Numerical linear algebra and optimization tools for bioinformatics

Nick Henderson, Ding Ma, Yuekai Sun, and Michael Saunders
Institute for Computational and Mathematical Engineering, Stanford University

Bioinformatics problems

- **Conservation analysis**
  Find subgroups conserved by biological systems
- **Flux balance analysis (FBA)**
  Find fluxes in multiscale metabolic networks

Numerical software

- **Sparse matrix factorization**
  LUSOL $S = LDU$ with L and possibly U well-conditioned
  SPQR (Davis 2011) $SP = QR$ or $S^TP = QR$
- **Sparse linear equations and least squares**
  MINRES, MINRES-QLP Symmetric $Ax = b$
  LSQR, LSMR, LSRN $Ax = b$, min $\|Ax - b\|$
- **LP**
  MINOS, SQOPT, CPLEX, Gurobi, Mosek, Soplex, Xpress, ...
- **NLP**
  MINOS, SNOPT, PDCO, CONOPT, IPOPT, Knitro, ...
- **Multiscale LP and NLP**
  quad-MINOS

Conservation analysis $S =$ stoichiometric matrix

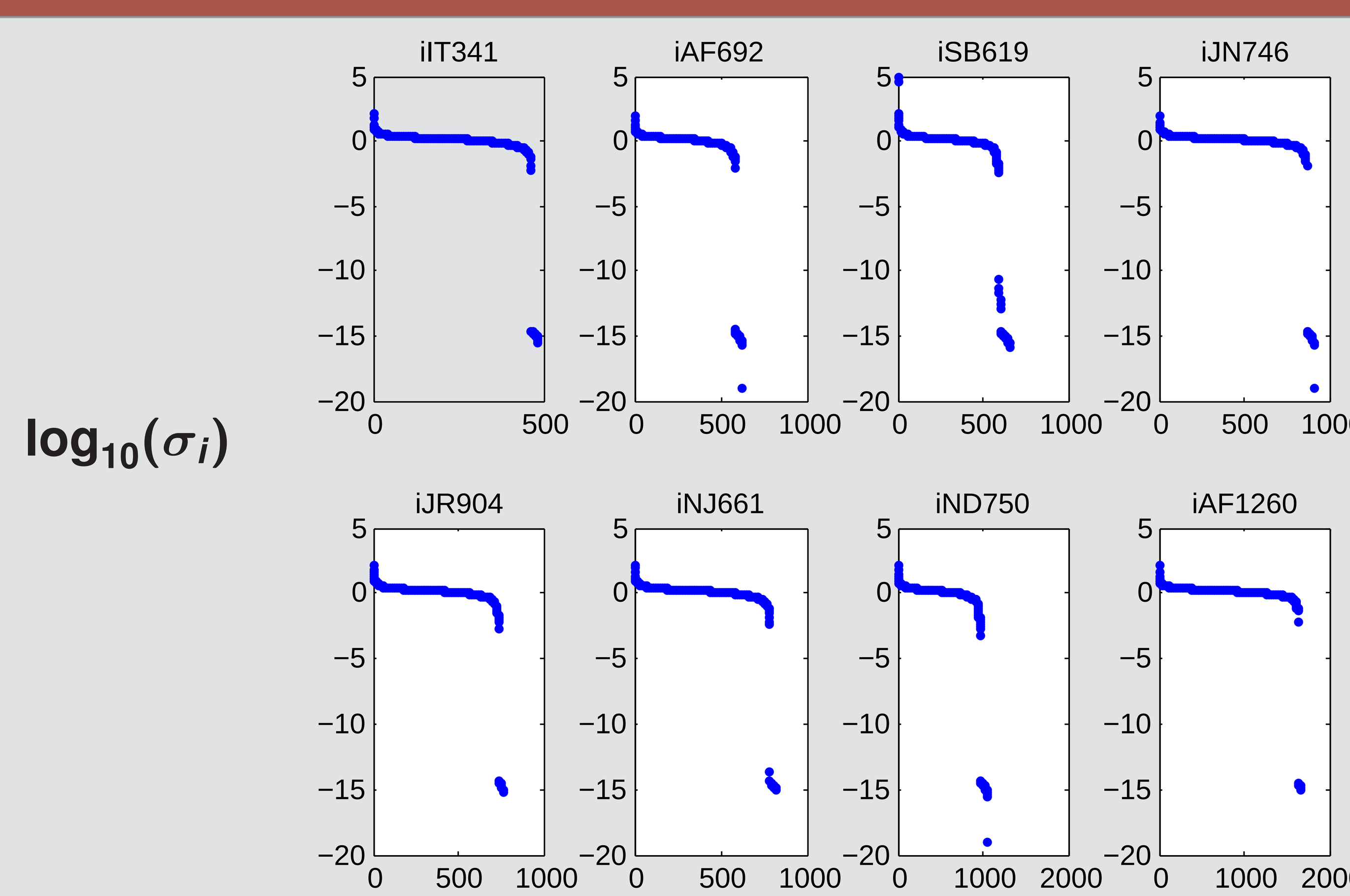
Reduces to finding rank(S) and null(S^T):

$$0 = \frac{d}{dt}\{z^T c(t)\} = z^T \frac{dc(t)}{dt} = z^T S v(t)$$

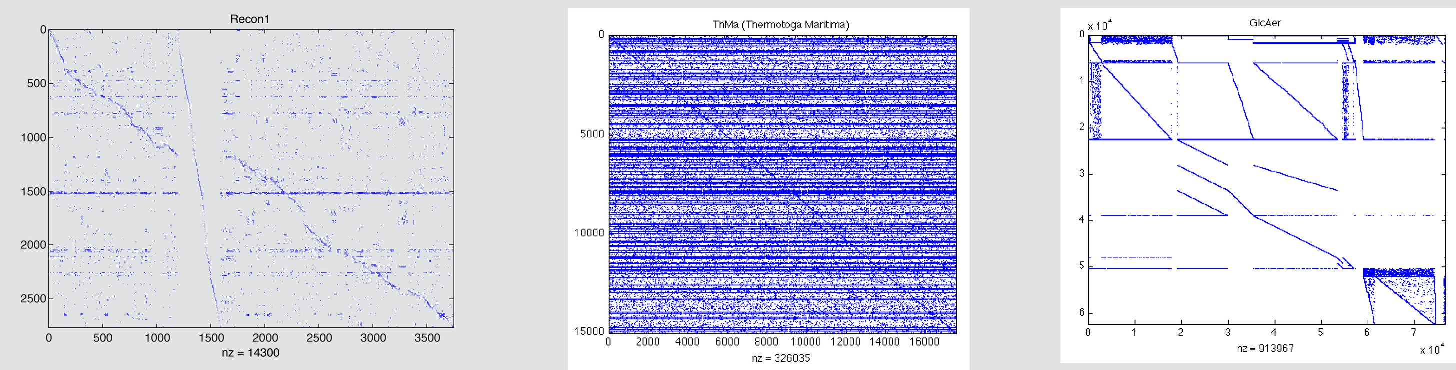
z is a conserved moiety (group of chemical species) Need $S^T z = 0$

- Partition rows of S into dependent and independent species: $PS = \begin{pmatrix} S_{\text{ind}} \\ S_{\text{dep}} \end{pmatrix}$
- Compute *link matrix* N that describes the relations among concentrations of dependent/independent species: $S = NS_{\text{ind}}$

Singular values of models 1–8 Dense SVD of S^T



Models 9, 10, 11



rank(S) by QR SPQR

Householder QR factorization $SP = QR$

- $P =$ col perm $Q^TQ = I$ R diagonal rank(S) = rank(R)
- Nearly as reliable as SVD
- Dense QR Vallabhajosyula, Chickarmane, Sauro (2006)
- Sparse QR (SPQR) now available Davis (2011)
- model 9 (Recon1) 2800 × 3700 0.1 secs
  model 10 (ThMa) 15000 × 18000 2.5 secs
  model 11 (GlcAer) 62000 × 77000 0.2 secs(!)

rank(S) by LDU LUSOL

Sparse LU with Threshold Rook Pivoting $P_1SP_2 = LDU$

- $P_1, P_2 =$ perms D diagonal rank(S) ≈ rank(D)
  L, U well-conditioned
- $L_{ii} = U_{ii} = 1$, $|L_{ij}|$ and $|U_{ij}| \leq 2.0$ (say)
- LUSOL: Main engine in optimizers MINOS, SQOPT, SNOPT
- model 9 (Recon1) 2800 × 3700 0.1 secs
  model 10 (ThMa) 15000 × 18000 4.1 secs
  model 11 (GlcAer) 62000 × 77000 186 secs

SVD, SPQR, LUSOL on stoichiometric S

SVD and sparse QR

model	m	n	rank(S) SVD	rank(S) SPQR	nnz(S)	nnz(Q)	nnz(R)	time SVD	time SPQR
Recon1	2766	3742	2674	2674	14300	2750	21093	17.5	0.1
ThMa	15024	17582	14983	14983	326035	844096	10595016	11hrs	2.5
GlcAer	62212	76664	?	62182	913967	1287	916600	infty	0.2

LUSOL with Threshold Rook Pivoting: $S = LDU$, $|L_{ij}|, |U_{ij}| \leq 2.0$

model	m	n	rank(S)	nnz(S)	nnz(L)	nnz(U)	time
Recon1	2766	3742	2674	14300	4280	16463	0.1
ThMa	15024	17582	14983	326035	30962	346122	4.1
GlcAer	62212	76664	62182	913967	635571	1810491	186.2

LUSOL with Threshold Partial Pivoting: $S = LU$, $|L_{ij}| \leq 2.0$

	rank(S)	nnz(S)	nnz(L)	nnz(U)	time
Recon1	2674	14300	721	13585	0.1
ThMa	14983	326035	7779	324483	0.2
GlcAer	62182	913967	533	913781	0.4

LUSOL with TPP finds rank(S) efficiently, and also S_{ind}, S_{dep}

Multiscale FBA quad-precision MINOS

Step 1: double-MINOS, cold start, scaling

```
Problem name                   GlcAerWT   EXIT -- the problem is infeasible
No. of iterations                 62856   Objective value      -2.4489880182E+04
No. of infeasibilities               41   Sum of infeas         1.5279397622E+01
No. of degenerate steps           33214   Percentage                       52.84
Max x          (scaled)     68680 4.4E+06   Max pi        (scaled)    54979 1.4E+02
Max x                       62607 1.0E+09   Max pi                    25539 3.0E+02
Max Prim inf(scaled)       134382 6.5E+00   Max Dual inf(scaled)      70913 1.2E-05
Max Primal infeas          129844 1.0E+04   Max Dual infeas           23177 2.0E-05
Time for solving problem        9707.28 seconds
```

Step 2: quad-MINOS, warm start, scaling

```
Problem name                   GlcAerWT   EXIT -- optimal solution found
No. of iterations                  5580   Objective value      -7.0382449681E+05
No. of degenerate steps            4072   Percentage                       72.97
Max x          (scaled)     59440 3.7E+00   Max pi        (scaled)    40165 8.1E+11
Max x                       61436 6.3E+07   Max pi                    25539 2.4E+07
Max Prim inf(scaled)        83602 3.8E-16   Max Dual inf(scaled)      11436 4.4E-19
Max Primal infeas           83602 1.7E-07   Max Dual infeas           24941 8.6E-27
Time for solving problem        3995.58 seconds
```

Step 3: quad-MINOS, warm start, no scaling

```
Problem name                   GlcAerWT   EXIT -- optimal solution found
No. of iterations                     4   Objective value      -7.0382449681E+05
No. of degenerate steps               0   Percentage                        0.00
Max x                       61436 6.3E+07   Max pi                    25539 2.4E+07
Max Primal infeas          142960 1.3E-19   Max Dual infeas            6267 9.4E-22
Time for solving problem          60.07 seconds
```

References

- Davis (2011). Algorithm 915, SuiteSparseQR: Multifrontal multithreaded rank-revealing sparse QR factorization, *ACM TOMS* 38(1), 8:1–8:22
- Fleming, Maes, Saunders, Ye, and Palsson (2012)., A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks, *J. Theor. Biol.* 292:71–77
- Gill, Murray, and Saunders (2005).
  SNOPT: An SQP algorithm for large-scale constrained optimization,
  SIAM Review 47(1):99–131 (includes description of LUSOL)
- Henderson (2013). Matlab interface to LUSOL,
  https://github.com/nwh/lusol
- Henderson, Ma, Saunders, Sun, Fleming, and Thiele (2014).
  Conservation analysis of genome-scale biochemical networks,
  Bioinformatics, in preparation
- Ma and Saunders (2014). Solving multiscale linear programs using the simplex method in quadruple precision, Springer, to appear
- Vallabhajosyula, Chickarmane, and Sauro (2006). Conservation analysis of large biochemical networks, *Bioinform.* 22(3), 346–353
- http://web.stanford.edu/group/SOL/multiscale/
  http://web.stanford.edu/group/SOL/software.html

Funding

DOE DE-FG02-09ER25917
NIH U01-GM102098

{nwh,dingma,yuekai,saunders}@stanford.edu