A Symbolic Computation Pipeline to Compile Mathematical Elementary Functions into Chemical Reaction Networks

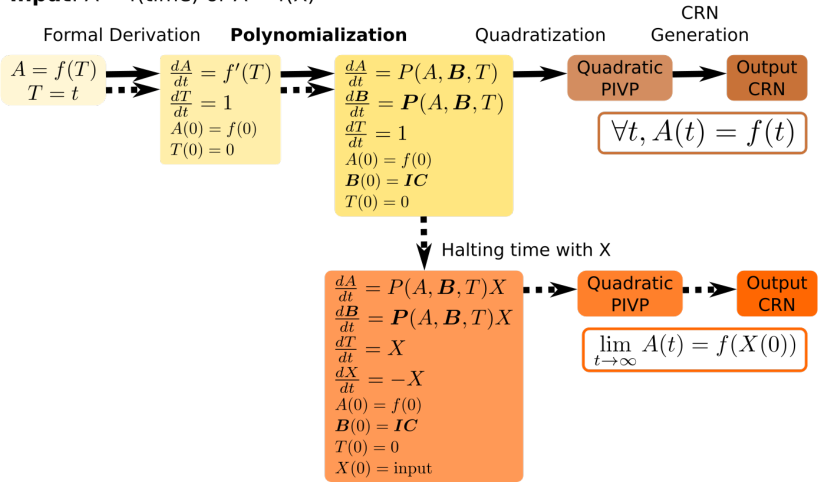
François Fages
joint work with Mathieu Hemery and Sylvain Soliman

Inria Saclay Ile de France,
EPI Lifeware: *Computational systems biology and optimization*

CMSB 2021, SIAM AG 2021, CASC 2021

Compilation Pipeline of Mathematical Functions into CRNs

Input: A = f(time) or A = f(X)



Theorem. (F, Le Guludec, Bournez, Pouly, CMSB 2017)
Turing-completeness of finite continuous CRN

Example for a Function of Time

Input: $A(t) = \log(1+t^2)$

ODE: $\frac{dA}{dt} = \frac{2t}{1+t^2} \qquad A(0) = 0$

Introduce $B = \frac{1}{1+t^2}$: $\frac{dB}{dt} = \frac{-2t}{(1+t^2)^2} = -2tB^2 \qquad B(0) = 1$

PODE: $\frac{dA}{dt} = 2.T.B \qquad \frac{dB}{dt} = -2.T.B^2 \qquad \frac{dT}{dt} = 1$

$A(0) = T(0) = 0, \; B(0) = 1$

Introduce $BT = B.T$ and removing T:

Quadratic ODE: $\frac{dA}{dt} = 2.BT \qquad \frac{dB}{dt} = -2.BT.B \qquad A(0) = 0$

$\frac{d(BT)}{dt} = \frac{dB}{dt}.T + B\frac{dT}{t} = -2.BT^2 + B \qquad B(0) = 1, BT(0) = 0$

Output abstract CRN: $A(0) = BT(0) = 0, \; B(0) = 1$

$BT \xrightarrow{2} A + BT \quad B + BT \xrightarrow{2} BT \quad B \xrightarrow{1} B + BT \quad 2.BT \xrightarrow{2} BT$

Example for a Function of Time

Input: $A(t) = \log(1+t^2)$

ODE: $\frac{dA}{dt} = \frac{2t}{1+t^2} \qquad A(0) = 0$

Introduce $B = \frac{1}{1+t^2}$: $\frac{dB}{dt} = \frac{-2t}{(1+t^2)^2} = -2tB^2 \qquad B(0) = 1$

PODE: $\frac{dA}{dt} = 2.T.B \qquad \frac{dB}{dt} = -2.T.B^2 \qquad \frac{dT}{dt} = 1$

$$A(0) = T(0) = 0,\ B(0) = 1$$

Introduce $BT = B.T$ and removing T:

Quadratic ODE: $\frac{dA}{dt} = 2.BT \qquad \frac{dB}{dt} = -2.BT.B \qquad A(0) = 0$

$$\frac{d(BT)}{dt} = \frac{dB}{dt}.T + B\frac{dT}{t} = -2.BT^2 + B \qquad B(0) = 1, BT(0) = 0$$

Output abstract CRN: $A(0) = BT(0) = 0,\ B(0) = 1$

$$BT \xrightarrow{2} A + BT \quad B + BT \xrightarrow{2} BT \quad B \xrightarrow{1} B + BT \quad 2.BT \xrightarrow{2} BT$$

Example for a Function of Time

Input: $A(t) = \log(1+t^2)$

ODE: $\frac{dA}{dt} = \frac{2t}{1+t^2} \qquad A(0) = 0$

Introduce $B = \frac{1}{1+t^2}$: $\frac{dB}{dt} = \frac{-2t}{(1+t^2)^2} = -2tB^2 \qquad B(0) = 1$

PODE: $\frac{dA}{dt} = 2.T.B \qquad \frac{dB}{dt} = -2.T.B^2 \qquad \frac{dT}{dt} = 1$

$A(0) = T(0) = 0,\ B(0) = 1$

Introduce $BT = B.T$ and removing T:

Quadratic ODE: $\frac{dA}{dt} = 2.BT \qquad \frac{dB}{dt} = -2.BT.B \qquad A(0) = 0$

$\frac{d(BT)}{dt} = \frac{dB}{dt}.T + B\frac{dT}{t} = -2.BT^2 + B \qquad B(0) = 1, BT(0) = 0$

Output abstract CRN: $A(0) = BT(0) = 0,\ B(0) = 1$

$BT \xrightarrow{2} A + BT \quad B + BT \xrightarrow{2} BT \quad B \xrightarrow{1} B + BT \quad 2.BT \xrightarrow{2} BT$

Example for a Function of Time

Input: $A(t) = \log(1+t^2)$

ODE: $\frac{dA}{dt} = \frac{2t}{1+t^2} \qquad A(0) = 0$

Introduce $B = \frac{1}{1+t^2}$: $\frac{dB}{dt} = \frac{-2t}{(1+t^2)^2} = -2tB^2 \qquad B(0) = 1$

PODE: $\frac{dA}{dt} = 2.T.B \qquad \frac{dB}{dt} = -2.T.B^2 \qquad \frac{dT}{dt} = 1$

$A(0) = T(0) = 0,\ B(0) = 1$

Introduce $BT = B.T$ and removing T:

Quadratic ODE: $\frac{dA}{dt} = 2.BT \qquad \frac{dB}{dt} = -2.BT.B \qquad A(0) = 0$

$\frac{d(BT)}{dt} = \frac{dB}{dt}.T + B\frac{dT}{t} = -2.BT^2 + B \qquad B(0) = 1, BT(0) = 0$

Output abstract CRN: $A(0) = BT(0) = 0,\ B(0) = 1$

$BT \xrightarrow{2} A + BT \quad B + BT \xrightarrow{2} BT \quad B \xrightarrow{1} B + BT \quad 2.BT \xrightarrow{2} BT$

Example for a Function of some Input Molecular Species

Input: $A(x) = \log(1 + x^2)$

ODE: $\frac{dA}{dt} = X.\frac{2t}{1+t^2} \quad \frac{dX}{dt} = -X \quad X(0) = x,\ A(0) = 0$

PODE: $\frac{dX}{dt} = -X \quad \frac{dA}{dt} = 2.T.B.X \quad \frac{dB}{dt} = -2.T.B^2.X \quad \frac{dT}{dt} = X$
$X(0) = x,\ A(0) = 0,\ B(0) = 1,\ T(0) = 0$

Quadratic ODE: by introducing BX and TBX variables
and by removing T and B

Output abstract CRN:
$X(0) = BX(0) = x \quad TBX(0) = A(0) = 0$

$$X \xrightarrow{1} \emptyset \qquad BX + X \xrightarrow{1} BX + X + TBX$$

$$BX \xrightarrow{1} \emptyset \qquad BX + TBX \xrightarrow{2} TBX$$

$$TBX \xrightarrow{1} \emptyset \qquad 2.TBX \xrightarrow{2} TBX \qquad TBX \xrightarrow{2} A + TBX$$

Example for a Function of some Input Molecular Species

Input: $A(x) = \log(1 + x^2)$

ODE: $\frac{dA}{dt} = X.\frac{2t}{1+t^2} \qquad \frac{dX}{dt} = -X \qquad X(0) = x,\ A(0) = 0$

PODE: $\frac{dX}{dt} = -X \quad \frac{dA}{dt} = 2.T.B.X \quad \frac{dB}{dt} = -2.T.B^2.X \quad \frac{dT}{dt} = X$

$X(0) = x,\ A(0) = 0,\ B(0) = 1,\ T(0) = 0$

Quadratic ODE: by introducing BX and TBX variables

and by removing T and B

Output abstract CRN:

$X(0) = BX(0) = x \quad TBX(0) = A(0) = 0$

$$X \xrightarrow{1} \emptyset \qquad BX + X \xrightarrow{1} BX + X + TBX$$

$$BX \xrightarrow{1} \emptyset \qquad BX + TBX \xrightarrow{2} TBX$$

$$TBX \xrightarrow{1} \emptyset \qquad 2.TBX \xrightarrow{2} TBX \qquad TBX \xrightarrow{2} A + TBX$$

Example for a Function of some Input Molecular Species

Input: $A(x) = \log(1 + x^2)$

ODE: $\frac{dA}{dt} = X.\frac{2t}{1+t^2} \qquad \frac{dX}{dt} = -X \qquad X(0) = x,\ A(0) = 0$

PODE: $\frac{dX}{dt} = -X \quad \frac{dA}{dt} = 2.T.B.X \quad \frac{dB}{dt} = -2.T.B^2.X \quad \frac{dT}{dt} = X$
$X(0) = x,\ A(0) = 0,\ B(0) = 1,\ T(0) = 0$

Quadratic ODE: by introducing BX and TBX variables
and by removing T and B

Output abstract CRN:
$X(0) = BX(0) = x \quad TBX(0) = A(0) = 0$

$$X \xrightarrow{1} \emptyset \qquad BX + X \xrightarrow{1} BX + X + TBX$$

$$BX \xrightarrow{1} \emptyset \qquad BX + TBX \xrightarrow{2} TBX$$

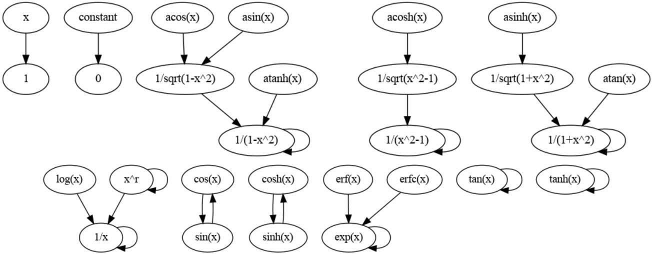
$$TBX \xrightarrow{1} \emptyset \qquad 2.TBX \xrightarrow{2} TBX \qquad TBX \xrightarrow{2} A + TBX$$

Polynomialization Algorithm (Hemery F Soliman CMSB 2021)

Lemma Terminates for any finite set F of formally differentiable functions over the reals whenever their derivatives belong to the algebra of F over $\mathbb{R}$.

Proposition Terminates on elementary functions over the reals, with at most a linear number of introduced variables and quadratic time complexity for linear size derivatives.

Proof $1/x$ for negative powers and derivatives' dependency graph:



Folklore Quadratization Theorem

Theorem. Any function generated by a PIVP (polynomial initial value pb) can be generated by a PIVP of degree at most two.

Quadratization algorithm in $O(d^n)$ [Carothers et al. 2005]

Input: PIVP with n variables $\{x_1, \ldots, x_n\}$ max powers $d_1, \ldots, d_n$

• Introduce $v_{i_1,\ldots,i_n} = x_1^{i_1} x_2^{i_2}, \ldots, x_n^{i_n}$ for all i_j, $0 \leq i_j \leq d_j$, $1 \leq j \leq n$ satisfying $i_k > 0$ for some k

• Their derivatives can be expressed with monomials of degree at most 2

Output: quadratic PIVP with same output function on original variables $v_{1,0,\ldots,0}(t), \ldots, v_{0,\ldots,0,1}(t)$.

Quadratization Minimization Problem:

Quadratic Transformation Problem using Carothers Monomials

QTP input: A PIVP on n variables $X = \{x_i\}_{0 \leq i \leq n-1}$ with a distinguished output variable x_0.

QTP output: the minimum number k of variables for Carothers monomials $f_j(X)$ such that $\{x_0, f_1(X), \ldots, f_k(X)\}$ defines an equivalent quadratic PIVP for computing output x_0.

Associated decision problem

QTDP input: QTP input with given number k

QTDP output: existence of a quadratic form using k variables

Inria informatics mathematics

Complexity of the Quadratization Problem

QTDP input: succinct symbolic representation of PIVP

nsQTDP input: non-succinct representation of PIVP in matrix form containing all Carothers monomials

Proposition. nsQTDP $\in$ NP. QTDP $\in$ NEXP.

Theorem. (Hemery F Soliman CMSB 2020)
nsQTDP is NP-complete. nsQTP is NP-hard.

Proof. By reduction of the vertex set covering problem.

In symbolic succinct representation:

Conjecture. The QTDP is NEXP-complete. QTP is NEXP-hard.

Implementation in BIOCHAM and Evaluation

Quadratization minimizes the number of CRN variables (NP-hard)

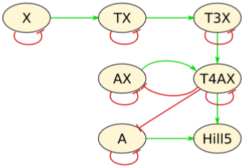
- MAXSAT solver (RC2) used either directly (`sat_species`) or
- after heuristics to restrict the subset of variables (`fastnSAT`)

	fastnSAT			sat_species		
Function	time (ms)	number of species	number of reactions	time (ms)	number of species	number of reactions
Hill1	80	4	5	85	3	3
Hill2	90	6	10	82	5	8
Hill3	100	6	10	115	6	12
Hill4	100	7	13	162	7	13
Hill5	110	8	16	550	7	11
Hill10	160	13	31	timeout		
Hill20	380	23	61	timeout		
Logistic	80	3	5	85	3	5
Double exp.	80	3	4	85	3	4
Gaussian	85	3	4	85	3	4
Logit	95	4	7	100	4	6

Hill5 as a Synthetic Analog of the MAPK Signalling CRN

Natural MAPK CRN: 22 species, 30 reactions, Hill5 [Huang Ferrel 96]
(real enzymes, reverse reactions)

Synthetic Hill5 CRN: 7 species, 11 reactions
(abstract species, no reverse reactions)



$$x \xrightarrow{1} \emptyset \qquad Ax \xrightarrow{1} \emptyset$$

$$Tx \xrightarrow{1} \emptyset \qquad T3x \xrightarrow{1} \emptyset$$

$$T4Ax \xrightarrow{1} \emptyset \qquad A + T4Ax \xrightarrow{5} T4Ax + hill5$$

$$Ax + T4Ax \xrightarrow{5} T4Ax \qquad 2.x \xrightarrow{1} Tx + 2.x$$

$$2.Tx \xrightarrow{3} T3x + 2.Tx \qquad Ax + T3x \xrightarrow{4} Ax + T3x + T4Ax$$

$$2.T4Ax \xrightarrow{5} T4Ax \qquad x(0) = \text{input}$$

$$A(0) = 1 \qquad Ax(0) = \text{input}$$

Summary and Open Problems

- Pipeline for compiling any elementary function in a finite CRN

$O(n^2)$ polynomialization algorithm [Hemery F Soliman CMSB 2021]

$O(d^n)$ quadratization no species minimization [Carothers et al. 2005]

Carothers-minimal quadratization pb [Hemery F Soliman CMSB 2020]:

– NP-complete in non-succinct (matricial) repr.
– conjectured NEXP-complete in succinct (symbolic) repr.
– MAXSAT algorithm minimizing Carothers' monomials
– Heuristics to restrict the set of Carothers monomials

- Optimal monomial quadratization [Bychkov Pogudin IWOCA 2021]

– Branch&bound algorithm
– Carothers' monomials lead to suboptimal solutions

- Better solution with non-monomial quadratization [Alauddin 2021]

– no algorithm