General-purpose Software for Systems Biology

Ron Estrin, Ronan Fleming, Nick Henderson, San Kim, Ding Ma, Yuekai Sun, Ines Thiele, Laurence Yang, Bernhard Palsson, and Michael Saunders

Stanford University, University of Luxembourg, and UC San Diego

Abstract

We summarize research and general-purpose software packages that have applications to multiscale modeling in systems biology [5]. They include

- lusolZ: Sparse-matrix factorizations for rank and nullspace computations [2, 3]
- PNOPT: An optimization method for composite smooth and nonsmooth functions [7]
- double-MINOS, quad-MINOS, PDCO: Constrained optimization routines for linear and nonlinear flux balance analysis (FBA) and general optimization using double- and quadruple-precision [1, 4, 6, 8, 9]

The need for quadruple precision

Carrying somewhat more precision in the arithmetic than twice the precision carried in the data and available for the result will vastly reduce embarrassment due to roundoff-induced anomalies

Default evaluation in quad is the humane option

— *William Kahan (2012)*

Quad datatypes are now available

GCC provides `real(16)` and `float128` in gfortran, C, C++

This is the humane option for producing quad software

- double-MINOS, quad-MINOS called from f77, f90, python
- C++ PDCO will switch from double to quad at runtime

Linear ME model of *E. coli* double-MINOS, quad-MINOS LP

Problem GlcAerWT (Thiele, Fleming, et al. 2012), **68300** $\times$ **76664**

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
```

Nonlinear ME model variables μ, v, w

μ = growth rate

$$\begin{aligned} &\max\ \mu \\ &\quad \text{st } \mu Av + Bv = 0 \\ &\qquad\quad Sv\ \ \ = b \\ &\quad \text{bounds on } v \end{aligned} \quad \equiv \quad \begin{aligned} &\max\ \mu \\ &\quad \text{st } \mu Av + w = 0 \\ &\qquad\quad Bv - w = 0 \\ &\qquad\quad Sv\ \ \ = b \\ &\quad \text{bounds on } v\text{, none on } w \end{aligned}$$

A and B overlap

solveME (Yang et al. [9]) 11386 $\times$ 18755 quad-MINOS NLP

```
Calling minoss. Warm start with provided basis (hs)
Itn       32 -- linear constraints satisfied.
Calling funcon.  mu =  0.832815729997476367249118875820191994
nnCon, nnJac, neJac         2629       16126       14322
funcon  sets    14322    out of    14322    constraint gradients.
funobj  sets        1    out of        1    objective  gradients.

Major minor  step     objective  Feasible Optimal   nsb    ncon penalty BSwap
    1    32T 0.0E+00  8.32816E-01 4.3E-13 1.0E+03     0       4 1.0E+02      0

   19    40T 1.0E+00  8.32816E-01 2.5E-16 1.0E-03     0     743 1.0E+02      0
   20    40T 1.0E+00  8.32816E-01 1.0E-21 9.3E-04     0     784 1.0E+02      0

   23    40T 1.0E+00  8.55337E-01 3.4E-07 5.7E-05     0     907 1.0E+02      0
   24    40T 1.0E+00  8.55664E-01 2.1E-08 6.6E-07     0     948 1.0E+02      0
Itn      979 --      10  nonbasics set on bound, basics recomputed
   25    11  1.0E+00  8.55664E-01 7.0E-17 1.8E-11     0     961 1.0E+02      0
   26     0  1.0E+00  8.55664E-01 9.3E-19 8.0E-29     0     962 1.0E+01      0

EXIT -- optimal solution found

No. of iterations                 979   Objective value      8.5566388920E-01
No. of major iterations            26   Linear objective     0.0000000000E+00
Penalty parameter            1.000000   Nonlinear objective  8.5566388920E-01
No. of calls to funobj              962   No. of calls to funcon            962
No. of superbasics                    0   No. of basic nonlinears          7896
Max x       (scaled)    12918 4.5E+01   Max pi       (scaled)     9674 3.3E+04
Max x                   14520 4.5E+01   Max pi                    4138 3.8E+03
Max Prim inf(scaled)        0 0.0E+00   Max Dual inf(scaled)         1 8.1E-24
Max Primal infeas           0 0.0E+00   Max Dual infeas           6827 1.1E-24
Nonlinear constraint violn    1.4E-19
```

Optimal control problem modeling a spring/mass/damper

$$\min\ f(y,z,u) = \frac{1}{2}\sum_{t=0}^{T} z_t^2$$

$$y_{t+1} = y_t - 0.01y_t^2 - 0.004z_t + 0.2u_t$$
$$z_{t+1} = z_t + 0.2y_t$$

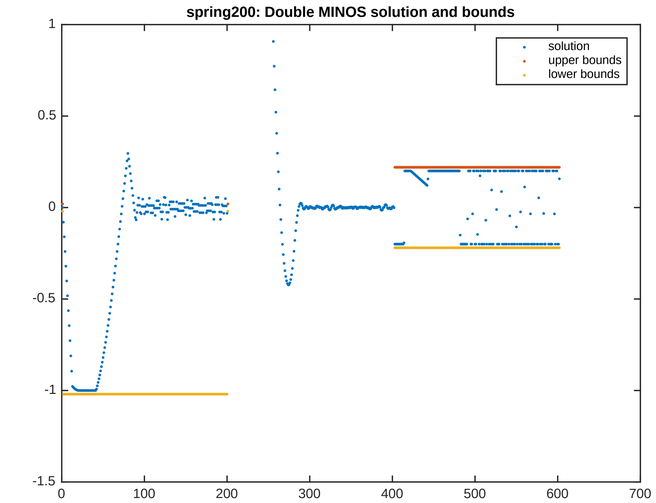
$$-1 \le y_t \qquad y_0 = 0,\ y_T = 0$$
$$-0.2 \le u_t \le 0.2 \qquad z_0 = 10$$

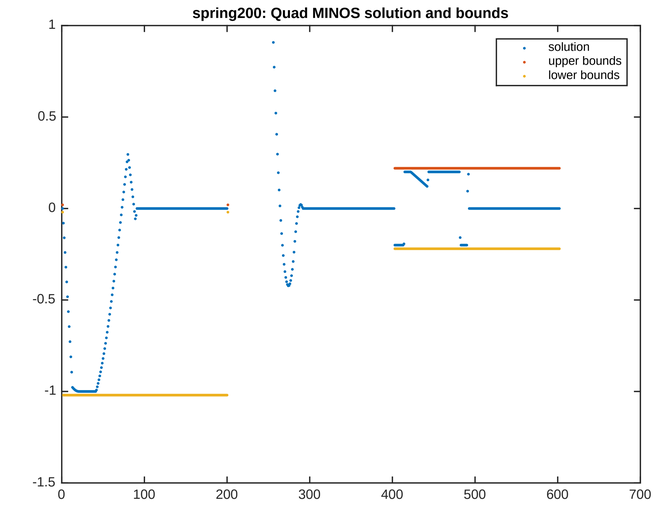
	Opt tol	Majors	Minors	Superbasics	Objective	Time
double	1e-06	13	576	18	1186.3839	0.05
quad	1e-15	31	1282	113	1186.3820	2.75

quad-MINOS gives an unexpectedly "clean" solution
(many variables exactly zero, including control variables u_t)

Funding

NIH U01GM102098
DOE ER65524





References

1 The COBRA Toolbox. `https://github.com/opencobra/cobratoolbox`, 2016.

2 LUSOL sparse LU factorization package. Fortran 77 code with Matlab interface, `http://stanford.edu/group/SOL/software/lusol`, 2013.

3 lusolZ: Sparse nullspace computations using SuiteSparseQR or LUSOL. Matlab code, `http://stanford.edu/group/SOL/software/lusolZ`, 2016.

4 D. Ma, L. Yang, R. M. T. Fleming, I. Thiele, B. O. Palsson, and M. A. Saunders. Reliable and efficient solution of genome-scale models of metabolism and macromolecular expression. *Sci. Rep., 7, 40863*, `doi:10.1038/srep40863`, 2017.

5 Multiscale systems biology collaboration. `http://stanford.edu/group/SOL/multiscale/`, 2016.

6 PDCO: Primal-Dual interior method for Convex Objectives and linear constraints. Matlab code, `http://stanford.edu/group/SOL/software/pdco/`, 2015.

7 PNOPT: Proximal Newton Optimizer. `http://stanford.edu/group/SOL/software/pnopt/`, 2015.

8 quad-PDCO: Primal-Dual interior method for Convex Objectives and linear constraints. C++ code combining double- and quad-precision, in preparation.

9 L. Yang, D. Ma, A. Ebrahim, C. J. Lloyd, M. A. Saunders, and B. O. Palsson. solveME: fast and reliable solution of nonlinear ME models. *BMC Bioinformatics*, 17(391), `doi:10.1186/s12859-016-1240-1`, 2016.