_i}, \quad (27)
\]
\[
F(\theta) = \begin{bmatrix}
a_2 & a_3 & 0 & \ldots & 0 & 0 & 0 \\
a_1 & a_2 & a_3 & \ldots & 0 & 0 & 0 \\
0 & a_1 & a_2 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & a_2 & a_3 & 0 \\
0 & 0 & 0 & \ldots & a_1 & a_2 & a_3 \\
0 & 0 & 0 & \ldots & 0 & a_1 & a_2 \\
\end{bmatrix}, \quad B(\theta) = \begin{bmatrix}
a_1 & 0 \\
0 & 0 \\
0 & 0 \\
\vdots & \vdots \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\end{bmatrix}, \quad (28)
\]
\[
\frac{\partial F(\theta)}{\partial \theta_i} = \begin{bmatrix}
\frac{\partial a_2}{\partial \theta_1} & \frac{\partial a_3}{\partial \theta_1} & 0 & \ldots & 0 & 0 & 0 \\
\frac{\partial a_1}{\partial \theta_1} & \frac{\partial a_2}{\partial \theta_1} & \frac{\partial a_3}{\partial \theta_1} & \ldots & 0 & 0 & 0 \\
0 & \frac{\partial a_1}{\partial \theta_1} & \frac{\partial a_2}{\partial \theta_1} & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & \frac{\partial a_2}{\partial \theta_i} & \frac{\partial a_3}{\partial \theta_i} & 0 \\
0 & 0 & 0 & \ldots & \frac{\partial a_1}{\partial \theta_i} & \frac{\partial a_2}{\partial \theta_i} & \frac{\partial a_3}{\partial \theta_i} \\
0 & 0 & 0 & \ldots & 0 & \frac{\partial a_1}{\partial \theta_i} & \frac{\partial a_2}{\partial \theta_i} \\
\end{bmatrix}, \quad \frac{\partial B(\theta)}{\partial \theta_i} = \begin{bmatrix}
\frac{\partial a_1}{\partial \theta_i} & 0 \\
0 & 0 \\
0 & 0 \\
\vdots & \vdots \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
\end{bmatrix}. \quad (29)
\]

Given that \( \theta_1 = \nu \) and \( \theta_2 = \alpha \), we get \( \frac{\partial a_1}{\partial \nu} = \frac{\Delta t}{2\Delta x}, \quad \frac{\partial a_2}{\partial \nu} = 0, \quad \frac{\partial a_3}{\partial \nu} = -\frac{\Delta t}{2\Delta x}, \quad \frac{\partial a_1}{\partial \alpha} = \frac{\Delta t}{\Delta x^2}, \quad \frac{\partial a_2}{\partial \alpha} = -\frac{2\Delta t}{\Delta x^2} \) and \( \frac{\partial a_3}{\partial \alpha} = \frac{\Delta t}{\Delta x^2} \).

4. Numerical experiment

Consider the following problem:

\[
\frac{\partial c}{\partial t} + v \frac{\partial c}{\partial x} = \alpha \frac{\partial^2 c}{\partial x^2}, \quad 0 < x < \pi, 0 < t < +\infty, \quad (30)
\]
\[
c(x,0) = \exp\left(\frac{\nu t}{2\alpha}\right) \sin x, \quad 0 \leq x \leq \pi, \quad (31)
\]
\[
c(0,t) = 0, c(\pi,t) = 0, \quad 0 < t < +\infty, \quad (32)
\]

where \( c(x,t) \) is the concentration of the pollutant in a one-dimensional flow, (31) is the initial concentration of the pollutant, and boundary conditions (32) correspond to the case of absorbing walls.

Exact system solution (30)–(32):

\[
c(x,t) = \exp\left[ \frac{v}{2\alpha} \left( x - \frac{vt}{2} \right) \right] \sin x \exp(-\alpha t). \quad (33)
\]
Let $v = 2, \alpha = 1$. The exact solution graph for this case is shown in figure 1.

Now, suppose the values $v$ and $\alpha$ in equation (30) are unknown, and it is required to determine them according to the data of noisy measurements coming from sensors located in the internal points of a certain regular grid. Let a grid with 10 points be given ($N = 10, n = 8, \Delta x = \frac{\pi}{10}$) along the $Ox$ axis with time step $\Delta t = \frac{\Delta x^2}{4\alpha}$.

In the model (11), measurements are taken at each internal point of the grid, i.e., $H = I_{8 \times 8}$, and measurement error $\xi_k$ is modeled by a sequence of random variables of the normal distribution with zero mean and a covariance matrix $R = 0.003I_{8 \times 8}$. The identification of an unknown vector parameter $\theta = (v, \alpha)^T$ was carried out in Matlab using the minimization function `fminunc`. The estimate $\hat{\theta} = (\hat{v}, \hat{\alpha})^T$ of the parameter $\theta$ was given the initial value $\theta_0 = (2.5, 1.5)^T$. Table 1 presents the identification results from 100 observations for a series of 10 experiments.

Table 1. Parameter identification results

| MEAN$\hat{v}$ | MEAN$\hat{\alpha}$ | RMSE$\hat{v}$ | RMSE$\hat{\alpha}$ | MAPE$\hat{v}$ | MAPE$\hat{\alpha}$ | $||\theta - \hat{\theta}||$ | $\frac{||\theta - \hat{\theta}||}{||\theta||} \cdot 100\%$ |
|---------------|---------------------|---------------|---------------------|---------------|---------------------|--------------------------|----------------------------------|
| 1.9441        | 1.1807              | 0.2991        | 0.2586              | 12.8972       | 18.0657             | 0.3341                   | 14.9401                          |
5. Conclusion
The paper proposes an approach to the numerical identification of the convection velocity and the diffusion coefficient in the convection-diffusion transport model under noisy measurements conditions, based on the transition to a linear discrete stochastic model in state space and the use of the gradient (iterative) method of numerical minimization of the logarithmic likelihood function. Further research will be aimed at the development of new numerical methods for identifying parameters of the convective-diffusion transport models.

6. Acknowledgments
The reported study was funded by RFBR and Ulyanovsk region, project number №9-41-730009.

7. References
[1] Leontiev A I 1997 The Theory of Heat and Mass Transfer (Moscow: Bauman Moscow State Technical University) 683.
[2] Farlow S J 1982 Partial Differential Equations for Scientists and Engineers (New York: Wiley) 402.
[3] Matveev M G, Sirota E A, Semenov M E and Kopytin A V 2017 Verification of the convective diffusion process based on the analysis of multidimensional time series (DAMID/RCDL) 433-437.
[4] Matveev M G, Kopytin A V and Sirota E A 2018 Combined method for identifying the parameters of a distributed dynamic model Proc. IV Int. Conf. on Information Technology and Nanotechnology 1651-1657.
[5] Tsyganov A V, Tsyganova Yu V, Kuvshinova A N and Tapia Garza U R 2018 Metaheuristic algorithms for identification of the convection velocity in the convection-diffusion transport model CEUR Workshop Proceedings 188-196.
[6] Vasilyev V P 1982 Numerical Methods for Solving Extreme Problems (M: Mir) 372.
[7] Tsyganov A V, Tsyganova Yu V and Kuvshinova A N 2019 J. Phys.: Conf. Series 1368 042029.
[8] Denisov A M 1994 Introduction to the Theory of Inverse Problems (Moscow: Moscow State University) 208.
[9] Astrom K J 1980 Automatica 16(5) 551-574.
[10] Grewal M S 2001 Theory and Practice Using Matlab (New Jersey: Prentice-Hall) 401.