Research Article

Olga Krivorotko*, Sergey Kabanikhin, Shuhua Zhang and Victoriya Kashtanova

Global and local optimization in identification of parabolic systems

https://doi.org/10.1515/jiip-2020-0083
Received July 7, 2020; accepted August 20, 2020

Abstract: The problem of identification of coefficients and initial conditions for a boundary value problem for parabolic equations that reduces to a minimization problem of a misfit function is investigated. Firstly, the tensor train decomposition approach is presented as a global convergence algorithm. The idea of the proposed method is to extract the tensor structure of the optimized functional and use it for multidimensional optimization problems. Secondly, for the refinement of the unknown parameters, three local optimization approaches are implemented and compared: Nelder–Mead simplex method, gradient method of minimum errors, adaptive gradient method. For gradient methods, the evident formula for the continuous gradient of the misfit function is obtained. The identification problem for the diffusive logistic mathematical model which can be applied to social sciences (online social networks), economy (spatial Solow model) and epidemiology (coronavirus COVID-19, HIV, etc.) is considered. The numerical results for information propagation in online social network are presented and discussed.

Keywords: Inverse problem, parameter estimation, optimization, regularization, gradient method, social network, partial differential equations, tensor train, tensor train decomposition

MSC 2010: 35R30, 65N21, 65K10, 35K61, 91C15

1 Introduction

We consider the initial boundary value problem for the system of partial differential equations of parabolic type

$$\frac{\partial y_j}{\partial t} = d_j \frac{\partial^2 y_j}{\partial x^2} + \mu_j(y, \varphi), \quad t \in (t_0, T), \quad x \in (l, L), \tag{1.1a}$$

$$y_j(x, t_0) = \psi_j(x), \qquad x \in (l, L), \tag{1.1b}$$

$$\left.\frac{\partial y_j}{\partial x}\right|_{x=l} = \left.\frac{\partial y_j}{\partial x}\right|_{x=L} = 0, \qquad t \in (t_0, T). \tag{1.1c}$$

Here $j = 1, \dots, N$, $y(x, t) = (y_1(x, t), \dots, y_N(x, t))$, $d_j \geq 0$ and $\varphi(t) = (\varphi_1, \dots, \varphi_M)$ are unknown coefficients, $\psi_j(x)$ are unknown initial conditions, $l > 0$, μ_j is a continuous function.

Suppose that the additional information can be measured:

$$y_j(x_i, t_k; q) = F^j_{ik}, \quad k = 1, \dots, K, \ i = 1, \dots, N_x, \ j \in J \subset \{1, \dots, N\}. \tag{1.2}$$

***Corresponding author: Olga Krivorotko,** Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Novosibirsk State University, Novosibirsk, Russia, e-mail: krivorotko.olya@mail.ru. http://orcid.org/0000-0003-0125-4988
Sergey Kabanikhin, Institute of Computational Mathematics and Mathematical Geophysics of SB RAS, Sobolev Institute of Mathematics of SB RAS, Novosibirsk State University, Novosibirsk, Russia, e-mail: ksi52@mail.ru
Shuhua Zhang, Tianjin University of Finance and Economics, Tianjin, P. R. China, e-mail: shuhua55@126.com
Victoriya Kashtanova, Novosibirsk State University, Novosibirsk, Russia, e-mail: vikakashtanova@ya.ru

We investigate the problem of identification of coefficients and initial data $q = (\varphi(t), d, \psi(x)) \in Q$ (inverse problem) from (1.1a)–(1.1c) using additional information F^j_{ik} (1.2). Here Q is the set of admissible solutions of the inverse problem. Three applications of mathematical model (1.1a)–(1.1c) that describe information propagation in online social networks, stock dynamics in the economy and infectious disease propagation in epidemiology (tuberculosis, HIV, Wuhan coronavirus 2019-nCoV, etc.) are given as examples in Section 3.

Inverse problem (1.1a)–(1.1c), (1.2) can be rewritten in operator form $A(q) = f$, where $A\colon Q \to F$ is an injective operator, $q \in Q$, $f \in F$, F is a Euclidean space of data $f = (F^j_{11}, \dots, F^j_{N_x K})_{j\in J}$. The inverse problem $A(q) = f$ is ill-posed for many applications, i.e. its solution may not exist and/or its solution is non-unique and/or unstable to errors in measurements (1.2) [21, 26]. In the current paper, the inverse problem is reduced to the minimization problem

$$q^\star = \operatorname*{argmin}_{q\in Q} J(q), \quad J(q) = \langle A(q) - f, A(q) - f\rangle, \tag{1.3}$$

where the functional $J(q)$ characterizes the quadratic deviation of the model data from the experimental ones (see Section 3).

There exist many optimization methods for solving minimization problem (1.3) that could be divided into three groups: local, global and hybrid (see Section 4 for more details). But in real applications, the vector of parameters q could be huge and classical nature inspired algorithms (genetic algorithm, simulation annealing, particle swarm optimization, etc.), gradient approaches and so on require large computing resources and do not guarantee finding a global minimum. We apply a more powerful approach for solving multidimensional minimization problem (1.3) named tensor train decomposition that is based on a tensor representation of the initial functional and applying the tensor decomposition properties and tensor interpolations [32–34, 41]. The advantages of this method consist in implementation of main algebra operations and multi-parallel architecture for the computation. Since tensor train optimization finds the area of global minima of (1.3), we apply and compare three local methods: Nelder–Mead that does not require the gradient of functional $J(q)$ and two types of gradient approaches for which we derive the evident formula of the gradient of functional $J(q)$ that is based on a solution of the corresponding adjoint problem (Section 4).

We numerically investigate the problem of parameter identification in the mathematical model of social networks based on a parabolic equation. The comparison and domain of applicability of the investigated methods are analyzed and discussed (Section 5).

2 Brief historical review of mathematical modeling in social sciences

The specifics of the dissemination of information in society and the development of socially significant diseases (tuberculosis, HIV/AIDS, coronavirus COVID-19) depend on the region. However, the statistically calculated parameters (for example, the probability of the appearance of information in the social network, the rate of infection development, the mortality parameters, etc.) are averages. The investigation of model populations, on the one hand, can lead to erroneous conclusions (since real populations are significantly heterogeneous) and, on the other hand, it cannot serve as a tool for assessing the current situation.

One of the most effective methods of monitoring and managing social and epidemiological processes is mathematical modeling, namely the development and identification of mathematical models that describe the processes of information dissemination in social networks and infections in the population. Such models are described by systems of differential equations, the coefficients of which characterize the distribution of information, population and disease development. To control information in social networks and epidemics in individual regions and economic processes, it is necessary to refine the model coefficients and initial data by some additional information.

Ordinary differential equations. In recent years, online social networks, such as Twitter and Facebook, have become the main source of information exchange. A large amount of data available to researchers

has increased interest in studying the process of information dissemination in online networks. One of the approaches to construct mathematical models of social processes is the principle of constructing mathematical models of epidemiological processes (in particular, describing socially significant diseases, such as tuberculosis, HIV/AIDS) based on the chamber structure and probabilistic transitions between homogeneous groups [12] (mathematical models of Wuhan coronavirus 2019-nCoV [8, 39], tuberculosis [3, 6, 36], HIV [16]). The spread of the virus in a homogeneous network based on the epidemiological model was investigated in [20]. Such mathematical models are based on systems of ordinary differential equations (mathematical model (1.1a)–(1.1c) with $d_j \equiv 0$), the coefficients of which in many cases are unknown or given approximately. This leads to solving of inverse problems. In many models, it is assumed that the social system is homogeneous, and individuals are in equal status (the same degree of spread, the probability of infection, etc.), which does not take into account the specifics of the process. This entails the development of more complex mathematical models.

Partial differential equations. To build a more complete picture of the development of social processes, it is necessary to take into account migration, age data and changes in time. Such models are described by partial differential equations (PDE). F. Wang, H. Wang and K. Xu [38] proposed to use partial differential equations built on intuitive cyber-distance among online users to study both temporal and spatial patterns of information diffusion process in online social networks. A detailed review of mathematical models for social networks is given in [19]. The PDE-based models for online social networks in [30] are spatial dynamical systems (1.1a)–(1.1c) that take into account the influence of the underlying network structure as well as information contents to predict information diffusion over both temporal and spatial dimensions. In the paper [10], a non-autonomous diffusive logistic model with indefinite weight and Robin boundary condition is developed to describe information diffusion in online social networks. The model is validated with a real dataset from an online social network, Digg.com, and the simulation shows that the logistic model with Robin boundary condition is able to more accurately predict the density of influenced users.

Stochastic differential equations. Real social systems are always exposed to external influences that are not completely understandable or impossible for explicit models (enzymatic processes, energy needs, smoking, stress effects, information wars, etc.), and therefore there is a growing need to expand deterministic models to models which cover more complex variations in dynamics. A natural continuation of the models of deterministic differential equations is a system of stochastic differential equations, where the corresponding parameters are modeled as suitable random processes, or stochastic processes are added to the equations of the motion system. An analysis of stochastic differential equations and numerical studies of the solution of a direct problem are given in the works of P.-L. Lions [28, 29] and H. T. Banks [4, 5].

However, each social network has its own platform and structure, and therefore the parameters that characterize these indicators vary. For the best result of modeling and information control, it is necessary to identify the coefficients and initial conditions (inverse problem) for each specific case. In the papers of S. I. Kabanikhin, O. I. Krivorotko and co-authors [22, 23], the inverse problem for ordinary differential equations describing the spread of tuberculosis and HIV in the regions of the Russian Federation and the inverse problem of immunology (intracellular dynamics of HIV) [24] were solved. An analysis of the structural identifiability of the mathematical models under study, which makes it possible to identify the region of the correctness of the inverse problem, is given in the paper [25].

We concentrate on PDE-based mathematical models arising in social sciences such as social networks, epidemiology/immunology, economics.

3 Examples and variational formulation of inverse problem for PDE

Mathematical model (1.1a) can describe the dynamics of social, economical and epidemiological processes. For example, in *social networks*, $j = 1$ and parameters and functions in problem (1.1a)–(1.1c) have the following interpretation [10]:

- $y(x, t)$ is a density of influenced users with a distance of x at time t;

- d represents the popularity of information which promotes the spread of the information through non-structure based activities such as search;
- $\mu(y, \varphi)$ is a local growth function (death and birth) that has, e.g., the form $\mu(y, \varphi) = r(t)y(x, t)(1 - \frac{y(x,t)}{K_{\text{cap}}})$;
- $r(t)$ represents the intrinsic growth rate of influenced users with the same distance and measures how fast the information spreads within the users with the same distance;
- K_{cap} is a carrying capacity, which is the maximum possible density of influenced users at a given distance;
- l and L represent the lower and upper bounds of the distances between the source and other social network users;
- $\psi(x) \geq 0$ is the unknown initial density function; each information has its own unique initial function.

Mathematical model (1.1a) arises in modern *economy* and is called the spatial Solow model [7, 35] ($j = 1$):

- $y(x, t)$ denotes the capital stock held by the representative household located at x and time t;
- $\mu(y, \varphi) = A(x, t)f(y(x, t)) - \varepsilon y(x, t)$;
- $A(x, t)$ denotes the technological level at x and time t;
- $f(y(x, t))$ is the production function that is assumed to be non-negative, increasing and concave;
- ε is the depreciation rate;
- $\psi(x) \geq 0$ is an initial capital distribution.

PDE (1.1a) for age-structured populations describes the *epidemiological* process and is called Kermack–McKendrick equation [27]:

- $y_j(x, t) \geq 0$ represents the density of the population (infected and recovered individuals of different groups) of age x and time t;
- $\mu_j(y, \varphi)$ is a local growth function that describes the behavior of the j-th group;
- the sparsity coefficient d_j characterizes the migration at the j-th group;
- $\psi_j(x) \geq 0$ is an initial population at the j-th group.

Inverse problem (1.1a)–(1.1c), (1.2) is ill-posed, and most often its solutions are unstable to small perturbations in the data [21]. One of the types of solving ill-posed problems is reducing inverse problems to a variational formulation (1.3). For this purpose, we consider the misfit function

$$J(q) = \gamma \sum_{j \subset J} \sum_{k=1}^{K} \sum_{i=1}^{N_x} |y_j(x_i, t_k; q) - F_{ik}^j|^2, \quad \gamma = \frac{(T - t_0) \cdot (L - l)}{K \cdot N_x}, \tag{3.1}$$

that should to be minimized. Finally, the solution $q^\star$ of inverse problem (1.1a)–(1.1c), (1.2) can be considered as the solution of an optimization problem [21]:

$$q^\star = \operatorname*{argmin}_{q \in Q} J(q). \tag{3.2}$$

4 Optimization methods

There are several methods for solving minimization problem (3.1), and they can be split into three groups: local, global and hybrid optimization methods. An extensive review of the existing methods, including their classifications and properties, can be found in [1]. There are two groups of local optimization methods: the direct ones which do not require the derivative of the objective function (e.g., Hooke–Jeaves method, Nelder–Mead method) and the gradient based methods (e.g., the Gauss–Newton and Levenberg–Marquard methods). Although the local optimization methods usually work fast and their convergence to a local minimum can be proved, their main drawback is that they can miss the global minimum, especially in the case of a high dimensionality of the parameter space. Global optimization methods can be used to explore rather large regions of the parameter space. However, they usually are slow and have no theoretical proof of convergence. They can be formulated as stochastic (simulated annealing, evolutionary algorithms) or deterministic (covering methods) techniques [1] such as the method of nonuniform coverings proposed by Y. G. Evtushenko for functions that comply with the Lipschitz condition [13]. Evolutionary algorithms are usually most suitable for a large search space because they escape local minima and are intrinsically parallel. The covering methods require

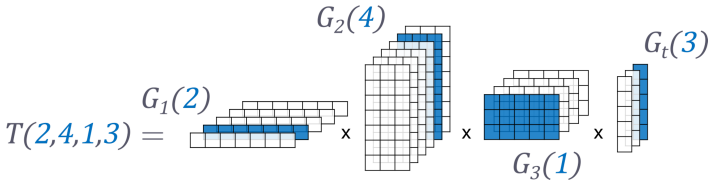

Figure 1: Illustration of the evaluation of one element of a fourth-order tensor (with four indices) having a tensor train decomposition. G_1, G_2, G_3, G_t are the tensor carriages of this decomposition. $T(2, 4, 1, 3) \in \mathrm{R}$ is obtained by the product of one row vector $G_1(2)$, two matrices $G_2(4)$ and $G_3(1)$, and one column vector $G_t(3)$. The dimensions of these matrices are respectively 1×7, 7×3, 3×5 and 5×1. $G_1(2)$, $G_2(4)$, $G_3(1)$ and $G_t(3)$ are the extraction of one layer (identified by a darker shade) in the tensor G_1, G_2, G_3, G_t, respectively.

some prior information about the function and can locate the optima with the given accuracy. For example, the main idea of the method of nonuniform coverings consists in division of the solution set into subsets whose union coincides with the original set. The misfit function on different subsets has certain properties (it satisfies the Lipschitz condition, convexity, existence of the second derivative, etc.), which speeds up the calculations [14, 15]. Hybrid methods are based on the following idea: global optimization is used to explore the parameter space to locate the starting points for further local optimization [23]. To reduce the complexity of the parameter estimation task, there exist several support techniques such as constraining the parameter space, data smoothing and others [9].

The majority of the state-of-the-art global optimization methods are based on the tensor representation of an initial functional and applying the tensor decomposition properties and tensor interpolations [32–34], and we use it in numerical calculations and show its advantages.

4.1 Tensor train decomposition

Optimization of functional (3.1) is a complex problem because it is a problem of global optimization for a multidimensional, usually non-convex functional with large dimension and a huge number of local minimums. The tensor train (TT) approach [33] is based on decomposition of a multidimensional tensor $T \in R^{n_1 \times n_2 \times \cdots \times n_p}$ to the "train" of the *carriages* tensors,

$$T(i_1; i_2; \ldots; i_p) = \sum_{\alpha_0=1;\ldots;\alpha_p=1}^{r_0;\ldots;r_p} G_1(\alpha_0; i_1; \alpha_1) \cdot \ldots \cdot G_p(\alpha_{p-1}; i_p; \alpha_p).$$

An example of TT decomposition is illustrated in Figure 1. Although this approach is metaheuristic and cannot guarantee that the global optimum is reached, it uses structure of the functional and, generally, works faster and more robustly than other metaheuristic and stochastic methods [41].

4.1.1 TT algorithm for optimization problem

Consider global minimization problem (3.2). This problem could be transformed to an equivalent problem of the magnitude maximization of the continuous and monotonous mapping of J to the interval $[0; +\infty)$,

$$q^\star = \operatorname*{argmin}_{q \in Q} |g(q)|, \quad g(q) = \operatorname{arcctg}\{J(q)\}. \tag{4.1}$$

Here Q is a initial domain of the function J and can be represented as a p-dimensional parallelepiped ($p = M + N + 1$). So $q = (q_1, \ldots, q_p)^T$ is a vector, and suppose that each element q_j lies in the interval $[a_j, b_j]$, $j = 1, \ldots, p$. Introduce the uniform grid on each interval $[a_j, b_j]$ with fixed step $h_j = \frac{b_j - a_j}{n-1}$, $j = 1, \ldots, p$ and n nodes in each direction.

The misfit function values on the grid form the tensor $T \in R^{n \times \cdots \times n}$ with elements

$$T(i_1, \ldots, i_p) = g(q_1^{(i_1)}, \ldots, q_p^{(i_p)}).$$

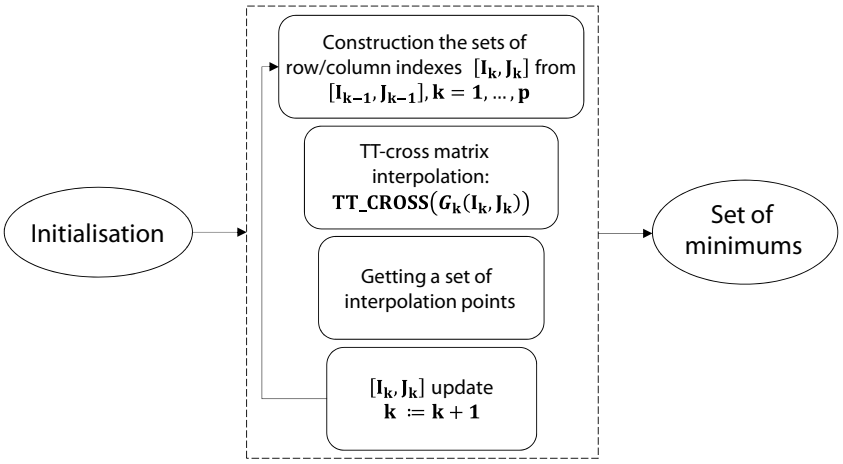

Figure 2: The general scheme of TT decomposition algorithm.

Here $q_j^{(i_j)}$ is the i_j-th point on the grid for element q_j. Introduce the vector of indexes $i = (i_1, \dots, i_p)$, and rewrite the formula as $T(i) = g(q^i)$. Then the discrete minimization problem is represented in the form

$$i = \underset{j=1,\dots,n^p}{\operatorname{argmin}} |T(j)|. \tag{4.2}$$

If the grid is fine enough, then the solutions of (4.1) and (4.2) are expected to be close.

Problem (4.2) consists in finding the maximal magnitude element of a p-dimensional tensor $T \in R^{n\times\cdots\times n}$. For example, if a number of coefficients is equal to 20 and a number of mesh nodes in space x is equal to 100, then the number of unknown parameters is equal to 121. If $n = 100$, then the number of elements of tensor T is equal to 100^{121}.

In this case, to reduce the problem complexity, the technique based on the TT cross interpolation machinery, which exploits the matrix cross interpolation algorithm [17, 18, 37] applied to heuristically selected submatrices in the unfolding matrices of the given tensor, could be used. The obtained method, named TT algorithm, takes only $O(pnr_{\max}^3)$ arithmetic operations, $O(pnr_{\max}^2)$ function calculations and $O(nr_{\max})$ local optimizations, where $r_{\max}$ is the maximum rank of the used tensors.

The TT global optimization method iteratively performs the following steps:

- already inspected points are used to generate submatrices of the unfolding matrices;
- these submatrices are approximated by the matrix cross approximation method with rank bounded from above by $r_{\max}$;
- the interpolation points and local minima in their vicinity (projected to the grid) are used to form new sets of "hopefully better" points;
- the sets of points are extended by the points from "neighboring" unfolding matrices and by $r_{\max}$ points considered as the best of all inspected values.

The general scheme of this method is presented in Figure 2; a more detailed description of the algorithm is available in [40, 41].

4.2 Gradient type methods

Most gradient methods are reduced to the iteration process $q_{n+1} = q_n - \alpha_n J'(q_n)$, where α_n is a descent parameter, $J'(q_n)$ is the gradient of the misfit function at point q_n. Note that convergence of the gradient method depends on the initial approximation q_0, and as a result, one has the local convergence theorem [21, 26].

The dependence on the vector of parameters q shows at $y(x, t; q) := y(x, t)$. Function $\hat{y}\colon (t_0, T) \mapsto H^2(l, L)$ is a mapping, associated by the function $y(x, t)$, that is, $[\hat{y}(t)](x) := y(x, t)$. Suppose that $\psi, \psi + \delta\psi \in H^2(l, L)$, $\delta y(x, t; \delta q) := y(x, t; q + \delta q) - y(x, t; q)$, where $y(x, t; q) \in C(t_0, T; H^2(l, L))$ is a solution of problem (1.1a)–

(1.1c) with $\psi \in H^2(l, L)$. Then the deviation $\delta y := \delta y(x, t; \delta q)$ satisfies the following initial boundary value problem with an accuracy up to the terms of order $o(|\delta q|^2)$:

$$\begin{cases} \mathcal{L}\delta y_j = 0, & t \in (t_0, T), \quad x \in (l, L), \\ \delta y_j(x, t_0) = \delta\psi_j(x), & x \in (l, L), \\ \left.\dfrac{\partial \delta y_j}{\partial x}\right|_{x=l} = \left.\dfrac{\partial \delta y_j}{\partial x}\right|_{x=L} = 0, & t \in (t_0, T). \end{cases}$$

Here

$$\mathcal{L}\delta y_j := \frac{\partial \delta y_j}{\partial t} - d_j \frac{\partial^2 \delta y_j}{\partial x^2} - a_j(\delta y, \delta\varphi, y, \delta d),$$

$$a_j(\delta y, \delta\varphi, y, \delta d) = \sum_{n=1}^{N} P_{jn}\delta y_n + \sum_{m=1}^{M} R_{jm}\delta\varphi_m + \delta d_j \frac{\partial^2 y_j}{\partial x^2}, \quad j = 1, \dots, N,$$

$P = \left\{\frac{\partial \mu(y,\varphi)}{\partial y}\right\}$ and $R = \left\{\frac{\partial \mu(y,\varphi)}{\partial \varphi}\right\}$ are Jacobi matrices of vector function μ at spaces $\mathbf{R}^N \times \mathbf{R}^N$ and $\mathbf{R}^N \times \mathbf{R}^M$, respectively.

Proposition 1. *The gradient of misfit function* (3.1) *has the form*

$$J'(q) = \left(\int_l^L R^T \Psi(x, t)\, dx, \int_l^L \int_{t_0}^T \frac{\partial^2 y}{\partial x^2}(x, t)\Psi(x, t)\, dt\, dx, \Psi(x, t_0) \right)^T, \tag{4.3}$$

where vector function $\Psi(x, t)$ *satisfies the adjoint problem* ($j = 1, \dots, N$)

$$\begin{cases} \dfrac{\partial \Psi_j}{\partial t} = -d_j \dfrac{\partial^2 \Psi_j}{\partial x^2} - \sum\limits_{n=1}^{N} P_{nj}\Psi_n + B_j, & t \in (t_0, T), \quad x \in (l, L), \\ \Psi_j(x, T) = 0, & x \in (l, L), \\ \left.\dfrac{\partial \Psi_j}{\partial x}\right|_{x=l} = \left.\dfrac{\partial \Psi_j}{\partial x}\right|_{x=L} = 0, & t \in (t_0, T). \end{cases} \tag{4.4}$$

Here

$$B_j = 2\gamma \sum_{k=1}^{K} \sum_{i=1}^{N_x} \int_{t_0}^{T} \int_{l}^{L} (y_j(x, t; q) - F_{ik}^j)\delta(t - t_k)\delta(x - x_i)\, dx\, dt,$$

and $\delta(t - t_k)$ *is a Dirac delta function.*

Proof. Consider misfit function (3.1) variation, and apply the rule $a^2 - b^2 = (a - b)(a + b)$,

$$\begin{aligned} \delta J := J(q + \delta q) - J(q) &= \gamma \sum_{k=1}^{K} \sum_{i=1}^{N_x} 2(y(x_i, t_k; q) - F_{ik})\delta y(x_i, t_k; q) + \gamma \sum_{k=1}^{K} \sum_{i=1}^{N_x} |\delta y(x_i, t_k; \delta q)|^2 \\ &= \gamma \sum_{k=1}^{K} \sum_{i=1}^{N_x} \int_{t_0}^{T} \int_{l}^{L} 2(y(x, t; q) - F_{ik})\delta y(x, t; q)\delta(t - t_k)\delta(x - x_i)\, dx\, dt \\ &\quad + \gamma \sum_{k=1}^{K} \sum_{i=1}^{N_x} |\delta y(x_i, t_k; \delta q)|^2. \end{aligned} \tag{4.5}$$

Consider the scalar product in $L_2((0, T) \cup (l, L))$ space

$$\langle \mathcal{L}\delta y_j, \Psi \rangle = \left\langle \frac{\partial \delta y_j}{\partial t}, \Psi_j \right\rangle - d_j \left\langle \frac{\partial^2 \delta y_j}{\partial x^2}, \Psi_j \right\rangle - \langle a_j(\delta y, \delta\varphi, y, \delta d), \Psi_j \rangle. \tag{4.6}$$

Write in detail each term of the equation (4.6). Using differentiation by parts and the initial conditions of direct problem (1.1b) and adjoint problem (4.4), we get the following expression for the first term:

$$\left\langle \frac{\partial \delta y_j}{\partial t}, \Psi_j \right\rangle = \int_l^L \int_{t_0}^T \frac{\partial \delta y_j}{\partial t}\Psi_j\, dt\, dx = -\left\langle \delta y_j, \frac{\partial \Psi_j}{\partial t} \right\rangle - \int_l^L \delta\psi_j(x)\Psi_j(x, t_0)\, dx.$$

The second term in (4.6) after using differentiation by parts twice and the boundary conditions of direct problem (1.1c) and adjoint problem (4.4) has the form

$$d_j\Big\langle \frac{\partial^2 \delta y_j}{\partial x^2}, \Psi_j \Big\rangle = d_j \int_l^L \int_{t_0}^T \frac{\partial^2 \delta y_j}{\partial x^2} \Psi_j \, dt \, dx = \Big\langle \delta y_j, d_j \frac{\partial^2 \Psi_j}{\partial x^2} \Big\rangle.$$

The last term in (4.6) is rewritten as follows using the linearity of a scalar product:

$$\begin{aligned}\langle a_j(\delta y, \delta\varphi, y, \delta d), \Psi_j\rangle &= \Big\langle \sum_{n=1}^N P_{jn}\delta y_n, \Psi_j \Big\rangle + \Big\langle \sum_{m=1}^M R_{jm}\delta\varphi_m, \Psi_j \Big\rangle + \Big\langle \delta d_j \frac{\partial^2 y_j}{\partial x^2}, \Psi_j \Big\rangle \\ &= \Big\langle \delta y_j, \sum_{n=1}^N P_{nj}\Psi_n \Big\rangle + \Big\langle \delta\varphi_j, \sum_{m=1}^N R_{mj}\Psi_m \Big\rangle + \delta d_j \Big\langle \frac{\partial^2 y_j}{\partial x^2}, \Psi_j \Big\rangle.\end{aligned}$$

Collecting all equations to expression (4.6), using adjoint problem (4.4) and noting that formula (4.6) is equal to zero, we get the following equation for $j = 1, \dots, N$:

$$B_j \delta y_j(x, t; q) = \int_l^L \delta\psi_j(x)\Psi_j(x, t_0)\, dx + \delta d_j \int_{t_0}^T \int_l^L \frac{\partial^2 y_j}{\partial x^2}\Psi_j \, dx \, dt + \int_{t_0}^T \delta\varphi_j \int_l^L \sum_{m=1}^N R_{mj}\Psi_m \, dx \, dt.$$

Using formula (4.5), note that

$$\sum_{j=1}^N B_j \delta y_j(x, t; q) = \delta J - \gamma \sum_{k=1}^K \sum_{i=1}^{N_x} |\delta y(x_i, t_k; \delta q)|^2.$$

Suppose that $\gamma \sum_{k=1}^K \sum_{i=1}^{N_x} |\delta y(x_i, t_k; \delta q)|^2 \approx o(\|\delta q\|^2)$. After comparison with the Freshet derivative formula for the misfit function $\delta J = \langle J', \delta q\rangle + o(\|\delta q\|^2)$, we get the gradient $J'(q)$ (see (4.3)) from the space $\mathbf{R}^{M+N+1}$. □

4.2.1 Algorithm of the gradient method of minimum errors

The type of the gradient method depends on descent parameter α_n. We propose the algorithm of a gradient method of minimum errors [23] as follows (Algorithm 1).

(1) Set an initial approximation vector q_0 and stopping parameter $\varepsilon > 0$. Suppose that we have q_n. Show how to get the next approximation q_{n+1}.
(2) Check the stop condition: if $J(q_n) < \varepsilon$, then q_n is an approximate solution of inverse problem (1.1a)–(1.1c), (1.2). Otherwise, go to step (3).
(3) Solve direct problem (1.1a)–(1.1c) for a given set of the parameters q_n by an explicit finite difference scheme of second-order approximation, and get $y_j(x_i, t_k; q_n)$, $i = 1, \dots, N_x$, $k = 1, \dots, K$, $j \subset J$.
(4) Solve adjoint problem (4.4) by an explicit finite difference scheme of second-order approximation, and get the solution $\Psi_j(x, t)$, $j \subset J$.
(5) Determine the gradient of misfit function $J(q_n)$ by formula (4.3).
(6) Calculate the descent parameter $\alpha_n = 2J(q_n)/\|J'(q_n)\|$ for the minimum errors gradient method.
(7) Calculate the next approximation $q_{n+1} = q_n - \alpha_n J'(q_n)$, and go to step (2).

4.2.2 Adaptive gradient algorithm

The adaptive gradient algorithm (AdaGrad) is a modified gradient descent algorithm with per-parameter learning rate, first published in 2011 [11]. It "adapts" the learning rate to the parameters, performing smaller updates (i.e. low learning rates) for parameters associated with frequently occurring features, and larger updates (i.e. high learning rates) for parameters associated with infrequent features.

Result: Get the approximate solution q_{n+1}
Set stopping parameter $\varepsilon > 0$, iteration number $n = 0$;
Read the values q_0;
Calculate $J(q_0)$ using (3.1);
while $J(q_n) > \varepsilon$ **do**
  $y_j(x, t; q_n) \leftarrow direct\_problem(q_n)$;
  Calculate $J(q_n)$ using (3.1);
  $\Psi_j(x, t; q_n) \leftarrow adjoint\_problem(y_j(x, t; q_n), F^j_{ik}, q_n)$;
  Calculate $J'(q_n)$ using (4.3);
  Calculate descent parameter $\alpha_n = 2J(q_n)/\|J'(q_n)\|$;
  $q_{n+1} \leftarrow q_n - \alpha_n J'(q_n)$;
  $n \leftarrow n + 1$;
end

Algorithm 1: Gradient method of minimum errors.

Result: Get the approximate solution q_{n+1}
Set descent parameter $\alpha > 0$, stopping parameter $\varepsilon > 0$, iteration number $n = 0$ and $G = I$ (identity matrix);
Read the values q_0;
Calculate $J(q_0)$ using (3.1);
while $J(q_n) > \varepsilon$ **do**
  $y_j(x, t; q_n) \leftarrow direct\_problem(q_n)$;
  Calculate $J(q_n)$ using (3.1);
  $\Psi_j(x, t; q_n) \leftarrow adjoint\_problem(y_j(x, t; q_n), F^j_{ik}, q_n)$;
  Calculate $J'(q_n)$ using (4.3);
  Update G using (4.7);
  Calculate descent parameter $\alpha_n = \alpha\, \mathrm{diag}(G)^{-\frac{1}{2}}$;
  $q_{n+1} \leftarrow q_n - \alpha_n \circ J'(q_n)$;
  $n \leftarrow n + 1$;
end

Algorithm 2: Adaptive gradient method (AdaGrad).

This method also depends on descent parameter α but, in this case, this is multiplied with the elements of a vector $G_{i,j}$ which is the diagonal of the outer product matrix

$$G = \sum_{n=1}^{N} g_n g_n^T, \tag{4.7}$$

where $g_n = \nabla J(q_n)$ is the gradient at iteration n. The diagonal is given by

$$G_{j,j} = \sum_{n=1}^{N} g_{n,j}^2.$$

The AdaGrad algorithm is as follows [11] (Algorithm 2).

(1) Set an initial approximation vector q_0, descent parameter $\alpha > 0$ and stopping parameter $\varepsilon > 0$. Suppose that we have q_n. Show how to get the next approximation q_{n+1}.
(2) Check the stop condition: if $J(q_n) < \varepsilon$, then q_n is an approximate solution of inverse problem (1.1a)–(1.1c), (1.2). Otherwise, go to step (3).
(3) Solve direct problem (1.1a)–(1.1c) for a given set of the parameters q_n by an explicit finite difference scheme of second-order approximation, and get $y_j(x_i, t_k; q_n)$, $i = 1, \ldots, N_x$, $k = 1, \ldots, K$, $j \subset J$.

(4) Solve adjoint problem (4.4) by an explicit finite difference scheme of second-order approximation, and get the solution $\Psi_j(x,t)$, $j \subset J$.
(5) Determine the gradient of misfit function $J(q_n)$ by formula (4.3).
(6) Update the outer product matrix by formula (4.7).
(7) Calculate the descent parameter $\alpha_n = \alpha \operatorname{diag}(G)^{-\frac{1}{2}}$.
(8) Calculate the next approximation $q_{n+1} = q_n - \alpha_n \circ J'(q_n)$, and go to step (2).

5 Numerical solution of inverse problem for the mathematical models of a social network

We apply the proposed optimization algorithms to the inverse problem for the mathematical diffusive logistic model of type (1.1a) arising in online social networks [38] (number of equations $N = 1$),

$$\begin{cases} \dfrac{\partial y}{\partial t} = d\dfrac{\partial^2 y}{\partial x^2} + r(t)y(x,t)\Big(1 - \dfrac{y(x,t)}{K_{\text{cap}}}\Big), & t \in (1,T), \quad x \in (l,L), \\ y(x,1) = \psi(x), & x \in (l,L), \\ \dfrac{\partial y}{\partial x}\Big|_{x=l} = \dfrac{\partial y}{\partial x}\Big|_{x=L} = 0, & t \in (1,T). \end{cases} \tag{5.1}$$

$$y(x_i,t_k) = F_{ik}, \quad i = 1,\dots,N_x,\ k = 1,\dots,K. \tag{5.2}$$

The description of all parameters and functions is presented in Section 3 and Table 1. Here we put

$$r(t) = \frac{\beta_2}{\beta_1} - e^{-\beta_1(t-1)}\Big(\frac{\beta_2}{\beta_1} - \beta_3\Big).$$

Since the initial density of influenced users $\psi(x)$ depends on social network and information type, we get ψ_i, $i = 1,\dots,N_x$, from [38] (describing the situation of the information network Digg.com) as an "exact" initial condition and then approximate it on (l,L).

Inverse problem (5.1), (5.2) consists in determination of the vector function

$$q = (d, K_{\text{cap}}, \beta_1, \beta_2, \beta_3, \psi_1, \dots, \psi_{N_x})$$

from (5.1) using additional information (5.2). For getting the synthetic inverse problem data (5.2), we set an "exact" solution q_{ex} according to Table 1 and solve direct model (5.1) with initial condition as in [38] by an explicit finite difference scheme of the second-order approximation, with a grid of 50 points on the x-axis and, according to the Courant condition, 500 points on the t-axis. We set the distance from source of information $N_x = 6$ on friendship interval $l = 1$, $L = 6$ and measure the density of influenced users every hour from $t_1 = 3$ to $t_6 = 10$, $K = 8$, during $T = 24$ hours.

To obtain the best results, we combine the algorithms according to the pipeline: a global optimization algorithm to get a good initial approximation, then a local optimization method to get a result. We use the local optimization gradient methods described in Section 4.2 and the Nelder–Mead method based on function comparison [31].

Symbol	Description	Average value
d	Popularity of information	0.01
K_{cap}	Carrying capacity	25
β_1	Rate of decline in information over time	1.5
β_2	Residual speed	0.375
β_3	Initial growth rate of the number of influenced users	1.65

Table 1: The description and values of parameters of the mathematical model (5.1).

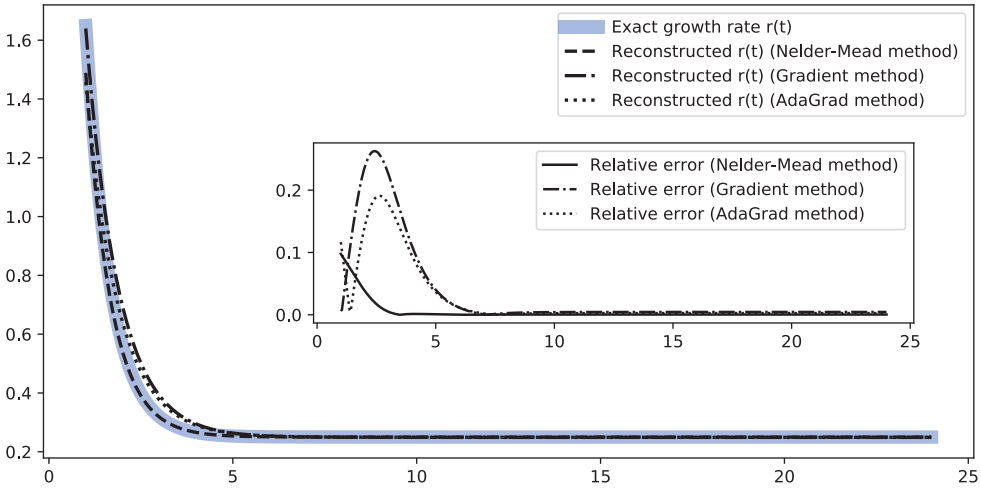

Figure 3: The exact and reconstruction functions $r(t)$ with parameters from Table 2 and the relative error $E(r)$.

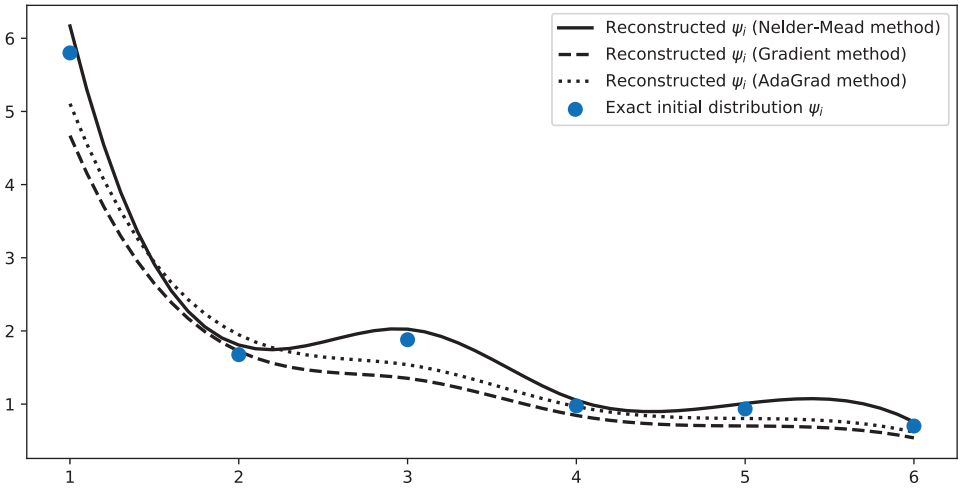

Figure 4: Reconstruction of function $\psi(x)$ and "exact" points ψ_i, $i = 1, \dots, 6$. Here the relative error $E_{\text{NM}}(\psi) = 0.0668$, $E_{\text{GM}}(\psi) = 0.1978$ and $E_{\text{ADA}}(\psi) = 0.128$ for combined methods based on Nelder–Mead and gradient methods respectively.

Suppose that each element q_j, $j = 1, \dots, 11$, lies in the interval $[0, 6]$ and the uniform grid on each interval for the TT method is 256 nodes in each direction ($r_{\max} = 10$). Solving the inverse problem under these conditions by the combined method, we get the following reconstructions of functions $r(t)$ and $\psi(x)$ (Figures 3 and 4). Introduce the relative error for reconstruction of parameters

$$E(r) = \frac{\|r_{\text{pred}} - r_{\text{ex}}\|_{L_2}}{\|r_{\text{ex}}\|_{L_2}}.$$

Figure 7 illustrates the predicting results for an example news story with the proposed model, where the x-axis is the distance between users, while the y-axis represents the density of influenced users within each distance. The solid width lines denote the actual observations for the density of influenced users for a variety of time periods (i.e., 1-hour, 2-hours, 3-hours, 4-hours and 5-hours), while the dashed lines illustrate the predicted density of influenced users by the model for reconstructed parameters q. As we can see, the proposed model is able to accurately predict the density of influenced users with different distance over time (see the real measurements in [38]).

Table 2 gives the numerical values of parameter reconstruction during the numerical calculations with their relative approximation errors $E(q)$. The table shows that the Nelder–Mead method outperforms the gradient based methods, which have converged to the closest local minimum to the tensor train's solution.

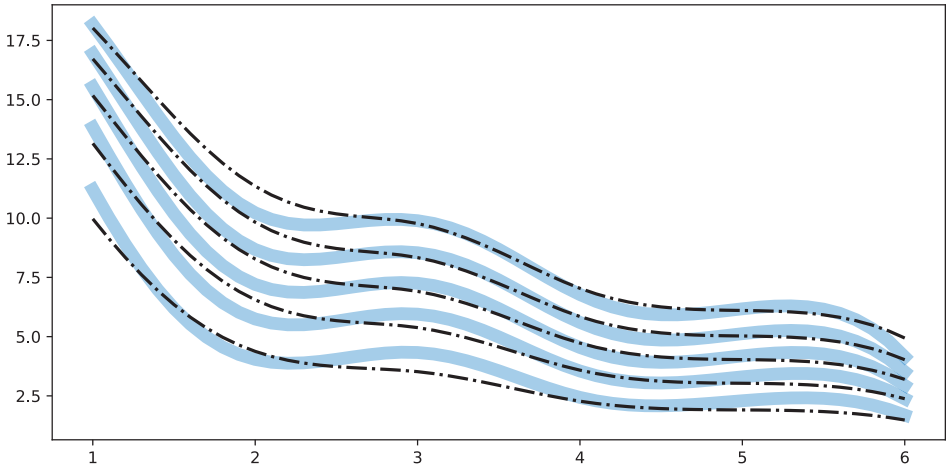

Figure 5: TT and gradient method of minimum errors: comparison of the solutions to direct problem (5.1) (blue lines – exact parameters, black lines – reconstructed parameters from Table 2).

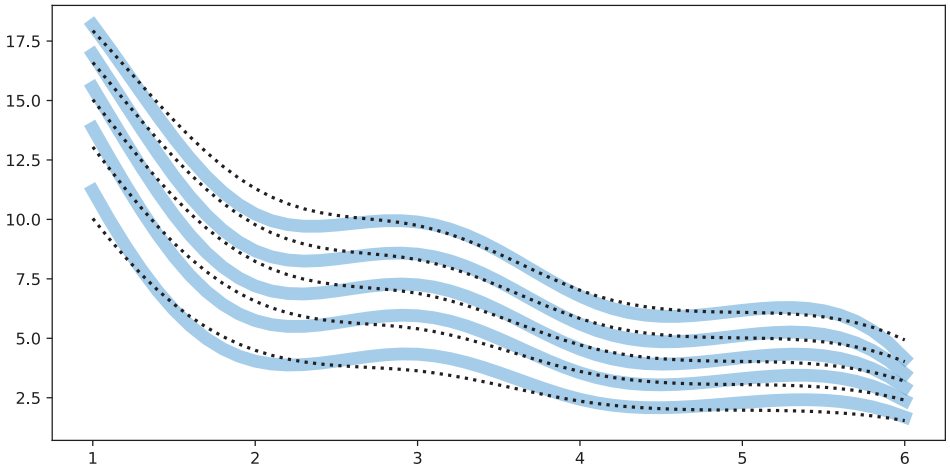

Figure 6: TT and AdaGrad: comparison of the solutions to direct problem (5.1) (blue lines – exact parameters, black lines – reconstructed parameters from Table 2).

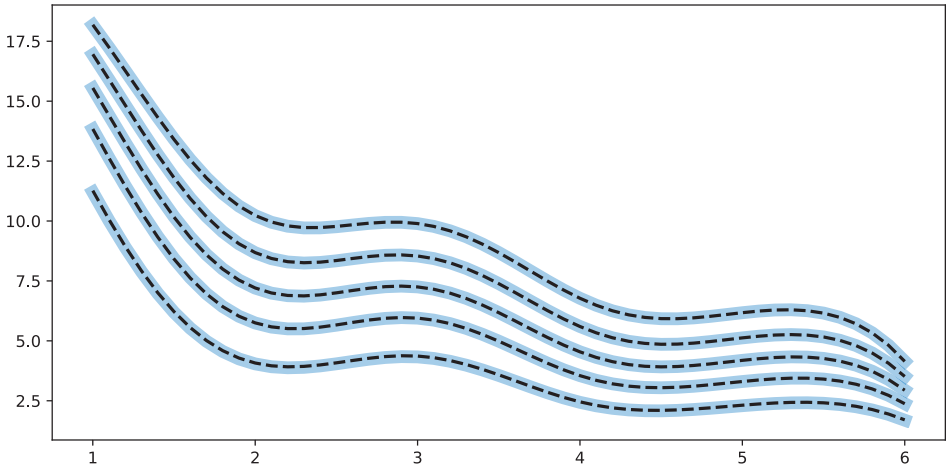

Figure 7: TT and Nelder–Mead: comparison of the solutions to direct problem (5.1) (blue lines – exact parameters, black lines – reconstructed parameters from Table 2).

		TT method		Gradient method		AdaGrad method		Nelder–Mead method	
Symbol	Exact value	Value	Relative error	Value	Relative error	Value	Relative error	Value	Relative error
d	0.01	0	1	0	1	0	1	0.01	$6 \cdot 10^{-3}$
K_{cap}	25	23.18	$7.3 \cdot 10^{-2}$	24.79	$8.4 \cdot 10^{-3}$	25	$4.5 \cdot 10^{-5}$	25	$8.1 \cdot 10^{-5}$
β_1	1.5	1.12	$2.5 \cdot 10^{-1}$	1.137	$2.42 \cdot 10^{-1}$	1.125	$2.5 \cdot 10^{-1}$	1.45	$3.3 \cdot 10^{-2}$
β_2	0.375	0.294	$2.16 \cdot 10^{-1}$	0.2888	$2.3 \cdot 10^{-1}$	0.28	$2.5 \cdot 10^{-1}$	0.362	$3.4 \cdot 10^{-2}$
β_3	1.65	1.724	$4.5 \cdot 10^{-2}$	1.64	$6 \cdot 10^{-3}$	1.457	$1.16 \cdot 10^{-1}$	1.49	$9.7 \cdot 10^{-2}$

Table 2: The results of solving of inverse problem (5.1), (5.2) by TT, gradient type and Nelder–Mead methods.

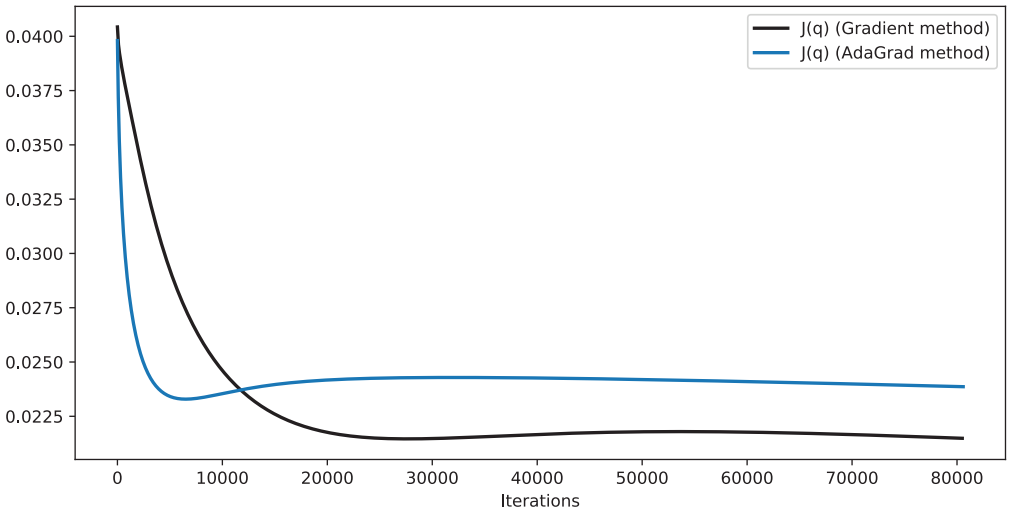

Figure 8: Dependence of $J(q_n)$ on iteration number n (black line – gradient method of minimum errors, blue line – AdaGrad).

Despite the AdaGrad solution has larger relative error, the method has reached this local solution point much faster than the method of minimum errors (Figure 8).

6 Conclusion

The combined optimization algorithm for solving a multi-parameter inverse problem for the mathematical model of a PDE of parabolic type arising in social networks, epidemiology and economy is investigated. The inverse problem consists in identification of coefficients in a PDE and an initial condition of the initial boundary value problem for the PDE using additional measurements of the solution of the direct problem in fixed points of one-dimensional space and time. The considered inverse problem is ill-posed, i.e. the solution of the inverse problem is non-unique and unstable. We reduce the inverse problem to the minimization of the least-squares misfit function. There exists a wide class of optimization methods for a multi-parameter minimization problem. We choose the tensor train decomposition approach as a more appropriate method for solving multidimensional minimization problems. The idea of the proposed method is to extract the tensor structure of the optimized functional and use it for optimization. For a more accurate reconstruction, we apply the methods of local optimization: gradient method of minimum errors, AdaGrad method and Nelder–Mead method. The formula of approximate gradient of the misfit function based on the solution of an adjoint problem is derived.

For numerical experiments the inverse problem for the diffusive logistic mathematical model describing online social networks is solved by combination of tensor train optimization and local methods. Such mathematical model describe information propagation in Digg.com [38], Facebook and Twitter using the TALISMAN database [2], etc. All calculations were implemented on Python on the Google Cloud Platform with a virtual

machine with 15 GB of RAM. It took about 18 million of function evaluations for the tensor train method and about 80 thousand of function evaluations for the gradient methods. However, taking into account that, in the tensor train method, the functions are calculated in parallel, the calculations took much less machine time than in the gradient method.

Nevertheless, we reached a good result by using the combined method; we could improve the quality of parameter reconstruction with the tensor train method by using a more detailed grid ($n = 1000$). But in this case, we need to use a more powerful computing machine.

Funding: This work is supported by the Russian Science Foundation (grant No. 18-71-10044).

References

[1] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp and J. G. Blom, Systems biology: Parameter estimation for biochemical models, *FEBS Journal* **276** (2009), no. 4, 886–902.

[2] A. A. Avetisyan, M. D. Drobyshevskiy, D. Y. Turdakov and T. Ghukasyan, Methods for information diffusion analysis, *Program. Comput. Softw.* **45** (2019), no. 7, 372–380.

[3] K. K. Avilov and A. A. Romanyukha, Mathematical modeling of the spread of tuberculosis and the detection of those infected, *Avtomat. i Telemekh.* (2007), no. 9, 145–160.

[4] H. T. Banks, A. F. Karr, H. K. Nguyen and J. R. Samuels, Jr., Sensitivity to noise variance in a social network dynamics model, *Quart. Appl. Math.* **66** (2008), no. 2, 233–247.

[5] H. T. Banks, K. Rehm and K. L. Sutton, Dynamic social network models incorporating stochasticity and delays, *Quart. Appl. Math.* **68** (2010), no. 4, 783–802.

[6] S. Blower and T. Porco, Quantifying the intrinsic transmission dynamics of tuberculosis, *Theoret. Popul. Biol.* **54** (1998), no. 2, 117–132.

[7] C. Camacho, B. Zou and M. Briani, On the dynamics of capital accumulation across space, *European J. Oper. Res.* **186** (2008), no. 2, 451–465.

[8] Y. Chen, J. Cheng, Y. Jiang and K. Liu, A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification, *J. Inverse Ill-Posed Probl.* **28** (2020), no. 2, 243–250.

[9] I.-C. Chou and E. O. Voit, Recent developments in parameter estimation and structure identification of biochemical and genomic systems, *Math. Biosci.* **219** (2009), no. 2, 57–83.

[10] G. Dai, R. Ma, H. Wang, F. Wang and K. Xu, Partial differential equations with Robin boundary condition in online social networks, *Discrete Contin. Dyn. Syst. Ser. B* **20** (2015), no. 6, 1609–1624.

[11] J. Duchi, E. Hazan and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, *J. Mach. Learn. Res.* **12** (2011), 2121–2159.

[12] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets*, Cambridge University, Cambridge, 2010.

[13] Y. G. Evtushenko, Numerical methods for finding global extrema (Case of a non-uniform mesh), *USSR Comput. Math. Math. Phys.* **11** (1971), no. 6, 38–54.

[14] Y. G. Evtushenko, V. U. Malkova and A. A. Stanevichyus, Parallelization of the process of searching for the global extremum, *Avtomat. i Telemekh.* (2007), no. 5, 46–58.

[15] Y. G. Evtushenko, V. U. Malkova and A. A. Stanevichyus, Parallel search for the global extremum of functions of several variables, *Zh. Vychisl. Mat. Mat. Fiz.* **49** (2009), no. 2, 255–269.

[16] B. Fatma and P. Fatma, Mathematical modelling of HIV epidemic and stability analysis, *Adv. Difference Equ.* **2014** (2014), DOI 10.1186/1687-1847-2014-95.

[17] S. A. Goreinov, E. E. Tyrtyshnikov and N. L. Zamarashkin, A theory of pseudoskeleton approximations, *Linear Algebra Appl.* **261** (1997), 1–21.

[18] S. A. Goreĭnov, N. L. Zamarashkin and E. E. Tyrtyshnikov, Pseudo-skeleton approximations of matrices, *Dokl. Akad. Nauk* **343** (1995), no. 2, 151–152.

[19] A. Guille, H. Hacid, C. Favre and D. Zighed, Information diffusion in online social networks: A survey, *SIGMOD Record.* **42** (2013), 17–28.

[20] J. M. Hyman and J. Li, Differential susceptibility and infectivity epidemic models, *Math. Biosci. Eng.* **3** (2006), no. 1, 89–100.

[21] S. I. Kabanikhin, Definitions and examples of inverse and ill-posed problems, *J. Inverse Ill-Posed Probl.* **16** (2008), no. 4, 317–357.

[22] S. I. Kabanikhin and O. I. Krivorotko, Identification of biological models described by systems of nonlinear differential equations, *J. Inverse Ill-Posed Probl.* **23** (2015), no. 5, 519–527.

[23] S. I. Kabanikhin, O. I. Krivorotko and V. Kashtanova, A combined numerical algorithm for reconstructing the mathematical model for tuberculosis transmission with control programs, *J. Inverse Ill-Posed Probl.* **26** (2018), no. 1, 121–131.
[24] S. I. Kabanikhin, O. I. Krivorotko, D. V. Yermolenko, V. A. Latyshenko and V. N. Kashtanova, Inverse problems of immunology and epidemiology, *Eurasian J. Math. Comput. Appl.* **5** (2017), no. 2, 14–35.
[25] S. I. Kabanikhin, D. A. Voronov, A. A. Grodz and O. I. Krivorotko, Identifiability of mathematical models in medical biology, *Vavilov J. Gen. Breeding* **6** (2016), no. 8, 838–844.
[26] B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Radon Ser. Comput. Appl. Math. 6, Walter de Gruyter, Berlin, 2008.
[27] B. L. Keyfitz and N. Keyfitz, The McKendrick partial differential equation and its uses in epidemiology and population study, *Math. Comput. Modelling* **26** (1997), no. 6, 1–9.
[28] J.-M. Lasry and P.-L. Lions, Mean field games, *Jpn. J. Math.* **2** (2007), no. 1, 229–260.
[29] J.-M. Lasry, P.-L. Lions and O. Gueant, Application of mean field games to growth theory, preprint (2008), https://hal.archives-ouvertes.fr/hal-00348376/document.
[30] C. Lei, Z. Lin and H. Wang, The free boundary problem describing information diffusion in online social networks, *J. Differential Equations* **254** (2013), no. 3, 1326–1341.
[31] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.* **7** (1965), no. 4, 308–313.
[32] I. Oseledets and E. Tyrtyshnikov, TT-cross approximation for multidimensional arrays, *Linear Algebra Appl.* **432** (2010), no. 1, 70–88.
[33] I. V. Oseledets, Tensor-train decomposition, *SIAM J. Sci. Comput.* **33** (2011), no. 5, 2295–2317.
[34] I. V. Oseledets and E. E. Tyrtyshnikov, Breaking the curse of dimensionality, or how to use SVD in many dimensions, *SIAM J. Sci. Comput.* **31** (2009), no. 5, 3744–3759.
[35] R. Solow, A contribution to the theory of economic growth, *Q. J. Econ.* **70** (1956), 65–94.
[36] J. M. Trauer, J. T. Denholm and E. S. McBryde, Construction of a mathematical model for tuberculosis transmission in highly endemic regions of the Asia-pacific, *J. Theoret. Biol.* **358** (2014), 74–84.
[37] E. E. Tyrtyshnikov, Incomplete cross approximation in the mosaic-skeleton method, *Computing* **64** (2000), 367–380.
[38] F. Wang, H. Wang and K. Xu, Diffusive logistic model towards predicting information diffusion in online social networks, in: *32nd International Conference on Distributed Computing Systems Workshops* (ICDCSW), IEEE Press, Piscataway (2012), 133–139.
[39] J. Zhang, J. Lou, Z. Ma and J. Wu, A compartmental model for the analysis of SARS transmission patterns and outbreak control measures in China, *Appl. Math. Comput.* **162** (2005), 909–924.
[40] D. A. Zheltkov, I. V. Oferkin, E. V. Katkova, A. V. Sulimov, V. B. Sulimov and E. E. Tyrtyshnikov, TTDock: A docking method based on tensor train decompositions, *Vychisl. Met. Programm* **14** (2013), 279–291.
[41] V. V. Zheltkova, D. A. Zheltkov, Z. Grossman, G. A. Bocharov and E. E. Tyrtyshnikov, Tensor based approach to the numerical treatment of the parameter estimation problems in mathematical immunology, *J. Inverse Ill-Posed Probl.* **26** (2018), no. 1, 51–66.