PLANNING IDENTIFICATION EXPERIMENTS FOR CELL SIGNALING PATHWAYS USING SENSITIVITY ANALYSIS

Krzysztof Fujarewicz

Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
krzysztof.fujarewicz@polsl.pl

Keywords: Cell signaling pathways, sensitivity analysis, experiment design.

1. INTRODUCTION

One of possible approaches to modeling of cell signaling pathways is to use a set of nonlinear ODEs [de Jong 2002]. In order to estimate unknown parameters of such a model several experiments are performed, during which concentrations of part of variables are measured at rare discrete time moments. Usually, a blotting technique is used. In this work we focus on choosing optimal time moments for experiments. This problem has been investigated in the related literature. To solve it a matrix of correlation coefficients between sensitivities of measurements with respect to identified parameters is calculated [Jacquez, Greif 1985], [Jacquez 1998]. Then one tries to choose such time moments for which sensitivities are "less correlated". The standard approach to optimization is the non-gradient Gauss-Seidl technique.
In this work we calculate the gradient of the function of the correlation matrix with respect to times of measurements, then we propose a gradient-based algorithm.

2. PROBLEM FORMULATION

Let us consider a model of a cell signaling pathway in a form of a set of non-linear ODEs:

$$\dot{x} = f(x,u,\theta); \quad x(0) = x_0 \tag{1}$$

where x is a vector of state variables, u is an input signal and $\theta \in R^p$ is a vector of identified parameters.

The output equation is as follows

$$y = g(x,u) \tag{2}$$

For the simplicity of notation let us assume there is only one output variable which is measured at times $t_1, t_2, \ldots, t_n$ giving instantaneous values

$$y(t_i) = g(x(t_i), u(t_i)) = g_i; \quad i = 1,2,\ldots,n \tag{3}$$

After performing experiments one obtains observations

$$z(t_i) = y(t_i) + \varepsilon_i; \quad i = 1,2,\ldots,n \tag{4}$$

where ε_i is an error of zero mean and variance σ_i^2. We assume that we have initial (rough) estimation of parameters $\theta_i^0, \theta_2^0, \ldots, \theta_n^0$ for which measured variable (3) takes values $g_i^0, g_2^0, \ldots, g_n^0$.
We build the sensitivity matrix [Jacquez, Greif 1985], [Jacquez 1998] as follows:

$$G = \begin{bmatrix} \frac{\partial g_i^0}{\partial \theta_1} & \cdots & \frac{\partial g_i^0}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^0}{\partial \theta_1} & \cdots & \frac{\partial g_n^0}{\partial \theta_p} \end{bmatrix} \tag{5}$$

and based on it the Fisher information matrix

$$I = G^T \Sigma^{-1} G \tag{6}$$

where $\Sigma^{-1} = diag[1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_n]$.

If the determinant of I is non-zero then I^{-1} is proportional to the covariance matrix of the estimates of θ. We want the covariance matrix to me small. The criterion widely used is to maximize the determinant of I. This is so called *D*-optimal design [Jacquez 1998]. The standard approach is to maximize $\det(I)$ by finding optimal times $t_1, t_2, \ldots, t_n$ using non-gradient Gauss-Seidl technique. To formulate a gradient-based algorithm we formulate following problem:

Problem. Find derivatives

$$\frac{\partial \det(I)}{\partial t_i}; \quad i = 1, 2, \ldots, n \tag{7}$$

3. PROBLEM SOLUTION

Note that $\det(I)$ is a function of the sensitivity matrix G so the problem stated above is to find the "sensitivity function of the function of other sensitivity functions".

For particular time moment t_i one may write

$$\frac{\partial \det(I)}{\partial t_i} = \sum_{j=1}^{p} \frac{\partial \det(I)}{\partial \left(\partial g_i^0 / \partial \theta_j \right)} \cdot \frac{\partial \left(\partial g_i^0 / \partial \theta_j \right)}{\partial t_i} \tag{8}$$

Let us denote the first factor under the sum (8) by q_{ji} and the second factor by r_{ji}. Then let us build matrices $Q = [q_{ji}], R = [r_{ji}], \; Q, R \in R^{n \times p}$.

It can be shown that whole matrix Q may be calculated as follows

$$Q = 2\Sigma^{-1} G \operatorname{adj}(I) \tag{9}$$

The element r_{ji} is the derivative w.r.t. time of the output of the following sensitivity model for the original model (1),(2):

$$\begin{aligned} \dot{\bar{x}} &= f_x(t)\bar{x} + f_u(t)\bar{u} + f_\theta(t)\bar{\theta}; \; \bar{x}(0) = 0 \\ \bar{y} &= g_x(t)\bar{x} + g_u(t)\bar{u} \end{aligned} \tag{10}$$

taken at time t_i where the sensitivity is calculated for θ_j which means the variation $\bar{\theta}$ is a vector of zeros except one element number j which equals 1.

Unfortunately, practical using of (10) requires numerical derivation w.r.t. time. Hopefully, it is possible to derive following formula

$$r_{ji} = \left(g_{xx}(t_i)\bar{x}(t_i) + g_x(t_i) \right) \dot{\bar{x}}(t_i) \tag{11}$$

that does not require any numerical derivation because both $\bar{x}$ and $\dot{\bar{x}}$ appear in the sensitivity model (10).

In order to maximize $\det(I)$ one can combine results (9) with (11) and compute all derivatives (7) and use any gradient-based optimization algorithm.

4. CONCLUSIONS

This work is concerned with the problem of optimal design of experiments in sense of finding optimal times for measurements. The approach has one drawback at first glance. To design the experiment, which is performed in order to estimate parameters, we need initial (rough) estimations of parameters. It may be hard to guess such initial values of parameters and the experiment may be designed for quite different point in the parameter space. However, in practice several experiments for different variables are conducted and it is possible to plan each experiment based on all previous experiments starting with first non-optimally designed experiment.

Having such a tool for experiment planning, two closely related problems may be solved: how many time moments are enough to estimate parameters and what variable (concentration of proteins, protein complexes or mRNA) should be measured.

ACKNOWLEDGEMENT

This work has been supported by Polish Ministry of Education and Science under grant 3T11A 019 29.

REFERENCES

Jacquez, J.A. and P. Greif, "Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design," Math. Biosci., vol. 77, pp. 201–227, 1985.

Jacquez, J.A., Designs of experiments, J. Franklin Inst., vol. 335B, No. 2., 259-279, 1998.

de Jong, H., Modeling and Simulation of Genetic Regulatory Systems: A Literature Review, Journal of Computational Biology, Vol. 9, No 1, pp. 67–103, 2002.