Finding Relevant Elementary Modes in the Flux Cone

Benjamin Pomeranz[1,2], Bree Cummins[1]

[1] MSU Dept. of Mathematical Sciences [2] Brown University

Motivation

Metabolic models can be used to predict and understand the behavior of small species such E. coli or human blood cells. Elementary flux modes (EFMs) are fundamental in interpreting results from metabolic models. Unfortunately, the enumeration of the complete set of EFMs is computationally prohibitive in genome scale models.

Metabolic Modelling and Elementary Modes

A metabolic network is represented mathematically by stoichiometric matrix $S \in \mathbb{R}^{|\text{metabolites}| \times |\text{reactions}|}$, where

$$S_{i,j} = \text{quantity of metabolite } i \text{ consumed or produced by reaction } j$$

We then represent the state of a metabolism with a flux vector $\mathbf{v}$, where

$$v_j = \text{the flux through reaction } j$$

If we assume that the metabolism is in steady state and that all reactions are irreversible and thus have positive flux, we get the *feasible flux cone*

$$\mathcal{F} = \{\mathbf{v}; S\mathbf{v} = 0, \mathbf{v} \geq 0\}$$

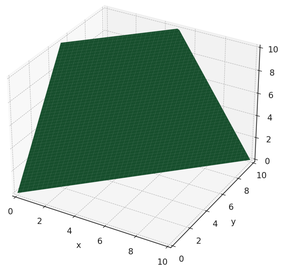


Figure 1: Flux cone generated by EFMs $(1, 1, 0)$ and $(0, 1, 2)$

Then, the EFMs of a stoichiometric can be equivalently thought of as

- The minimal metabolic pathways which are feasible and steady state
- The *extreme rays* which generate the feasible flux cone

An algorithm for finding EFMs

In 2011, Jungers et al. published an algorithm for decomposing a flux vector $\mathbf{v}$ into a minimal set of EFMs:

1. By adding constraint $\sum_i v_i = a$ to $\mathcal{F}$, we get a bounded polytope $\mathcal{P}$.
2. We maximize some random direction $\mathbf{d}$ in $\mathcal{P}$ using an LP, which always returns an EFM $\mathbf{e_1}$.
3. We maximize distance from $\mathbf{e_1}$ along the line $\overleftrightarrow{\mathbf{e_1 v}}$ using another LP. We land on a lower dimensional face $\mathcal{P}_1$, and repeat.

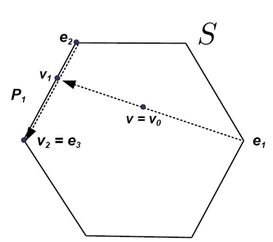


Figure 2: An illustration from Jungers et al. '11 of their algorithm.

Methods

We modified the original algorithm (henceforth OA) to be faster and more numerically stable, as well as objective driven.

- In step 2, instead of randomly selecting an objective $\mathbf{d}$ we let $\mathbf{d} = \mathbf{v}_i + \lambda \mathbf{O}$. $\mathbf{O}$ is an objective vector representing a linear combination of reactions to maximize (e.g. biomass production) and λ is a weighting factor.
- We replace step 3 with a simpler, deterministic calculation which we prove yields the same face of the flux cone.

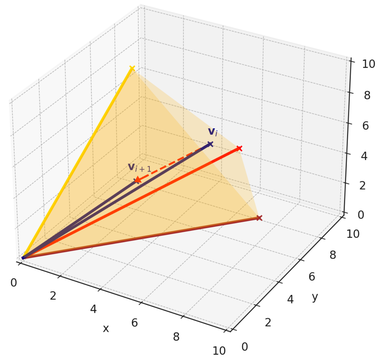


Figure 3: To go from $\mathbf{v}_i$ to $\mathbf{v}_{i+1}$ without an LP, we subtract off the maximal amount of e (the orange edge) that maintains $\mathbf{v}_{i+1} \geq 0$.

Data Collection

We ran the algorithm on 3 models: Cobra's E. coli Core and full E. coli models as well as the mutualistic model of Desulfovibrio vulgaris and Methanococcus maripaludis from Stolyar et al. '07.

Model Name	Reactions	Metabolites	Enumerable?
E. coli core	72	95	Yes
E. coli full	1805	2583	No
Mutualism of D.v. and M	163	188	Yes

For the smaller models, we blocked several exchange reactions to simulate a resource limited medium. This allowed us to compute the complete set of EFMs using *efmtool*. We uniformly sampled flux vectors to decompose using Cobra.

Given a flux vector $\mathbf{v}$, an EFM $\mathbf{e}$, and an objective $\mathbf{O}$, one can calculate the *score* of $\mathbf{e}$, which is the amount of the objective achieved by $\mathbf{v}$ which can be explained by $\mathbf{e}$ in an EFM decomposition of $\mathbf{v}$:

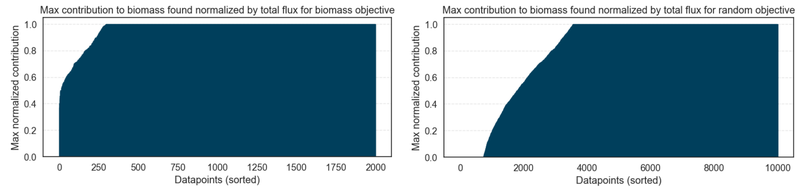
$$s(\mathbf{e}) = \frac{(\mathbf{e} \cdot \mathbf{O})(\min_i \frac{\mathbf{v}_i}{e_i})}{\mathbf{v} \cdot \mathbf{O}}$$

We then computed the first k EFMs in the decomposition, where $k \approx 20$. Then, we compared the score of the best EFM we found in our decomposition with the best scoring EFM in the completely enumerated set of all EFMs.

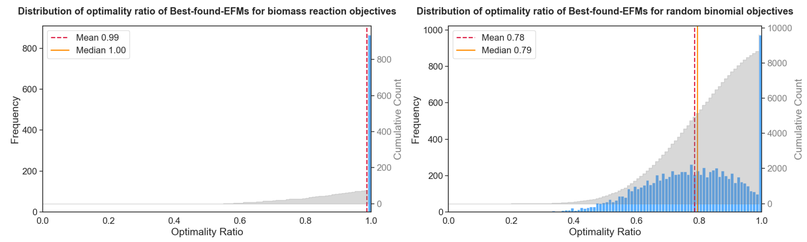
	random	obj_name	obj_weight	flux_vector	runtime	first_efms_scores	max_efm_score	first_coeffs	model
544	False	random_binomial_8	1.0	[-568.6408652606524, -43.83386123304157, -4.52...	0.209146	[130.5389207060518, 325.0794740549095, 0.0, 24...	601.843109	[0.03868404005948981, 0.06572817042878, 0.05...	textbook
3202	False	random_singleton_9	1000.0	[-179.08148323264074, -74.10550824107285, -46.0...	0.209205	[0.7176664706551411, 0.0, 0.0, 0.0, 0.0, 0.0, ...	0.717666	[0.018047913944818462, 0.060087009894202872, 0....	textbook
3334	False	random_binomial_5	1000.0	[-329.36308117393972, -104.07784733388516, -8.76...	0.226986	[437.0990674445581, 158.93635495833118, 647.74...	1078.561141	[0.08019771402344866, 0.008990092101660413, 0....	textbook
4050	True	random_singleton_8	NaN	[-464.44840417354131, -319.8097546637514, -175....	0.213018	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...	303.512192	[0.06237357229813094, 0.011781967480568948, 0....	textbook
5817	False	Biomass	1.0	[-358.2442076374065, -180.97403877435784, -49....	0.227721	[0.0, 0.0, 0.011188550850509698, 0.0, 0.0, 0.0,...	0.011189	[0.12190778925703694, 0.043179092605541744, 0....	textbook

Figure 4: Example data from collection the E. coli core model

Results and Ongoing Work



When we run the algorithm in E. coli core with objective based search for biomass, we are able to find an EFM which completely accounts for the biomass synthesis of the model in an EFM decomposition 73.1 % of the time, as opposed to 53.6% in the random algorithm. Moreover, none of the objective based decompositions had maximal contribution less than .5, while 17.95% of the random decompositions did not reach this threshold.



The optimality ratio is the score of the best EFM found by the algorithm, divided by the best possible score among all EFMs. We find that the algorithm performs much worse on binomial objectives (where each reaction has value 1 in the objective with probability $1/10$) than it does on single reaction objectives, and in turn performs far better on maximizing biomass reaction than it does arbitrary reactions.

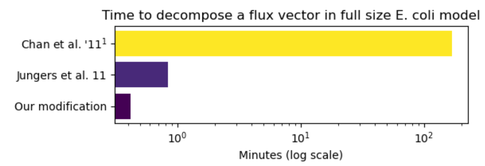


Figure 5: [1] is a conservative estimate based upon the linear relationship they find between runtime and $n_{\text{EFMs}} \times n_{\text{active reactions}}$.

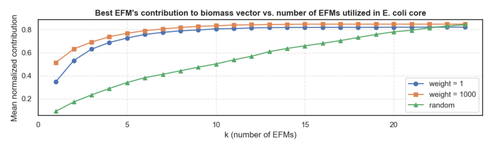


Figure 6: Interestingly, while we find that searching with objective leads to finding a better EFM faster, we find that with a longer random search the algorithm performs approximately as well.

This project remains ongoing. We are measuring performance on FBA based flux vectors rather than random ones and comparing results across models. We also plan to run the algorithm on a large model, where EFMs cannot be completely enumerated, to explain some phenotypes using mode analysis.

Acknowledgments

Thank you to Nate Paul Rissin, Aubrey Kimmel, Tianyu Zhang, Isaac Klapper, and especially Bree Cummins for the conversations and advice which made this project possible. This research is supported by NSF Grant #2325171.

NSF Grant #2325171