

# Multi-Enzyme Pathway Optimisation Through Star-Shaped Reachable Sets

Stanislav Mazurenko<sup>1(✉)</sup>, Jiri Damborsky<sup>1,2</sup>, and Zbynek Prokop<sup>1,2</sup>

<sup>1</sup> Faculty of Science, Research Centre for Toxic Compounds in the Environment  
RECETOX, Loschmidt Laboratories, Masaryk University,  
Kamenice 753/5, 625 00 Brno, Czech Republic  
[stan.mazurenko@gmail.com](mailto:stan.mazurenko@gmail.com)

<sup>2</sup> St. Anne's University Hospital, Pekarska 53, 656 91 Brno, Czech Republic

**Abstract.** This article studies the time evolution of multi-enzyme pathways. The non-linearity of the problem coupled with the infinite dimensionality of the time-dependent input usually results in a rather laborious optimization. Here we discuss how the optimization of the input enzyme concentrations might be efficiently reduced to a calculation of reachable sets. Under some general conditions, the original system has star-shaped reachable sets that can be derived by solving a partial differential equation. This method allows a thorough study and optimization of quite sophisticated enzymatic pathways with non-linear dynamics and possible inhibition. Moreover, optimal control synthesis based on reachable sets can be implemented and was tested on several simulated examples.

**Keywords:** Enzyme kinetics, Optimal control, Synthetic biology, Metabolic networks, Non-linear dynamics

## 1 Introduction

### 1.1 Multi-Enzyme Pathways

In this paper, we consider a set of chemical reactions catalysed by several enzymes. Such reactions take place inside cells and are also used in synthetic biology, e.g. in manufacturing of chemical compounds, biodegradation, medicine, etc. Currently, there are large databases of enzymes based on which pathways can be constructed to turn given substrates into desired products [1]. The enzyme kinetic optimisation of these processes is high on the agenda as it may lead to a substantial economy of time and consumables. Such optimisation may

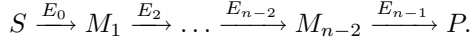
---

This research was supported by the National Sustainability Programme of the Czech Ministry of Education, Youth and Sports (LO1214) and the RECETOX research infrastructure (LM2011028).

**Electronic supplementary material** The online version of this chapter (doi:[10.1007/978-3-319-60816-7\\_2](https://doi.org/10.1007/978-3-319-60816-7_2)) contains supplementary material, which is available to authorized users.

also provide insights into the evolution of cells since some studies suggest that optimal pathways are evolutionarily advantageous and can be predicted based on the genetic information of living cells [2].

We consider an  $n$ -step chemical reaction in which the state variables are the concentrations of metabolites produced and consumed in the course of the reaction:



The control here are the concentrations of enzymes  $E_i$ , the sum of which is limited from above. We will prove that under some general assumptions about the rate equations, one can expect the set of all the possible states of such systems to be star-shaped at any point in time. As a result, an optimisation of the pathway using star-shaped reachable sets [3] can be implemented to obtain the maximum concentration of the final product and the corresponding optimal profile of enzymes.

## 1.2 Mathematical Setup

For a pathway consisting of  $n$  consecutive steps, we will use the following notations:  $e_i$  is the concentration of the enzyme responsible for step  $i$ ;  $x_i$  is the metabolite concentration;  $f_i(x, t)$ ,  $\mathbf{x} = (x_1, \dots, x_n)$ , is the reaction rate per unit of the enzyme concentration  $e_i$ . We assume that  $f_i$  includes all the individual kinetic parameters such as  $k_{cat}$  and  $K_M$  and may depend on the concentrations of all the metabolites involved (e.g., systems with cross-inhibition are included). Moreover, the dependence of all the variables on  $t$  is implied in all the cases below, but we will omit this explicit notation for the sake of simplicity. In practice, all the rates  $f_i$  are non-linear, which significantly complicates any treatment of such systems.

According to enzyme kinetics, the time evolution of a multi-enzyme system over the time  $t \in [0, T]$  can be described as follows:

$$\begin{cases} \dot{x}_1 = e_0 f_0(\mathbf{x}) - e_1 f_1(\mathbf{x}), \\ \dot{x}_2 = e_1 f_1(\mathbf{x}) - e_2 f_2(\mathbf{x}), \\ \dots \\ \dot{x}_{n-1} = e_{n-2} f_{n-2}(\mathbf{x}) - e_{n-1} f_{n-1}(\mathbf{x}), \\ \dot{x}_n = e_{n-1} f_{n-1}(\mathbf{x}). \end{cases} \quad (1)$$

In order to make sure that none of the concentrations becomes negative, we will require that for any metabolite  $i$  the rate  $f_{i-1}$  is non-negative and  $f_i$  is non-positive at  $x_i = 0$ . In other words, metabolite  $i$  is not consumed when its concentration is already zero.

We will consider the following control set:

$$e \in E = \left\{ (e_0, \dots, e_{n-1}) \left| e_i \geq 0, i = 0..n-1, \sum_{i=0}^{n-1} e_i \leq E_{max} \right. \right\},$$

which indicates that at any moment in time the total enzyme concentration must not exceed a certain predefined value  $E_{max}$ . This limitation, for example,

describes limited resources of a cell that force it to choose which enzyme to produce or maintain at any point in time.

As far as the starting points are concerned, we will consider the following two most wide-spread frameworks: (A) all  $x_i(0) = 0$  and  $f_0 \geq 0$  (there is a constant supply of the initial substrate); or (B) the initial concentration  $x_1(0) = 1$ ,  $x_i(0) = 0$  for  $i = 2..n$ , and  $f_0 \equiv 0$  (the first metabolite is the initial substrate being consumed in the course of the reaction).

Finally, we will assume that the standard existence and uniqueness results hold for the solutions to (1) over the whole relevant time interval for any measurable input  $\bar{e} \in E$  [4, 5], which is usually the case in enzyme kinetics since the state vector denotes real concentrations limited from above and below. We will provide some examples of such systems in the following sections.

### 1.3 Optimal Control

In this framework, several objectives for optimal control are possible. Usually, one is interested in maximizing the final product, which can be formulated either as the minimization of the transition time  $t_f$  to drive  $x_n$  to some predefined level [6] or by maximizing  $x_n$  at a fixed point in time [7]. Other definitions of the transition time are also possible [8–10]. Moreover, a multi-objective optimization problem can also occur [11]. For the sake of simplicity, we will be considering the maximization of the final product at a given point in time although more general target functions can also be used (see below).

There are two main groups of methods commonly used to find optimal solutions: the so-called direct and indirect methods. The former usually imply a transformation of the original problem into non-linear programming by time-discretization and approximation of the control variables either alone or together with the states (for a comprehensive review see [11]). The advantages include a great variety of solvers, a general applicability, and an intuitive implementation. Nonetheless, these methods require some preliminary proof of the existence and stability of the solution. Moreover, global optima finders are much more computationally expensive than local ones, and due to the innate infinite dimensionality, the costs of refining the grid are high. Finally, if the target function is changed, e.g. to account for other metabolites, the entire calculation has to be repeated.

The indirect methods suggest analytical treatment of the problem, e.g., by using Pontryagin’s maximum principle [2, 6, 9, 10]. The main advantages include a more comprehensive analysis of the system behaviour and simpler numerical methods. However, Pontryagin’s maximum principle is only a necessary condition, and the exact analytical solutions are usually difficult to obtain even in the case of simple linear systems. The proof of a global maximum is again complicated, and any change of the model, e.g. addition of cross-inhibition, may completely invalidate the analysis.

In this article, we suggest an alternative indirect method based on exact reachable sets [12, 13], i.e. the states of a multi-enzyme system reachable from the initial point for all the possible enzyme profiles. While this method is more

computationally intensive than the maximum principle, it provides the time-evolution of the system in full since all the possible states are analysed. This allows for some flexibility in choosing the target functional after the calculation of reachable sets. Optimal control synthesis may be implemented in various ways once the sets are calculated, and the global optimality is implied automatically. No change to the model will require any qualitative re-analysis. Moreover, geometric state constraints may be taken into account, which extends the applicability of the method to, e.g. the problems with metabolite constraints due to metabolite toxicity. Finally, given some relatively broad assumptions about the reaction rates, the reachable sets are star-shaped, which reduces the problem dimensionality by one and enhances its computational efficiency and applicability. The summary table comparing the approaches mentioned above is given in the Supplement (table S1).

## 2 Star-Shaped Reachable Sets

We will now briefly define reachable sets and their applications to optimization, provide the evolution theorems for star-shaped sets, and formulate the main theoretical result for the systems in question.

### 2.1 Reachable Sets and Optimization

Reachable sets provide an important tool for the analysis of the time evolution of systems as they demonstrate how systems might behave given every possible control input. In order to demonstrate a general idea, consider the following differential inclusion:

$$\dot{\mathbf{x}} \in F(t, \mathbf{x}), \quad \mathbf{x}(t_0) \in X_0, \quad t \in T = [t_0, t_1], \quad (2)$$

where  $X_0$  is a compact subset of  $R^n$  and  $F$  is a continuous multivalued map from  $T \times R^n$  to compact convex subsets of  $R^n$ . For instance, (1) can be formulated in the above terms if one takes the union of the right-hand side of the equations over  $\mathbf{e} \in E$ . This differential equation generates a bundle of trajectories; consequently, its behaviour may be translated into that of the bundle. Let the reachable set  $X[t]$  be the set of all possible states of the system at time  $t$ . The intuitive strategy to find  $X[t]$  by inserting different values from  $F(t, \mathbf{x})$  may work only if an explicit analytical solution is available, which is hardly ever the case even for linear systems. However, under some general assumptions on  $F$ , the reachable set can be found as the solution to an evolutionary equation [14]. While this equation is usually difficult to solve, a great variety of methods has been developed to calculate such sets [12, 13, 15].

In this paper, we will use the fact that under some general assumptions (see the Supplement), inclusion 2 has reachable sets that are star-shaped [16, 17], i.e. they are compact, and for any  $\lambda \in [0, 1]$  the set  $\lambda X[t] \subseteq X[t]$ . Such sets are uniquely defined by their radial function:

$$r(\mathbf{l}, t) = r(\mathbf{l}|X[t]) = \max\{\lambda \geq 0 : \lambda \mathbf{l} \in X[t]\}$$

that is the viscosity solution to the following partial differential equation on an  $n$ -dimensional sphere  $S^n$  :

$$\frac{\partial r}{\partial t} = \rho \left( -\frac{\partial_s r}{\partial t} + r \mathbf{l} \left| \frac{1}{r} F(t, r \mathbf{l}) \right. \right), \quad (3)$$

where  $\rho(\mathbf{l}|F) = \sup\{\sum_i l_i y_i | \mathbf{y} \in F\}$  is the support function. This result, together with viscosity methods [18, 19], provides a powerful tool for an exact calculation of reachable sets, e.g. for multi-enzyme reactions as demonstrated below.

As soon as one calculates the reachable set  $X[t]$ , the optimal solution to maximizing  $x_n$  at time  $T$  is tantamount to finding the point in  $X[T]$  with the maximal value of coordinate  $x_n$ . In general, any target function dependent only on the final metabolite concentrations can be used since given  $X[T]$ , the initial optimal control problem turns into a relatively simple optimization of the function over the set  $X[T]$ . And once the optimal point has been found, one may apply control synthesis strategies to find the control profile that will lead the system to this optimum [3].

## 2.2 Star-Shaped Sets Generated by Multi-Enzyme Pathway

We will now apply the results of the previous subsection to the multi-enzyme systems (1) for initial conditions (A), i.e. some constant supply of the substrate, and (B), in which the first substrate is being consumed without any supply. The direct adaptation of Assumption S to (1) leads to the following results:

**Proposition 1.** *Suppose for system (1) with initial condition (A) the rate functions  $f_i(\mathbf{x})$  are Lipschitz-continuous with the constant independent of  $t$ . If for any  $\lambda \in (0, 1]$  and  $\mathbf{x} : f_i(\lambda \mathbf{x}) \neq 0 \Rightarrow 0 \leq \lambda f_i(\mathbf{x})/f_i(\lambda \mathbf{x}) \leq 1$ , the radial function of the reachability set  $r(\mathbf{l}, t) = r(\mathbf{l}|X[t])$  is the pointwise limit of  $r_\varepsilon(\mathbf{l}, t)$  for any  $\mathbf{l} \in S^n$  and  $t \in [0, T]$ , where  $r_\varepsilon(\mathbf{l}, t)$  is the viscosity solution to the following equation on  $S^n \times [0, T]$  :*

$$\frac{\partial r_\varepsilon}{\partial t} = E_{max} \max_i \left\{ f_i(r_\varepsilon \mathbf{l}) \left( \frac{1}{r_\varepsilon} \left( \frac{\partial_s r_\varepsilon}{\partial t} - \frac{\partial_s r_\varepsilon}{\partial l_{i+1}} \right) - l_i + l_{i+1} \right) \right\},$$

$$r_\varepsilon(\mathbf{l}, 0) = \varepsilon \rightarrow +0.$$

(here for  $i = 0$  symbols  $\partial_s r/\partial l_i$  and  $l_i$  should be omitted).

As far as initial condition (B) is concerned, we will replace the coordinate  $x_1$  with  $x_1^* = x_1 - 1$ . If in addition to the above we require that  $f_i$  is non-negative and non-decreasing in  $x_1$ , the following holds:

**Corollary 1.** *Suppose for (1) the initial concentration  $x_1(0) = 1$ ,  $x_i(0) = 0$  for  $i = 2..n$ , and  $f_0 \equiv 0$ . Moreover, suppose that in addition to the requirements of Proposition 1 on  $f_i$ , the  $f_i$  that depend on  $x_1$  are non-negative and non-decreasing in  $x_1$ . Then for (1) with the new coordinate  $x_1^* = x_1 - 1$  Proposition 1 holds.*

The proofs of the statements above are given in the Supplement.

### 2.3 Examples

Here we will list the examples of (1) relevant to the enzyme kinetics, for which Proposition 1 holds:

1. Linear mass-action kinetics  $f_i(\mathbf{x}) = k_i x_i$ ;
2. Michaelis-Menten kinetics:  $f_i(\mathbf{x}) = k_i x_i / (K_i + x_i)$ , with substrate inhibition:  $f_i(\mathbf{x}) = k_i x_i / (K_i + x_i + N_i x_i^2)$ , or with cross-inhibition:  $f_i(\mathbf{x}) = k_i x_i / (K_i + \sum_j N_{ij} x_j)$ ;
3. Power law  $f_i(\mathbf{x}) = k_i x_i^c$  with  $c \in (0, 1)$ ;

All the above functions may be present in any combination, thereby providing a significant flexibility for the model selection.

Moreover, the same enzyme can be used in different steps if the following additional requirement holds: for any enzyme  $e$  used in several reactions the value  $\lambda f_i(\mathbf{x}) / f_i(\lambda \mathbf{x})$  is independent of  $i$  for the respective  $i$ 's. This will be the case, e.g. in Michaelis-Menten kinetics since the free enzyme, and consequently, the denominator of  $f_i$ , will be the same across the respective  $i$ 's. Reversible reactions are also covered. In other cases when the star-shapedness cannot be guaranteed, one may still use general reachable set methods [13], albeit forgoing the advantage of the reduced dimensionality.

We will now proceed to several examples.

*Example 1.* The first example is a three-metabolite scheme with a constant supply of substrate zero, and it demonstrates the standard bang-bang optimal profile [2, 9, 10] (Fig. 1):

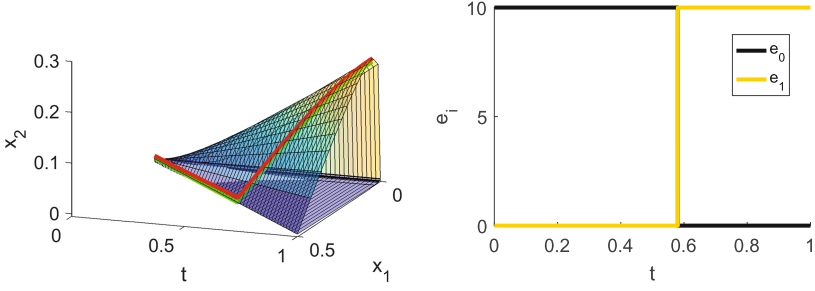
$$\begin{cases} \dot{x}_1 = \frac{0.1x_0}{1+x_0}e_0 - \frac{0.1x_1}{0.1+x_1}e_1, \\ \dot{x}_2 = \frac{0.1x_1}{0.1+x_1}e_1. \end{cases}, \quad t \in [0, 1], \quad x_0 \equiv 1, \quad E_{max} = 10. \quad (4)$$

This switching between the two regimes stems from the intuitive fact that the rate of the reaction is increasing with the increase in  $x_1$ . As a result, the optimal strategy is to accumulate  $x_1$  first and then to switch to production of  $x_2$ .

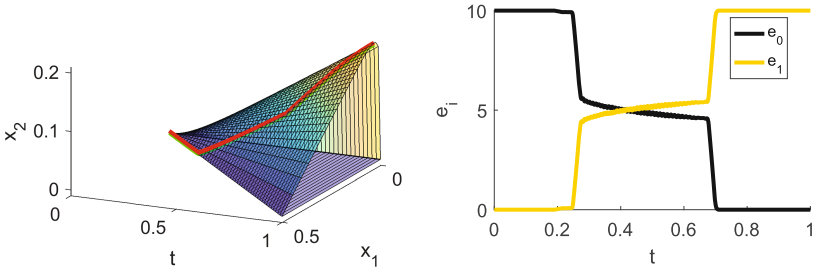
*Example 2.* The second example is a modification of the previous case with a substrate inhibition of enzyme  $e_1$  (Fig. 2):

$$\begin{cases} \dot{x}_1 = \frac{0.1x_0}{1+x_0}e_0 - \frac{0.1x_1}{0.1+x_1+5x_1^2}e_1, \\ \dot{x}_2 = \frac{0.1x_1}{0.1+x_1+5x_1^2}e_1. \end{cases}, \quad t \in [0, 1], \quad x_0 \equiv 1, \quad E_{max} = 10. \quad (5)$$

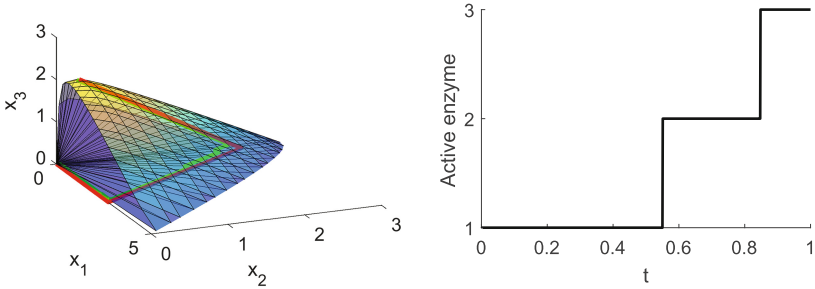
Now, the simple accumulation of  $x_1$  will not yield an optimal solution; due to the inhibition, the reaction rate would decrease for large values of  $x_1$ . Hence,  $e_1$  should be switched on earlier and not to its maximal value as can be seen from the optimal control synthesis in Fig. 2.



**Fig. 1.** The reachable tube of Example 1 (left) and the synthesized optimal control (right). The red line is the synthesized trajectory from the point with the maximal coordinate  $x_2$  backward in time. The green line is the trajectory from the origin calculated with the filtered optimal control. The calculation time on a regular desktop was 7 s.



**Fig. 2.** The reachable tube of Example 2 (left) and the synthesized optimal control (right). The red line is the synthesized trajectory from the point with the maximal coordinate  $x_2$  backward in time. The green line is the trajectory from the origin calculated with the filtered optimal control. The calculation time on a regular desktop was 7 s.



**Fig. 3.** The reachable set of Example 3 at  $t = 1$  (left) and the synthesized optimal control (right). The red line is the synthesized trajectory from the point with the maximal coordinate  $x_3$  backward in time. The green line is the trajectory from the origin calculated with the filtered optimal control. The calculation time on a regular desktop was 57 s.

*Example 3.* Finally, we will also consider a three-dimensional example to demonstrate the calculability of the method (Fig. 3):

$$\begin{cases} \dot{x}_1 = \frac{x_0}{1+x_0}e_0 - \frac{2x_1}{2+x_1}e_1, \\ \dot{x}_2 = \frac{2x_1}{2+x_1}e_1 - \frac{3x_2}{1+x_2}e_2, \\ \dot{x}_3 = \frac{3x_2}{1+x_2}e_2. \end{cases}, t \in [0, 1], x_0 \equiv 1, E_{max} = 10. \quad (6)$$

In general, the curse of dimensionality leads to a significant increase in computational costs as the dimensionality of  $x$  increases, in contrast to direct methods that are sensitive to the dimensionality of the control vector. The star-shaped sets partially alleviate the problem by reducing the dimensionality by one, which is why a two-dimensional grid was used in this example. Thus, the calculations for systems with up to 5–6 state variables can be performed on a regular desktop in a reasonable time. Otherwise, approximation techniques, e.g., ellipsoidal calculus [12] or zonotopes [15], might be used.

### 3 Conclusions

In this work, we studied a multi-enzyme optimization problem. We demonstrated that under some general assumptions, the reachable sets of such a problem are star-shaped. Further, we constructed reachable sets using their radial function that is a viscosity solution to a certain partial differential equation. By doing so, we were able to visualize the time-evolution of the system given all possible enzyme profiles. Once calculated, the reachability tube provides means for optimal control synthesis. Finally, we considered several examples that verified results obtained by other authors using different techniques as well as provided some new insights into the behavior of more sophisticated multi-enzyme pathways, e.g. the ones with inhibition.

### References

1. Carbonell, P., Parutto, P., Herisson, J., Pandit, S.B., Faulon, J.L.: XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res.* **42**(W1), W389–W394 (2014)
2. Klipp, E., Heinrich, R., Holzhütter, H.G.: Prediction of temporal gene expression. *Eur. J. Biochem.* **269**(22), 5406–5413 (2002)
3. Mazurenko, S.: Partial differential equation for evolution of star-shaped reachability domains of differential inclusions. *Set-Valued Variational Anal.* **24**(2), 333–354 (2016)
4. Filippov, A.: On certain questions in the theory of optimal control. *J. Soc. Ind. Appl. Math. Ser. A Control* **1**(1), 76–84 (1962)
5. Aubin, J.P., Cellina, A.: *Differential Inclusions: Set-Valued Maps and Viability Theory*, vol. 264. Springer, Heidelberg (1984)
6. Bayon, L., Otero, J.A., Ruiz, M.M., Suárez, P.M., Tasis, C.: Sensitivity analysis of a linear and unbranched chemical process with  $n$  steps. *J. Math. Chem.* **53**(3), 925–940 (2015)



7. Dvorak, P., Kurumbang, N.P., Bendl, J., Brezovsky, J., Prokop, Z., Damborsky, J.: Maximizing the efficiency of multienzyme process by stoichiometry optimization. *ChemBioChem* **15**(13), 1891–1895 (2014)
8. Llorens, M., Nuño, J.C., Rodríguez, Y., Meléndez-Hevia, E., Montero, F.: Generalization of the theory of transition times in metabolic pathways: a geometrical approach. *Biophys. J.* **77**(1), 23–36 (1999)
9. Bartl, M., Li, P., Schuster, S.: Modelling the optimal timing in metabolic pathway activation – use of Pontryagin’s Maximum Principle and role of the Golden section. *Biosystems* **101**(1), 67–77 (2010)
10. Oyarzun, D.A., Ingalls, B.P., Middleton, R.H., Kalamatianos, D.: Sequential activation of metabolic pathways: a dynamic optimization approach. *Bull. Math. Biol.* **71**(8), 1851–1872 (2009)
11. Hijas-Liste, G.M., Klipp, E., Balsa-Canto, E., Banga, J.R.: Global dynamic optimization approach to predict activation in metabolic pathways. *BMC Syst. Biol.* **8**(1), 1 (2014)
12. Kurzhanski, A.B., Varaiya, P.: Dynamics and Control of Trajectory Tubes. Theory and Computation. Birkhauser, Basel, (2014)
13. Mitchell, I.M., Bayen, A.M., Tomlin, C.J.: A time-dependent Hamilton-Jacobi formulation of reachable sets for continuous dynamic games. *IEEE Trans. Autom. Control* **50**(7), 947–957 (2005)
14. Panasyuk, A.I., Panasyuk, V.I.: An equation generated by a differential inclusion. *Math. Notes Acad. Sci. USSR* **27**(3), 213–218 (1980)
15. Althoff, M., Stursberg, O., Buss, M.: Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes. *Nonlinear Anal. Hybrid Syst.* **4**(2), 233–249 (2010)
16. Mazurenko, S.S.: Viscosity Solutions to Evolution of Star-Shaped Reachable Sets (2016). Submitted and is currently under review
17. Kurzhanski, A.B., Filippova, T.F.: On the theory of trajectory tubes - a mathematical formalism for uncertain dynamics, viability and control. In: *Advances in Nonlinear Dynamics and Control*, pp. 122–188. Birkhauser, Boston (1993)
18. Crandall, M.G., Lions, P.L.: Two approximations of solutions of Hamilton-Jacobi equations. *Math. Comput.* **43**(167), 1–19 (1984)
19. Souganidis, P.E.: Approximation schemes for viscosity solutions of Hamilton-Jacobi equations. *J. Differ. Eqn.* **59**(1), 1–43 (1985)

11th International Conference on Practical Applications  
of Computational Biology & Bioinformatics

Fdez-Riverola, F.; Mohamad, M.S.; Rocha, M.; De Paz,  
J.F.; Pinto, T. (Eds.)

2017, XIV, 330 p. 97 illus., Softcover

ISBN: 978-3-319-60815-0