Satisfying Flux Balance and Mass-Action Kinetics in a Network of Biochemical Reactions

Speakers	Michael Saunders and Ines Thiele
Senior Personnel	Ronan Fleming, Bernhard Palsson, Yinyu Ye
Collaborators	Santiago Akle, Onkar Dalal, Joshua Lerman, Yuekai Sun, Nicole Taheri

ICME	Stanford University
Systems Biology Research Group	University of California, San Diego
Center for Systems Biology	University of Iceland

Genomic Sciences Annual PI Meeting
Hyatt Regency Crystal City, VA
April 10–13, 2011

Abstract

Computational modeling of biochemical reaction networks will become increasingly dependent on large-scale numerical optimization as the network dimensionality continues to rise. Linear optimization, the mainstay for flux balance analysis, is trivial for well scaled metabolic models, but nontrivial for integrated metabolic and macromolecular synthetic networks because of the size of the stoichiometric matrix and the wide temporal scales involved. We describe the first such reconstruction for E. coli, which accounts for 43% of all its genes, and a similar reconstruction for T. maritima as well as numerical challenges associated with flux balance analysis.

Beyond linear optimization, we have established that a sequence of parametric convex optimizations can be used to model the inherently nonlinear kinetic relationship between reaction flux and metabolite concentration. Convex optimization preserves many of the attractive numerical features of linear optimization algorithms. For example, given a model of a biochemical network, one can establish a priori whether a solution will exist. The non-existence of a solution is an indication of a malformed model arising from incorrect stoichiometry, reaction directionality, etc.

Abstract, contd

With a sequence of convex optimization problems, establishing the existence of a fixed point for such a sequence is equally important for indicating when a model is malformed. We summarize recent results establishing necessary conditions for existence of a flux and concentration at such a fixed point. The approach also leads to a numerically tractable algorithm for simultaneous satisfaction of flux balance and mass-action kinetics for a well scaled stoichiometric model, subject to certain conditions.

The overall objective of this work is to provide the systems biology community with a new wave of computationally efficient algorithms for physico-chemically realistic modeling of genome-scale biochemical networks. We highlight open source algorithms already released.

Flux Balance Analysis

Leads to linear programming problems involving the stoichiometric matrix S (Palsson 2006):

$$\begin{aligned} \underset{v_f, v_r, v_e}{\text{minimize}} \quad & d^T v_e \\ \text{subject to} \quad & Sv_f - Sv_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \quad \ell \leq v_e \leq u \end{aligned}$$

Flux Balance Analysis

Leads to linear programming problems involving the stoichiometric matrix S (Palsson 2006):

$$\begin{aligned} \underset{v_f, v_r, v_e}{\text{minimize}} \quad & d^T v_e \\ \text{subject to} \quad & Sv_f - Sv_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \quad \ell \leq v_e \leq u \end{aligned}$$

Treat as a challenging large-scale linear program:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & d^T x \\ \text{subject to} \quad & Ax = b, \quad \ell \leq x \leq u \end{aligned}$$

Badly scaled stoichiometric constraints

When modeling biological systems, we often encounter reactions with large coefficients:

$$A + 10^6 B \to C + D$$

We replace these reactions with a sequence of reactions

$$\begin{aligned} A + 100\hat{B}_2 &\to C + D \\ 100\hat{B}_1 &\to \hat{B}_2 \\ 100B &\to \hat{B}_1 \end{aligned}$$

Badly scaled coupling constraints

Recently, reaction coupling has been used to represent dependencies between synthesis and utilization [Thiele et al., Biophys. J., 98(10), 2072 (2010)]:

$$\alpha \leq \frac{v_1}{v_2} \leq \beta$$

Badly scaled coupling constraints

Recently, reaction coupling has been used to represent dependencies between synthesis and utilization [Thiele et al., Biophys. J., 98(10), 2072 (2010)]:

$$\alpha \leq \frac{v_1}{v_2} \leq \beta$$

- These are linear constraints:

$$\alpha v_2 - v_1 \leq 0 \qquad v_1 - \beta v_2 \leq 0$$

Badly scaled coupling constraints

Recently, reaction coupling has been used to represent dependencies between synthesis and utilization [Thiele et al., Biophys. J., 98(10), 2072 (2010)]:

$$\alpha \leq \frac{v_1}{v_2} \leq \beta$$

- These are linear constraints:

$$\alpha v_2 - v_1 \leq 0 \qquad v_1 - \beta v_2 \leq 0$$

- We replace $v_1 - 10^6 v_2 \leq 0$ by a sequence of constraints:

$$v_1 - 100 s_1 \leq 0$$
$$s_1 - 100 s_2 \leq 0$$
$$s_2 - 100 v_2 \leq 0$$

Example data

$$Ax = b$$

Organism	Problem	Dimensions
T. maritima	ME_2011	18000×17600
E. coli	ME_2009	21000×25000
E. coli	ME_2010	68000×77000

CPLEX simplex and barrier LP solvers applied with Feasibility and Optimality tolerances 10^{-9}

Experimental results

CPLEX simplex solver

	original		with rebuild	
Problem	$\|b - Ax\|_1$	Time	$\|b - Ax\|_1$	Time
TM 2011	2.6×10^{-8}	0.9	4.6×10^{-9}	1.2
EC 2009	4.6×10^{-7}	1.1	1.7×10^{-7}	2.1
EC 2010	1.3×10^{-4}	242.8	2.9×10^{-6}	292.3

Experimental results

CPLEX simplex solver

	original		with rebuild	
Problem	$\|b - Ax\|_1$	Time	$\|b - Ax\|_1$	Time
TM 2011	2.6×10^{-8}	0.9	4.6×10^{-9}	1.2
EC 2009	4.6×10^{-7}	1.1	1.7×10^{-7}	2.1
EC 2010	1.3×10^{-4}	242.8	2.9×10^{-6}	292.3

CPLEX barrier solver

	original		with rebuild	
Problem	$\|b - Ax\|_1$	Time	$\|b - Ax\|_1$	Time
TM 2011	7.9×10^{-9}	7.9	3.15×10^{-9}	9.8
EC 2009	3.0×10^{-5}	32.0	4.75×10^{-6}	14.0
EC 2010	1.4×10^{-1}	384.0	3.42×10^{-6}	93.3

Implications for Flux Balance Analysis

- Our procedure addresses the numerical difficulties associated with badly scaled linear systems that arise in systems biology

Implications for Flux Balance Analysis

- Our procedure addresses the numerical difficulties associated with badly scaled linear systems that arise in systems biology
- Users can more accurately model biological systems by enforcing ill-scaled constraints

Implications for Flux Balance Analysis

- Our procedure addresses the numerical difficulties associated with badly scaled linear systems that arise in systems biology
- Users can more accurately model biological systems by enforcing ill-scaled constraints
- Remember to switch off the CPLEX Presolve

Implications for Flux Balance Analysis

- Our procedure addresses the numerical difficulties associated with badly scaled linear systems that arise in systems biology
- Users can more accurately model biological systems by enforcing ill-scaled constraints
- Remember to switch off the CPLEX Presolve
- A further card up sleeve: 128-bit f90 version of SQOPT

Implications for Flux Balance Analysis

- Our procedure addresses the numerical difficulties associated with badly scaled linear systems that arise in systems biology
- Users can more accurately model biological systems by enforcing ill-scaled constraints
- Remember to switch off the CPLEX Presolve
- A further card up sleeve: 128-bit f90 version of SQOPT
- Avoiding large numbers in S is essential for rank estimation (e.g. to evaluate conserved moieties correctly)

Entropy problem with parameter p

$$\begin{aligned} \underset{v_f, v_r > 0}{\text{minimize}} \quad & v_f^T(\log v_f + p - e) + v_r^T(\log v_r + p - e) \\ \text{subject to} \quad & Sv_f - Sv_r = b \qquad : y \end{aligned}$$

Entropy problem with parameter p

$$\begin{array}{ll} \underset{v_f, v_r > 0}{\text{minimize}} & v_f^T(\log v_f + p - e) + v_r^T(\log v_r + p - e) \\ \text{subject to} & Sv_f - Sv_r = b \qquad : y \end{array}$$

Optimality condition

$$S^T y \;\propto\; \log(v_r \,./\, v_f)$$

Entropy problem with parameter p

$$\begin{aligned} \underset{v_f, v_r > 0}{\text{minimize}} \quad & v_f^T(\log v_f + p - e) + v_r^T(\log v_r + p - e) \\ \text{subject to} \quad & Sv_f - Sv_r = b \qquad : y \end{aligned}$$

Optimality condition

$$S^T y \;\propto\; \log(v_r \,./\, v_f)$$

Optimal $v_f,\ v_r$ are thermodynamically feasible

First numerically scalable method (polynomial complexity)

[Fleming et al., A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks, 2011]

Homogeneous network (stepping stone to $b \neq 0$)

$$\begin{bmatrix} S & -S \end{bmatrix} \begin{bmatrix} v_f \\ v_r \end{bmatrix} = 0 \qquad S \equiv R - F$$

Homogeneous network (stepping stone to $b \neq 0$)

$$\begin{bmatrix} S & -S \end{bmatrix} \begin{bmatrix} v_f \\ v_r \end{bmatrix} = 0 \qquad S \equiv R - F$$

$$(R - F)v_f = (R - F)v_r \qquad v_f = \mathsf{diag}\,(k_f) \begin{bmatrix} \prod_i c_i^{F_{i1}} \\ \vdots \\ \prod_i c_i^{F_{in}} \end{bmatrix}$$

$$v_r = \dots$$

Homogeneous network (stepping stone to $b \neq 0$)

$$\begin{bmatrix} S & -S \end{bmatrix} \begin{bmatrix} v_f \\ v_r \end{bmatrix} = 0 \qquad S \equiv R - F$$

$$(R-F)v_f = (R-F)v_r \qquad v_f = \mathsf{diag}\,(k_f) \begin{bmatrix} \prod_i c_i^{F_{i1}} \\ \vdots \\ \prod_i c_i^{F_{in}} \end{bmatrix}$$

$$v_r = \dots$$

$$Fv_r + Rv_f = Fv_f + Rv_r \qquad v(c) = \begin{bmatrix} \prod_i c_i^{Y_{i1}} \\ \vdots \\ \prod_i c_i^{Y_{in}} \end{bmatrix}$$

Homogeneous network (stepping stone to $b \neq 0$)

$$\begin{bmatrix} S & -S \end{bmatrix} \begin{bmatrix} v_f \\ v_r \end{bmatrix} = 0 \qquad S \equiv R - F$$

$$(R - F)v_f = (R - F)v_r \qquad v_f = \mathsf{diag}\,(k_f) \begin{bmatrix} \prod_i c_i^{F_{i1}} \\ \vdots \\ \prod_i c_i^{F_{in}} \end{bmatrix}$$

$$v_r = \ldots$$

$$Fv_r + Rv_f = Fv_f + Rv_r \qquad v(c) = \begin{bmatrix} \prod_i c_i^{Y_{i1}} \\ \vdots \\ \prod_i c_i^{Y_{in}} \end{bmatrix}$$

$$YA^T v(c) = YDv(c) \qquad \log v(c) = Y^T \log c$$

We want $YDv = YA^Tv$ and $Y^T \log c = \log v$

Entropy problem with parameter μ

$$\begin{aligned} \underset{v>0}{\text{minimize}} \quad & v^T D \log v - e^T D v \\ \text{subject to} \quad & YDv = \mu \qquad : y \end{aligned}$$

We want $YDv = YA^Tv$ and $Y^T \log c = \log v$

Entropy problem with parameter μ

$$\begin{aligned} \underset{v>0}{\text{minimize}} \quad & v^T D \log v - e^T D v \\ \text{subject to} \quad & YDv = \mu \qquad : y \end{aligned}$$

Optimality condition

$$DY^T y(\mu) = D \log v(\mu) \quad \Leftrightarrow \quad Y^T y(\mu) = \log v(\mu)$$

We want $YDv = YA^Tv$ and $Y^T \log c = \log v$

Entropy problem with parameter μ

$$\begin{aligned} \underset{v>0}{\text{minimize}} \quad & v^T D \log v - e^T D v \\ \text{subject to} \quad & YDv = \mu \qquad : y \end{aligned}$$

Optimality condition

$$DY^T y(\mu) = D \log v(\mu) \quad \Leftrightarrow \quad Y^T y(\mu) = \log v(\mu)$$

We would like the fixed point condition

$$\mu = YA^T v(\mu)$$

Fixed point iteration

$v \leftarrow v_0 \qquad \mu \leftarrow YA^T v$

while $\|YDv - YA^T v\|_\infty > \tau$ **do**

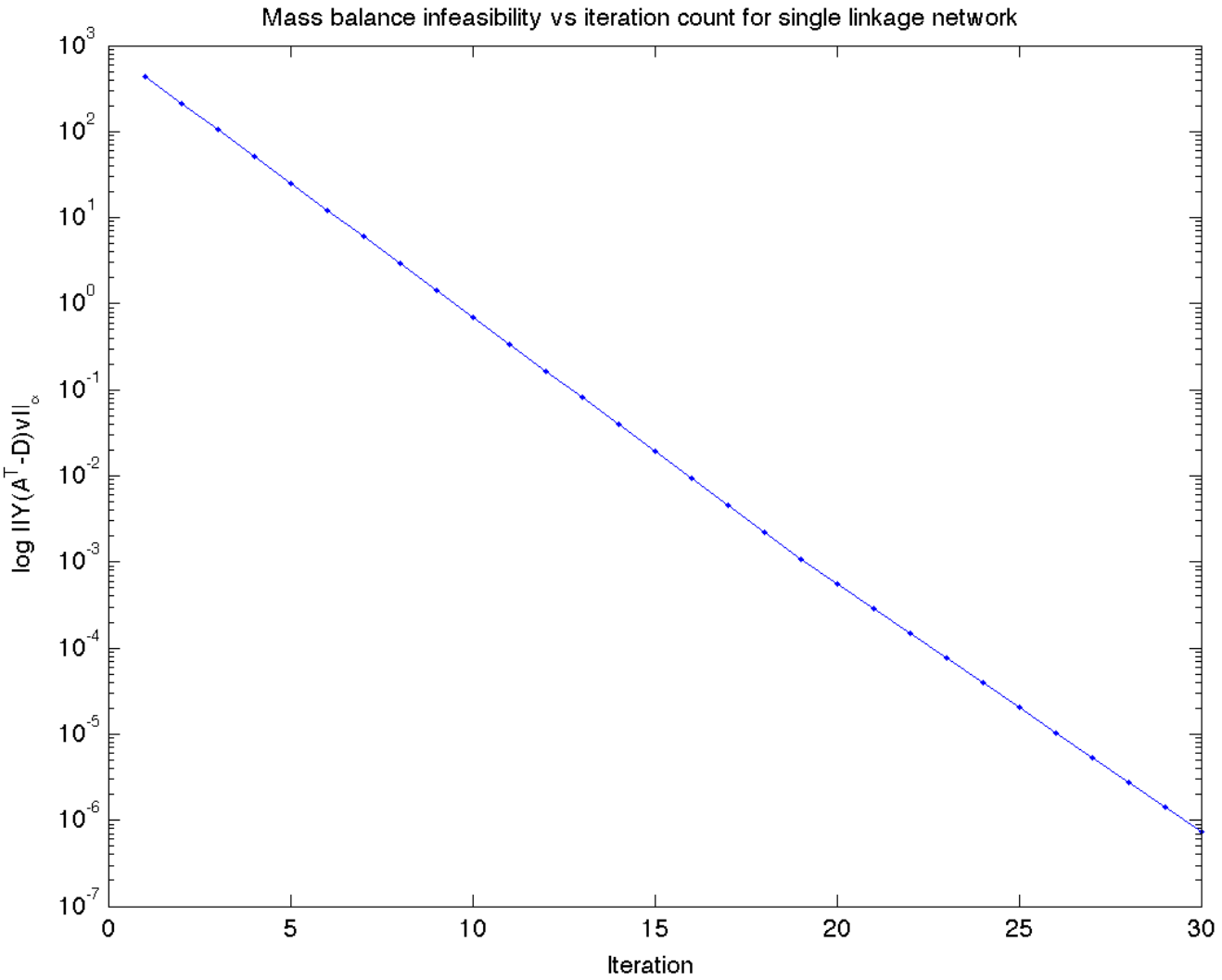
Solve Entropy subproblem

$$\begin{array}{ll} \min & v^T D \log v - e^T D v \\ \text{st} & YDv = \mu \end{array}$$

$\mu \leftarrow (1-\theta)\mu + \theta YA^T v$

end

Fixed point iteration, $\theta = 0.5$



LUSOL: A sparse LU package

Factorize $m \times n$ sparse matrix $S = LDU$
with permutations to preserve stability and sparsity

L and U triangular $\quad L_{ii} = 1$, $\quad D$ diagonal, $\quad U_{ii} = 1$

LUSOL: A sparse LU package

Factorize $m \times n$ sparse matrix $S = LDU$
with permutations to preserve stability and sparsity

L and U triangular $\quad L_{ii} = 1$, $\quad D$ diagonal, $\quad U_{ii} = 1$

Uses: Find rank$(S) < m < n$
Remove redundant rows
Form nullspace basis

LUSOL: A sparse LU package

Factorize $m \times n$ sparse matrix $S = LDU$
with permutations to preserve stability and sparsity

L and U triangular $\quad L_{ii} = 1$, $\quad D$ diagonal, $\quad U_{ii} = 1$

Uses: Find rank$(S) < m < n$
Remove redundant rows
Form nullspace basis

- Threshold partial pivoting keeps L well-conditioned

$$|L_{ij}| \leq \texttt{Ltol} \quad \text{(5.0 or 2.0 say)}$$

LUSOL: A sparse LU package

Factorize $m \times n$ sparse matrix $S = LDU$
with permutations to preserve stability and sparsity

L and U triangular $\quad L_{ii} = 1, \quad$ D diagonal, $\quad U_{ii} = 1$

Uses: Find rank$(S) < m < n$
Remove redundant rows
Form nullspace basis

- Threshold partial pivoting keeps L well-conditioned

$$|L_{ij}| \leq \texttt{Ltol} \quad \text{(5.0 or 2.0 say)}$$

- Threshold rook pivoting controls L and U

$$|L_{ij}| \text{ and } |U_{ij}| \leq \texttt{Ltol} \quad \text{(2.0 or 1.5 say)}$$

Stoichiometric matrix (rank-deficient)

$$S = \begin{bmatrix} 5 & 0 & -10 & 0 & 5 & 0 \\ 0 & 3 & 0 & 6 & -6 & 0 \\ -10 & 0 & 21 & 1 & -10 & 1 \\ 0 & -6 & -1 & -13 & 12 & -1 \end{bmatrix}$$

Stoichiometric matrix (rank-deficient)

$$S = \begin{bmatrix} 5 & 0 & -10 & 0 & 5 & 0 \\ 0 & 3 & 0 & 6 & -6 & 0 \\ -10 & 0 & 21 & 1 & -10 & 1 \\ 0 & -6 & -1 & -13 & 12 & -1 \end{bmatrix}$$

SVD

$S = UDV^T$ $\quad U, V$ orthogonal $\quad D = \begin{bmatrix} 29.5 & & & \\ & 18.8 & & \\ & & 0.88 & \\ & & & \epsilon_1 \end{bmatrix}$

Stoichiometric matrix (rank-deficient)

$$S = \begin{bmatrix} 5 & 0 & -10 & 0 & 5 & 0 \\ 0 & 3 & 0 & 6 & -6 & 0 \\ -10 & 0 & 21 & 1 & -10 & 1 \\ 0 & -6 & -1 & -13 & 12 & -1 \end{bmatrix}$$

SVD

$$S = UDV^T \qquad U, V \text{ orthogonal} \qquad D = \begin{bmatrix} 29.5 & & & \\ & 18.8 & & \\ & & 0.88 & \\ & & & \epsilon_1 \end{bmatrix}$$

LUSOL with threshold rook pivoting `Ltol`= 2.0

$$S = LDU = \begin{bmatrix} 1 & & & \\ & 1 & & \\ -2 & & 1 & \\ & -2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 5 & & & \\ & 3 & & \\ & & 1 & \\ & & & \epsilon_2 \end{bmatrix} \begin{bmatrix} 1 & & -2 & & 1 & \\ & 1 & & 2 & -2 & \\ & & 1 & 1 & & 1 \\ & & & x & x & x \end{bmatrix}$$

LUSOL with rook pivoting and reducing Ltol

Sorted elements of D:

SVD	29.5	18.8	0.9	ϵ
Ltol 2.0	5.0	3.0	1.0	ϵ
1.5	21.0	-13.0	0.2	ϵ
1.1	21.0	-13.0	0.2	ϵ

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$
- PDCO: Matlab
  Barrier method for LP and Entropy problems

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$
- PDCO: Matlab
  Barrier method for LP and Entropy problems
- MINRES, SYMMLQ, LSQR, LSMR: f77, f90, Matlab
  Iterative solvers for $Ax = b$, $\min \|Ax - b\|$

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$
- PDCO: Matlab
  Barrier method for LP and Entropy problems
- MINRES, SYMMLQ, LSQR, LSMR: f77, f90, Matlab
  Iterative solvers for $Ax = b$, $\min \|Ax - b\|$
- SQOPT, SNOPT: f77, f90 coming
  General-purpose solvers for LP, QP, NLP

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$
- PDCO: Matlab
  Barrier method for LP and Entropy problems
- MINRES, SYMMLQ, LSQR, LSMR: f77, f90, Matlab
  Iterative solvers for $Ax = b$, $\min \|Ax - b\|$
- SQOPT, SNOPT: f77, f90 coming
  General-purpose solvers for LP, QP, NLP
- Systems Optimization Laboratory
  `http://www.stanford.edu/group/SOL/software.html`

Available software

- LUSOL: Fortran 77, C, f90 coming
  New Matlab mex interface (Nick Henderson, ICME, Stanford)
- Nullspace functions:
  Compute products $p = Zv$ or $q = Z^T y$, where $S^T Z = 0$
- PDCO: Matlab
  Barrier method for LP and Entropy problems
- MINRES, SYMMLQ, LSQR, LSMR: f77, f90, Matlab
  Iterative solvers for $Ax = b$, $\min \|Ax - b\|$
- SQOPT, SNOPT: f77, f90 coming
  General-purpose solvers for LP, QP, NLP
- Systems Optimization Laboratory
  `http://www.stanford.edu/group/SOL/software.html`
- and more (COBRA Toolbox and extensions)