Inference of non-linear genomic dynamics

Ernst Wit
University of Groningen

e.c.wit@rug.nl
http://www.math.rug.nl/∼ernst

23 January 2014

Law of Mass Action

Given a basic reaction

$$A + B \overset{k_{-1},k_1}{\longleftrightarrow} C$$

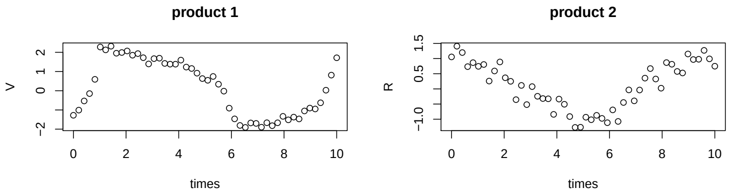
the rate of forward and backward and backward reactions is **linearly** proportional with concentration A, B and C respectively:

$$\frac{d[A]}{dt} = k_{-1}[C] - k_1[A][B].$$

rijksuniversiteit groningen

What we want

Given noisy data



and a ODE description of the system

$$\frac{dV}{dt} = c\left(V - \frac{V^3}{3} + R\right), \qquad \frac{dR}{dt} = \frac{1}{c}(V - a + bR)$$

can we infer the dynamic parameters of the system a, b, c?

Dynamic parameters

Parametrized ODE describes *background* knowledge of system; dynamic parameters describe *actual* system.

Elements of the problem

Differential equation

The change in the concentration of some element is described by

$$P_\theta x(t) = f(x, u|\beta, t),$$

where

- $P_\theta = \sum_{k=1}^{d} \theta_k D^{k-1}$, $d > 1$ and $D^k x(t) = \frac{d^{(k)}}{dt^k} x(t)$,
- f describes dynamics type.
- u are known inputs, β describes actual dynamics.

Data

Random sample of noisy observations of state variable $x(t)$

$$y_i \sim N(x(t_i), \sigma^2), \qquad \text{for } t_1, \ldots, t_m$$

Maximum likelihood

Maximum likelihood

Maximize

$$\ell(\theta) = \frac{-1}{2\sigma^2}\sum_{j=1}^{m}(y_j - x(t_j))^2 \quad subject \quad to \quad P_\theta x(t) = f(x, u|\beta, t)$$

Regularized likelihood

Definition: *regularized likelihood*

$$\ell_\lambda(\theta) = -\frac{1}{2\sigma^2}\sum_{j=1}^{m}(y_j - x(t_j))^2 + \lambda\|x\|^2_{\mathcal{H}}$$

where $\lambda > 0$ and

$$\|x\|^2_{\mathcal{H}} = \int_T (P_\theta x(t) - f(x, u|\beta, t))^2 dt.$$

Initially, we consider:

$$f = 0.$$

Regularized likelihood can be rewritten using Green function K of P^*P:

$$\ell_\lambda(\theta) = \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{K}_\theta\alpha)^T(\mathbf{y} - \mathbf{K}_\theta\alpha) + \lambda\alpha^T\mathbf{K}_\theta\alpha$$

Explicit solution to the problem

Representer Theorem [Kimeldorf and Wahba, 1970]

Minimizer of $\ell_\lambda(\theta)$ exists, is unique and admits a representation:

$$x^*(t) = \sum_{j=1}^{m} \alpha_j K_\theta(t_j, t), \quad \forall t \in T \, .$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)^T$ is the solution of the linear system

$$\alpha = (\lambda \sigma^2 \mathbf{I}_n + \mathbf{K}_\theta)^{-1} \mathbf{y}$$

where $(\mathbf{K}_\theta)_{ij} = K(t_i, t_j)$, and $\mathbf{y} = (y_1, \ldots, y_m)^T$.

- We can then maximize $\ell_\lambda(\theta)$ directly w.r.t. θ.

R package odest

We have implemented an R-package for general ODEs of type

$$P_\theta x(t) = f(x, u|\beta, t)$$

- Requires observations at specific time points.
- Requires LHS differential operator;
- Requires RHS function;
- Can deal with multivariate ODEs.
- Doesn't need observations on all dimensions.
- Doesn't need derivatives.