Dynamic Optimization Approach for Genome-Scale Investigation of Metabolic Reprogramming

Biological mathematics course projects
Fall 20-21

Fateme Safaeifard, PhD student of Biophysics, IBB, University of Tehran

Parisa Vosooq Nejad PhD student of Mathematics, Sharif University of Technology

Saeed Aghamiri, PhD student of Artificial Intelligence, Sharif University of Technology

Supervised by:

Mojtaba Tefagh

Department of Mathematics, Sharif University of Technology

Metabolic reprogramming is a critical phenomenon in cellular processes associated with development, morphogenesis and cancer. In addition to the alteration of metabolites concentration, the method of dynamic optimization is capable of providing the dynamic profile of metabolic enzymes in the cells. However, the application of this method in the systems as large as metabolic networks suffers from drastic computational challenges. Here, we use techniques derived from distributed optimal control theory to set up a comprehensive landscape of metabolism during cellular reprogramming and differentiation. The proposed approach and the expected outputs introduce potential targets for controlling metabolic plasticity and help for directing cell-level evolution and engineering biological circuits.

We introduce distributed optimal control (DOC) as a mathematical framework for dynamical control of large scale metabolic models. This approach tackles the network components as separate dynamic objects which stacked together in the shade of system objectives. DOC problems can be solved in the form of a nonlinear programming so that in each time step, the time-dependent probability density function of the agents is optimized which determines the agent's optimal feedback control based on the actual state of the system (the current agent distribution).

Determination of an appropriate temporal objective function would be one of the key steps of the work. Finally, experimental data correspond to enzymes and metabolites concentrations should be used for evaluating the optimization model predictions. A verified model of E. coli kinetic model of metabolism can be used as model input and the outcomes would reveal some potential mechanisms of metabolic control in E. coli as a prokaryotic model organism.

Integration of network dynamics in metabolic networks analysis

Understanding the relationship between metabolism and key cellular processes such as environmental adaptation requires insights into the concept of optimization in the evolution of metabolic regulatory networks. While in the past the main focus of researches has been on regulating metabolic pathways based on steady state assumption, dynamic optimization has proven itself as an ideal tool for determining regulatory strategies for metabolic pathways in response to environmental changes. Numerous studies have been conducted to determine the optimal regulatory strategies and to identify the main control points in metabolic networks. (1) Wessely et al. Presented a dynamic optimization problem to determine the translational regulation program. In their approach, based on constraints such as reaction kinetics, growth rate, initial values of intermediate metabolites, concentrations of enzymes and a certain concentration range of products, the studied metabolic pathway are restricted and evolve. The objective function of the optimization problem is minimizing the cost of enzymes expression while metabolic goal can be reached (2). Hijas et al. used a dynamic optimization problem to study the feedback inhibition mechanism in the metabolic pathway regulation program. The constraints of their optimization problem include a competitive inhibition of certain enzymes by the metabolic pathway product. The existence of this competitive product affects reaction dynamics. The results of model simulations show that the control of enzymes at the branch of metabolic pathways has been selected during the evolution of pathways. (3)

Cell growth can depend on the concentration of products of some metabolic pathways. With this biological justification, in the dynamic optimization problem proposed by Bartl, et al optimization of metabolic pathway products considered as the model objective function. In this study, in addition to the dynamics of reactions, including the biomass production reaction, cell potential for the synthesis of ribosomal free proteins and cellular capacity for the synthesis of enzymes were considered as system constraints. Simulations based on this problem show that increasing the synthesis capacity of free proteins leads to a simultaneous change in the expression of pathway enzymes, while increasing the synthesis rate of individual enzymes leads to regulating the expression of pathway enzymes in a stepwise manner. (4, Fig. 1)

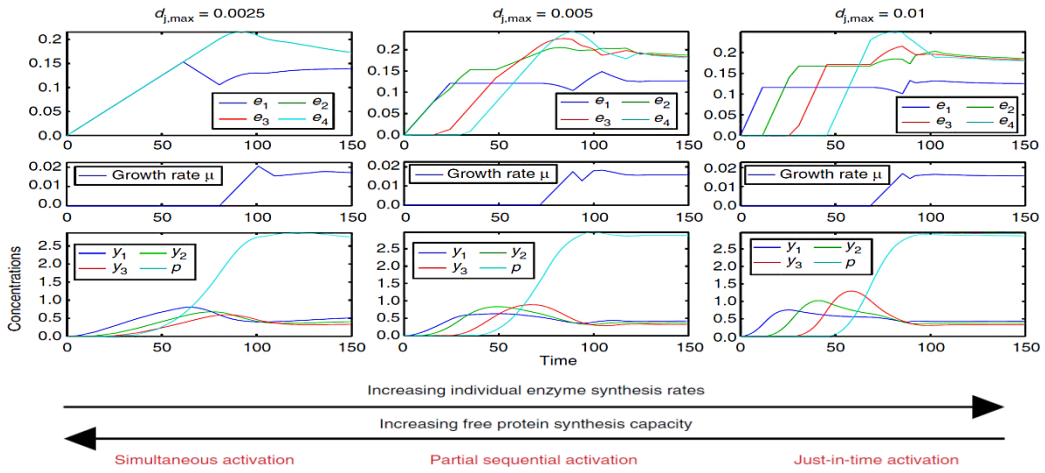

Figure 1 Dependence of metabolic pathway control enzyme control strategy on cellular protein synthesis capacity

Dynamic optimization framework for modeling metabolic pathways has shown its potential in multi-objective optimization problems. One of such efforts aims for modeling the metabolic shift mechanism in Saccharomyces cerevisiae. This model is limited by optimizing cell life time and minimizing changes in the concentration of enzymes in the studied metabolic pathway with constraints such as upper bound concentration of metabolites. The modeling results show that minimizing the cost of controlling the expression of enzymes up to a threshold value will not significantly affect cell lifetime. (5)

Finally, efforts of Waldherr et al. to integrate the metabolic network with changes in biomass extent and composition, in addition to predicting changes in enzymatic expression and reaction flux, was able to predict the composition of biomass in a time-dependent manner. Reducing the computational cost of this method as an extension of the classical FBA method has made it possible to apply it to more complex networks. (6)

Numerical methods for solving dynamic optimization problems in systems biology

There are three approaches for numerical solution of nonlinear optimal control problems: dynamic programming, indirect and direct methods. Dynamic programming is not popular due to the curse of dimensionality, so the latter two are promising methods for complex problems. Indirect approaches take the advantage of Pontryagin's necessary conditions to transform the original problem into a multi-point boundary value problem, Direct methods transform the optimal control problem into a nonlinear programming problem based on the discretization of either the control, or both the control and the states variables, the first known as the sequential strategy and the later as the simultaneous or complete strategy (7,8).

Sequential strategy, also known as control vector parametrization, usually uses low order Lagrange polynomials, the coefficients of which are the control variables. This gives a problem including an outer non-linear programming (NLP) problem and an inner initial value problem (IVP) where the cost function evaluated iteratively and the control variables are approximated and the system is integrated for each time.

Polynomial coefficients correspond to time invariant parameters that determine the controls. Therefore, we have a non-linear programming problem with dynamic (the model) and algebraic constraints, where the decision variables as unknown parameters (9).

In **complete parameterization**, both states and controls are discretized by dividing the time into intervals. In the most popular approach the solution is transcribed into a non-linear programming by means of low-order polynomial approximations and a K-stage Runge-Kutta theorem is used for integration steps. Indeed this method transforms the infinite dimensional problem into a large NLP without requiring the system integration in each solution iteration (10).

Bartl et al. (11) used quasi-sequential approaches for determination of optimal control variables and to investigate parameters of pathway activation model. While those deterministic and gradient-based methods enable a fast convergence to local optimal solutions of NLPs, global methods identify the global optimum in more robust manner and stochastically investigate solution space however in the cost of computation time. De Hijas-Liste et al. (5,12) applied a hybrid of local and global models for dynamic optimization for more efficient multi objective optimization of metabolic networks.
In a recent work, two phase strategy for numerical optimal control was used where in phase (I), a hybrid stochastic-deterministic method based on control vector parameterization used to reach the global solution and in phase (II), a complete discretization fast local method used for computational enhancement (8).

The importance of numerical approach for the effects of constraints on optimal solution

In optimal control the equivalent of the shadow price is the adjoint variable. The Lagrange multipliers are discrete approximation of the adjoint variables.

Considering the original infinite-dimensional optimal control problem in the Lagrangian form:

$$\mathcal{L}(\mathbf{y}, \boldsymbol{\alpha}) = \mathbf{F}(\mathbf{y}) - \boldsymbol{\alpha}^T \mathbf{c}(\mathbf{y}) = \mathbf{F}(\mathbf{y}) - \sum_{i=1}^{m} \alpha_i c_i(\mathbf{y})$$

The transformed problem to a discretized (infinite-dimensional) NLP problem would be in the form of:

$$L(\mathbf{y}, \boldsymbol{\lambda}) = \mathbf{F}(\mathbf{y}) - \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{y}) = \mathbf{F}(\mathbf{y}) - \sum_{i=1}^{m} \lambda_i c_i(\mathbf{y})$$

The difference between to adjoint variables and lagrange multiplier in two Lagrangian forms (L & £) depends on how the numerical method for approximation of continuous variables (8).

Distributed Optimal Control for dynamic optimization of metabolic network

In distributed optimal control, the cost function can represent a more general form of objectives than other methods which can provide solutions with strong couplings between the agent behavior and control laws. Furthermore, DOC approach is applicable for the macroscopic descriptions, other than expectation of the agent distribution such as the agent probability density function and its moments, thereby covers a broad range of collaborative behaviors and objectives. It does not need for assigning the agent distribution a priori. Instead, optimizes system behavior that is constrained

by microscopic dynamics thus needs for a precise macroscopic evolution equation and the corresponding restriction operator that characterize the multiscale system to reduce the computational complexity of the problem.
As a result, coupled agent objectives and control laws can be considered over large spatial and time scales in feasible computational frame (13).

We assumed that the metabolite cooperate toward cellular objectives by modifying enzymes expression as microscopic control such that, at large networks and in broad temporal scales, the metabolic system performance over a time interval ($\boldsymbol{T_0}$, $\boldsymbol{T_f}$] can be expressed as an integral cost function of $\boldsymbol{u_i}$ and a macroscopic state variable $X(t) = p(\boldsymbol{x_i}, t)$:

$$J = \phi[p(\mathbf{x}_i, T_f)] + \int_{T_0}^{T_f} \int_{\chi} \mathcal{L}[p(\mathbf{x}_i, t), \mathbf{u}_i(t), t] d\mathbf{x}_i dt,$$

where p is chosen to be a time-varying PDF as a restriction operator. we suppose metabolites can be described by a small system of SDEs in the form:

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}[\mathbf{x}_i(t), \mathbf{u}_i(t), t] + \mathbf{G}\mathbf{w}_i(t), \quad \mathbf{x}_i(T_0) = \mathbf{x}_{i0},$$

where $\boldsymbol{x_i}$ and $\boldsymbol{u_i}$ correspond to the microscopic agent state and control, respectively, in this paradigm an additive Gaussian disturbance vector and the constant matrix **G** are included for random variations of parameters (14).

Discretization of Distributed Optimal Control problem

Parameterization of metabolites PDF over the solution domain can represented as a mixture model of Gaussian distribution:

$$f_j(\mathbf{x}_i, t) = \frac{1}{(2\pi)^{n/2} \mid \Sigma_j \mid^{1/2}} e^{[-(1/2)(\mathbf{x}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j)]}$$

Which provides an approximation as a superposition of z PDFs components, their corresponding proportions or weights, denoted by w. at any $t \in (\boldsymbol{T_0}, \boldsymbol{T_f}]$ the optimal agent distribution can be represented as

$$p(\mathbf{x}_i, t) = \sum_{j=1}^{z} w_j(t) f_j(\mathbf{x}_i, t),$$

z is an arbitrary fixed parameter. time-varying optimal parameters of the model determines the agent distribution p.

Determination of the parameters must hold the component densities positive and with summation equals 1 (normalization condition) for all times:

$$p_k = p(\mathbf{x}_i, t_k) = \sum_{j=1}^{z} w_j(t_k) f_j(\mathbf{x}, t_k)$$
$$\equiv \sum_{j=1}^{z} w_{jk} \frac{1}{(2\pi)^{n/2} \left|\Sigma_{jk}\right|^{1/2}} e^{[-(1/2)(\mathbf{x}_i - \boldsymbol{\mu}_{jk})^T \boldsymbol{\Sigma}_{jk}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_{jk})]}.$$

As well, for state- space discretization we consider the metabolites PDF as a conserved quantity with temporal evolution which is governed by the advection-diffusion equation and ignoring the input noise we have:

$$\int_{\chi} p(\mathbf{x}_i, t)\, d\mathbf{x}_i = 1,$$

$$\mathbf{v}_i(t) = \mathbf{f}[\mathbf{x}_i(t), \mathbf{u}_i(t), t]$$

$$\frac{\partial p}{\partial t} = -\nabla \cdot [p(\mathbf{x}_i, t)\, \mathbf{v}_i(t)] + \nabla \cdot [(\mathbf{G}\mathbf{G}^T) \nabla p(\mathbf{x}_i, t)]$$

$$\frac{\partial p}{\partial t} + (\nabla p) \cdot \mathbf{f} + p(\nabla \cdot \mathbf{f}) = 0,$$

Using finite- volume approach we divide the state-space X into FVs by a constant interval $\Delta\mathbf{x}$ which centered about the points $\boldsymbol{x_l}$,with $l = 1, \ldots, L$

$$\rho_k \triangleq -\int_S [p_k\, \mathbf{f}(p_{l,k}, \mathbf{u}_{l,k}, t_k)] \cdot \hat{\mathbf{n}}\, dS,$$

Where

$$p_{l,k,} = p(\mathbf{x}_l, t_k)$$

$$\mathbf{u}_{l,k} = \mathbf{c}[p(\mathbf{x}_l, t_k)]$$

Therefor:

$$p_{k+1} = p_k + \Delta t \rho_k,$$

Finally, the discretized DOC problem as the finite-dimensional NLP would be (13-14):

$$\begin{aligned}
\min J_D = & \sum_{j=1}^{n} \Delta \mathbf{x}_{(j)} \sum_{l=1}^{L} [\phi_{l,K} + \Delta t \sum_{k=1}^{K} \mathcal{L}(p_{l,k}, \mathbf{u}_{l,k}, t_k)], \\
\text{subject to } & p_{k+1} - p_k - \Delta t \rho_k = 0,\ k = 1, \ldots, K, \\
& \sum_{j=1}^{n} \Delta \mathbf{x}_{(j)} \sum_{l=1}^{L} p_{l,k} - 1 = 0,\ k = 1, \ldots, K, \\
& p_{l,0} = p_0(\mathbf{x}_l),\ \mathbf{x}_l \in \mathcal{X}, \\
& p_{l,k} = 0,\ \mathbf{x}_l \in \partial\mathcal{X},\ k = 1, \ldots, K,
\end{aligned}$$

Which is a finite dimensional NLP can be solved using an SQP algorithm by representing as a sequence of unconstrained quadratic programming subproblems (15, Fig. 2).

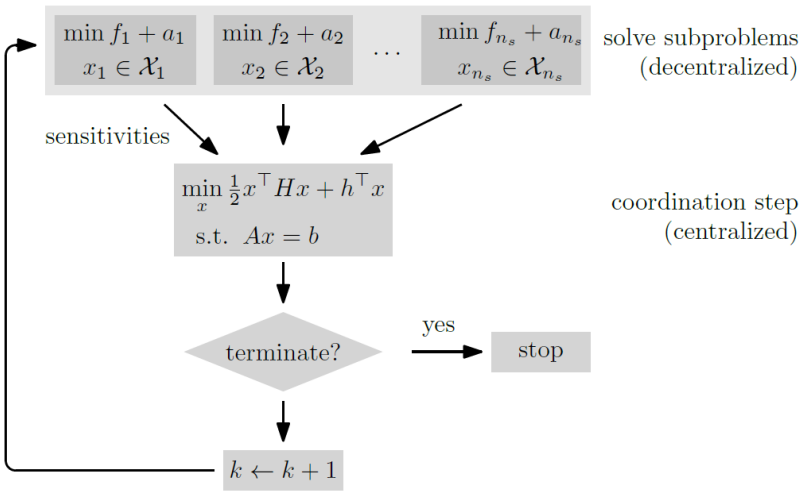

Figure 2 Simplified flow chart of standard NLP solution

Concluding remarks

In this study, the integrative approaches of metabolic network dynamics and constraint based models was investigated in the closure of optimal control theory. The basis of the methods in this integrative paradgim is the use of kinetic models as additional constraints in the stoichiometric modeling of the metabolic network. Because of limited access to the kinetic information associated with reactions, including information on kinetics of enzymes and their concentrations, optimization of metabolic network dynamics often remains at the level of model metabolic pathways, including multiple reactions or, ultimately, limited and interconnected pathways. However, distributed optimal control introduced in this paper, seems to be able for analyzing wider networks compared to other methods due to the potential for using phenotypic information while tracing microscopic evolution of microscopic transitions.
Therefore, unlike methods for integration of regulation in metabolic analysis (16-18), there is no need to know the regulatory interactions extensively, because the proposed model can use only, the metabolic cost minimization of enzymes production as cellular goal. In fact, there can be two different approaches for modeling metabolic gene regulation based on enzymes expression: reconstructing metabolic models using the transcriptional constraints which directly used in reconstruction algorithm, and using the constraints on the capacity and cost of production, the same method studied in this paper.

A unique feature of DOC method is the prediction of temporal changes of gene expression based on optimization concepts, without the use of information based on regulatory interactions. This feature, along with the limited dependence on kinetic data, improves the computational efficiency of the model compared to other dynamic methods and makes the analysis of relatively complex networks possible.
In general, the integration of appropriate constraints related to regulatory mechanisms in the optimization of metabolic networks is still an open issue in the modeling of these networks. However, DOC method along with other proposed models, shows a promising framework for tracing back the plasticity of the metabolic system in relation to environmental changes.

References:

1. Ewald, Jan, Martin Bartl, and Christoph Kaleta. "Deciphering the regulation of metabolism with dynamic optimization: an overview of recent advances." *Biochemical Society Transactions* 45.4 (2017): 1035-1043.

2. Wessely, Frank, et al. "Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs." *Molecular systems biology* 7.1 (2011): 515.

3. de Hijas-Liste, Gundián M., et al. "Optimal programs of pathway control: dissecting the influence of pathway topology and feedback inhibition on pathway regulation." *BMC bioinformatics* 16.1 (2015): 163.

4. Bartl, Martin, et al. "Dynamic optimization identifies optimal programmes for pathway regulation in prokaryotes." *Nature communications* 4 (2013): 2243.

5. de Hijas-Liste, Gundián M., et al. "Global dynamic optimization approach to predict activation in metabolic pathways." *BMC systems biology* 8.1 (2014): 1.

6. Waldherr, Steffen, Diego A. Oyarzún, and Alexander Bockmayr. "Dynamic optimization of metabolic networks coupled with gene expression." *Journal of theoretical biology*365 (2015): 469-485

7. Ewald, Jan, Martin Bartl, and Christoph Kaleta. "Deciphering the regulation of metabolism with dynamic optimization: an overview of recent advances." *Biochemical Society Transactions* 45.4 (2017): 1035-1043.

8.Tsiantis, Nikolaos, and Julio R. Banga. "Using optimal control to understand complex metabolic pathways." *bioRxiv* (2020)

9. Shokry, A. and Espuña, A. (2014) Sequential dynamic optimization of complex nonlinear processes based on kriging surrogate models. ProcediaTechnol. 15, 376–387 doi:10.1016/j.protcy.2014.09.092

10 Biegler, L.T. (2007) An overview of simultaneous strategies for dynamic optimization. Chem. Eng. Process. 46, 1043–1053 doi:10.1016/j.cep.2006.06.021

11. Bartl, Martin, et al. "Dynamic optimization identifies optimal programmes for pathway regulation in prokaryotes." *Nature communications* 4 (2013): 2243.

12. de Hijas-Liste, Gundián M., et al. "Optimal programs of pathway control: dissecting the influence of pathway topology and feedback inhibition on pathway regulation." *BMC bioinformatics* 16.1 (2015): 163.

13. Ferrari, Silvia, et al. "Distributed optimal control of multiscale dynamical systems: a tutorial." *IEEE Control Systems Magazine* 36.2 (2016): 102-116.

14. G. Foderaro, S. Ferrari, and T. A. Wettergren, “Distributed optimal control of sensor networks for dynamic track coverage,” submitted for publication

15. Engelmann, Alexander, et al. "ALADIN-$\alpha $--An open-source MATLAB toolbox for distributed non-convex optimization." *arXiv preprint arXiv:2006.01866* (2020).

16. Covert, Markus W., and Bernhard Ø. Palsson. "Transcriptional regulation in constraints-based metabolic models of Escherichia coli." *Journal of Biological Chemistry* 277.31 (2002): 28058-28064.

17. Covert, Markus W., Christophe H. Schilling, and Bernhard Palsson. "Regulation of gene expression in flux balance models of metabolism." *Journal of theoretical biology* 213.1 (2001): 73-88.

18. Lerman, Joshua A., et al. "In silico method for modelling metabolism and gene product expression at genome scale." *Nature communications* 3 (2012): 929

“Many thanks Dr. ***Mojtaba Tefagh*** for all support and guidance”