A DYNAMICAL MODEL OF A COGNITIVE FUNCTION: ACTION SELECTION

Özkan Karabacak, N. Serap Sengör

Faculty of Electrical and Electronics Engineering,
Istanbul Technical University, Istanbul, Turkey

Abstract: A model of cortex-basal ganglia-thalamus-cortex loop is presented. Even though the mentioned loop has been proposed to take part in different cognitive processes, the model exploits only the "action selection" function. The analysis of the model is based on the stability analysis of a non-linear dynamical system. Thus a biologically valid model of a cognitive function is given and its analysis is accomplished using system theory tools.

Keywords: dynamic modelling, difference equations, stability analysis, stability domains, discriminators.

1. INTRODUCTION

Modelling the cognitive processes has at least twofold advantage, inspiration of new engineering concepts and understanding the mechanisms behind the cognitive behaviour. Simple models of cognitive processes may be restrictive in explaining the considered phenomena in whole, but could give clues for the observed dysfunctions.

In this paper, "action selection" function of basal ganglia will be examined from systems level point of view. Basal ganglia are a group of sub-cortical brain structure, which take part not only in motor functions but also in cognitive processes as action gating, action selection, sustaining working memory representations and sequence processing (Prescott, *et al*., 2003; Gurney, *et al.*, 2001; Taylor and Taylor, 2000). While, the proposed model is simple compared to already existing realistic models in the sense of neurobiology (Gurney, *et al.*, 2001; Taylor and Taylor, 2000; Hasan, *et al.,* 2004), it is much more complicated than models accomplishing its "action selection" function by a MAX-NET structure (Kaplan, *et al.,* 2003). The simplicity of the model eased the analysis, thus provided some explanation on the effect of the internal connection of basal ganglia structures and dopamine discharge. By perturbing parameters corresponding to interconnection weights and dopamine discharge, dysfunction of "action selection" property is obtained. Thus one function of a sub-cortical structure is modelled as a non-linear, dynamical, discrete time system and its analysis is carried out utilizing the concept of fixed points, LaSalle's invariance principle, constructing domains of attraction, etc.

In the second section, following a brief explanation of cortex-basal ganglia-thalamus-cortex loop (C-BG-TH-C), the model will be introduced and the stability analysis of the system will be given. The effect of parameters on the behaviour of the model will be considered. In the third section, how the model accomplishes action selection will be explained, the domains of attraction will be illustrated and how they could be interpreted to explain the "action selection" will be discussed. Furthermore, the model can be utilized in discriminating more than two actions and this property is obtained in subsection 3.2. The effect of parameters on "action selection" disabling the selection between competing actions also will be explained. In the fourth section, the dopamine effect will be introduced and its effect on "action selection" will be presented.

2. A MODEL FOR CORTEX-BASAL GANGLIA-THALAMUS-CORTEX LOOP

Basal ganglia (BG), most thoroughly studied neural structure, once was thought to be effective only in motor control but now its role in cognitive processes is more appreciated (Packard and Knowlton, 2002). The dysfunctions of this structure exploit itself especially in brain disorders as Parkinson's disease, Huntington's disease and schizophrenia. Existence of at least five different loops of C-BG-TH-C has been suggested (Alexander and Crutcher, 1990). In each of these loops different substructures of cortex and BG are employed. The principle substructures of BG are proposed to be the striatum (STR), the subthalamic nucleus (STN), the globus pallidus internal and external (GP_i, GP_e), the substantia nigra pars reticulata and compacta (SN_r, SN_c). As different substructures are active for different functions (Gurney, *et al.*, 2001; Taylor and Taylor, 2000), only those exploited in Figure 1 are considered in the proposed model. The main input components of BG, STR and STN, and the main output components, GP_i and SN_r, are all considered in the model. The main effect of BG on thalamus is inhibitory thus in the model the connection from BG to thalamus is negative, whereas the connections from cortex to BG are positive and these are shown as excitatory connections in Figure 1. As system level of modelling is considered the number of neurons for each structure is minimized. To carry out the analysis, first only one neuron for each structure in C-BG-TH-C loop is considered then in the following sections the number of neurons will be multiplied according to function considered.

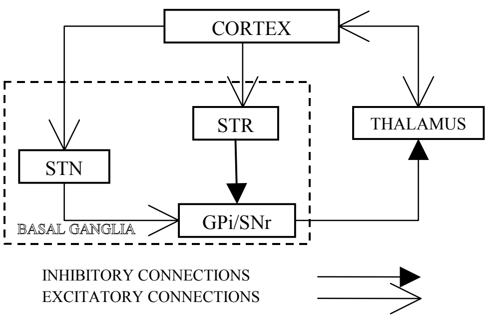


Fig. 1. BG-TH-C Loop

The principal structures of basal ganglia considered in the model are STR, STN, and GP_i/SN_r and they are denoted in equation (1) by $r(k)$, $n(k)$, and $d(k)$, respectively. The other structures, which have connection with basal ganglia are denoted by $m(k)$ and $p(k)$, and they correspond to thalamus and cortex, respectively.

$$\begin{bmatrix} r(k+1) \\ n(k+1) \\ d(k+1) \\ m(k+1) \\ p(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} r(k) \\ n(k) \\ d(k) \\ m(k) \\ p(k) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ -a & b & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} g(r(k)) \\ f(n(k)) \\ f(d(k)) \\ f(m(k)) \\ f(p(k)) \end{bmatrix} \qquad (1)$$

Due to biological structure considered, the parameters *a, b* are positive quantities.

The subsystem given by equation (1) is rewritten in compact form as follows:

$$x(k+1) = \Gamma x(k) + \Lambda F(x(k)) \qquad (\Sigma_1)$$

In system Σ_1, $F : \Re^5 \to \Re^5$ and $\Gamma, \Lambda \in \Re^{5\times 5}$. $f(x)$ is defined as $f(x) \hat{=} \frac{1}{2}(\tanh 2(x-0.6)+1)$ and it is illustrated in Figure 2. The function $g(x)$ is determined as $f(x-\theta)$ in order to model dopamine effect. The value of θ is fixed to 0.3 in the following except section 4, where its effect on "action selection" is considered.

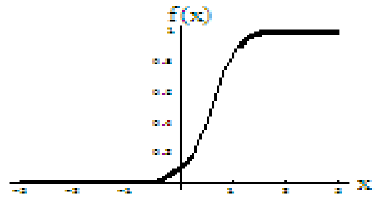


Fig. 2. Activation function of a neuron and its saturation regions (bold)

The subsystem Σ_1 is a discrete-time, non-linear dynamical system. The solution of subsystem Σ_1 is bounded if $\lambda < 1$. The solution of Σ_1 can be written as follows:

$$x(n) = \Gamma^n x(0) + \Gamma^{n-1}\Lambda F(x(0)) + \Gamma^{n-2}\Lambda F(x(1)) + \ldots + \Gamma\Lambda F(x(n-2)) + \Lambda F(x(n-1)) \qquad (2)$$

$$\|x(n)\| \le \|\Gamma^n\|\|x(0)\| + \|\Gamma^{n-1}\|\|\Lambda F(x(0))\| + \|\Gamma^{n-2}\|\|\Lambda F(x(1))\| + \ldots + \|\Gamma\|\|\Lambda F(x(n-2))\| + \|\Lambda F(x(n-1))\| \qquad (3)$$

Let $\xi \hat{=} \max_{x\in\Re^5}\{\|\Lambda F(x)\|\}$

Such a ξ exists because $F(.)$ is bounded. Then,

$$\|x(n)\| \le \|\Gamma\|^n \|x(0)\| + \left(\|\Gamma\|^{n-1} + \|\Gamma\|^{n-2} + \ldots + \|I\|\right)\cdot \xi \qquad (4)$$

As long as $|\lambda| < 1$, the second term on the right hand side of equation (4) will converge to $\frac{1}{1-\lambda}\xi$ as $n \to \infty$. Thus the solution of Σ_1 is bounded. In the following, during simulations λ is taken as *0.5*.

Following LaSalle's invariance theorem (LaSalle, 1986) it can be shown that the solutions of Σ_1 converge to an invariant set.

Proposition 1: The solutions of Σ_1 converge to an invariant set M when $|\lambda| < 1$.

Proof: $V(x) \hat{=} x^T P^T P x$, where P is the annihilator of Γ and Λ, i.e., $P\Gamma = P\Lambda = 0$.

$$V(x(k+1)) - V(x(k)) = x(k+1)^T P^T P x(k+1) - x(k)^T P^T P x(k), \qquad (5)$$

$$= \left(\Gamma x + \Lambda F(x)\right)^T P^T P\left(\Gamma x + \Lambda F(x)\right) - x^T P^T P x \qquad (6)$$

Since $P\Gamma$ and $P\Lambda$ are zero matrices,

$$V(x(k+1)) - V(x(k)) = -x^T P^T P x \le 0, \quad \forall x \in \Re^5 .$$

As $V(x)$ is bounded since $|\lambda| < 1$ and $V(x(k+1)) - V(x(k)) \le 0$, $\forall x \in \Re^5$ the solutions of Σ_1 converge to an invariant set due to LaSalle's invariance theorem. □

From LaSalle's invariance theorem (LaSalle 1986), the invariant set is contained in the set $E = \{x \in \Re^5 | V(x(k+1)) - V(x(k)) = 0\}$. Now, there is need to show what the invariant set is composed of.

A fixed point x^* of subsystem Σ_1 satisfies $x^* = \Gamma x^* + \Lambda F(x^*)$. The set E contains the fixed points since the difference $V(x(k+1)) - V(x(k))$ is equal to zero when $x(k+1) - x(k) = \Gamma x + \Lambda F(x) - x = 0$ and by definition fixed point satisfies this equality. But set E contains more elements since there are other possible solutions of $V(x(k+1)) - V(x(k)) = 0$. In the model fixed points will correspond to action selection, so it is important to determine the regions where the fixed points are present.

To find out where the fixed points of subsystem Σ_1 are, one approach is to find the region where the right hand side of equation (1) is a contraction mapping (Vidyasagar, 1993). The following proposition states that the right hand side of equation (1) is a local contraction mapping. Thus a way of figuring out the regions where only fixed points exist will be given.

Proposition 2: The mapping $T(.) : \Re^5 \to \Re^5$, where $T(x) \hat{=} \Gamma x + \Lambda F(x)$ is a contraction mapping in a region R, if $\|\Gamma\| + \alpha\|\Lambda\| < 1$.

Proof: The mapping $T(.)$ is a contraction if $\|T(x) - T(y)\| \le \gamma\|x - y\|$, $\forall x, y \in R$ and $0 < \gamma < 1$.

$$\|\Gamma x + \Lambda F(x) - \Gamma y - \Lambda F(y)\| \le \|\Gamma(x - y)\| + \|\Lambda(F(x) - F(y))\|$$

As $F(x)$ is a continuous operator, from mean value theorem the inequality $\|F(x) - F(y)\| \le \alpha\|x - y\|$ holds, where $\alpha \hat{=} \max_{x \in \Re}\left(\left\|\frac{df(x)}{dx}\right\|, \left\|\frac{dg(x)}{dx}\right\|\right)$. Thus, $\|T(x) - T(y)\| \le (\|\Gamma\| + \alpha\|\Lambda\|)\|x - y\|$ in a region R where $\|\Gamma\| + \alpha\|\Lambda\| < 1$. □

The region R will be in the saturation regions of $F(x)$, since $\|\Gamma\| = \lambda$, $\|\Lambda\| = \max(1.8478, \sqrt{a^2 + b^2})$. One case is when $\lambda = 0$ and $\|\Lambda\| = 1.8478$. The value of α has to be less than $\|\Lambda\|^{-1}$, i.e., $\alpha < 0.5412$ in this case and the region R corresponds to subsets of $\Re^5$, where components of x are either less than 0.188 or greater than 1.311.

The fifth component of the state vector of system Σ_1 corresponds to cortex neuron and this component is observed since any activity occuring will be as a result of activation at cortex. If this component, i.e., p is nearly zero, while the system is converging to a fixed point, the loop characterized by subsystem Σ_1 is regarded as unactivated or inhibited. Thus this fixed point is defined as "passive point". When component corresponding to cortex neuron converges to a fixed point with a high value, the loop is regarded as being active and this fixed point is called as "active point". Even though the behaviour of subsystem Σ_1 is classified according to the value of fifth component of state vector, the other components also take fixed values. Thus the fixed points of the subsystem Σ_1 are classified according to the activation of cortex. From the simulation results, it is observed that Σ_1 has at least one of the two type fixed points, which stay almost near the same points even though parameter values a, b change. The first one is in the neighbourhood of the point $[0.1 \quad 0.1 \quad 0.2 \quad -0.1 \quad 0.1]^T$ and the second one is in the neighbourhood of the point $[0.9 \quad 0.9 \quad 0.2 \quad 1.8 \quad 1.8]^T$. These points are regarded as passive and active points, respectively. Not all components satisfy the constraint given by proposition 1, but still they are fixed points. This is acceptable, since proposition 1 only gives restrictive sufficient condition. The simulation results also reveal that for different *a, b* values the subsystem converges to different fixed points. For some *a, b* values the subsystem Σ_1 converges either to a passive point or to an active point according to the initial conditions as exploited in Figure 3. The region S_p in Figure 3 includes the parameter values for which Σ_1 converges to a passive point for any initial condition. For the parameter values in region S_a the fixed point

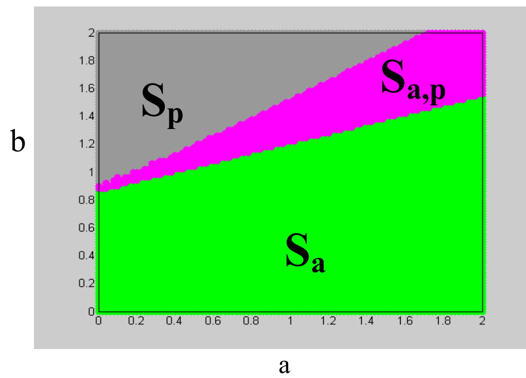


Fig. 3. Dependence of solutions on parameters

of Σ_1 is always an active point and the region $S_{a,p}$ includes the parameter values for which Σ_1 has both active and passive points. That is the state converges to active points for some initial conditions while it converges to passive points for other initial conditions. These regions are approximately determined as follows:

$S_a: \ b < 0.34 \cdot a + 0.87$ (7)

$S_p: \ b > 0.65 \cdot a + 0.9$ (8)

$S_{a,p}: \ 0.34 \cdot a + 0.87 < b < 0.65 \cdot a + 0.9$ (9)

3. THE PROPOSED MODEL FOR ACTION SELECTION

The subsystem Σ_1, which is composed of five neurons, each corresponding to a substructure in BG-TH-C loop is revealed to be either active or passive. Competition between actions can be generated considering more than one such subsystem. In this case the subsystems are connected crosswise by the excitatory connections from the STN neuron of one loop to GPi/SNr neuron of the other loop. Such a model for two loops is illustrated in Figure 4.

In the section 3.1 two competing actions are considered because of the simplicity of analyzing the whole system and understanding the underlying mechanisms. In the subsection 3.2 more than two loops are combined to exploit selection between more actions.

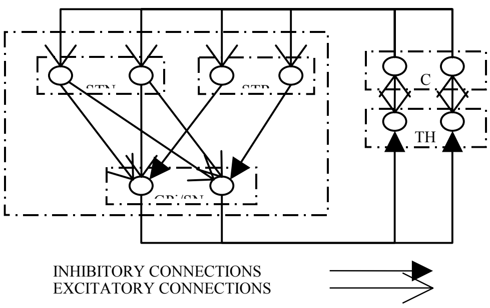


Fig. 4. Connected C- BG-TH-C loops

3.1 Action Selection Between Two Actions

To model the "action selection" between two competing actions, two of subsystems Σ_1 are connected as in equation (10):

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} \Gamma & 0 \\ 0 & \Gamma \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} \Lambda & \Pi \\ \Pi & \Lambda \end{bmatrix} \begin{bmatrix} F(x_1(k)) \\ F(x_2(k)) \end{bmatrix} \quad (10)$$

Here, $\Pi \in \Re^{5\times5}$ and its elements are zero except the one on third row second column, which is denoted by *c*. This parameter *c* is the weight of the binding connection between loops.

A compact form of this system is stated as follows:

$$x_\Pi(k+1) = \Gamma_\Pi x_\Pi(k) + \Lambda_\Pi F(x_\Pi(k)) \qquad (\Sigma_2)$$

As the system Σ_2 is autonomous, the only way to present an input is by means of initial conditions. So the actions to be selected are introduced to the system as initial conditions. Only the components of initial conditions corresponding to cortex, i.e., p, are different than zero. Selecting initial condition this way is also in agreement with what Hirsch stated, "The initial values of the non-input units are generally reset to the same conventional value (usually zero) each time the net is run" (Hirsh, 1989).

Similar to Σ_1, it is possible to show that the solutions are bounded, the system is stable in the sense of LaSalle and there exists fixed points of the system Σ_2.

Proposition 3: The solutions of the system Σ_2 converge to an invariant set M_Π when $|\lambda| < 1$.

The Liapunov function for the system Σ_2 is given by equation (11)

$$V_\Pi(x) \triangleq x^T P_\Pi^T P_\Pi x \qquad (11)$$

where, $P_\Pi \triangleq \begin{bmatrix} P & P \\ P & P \end{bmatrix}$.

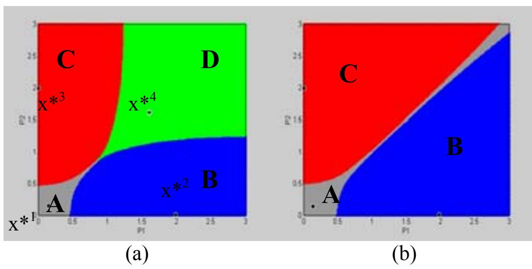
Proposition 4: The mapping $T_\Pi(.): \Re^{10} \to \Re^{10}$, where $T_\Pi(x) \triangleq \Gamma_\Pi x + \Lambda_\Pi F(x)$ is a contraction mapping in a region R, if $\|\Gamma_\Pi\| + \alpha \|\Lambda_\Pi\| < 1$.

As $\|\Lambda_\Pi\| = \max(1.8478, \sqrt{a^2 + b^2 + c^2 + 2bc})$, again the fixed points are in the saturation regions.

Combining two loops the maximum number of the fixed points increases from 2 to 4 as exploited in Figure 5a. These fixed points are denoted by x^{*1}, x^{*2}, x^{*3} and x^{*4}, and each correspond to a different behaviour of the system as follows:

x^{*1} : both of the subsystems are passive
x^{*2} : only the first subsystem is active
x^{*3} : only the second subsystem is active
x^{*4} : both of the subsystems are active

The domains of attraction of these fixed points are



(a) (b)

Fig. 5. (a) Domains of attraction for a=1.5, b=1, c=0.8
(b) Domains of attraction for a=1.5, b=1, c=0.9

named A for x^{*1}, B for x^{*2}, C for x^{*3} and D for x^{*4} and for parameter value a=1.5, b=1, c=0.8 they are illustrated in Figure 5a. For different initial conditions the system Σ_2 converges to different fixed points. For example, if the initial condition is $p_1 = 0.5$, $p_2 = 1$, it converges to x^{*3}. Thus the second action is selected. If the initial state of the system is in region A or D, the system cannot discriminate actions. In the first case none of the actions and in the second case both actions are generated. If there exist all of these regions A, B, C and D like in Figure 5a, the system cannot be considered as a good discriminator.

However existence of the region D gives an idea about how to construct a soft discriminator, which can select more than one action if necessary. This would correspond to case when actions to be selected have great values of initial conditions that are close.

For some proper values of *a, b* and *c* the system behaves like a strict discriminating network because the region D vanishes. This case is shown in Figure 5b.

To understand how the region D vanishes, consider the solutions of Σ_2 with the initial conditions $p_1(0)=p_2(0)$, which are included either in A or D. Because all the other neurons are initially reset to zero, $p_1(0)=p_2(0)$ implies $x_1(0)=x_2(0)$. In this case the system Σ_2 behaves like the system $\Sigma_2^{'}$ below:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} \Gamma & 0 \\ 0 & \Gamma \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} \Lambda+\Pi & 0 \\ 0 & \Lambda+\Pi \end{bmatrix} \begin{bmatrix} F(x_1(k)) \\ F(x_2(k)) \end{bmatrix} \qquad (\Sigma_2^{'})$$

The system $\Sigma_2^{'}$ consists of two disconnected subsystems $\Sigma_1^{'}$:

$$x(k+1) = \Gamma x(k) + \Lambda' F(x(k)) \qquad (\Sigma_1^{'})$$

where $\Lambda' \hat{=} \Lambda+\Pi$. It is enough to analyze $\Sigma_1^{'}$ in order to understand how the system Σ_2 behaves for the initial values $x_1(0)=x_2(0)$. To analyze the system $\Sigma_1^{'}$ is easy now, because the subsystem Σ_1 has been already analyzed and the systems Σ_1 and $\Sigma_1^{'}$ differs only in one parameter which is on third row second column of Λ and Λ'. In Λ it is *b* and in Λ' it is *b*+*c*. All the other components of Λ and Λ' are same. Thus, the inequalities (7-9) can be used for system $\Sigma_1^{'}$ by substituting *b*+*c* for *b*:

$$S_a{}' : b+c < 0.34\cdot a + 0.87 \qquad (12)$$

$$S_p{}' : b+c > 0.65\cdot a + 0.9 \qquad (13)$$

$$\mathrm{S_{a,p}}' : 0.34\cdot \mathrm{a} + 0.87 < b+c < 0.65\cdot a + 0.9 \qquad (14)$$

The system $\Sigma_1^{'}$ converges to active point, passive point and either to passive or to active point, for cases given by Equations (12), (13), (14), respectively.

If the inequality (12) is satisfied, $\Sigma_1^{'}$ converges to an active point for all initial conditions. Thus $\Sigma_2^{'}$ converges to such a point where x_1 and x_2 are both active. In this case both actions are selected. This means all the points with $x_1(0)=x_2(0)$ are included in the region D, so the region A does not exist and the stable fixed point that corresponds to the region where no selection is possible is x*[4]. Similarly, if the parameters satisfy the inequality (13) the region D does not exist as for a=1.5, b=1, c=0.9 and this is shown in Figure 5b. If they satisfy the inequality (14) both regions A and D exist as for a=1.5, b=1, c=0.8 which is shown in Figure 5a.

3.2 Action Selection Between More Than Two Actions

To model the "action selection" between n competing actions, n subsystems Σ_1 are connected:

$$\begin{bmatrix} x_1(k+1) \\ \vdots \\ x_\ell(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} \Gamma & 0 & 0 & \cdots & 0 \\ 0 & \Gamma & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \Gamma \end{bmatrix} \begin{bmatrix} x_1(k) \\ \vdots \\ x_\ell(k) \\ \vdots \\ x_n(k) \end{bmatrix} + \begin{bmatrix} \Lambda & \Pi & \Pi & \cdots & \Pi \\ \Pi & \Lambda & \ddots & \vdots & \vdots \\ \Pi & \ddots & \ddots & \ddots & \Pi \\ \vdots & \ddots & \ddots & \ddots & \Pi \\ \Pi & \cdots & \Pi & \Pi & \Lambda \end{bmatrix} \begin{bmatrix} F(x_1(k)) \\ \vdots \\ F(x_\ell(k)) \\ \vdots \\ F(x_n(k)) \end{bmatrix} \qquad (\Sigma_n)$$

In system Σ_n, without losing generality, the first ℓ subsystems correspond to selected actions, so Σ_n is supposed to select ℓ actions from n actions. To fulfil this aim Σ_n should have stable fixed points x* where ℓ winning subsystems will have active points, while the losing ones will have passive points. Thus the stable fixed points x* are composed of active and passive fixed points of Σ_1. If there exist such stable fixed points, Σ_n will necessarily converge to these points for some initial values which have ℓ number of components p with same values and others zero. From simulation results, it is observed that the winning ℓ subsystems p components corresponding to cortex have great values while others are nearly zero. Thus the behavior of winning subsystems can be analysed similar to $\Sigma_2^{'}$. Only these ℓ winning subsystems are considered and they are expressed as follows:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_\ell(k+1) \end{bmatrix} = \begin{bmatrix} \Gamma & 0 & \dots & 0 \\ 0 & \Gamma & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \Gamma \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_\ell(k) \end{bmatrix} + \begin{bmatrix} \Lambda+(\ell-1)\cdot\Pi & 0 & \dots & 0 \\ 0 & \Lambda+(\ell-1)\cdot\Pi & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Lambda+(\ell-1)\cdot\Pi \end{bmatrix} \begin{bmatrix} F(x_1(k)) \\ F(x_2(k)) \\ \vdots \\ F(x_\ell(k)) \end{bmatrix} \qquad (\Sigma_\ell^{'})$$

As the effect of $(n-\ell)$ losing subsystems on these negligible, they are not considered.

The system $\Sigma_\ell^{'}$ consists of ℓ disconnected subsystems $\Sigma_1^{'}$ where $\Lambda' \hat{=} \Lambda+(\ell-1)\cdot\Pi$. Thus inequalities (7-9) can be used again but in this case substituting $b+(\ell-1)c$ for b:

$$S_{a\,\ell}^{'} : b+(\ell-1)\cdot c < 0.34\cdot a + 0.87 \qquad (15)$$

$$S_{p\,\ell}^{'} : b+(\ell-1)\cdot c > 0.65\cdot a + 0.9 \qquad (16)$$

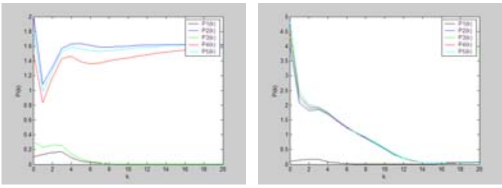
$$S_{a,p\,\ell}^{'} : 0.34\cdot \mathrm{a} + 0.87 < b+(\ell-1)\cdot c < 0.65\cdot a + 0.9 \qquad (17)$$

Due to the approximation done above, these inequalities hold approximately. As an example five competing actions are considered and the expectation is that, for different initial conditions, up to three actions would be selected. Then the discriminator is Σ_5 and the parameters should be chosen to satisfy the inequalities below where ℓ is taken 4 and 3, in inequalities (18) and (19):

$$b+(4-1)\cdot c>0.65\cdot a+0.9 \tag{18}$$

$$0.34\cdot \mathrm{a}+0.87<b+(3-1)\cdot c<0.65\cdot a+0.9 \tag{19}$$

The parameter values *a*=1.5, *b*=1, *c*=0.35 satisfy the above inequalities. For different initial conditions, Σ_5 selects one, two, three or none of five competing actions. The case corresponding to selecting three actions is illustrated in Figure 6a. Because *a*, *b* and *c* satisfy the inequality (18), Σ_5 never selects four of five actions, even if for four of five actions great values of initial conditions are taken (Figure 6b).



(a) (b)

Fig. 6. Solutions of Σ_5 for

(a): $p_1(0)=0.1$, $p_2(0)=2$, $p_3(0)=0.3$, $p_4(0)=1.5$, $p_5(0)=1.8$
(b): $p_1(0)=0.1$, $p_2(0)=4$, $p_3(0)=4.3$, $p_4(0)=4.5$, $p_5(0)=4.8$

4. THE EFFECT OF DOPAMINE ON ACTION SELECTION

In the preceding sections, the effect of parameters *a*, *b*, *c* which correspond to interconnection weights in C-BG-TH-C loop, on "action selection" has been investigated. In this section, the similar effect of θ on action selection will be illustrated just by figures. From neurobiogical point of view θ parameter corresponds to dopamine discharge. Change in this parameter effects the domains of attraction like the parameters *a*, *b*, *c*. Thus in Figure 7, the figure related with $\theta=0.3$ corresponds to a good discriminator. Whereas the figures of $\theta=0.2$ and $\theta=0.8$ exploit that for these parameter values the system is not a good discriminator as there are either two regions where discrimination is not possible or no region where discrimination is possible.

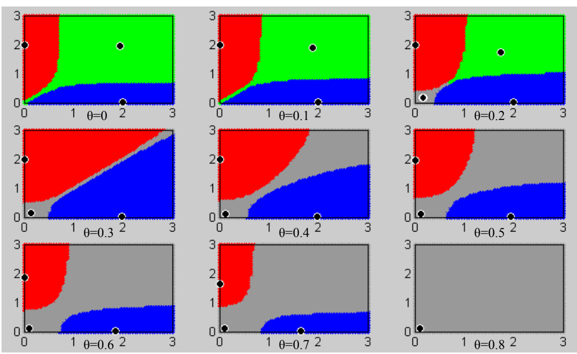


Fig. 7. Domains of attraction for different values of θ

5. CONCLUSION

A non-linear discrete time system is proposed as a model of C-BG-TH-C loop. The proposed model is not only capable in explaining the "action selection" function of C-BG-TH-C loop but also exploits dysfunction of "action selection" as interconnection weights and dopamine discharge changes. In order to show that the non-linear, discrete time system has stable fixed points proposition 1 is given which is based on LaSalle's invariance principle. To find out the place of fixed points, proposition 2, which is based on Banach fixed-point theorem, is stated. The fixed points of the non-linear system change as parameters change and how these fixed points can be interpreted for "action selection" function is explained using the attraction domains of the fixed points. As for "action selection", a system is obtained by interconnecting the proposed model and propositions similar to proposition 1 and 2 are given. Parameter values for proper "action selection" are given for two and more competing actions.

REFERENCES

Alexander, G.E., and M.D. Crutcher (1990), Functional architecture of basal ganglia circuits: neural substrates of parallel processing, *Trends in Neuroscience,* **13**, pp. 266-271.

Gurney, K., T. J. Prescott and P. Redgrave (2001), A computational model of action selection in the basal ganglia. I. A new functional anatomy, *Biological Cybernetics,* **84,** pp. 401-410.

Hasan, D., Ö. Karabacak and N. S. Sengör (2004), A computational model of reinforcement learning at basal ganglia. In: *Abstract Book, Modelling Mental Processes and Disorders,* (S. Pögün, H. Liljenström (Ed)), pg. 48, Ege University, Izmir.

Hirsch, M.W., (1989) Convergent activation dynamics in continuous time networks, *Neural Networks,* **2**, pp. 331-349.

Kaplan, G. B., N. S. Sengör, H. Gürvit and C. Güzelis (2003), Modelling stroop effect by a connectionist model. In: *Proceedings of ICANN/ICONIP 2003* (O. Kaynak, E. Apaydın, E. Oja, L. Xu (Ed)), pp. 457-460, Bogazici University, Istanbul.

Lasalle, J.P. (1986), *The stability and control of discrete processes,* chapter1, Springer-Verlag, New York.

Packard, M. G., and B. J. Knowlton (2002), Learning and memory function of the basal ganglia. *Annual review of neuroscience,* **25,** pp. 563-593.

Prescott, T. J., K. Gurney and P. Redgrave (2003), Basal Ganglia. In: *The handbook of brain theory and neural networks* (M. A. Arbib (Ed)), pp. 147-151, A Bradford Book, The MIT Press, Cambridge.

Taylor, J.G., and N.R. Taylor (2000), Analysis of recurrent cortico-basal ganglia-thalamic loop for working memory, *Biological Cybernetics*, **82**, pp. 415-432.

Vidyasagar, M. (1993), *Nonlinear Systems Analysis,* chapter2, Prentice Hall, New Jersey.