

Determining MinSpan vectors of COBRA model

Author: Aarash Bordbar

Affiliation: Sinopla Biosciences, San Diego, CA USA

Reviewer: James T. Yurkovich

INTRODUCTION

In this tutorial, we show how to calculate MinSpan vectors [1] for a COBRA model. COBRA models are predominantly studied under steady-state conditions, thus the null space of the S matrix is of high interest. MinSpan vectors represent the sparsest linear basis of the null space \mathcal{N} while still maintaining the biological and thermodynamic constraints of the COBRA model (Figure 1). The S matrix has dimensions $(m \times n)$ where m is the number of metabolites and n is the number of reactions. The linear basis of the null space \mathcal{N} has dimensions $(n \times r)$ where r is the rank of S . Thus this algorithm calculates $n-r$ vectors that are linearly independent of each other and also are minimal. For further info on MinSpan, its derivation, implementation, and uses, see Bordbar et al. 2014 [1].

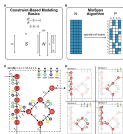


Figure 1 | Overview of the MinSpan algorithm. (A) A metabolic network is mathematically represented as a stoichiometric matrix S . Reaction fluxes (v) are determined assuming steady state. All potential flux states lie in the null space \mathcal{N} . (B) The MinSpan algorithm determines the shortest, independent pathways of the metabolic network by decomposing the null space of the stoichiometric matrix to form the sparsest basis. (C) A simplified model for glycolysis and the TCA cycle is presented with 14 metabolites, 18 reactions, and a 4-dimensional null space. Reversible reactions are shown. (D) The four pathways calculated by MinSpan for the simplified model are presented, two of which recapitulate glycolysis and the TCA cycle, while the other two represent other possible metabolic pathways. The flux direction of a pathway through reversible reactions are shown as irreversible reactions.

MATERIALS

Equipment Setup

Running the MinSpan algorithm requires the installation of a mixed-integer linear programming (MILP) solver. We have used Gurobi v6+ (<http://www.gurobi.com/downloads/download-center>) which is freely available for academic use. This tutorial and the algorithm has been rigorously tested and support Gurobi v6+. `decsMinSpan` will not work with GLPK; other solvers are untested.

Implementation

For MinSpan vectors to be calculated, the model must (1) consist of only reactions that are able to carry flux under that particular condition, (2) allow for the trivial solution ($v = 0$) to be feasible, and (3) have the biomass function removed. `decsMinSpan` will automatically check and complete the first two modifications, but the biomass must be removed manually.

The algorithm is an iterative pruning of null space basis vectors to the sparsest possible matrix. The problem is NP-hard, meaning that an optimal solution is not guaranteed for large COBRA models; an approximate solution is found by setting a time limit on the MILP calculation.

Procedure

In this example, we will calculate the MinSpan vectors for the *E. coli* core network.

Ensure that the Gurobi MILP and LP solvers are working:

```
test1 = checkGurobiSolver('gurobi', 'MILP');
test2 = checkGurobiSolver('gurobi', 'LP');
test3 = checkGurobiSolver('gurobi', 'QP');
if test1 == 0 || test2 == 0 || test3 == 0
    error('minspan only works with gurobi version 6+');
end
```

Load the core model:

```
global CSTR;
modelFileName = 'ecoli_core_model.mat';
modelDirectory = getDistributedModelFolder(modelFileName); % Look up the folder for the distributed Models.
modelPathName = [modelDirectory filesep modelFileName]; % Get the full path. Necessary to be sure, that the right model is loaded
model = readModel(modelPathName);
```

