A method of generative model design based on irregular data in application to heat transfer problems

Nikolay Bykov, Alexander Hvatov, Anna Kalyuzhnaya and Alexander Boukhanovsky
ITMO University, Kronversky pr. 49, 197101, St. Petersburg, Russia
E-mail: nbykov2006@yandex.ru

Abstract. The paper presents the results of applying the generative design method to reconstruct a model driven by irregular data in the form of a partial differential equation. The problem of non-stationary heating of a metal rod is considered as an example. The finite element method is used to construct templates of possible elements of the derived differential equation. The optimization algorithm for determining a model design is based on the procedure of best subset selection. The efficiency of the method for derivating a model of the thermal conduction process based on irregular data is shown. As irregular data, data obtained on spatial meshes with alternating steps of different size, or a step changing in arithmetic progression, are considered.

1. Introduction
The development of methods of the generative model design (GMD), in particular, data-driven models, is a promising direction for creating toolkit for monitoring and predicting complex multiphysical man-made or natural processes. Data-driven models have become widespread in various applications of neural networks and machine learning methods [1, 2]. The idealized task is to reconstruct an accurate and well-interpreted model of a particular process at the same time. The generative design of models in the form of differential equations is a promising means of solving this problem [3, 4, 5, 6, 7].

The GMD methods have been intensively developed in the last few years. The methods were used for reconstruction of partial differential equations (PDE) that describe classical problems of mathematical physics: the wave equation, the Burgers equation, the Korteweg de Vries equation, with derivative term coefficients of the order of unity [7], as well as to solve real problems of ocean dynamics [8, 9]. In these studies, the data presented on regular spatial meshes were considered. The studies demonstrated the high efficiency of the GMD methods.

The GMD algorithm consists of several procedures. The central place among which is (i) the construction of templates of elements (terms) of a potential model based on the available data, (ii) an optimization procedure for selecting the most significant elements.

In the works devoted to the problem of the GMD application, the finite difference method (FDM) is used to discretize partial differential equations (PDE), which is necessary for calculating the templates of elements of the derived model. In this paper, at the stage of template calculation, it is proposed to use the finite element method (FEM) [10, 11], which makes it relatively easy to take into account the irregularity of the data. As a test, the problem of heating a metal rod is considered. To obtain synthetic data, a heat transfer equation is used, describing a non-stationary one-dimensional heat conductivity process. The equation has an analytical solution. For reconstruction procedure, data arrays on irregular meshes are
obtained. Several types of meshes are considered: with alternating step size along the spatial coordinate and with a step size increasing in arithmetic progression with a given progression difference.

To obtain a sparse vector of coefficients of the reconstructed dependence of the response on predictors, the LASSO technique has been widely used in regression analysis over the past two decades [12]. The advantages of the LASSO procedure include the ability to take into account a very large number of predictors. The disadvantages include the need for a reasonable choice of a penalty hyperparameter to filter out insignificant predictors, as well as, in some cases, the low accuracy of the obtained results. The differential equations underlying the mathematical model of most physical processes usually contain only a few terms. As a rule, it is enough to take into account up to 10-20 terms (elements) in the template of a potential model within the GMD method. In this case, the method of the selection of best subset of variables [13], which does not require the introduction and justification of auxiliary parameters, can be used as an optimization procedure. This approach is used in this work.

The present work has two objectives. First, to apply the GMD technique to derive the PDE based on the available data on the non-stationary process of thermal conduction in a metal rod. Secondly, to find out the effectiveness of the GDM method for solving a real problem in the case of irregular data. Irregular data refers to data obtained on an irregular spatial mesh.

2. Object of the research
As an object of the research, the problem of heating a thermally insulated rod is considered. The non-stationary process of heat conduction in a rod in a one-dimensional formulation is described by the equation [14]

$$\frac{\partial T}{\partial t} - \alpha \frac{\partial^2 T}{\partial x^2} = 0,$$

(1)

where \(t\) is the time, \(x\) is the coordinate, \(T\) is the temperature, \(\alpha\) is the coefficient of thermal diffusivity of rod material. The initial and boundary conditions have the form

$$T(x,0) = T_0, \quad T(x=0,t) = T(x=1,t) = T_b.$$

Here \(T_0\) is the initial rod temperature, \(T_b\) is the fixed temperature of rod ends.

The analytical solution of the problem for given initial and boundary conditions can be obtained by the method of separation of variables [14]:

$$T(x,t) = T_b - \sum_{k=1}^{K} \frac{T_b - T_0}{(2k-1)\pi} \sin[(2k-1)\pi x] \exp[-\alpha(2k-1)^2\pi^2 t],$$

(3)

where \(K\) is the number of terms of the equation (in calculations \(K = 100\)).

In this paper, heating of a copper rod is considered. The coefficient of thermal diffusivity is \(\alpha = 10^{-4}\) m²/s, \(T_b = 700\)K and \(T_0 = 600\)K. The analytical solution for these parameters is shown in Figure 1 for three times \(t = 500, 1000, 1500\)s. The length of the rod is assumed to be equal to 1m.

3. The procedure of generative model design
In order to determine the design of equation for heat transfer problem, the following general form of an equation is considered in this paper

$$\alpha_1 \frac{\partial^2 T}{\partial t^2} + \alpha_2 \frac{\partial T}{\partial t} + \alpha_3 \frac{\partial^2 T}{\partial x^2} + \alpha_4 \frac{\partial T}{\partial x} = 0,$$

(4)

where \(\alpha_1, \alpha_2, \alpha_3, \alpha_4\) are unknown coefficients.
Figure 1. Temperature distributions along the rod according to equation (3) for time 500 s (1), 1000 s (2) and 1500 s (3). The number of nodes is 201.

To obtain a discretized template, the present paper proposes to use the finite element method [10, 11]. The approximate solution of the equation (4) is sought as a combination of the nodal values of the desired function

$$T = \sum_{j=1}^{N} T_j \phi_j(x),$$

where $T_j$ is the temperature in node $j$, $\phi_j$ is the interpolation function, $N$ is the number of nodes ($N-1$ is the number of elements).

In order to determine the nodal values $T_j$, the integral of weighted residual over the entire computational domain (in the case under consideration $[0; 1]$) is equated to zero:

$$\int_{0}^{1} \phi_m(x) \left[ \alpha_1 \frac{\partial^2 T}{\partial t^2} + \alpha_2 \frac{\partial T}{\partial t} + \alpha_3 \frac{\partial^2 T}{\partial x^2} + \alpha_4 \frac{\partial T}{\partial x} \right] dx = 0.$$  \hspace{1cm} (6)

In the Galerkin formulation, the weight function has the same form as the interpolation function. In the present paper, first order polynomials (linear interpolation) are considered as interpolation functions. Substituting the functions in (6) leads to the following system of algebraic equations

$$\alpha_1 a_{j,1} + \alpha_2 b_{j,1} + \alpha_3 c_{j,1} + \alpha_4 d_{j,1} = 0,$$

$$a_{j,1} = \frac{\Delta x_j}{6} \left[ \frac{T_{j+1}^{n-1} - 2T_{j-1}^{n} + T_{j-1}^{n+1}}{\Delta t^2} \right] + \frac{\Delta x_j + \Delta x_{j+1}}{3} \left[ \frac{T_{j+1}^{n-1} - 2T_{j}^{n} + T_{j+1}^{n+1}}{\Delta t^2} \right] + \ldots$$
\[ b_{jj} = \frac{\Delta x_j}{6} \left[ \frac{T_{j-1}^{n+1} - T_{j-1}^n}{2\Delta} \right] + \frac{\Delta x_j + \Delta x_{j+1}}{3} \left[ \frac{T_{j+1}^{n+1} - T_{j}^n}{2\Delta} \right] + \frac{\Delta x_{j+1}}{6} \left[ \frac{T_{j+1}^{n+1} - T_{j+1}^n}{2\Delta} \right], \]  
\[ c_{jl} = \frac{\beta_j^{im}}{\Delta x_j} T_{jl}^{n-1} - \beta_j^{im} \left( \frac{1}{\Delta x_j} + \frac{1}{\Delta x_{j+1}} \right) T_{jl}^n + \frac{\beta_j^{im}}{\Delta x_{j+1}} T_{jl}^{n+1} + + \frac{1 - \beta_j^{im} - \beta_j^{im}}{\Delta x_j} T_{jl}^{n+1} - (1 - \beta_j^{im} - \beta_j^{im}) \left( \frac{1}{\Delta x_j} + \frac{1}{\Delta x_{j+1}} \right) T_{jl}^n + \frac{1 - \beta_j^{im} - \beta_j^{im}}{\Delta x_{j+1}} T_{jl}^{n+1}, \]  
\[ d_{jl} = \frac{\beta_j^{im}}{\Delta x_j} \left[ \frac{T_{jl+1}^{n+1} - T_{jl}^{n+1} - 2T_{jl+1}^n + T_{jl+1}^{n-1}}{2\Delta^2} \right] + (1 - \beta_j^{im} - \beta_j^{im}) \left[ \frac{T_{jl+1}^{n+1} - T_{jl+1}^n - 2T_{jl+1}^n + T_{jl+1}^{n-1}}{2\Delta^2} \right] + \frac{\beta_j^{im}}{\Delta x_{j+1}} \left[ \frac{T_{jl+1}^{n+1} - T_{jl+1}^n - 2T_{jl+1}^n + T_{jl+1}^{n-1}}{2\Delta^2} \right]. \]  

A three-layer approximation is used in (7) to discretize temporal derivatives in (6). The index \( j \) denotes the node number and varies from 2 to \((N - 1)\) taking into account the specified stationary boundary values of the unknown function \( T \) in nodes \( j = 1 \) and \( j = N \). The index \( l \) denotes a time slice. One time slice consists of three neighbor time layers \( n - 1, \) \( n \) and \( n + 1 \). The number of time slices considered is \( L \) (in the present calculations, \( L = 1 \)). The \( \beta_j^{im} \) parameters regulate the degree of implicitness (the contribution to the computational templates of data from different time points). In the calculations, \( \beta_j^{im} = 0, \beta_1^{im} = 0 \).

In vector form, the expression (7) can be represented as:

\[ \alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 + \alpha_4 V_4 = Z_0. \]  

Here vector \( V_1 \) consists of elements \( a_{jl} \) (index \( j \) changes from 2 to \((N - 1)\), index \( l \) from 1 to \( L \))

\[ V_1^T = (v_{11}, v_{12}, v_{13}, \ldots, v_{1(N-2)\times L}), \]

\[ = (a_{21}, a_{31}, a_{41}, \ldots, a_{(N-1)1}, a_{22}, a_{32}, a_{42}, \ldots, a_{(N-1)2}, a_{23}, a_{33}, a_{43}, \ldots, a_{(N-1)3}), \]

accordingly the vector \( V_2 \) consists of elements \( b_{jl} \), \( V_3 \) consists of elements \( c_{jl} \), \( V_4 \) consists of elements \( d_{jl} \), \( Z_0 \) is zero vector.

For the application of statistical learning methods [13], it is convenient to rewrite the expression (12) in the form:

\[ Y = \alpha_0 - \alpha_1 V_1 - \alpha_3 V_3 - \alpha_4 V_4, \]

where \( Y = \alpha_2 V_2, \alpha_0 = 0 \). Estimates of \( \alpha \) coefficients can be made using least-squares (LS) method. The statistical analysis was performed using the R package [15].

The represented below the GMD algorithm consists of the following steps:

1) Using the generated data (see Section 2), the components of the vectors \( V_i \) are calculated by the use of equations (8-11).

2) The coefficient \( \alpha_2 \) is assumed to be equal to unity. Other coefficients are calculated by the LS method. The value of the \( t \)-criterion and the \( p \)-value (the probability of observing a value equal to or greater than \(|t|\)) are determined.
3) The procedure of the selection of best subset of coefficients is applied [13, 15], which allows to filter out insignificant coefficients. In the case of equation (13), the number of variables (predictors) is \( P = 3 \). The procedure involves iterating through \( 2^P \) possible arrangements with one, two or three terms (elements). For each fixed number of terms, the possible variants of the terms are sorted out and the optimal model is selected based on the calculation of the smallest sum of the square of the residuals. Next, a single optimal model is selected, for example, using the \( C_P \) statistic estimation of the mean squared error on the control sample [13]:

\[
C_P = \frac{1}{n_o} (\text{RSS} + 2d\hat{\sigma}^2),
\]

where \( \hat{\sigma}^2 \) is the estimation of the variance of residuals associated with each response value in (13), \( \text{RSS} \) is the residual sum of squares, \( d \) is the number of predictors in the model, \( n_o = (N - 2) \times L \) is the number of observations. The \( C_P \) statistics adds penalty equals to \( 2d\hat{\sigma}^2 \) to \( \text{RSS} \), calculated on training sample. Thus it takes into account the property of the error on the training sample to underestimate the error on the control sample. The penalty increases as the number of predictors included in the model increases.

4. Application of the method for deriving the heat transfer equation from irregular data

Three cases corresponding to different types of data obtained using the expression (3) are considered. Case 1. The regular mesh with number of nodes \( N \) and space step

\[
\Delta x = L_S/(N-1).
\]

Here \( L_S = 1 \text{m} \) is the rod length.

Case 2. The irregular mesh with alternating steps \( \Delta x_{1c} \) and \( \Delta x_{2c} \):

\[
\Delta x_{1c} = 2\gamma \Delta x = 2\gamma L_S/(N-1),
\]

\[
\Delta x_{2c} = 2(1-\gamma)\Delta x = 2(1-\gamma)L_S/(N-1),
\]

where \( \gamma \) is the parameter of irregularity. The value \( \gamma = 0.475 \), defining a difference in the size of neighbor cells of 10\%, is considered.

Case 3. The irregular mesh with number of nodes \( N \) and distance between nodes varied in arithmetic progression with progression step \( d_a \):

\[
\Delta x_i = \Delta x_1 + (i-1)d_a,
\]

\[
d_a = 2\left(\frac{L_S}{N-1} - \Delta x_1\right)/(N-2).
\]

Here \( \Delta x_1 \) is the distance between nodes 0 and 1. The values \( \Delta x_1 = 0.0049005 \) and \( d = 10^{-6} \) are considered. The size of the first and last cells differs by 4\%.

It is assumed that \( N = 201 \) for all cases.

The analysis showed the correlation of data on different time slices, which is typical for many non-stationary problems [13]. The paper considers one time slice corresponding to the time of 1500 s, the step between the neighbor time layers is 1 s. Additional internal tests showed a more accurate recovery of the model in the case of the consideration only internal nodes, excluding nodes adjacent to the boundaries, the conditions at which are stationary according to (2). Therefore, the real number of observations \( n_o \) was 163 (19 < \( i < 182 \)).

Table 1 shows the recovered by the LS-method coefficients as well as the \( t \)-criterion values and the corresponding \( p \)-values.

In the case of a uniform mesh (case 1), all estimates of the coefficients are reproduced. However, the coefficient \( \alpha_3 \) has significantly lower \( p \)-value. One can expect that the term (element) with this coefficient
Table 1. The coefficients of the equation and the results of statistical analysis.

| Case  | \( \alpha_0 \) | \( \alpha_1 \) | \( \alpha_3 \) | \( \alpha_3 \) |
|-------|----------------|----------------|----------------|----------------|
| 1     | \( 3.60 \cdot 10^{-12} \) | \( -1.54 \cdot 10^{-1} \) | \( -1.0 \cdot 10^{-4} \) | \( 6.16 \cdot 10^{-20} \) |
| \( t \) | 0.18 | -0.51 | -3355.0 | 0.0 |
| \( p \) | 0.86 | 0.61 | < \( 2 \cdot 10^{-16} \) | 1.0 |
| 2     | \( 7.02 \cdot 10^{-12} \) | \( -1.36 \cdot 10^{-1} \) | \( -1.0 \cdot 10^{-4} \) | \( 1.04 \cdot 10^{-12} \) |
| \( t \) | 0.29 | -0.37 | -2776.37 | 0.11 |
| \( p \) | 0.78 | 0.71 | < \( 2 \cdot 10^{-16} \) | 0.92 |
| 3     | \( -9.39 \cdot 10^{-12} \) | \( -1.36 \cdot 10^{-1} \) | \( -1.0 \cdot 10^{-4} \) | \( 1.07 \cdot 10^{-10} \) |
| \( t \) | -0.45 | 0.44 | -3286.22 | 12.94 |
| \( p \) | 0.65 | 0.66 | < \( 2 \cdot 10^{-16} \) | < \( 2 \cdot 10^{-16} \) |

Table 2. Results of applying the best subset selection procedure.

| Case | \( \alpha_1 \) | \( \alpha_3 \) | \( \alpha_4 \) | \( C_P \) |
|------|----------------|----------------|----------------|--------|
| 1    | *              |               | 0.26           |        |
|      | *               |               | 2.00           |        |
|      | * *             |               | 4.00           |        |
| 2    | *              |               | 0.15           |        |
|      | *               |               | 2.01           |        |
|      | * *             |               | 4.00           |        |
| 3    | *              |               | 167.95         |        |
|      | * *             |               | 2.20           |        |
|      | * * *           |               | 4.00           |        |

will have the significant impact on the response \( Y \). This hypothesis is supported by the results shown in table 2. For case 1 of \( 2^3 = 8 \) possible combinations, three optimal sets among variants of one, two and three terms are marked by asterisks. Value of \( C_P \) parameter is minimum for the first set with one term with coefficient \( \alpha_3 \). The results of a procedure of the best subset selection for case 1 are shown in table 3. The results fully correspond to the desired model.

For case 2 with "alternating" steps in space (with step ratio 0.475/0.525 = 0.9), the situation is similar. The difference in steps does not affect the derived model quality (see table 3).

For case 3 with a step varying in arithmetic progression according to the results of statistical analysis (table 2), a model with two coefficients \( \alpha_3 \) and \( \alpha_4 \) can be preferred (table 3). However, first, the value of the coefficient \( \alpha_4 \) is negligible compared to \( \alpha_3 \) (the norms of vectors \( V_3 \) and \( V_4 \) are of the same order). Secondly, the data in table 1 indicate a significantly greater influence on the response of the term with \( \alpha_3 \). Therefore, we can recognize the results of the model recovery process as good.
### Table 3. The results of generative model design application.

| Case | $\alpha_0$   | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ |
|------|--------------|------------|------------|------------|------------|
| 1    | $-5.82 \cdot 10^{-12}$ | 0          | 1          | $-1.0 \cdot 10^{-4}$ | 0          |
| 2    | $-1.33 \cdot 10^{-12}$ | 0          | 1          | $-1.0 \cdot 10^{-4}$ | 0          |
| 3    | $-1.04 \cdot 10^{-12}$ | 0          | 1          | $-1.0 \cdot 10^{-4}$ | $1.06 \cdot 10^{-10}$ |
| Exact value | 0 | 0 | 1 | $-1.0 \cdot 10^{-4}$ | 0 |

### 5. Conclusions

The method of the generative model design is a promising tool for constructing a data-driven process model based on a partial differential equation. In this paper, the GMD is used to derive the PDE, which describes the non-stationary process of heat conduction. It has been shown that the GMD qualitatively reproduces the model on both regular and irregular data. As irregular data, spatial-temporal distributions of temperature on irregular meshes with alternating steps of different length and with step changing in arithmetic progression are considered. Within the framework of the GMD algorithm for obtaining discretized templates of model elements on an irregular grid, the use of the finite element method is proposed, and for optimizing the model structure – the method of the best subset selection. In the future, it is planned to take into account the presence of statistical noise in the source data used to reconstruct the heat transfer equation.

### 6. Acknowledgements

This research is financially supported by The Russian Science Foundation, Agreement N 21-11-00296, https://rscf.ru/en/project/21-11-00296/

### References

[1] Hutter F, Kotthoff L and Vanschoren J (Eds) 2019 Automated Machine Learning (Switzerland: Springer)
[2] Kalyuzhnaya A V, Nikitin N O, Vychuzhanin P, Hvatov A and Boukhanovsky A V 2020 Proc. Genetic and Evolutionary Computation Conf. Comp. (Electronic-only Conference) p 43
[3] Rudy S H, Alla A, Brunton S L and Kutz J N 2017 SIAM Journal on Applied Dynamical Systems 18 643
[4] Rudy S H, Brunton S L, Proctor J L and Kutz J N 2017 Data-driven discovery of partial differential equations Science Advances 3 e1602614
[5] Schaeffer H 2017 Learning partial differential equations via data discovery and sparse optimization Proc. R. Soc. A 473 20160446
[6] Kondrashov D, Chekroun M D and Ghil M 2015 Physica D: Nonlinear Phenomena 297 33
[7] Maslyaev M, Hvatov A and Kalyuzhnaya A 2019 Proc. Int. Conf. on Computational Science (Cham:Springer) p 635
[8] Maslyaev M and Hvatov A 2019 Procedia Computer Science 156 367
[9] Maslyaev M, Hvatov A and Kalyuzhnaya A 2020 Proc. Int. Conf. on Computational Science (Cham: Springer) p 86
[10] Zienkiewicz O C, Taylor R L and Zhu J Z 2013 The Finite Element Method: Its Basis and Fundamentals. Seventh Edition (Elsevier)
[11] Zienkiewicz O C, Taylor R L and Nithiarasu P 2014 The Finite Element Method for Fluid Dynamics. Seventh Edition (Elsevier)
[12] Tibshirani R 1996 J. R. Statist. Soc. B 58 267
[13] James G, Witten D, Hastie T and Tibshirani R 2013 An Introduction to Statistical Learning (New York: Springer)
[14] Samarsky A A and Vabishchevich P N 2003 Computational Heat Transfer (Moscow: Editorial URSS)
[15] R Core Team 2020 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing (Austria)