Numerical linear algebra and optimization tools for bioinformatics

Michael Saunders, Santiago Akle, Ding Ma, Yuekai Sun,
Ronan Fleming, and Ines Thiele

SOL and ICME, Stanford University
Luxembourg Centre for Systems Biomedicine, University of Luxembourg

2013 BMES Annual Meeting
BIOINFORMATICS, COMPUTATIONAL & SYSTEMS BIOLOGY
Computational Bioengineering

Seattle, WA, Sep 25–28

Abstract

Computational models often require the solution of large systems of linear equations $Ax = b$ or least-squares problems $Ax \approx b$ or more challenging optimization problems involving large sparse matrices.

For example, the modeling of biochemical reaction networks in systems biology may depend on determining the rank of large stoichiometric matrices, and on accurate solution of large multiscale linear programs, as in Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA). A thermodynamically feasible set of fluxes can be obtained by solving a similar large optimization problem that has a negative entropy objective function.

We describe some general-purpose algorithms and software that have provided efficient and reliable solutions for important problems in systems biology, and are likely to find broader application.

SOL
Systems Optimization Laboratory
Stanford University

SOL

- Founded 1974 by George Dantzig and Richard Cottle
- Dantzig, Alan Manne: economic models (linear & nonlinear)
- Gill, Murray, Saunders, Wright: Software for optimization

SOL

- Founded 1974 by George Dantzig and Richard Cottle
- Dantzig, Alan Manne: economic models (linear & nonlinear)
- Gill, Murray, Saunders, Wright: Software for optimization

Recent collaborators:

- Philip Gill (UC San Diego)
  Optimization software NPSOL, QPOPT, SQOPT, SNOPT
- Ronan Fleming, Ines Thiele (UCSD, Iceland, Luxembourg)
  Flux balance analysis (FBA), Flux variability analysis (FVA)
  Rank and nullspace of stoichiometric matrices
  Nonequilibrium fluxes in metabolic networks
- Bernhard Palsson (UCSD)
  FBA and FVA

SOL

- Founded 1974 by George Dantzig and Richard Cottle
- Dantzig, Alan Manne: economic models (linear & nonlinear)
- Gill, Murray, Saunders, Wright: Software for optimization

Recent collaborators:

- Philip Gill (UC San Diego)
  Optimization software NPSOL, QPOPT, SQOPT, SNOPT
- Ronan Fleming, Ines Thiele (UCSD, Iceland, Luxembourg)
  Flux balance analysis (FBA), Flux variability analysis (FVA)
  Rank and nullspace of stoichiometric matrices
  Nonequilibrium fluxes in metabolic networks
- Bernhard Palsson (UCSD)
  FBA and FVA

Funding: ONR, AFOSR, ARO, DOE, NSF, AHPCRC, . . . ,
DOE DE-FG02-09ER25917, NIH U01-GM102098

Sparse linear equations $Ax = b$ and least squares problems $Ax \approx b$

Problem types and software packages

Iterative methods for $Ax \approx b$

A may be a sparse matrix
or an operator for computing Av and/or $A^T w$

A may have any rank

Iterative methods for $Ax \approx b$

A may be a sparse matrix
or an operator for computing Av and/or $A^T w$

A may have any rank

- Symmetric $Ax = b$ $A = \square$ CG, MINRES-QLP

Iterative methods for $Ax \approx b$

A may be a sparse matrix
or an operator for computing Av and/or $A^T w$

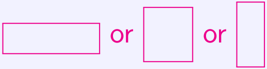
A may have any rank

- Symmetric $Ax = b$ $A =$ □ CG, MINRES-QLP
- min $\|Ax - b\|$ $A =$ ▭ or □ or ▯ LSMR

Iterative methods for $Ax \approx b$

A may be a sparse matrix
or an operator for computing Av and/or $A^T w$

A may have any rank

- Symmetric $Ax = b$ $A =$  CG, MINRES-QLP
- min $\|Ax - b\|$ $A =$  LSMR
- Tall skinny min $\|Ax - b\|$ $A =$  LSRN

Sparse direct methods for $Ax = b$

• $A = LDU$	LUSOL (Stanford)
• $A = QR$	SPQR (Tim Davis, UFL)
• Many sparse solvers	HSL Library (RAL, UK)

Sparse direct methods for $Ax = b$

- $A = LDU$ LUSOL (Stanford)
- $A = QR$ SPQR (Tim Davis, UFL)
- Many sparse solvers HSL Library (RAL, UK)

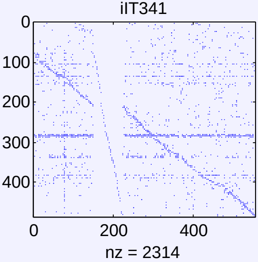
Dense factorizations

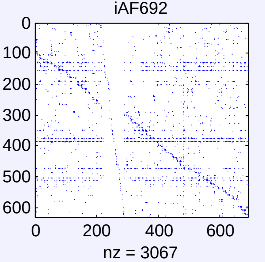
- SVD Golub & Reinsch
- GSVD (2 matrices) Van Loan; Paige & Saunders
- HOGSVD (N matrices) Ponnapalli, Saunders, Van Loan, Alter
  Saturday 1:30pm session
- Parafac etc (tensors) Acar et al., previous talk!

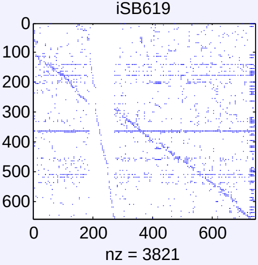
Stoichiometric matrices

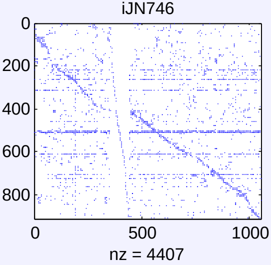
Rows: Chemical species
Cols: Chemical reactions

Models 1, 2, 3, 4 (all similar)

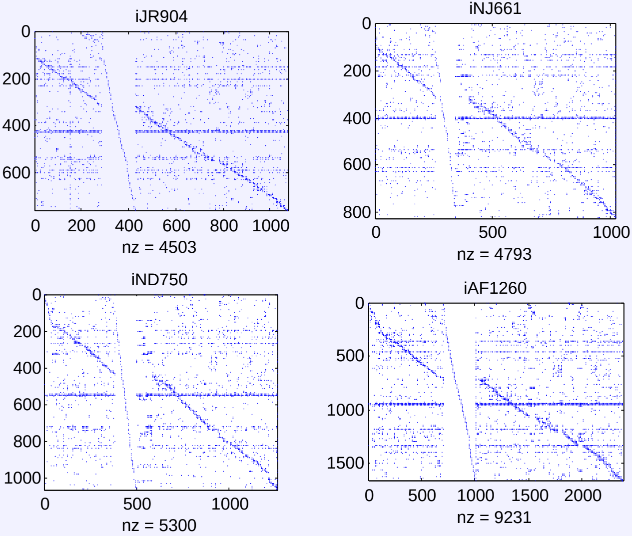




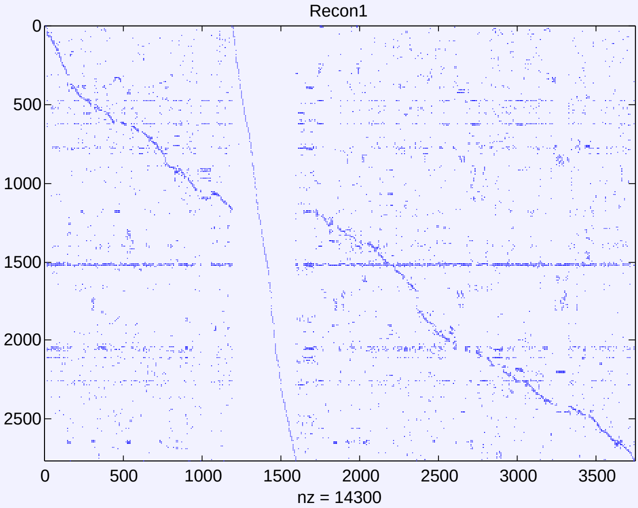




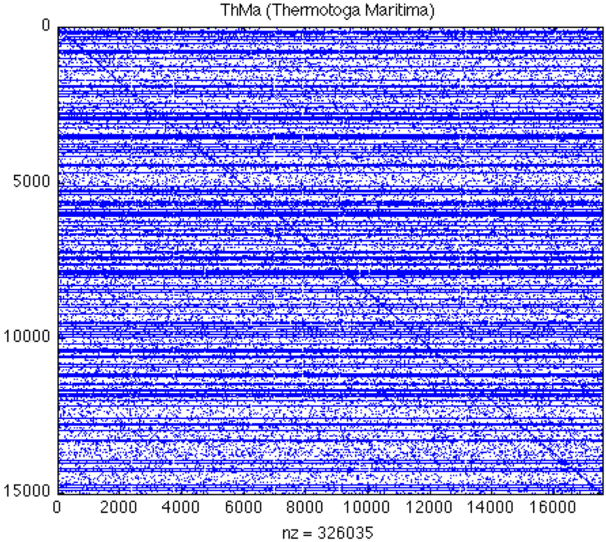
Models 5, 6, 7, 8 (all similar)



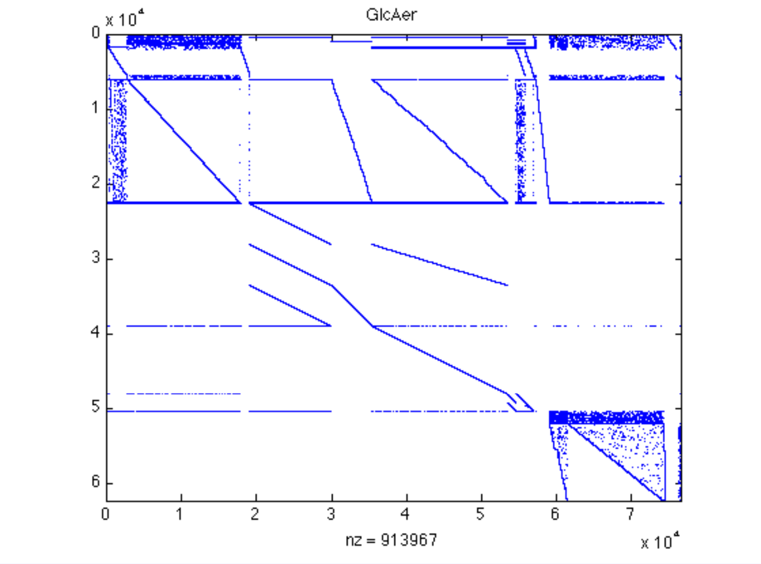
Model 9 (Recon1)



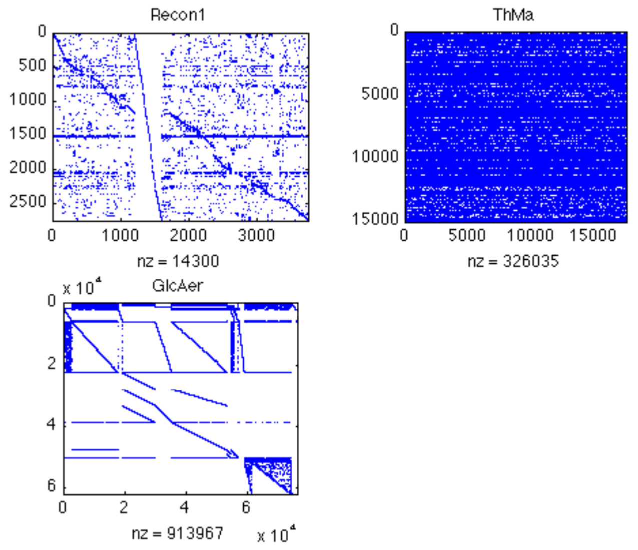
Model 10 (ThMa = Thermotoga maritima)



Model 11 (GlcAer)



Models 9, 10, 11



Rank of stoichiometric matrices

Conservation analysis for biochemical networks

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)

Conservation analysis reduces to finding rank(S) and null(S^T)

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

where z is a conserved moiety (group of chemical species)

Requires $S^T z = 0$

Finding rank(S) and null(S^T)

Conservation analysis reduces to finding rank(S) and null(S^T)

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

where z is a conserved moiety (group of chemical species)

Requires $S^T z = 0$

Also part of conservation analysis:

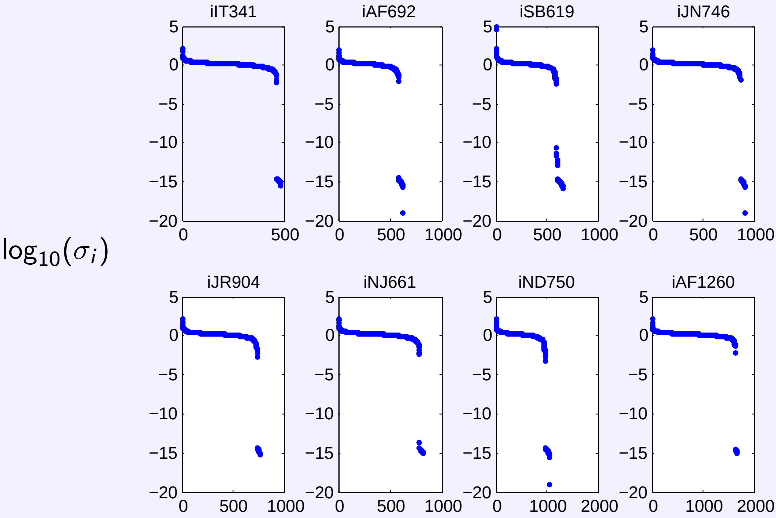
- Partitioning the rows (species) of S into dependent and independent rows (species)
- Computing a link matrix that describes the relations among the concentrations of dependent and independent species

rank(S) by SVD

Singular value decomposition $S = UDV^T$

- $U^TU = I$ $V^TV = I$ 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 ∞

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



rank(S) by QR

Householder QR factorization $\quad SP = QR$

- $P =$ col perm $\quad Q^TQ = I \quad R$ diagonal $\quad \text{rank}(S) = \text{rank}(R)$
- Nearly as reliable as SVD

rank(S) by QR

Householder QR factorization $SP = QR$

- $P =$ col perm $Q^TQ = I$ R diagonal $\text{rank}(S) = \text{rank}(R)$
- 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	0.0 secs
model 10 (ThMa)	15000×18000	2.5 secs
model 11 (GlcAer)	62000×77000	0.2 secs(!)

rank(S) by LDU

Sparse LU with Threshold Rook Pivoting (TRP) $P_1 S P_2 = LDU$

- $P_1, P_2 =$ 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, $\ldots$

rank(S) by LDU

Sparse LU with Threshold Rook Pivoting (TRP) $P_1 S P_2 = LDU$

- $P_1, P_2 =$ 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 0.0 secs
  model 10 (ThMa) 15000 × 18000 4.0 secs
  model 11 (GlcAer) 62000 × 77000 158 secs

```
       |     m      n       rank(S) |                            |  time
 model |                  SVD   SPQR | nnz(S)  nnz(Q)    nnz(R) |  SVD  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
```

```
       |     m      n       rank(S) |                         |   time
 model |                  SVD   SPQR | nnz(S)  nnz(Q)   nnz(R) |  SVD  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
```

```
       |      m       n      rank(S') |                         |   time
 model |                   SVD    SPQR | nnz(S)  nnz(Q)   nnz(R) |  SVD  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
```

Large-scale LP, QP, NLP

SQOPT

QP
$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{minimize}} && c^Tx + \tfrac{1}{2}(x-x_0)^TH(x-x_0) \\ &\text{subject to} && \ell \le \begin{pmatrix} x \\ Ax \end{pmatrix} \le u \end{aligned}$$

- Includes LP ($H = 0$)
- H symmetric positive semidefinite
- A large and sparse
- LUSOL is main engine (for satisfying the constraints)
- Fortran 77 ($\Rightarrow$ C implementation via f2c translator)
- Fortran 90 version can compile with double precision (normal) or quadruple precision (for astounding results!)

SNOPT

NLP

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \phi(x)$$

$$\text{subject to} \quad \ell \leq \begin{pmatrix} x \\ Ax \\ c(x) \end{pmatrix} \leq u$$

- $\phi(x)$ smooth nonlinear objective function
- $c(x)$ = vector of smooth nonlinear functions
- Best if gradients of nonlinear functions are known
- Uses SQOPT to solve sequence of QP subproblems

Flux Balance Analysis (FBA) on Thermotoga maritima

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	18210×17535
Nonzero S_{ij}	33602
max and min $\lvert S_{ij} \rvert$	2×10^4 and 3×10^{-6}

Flux Balance Analysis (FBA) on Thermotoga maritima

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	18210×17535
Nonzero S_{ij}	33602
max and min $\lvert S_{ij} \rvert$	2×10^4 and 3×10^{-6}

SQOPT in double precision (15 digits)

```
Feasibility tol          1e-6
Optimality  tol          1e-6
```

Flux Balance Analysis (FBA) on Thermotoga maritima

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	18210×17535
Nonzero S_{ij}	33602
max and min $\lvert S_{ij} \rvert$	2×10^4 and 3×10^{-6}

SQOPT in double precision (15 digits)

```
Feasibility tol          1e-6
Optimality  tol          1e-6
```

SQOPT in quad precision (32 digits)

```
Feasibility tol          1e-15
Optimality  tol          1e-15
```

Flux Balance Analysis (FBA) on Thermotoga maritima

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	18210×17535
Nonzero S_{ij}	33602
max and min $\lvert S_{ij} \rvert$	2×10^4 and 3×10^{-6}

SQOPT in double precision (42 secs)

```
SQOPT  EXIT  10 -- the problem appears to be infeasible

Problem name                     ThMa
No. of iterations               18500   Objective value      8.2286249495E-07
No. of infeasibilities              9   Sum of infeas        1.9606461069E-03
No. of degenerate steps         11611   Percentage                      62.76
Max x       (scaled)     3482 8.2E+00   Max pi      (scaled)    18210 9.8E-01
Max x                    5134 5.9E+00   Max pi                  18210 1.0E+00
Max Prim inf(scaled)    32832 1.3E-03   Max Dual inf(scaled)    16417 1.0E+00
Max Primal infeas       32832 5.6E-06   Max Dual infeas         32669 2.3E+02
```

Flux Balance Analysis (FBA) on Thermotoga maritima

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \leq v \leq u$$

S rows and cols	18210×17535
Nonzero S_{ij}	33602
max and min $\lvert S_{ij} \rvert$	2×10^4 and 3×10^{-6}

Restart SQOPT in quad precision (36 secs)

```
SQOPT  EXIT   0 -- finished successfully

Problem name                       ThMa
No. of iterations                   498   Objective value      8.7036461686E-07
No. of infeasibilities                0   Sum of infeas        0.0000000000E+00
No. of degenerate steps             220   Percentage                      44.18
Max x       (scaled)       3482 8.2E+00   Max pi      (scaled)      2907 1.3E+00
Max x                      5134 5.9E+00   Max pi                   15517 1.1E+00
Max Prim inf(scaled)      16475 5.2E-28   Max Dual inf(scaled)     13244 1.9E-32
Max Primal infeas         16475 5.2E-29   Max Dual infeas          13244 4.8E-33
```

Flux Balance Analysis (FBA) on GlcAer

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	68300×76664
Nonzero S_{ij}	926357
max and min $\lvert S_{ij} \rvert$	8×10^5 and 5×10^{-5}

SQOPT in quad precision cold start, no scaling (30786 secs)

```
SQOPT  EXIT   0 -- finished successfully

Problem name                    GlcAer
No. of iterations                84685   Objective value      -7.0382454070E+05
No. of degenerate steps          62127   Percentage                       73.36
Max x                    61436 6.3E+07   Max pi                   25539 2.4E+07
Max Primal infeas        72623 3.0E-21   Max Dual infeas          17817 2.7E-21
```

Flux Balance Analysis (FBA) on GlcAer

$$\min c^T v \quad \text{subject to} \quad Sv = 0, \quad \ell \le v \le u$$

S rows and cols	68300×76664
Nonzero S_{ij}	926357
max and min $\lvert S_{ij} \rvert$	8×10^5 and 5×10^{-5}

SQOPT in quad precision   cold start, with scaling (4642 secs)

```
SQOPT  EXIT   0 -- finished successfully

Problem name                      GlcAer
No. of iterations                  37025   Objective value     -7.0382454070E+05
No. of infeasibilities                 1   Sum of infeas        6.9661927856E-16
No. of degenerate steps            28166   Percentage                      76.07
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)       81918 7.0E-16   Max Dual inf(scaled)     59325 1.5E-17
Max Primal infeas          81918 1.3E-07   Max Dual infeas          27953 2.0E-22
```

PDCO
Primal-dual interior method for convex optimization

PDCO (Matlab primal-dual convex optimizer)

$$\boxed{\begin{array}{ll} \underset{x}{\text{minimize}} & \phi(x) \\ \text{subject to} & Ax = b, \quad \ell \le x \le u, \end{array}}$$

where $\phi(x)$ is convex with known gradient and Hessian.

A may be a sparse matrix or an operator for Av and $A^T w$

e.g. Basis Pursuit (BP and BPDN) Chen, Donoho, Saunders 2001

PDCO (Matlab primal-dual convex optimizer)

$$\boxed{\begin{array}{ll} \underset{x}{\text{minimize}} & \phi(x) \\ \text{subject to} & Ax = b, \quad \ell \le x \le u, \end{array}}$$

where $\phi(x)$ is convex with known gradient and Hessian.

A may be a sparse matrix or an operator for Av and $A^T w$

e.g. Basis Pursuit (BP and BPDN) Chen, Donoho, Saunders 2001

To ensure unique solutions, PDCO solves regularized problems:

$$\boxed{\begin{array}{ll} \underset{x,\,r}{\text{minimize}} & \phi(x) + \frac{1}{2}\left\|D_1 x\right\|^2 + \frac{1}{2}\left\|r\right\|^2 \\ \text{subject to} & Ax + D_2 r = b, \quad \ell \le x \le u, \end{array}}$$

where D_1, D_2 are diagonal and positive-definite.

PDCO (Matlab primal-dual convex optimizer)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \phi(x) \\ \text{subject to} \quad & Ax = b, \quad \ell \le x \le u, \end{aligned}$$

where $\phi(x)$ is convex with known gradient and Hessian.
A may be a sparse matrix or an operator for Av and $A^T w$
e.g. Basis Pursuit (BP and BPDN) Chen, Donoho, Saunders 2001

To ensure unique solutions, PDCO solves regularized problems:

$$\begin{aligned} \underset{x,\,r}{\text{minimize}} \quad & \phi(x) + \tfrac{1}{2}\|D_1 x\|^2 + \tfrac{1}{2}\|r\|^2 \\ \text{subject to} \quad & Ax + D_2 r = b, \quad \ell \le x \le u, \end{aligned}$$

where D_1, D_2 are diagonal and positive-definite.

Typically $D_1 = \gamma I$ $\quad\quad \gamma = 10^{-3}$ or 10^{-4}
Same for D_2 if $Ax = b$ should be satisfied accurately
For least-squares problems $D_2 = I$

PDCO applied to FBA

FBA

$$\begin{array}{ll} \underset{v_f, v_r, v_e}{\text{minimize}} & d^T v_e \\ \text{subject to} & S v_f - S v_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \qquad \ell \leq v_e \leq u \end{array}$$

- Flux Balance Analysis = LP problem (Palsson 2006)
- d optimizes a biological objective
  e.g., maximize replication rate in unicellular organisms
- v_e = exchange fluxes = sources and sinks of chemicals

PDCO applied to FBA

FBA

$$\begin{aligned} \underset{v_f, v_r, v_e}{\text{minimize}} \quad & d^T v_e \\ \text{subject to} \quad & S v_f - S v_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \qquad \ell \leq v_e \leq u \end{aligned}$$

- Flux Balance Analysis = LP problem (Palsson 2006)
- d optimizes a biological objective
  e.g., maximize replication rate in unicellular organisms
- v_e = exchange fluxes = sources and sinks of chemicals
- PDCO works with $A = \begin{bmatrix} S & -S & S_e \end{bmatrix}$ then $LL^T = AD^2A^T$
  (sparse Cholesky with D increasingly ill-conditioned)

PDCO applied to FBA

$$\boxed{\text{FBA} \qquad \begin{array}{ll} \underset{v_f, v_r, v_e}{\text{minimize}} & d^T v_e \\ \text{subject to} & S v_f - S v_r + S_e v_e = 0 \\ & v_f, v_r \geq 0, \qquad \ell \leq v_e \leq u \end{array}}$$

- Flux Balance Analysis = LP problem (Palsson 2006)
- d optimizes a biological objective
  e.g., maximize replication rate in unicellular organisms
- v_e = exchange fluxes = sources and sinks of chemicals
- PDCO works with $A = \begin{bmatrix} S & -S & S_e \end{bmatrix}$ then $LL^T = AD^2A^T$
  (sparse Cholesky with D increasingly ill-conditioned)
- Solution is $v^* = v_f^* - v_r^*$ and v_e^*

PDCO applied to Entropy problem

$$\boxed{\begin{array}{lll}\text{EP} & \underset{v_f,v_r}{\text{minimize}} & v_f^T(\log v_f + c - e) + v_r^T(\log v_r + c - e) \\ & \text{subject to} & Sv_f - Sv_r = -S_e v_e^* \\ & & v_f, v_r > 0\end{array}}$$

- $c =$ any vector, $e = (1, 1, \ldots, 1)^T$
  $v_e^* =$ optimal exchange fluxes from FBA
- Entropy objective function is strictly convex
- Solution v_f^*, v_r^* is thermodynamically feasible
  (satisfies energy conservation and 2nd law of thermodynamics)

Fleming, Maes, Saunders, Ye, Palsson (2012)

References

- S. S. Chen, D. L. Donoho, M. A. Saunders (2001).
  Atomic decomposition by basis pursuit,
  SIAM Review 43(1):129–159.
- S.-C. Choi, C. C. Paige, M. A. Saunders (2011).
  MINRES-QLP: a Krylov subspace method for indefinite or singular symmetric systems,
  SIAM Journal on Scientific Computing 33(4):1810–1836.
- T. A. Davis (2013).
  Algorithm 9xx: SuiteSparseQR, a multifrontal multithreaded sparse QR factorization package, *ACM TOMS*, submitted.
- R. M. T. Fleming, C. M. Maes, M. A. Saunders, Y. Ye, B. Ø. Palsson (2012).
  A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks,
  Journal of Theoretical Biology 292:71–77.
- D. C.-L. Fong and M. A. Saunders (2012).
  LSMR: An iterative algorithm for sparse least-squares problems,
  SIAM J. Scientific Computing 33(5):2950–2971.
- P. E. Gill, W. Murray, M. A. Saunders, M. H. Wright (1987).
  Maintaining LU factors of a general sparse matrix,
  Linear Algebra and its Applications, 88/89:239–270.

References (contd)

- 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`.
- S. P. Ponnapalli, M. A. Saunders, C. F. Van Loan, O. Alter.
  A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms,
  PLoS ONE 6(12): e28072 (2011). doi:10.1371/journal.pone.0028072, 11 pp.
- R. R. Vallabhajosyula, V. Chickarmane, H. M. Sauro (2005).
  Conservation analysis of large biochemical networks,
  Bioinformatics 22(3):346–353.

Conclusions

- $Ax \approx b$ iterative solvers: MINRES-QLP, LSMR, LSRN
  $Ax = b$ direct solvers: LUSOL and many others

Conclusions

- $Ax \approx b$ iterative solvers: MINRES-QLP, LSMR, LSRN
  $Ax = b$ direct solvers: LUSOL and many others
- Numerical rank of stoichiometric S is clearly defined
  SPQR and LUSOL (threshold rook pivoting) seems reliable
  SPQR and LUSOL on S is usually faster than on S^T
  SPQR is extremely fast (except if even 1 dense row)
  LUSOL with rook pivoting is more sparse

Conclusions

- $Ax \approx b$ iterative solvers: MINRES-QLP, LSMR, LSRN
  $Ax = b$ direct solvers: LUSOL and many others
- Numerical rank of stoichiometric S is clearly defined
  SPQR and LUSOL (threshold rook pivoting) seems reliable
  SPQR and LUSOL on S is usually faster than on S^T
  SPQR is extremely fast (except if even 1 dense row)
  LUSOL with rook pivoting is more sparse
- FBA (Flux balance analysis):
  SQOPT (double precision) + restart SQOPT (quad precision)
  should be effective on very large models

Conclusions

- $Ax \approx b$ iterative solvers: MINRES-QLP, LSMR, LSRN
  $Ax = b$ direct solvers: LUSOL and many others
- Numerical rank of stoichiometric S is clearly defined
  SPQR and LUSOL (threshold rook pivoting) seems reliable
  SPQR and LUSOL on S is usually faster than on S^T
  SPQR is extremely fast (except if even 1 dense row)
  LUSOL with rook pivoting is more sparse
- FBA (Flux balance analysis):
  SQOPT (double precision) + restart SQOPT (quad precision)
  should be effective on very large models
- FBA + thermodynamically feasible solution:
  PDCO with entropy objective

Conclusions

- $Ax \approx b$ iterative solvers: MINRES-QLP, LSMR, LSRN
  $Ax = b$ direct solvers: LUSOL and many others
- Numerical rank of stoichiometric S is clearly defined
  SPQR and LUSOL (threshold rook pivoting) seems reliable
  SPQR and LUSOL on S is usually faster than on S^T
  SPQR is extremely fast (except if even 1 dense row)
  LUSOL with rook pivoting is more sparse
- FBA (Flux balance analysis):
  SQOPT (double precision) + restart SQOPT (quad precision)
  should be effective on very large models
- FBA + thermodynamically feasible solution:
  PDCO with entropy objective
- `http://www.stanford.edu/group/SOL/`
  `http://www.stanford.edu/group/SOL/multiscale/`

Future work

- Randomized numerical linear algebra
  - How to design *statistically aware* algorithms for matrix computations?
  - How to parallelize algorithms to handle truly massive data sets?
  - For example, LSRN

Future work

- Randomized numerical linear algebra
    - How to design *statistically aware* algorithms for matrix computations?
    - How to parallelize algorithms to handle truly massive data sets?
    - For example, LSRN
- High-dimensional statistics
    - How to make valid inference when the number of problem parameters is much larger than the sample size?
    - How to construct confidence regions and obtain p-values in this setting?

Acknowledgements

- Tim Davis, UFL (SPQR)
- Nick Henderson, Stanford (Matlab interface to LUSOL)
- Philip Gill, Elizabeth Wong, UCSD (SNOPT9)
- Joshua Lerman, UCSD (Thermotoga maritima FBA model)
- Orly Alter, U Utah (Track organizer!)
- DOE DE-FG02-09ER25917, NIH U01-GM102098