

# Interpolation Effects on Gene Network Modeling by Zhegalkin Polynomials

## Contact:

Prof. Fabian Theis  
Computational Modeling in Biology  
Institute of Bioinformatics and Systems Biology  
Helmholtz Zentrum Muenchen

## Prerequisites:

Basic knowledge in model and parameter identification, reasonable Mathematical background, basic knowledge of discrete/Boolean modeling, some experience with Matlab.

## Background:

Gene expression measurements are usually available for only a few selected time points. For certain modeling method, in order to infer gene network models from the given data set, these measurements are usually interpolated using various interpolation methods e.g., linear or cubic splines etc. However there is no guarantee that these interpolations provide a good approximation to the real gene expression at the missing time points.

The gene network modeling method based on Zhegalkin Polynomial identification (Faisal *et al*, 2009) is one such method that has on one hand the ability to render a quantitative nonlinear model from the given gene expression data and on the other hand preserves the qualitative behavior of gene networks which is similar to Boolean networks.

Zhegalkin Polynomial based modeling method has so far been used by making use of the above mentioned interpolation methods.

## Objectives of the thesis:

It is important to systematically evaluate interpolation techniques first on toy examples later on real data in order to gain knowledge about the impact of interpolations on the model inferred by the method. In particular we ask the following questions:

Is it possible to work in specific function basis without actually interpolating the data set?  
Can this be integrated in the likelihood after formulation of inference as Max. Likelihood estimation?

## Schedule:

- Literature review
- Fixing interpolation techniques and inferring the model after interpolation.
- Working with the Zhegalkin Polynomial's basis

## Preliminary Exercise:

- *Theoretical*: Convert all Boolean functions of two variables to Zhegalkin Polynomials and all Zhegalkin Polynomials of two variables to Boolean functions (read: Faisal *et al*, 2009).
- *Matlab*: Randomly generate expression data set for a set of 3 genes for selected time steps 0hr, 4hrs, 12hrs and 24 hrs. All the expression values must lie in the unit interval. Interpolate them using cubic interpolation. Choose one gene and compute the polynomial that describes its behavior best with the help of other two genes. (read: Faisal *et al*, 2009).