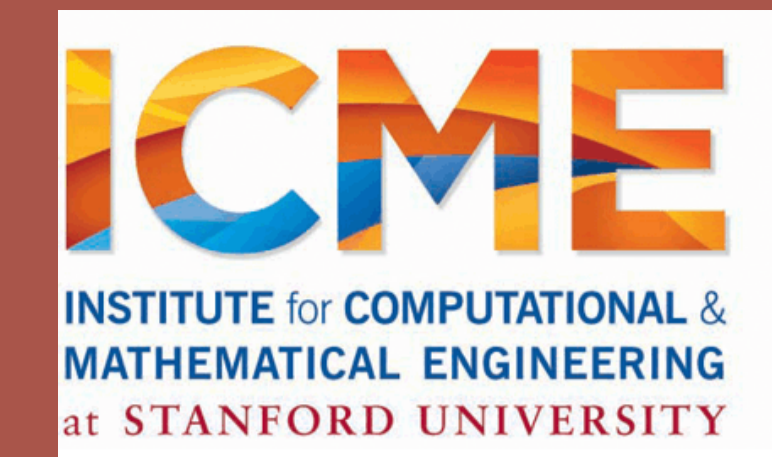# Conservation analysis of genome-scale biochemical networks

Nick Henderson, Ding Ma, Yuekai Sun, and Michael Saunders

Institute for Computational and Mathematical Engineering, Stanford University

ICME
INSTITUTE for COMPUTATIONAL &
MATHEMATICAL ENGINEERING
at STANFORD UNIVERSITY

Conservation analysis

Goal: find subgroups conserved by biological systems

▶ Examples:
  ▶ adenine nucleotide moiety (ADP, ATP, AMP)
  ▶ NAD/NADH
  ▶ CoA/Acetyl-CoA
▶ An important preliminary step in
  ▶ evaluating drug targets
  ▶ analyzing the transient behavior of biochemical networks

Finding rank(S) and null(S^T) $S =$ stoichiometric matrix

**Conservation analysis 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 $S^T z = 0$
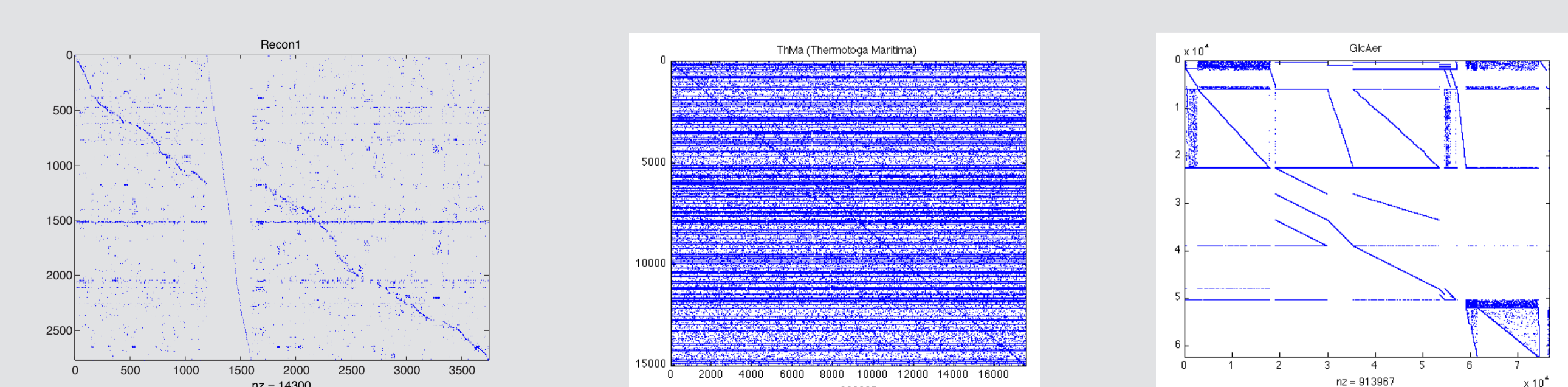
Also part of conservation analysis:
▶ Partitioning the rows (species) of S into
  dependent and independent rows (species)
▶ Computing a link matrix that describes the relations among the
  concentrations of dependent and independent species

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

$\log_{10}(\sigma_i)$

iIT341 iAF692 iSB619 iJN746
iJR904 iNJ661 iND750 iAF1260
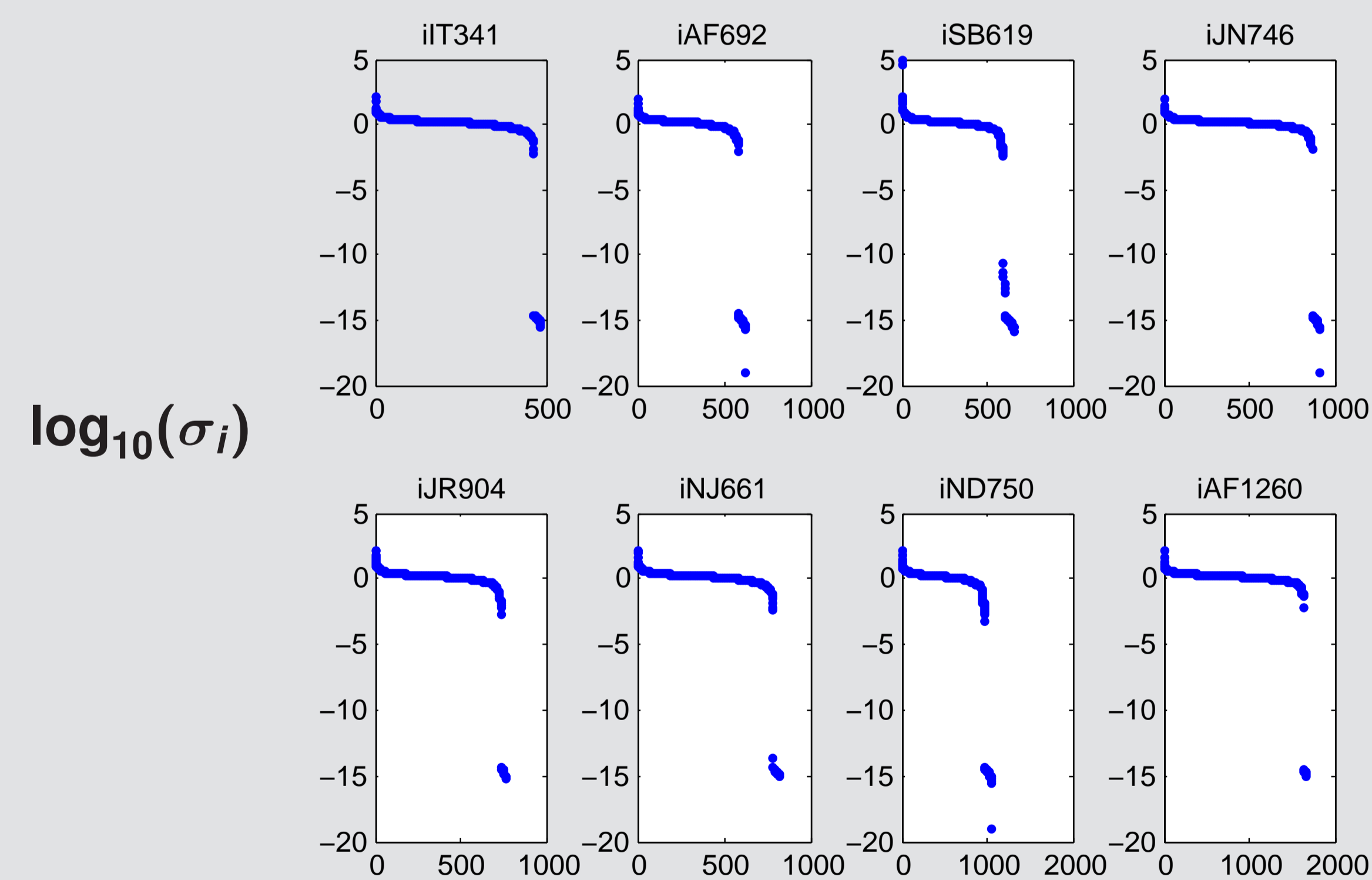
Models 9, 10, 11

rank(S) by SVD   Matlab

Singular value decomposition $S = UDV^T$

▶ $U^T U = I$ $V^T V = I$ D diagonal   rank(S) = rank(D)
▶ Ideal for rank-estimation but U, V are dense
▶ model 9 (Recon1) 2800×3700 17 secs
  model 10 (ThMa) 15000×18000 11 hours
  model 11 (GlcAer) 62000×77000 ∞

rank(S) by QR   SPQR

Householder QR factorization $SP = QR$

▶ $P =$ col perm $Q^T Q = I$ R diagonal   rank(S) = rank(R)
▶ Nearly as reliable as SVD
▶ Dense QR used by Vallabhajosyula, Chickarmane, Sauro (2005)
▶ Sparse QR (SPQR) now available: Davis (2013)
▶ 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 $P_1 S P_2 = LDU$

▶ $P_1, P_2 =$ perms D diagonal   rank(S) $\approx$ rank(D)
  L, U well-conditioned
▶ $L_{ii} = U_{ii} = 1$
  $|L_{ij}|$ and $|U_{ij}| \leq$ `factol` $= 4$ or 2 or 1.2, 1.1, . . .
▶ LUSOL: Main engine in sparse linear/nonlinear optimizers
  MINOS, SQOPT, SNOPT
▶ model 9 (Recon1) 2800×3700 0.1 secs
  model 10 (ThMa) 15000×18000 4.0 secs
  model 11 (GlcAer) 62000×77000 158 secs

References

▶ T. A. Davis (2013). Algorithm 9xx: SuiteSparseQR, a multifrontal multithreaded sparse QR
  factorization package, *ACM TOMS*
▶ P. E. Gill, W. Murray, M. A. Saunders (2005).
  SNOPT: An SQP algorithm for large-scale constrained optimization,
  SIAM Review 47(1):99–131 (includes description of LUSOL)
▶ N. W. Henderson (2013).
  Matlab interface to LUSOL, https://github.com/nwh/lusol/tree/master/matlab

SVD, SPQR, LUSOL on S

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

factol = 2.00   S = LDU

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

factol = 4.00   S = LDU

model	m	n	rank(S)	nnz(S)	nnz(L)	nnz(U)	time
Recon1	2766	3742	2674	14300	2701	12896	0.1
ThMa	15024	17582	14983	326035	36350	330485	4.0
GlcAer	62212	76664	62182	913967	427456	1584188	157.9

SVD, SPQR, LUSOL on S^T

model	m	n	rank(S') SVD	rank(S') SPQR	nnz(S)	nnz(Q)	nnz(R)	time SVD	time SPQR
Recon1	3742	2766	2674	2674	14300	107935	36929	17.2	0.1
ThMa	17582	15024	14983	14983	326035	624640	605888	11hrs	0.7
GlcAer	76664	62212	?	62182	913967	3573696	4038988	infty	2.7

factol = 2.00   S' = LDU

model	m	n	rank(S')	nnz(S)	nnz(L)	nnz(U)	time
Recon1	3742	2766	2674	14300	12832	7421	0.3
ThMa	17582	15024	14983	326035	501198	358601	37.8
GlcAer	76664	62212	62182	913967	1996892	709448	586.0

factol = 4.00   S' = LDU

model	m	n	rank(S')	nnz(S)	nnz(L)	nnz(U)	time
Recon1	3742	2766	2674	14300	9811	6093	0.2
ThMa	17582	15024	14983	326035	410290	355475	14.8
GlcAer	76664	62212	62182	913967	1823067	711906	791.2

Funding