Elementary Flux Modes

State-of-the-art Implementation and Scope of Application

Marco Terzer and Jörg Stelling

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Computational Science, ETH Zurich, 8092 Zurich, Switzerland

Contact: {marco.terzer,joerg.stelling}@inf.ethz.ch

Introduction & Motivation

Motivation

- Computational analysis of cellular metabolic networks
- Modeling/analysis methods: structural versus dynamic
- Structural/constraint-based approaches: rely on stoichiometry only, no kinetic parameters needed

Constraints

$\boldsymbol{r}$: flux distribution, one flux value for every reaction
$\boldsymbol{N}$: stoichiometric matrix, network invariant

- *Thermodynamic constraints:* some reactions are irreversible, they have a non-negative flux value $\rightarrow \boldsymbol{r}_{irrev} \geq \boldsymbol{0}$
- *Quasi steady state:* concentrations of (internal) metabolites assumed to be constant $\rightarrow \boldsymbol{N} \cdot \boldsymbol{r} = \boldsymbol{0}$

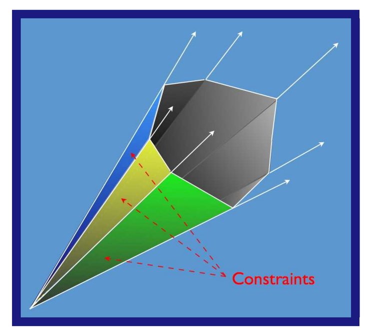
$\rightarrow$ Flux distributions are restricted by constraints

Pathway Analysis

- Solution space of flux distributions: *described* by constraints
- Elementary flux modes (EFMs):
  - Set of flux vectors *spanning* the solution space
  - Unique and minimal (non-decomposable)
- Extreme pathways (EPs): concept closely related to EFMs

Mathematics

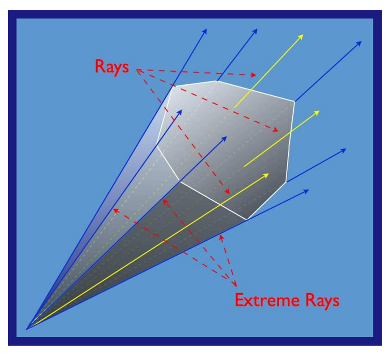
Polyhedral Cone



- Each constraint defines a halfspace through the origin
- Solution space: intersection of halfspaces, a *polyhedral cone* P
- Formally: $P = \{\boldsymbol{x} \,|\, \boldsymbol{A}\,\boldsymbol{x} \geq \boldsymbol{0}\}$

Extreme Rays

- From Minkowski's Theorem: an alternative, *constructive* definition for polyhedral cones exists [1]
  $\rightarrow P$: all non-negative linear combinations of *extreme rays*
- Formally: $P = \{\boldsymbol{x} \,|\, \boldsymbol{x} = \boldsymbol{R}\,\boldsymbol{c} \text{ for some } \boldsymbol{c} \geq \boldsymbol{0}\}$



Metabolic Context

- Flux distributions correspond to rays
- Elementary flux modes correspond to extreme rays (columns in $\boldsymbol{R}$)
- Constraints: rows in $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{N} \\ -\boldsymbol{N} \\ \boldsymbol{I} \end{bmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \end{matrix}\right\} \boldsymbol{N} \cdot \boldsymbol{r} = \boldsymbol{0} \\ \left.\right\} \quad \boldsymbol{r} \geq \boldsymbol{0} \end{matrix}$$

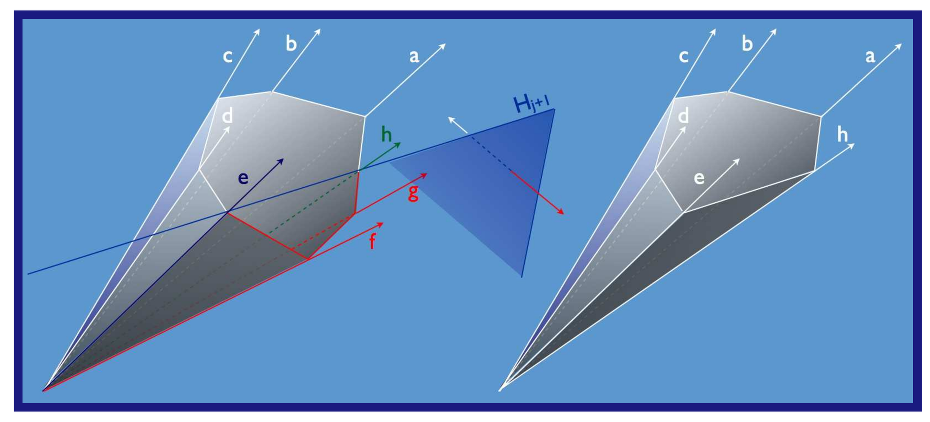
Algorithm

Input
Representation matrix $\boldsymbol{A}$
$P = \{\boldsymbol{x} \,|\, \boldsymbol{A}\,\boldsymbol{x} \geq \boldsymbol{0}\}$

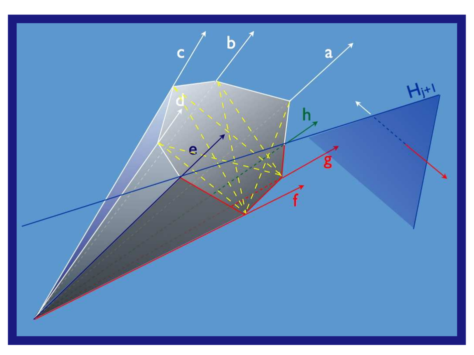
Output
EFMs as columns in $\boldsymbol{R}$
$P = \{\boldsymbol{x} \,|\, \boldsymbol{x} = \boldsymbol{R}\,\boldsymbol{c} \text{ for some } \boldsymbol{c} \geq \boldsymbol{0}\}$

Outline

1. *Initialization Step*
   Choose submatrix $\boldsymbol{A}_d$ of $\boldsymbol{A}$ with corresponding $\boldsymbol{R}_d$, defining a polyhedral cone P_d, which encloses P
   $\rightarrow d$ inequalities are already considered
2. *Iteration Step*
   Construct P_{j+1} from P_j
   $\rightarrow$ choose next inequality $j+1$, intersect hyperplane with P_j
   $\rightarrow$ continue until all inequalities (constraints) are considered



Elementarity & Adjacency



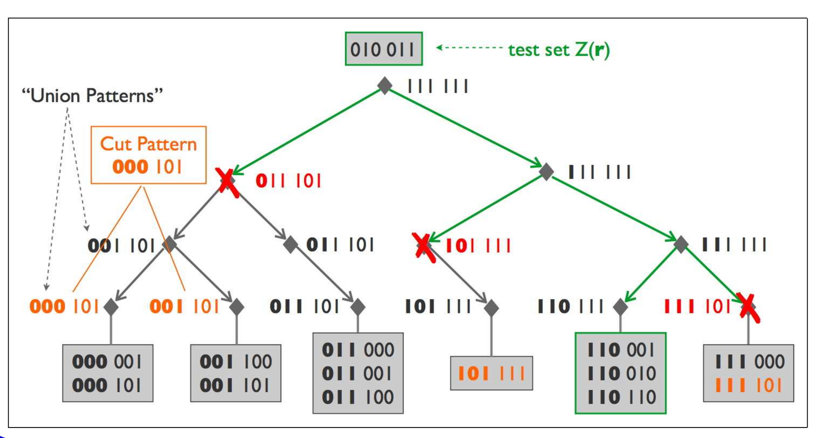
- Two rays from opposing sides of the new hyperplane are combined
  $\rightarrow$ only adjacent extreme rays generate new extreme rays
- Let $Z(\boldsymbol{r})$ be the zero set of a ray $\boldsymbol{r}$, the reactions not occurring in $\boldsymbol{r}$. For a new ray $\boldsymbol{r}$ created from $\boldsymbol{r}'$ and $\boldsymbol{r}''$: $Z(\boldsymbol{r}) = Z(\boldsymbol{r}') \cap Z(\boldsymbol{r}'')$.
  Two tests are possible:
  1. *Combinatorial:* test the new ray $\boldsymbol{r}$ against already accepted ones.
     $\boldsymbol{r}$ is elementary $\iff$ no $\boldsymbol{r}' \neq \boldsymbol{r}$ exists with $Z(\boldsymbol{r}) \subset Z(\boldsymbol{r}')$
  2. *Algebraic:* the rank of a submatrix $\boldsymbol{A}_{Z(\boldsymbol{r})}$ determines elementarity

Basic Approach (with combinatorial test)

$$\left.\begin{matrix} \text{i) All candidates: } O(n^2) \\ \text{ii) Adjacency test: } O(n) \end{matrix}\right\} O(n^3) \text{ per iteration}$$

Indexed Combinatorial Test

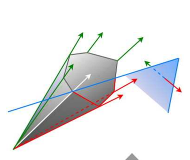
Indexed search with k-d-tree like structure
$\rightarrow$ *pattern trees*, see [6]



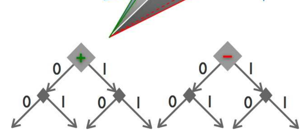
Candidate Narrowing

- Two *pattern trees* for kept / removed extreme rays
- Candidates: combine each element beneath green root with each element beneath red root
  - Recurse with children, 4 combinations
  - On recursing: one test covering whole subtree
  - *Idea:* intersect union patterns of both nodes

$\rightarrow$ Intersection of union $\equiv$ unifying all intersections $\equiv$ union of test sets $\Rightarrow$ if union of them fails, all fail



Dual-Core Processors



- Use as many threads as cores
- Before recursing, test if a free thread is available

Scope of Application

Comparing Constraint-based Approaches

Approach	*Constraints:* *Quasi steady state*	*Thermo-dynamics*	*Optimality*	*Applications:* *Functional pathways*	*Optimal operation*	*Reaction importance*	*Reaction correlation*	*Network function*	*Robustness*	*Computational Costs*	*Solutions*
Kernel	√	–	–	–	–	–	(√)	(√)	–	low	all
FBA	√	√	√	–	√	(√)	–	√	(√)	low	single
MoMA	√	√	√	–	√	(√)	–	√	(√)	medium	single
EFMs/EPs	√	√	–	√	√	√	√	√	√	high	all

From: S. Klamt and J. Stelling, Stoichiometric and constraint-based modeling. In *System Modeling in Cellular Biology* [4]

EFM Specialities

- Natural decomposition of complex metabolic networks into elementary units
- Semi-quantitative prediction of gene expression patterns with multi-objective optimization: e.g. flexibility (robustness) versus efficiency
- Detection of *all* qualitatively different (optimal) flux vectors for a given optimization function

Drawbacks & Limits

- Combinatorial explosion, vast number of EFMs for large networks
- To date, genome-scale networks not computable
- FBA & MoMA better suited and much more efficient for finding particular solutions. For instance, FBA provides similarly good predictions of mutant phenotypes

Results

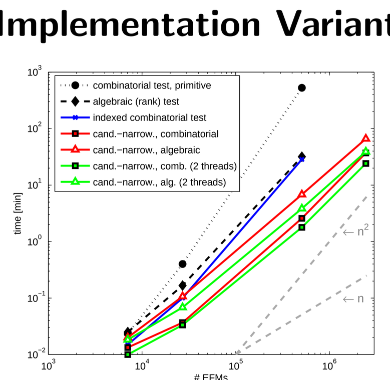
Network

- Variants of a stoichiometric model for the central metabolism of *Escherichia coli* [5]

Implementation Issues

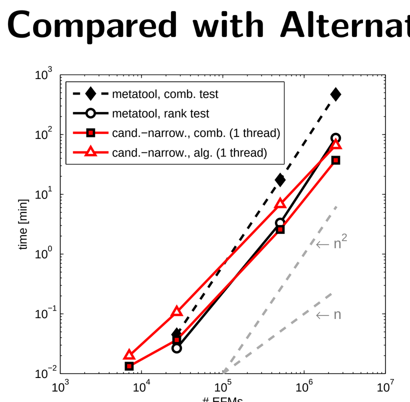
- The processing order of the inequalities has high impact on the computation time. Empirical tests propose lexicographical ordering [1]. Variants of such orderings have been applied
- Network compression techniques mostly identical to those presented in [2] have been used
- From the algebraic elementarity test, an upper bound for the number of reactions participating in an EFM can be derived. This prerequisite is always checked before doing the real test. For candidate narrowing, applying only the prerequisite test before recursing yielded best performance
- Only binary representations of modes were calculated, see [2]
- Implemented in Java, tested on a Linux machine with an Intel Dual-Core processor 6400 at 2.13 GHz, using a Java 5 virtual machine with max. 2 GB memory

Implementation Variants



- Significant speedup w. candidate narrowing
- Combinatorial test still fastest, but better scalability with algebraic
- Dual-Core: speedup of 1/2—3/4 with 2 threads

Compared with Alternative Implementations



- Better scalability with candidate narrowing
- Large networks: both test versions faster, even with Java vs. C
- Metatool 5.0 benchmarks: see [3], Table 1

Conclusions

Conclusions

- Significant performance improvements with *pattern trees*, mainly by *candidate narrowing*
- Candidate narrowing is applicable to both adjacency tests, the *combinatorial* test is currently still somewhat faster, but the *algebraic* test features better scalability
- Candidate narrowing is well suited for parallelization, which has been shown with a multi-threaded algorithm version on a common Dual-Core processor. Extension to multiple processors straight forward
- The current implementation competes well with alternative implementations and is to the best of our knowledge the fastest EFM algorithm today
- Pathway analysis represents a powerful tool for the structural analysis of metabolic networks. However, combinatorial complexity constricts its application to medium scale networks. FBA and MoMA might be suitable, even though less comprehensive alternatives

References

[1] K. Fukuda and A. Prodon. Double description method revisited. In *Combinatorics and Computer Science*, pages 91–111, 1995.

[2] J. Gagneur and S. Klamt. Computation of elementary modes: A unifying framework and the new binary approach. *BMC Bioinformatics*, 5:175, 2004.

[3] S. Klamt, J. Gagneur, and A. von Kamp. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *IEE Proc. Systems Biol.*, 152:249–55, 2005.

[4] S. Klamt and J. Stelling. Stoichiometric and constraint-based modeling. In *System Modeling in Cellular Biology*, pages 73–96. MIT Press (Cambridge / MA), 2006.

[5] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, 2002.

[6] Marco Terzer and Jörg Stelling. Accelerating the computation of elementary modes using pattern trees. In *WABI*, volume 4175 of *Lecture Notes in Computer Science*, pages 333–343. Springer, 2006.