LU preconditioning for full-rank and singular sparse least squares

Nick Henderson, Ding Ma, and Michael Saunders

ICME and MS&E, Stanford University

SIAM Conference on Applied Linear Algebra

Atlanta, GA

October 26-30, 2015

Abstract

Sparse QR factorization is often the most efficient approach to solving sparse least-squares problems $\min \|Ax - b\|$, especially since the advent of Davis's SuiteSparseQR software. For cases where QR factors are unacceptably dense, we consider sparse LU factors from LUSOL for preconditioning an iterative solver such as LSMR.

LUSOL computes factors of the form $P_1 A P_2 = LU$, with permutations chosen to preserve sparsity while ensuring L is well-conditioned. For full-rank problems, one can right-precondition with either U or B, where B is a basis from the rows of A defined by P_1 [Saunders, 1979]. More recently, Arioli and Duff have recommended that B be chosen to have maximum volume.

We experiment with LUSOL and LSMR on many realistic examples. For singular problems we make use of LUSOL's threshold rook pivoting option, and investigate whether threshold complete pivoting increases the volume of B in a useful way.

Partially supported by the
National Institute of General Medical Sciences
of the National Institutes of Health (NIH)
Award U01GM102098

Background

Direct and iterative methods for least squares

Dense $\min \|Ax - b\|$

- QR — full rank (Golub 1965)
- QR — tall and skinny (MapReduce: Benson, Gleich & Demmel 2013)
- SVD — singular

Sparse $\min \|Ax - b\|^2 + \delta^2 \|x\|^2$

- SuiteSparseQR (Davis 2011) — full rank or singular
- LSQR (Paige & S 1982) — same
- LSMR (Fong & S 2011) — same
- EN-LSQR (Arioli & Orban 2013) — SQD version
- LSRN (Meng, S, & Mahoney 2014) — dense or sparse tall and skinny

LU preconditioning

Dense $\min \|Ax - b\|$, full rank

$PA = LU$ Partial Pivoting $|L_{ij}| \le 1$

Peters and Wilkinson 1970 $L^T L y = L^T P b$, $Ux = y$

Sparse $\min \|Ax - b\|$, full rank

$P_1 A P_2 = LU$ Threshold Partial Pivoting (TPP) $|L_{ij}| \le 1.1$

S, 1979	LSQR	preconditioner U	row-oriented TPP	
Arioli & Duff 2015	EN-LSQR	preconditioner B	MA58	TPP, TRP
Today	LSMR	preconditioner I, U, B	LUSOL	TPP, TRP, TCP

LU preconditioning for $\min \|Ax - b\|$

Peters and Wilkinson 1970 Dense A

LU with Partial Pivoting

- $PA = LU \Rightarrow \|Ax - b\| = \|LUx - Pb\|$ P is likely to keep L well-conditioned
- $\min \|Ax - b\| \equiv \min \|Ly - Pb\|$
- $L^T Ly = L^T Pb$ is ok
- $Ux = y$

Saunders 1979 sparse A

LSQR with preconditioner U

- $PA = LU = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} U = \begin{bmatrix} B \\ N \end{bmatrix}$ $|L_{ij}| \le 1.1 \Rightarrow L$ probably well-conditioned
- Primitive row-oriented LU
- LSQR improved on illc1033

Suggested LSQR with preconditioner B

- $\|Ax - b\| = \left\| \begin{bmatrix} B \\ N \end{bmatrix} x - Pb \right\| \Rightarrow \left\| \begin{bmatrix} I \\ NB^{-1} \end{bmatrix} y - Pb \right\|, \quad Bx = y$
- $NB^{-1} = L_2 L_1^{-1}$
- $\|NB^{-1}\|$ should not be large $\Rightarrow \begin{bmatrix} I \\ NB^{-1} \end{bmatrix}$ should be well-conditioned
- Can do sparser $B = LU$ once B is found

Arioli & Duff 2015 sparse A

Modern sparse $PA = LU$ (column perm also)

- $PA = LU = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} U = \begin{bmatrix} B \\ N \end{bmatrix}$ $|L_{ij}| \le 1.1 \Rightarrow L$ probably well-conditioned
- $\|Ax - b\| = \left\| \begin{bmatrix} B \\ N \end{bmatrix} x - Pb \right\| \Rightarrow \left\| \begin{bmatrix} I \\ NB^{-1} \end{bmatrix} y - Pb \right\|, \quad Bx = y$
- Focused on $\det(U) = \text{volume}(B)$ large
- $\|NB^{-1}\|$ should not be large $\Rightarrow$ $\begin{bmatrix} I \\ NB^{-1} \end{bmatrix}$ should be well-conditioned
- B should be good preconditioner
- $Bx_0 = b_B$ gives starting point x_0

Tikhonov

Take advantage of 0 in rhs

- $PA = \begin{bmatrix} B \\ N \end{bmatrix}$, form $Pb = \begin{bmatrix} b_B \\ b_N \end{bmatrix}$
- Solve $Bx_0 = b_B$, form $r_N = b_N - Nx_0$
- Solve $\min \left\| \begin{bmatrix} NB^{-1} \\ I \end{bmatrix} z - \begin{bmatrix} r_N \\ 0 \end{bmatrix} \right\|$ rank(NB^{-1}) is important, not rank(AB^{-1})
  Theoretically, LSQR itns $\leq \min\{m-n, n\}$
  $x = x_0 + z$
- Arioli and Duff 2015:
  If $Ax = b$ is consistent, $r_N = 0$ and $z = 0$ (0 iterations)
  If $Ax \approx b$, $r_N \approx 0$ and need few iterations
- Needs correct stopping rule! Must allow for small $\|\text{rhs}\|$!

Arioli & Duff 2015

Four-step method

1. MA58: find P such that $PA = \begin{bmatrix} B \\ N \end{bmatrix}$ TPP with $|L_{ij}| < 1.1$ (major finding)
2. Goreinov, Tyrtyshnikov, et al. 2001, 2010: increase volume(B) by interchanging rows of B and N
3. MA48: find sparser $B = LU$
4. EN-LSQR (Arioli & Orban 2013): solve SQD system

$$\begin{bmatrix} I & NB^{-1} \\ B^{-T}N^T & -I \end{bmatrix} \begin{bmatrix} r_N \\ z \end{bmatrix} = \begin{bmatrix} b_N \\ -b_B \end{bmatrix}$$

with $z_0 = b_B$, $Bx = z$ (equivalent to Tikhonov)

LUSOL
TPP, TRP, TCP
Threshold partial/rook/complete pivoting

Needed for MINOS, SNOPT, PATH, ... basis handling

Also for $\begin{bmatrix} B & S \end{bmatrix} P = \begin{bmatrix} \bar{B} & \bar{S} \end{bmatrix}$

via TPP on $\begin{pmatrix} B^T \\ S^T \end{pmatrix}$ **keeping** L **well-conditioned**

LU with threshold pivoting keeping L and/or U well-conditioned

Best to think of

$$P_1 A P_2 = LDU \qquad \text{like } A = U \Sigma V^T$$

$$L = \begin{bmatrix} L_1 & \\ L_2 & I \end{bmatrix} \qquad D = \begin{bmatrix} D_1 & \\ & 0 \end{bmatrix} \qquad U = \begin{bmatrix} U_1 & U_2 \\ & I \end{bmatrix}$$

L, U have unit diags

τ = threshold pivot tolerance (to preserve stability) $\quad \tau = 1/u$

$\tau = 1.1$ here

TRP and TCP are rank-revealing

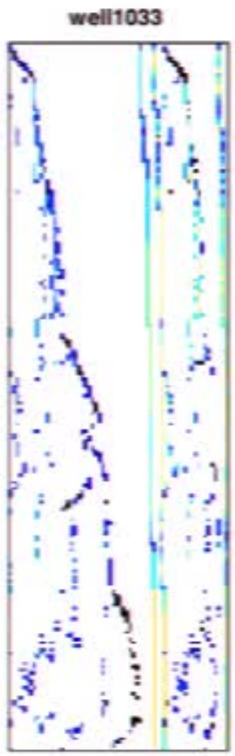
TPP	$\lvert L_{ij}\rvert \le \tau$	L well-conditioned
TRP	$\lvert L_{ij}\rvert \le \tau, \quad \lvert U_{ij}\rvert \le \tau$	L, U well-conditioned
TCP	$\lvert L_{ij}\rvert \le \tau, \quad \lvert U_{ij}\rvert \le \tau, \quad \lvert D_{11}\rvert \gtrsim \lvert D_{22}\rvert \gtrsim \dots$	L, U well-conditioned

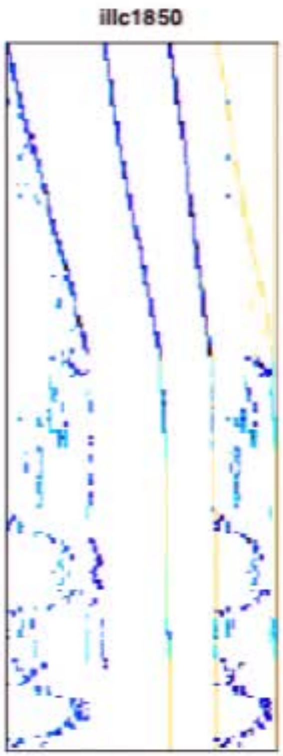
The DSIR Gravity Meter problems

D. Woodward, New Zealand, 1979

Problem	m	n	nnz
well1033	1033	320	4732
well1850	1033	712	8758
illc1033	1850	320	4732
illc1850	1850	712	8758

Gravity well1033 1033 × 320 illc1850 1850 × 712





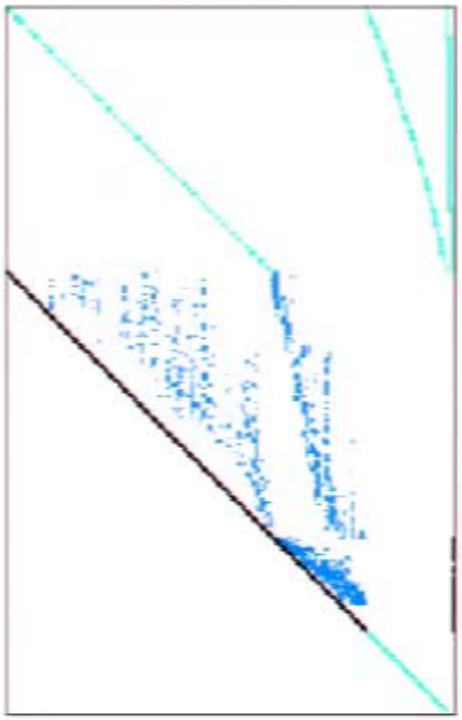
The PIGS problems

M. Hegland, CERFACS, 1993

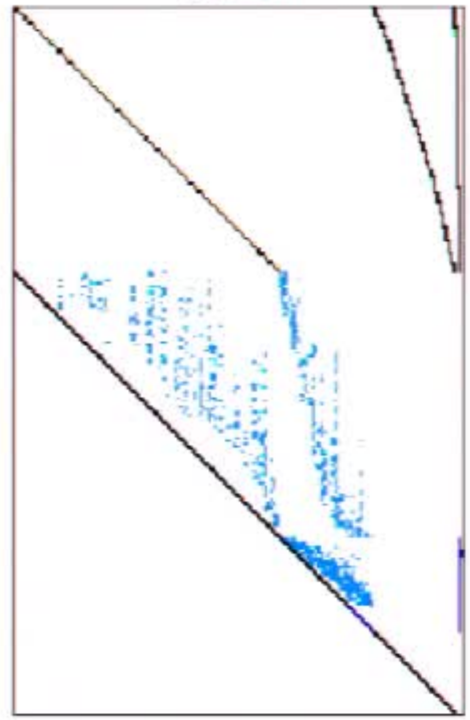
Problem	m	n	nnz
small	3140	1988	8510
small2	6280	3976	25530
medium	9397	6119	25013
medium2	18794	12238	75039
large	28254	17264	75018
large2	56508	34528	225054
very	174193	105882	463303
very2	348386	211764	1389909

PIGS small 3140 × 1988 small2 6280 × 3976

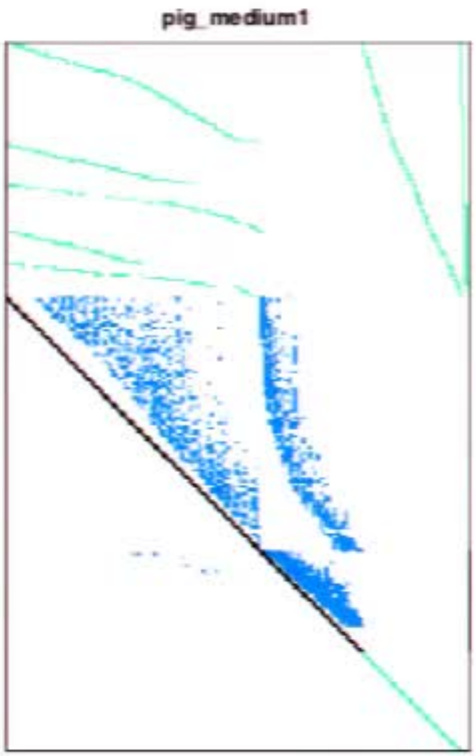
pig_small1

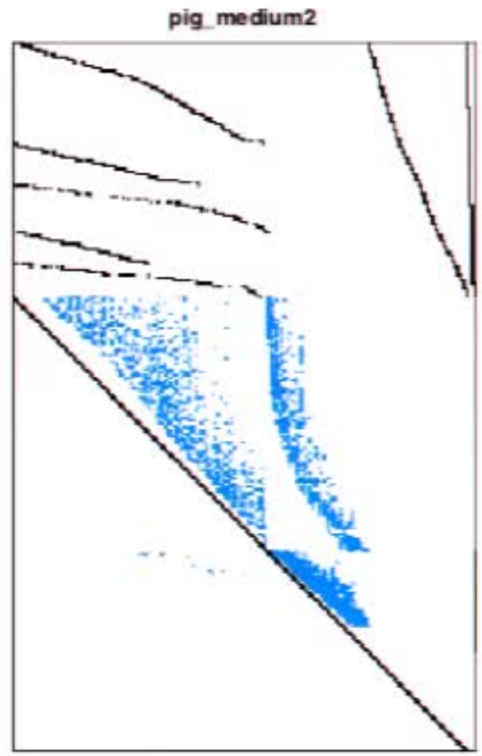


pig_small2

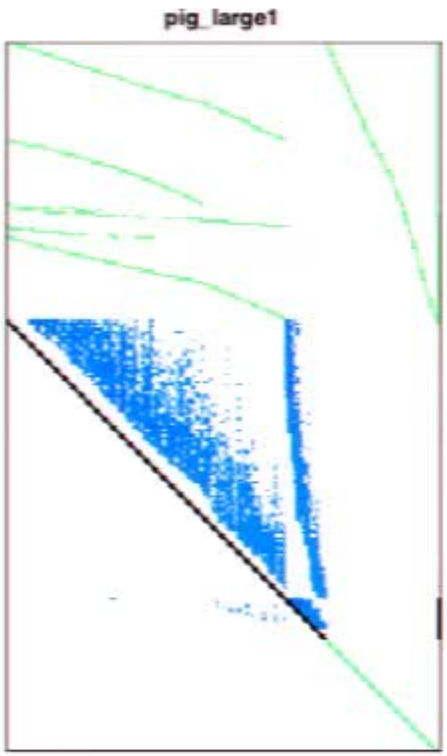


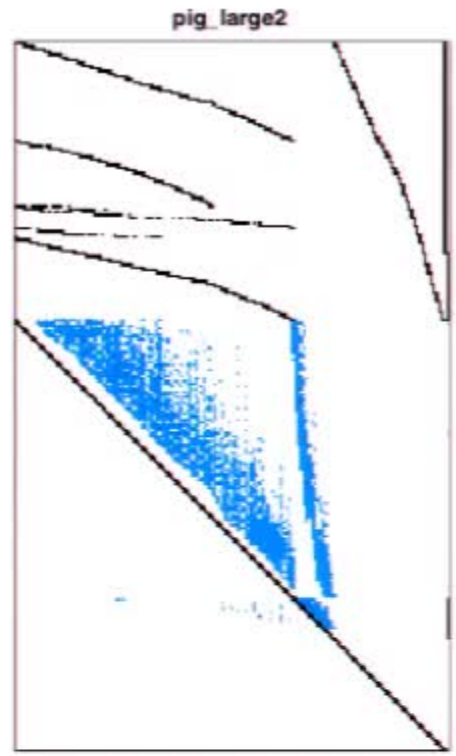
PIGS medium 9397×6119 medium2 18794×12238



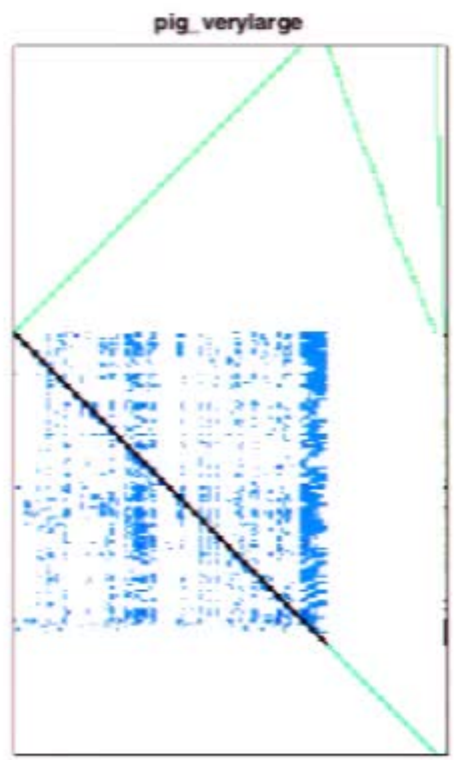


PIGS large 28254 × 17264 large2 56508 × 34528





PIGS verylarge 174193×105882 verylarge2 348386×211764



LSMR itns on $\min \|Ax - b\|$ with preconditioners I, U, B $\quad (x^* = \mathbf{1},\ b = Ax^*)$

	I	U TPP	U TRP	U TCP	B TPP	B TRP	B TCP
well1033	179	135	125	107	57	61	56
well1850	455	334	355	246	82	91	80
illc1033	3356	145	251	208	57	55	56
illc1850	2152	283	311	205	83	79	82
lp_osa_07	100	95	95	150	176	175	218
lp_osa_14	110	109	109	161	209	208	253
lp_osa_30	110	117	117	175	223	223	266
lp_osa_60	91	128	128	197	235	237	278
mesh_deform	599	161	154	142	273	273	246
small	205	228	224	140	90	88	88
small2	675	311	288	206	100	99	99
medium	241	301	283	196	137	142	141
medium2	801	318	318	309	135	139	136
large	250	454	441	305	214	216	218
large2	978	562	569	615	214	224	230
very	352	765	742	359	289	292	296
very2	1409	924	921	980	307	320	317

LSMR time on min $\|Ax - b\|$ with preconditioners I, U, B $\quad (x^* = \mathbf{1},\ b = Ax^*)$

	I	U TPP	U TRP	U TCP	B TPP	B TRP	B TCP
well1033	0.01	0.01	0.01	.005	.002	.003	.003
well1850	0.02	0.03	0.03	0.02	.007	.008	.007
illc1033	0.09	0.01	0.01	0.01	.003	.002	.002
illc1850	0.11	0.02	0.02	0.02	.007	.007	.007
lp_osa_07	0.07	0.07	0.07	0.11	0.12	0.12	0.14
lp_osa_14	0.16	0.18	0.18	0.27	0.32	0.30	0.36
lp_osa_30	0.34	0.43	0.41	0.62	0.71	0.68	0.80
lp_osa_60	0.66	1.1	1.1	1.7	1.8	1.8	2.1
mesh_deform	2.7	1.0	1.0	0.92	1.4	1.4	1.2
small	0.01	0.03	0.03	0.02	0.01	0.01	0.01
small2	0.11	0.09	0.09	0.06	0.03	0.03	0.03
medium	0.05	0.12	0.11	0.07	0.05	0.05	0.06
medium2	0.39	0.31	0.30	0.30	0.12	0.13	0.14
large	0.15	0.48	0.47	0.33	0.21	0.22	0.23
large2	1.4	1.7	1.6	1.8	0.55	0.59	0.67
very	1.4	6.3	6.1	3.0	2.1	2.2	2.4
very2	16.	23.	23.	24.	6.3	6.7	7.1

Largest Arioli & Duff example

Problem	m	n	nnz
Rav4	2880897	238324	558270

This is the only one for which QR (via Matlab backslash) was significantly slower

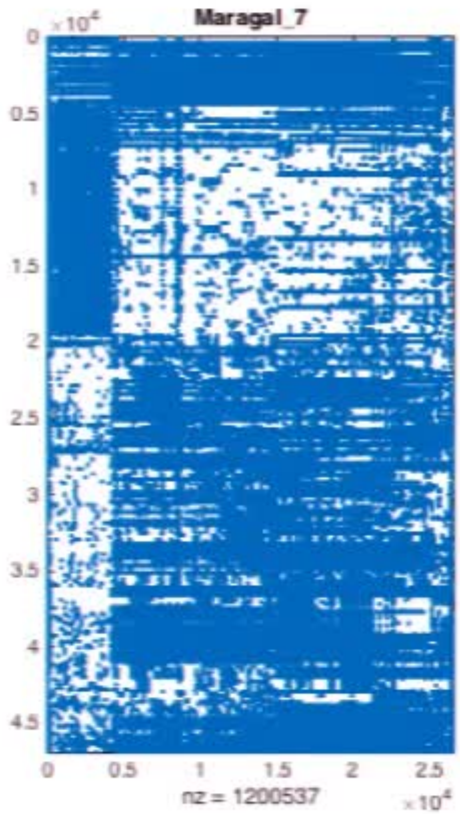
$$\text{nnz}(R) \approx 300\text{M}$$

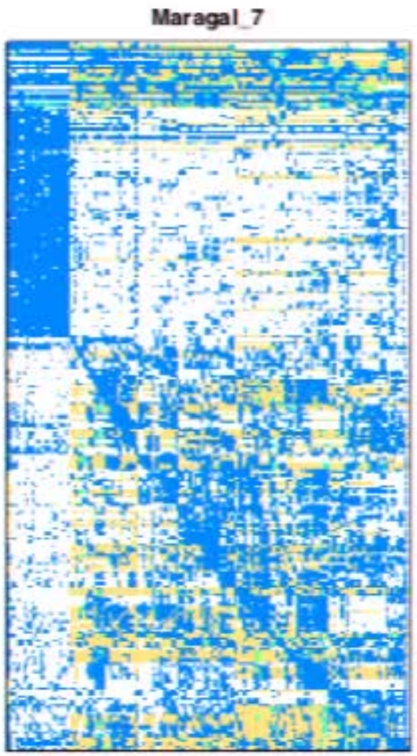
So far, our *raison d'être* for LU preconditioning

Largest rectangular A's in UFL collection

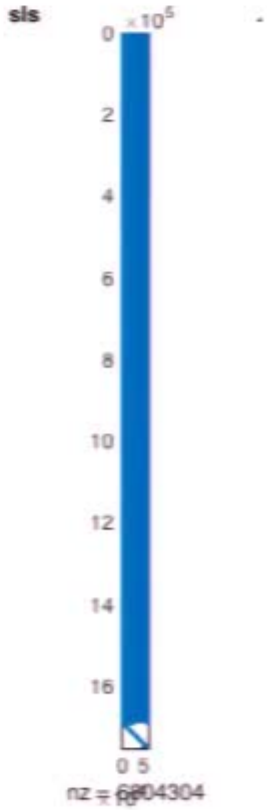
Problem	m	n	nnz
Maragal_7	46845	26564	1200537
landmark	71952	2704	1146848
ESOC	327062	37830	6019939
sls	1748122	62729	6804304
Rucci1	1977885	109900	7791168

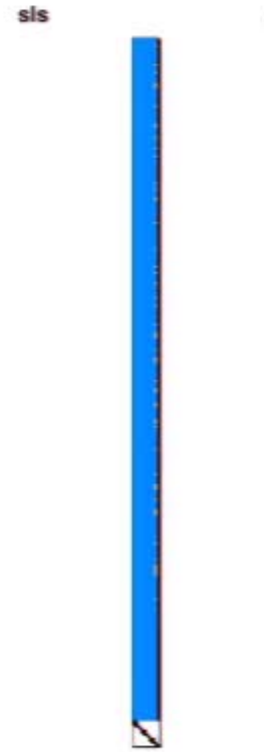
Maragal_7


Maragal_7
×10^4
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0
0.5
1
1.5
2
2.5
nz = 1200537
×10^4

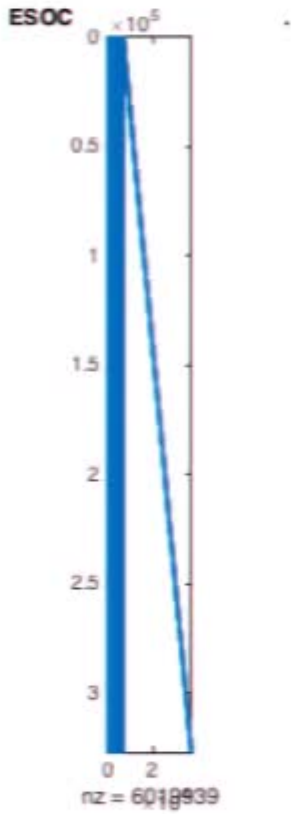

Maragal_7

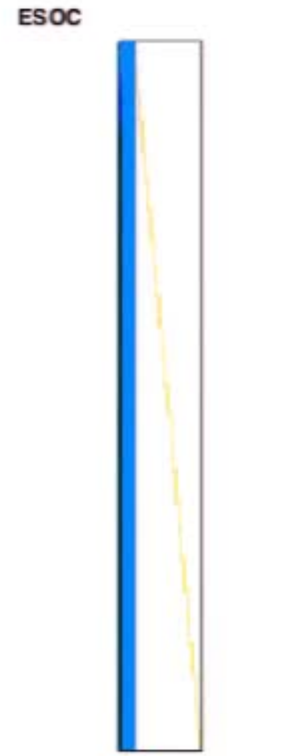
sls



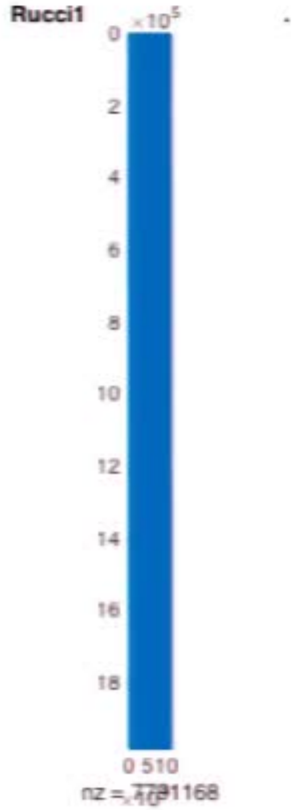


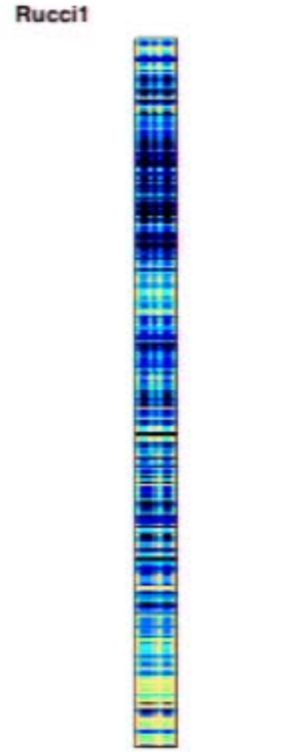
ESOC





Rucci1





Difficulties

- Maragal_7 is rank-deficient by 5721 or more
- landmark is rank-deficient by 31 or 32 (TPP) or 33 (TRP)
- landmark took 30 mins for TRP on A (only 0.1 secs on B)
- ESOC took 8 hours trying TPP on A with 50M storage for each LU array
  Needs more RAM, 8-byte integers

LSMR itns on min $\|Ax - b\|$ with preconditioners I, U, B $(b = \mathbf{1})$

	I	U TPP	U TRP	U TCP	B TPP	B TRP	B TCP
Maragal_7	6598						
landmark	20198						
ESOC	43502						
sls	201	473	473		713		
Rucci1	9441	4336	3707		663		

LSMR time on min $\|Ax - b\|$ with preconditioners I, U, B $(b = \mathbf{1})$

	I	U TPP	U TRP	U TCP	B TPP	B TRP	B TCP
Maragal_7	37						
landmark	92						
ESOC	1510						
sls	13	33	33		42.9		
Rucci1	516	343	290		412		