



Regularization theory in the study of generalization ability of a biological neural network model

Aleksandra Świetlicka¹

Received: 25 September 2018 / Accepted: 18 December 2018 /
Published online: 4 January 2019
© The Author(s) 2019

Abstract

This paper focuses on the generalization ability of a dendritic neuron model (a model of a simple neural network). The considered model is an extension of the Hodgkin-Huxley model. The Markov kinetic schemes have been used in the mathematical description of the model, while the Lagrange multipliers method has been applied to train the model. The generalization ability of the model is studied using a method known from the regularization theory, in which a regularizer is added to the neural network error function. The regularizers in the form of the sum of squared weights of the model (the penalty function), a linear differential operator related to the input-output mapping (the Tikhonov functional), and the square norm of the network curvature are applied in the study. The influence of the regularizers on the training process and its results are illustrated with the problem of noise reduction in images of electronic components. Several metrics are used to compare results obtained for different regularizers.

Keywords Kinetic model of neuron · Markov kinetic schemes · Lagrange multipliers · Generalization ability · Image processing · Noise reduction

Mathematics Subject Classification (2010) 68T05

1 Introduction

This article is the latest part of a series reporting the present author's research into a set of models of biological dendritic neurons and neurons with a point-like structure. The article presents the results of research into generalization ability of the stochastic

Communicated by: Pavel Solin

✉ Aleksandra Świetlicka
aleksandra.swietlicka@put.poznan.pl

¹ Institute of Automation and Robotics, Poznan University of Technology, ul. Piotrowo 3A, 60-965, Poznań, Poland

kinetic model of a biological dendritic neuron, which stems from the model proposed by A.L. Hodgkin and A.F. Huxley in [8]. Generalization ability of the stochastic kinetic model of a biological neuron with a point-like structure was studied in [16]. Models of biological neural networks were considered in [15], which showed the discretization of model equations in time and space, and in [13], where the training of neural network models was discussed, albeit without analyzing the generalization ability of these models.

The generalization ability determines whether a neural network is capable of returning a proper solution for data that did not appear in the training set. Moreover, ability of generalization prevents the training algorithm from over-training the neural network.

As was shown in [16], a biological neuron model has the ability of generalization. The model of a biological neural network understood as a biological dendritic neuron may be assumed to have enhanced capabilities compared to the model of a neuron with a point-like structure, as each and every point of a so-defined neural network may perform consecutive steps of the algorithm for which the network has been designed. Therefore, to be able to treat the results rendered by a neural network as valid, it is necessary to consider the generalization ability of the biological neural network model.

Currently, the most common method used to study the generalization ability of a model of any neural network is the method adopted from the regularization theory [2, 7]. This method consists in adding the so-called regularizer to the error function of the network. Depending on how the regularizer is defined, it can penalize the curvature of the neural network [10], smoothen and therefore stabilize the solution [7], or smoothen the curvature, making the output of the network locally invariant with respect to small perturbations of the input [6].

In this study, three kinds of regularizer are used: the penalty function, the Tikhonov functional, and the square norm of the network curvature. The penalty function is the sum of squared weights of the model [10], the Tikhonov functional is a linear differential operator related to input-output mapping [7], while the network curvature is defined as the second derivative of the output with respect to the input of the model (second-order differential operator) [6].

To show how the regularizer impacts the training process of the model of a biological neural network, this network was used to reduce noise in images of selected electronic components, such as a printed board or an integrated circuit. The noise was introduced, for example, by covering the camera lens with a thin plastic film.

The model of a biological neural network was implemented in MATLAB. The network was trained with the Lagrange multipliers method [1]. The definition of the problem made it possible to use the embedded MATLAB `fsolve` function.

The paper is organized as follows. A concise mathematical description of the biological neural network model is given in Section 2. The regularizers used in the study are described in Section 3. The application of the Lagrange multipliers method to train the neural network model is discussed in Section 4. Section 5 presents the results of the training process of the network. Section 6 completes the paper.

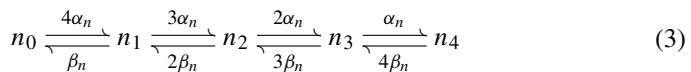
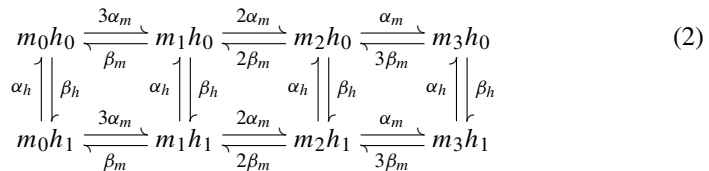
2 Model

The paper considers the stochastic kinetic model of a biological dendritic neuron (for simplicity, we will refer to it as a biological neural network throughout the paper). The original model, which was used to derive the kinetic model, was proposed and described in detail in the paper by A.L. Hodgkin and A.F. Huxley in 1952 [8]. The biological foundations of the model have been thoroughly studied and presented in a number of publications, e.g., in [3, 5, 8, 9, 14]; therefore, in this paper, we only provide a concise mathematical description of the model.

The main equation describing the considered model has the following form [13]:

$$\frac{a}{2R} \frac{\partial^2 V}{\partial x^2} = C \frac{\partial V}{\partial t} + g_{Na} [m_3 h_0] (V - V_{Na}) + g_K [n_4] (V - V_K) + g_L (V - V_L) \tag{1}$$

where V is the potential on the cell membrane, and the remaining terms (that do not include a derivative) represent the ion current components related to the respective types of ions (sodium, potassium, and chlorine). The variables $[m_3 h_0]$ and $[n_4]$ are components of the Markov kinetic schemes (2) and (3). Their values are obtained from the normal distribution. A detailed description of the procedure for obtaining these values is included, among others, in [5, 12–14]. The forms of transfer functions $\alpha_i(V)$ and $\beta_i(V)$, which appear in the kinetic schemes, are given in Table 1 [8].



The structure of the biological neural network considered in this paper is given in Fig. 1. The points on the structure represent places where the potential will be

Table 1 $\alpha_i(V)$ and $\beta_i(V)$ functions [8]

i	$\alpha_i(V)$	$\beta_i(V)$
n	$\frac{0.01 \cdot (10 - V)}{\exp\left(\frac{10 - V}{10}\right) - 1}$	$0.125 \cdot \exp\left(-\frac{V}{80}\right)$
m	$\frac{0.1 \cdot (25 - V)}{\exp\left(\frac{25 - V}{10}\right) - 1}$	$4 \cdot \exp\left(-\frac{V}{18}\right)$
h	$0.07 \cdot \exp\left(-\frac{V}{20}\right)$	$\frac{1}{\exp\left(\frac{30 - V}{10}\right) + 1}$

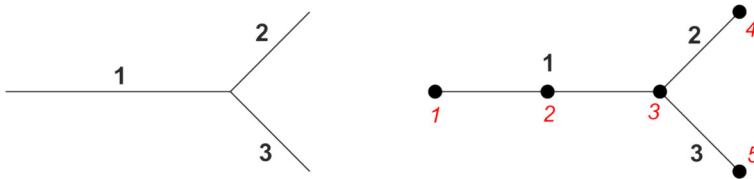


Fig. 1 Structure of the dendritic neuron [13]. Points on the structure represent places where the potential will be determined

determined. To make the mathematical description easier to comprehend, we introduce the symbol V_i for potential in a point placed on the branch where point i is, in time $t + \Delta t$ and in the respective point of branch x : $V_i = V_i^{(t+\Delta t)}(x)$; by analogy we have $V_{i\text{prev}} = V_i^{(t)}(x)$. This notation also applies to the variables $[m_3h_0]$ and $[n_4]$.

The following initial condition applies in the discretization of (1) [4],

$$\frac{\partial V_1^{(t)}(0)}{\partial x} = -\frac{R}{\pi a_1^2} i_0(t) \tag{4}$$

where [4]:

$$i_0(t) = I_0 t^2 e^{-10t} \tag{5}$$

and I_0 is a known value. The remaining boundary conditions are equal to 0.

Taking into consideration condition (4), we can write five equations, which describe all points of the model of the neural network given in Fig. 1 [13]:

$$\begin{aligned} F_1 &= C \frac{V_1 - V_{1\text{prev}}}{\Delta t} + g_{Na} [m_3h_0]_1 (V_1 - V_{Na}) + g_K [n_4]_1 (V_1 - V_K) \\ &\quad + g_L (V_1 - V_L) - \frac{a_1 I_0 t^2 e^{-10t}}{2\pi \Delta x} - a_1 \frac{V_2 - V_1}{2R \Delta x^2} \\ F_2 &= C \frac{V_2 - V_{2\text{prev}}}{\Delta t} + g_{Na} [m_3h_0]_2 (V_2 - V_{Na}) + g_K [n_4]_2 (V_2 - V_K) \\ &\quad + g_L (V_2 - V_L) - a_1 \frac{V_3 - 2V_2 + V_1}{2R \Delta x^2} \\ F_3 &= a_1^2 (V_3 - V_2) - a_2^2 (V_4 - V_3) - a_3^2 (V_5 - V_3) \\ F_4 &= C \frac{V_4 - V_{4\text{prev}}}{\Delta t} + g_{Na} [m_3h_0]_4 (V_4 - V_{Na}) + g_K [n_4]_4 (V_4 - V_K) \\ &\quad + g_L (V_4 - V_L) - a_2 \frac{V_4 - V_3}{2R \Delta x^2} \\ F_5 &= C \frac{V_5 - V_{5\text{prev}}}{\Delta t} + g_{Na} [m_3h_0]_5 (V_5 - V_{Na}) + g_K [n_4]_5 (V_5 - V_K) \\ &\quad + g_L (V_5 - V_L) - a_3 \frac{V_5 - V_3}{2R \Delta x^2} \end{aligned} \tag{6}$$

A detailed description of the discretization of a model of this type is given in [15].

It is noticeable that a biological neural network is modeled in a similar way to an artificial neural network. It has an input (current $i_0(t)$), output (the cell membrane potential V) while selected parameters of the model are treated as weights. In our case, the weights are g_{Na} , g_K , and g_L :

$$\begin{aligned} w_1 &= g_{Na} = \check{g}_{Na} e^{\check{g}_{Na}} \\ w_2 &= g_K = \check{g}_K e^{\check{g}_K} \\ w_3 &= g_L = \check{g}_L e^{\check{g}_L} \end{aligned} \tag{7}$$

where \check{g} are the initial values of conductances, given in Table 2, while \tilde{g} are proper weights of the model, which are updated during the training process.

3 Generalization terms

In this paper, three types of regularizer have been used to study the generalization ability of the biological neural network. As in dynamic neural networks, the regularizer is appended to the error function of a neural network [6]. Since the network considered in this paper is a five-point structure, the error function has the following form [15]:

$$E = \frac{1}{T} \sum_{i=1}^5 (V_i - V_i^*)^2 + \lambda E_\lambda. \tag{8}$$

The first term in 8 is the standard error of the neural network, where V_i^* is the reference voltage (the target of the training) in the i^{th} point of the neural network structure. The term λE_λ denotes the regularizer, where E_λ is the regularization functional and λ is the regularization parameter, which ensures a trade-off between the task performed by the regularizer and the accuracy of the network’s input-output mapping.

The first form of E_λ considered in this study is the so-called penalty function, given with the following formula:

$$E_\lambda = \frac{1}{T} \int_0^T \left(\sum_{i,j} w_{ij}^2 \right) dt = \frac{1}{T} \int_0^T (g_{Na}^2 + g_K^2 + g_L^2) dt. \tag{9}$$

To simplify the analysis, we have assumed that the weights are constants; therefore, we can omit the integral in (9).

Table 2 Parameters of Hodgkin-Huxley model

i	V_i [mV]	g_i [mS/cm ²]
Na	115	120
K	-12	36
L	10.6	0.3

The next considered form of E_λ is the Tikhonov functional:

$$E_\lambda = \frac{1}{2T} \int_0^T \|DV\|^2 dt = \frac{1}{2T} \int_0^T \left\| \frac{\partial V}{\partial I} \right\|^2 dt, \quad (10)$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Since in our case the derivative of potential V with respect to current I is a single value, the norm may be omitted in (10).

The last considered form of E_λ is the square norm of the network curvature (second-order differential operator):

$$E_\lambda = \frac{1}{2T} \int_0^T \left\| \frac{\partial^2 V}{\partial I^2} \right\|^2 dt. \quad (11)$$

The regularizers vary in the way they affect the neural network. The penalty function allows us to regularize the model's structure, which is essential since the overfitting of the neural network structure may lead to over-training the network. The Tikhonov functional introduces the additional assumption that the input-output mapping function should be smooth, which means that similar values of the input current I correspond to similar values of the output potential V [7]. Finally, as a result of minimization of the square norm of the network curvature, the derivative is locally invariant with respect to small changes in I . The application of this regularizer allows us to obtain a better mapping of the training set.

4 Training

The stochastic version of the kinetic model of a biological neural network is trained using the method of Lagrange multipliers. This method searches for the minimum (or maximum) of a certain function subject to specific constraints. In our case, we minimize the error function (8), while the constraints are represented by the system of equations (6). In order to find the minimum of function E , let us formulate the following auxiliary function:

$$L = E + \sum_{i=1}^5 \lambda_i F_i = \frac{1}{T} \sum_{i=1}^5 (V_i - V_i^*)^2 + \lambda E_\lambda + \sum_{i=1}^5 \lambda_i F_i \quad (12)$$

where λ_i are so-called Lagrange multipliers. We calculate the derivatives of function L with respect to all parameters of the model of the biological neural network. These parameters are $V_1, V_2, V_3, V_4, V_5, \tilde{g}_{Na}, \tilde{g}_K, \tilde{g}_L, \lambda_1, \lambda_2, \lambda_3, \lambda_4,$ and λ_5 . In the next step, we equate all derivatives to zero, thus obtaining a system of equations, whose solution is the minimum of function E . For clarity, the derivatives of (12) are given

below as three separate systems of Eqs. (13–15). Only the system of equations (14) changes depending on the form of regularizer used.

$$\left\{ \begin{aligned} \frac{\partial L}{\partial V_1} &= \frac{2}{T} (V_1 - V_1^*) - \lambda_1 \left(\frac{C}{\Delta t} + g_{Na} [m_3 h_0]_1 + g_K [n_4]_1 + g_L \right) \\ &\quad + \lambda_2 \frac{a_1}{2R(\Delta x)^2} \\ \frac{\partial L}{\partial V_2} &= \frac{2}{T} (V_2 - V_2^*) + \lambda_1 \frac{a_1}{2R(\Delta x)^2} \\ &\quad - \lambda_2 \left(\frac{C}{\Delta t} + g_{Na} [m_3 h_0]_2 + g_K [n_4]_2 + g_L + \frac{a_1}{R(\Delta x)^2} \right) \\ &\quad + \lambda_3 a_1^2 \\ \frac{\partial L}{\partial V_3} &= \frac{2}{T} (V_3 - V_3^*) + \lambda_2 \frac{a_1}{2R(\Delta x)^2} - \lambda_3 (a_1^2 + a_2^2 + a_3^2) \\ &\quad + \lambda_4 \frac{a_2}{2R(\Delta x)^2} + \lambda_5 \frac{a_3}{2R(\Delta x)^2} \\ \frac{\partial L}{\partial V_4} &= \frac{2}{T} (V_4 - V_4^*) + \lambda_3 a_2^2 \\ &\quad - \lambda_4 \left(\frac{C}{\Delta t} + g_{Na} [m_3 h_0]_4 + g_K [n_4]_4 + g_L + \frac{a_2}{2R} \right) \\ \frac{\partial L}{\partial V_5} &= \frac{2}{T} (V_5 - V_5^*) + \lambda_3 a_3^2 \\ &\quad - \lambda_5 \left(\frac{C}{\Delta t} + g_{Na} [m_3 h_0]_5 + g_K [n_4]_5 + \frac{a_3}{2R} \right) \end{aligned} \right. \tag{13}$$

$$\left\{ \begin{aligned} \frac{\partial L}{\partial g_{Na}} &= g_{Na} (-\lambda_1 [m_3 h_0]_1 (V_1 - V_{Na}) - \lambda_2 [m_3 h_0]_2 (V_2 - V_{Na}) \\ &\quad - \lambda_4 [m_3 h_0]_4 (V_4 - V_{Na}) - \lambda_5 [m_3 h_0]_5 (V_5 - V_{Na})) \\ &\quad + \lambda P_1 \\ \frac{\partial L}{\partial g_K} &= g_K (-\lambda_1 [n_4]_1 (V_1 - V_K) - \lambda_2 [n_4]_2 (V_2 - V_K) \\ &\quad - \lambda_4 [n_4]_4 (V_4 - V_K) - \lambda_5 [n_4]_5 (V_5 - V_K)) \\ &\quad + \lambda P_2 \\ \frac{\partial L}{\partial g_L} &= g_L (-\lambda_1 (V_1 - V_L) - \lambda_2 (V_2 - V_L) - \lambda_4 (V_4 - V_L) \\ &\quad - \lambda_5 (V_5 - V_L)) + \lambda P_3 \end{aligned} \right. \tag{14}$$

$$\left\{ \begin{aligned} \frac{\partial L}{\partial \lambda_1} &= C \frac{V_1 - V_{1prev}}{\Delta t} + g_{Na} [m_3 h_0]_1 (V_1 - V_{Na}) + g_K [n_4]_1 (V_1 - V_K) \\ &\quad + g_L (V_1 - V_L) - \frac{a_1 I_0 t^2 e^{-10t}}{2\pi \Delta x} - a_1 \frac{V_2 - V_1}{2R(\Delta x)^2} \\ \frac{\partial L}{\partial \lambda_2} &= C \frac{V_2 - V_{2prev}}{\Delta t} + g_{Na} [m_3 h_0]_2 (V_2 - V_{Na}) + g_K [n_4]_2 (V_2 - V_K) \\ &\quad + g_L (V_2 - V_L) - a_1 \frac{V_3 - 2V_2 + V_1}{2R(\Delta x)^2} \\ \frac{\partial L}{\partial \lambda_3} &= a_1^2 (V_3 - V_2) - a_2^2 (V_4 - V_3) - a_3^2 (V_5 - V_3) \\ \frac{\partial L}{\partial \lambda_4} &= C \frac{V_4 - V_{4prev}}{\Delta t} + g_{Na} [m_3 h_0]_4 (V_4 - V_{Na}) + g_K [n_4]_4 (V_4 - V_K) \\ &\quad + g_L (V_4 - V_L) - a_2 \frac{V_4 - V_3}{2R(\Delta x)^2} \\ \frac{\partial L}{\partial \lambda_5} &= C \frac{V_5 - V_{5prev}}{\Delta t} + g_{Na} [m_3 h_0]_5 (V_5 - V_{Na}) + g_K [n_4]_5 (V_5 - V_K) \\ &\quad + g_L (V_5 - V_L) - a_3 \frac{V_5 - V_3}{2R(\Delta x)^2} \end{aligned} \right. \tag{15}$$

If the regularizer is not applied, then $P_1 = P_2 = P_3 = 0$ (or equivalently $\lambda = 0$).

For the regularizer in the form of the penalty function, we have:

$$\left\{ \begin{aligned} P_1 &= \frac{2}{T} g_{Na}^2 \\ P_2 &= \frac{2}{T} g_K^2 \\ P_3 &= \frac{2}{T} g_L^2 \end{aligned} \right. \tag{16}$$

For the Tikhonov functional and for the square norm of the network curvature, P_1 , P_2 , and P_3 take the following form:

$$\begin{cases} P_1 = \frac{\frac{C}{\Delta t} P_{1prev} S - g_{Na} [m_3 h_0]_1 Q}{S^2} \\ P_2 = \frac{\frac{C}{\Delta t} P_{2prev} S - g_K [n_4]_1 Q}{S^2} \\ P_3 = \frac{\frac{C}{\Delta t} P_{3prev} S - g_L Q}{S^2} \end{cases} \tag{17}$$

where:

$$S = \frac{C}{\Delta t} + g_{Na} [m_3 h_0]_1 + g_K [n_4]_1 + g_L + \frac{a_1}{2R (\Delta x)^2} \tag{18}$$

$$Q = \frac{C}{\Delta t} \left(\frac{\partial V}{\partial I} \right)_{prev} + \frac{1}{2\pi \Delta x} \left(2a_1 I_0 t e^{-10t} - 10a_1 I_0 t^2 e^{-10t} \right) \tag{19}$$

for the Tikhonov functional, and:

$$Q = \frac{C}{\Delta t} \left(\frac{\partial^2 V}{\partial I^2} \right)_{prev} + \frac{1}{2\pi \Delta x} \left(2a_1 I_0 e^{-10t} - 40a_1 I_0 t e^{-10t} + 100a_1 I_0 t^2 e^{-10t} \right) \tag{20}$$

for the square norm of the network curvature.

5 Results

In order to assess the influence of the regularizer on the process of training a neural network model, we conducted a test consisting in noise reduction in images of electronic components. Figures 2(A₁) and 3(A₁) show two original images, to which noise was added in the following way: the lens was covered with thin plastic film before taking the picture (Fig. 2(B₁, C₁)), and the image was digitally blurred (Fig. 3(B₁, C₁)). The histograms in Figs. 2 and 3 (A₂, B₂, C₂) depict the difference between the original pictures and their distorted versions.

The task of the neural network model was to gradually reduce the noise in the images. The image was fed in the form of current I to the neural network input. The potential V obtained at any output point of the neural network structure was assumed to represent the denoised image, ideally as close to the original as possible. All final points of the considered structure rendered the same response as the neural network structure was symmetric.

In the test, we considered various values of the regularization factor λ ($\lambda = 0.01, 0.02, 0.1, 0.2, 0.5, 1, 2, 5, 10$), as well as the case $\lambda = 0$ (which is equivalent to not having a regularizer at all). We considered a number of measures to compare the results of the training process with and without the regularizer.

The first measure is the training time. Example training times of a biological neural network model for an approximately 100,000 pixel image are given in Table 3. Except for one case, the training time improves up to two times compared to the training time for the case without the regularizer ($\lambda = 0$).

The root mean square error (RMSE) is the second measure. Example RMSE results for a reconstructed image of approximately 100,000 pixels are given in Table 4. For $\lambda = 0$ (no regularizer used), RMSE is of the order of 10^{-2} . The addition

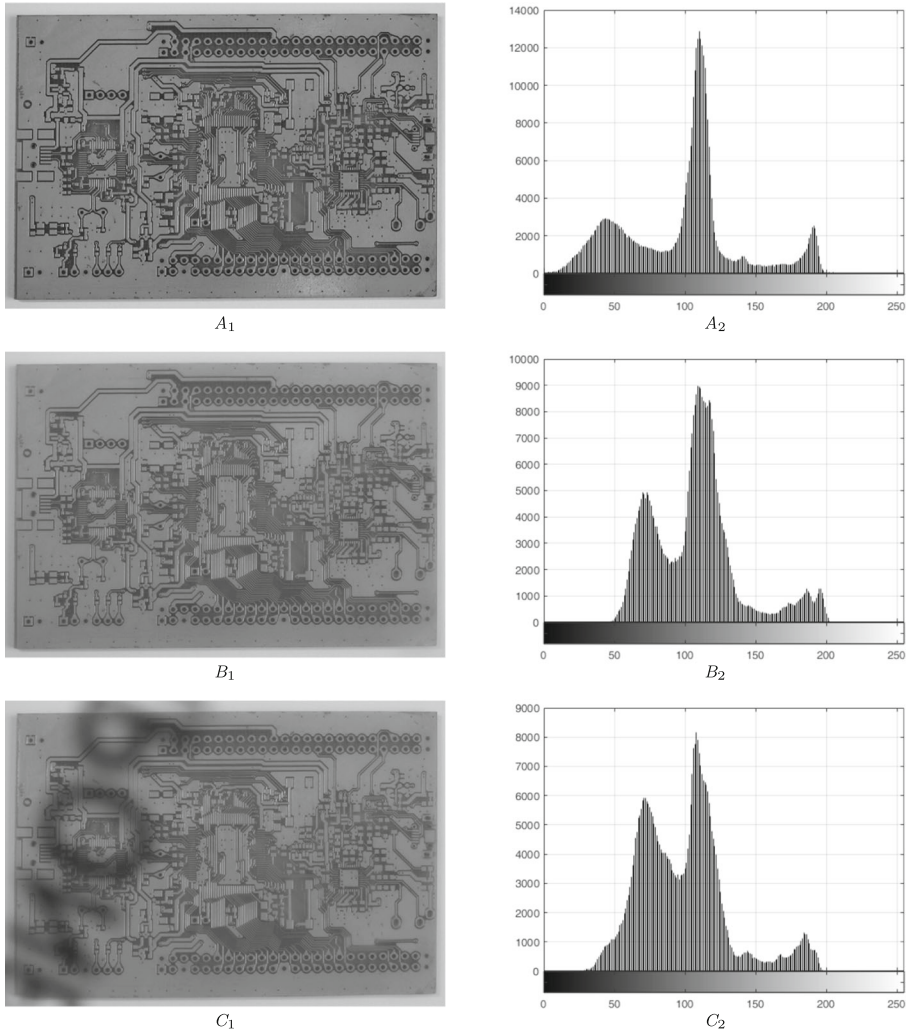


Fig. 2 Image of a circuit board (A_1) and two noisy versions of the image, (B_1) and (C_1), obtained by placing a thin plastic film against the camera lens. Figures (A_2), (B_2), and (C_2) are the histograms of (A_1), (B_1), and (C_1) respectively

of the penalty function and the Tikhonov functional to the error function had a negligible impact on RMSE for most values of λ , while the introduction of the square norm of the network curvature improved the RMSE by one order of magnitude.

Finally, the `fsolve` MATLAB function allows us to determine the so-called first-order optimality measure, which says how close the obtained solution is from the sought solution. This measure is related to the Karush-Kuhn-Tucker conditions of the method of Lagrange multipliers [11]. The measure must be zero at a minimum (a sufficient condition for optimality of a solution), but a point with the measure equal to zero may not be a minimum.

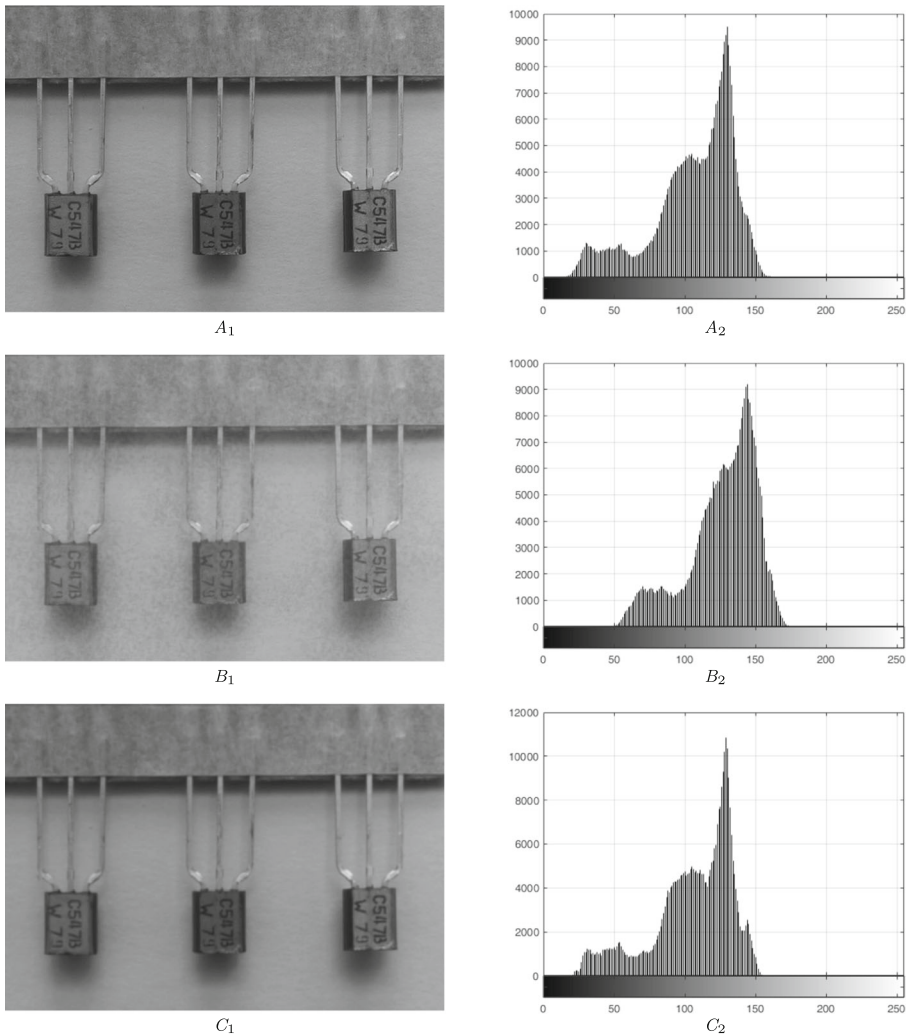


Fig. 3 Image of integrated circuits (A_1) and its two noisy versions, (B_1) and (C_1). Figures (A_2), (B_2), and (C_2) are the histograms of (A_1), (B_1), and (C_1), respectively

Figure 4 shows the minimal, maximal, and average distance between the vectors representing the original image and the image reconstructed by the model of a biological neural network. Except for individual cases, these values improve when a regularizer is added to the error function, compared to the case when $\lambda = 0$.

Figure 5 features example images which have been poorly reconstructed by the neural network model. The respective histograms in Fig. 5 show the difference between the reconstructions and the original images. The most difficult areas to be reconstructed are the edges and the dark parts of the images. Poor reconstruction of images may take place when the model is over-trained and consequently it cannot

Table 3 Example training time for a biological neural network model, for different regularizers and different values of λ

λ	Penalty	Tikhonov	Curvature
0	2224.8936		
0.01	1002.5165	3861.6031	1445.3085
0.02	999.6127	1242.6350	1000.4943
0.1	1005.8427	1272.6241	1002.3842
0.2	1010.0208	1302.0676	1002.4459
0.5	1010.1994	1283.0306	1001.4347
1	1011.1937	1282.3941	1008.6745
2	1010.9655	1265.2496	1003.3383
5	1010.3824	1264.6278	1005.6711
10	1010.8224	1260.9133	1010.0293

Table 4 Example values of RMSE between original image and images reconstructed by a biological neural network model, for different regularizers and different values of λ

λ	Penalty	Tikhonov	Curvature
0	0.0107		
0.01	0.0084	0.0118	0.0057
0.02	0.0115	0.0144	0.0034
0.1	0.0127	0.0122	0.0034
0.2	0.0076	0.0074	0.0033
0.5	0.0135	0.0149	0.0033
1	0.0114	0.0101	0.0033
2	0.0070	0.0105	0.0033
5	0.0160	0.0138	0.0033
10	0.0100	0.0081	0.0033

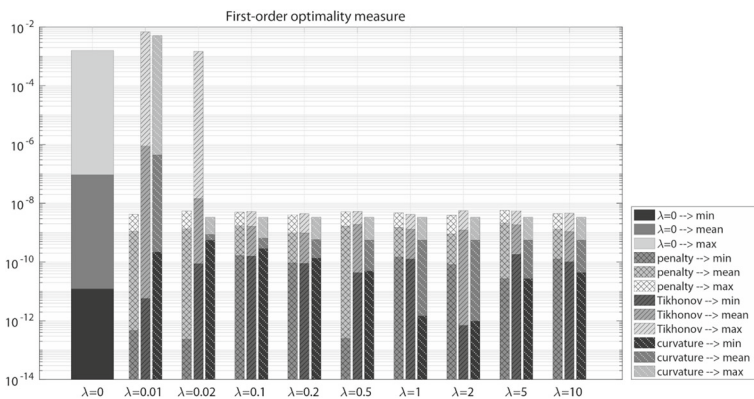


Fig. 4 Minimal, maximal, and mean distance between vectors representing the original image and the image reconstructed by a biological neural network model, for different regularizers and for $\lambda = 0$

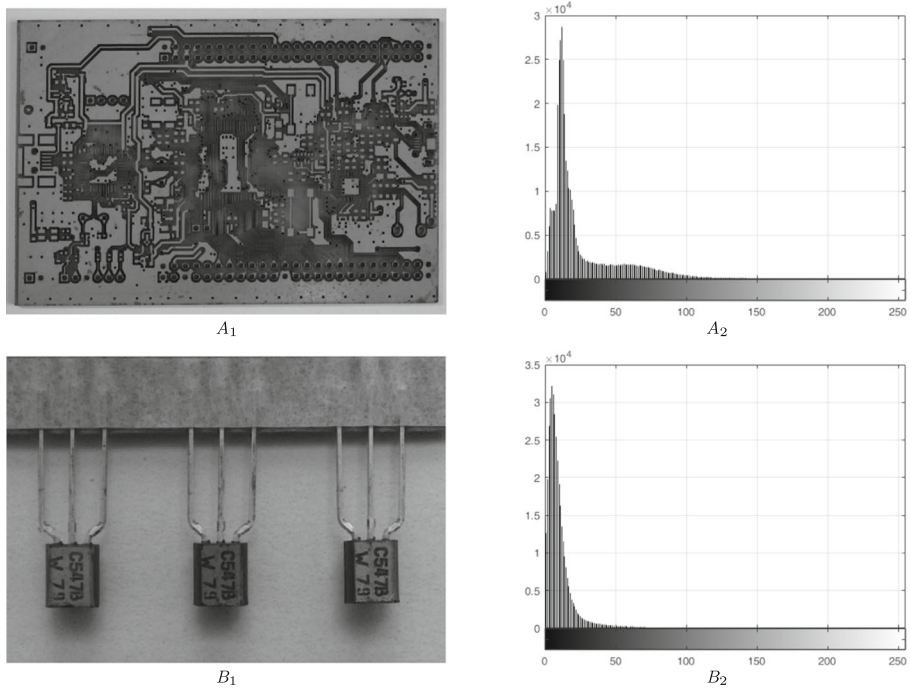


Fig. 5 Poor reconstructions of original images from Figs. 2 and 3 (A_1), (B_1)). Histograms (A_2) and (B_2) show the difference between the reconstructions and the original images

find a solution if the processed image is not in the training set. Also, during initialization, the values of weights are selected randomly; as a result, the algorithm which calculates the minimum error function may find a local minimum which it cannot leave. However, these situations are exceptional in case of the presented model of a biological neural network. In most cases, the network is able to reconstruct images which are very close to the original.

6 Summary

The paper analyzed the generalization ability of a stochastic kinetic model of a biological neural network. The neural network model was based on the Hodgkin-Huxley model extended with Markov's kinetic schemes, which describe in detail the processes taking place on the cell membrane. The generalization ability of the model was studied using a method known from the regularization theory, which consists in adding a regularizer to the error function of the biological neural network model.

The following forms of regularizer were studied: the penalty function, the Tikhonov functional, and the square norm of the network curvature. Different regularizers affect the training process of the neural network structure in a different way, which results from the way they change the neural network structure (it is minimized) or the input-output mapping (it is smoothened). The results presented in the paper

show that the use of the regularizer improves the measures used in comparisons, such as RMSE or the first-order optimality measure. The most visible effect is that of the square norm of the network curvature, thanks to which RMSE is improved by up to one order of magnitude, and the first-order optimality measure is improved by several orders of magnitude, from values of the order 10^2 to values of the order 10^8 .

Acknowledgements I would like to offer my special thanks to Krzysztof Kolanowski for taking Figs. 2 and 3(A₁), and to Agata Jurkowlaniec for preparing the images so that they could be used to train the biological neural network model.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Avriel, M.: *Nonlinear Programming: Analysis and Methods*, 2nd edn. Dover Publ Inc., USA (2003)
2. Bishop, C.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
3. Bower, J.M., Beeman, D.: *The Book of GENESIS, Exploring Realistic Neural Models with the GEneral NEural Simulation System*. Internet Edition (2003)
4. Cox, S.J., Griffith, B.E.: A fast, fully implicit backward Euler solver for dendritic neurons. Technical report, Rice University Department of Computational and Applied Mathematics (2000)
5. Destexhe, A., Mainen, Z.F., Sejnowski, T.J.: Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *J. Comput. Neurosci.* **3**(1), 195–230 (1994)
6. Galicki, M., Leistriz, L., Zwick, E.B., Witte, H.: Improving generalization capabilities of dynamic neural networks. *Neural Comput.* **16**, 1253–1282 (2004)
7. Haykin, S.O. *Neural Networks: A Comprehensive Foundation*, 2nd edn. Pearson Education, London (1998)
8. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544 (1952)
9. Koch, C.: *Biophysics of Computation: Information Processing in Single Neurons*. Chapter VI. Oxford University Press, New York (1999)
10. Krogh, A., Hertz, J.A.: A simple weight decay can improve generalization. In: *Advances in Neural Information Processing Systems 4*, pp. 950–957. Morgan Kaufmann (1992)
11. Nocedal, J., Wright, S.J. *Numerical Optimization*, 2nd edn. Springer, Berlin (2000)
12. Schneidman, E., Segev, I., Tishby, N.: Information capacity and robustness of stochastic neuron models. In: *Advances in Neural Information Processing*, vol. 12 (2000)
13. Świetlicka, A.: Trained stochastic model of biological neural network used in image processing task. *Appl. Math. Comput.* **267**, 716–726 (2015)
14. Świetlicka, A., Gugala, K., Jurkowlaniec, A., Śniatała, P., Rybarczyk, A.: The stochastic, Markovian, Hodgkin-Huxley type of mathematical model of the neuron. *Neural Network World* **25**(1), 219–239 (2015)
15. Świetlicka, A., Gugala, K., Pedrycz, W., Rybarczyk, A.: Development of the deterministic and stochastic Markovian model of a dendritic neuron. *Bioprocess Biosyst. Eng.* **37**, 201–216 (2017)
16. Świetlicka, A., Kolanowski, K., Kapela, R., Galicki, M., Rybarczyk, A.: Investigation of generalization ability of a trained stochastic kinetic model of neuron. *Appl. Math. Comput.* **319**, 115–124 (2018)