symmetry

MDPI

Article

On Optimization Techniques for the Construction of an Exponential Estimate for Delayed Recurrent Neural Networks

Vasyl Martsenyuk *,†,‡, **Stanislaw Rajba** ‡ and **Mikolaj Karpinski** ‡

Department of Computer Science and Automatics, Faculty of Mechanical Engineering and Computer Science, University of Bielsko-Biala, 43-309 Bielsko-Biala, Poland; rajbas@ath.bielsko.pl (S.R.); mkarpinski@ath.bielsko.pl (M.K.)

* Correspondence: vmartsenyuk@ath.bielsko.pl; Tel.: +48-338-279-264

† Current address: 2 Willowa, 43-309 Bielsko-Biala, Poland.

‡ These authors contributed equally to this work.

Received: 31 August 2020; Accepted: 15 October 2020; Published: 20 October 2020

check for updates

Abstract: This work is devoted to the modeling and investigation of the architecture design for the delayed recurrent neural network, based on the delayed differential equations. The usage of discrete and distributed delays makes it possible to model the calculation of the next states using internal memory, which corresponds to the artificial recurrent neural network architecture used in the field of deep learning. The problem of exponential stability of the models of recurrent neural networks with multiple discrete and distributed delays is considered. For this purpose, the direct method of stability research and the gradient descent method is used. The methods are used consequentially. Firstly we use the direct method in order to construct stability conditions (resulting in an exponential estimate), which include the tuple of positive definite matrices. Then we apply the optimization technique for these stability conditions (or of exponential estimate) with the help of a generalized gradient method with respect to this tuple of matrices. The exponential estimates are constructed on the basis of the Lyapunov–Krasovskii functional. An optimization method of improving estimates is offered, which is based on the notion of the generalized gradient of the convex function of the tuple of positive definite matrices. The search for the optimal exponential estimate is reduced to finding the saddle point of the Lagrange function.

Keywords: recurrent neural network; delayed differential equations; exponential estimation; optimization method; generalized gradient.

---

1. Introduction

Breakthrough results in the field of deep machine learning are obtained nowadays using recurrent neural networks (RNN). In particular, the construction of machine learning models for problems of image recognition with captioning, natural language processing and translation, was made possible by recurrent neural networks with Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). The paper [1] offers a description of such models using ordinary differential equations. Further research has to be related to the systems with time delays as they are modeling the memory within the network units. In [2] drawing from concepts in signal processing, they formally derived the canonical RNN formulation from differential equations.

Here our study of the RNN model is based on the system with multiple discrete and distributed time-varying delays

$$\dot{x}(t) = -Ax(t) + \sum_{k=1}^{r_1} W_{1,k} g(x(t - h_k(t))) + \sum_{m=1}^{r_2} W_{2,m} \int_{t-\tau_m(t)}^{t} g(x(\theta)) d\theta, \quad (1)$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $A = diag(a_1, a_2, ..., a_n)$ is a diagonal matrix with positive entries $a_i > 0$. For the i-th neuron $1/a_i$ can be interpreted as the activity decay constant (or time constant). $W_{1,k} = (w_{ij}^{1,k})_{n \times n}$, $k = \overline{1, r_1}$, $W_{2,m} = (w_{ij}^{2,m})_{n \times n}$, $m = \overline{1, r_2}$ are the synaptic connection weight matrices. The entries of $W_{1,k}$ and $W_{2,m}$ may be positive (excitatory synapses) or negative (inhibitory synapses). $g(x(t)) = [g_1(x(t)), g_2(x(t)), ..., g_n(x(t))]^\top \in \mathbb{R}^n$ is the non-decreasing activation function, which belongs to sector non-linear function class defined by

$$g_j(0) = 0 \quad \text{and} \quad 0 \leq \frac{g_j(\xi_1) - g_j(\xi_2)}{\xi_1 - \xi_2} \leq l_j, \quad l_j > 0, \quad (2)$$

$\xi_1, \xi_2 \in \mathbb{R}$, $\xi_1 \neq \xi_2$, $j \in \overline{1, n}$ and $x = 0$ is a fixed point of Equation (1). We let $L = diag(l_1, l_2, ..., l_n)$ is a diagonal matrix with positive entries $l_j > 0$.

The system (1) includes discrete and distributed time-varying delays, which are described with the help of the second and the third terms correspondingly.
The bounded differentiable functions $h_k(t)$ represent discrete delays of system with

$$0 \leq h_k(t) \leq h_{M,k},$$

and

$$\dot{h}_k(t) \leq h_{D,k} < 1, \quad (3)$$

$k = \overline{1, r_1}$, $t > 0$. Delays $h_k(t)$ and $\tau_m(t)$ have physical meaning as "controllable memory" if previous states of neurons effects on output only during some time intervals. $h_{M,k}$ and $h_{D,k}$ are bounds of the delay and its derivative for discrete delays.

The bounded functions $\tau_m(t)$ represent distributed delays of system with $0 \leq \tau_m(t) \leq \tau_{M,m}$, $m = \overline{1, r_2}$.

The bounded functions $h_k(t)$ and $\tau_m(t)$ represent axonal signal transmission delays. The condition (3) for derivative $\dot{h}_k(t)$ will be applied when estimating the upper right derivative of Lyapunov–Krasovskii functional (see, for example, [3]).

The initial conditions associated with system (1) are assumed to be

$$\begin{aligned} x(s) =& \phi(s), \quad s \in [-\tau_M, 0], \\ \tau_M :=& \max\left\{h_{M,k}, k = \overline{1, r_1}, \tau_{M,m}, m = \overline{1, r_2}\right\}, \end{aligned} \quad (4)$$

where $\phi(s) \in C[-\tau_M, 0]$.

Given any $\phi(s) \in C[-\tau_M, 0]$, under the assumption (2), there exists a unique trajectory of (1) starting from ϕ [4].

Here we use the Hopfield neural network model, which includes a diagonal matrix A with positive entries, that shows the self-connection of the neuron. That is the next state of the neuron is dependent on its current state and outputs of eventually all neurons. Such a diagonal matrix is traditionally applied in stability research of continuous-time RNNs [5]. On the other hand, if we used arbitrary matrix A, we would assume the next state of a neuron is dependent on its current state as well as the states of all other neurons, which means that the internal states of all neurons are accessible from outside. It contradicts that, for example, in the case of the LSTM unit only the hidden state vector (also known as the output vector) is seen.

Following the work [2], we may give the interpretation of the model (1) from the viewpoint of signal processing, leading us to "canonical" and "non-canonical" RNNs. Namely, we have $x(t)$, the state signal vector; $g(x(t))$, the readout signal vector, which is a warped version of the state signal vector; the bias parameters are omitted without loss of generality since they can be used in the transformation resulting in the homogeneous system (1). Initial state $\phi(s)$, $s \in [-\tau_M, 0]$ is considered as an input signal, thus, modeling one-to-many RNN architecture. In more general many-t0-many case, input signal vector $u(t)$, $t > 0$ can be used during the work of the RNN as an "input sequence".

Although RNNs can actually be described using difference equations, it makes sense to consider continuous-time equations that describe the operation of RNNs. This is due to the fact that differential equations make it possible to better describe and understand the dynamic processes that occur. In addition, with the help of differential equations it is possible to explicitly obtain the conditions for stabilization of recurrent neural networks. This is of great importance in the design of recurrent neural networks. Ref. [5] provide a comprehensive review of the research on the continuous-time recurrent neural networks focusing on the stability of Hopfield and Cohen–Grossberg neural networks.

Note also that the corresponding recurrent neural networks can be obtained by discretizing the models based on differential equations. Thus, work [2] shows the way to construct a RNN of the LSTM type starting from the corresponding model based on differential equations with delay and further through the discretization of the so-called canonical RNN.

Recurrent neural network models have been considered from the 1980s, after the pioneering work of Hopfield [6] modeling each neuron as a linear circuit consisting of a resistor and a capacitor. Two approaches can be differed when investigating the models of recurrent neural networks in the class of delayed differential equations. The first way means research of local stability with the help of comparison with the linearised system [7–10]. The conditions of the Hopf bifurcation were obtained in [10,11]. The second approach (which is called the direct Lyapunov's method) uses Lyapunov–Krasovkii functionals [3]. It allows us to get stability conditions constructively, which are formulated in the form of linear-matrix inequalities (LMIs). These stability conditions can be improved with the help of optimization of parameters of Lyapunov–Krasovskii functionals.

Exponential estimates of the solutions are very important when investigating the models of RNNs because they show the rate of convergence of calculations when recognizing input data. In the previous works [12,13], indirect method was developed allowing us to get exponential estimates in some general cases of the RNN models. It results in the numerical solution of quasipolynomial equation. It gives us a clear value of exponential decay, which, unfortunately, does not admit optimization and so, cannot be improved. In order to overcome this shortcoming, here we develop an optimization technique, which is based on the direct method for Liapunov–Krasovskii functionals of the special kind.

2. Exponential Estimate

Let $\Omega_n \subset \mathbb{R}^{n\times n}$ be a set of all symmetric positive definite matrices. It is an open convex cone because:

(a) convexity—for any $P_1 \in \Omega_n$, $P_2 \in \Omega_n$, $x \in \mathbb{R}^n$, and $\xi \in [0,1]$ we have $x^\top(\xi P_1 + (1-\xi)P_2)x = \xi x^\top P_1 x + (1-\xi)x^\top P_2 x > 0$;

(b) cone—for any $P \in \Omega_n$, $x \in \mathbb{R}^n$, and $\eta > 0$ we have $\eta x^\top P x > 0$.

$\bar{\Omega}_n^1$ is the part of the cone Ω_n contained inside the unit sphere, i.e., $\bar{\Omega}_n^1 := \{P \in \Omega_n : \|P\| \leq 1\}$. Here $\|P\|$ is Frobenius norm of the matrix $P \in \Omega_n$.

Lemma 1. *Reference [14] For any constant matrix $U \in \Omega_n$, scalar $\beta > 0$, vector function $u : [0,\beta] \to \mathbb{R}^n$ such that the integrations concerned are well-defined, then*

$$\Big(\int_0^\beta u^\top(s)ds\Big)U\Big(\int_0^\beta u(s)ds\Big) \leq \beta \int_0^\beta u^\top(s)Uu(s)ds.$$

Lemma 2. *Reference [15] Given any real matrices W_1, W_2, W_3 with appropriate dimensions and a scalar $\beta > 0$, $W_3 \in \Omega_n$, then the following inequality holds*

$$W_1^\top W_2 + W_2^\top W_1 \leq \beta W_1^\top W_3 W_1 + \beta^{-1} W_2^\top W_3^{-1} W_2.$$

In the following definitions, we assume that the trivial solution of (1) be the unique equilibrium point of the model (1).

Definition 1. *The trivial solution of (1) is globally asymptotically stable if for every solution $x(t)$ to the initial value problem (1)–(4), we have $x(t) \to 0$ as $t \to \infty$.*

Definition 2. *If there exist constants $\alpha > 0$, $K > 0$, and $T > 0$ such that every solution $x(t)$ to the initial value problem (1)–(4) always satisfies $\|x(t)\| \leq Ke^{-\alpha t}$ for all $t > T$, then the trivial solution of (1) is said to be globally exponentially stable.*

Our research is based on the following Lyapunov–Krasovskii functional, which is an extension of the one offered in the work [3] for the case of multiple delays

$$\begin{aligned} V[x_t(\cdot)] = e^{2\alpha t} x^\top(t) P x(t) + \sum_{k=1}^{r_1} \int_{t-h_k(t)}^{t} e^{2\alpha s} g^\top(x(s)) Q_k g(x(s)) ds \\ + \sum_{m=1}^{r_2} \tau_{M,m} \int_{-\tau_{M,m}}^{0} \int_{t+\theta}^{t} e^{2\alpha s} g^\top(x(s)) S_m g(x(s)) ds d\theta, \end{aligned} \tag{5}$$

where unknown constant $\alpha > 0$ and matrices P, Q_k, $k = \overline{1, r_1}$, S_m, $m = \overline{1, r_2}$ belong to Ω_n. Here we use traditional denotation of the element of the solution of (1) as the vector-interval $x_t(\cdot) := \{x(t+\theta) | \theta \in [-\tau_M, 0]\} \in C[-\tau_M, 0]$.

Theorem 1. *We assume that system (1) satisfies the following condition.*

H1. Let there exist constant $\alpha > 0$ and matrices P, Q_k, $k = \overline{1, r_1}$, S_m, $m = \overline{1, r_2}$, which belong to relint$(\bar{\Omega}_n^1)$, such that the symmetric matrix

$$\begin{aligned}
&\Gamma := \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix}, \\
&\Gamma_{11} := -2\alpha P + A^\top P + PA - L\Big(\sum_{k=1}^{r_1} Q_k + \tau_{M,m}^2 \sum_{m=1}^{r_2} S_m \Big) L \in \Omega_n, \\
&\Gamma_{12} := \left[\frac{e^{\alpha h_{M,1}}}{\sqrt{1 - h_{D,1}}} P W_{1,1} \quad \cdots \quad \frac{e^{\alpha h_{M,r_1}}}{\sqrt{1 - h_{D,r_1}}} P W_{1,r_1} \quad e^{\alpha \tau_{M,1}} P W_{2,1} \quad \cdots \quad e^{\alpha \tau_{M,r_2}} P W_{2,r_2} \right] \in \mathbb{R}^{n \times n(r_1 + r_2)}, \\
&\Gamma_{21} := \Gamma_{12}^\top \in \mathbb{R}^{n(r_1+r_2) \times n}, \\
&\Gamma_{22} := \begin{bmatrix} Q_1 & & & & & \Theta \\ & \ddots & & & & \\ & & Q_{r_1} & & & \\ & & & S_1 & & \\ & & & & \ddots & \\ \Theta & & & & & S_{r_2} \end{bmatrix} \in \Omega_{n(r_1+r_2)}, \\
&\quad \Theta \in \mathbb{R}^{n \times n} \text{ is matrix of zeroes,}
\end{aligned} \tag{6}$$

belong to $\Omega_{n(1+r_1+r_2)}$.

Then the trivial solution of (1) is globally asymptotically stable.

Proof. Estimating right upper derivative of the functional $V[x_t(\cdot)]$ along the solution of the system (1), we get

$$
\begin{aligned}
&\frac{dV^+[x_t(\cdot)]}{dt} \le e^{2\alpha t}\{x^\top(t)[2\alpha P - A^\top P - PA]x(t)\\
&+ \Big[\sum_{k=1}^{r_1} g^\top(x(t-h_k(t)))W_{1,k}^\top Px(t) + x^\top(t)P\sum_{k=1}^{r_1} W_{1,k}g(x(t-h_k(t)))\Big]\\
&+ \Big[\sum_{m=1}^{r_2}\int_{t-\tau_m(t)}^{t} g^\top(x(\theta)d\theta)W_{2,m}^\top Px(t) + x^\top(t)P\sum_{m=1}^{r_2} W_{2,m}\int_{t-\tau_m(t)}^{t} g(x(\theta))d\theta\Big]\}\\
&+ \sum_{k=1}^{r_1} e^{2\alpha t}\{g^\top(x(t))Q_k g(x(t))\\
&- e^{-2\alpha h_k(t)}g^\top(x(t-h_k(t)))Q_k g(x(t-h_k(t)))(1-h_{D,k})\}\\
&+ \sum_{m=1}^{r_2}\tau_{M,m}\{\tau_{M,m}e^{2\alpha t}g^\top(x(t))S_m g(x(t))\\
&- e^{2\alpha(t-\tau_{M,m})}\int_{t-\tau_m(t)}^{t} g^\top(x(s))S_m g(x(s))ds\}.
\end{aligned}
\tag{7}
$$

Applying Lemmas 1 and 2 for estimating counterparts of (7), we have

$$
\begin{aligned}
&\sum_{k=1}^{r_1}\{g^\top(x(t-h_k(t)))W_{1,k}^\top Px(t) + x^\top(t)PW_{1,k}g(x(t-h_k(t)))\}\\
&= \sum_{k=1}^{r_1}[e^{-\alpha h_{M,k}}(1-h_{D,k})^{1/2}g^\top(x(t-h_k(t)))][e^{\alpha h_{M,k}}(1-h_{D,k})^{-1/2}W_{1,k}^\top Px(t)]\\
&+ [e^{\alpha h_{M,k}}(1-h_{D,k})^{-1/2}x^\top(t)PW_{1,k}][e^{-\alpha h_{M,k}}(1-h_{D,k})^{1/2}g(x(t-h_k(t)))]\\
&\le \sum_{k=1}^{r_1}\{e^{2\alpha h_{M,k}}(1-h_{D,k})^{-1}x^\top(t)PW_{1,k}Q_k^{-1}W_{1,k}^\top Px(t)\\
&+ e^{-2\alpha h_{M,k}}(1-h_{D,k})g^\top(x(t-h_k(t)))Q_k g(x(t-h_k(t)))\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{m=1}^{r_2}\{\int_{t-\tau_m(t)}^{t} g^\top(x(\theta))d\theta W_{2,m}\top Px(t) + x^\top(t)PW_{2,m}\int_{t-\tau_m(t)}^{t} g(x(\theta))d\theta\}\\
&= \sum_{m=1}^{r_2}\{[e^{-\alpha\tau_{M,m}}\int_{t-\tau_m(t)}^{t} g^\top(x(\theta))d\theta][e^{\alpha\tau_{M,m}}W_{2,m}^\top Px(t)]\\
&+ [e^{\alpha\tau_{M,m}}x^\top(t)PW_{2,m}][\int_{t-\tau_m(t)}^{t} g(x(\theta))d\theta e^{-\alpha\tau_{M,m}}]\}\\
&\le \sum_{m=1}^{r_2}\{e^{-2\alpha\tau_{M,m}}\Big(\int_{t-\tau_m(t)}^{t} g^\top(x(\theta))d\theta\Big)S_m\Big(\int_{t-\tau_m(t)}^{t} g(x(\theta))d\theta\Big)\\
&+ e^{2\alpha\tau_{M,m}}x^\top(t)PW_{2,m}S_m^{-1}W_{2,m}^\top Px(t)\}\\
&\le \sum_{m=1}^{r_2}\{\tau_{M,m}e^{-2\alpha\tau_{M,m}}\int_{t-\tau_m(t)}^{t} g^\top(x(\theta))S_m g(x(\theta))d\theta\\
&+ e^{2\alpha\tau_{M,m}}x^\top(t)PW_{2,m}S_m^{-1}W_{2,m}^\top Px(t)\}.
\end{aligned}
$$

Finally we get

$$
\begin{aligned}
\frac{dV^{+}[x_t(\cdot)]}{dt} &\le e^{2\alpha t}x^{\top}(t)\{2\alpha P - A^{\top}P - PA \\
&+ \sum_{k=1}^{r_1}\Big[e^{2\alpha h_{M,k}}(1-h_{D,k})^{-1}PW_{1,k}Q_k^{-1}W_{1,k}^{\top}P + LQ_kL\Big] \\
&+ \sum_{m=1}^{r_2}\Big[e^{2\alpha\tau_{M,m}}PW_{2,m}S_m^{-1}W_{2,m}^{\top}P + \tau_{M,m}^2LS_mL\Big]\}x(t) \\
&\le -e^{2\alpha t}x^{\top}(t)\{\Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}\}x(t) \\
&= -e^{2\alpha t}x^{\top}(t)\Gamma/\Gamma_{22}x(t),
\end{aligned} \tag{8}
$$

where $\Gamma/\Gamma_{22} := \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}$ is the Schur complement of Γ in Γ_{22}. From the Schur complement it follows that the right side of (8) is negative definite if and only if $\Gamma \in \Omega_{n(1+r_1+r_2)}$ [16]. □

Corollary 1. *Provided that the condition H1 holds the trivial solution of (1) is globally exponentially stable as follows*

$$
\|x(t)\| \le \gamma(\alpha)|\phi|_{\tau_M}e^{-\alpha t}, \quad t > 0, \tag{9}
$$

where

$$
\begin{aligned}
\gamma(\alpha) :=& \lambda_{\min}^{-1/2}(P)\Big(\lambda_{\max}(P) + \sum_{k=1}^{r_1}\lambda_{\max}(Q_k)l_{\max}^2\frac{1-e^{-2\alpha h_{M,k}}}{2\alpha} \\
&+ \sum_{m=1}^{r_2}\tau_{M,m}\lambda_{\max}(LS_mL)\frac{2\alpha\tau_{M,m} - 1 + e^{-2\alpha\tau_{M,m}}}{4\alpha^2}\Big)^{1/2}, \\
l_{\max} :=& \max\{l_1, \dots, l_n\},
\end{aligned}
$$

$\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are minimal and maximal eigenvalues of the matrix. Here we use denotions of $\|\cdot\|$ as Euclidean norm in $\mathbb{R}^n$ and $|x(\cdot)|_{\tau_M} := \sup_{s\in[-\tau_M,0]}\|x(s)\|$ as the uniform convergence norm in $C[\tau_M, 0]$.

Proof. Firstly note that the inequality

$$
2\alpha\tau_{M,m} + e^{-2\alpha\tau_{M,m}} \ge 1,
$$

enables us that the square root expression in $\gamma(\alpha)$ is nonnegative for $\alpha > 0$. From Theorem 1 it follows that $V[x_t(\cdot)] \le V[\phi(\cdot)]$. Hence, we get

$$
\begin{aligned}
\lambda_{\min}(P)\|x(t)\|^2 &\le e^{-2\alpha t}V[x_t(\cdot)] \le e^{-2\alpha t}V[\phi(\cdot)] \\
&\le e^{-2\alpha t}\Big(\phi^{\top}(0)P\phi(0) + \sum_{k=1}^{r_1}\int_{-h_{M,k}}^{0}e^{2\alpha s}g^{\top}(x(s))Q_kg(x(s))ds \\
&+ \sum_{m=1}^{r_2}\tau_{M,m}\int_{-\tau_{M,m}}^{0}\int_{\theta}^{0}e^{2\alpha s}g^{\top}(\phi(s))S_mg(\phi(s))dsd\theta\Big) \\
&\le e^{-2\alpha t}\Big(\lambda_{\max}(P) + \sum_{k=1}^{r_1}\lambda_{\max}(Q_k)l_{\max}^2\int_{-h_{M,k}}^{0}e^{2\alpha s}ds \\
&+ \sum_{m=1}^{r_2}\tau_{M,m}\lambda_{\max}(LS_mL)\int_{-\tau_{M,m}}^{0}\int_{\theta}^{0}e^{2\alpha s}dsd\theta\Big)|\phi|_{\tau_M}^2 \\
&= e^{-2\alpha t}\Big(\lambda_{\max}(P) + \sum_{k=1}^{r_1}\lambda_{\max}(Q_k)l_{\max}^2\frac{1-e^{-2\alpha h_{M,k}}}{2\alpha} \\
&+ \sum_{m=1}^{r_2}\tau_{M,m}\lambda_{\max}(LS_mL)\frac{2\alpha\tau_{M,m} - 1 + e^{-2\alpha\tau_{M,m}}}{4\alpha^2}\Big)|\phi|_{\tau_M}^2.
\end{aligned}
$$

Finally, it yields

$$\lambda_{\min}(P)\|x(t)\|^2 \le e^{-2\alpha t}\gamma(\alpha)|\phi|^2_{\tau_M}.$$

□

3. Optimization Method

The condition H1 describes the main stability result. The matrix Γ presents operator, which is linear with the respect to the tuple of matrices $[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$. At the same time, the dependence of Γ on α is nonlinear. α is the parameter determining the exponential decay rate. Since optimization of Γ with the respect to $[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$ may be considered as linear matrix inequality, we can reduce the problem of the construction of exponential estimate like (9) to the convex programming. It is natural to assume that "the more positive definite" is the matrix Γ, the "more asymptotically stable" is the trivial solution. In turn, the solution is "more exponentially" stable. The positive definiteness of the matrix Γ is described with the help of minimum eigenvalue. Thus we result in the following optimization problem.

Further we apply optimization technique developed earlier for linear systems in [17]. Given $\alpha > 0$ we search the tuple of matrices $(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2})$ as a solution of the optimization problem

$$[P^\star, Q_1^\star, \ldots, Q_{r_1}^\star, S_1^\star, \ldots, S_{r_2}^\star] = \arg \inf_{\substack{[P, Q_1, \ldots, Q_{r_1}, \\ S_1, \ldots, S_{r_2}] \\ \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1}} \psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]. \tag{10}$$

Here $\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] = -\lambda_{\min}(\Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}])$.

We give some general conditions of existence of a solution of problem (10).

Definition 3. *The inner product of the tuples of matrices* $[P_1, Q_{1,1}, \ldots, Q_{r_1,1}, S_{1,1}, \ldots, S_{r_2,1}]$, $[P_2, Q_{1,2}, \ldots, Q_{r_1,2}, S_{1,2}, \ldots, S_{r_2,2}] \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ *is*

$$\begin{aligned}&\langle [P_1, Q_{1,1}, \ldots, Q_{r_1,1}, S_{1,1}, \ldots, S_{r_2,1}], [P_2, Q_{1,2}, \ldots, Q_{r_1,2}, S_{1,2}, \ldots, S_{r_2,2}] \rangle \\ &:= \sum_{i,j=1}^{n} \Big(p_{ij}^1 p_{ij}^2 + \sum_{k=1}^{r_1} q_{ij,k}^1 q_{ij,k}^2 + \sum_{m=1}^{r_2} s_{ij,m}^1 s_{ij,m}^2 \Big),\end{aligned}$$

where $P_\delta = \{p_{ij}^\delta\}$, $Q_{k,\delta} = \{q_{ij,k}^\delta\}$, $S_{m,\delta} = \{s_{ij,m}^\delta\}$, $i, j = \overline{1, n}$, $k = \overline{1, r_1}$, $m = \overline{1, r_2}$, $\delta = 1, 2$.

Definition 4. *The generalized gradient of the convex function* $\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$ *at the interior point* $[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ *is the tuple of matrices* $[D_0, E_{1,0}, \ldots, E_{r_1,0}, F_{1,0}, \ldots, F_{r_2,0}] \in \prod_{i=1}^{1+r_1+r_2} \mathbb{R}^{n\times n}$ *such that for all* $(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}) \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ *we have*

$$\begin{aligned}&\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \psi_0[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \\ &\quad \ge \langle [D_0, E_{1,0}, \ldots, E_{r_1,0}, F_{1,0}, \ldots, F_{r_2,0}), (P - P_0, Q_1 - Q_{1,0}, \ldots, Q_{r_1} - Q_{r_1,0}, S_1 - S_{1,0}, \ldots, S_{r_2} - S_{r_2,0}] \rangle.\end{aligned}$$

Let $\Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$ be a linear matrix-valued operator that maps the tuple of matrices $[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ into the $n(1 + r_1 + r_2) \times n(1 + r_1 + r_2)$ symmetric matrices Γ.

Denote by $\Delta_{ij} \in \mathbb{R}^{n\times n}$ the matrix in which the entries at positions (i, j) and (j, i) are units, and all the rest are zeroes.

Lemma 3. *The generalized gradient of the function* $\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] = -\lambda_{\min}(\Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}])$ *at the interior point* $[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ *is the tuple of matrices* $[D_0, E_{1,0}, \ldots, E_{r_1,0}, F_{1,0}, \ldots, F_{r_2,0}]$, *where* $D_0 = \{d_{ij}^0\}$, $E_{k,0} = \{e_{ij,k}^0\}$, $F_{m,0} = \{f_{ij,m}^0\}$, $i, j = \overline{1,n}$, $k = \overline{1,r_1}$, $m = \overline{1,r_2}$ *such that*

$$
\begin{aligned}
d_{ij}^0 &= \begin{cases} -z_0^\top \Gamma(\Delta_{ij}, \Theta_1, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i = j, \\ -\frac{1}{2} z_0^\top \Gamma(\Delta_{ij}, \Theta_1, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i \neq j, \end{cases} \\
e_{ij,k}^0 &= \begin{cases} -z_0^\top \Gamma(\Theta, \Theta_1, \ldots, \Theta_{k-1}, \Delta_{ij}, \Theta_{k+1}, \ldots, \Theta_{r_1}, \Theta_{r_1+1}, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i = j, \\ -\frac{1}{2} z_0^\top \Gamma(\Theta, \Theta_1, \ldots, \Theta_{k-1}, \Delta_{ij}, \Theta_{k+1}, \ldots, \Theta_{r_1}, \Theta_{r_1+1}, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i \neq j, \end{cases} \\
f_{ij,m}^0 &= \begin{cases} -z_0^\top \Gamma(\Theta, \Theta_1, \ldots, \Theta_{r_1}, \Theta_{r_1+1}, \ldots, \Theta_{m-1}, \Delta_{ij}, \Theta_{m+1}, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i = j, \\ -\frac{1}{2} z_0^\top \Gamma(\Theta, \Theta_1, \ldots, \Theta_{r_1}, \Theta_{r_1+1}, \ldots, \Theta_{m-1}, \Delta_{ij}, \Theta_{m+1}, \ldots, \Theta_{r_1+r_2}) z_0, & \text{if } i \neq j, \end{cases}
\end{aligned} \tag{11}
$$

where Θ_v *is matrix* Θ *at position* v,
z_0 *is the unit eigenvector corresponding to* $\lambda_{\min}(\Gamma(P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}))$.

Proof. Consider for $z \in \mathbb{R}^{n(1+r_1+r_2)}$

$$
\begin{aligned}
&\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \psi_0[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \\
&= -\min_{\|z\|=1} \{z^\top \Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] z\} + \min_{\|z\|=1} \{z^\top \Gamma[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] z\}.
\end{aligned}
$$

Assume the first quadratic form reaches its minimal value at $z = z_1$, and the second one at $z = z_0$. Then subtracting and adding the expression $z_0^\top \Gamma(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}) z_0$, we get

$$
\begin{aligned}
&\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \psi_0[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \\
&= -z_0^\top \Big(\Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \Gamma[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}]\Big) z_0 \\
&+ z_0^\top \Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] z_0 - z_1^\top \Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] z_1.
\end{aligned}
$$

Since $z_0^\top \Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] z_0 \geq z_1^\top \Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] z_1$, we have

$$
\begin{aligned}
&\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \psi_0[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \\
&\geq -z_0^\top \Big(\Gamma[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] - \Gamma[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}]\Big) z_0.
\end{aligned}
$$

Finally we use in the last inequality the representation of the matrices in the form

$$
P = \sum_{1 \leq i \leq j \leq n} p_{ij} \Delta_{ij}, \quad Q_k = \sum_{1 \leq i \leq j \leq n} q_{ij,k} \Delta_{ij}, \quad S_m = \sum_{1 \leq i \leq j \leq n} s_{ij,m} \Delta_{ij},
$$

and the linearity of Γ with respect to $(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2})$, which allows us to get the presentation of the generalized gradient in the form (11). □

When solving (10), we pass from a constrained problem to an unconstrained one. We define the penalty functions

$$
\psi_1(B) = \lambda_{\max}(B) - 1, \quad \psi_2(B) = -\lambda_{\min}(B), \quad B \in \Omega_n,
$$

and the corresponding Lagrange function

$$\mathcal{L}(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}, u) := \psi_0(P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}) + \sum_{\delta=1,2} \Big(u_{P,\delta}\psi_\delta(P) + \sum_{k=1}^{r_1} u_{Q_k,\delta}\psi_\delta(Q_k) + \sum_{m=1}^{r_2} u_{S_m,\delta}\psi_\delta(S_m) \Big), \tag{12}$$

where $u = \{u_{P,1}, u_{Q_1,1}, \ldots, u_{Q_{r_1},1}, u_{S_1,1}, \ldots, u_{S_{r_2},1}, u_{P,2}, u_{Q_1,2}, \ldots, u_{Q_{r_1},2}, u_{S_1,2}, \ldots, u_{S_{r_2},2}\} \in \mathbb{R}^{2(1+r_1+r_2)}$ are non-negative Lagrange multipliers.

Theorem 2. *Provided that the condition H1 holds, the function $\psi_0[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$ attains its minimum at the point $[P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}] \in \prod_{i=1}^{1+r_1+r_2} \bar{\Omega}_n^1$ if and only if the point $(P_0, Q_{1,0}, \ldots, Q_{r_1,0}, S_{1,0}, \ldots, S_{r_2,0}, u^0)$, where*

$$u^0 = (u^0_{P,1}, u^0_{Q_1,1}, \ldots, u^0_{Q_{r_1},1}, u^0_{S_1,1}, \ldots, u^0_{S_{r_2},1}, u^0_{P,2}, u^0_{Q_1,2}, \ldots, u^0_{Q_{r_1},2}, u^0_{S_1,2}, \ldots, u^0_{S_{r_2},2})^\top,$$

is a saddle point of the Lagrange function (12).

Proof. The objective function $\psi_0(\cdot)$ and the constraint functions $\psi_1(\cdot)$, $\psi_2(\cdot)$ are convex for the matrices $P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2} \in \Omega_n$. It follows from the convexity of maximum eigenvalue and the concavity of minimum eigenvalue of a symmetric matrix (see Example 3.10 in [18]). When proving the convexity of the function ψ_0 with respect to $[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}] \in \prod_{i=1}^{1+r_1+r_2} \Omega_n$ we also use the linear dependence of the operator Γ on $[P, Q_1, \ldots, Q_{r_1}, S_1, \ldots, S_{r_2}]$.

Due to the Karush–Kuhn–Tucker conditions for convex problems it is left to show that Slater conditions hold [18] (page 244). By virtue the assumption H1 there exists a tuple of matrices $[\bar{P}, \bar{Q}_1, \ldots, \bar{Q}_{r_1}, \bar{S}_1, \ldots, \bar{S}_{r_2}] \in \prod_{i=1}^{1+r_1+r_2} \Omega_n \cap \mathrm{dom}(\psi_0)$ such that

$$\psi_\delta(\bar{P}) < 0, \psi_\delta(\bar{Q}_1) < 0, \ldots, \psi_\delta(\bar{Q}_{r_1}) < 0, \psi_\delta(\bar{S}_1) < 0, \ldots, \psi_\delta(\bar{S}_{r_2}) < 0, \quad \delta = 1, 2,$$

is satisfied, and the Slater condition is true, which concludes the proof. □

4. Conclusions

The work is devoted to modeling and investigation of the architecture design for the delayed recurrent neural network basing on the delayed differential equations. The usage of discrete and distributed delays makes it possible to model the calculation of the next states using internal memory, which corresponds to the artificial recurrent neural network architecture used in the field of deep learning.

The paper proposes a method for constructing an exponential estimate of the solution of a model of a recurrent neural network using the Lyapunov–Krasovskii functional. This estimate is reduced to solving the corresponding linear matrix inequality. This is the most costly operation in terms of the computational complexity, however, this is where we can apply the optimization approach to find the optimal set of matrices from the viewpoint of the exponential estimate.

In contrast to the indirect method of constructing exponential estimates, which was proposed in previous works [12,13], in this study, the method based on the Lyapunov–Krasovskii functional allows the optimization of the estimate within the compact domain of positive definite matrices.

To this end, the concept of a generalized gradient of a convex function on a set of positive definite matrices is introduced. The constructive form of the generalized gradient for the minimal eigenvalue of the matrix Γ is presented. The Lagrange function for the unconditional optimization problem is constructed. In this case, the search for the optimal exponential estimate is reduced to finding the saddle point of the Lagrange function.

Author Contributions: Conceptualization, V.M. and M.K.; methodology, S.R.; software, V.M.; validation, V.M., S.R. and M.K.; formal analysis, V.M.; investigation, S.R.; resources, M.K.; data curation, V.M.; writing—original draft preparation, S.R.; writing—review and editing, M.K.; visualization, V.M.; supervision, S.R.; project administration, M.K.; funding acquisition, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the University of Bielsko-Biala.

Acknowledgments: The work is supported by the University of Bielsko-Biala, and the program number is K18/1b/UPBJ/2019-2020.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LSTM	Long Short-Term Memory
GRU	Gated Recurrent Units
RNN	Recurrent Neural Network
DDE	Delayed Differential equarion

References

1. Habiba, M.; Pearlmutter, B.A. Neural Ordinary Differential Equation based Recurrent Neural Network Model. *arXiv* **2020**, arXiv:2005.09807.
2. Sherstinsky, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]
3. Fang, S.; Jiang, M.; Wang, X. Exponential convergence estimates for neural networks with discrete and distributed delays. *Nonlinear Anal. Real World Appl.* **2009**, 702–714. [CrossRef]
4. Hale, J.K.; Lunel, S.M.V. *Introduction to Functional Differential Equations*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 99.
5. Zhang, H.; Wang, Z.; Liu, D. A Comprehensive Review of Stability Analysis of Continuous-Time Recurrent Neural Networks. *IEEE Trans. Neural Networks Learn. Syst.* **2014**, *25*, 1229–1262. [CrossRef]
6. Hopfield, J.J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 3088–3092. [CrossRef] [PubMed]
7. Wei, J.; Ruan, S. Stability and bifurcation in a neural network model with two delays. *Phys. D Nonlinear Phenom.* **1999**, *130*, 255–272. [CrossRef]
8. Ruan, S.; Wei, J. On the zeros of a third degree exponential polynomial with applications to a delayed model for the control of testosterone secretion. *Math. Med. Biol.* **2001**, *18*, 41–52. [CrossRef]
9. Ruan, S.; Wei, J. On the zeros of transcendental functions with applications to stability of delay differential equations with two delays. *Dyn. Contin. Discret. Impuls. Syst. Ser. A* **2003**, *10*, 863–874.
10. Yan, X.P.; Li, W.T. Stability and bifurcation in a simplified four-neuron BAM neural network with multiple delays. *Discret. Dyn. Nat. Soc.* **2006**, *2006*. [CrossRef]
11. Huang, C.; Huang, L.; Feng, J.; Nai, M.; He, Y. Hopf bifurcation analysis for a two-neuron network with four delays. *Chaos Solitons Fractals* **2007**, 795–812. [CrossRef]
12. Martsenyuk, V. On an indirect method of exponential estimation for a neural network model with discretely distributed delays. *Electron. J. Qual. Theory Differ. Equ.* **2017**, *2017*, 1–16. [CrossRef]
13. Martsenyuk, V. Indirect method of exponential convergence estimation for neural network with discrete and distributed delays. *Electron. J. Differ. Equ.* **2017**, *2017*, 1–16.
14. Gu, K. An integral inequality in the stability problem of time-delay systems. In Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187), Sydney, Australia, 12–15 December 2000. [CrossRef]
15. Sanchez, E.; Perez, J. Input-to-state stability (ISS) analysis for dynamic neural networks. *IEEE Trans. Circuits Syst. Fundam. Theory Appl.* **1999**, *46*, 1395–1398. [CrossRef]
16. Zhang, F. (Ed.) *The Schur Complement and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2005. [CrossRef]

17. Khusainov, D.Y.; Martsenyuk, V.P. Optimization method for stability analysis of delayed linear systems. *Cybern. Syst. Anal.* **1996**, *32*, 534–538. [CrossRef]
18. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004. [CrossRef]