The identification of the rules processes' regulation during restoration (regeneration) of dynamic homeostasis by methods of Adaptive Dynamic Programming

Valeriia Karieva, Sergey Lvov

V.N.Karazin Kharkiv National University

Statement of the problem of determining the strategy of liver regeneration based on the principles of evolutionary selection

Consider:

- the equation which describes the dynamics of populations of different types of liver cells
  $\bar{x}(t+1) = F(\bar{x}(t), \tau(t), \bar{\lambda}(t))$ (1)

$\bar{x}(t)$ - types of functional liver cells at moment t, $\tau(t)$ – given function of external toxicity (disturbing parameter), $\bar{\lambda}(t)$ – control parameters.

- the equation which describes the change in the functional state of the organism as a function of external toxicity and the functional state of the liver
  $\tilde{\Phi}(t+1) = \psi(\tau(t), \Phi(t))$ (2)

$\Phi(t+1) = \sum_{i=1}^{n} c_i(x_i(t), \tau(t))$ - general index of liver functionality

The cost function: $C = \sum_{t=0}^{T} \Phi(t)$

Criterion for the problem of optimal control

Consider evolutionary processes in a certain synthetic (artificial) situation.

A necessary condition for successful evolutionary selection is not a break of the evolutionary chain, but its result is a strategy for achieving suboptimal functional activity of the organism $\bar{\lambda}(t)$.

The example of a criterion for an optimal control problem that satisfies these requirements:

$\sum_{t=0}^{T}(K - \tilde{\Phi}(t))^2 + A\,\tilde{\Phi}(T') \leq E$ **(3)**

on condition $\sum_{t=T_0'}^{T'} I(\tilde{\Phi}(t)) \geq \mathrm{C}$, where $I\left(\tilde{\Phi}(t)\right) = \begin{cases} 1, (K - \tilde{\Phi}(t))^2 \leq E \\ 0, \quad\quad otherwise \end{cases}$ **(4)**.

T – the moment of the end of the organism' life cycle, $T_0'(T')$ – the moment of the beginning (end) of the reproductive period, K–optimal functional activity of the organism, E – constant of the reachability set of the organism' functional activity, A- weight constant, C- the constant of the maximum functionality during the reproductive period

Numerical experiments on the mathematical model

Single and constant toxic functions show that the above processes are not able to cope with the toxic factors that are accumulated in the body. The process of the body's functional state restoring requires the non-trivial strategy of liver regeneration.

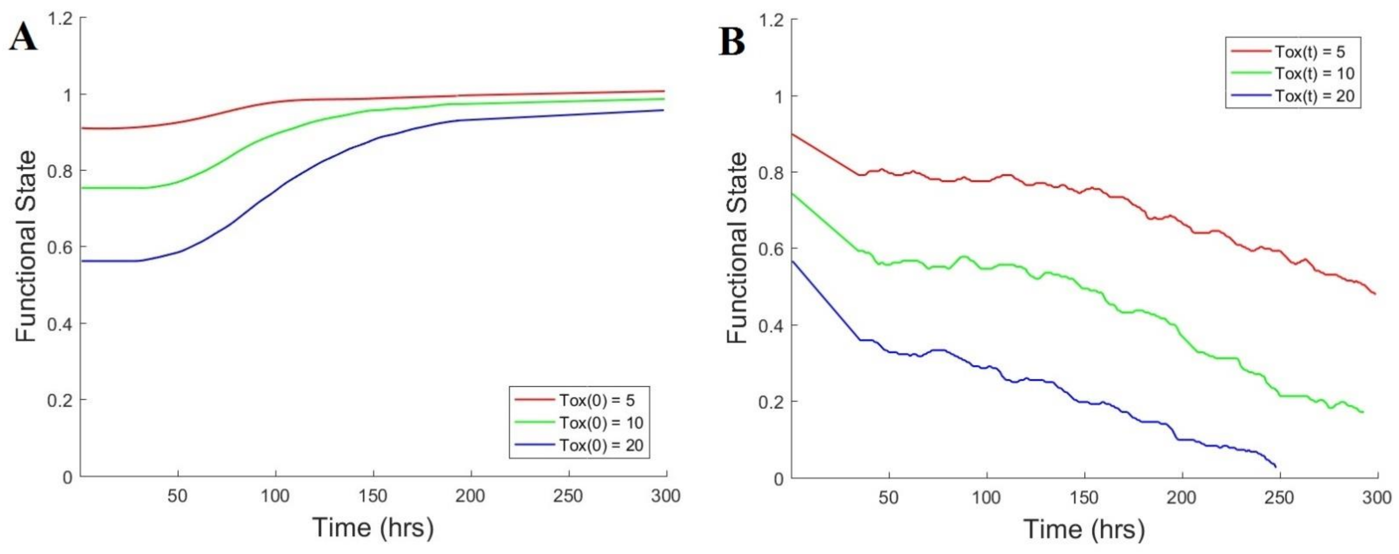

Fig.1. The body's functionality restoring due to the replication process in the case of single (A) and constant (B) toxic functions.

We simulated the liver regeneration following 70 % hepatectomy. We modeled three regenerating modes of response to hepatectomy: delayed, suppressed and enhanced. The obtaining calculations correspond to the biological regeneration process.

The process of the body's functional state restoring requires the non-trivial strategy of liver regeneration.

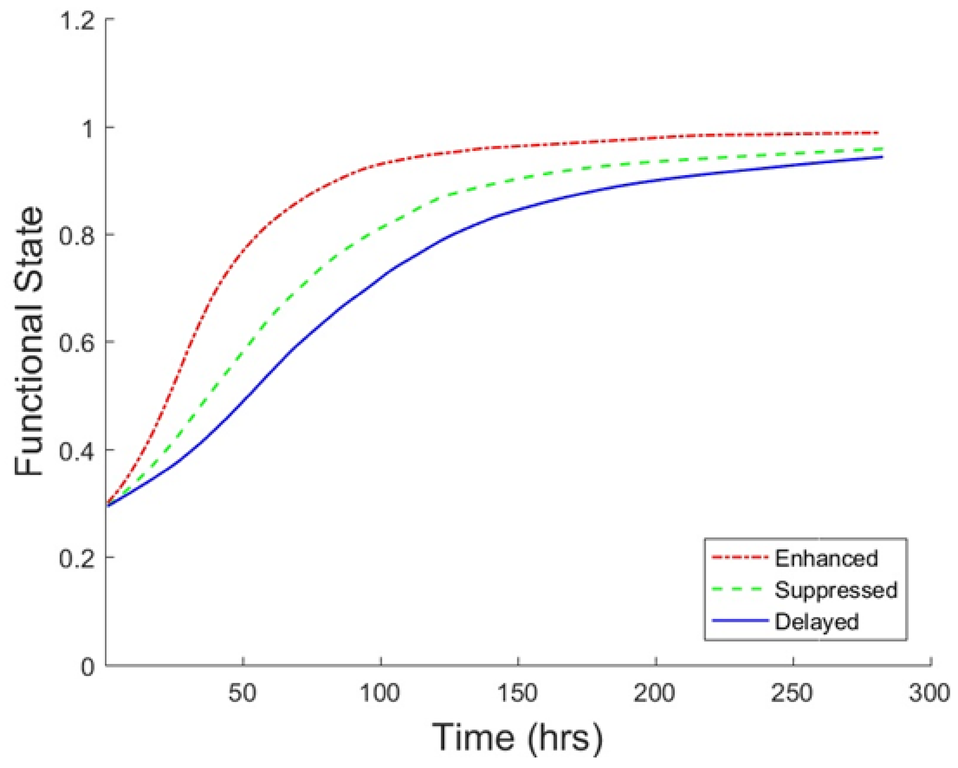

Fig.2. The body's functionality restoring after partial hepatectomy.

Approximation of the model

In theoretical biology the concept of "truth" path (strategy) is attempted to be formulated as a certain condition of extremality. In many cases this concept of extremality can be represented as an additive cost functional along the path which reduces the problem of finding the "true" path to the problem of dynamic programming on a discrete directed structure.

Biological systems are "soft systems" which cannot be solved exactly. We have to solve this problem approximately.

We propose to apply the methods of Reinforcement Learning and a practical implementation method known as Adaptive Dynamic Programming. These give us insight into the design of controllers for biological system that learns and exhibits optimal behavior.

For ease of analysis one often considers a class of discrete-time systems described by nonlinear dynamics in the affine state space difference equation form

$$\bar{x}(t+1) = f\big(\bar{x}(t)\big) + g\big(\bar{x}(t)\big)\bar{\lambda}(t)$$

with state $\bar{x}(t) \in \mathbb{R}^n$ and control input $\bar{\lambda}(t) \in \mathbb{R}^m$. The analysis of such forms is convenient and can be generalized to the general sampled data form $\bar{x}(t+1) = F(\bar{x}(t), \bar{\lambda}(t))$.

A control policy is defined as a function from state space to control space $h: \mathbb{R}^n \to \mathbb{R}^m$. That is, for every state $\bar{x}(t)$, the policy defines a control action $\bar{\lambda}(t) = h(\bar{x}(t))$.

Goal Directed Optimal Performance

The notion of goal-directed optimal behavior is captured by defining a performance measure or cost function

$V_h(\bar{x}(t)) = \sum_{i=t}^{\infty} \gamma^{i-t} r(\bar{x}(i), \bar{\lambda}(i))$

with $0 < \gamma \leq 1$ a discount factor and $\bar{\lambda}(t) = h(\bar{x}(t))$ feedback control policy.

The objective of optimal control theory is to select the policy that minimizes the cost to obtain

$$V^*\big(\bar{x}(t)\big) = \min_{h(.)}\left(\sum_{i=t}^{\infty} \gamma^{i-t} r\,(\bar{x}(i), h(\bar{x}(i))\right)$$

which is known as the optimal cost, or optimal value. Then, the optimal control policy is given by

$$h^*\big(\bar{x}(t)\big) = arg\min_{h(.)}\left(\sum_{i=t}^{\infty} \gamma^{i-t} r\,(\bar{x}(i), h(\bar{x}(i))\right)$$

Value Iteration (VI) Algorithm

However, the problem is to minimize not simply the one-step cost, but the sum of all discounted costs. This problem is generally very difficult or even impossible to solve exactly for general nonlinear systems. Various methods have been developed to simplify the solution of this optimization problem. We consider value iteration (VI) algorithm for solving this optimal problem.

Initialize. Select any control policy $h_0\,(\bar{x}(t))$, not necessarily admissible or stabilizing.

Value Update Step. Update the value using

$$V_{j+1}\big(\bar{x}(t)\big) = r\big(\bar{x}(t), h_j\,(\bar{x}(t))\big) + \gamma V_j\big(\bar{x}(t+1)\big)$$

Policy Improvement Step. Determine an improved policy using

$$h_{j+1}\big(\bar{x}(t)\big) = arg\min_{h()}\Big(r(\bar{x}(t), h\,(\bar{x}(t))) + \gamma V_{j+1}\big(\bar{x}(t+1)\big)\Big)$$

It has been shown that VI converges under certain situations. Note that VI does not require an initial stabilizing policy.

In fact, it is seen that Value Iteration is based on the fact that the Bellman Optimality Equation (BOE) $V^*\big(\bar{x}(t)\big) = \min_{h(.)}\Big(r(\bar{x}(t), h\,(\bar{x}(t))) + \gamma V^*\big(\bar{x}(t+1)\big)\Big)$

is also a fixed point equation. The interleaved steps of value update and policy improvement are the means of iterating the contraction map associated to BOE.

Generally, fixed point equations can be used, with suitable formulation, as a basis for on-line reinforcement learning algorithms that learn by observing data accrued along system trajectories.

References

[1] Karieva V. V., Lvov S. V. Mathematical model of liver regeneration processes: homogeneous approximation.Visnyk of V.N.Karazin Kharkiv National University Ser. "Mathematics, Applied Mathematics and Mechanics" - 2018. - Vol. 87. - pp. 29 - 41.

[2] Karieva V. V., Lvov S. V., Artyukhova L. P. Different strategies in the liver regeneration processes. Numerical experiments on the mathematical model.Visnyk of V.N.Karazin Kharkiv National University Ser. "Mathematics, Applied Mathematics and Mechanics" - 2020. - Vol. 91. - pp. 36 - 44.

[3]Lewis Frank L., Vrabie D. Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control.IEEE Circuits and Systems Magazine. - 2009. - Vol. 9, No.3 - pp. 32 - 50.