Inverse Dynamical Analysis of Gene Networks

James Lu[1]

[1] Johann Radon Institute for Computational and Applied Mathematics (RICAM)

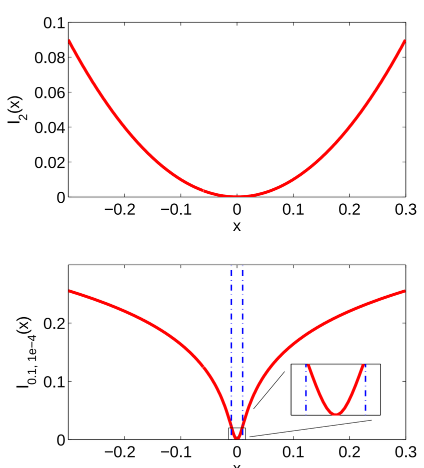
Collaborators: Heinz W. Engl (RICAM), Peter Schuster, Rainer Machnè (TBI, Vienna)

Methodology

- Maps: dynamical behavior → model mechanisms. May be applied at various stages of biological modelling.
- In initial modelling of gene regulation systems:
  - probe the possibility for the model to exhibit bistability or oscillations;
  - characterize parameter variations that can give rise to different qualitative dynamics.
- Given a plausable model:
  - identify mechanisms in model that can give rise to various *bifurcation phenotypes*: proposes new experiments to verify or falsify mathematical prediction;
  - once the model is satisfactory, design the system for desired dynamical characteristics.
- Mathematical methods:
  - inverse eigenvalue analysis;
  - inverse bifurcation analysis.
- Inverse problems are typically ill-posed: solution may not be unique and depend continuously on data.
- Various regularization techniques have been developed [1]. For biological applications, propose *sparsity-promoting* regularization to identify 'influential' mechanisms.

l_p-regularization, $p \leq 1$

While stabilizing ill-posed problems, regularization methods bring bias to the solution. Depending on the mathematical properties of the problem and the application of interest, different regularization techniques may be appropriate. Does one want to obtain a solution of the minimum Euclidean norm? Or is it more desirable to obtain a solution that is sparse, i.e., has as few non-zeros as possible? For biological applications, sparsity is useful when one wishes to identify a small number of parameters whose variation can be mapped to a wide range of system behavior.



- For positive p, consider functionals $l_p(x) = \sum_i |x_i|^p : \mathbb{R}^m \rightarrow \mathbb{R}$
- For the (limiting) case $p = 0$, the functional measures the number of non-zeros in x
- For linear problems under *sufficient-sparsity* condition, constrained minimization of l_0 functional can be obtained by solving the corresponding l_1 problem: D. L. Donoho and M. Elad, *PNAS* (2003)
- For nonlinear problems, consider smoothed l_p-functionals $l_{p,\epsilon}(x) = \sum_i (x_i^2 + \epsilon)^{p/2}$
- Open question: regularization properties of $l_{p,\epsilon}$ penalty

Invese Eigenvalue Analysis

Definition (IEP for Saddle-Node and Hopf bifurcations). *Denote scalars $\lambda_{SN} = \{0\}$ or $\lambda_H = \{\pm\omega i\}$. With equilibrium condition $f(x,\alpha) = 0$, determine parameter values α such that $\sigma(\frac{df}{dx}) \supset \lambda_{SN,H}$.*

- Least square formulation with regularization:

$$\min_{x,\alpha,U,T} \frac{1}{2}\|\frac{df}{dx} - UT(\Lambda_{SN,H};\tilde{T})U^H\|_{\mathcal{F}}^2 + \mu\|\alpha - \alpha^*\|_{l_p}^p \quad s.t. \quad f(x,\boldsymbol{\alpha}) = 0$$

- Numerical solution by lift-and-projection iterations
  - Approximate lift for non-symm matrices: Schur decomposition of the Jacobian, R. Orsi, *SIAM J. Matrix Anal. Appl.* (2006)
  - Projection: solve a nonlinear program, e.g. by Sequential Quadratic Programming (SQP)

Inverse Bifurcation Analysis

- Underlying strategy: mapping geometry of bifurcation diagrams → biochemical parameters [2, 3].
- Consider the splitting of m-dimensional parameter space $\alpha \subset \mathbb{R}^m$ into input and system parameters, $\alpha = (\alpha_i, \alpha_s) \in \alpha_i \times \alpha_s$.
- Let the *forward operator* F be a mapping in parameter space, taking a given point to its orthogonal projection on the bifurcation manifold, $\Sigma(\alpha_s)$. That is, $F(\alpha) = (\alpha_{\Sigma(\alpha_s)}\alpha_i, \alpha_s)$.
- Inverse bifurcation can be formulated as:

$$\min_{\alpha_s} J(\alpha) = \|F(\alpha)_i - \alpha_i\| \quad \text{s.t.} \quad \alpha_{\text{low}} \leq \alpha \leq \alpha_{\text{upp}}, \quad 0 \leq c(F(\alpha)_i),$$

where $\|\cdot\|$ denotes the l_2-norm and c represents k-dimensional nonlinear constraints. With the right and left eigensystems of the Jacobian for the vector field f denoted as $f_x v = \omega_{\text{crit}} v$, $f_x^T w = \bar{\omega}_{\text{crit}} w$, the expressions for adjoint solution in the case of saddle-node and Hopf bifurcations are given by:

$$\text{Saddle-node} : w^H f_\alpha, \qquad \text{Hopf} : \text{Re}\left\{\frac{1}{w^H v}\left(w^H f_{x\alpha} v - \lambda^H f_\alpha\right)\right\},$$

where $\lambda \equiv f_x^{-T}(w^H f_{xx} v)$ and superscript H denotes conjugate transpose.

Hierarchical identification strategy

In general there are multiple solutions to inverse eigenvalue/bifurcation problems; hierarchical strategy identifies a *sequence* of parameter sets, of possibly increasing cardinality.

Alg: HIER-PARAM-IDENT(α_s^0, MaxLev, p,ϵ)

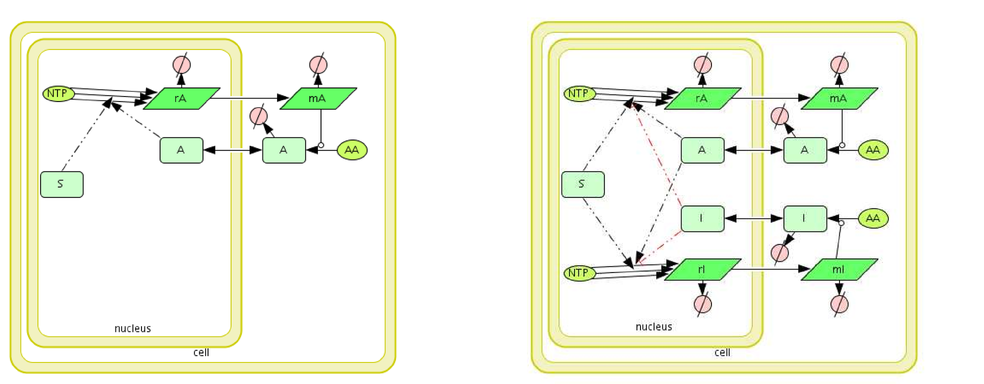
- Initialize: $s \leftarrow \{1, \cdots, m\}$, $I_{\text{identified}} \leftarrow \varnothing$
- For $j = 1, \cdots$, MaxLev
  - $I_{\text{rem}} \leftarrow s \setminus I_{\text{identified}}$
  - Solve $\alpha_{I_{\text{rem}}}^j \leftarrow$ ParamIden($\alpha_{I_{\text{rem}}}^0, p, \epsilon$)
  - $I_j \leftarrow \{i : |(\alpha_{I_{\text{rem}}}^j)_i| > \sqrt{\epsilon}\}$
  - $I_{\text{identified}} \leftarrow I_{\text{identified}} \cup I_j$

  End
- **Return** $\{\alpha_{I_1}^1, \alpha_{I_2}^2, \alpha_{I_3}^3 \cdots\}$

Applications

Evolutionary scenario of GATA transctiption factor system

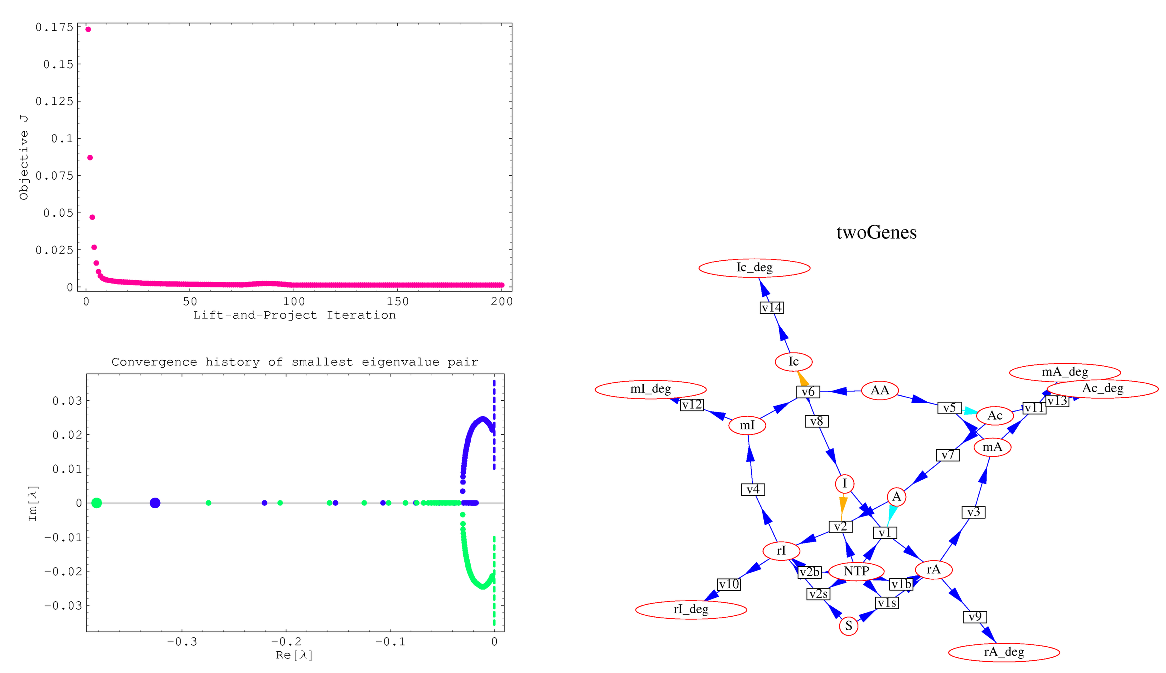
Consider the following evolutionary scenario: duplication of gene *a*, including its regulatory region but without its trans-activating domain, yielding competitive inhibitor gene *i*.



Evolution: from 1 to 2-gene network

Question: can oscillations emerge out of the duplication scenario? Can this new dynamical behavior occur with few asymmetries between A and I?

IEA: 1 asymmetry pair (cytoplasm-to-nucleus shuttling rate) is sufficient to induce oscillations



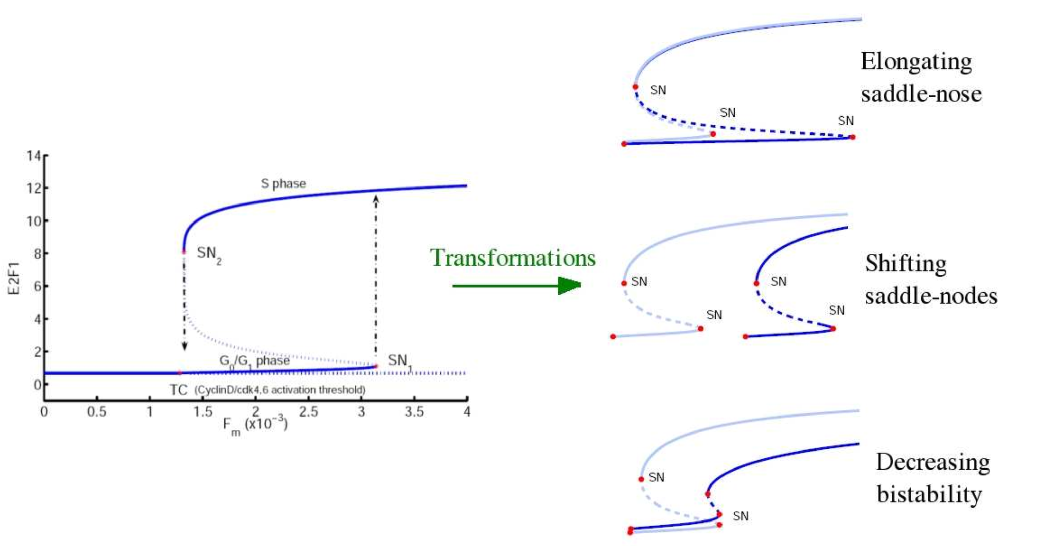
Algorithm convergence and identified reactions

Inferring core regulary mechnism in mammalian G_1/S module

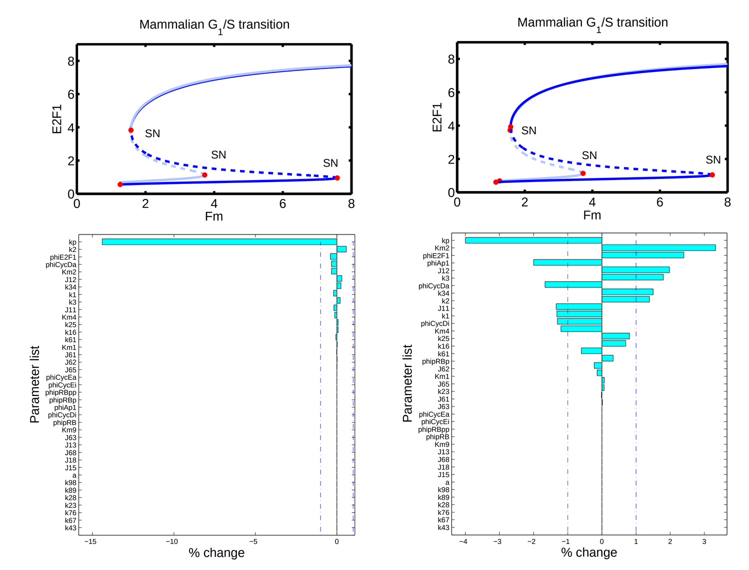
The model of the mammalian G_1/S transition by Swat *et al.* [4] consists of 9 chemical species, 25 reactions and 40 parameters, representing the transcription factor families AP-1, E2F, pRB, and cyclin/cyclin-dependent kinase complexes cyclin D/Cdk4,6 and cyclin E/Cdk2

Question: what can the regulation of bifurcation points be attributed to?

We consider mapping 3 modes of geometric variations to the parameter space:



Sparsity is important in identifying influential parameters; see following figure for an illustration.

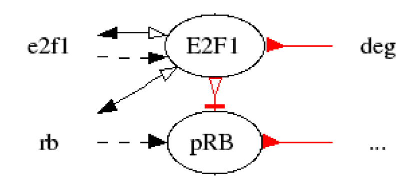


l_p vs. l_2 regularization

IBA: via hierarchical strategy, the following has been identified as core reactions, The double activator-inhibitor pair,

$$\frac{d}{dt}[\text{pRB}] = k_1 \frac{[\text{E2F1}]}{K_{m1} + [\text{E2F1}]} \frac{J_{11}}{J_{11} + [\text{pRB}]} \frac{J_{61}}{J_{61} + [\text{pRB}_p]} - k_{16}[\text{pRB}][\text{CycD}_a] + k_{61}[\text{pRB}_p] - \phi_{\text{pRB}}[\text{pRB}],$$

$$\frac{d}{dt}[\text{E2F1}] = k_p + k_2 \frac{a^2 + [\text{E2F1}]^2}{K_{m2}^2 + [\text{E2F1}]^2} \frac{J_{12}}{J_{12} + [\text{pRB}]} \frac{J_{62}}{J_{62} + [\text{pRB}_p]} - \phi_{\text{E2F1}}[\text{E2F1}]$$



as well as feedback in Cyclin D pair,

$$\frac{d}{dt}[\text{CycD}_i] = -k_{34}[\text{CycD}_i]\frac{[\text{CycD}_a]}{K_{m4} + [\text{CycD}_a]} + \cdots$$

$$\frac{d}{dt}[\text{CycD}_a] = k_{34}[\text{CycD}_i]\frac{[\text{CycD}_a]}{K_{m4} + [\text{CycD}_a]} + \cdots.$$

References

[1] H.W. Engl, M. Hanke, A. Neubauer *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.

[2] J. Lu, H. W. Engl, R. Machné, and P. Schuster, *Inverse bifurcation analysis of a model for the mammalian G_1/S regulatory module*, Proceedings of Bioinformatics in Research and Development '07, Lecture Notes in Bioinformatics, Springer-Verlag, 2007.

[3] J. Lu, H. W. Engl, and P. Schuster, *Inverse bifurcation analysis: application to simple gene systems*, Algor. Mole. Biol., 1/11 (2006).

[4] M. Swat, A. Kel, H. Herzel, *Bifurcation analysis of the regulatory modules of the mammalian G_1/S transition*, Bioinformatics, 20 (2004), pp. 1506–1511.