Solving large linear least squares problems with equality constraints

Jennifer Scott

STFC Rutherford Appleton Laboratory

and the University of Reading

Joint with Miroslav Tůma, Charles University, Prague

Sparse Days in St Girons 2022

Introduction to LSE problem

Assume $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{p\times n}$, with $m > n \gg p$.

Further assume A is large and sparse and C represents a few, possibly dense, linear constraints.

Given $b \in \mathbb{R}^m$ and $d \in \mathbb{R}^p$, the least squares problem with equality constraints (LSE problem) is

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2 \text{ subject to } Cx = d.$$

A solution exists if and only if $Cx = d$ is consistent.

Solution unique if and only if $\mathcal{A} = \begin{pmatrix} A \\ C \end{pmatrix}$ has full column rank.

Aims of our work

Revisit solution strategies; understand their strengths and weaknesses; propose new ideas and modifications designed for tackling large-scale LSE problems.

Particular emphasis on

- the possibility that the constraints may be dense
- requiring the constraints are tightly satisfied
- testing out ideas on real data (note: published literature related to LSE problems lacks numerical results, presumably because implementing proposed methods is tough)

Only time here for a quick overview and study summary.

Null-space approach

Classical technique (Hanson and Lawson 1969).

For LSE problems:

1. Find $x_1 \in \mathbb{R}^n$ such that $C\,x_1 = d$ (particular solution of constraint equation).
2. Construct $Z \in \mathbb{R}^{n\times(n-p)}$ of full column rank such that $C\,Z = 0$ (its columns form a basis for null-space $\mathcal{N}(C)$).
3. Solve the SPD normal equations of order $n-p$ $Z^T A^T A Z\,x_2 = (A\,Z)^T(b - A\,x_1)$.
4. Set $x = x_1 + Z\,x_2$.

Null-space approach: issues

- Must construct null-space basis Z.
- Traditional approach is to compute QR factorization of C^T and use to obtain Z. **But** resulting Z is dense.
- Thus $Z^T A^T AZ$ is dense so impractical to store in memory, expensive to solve.
- **Remedy:** Scott and Tůma (2022) developed method to compute sparse (banded) Z when C is “wide” using a QR algorithm applied to C that incorporates threshold column pivoting.
- Threshold parameter balances stability with sparsity of Z.

Null-space approach: results

Mixed findings: can work well but constraints not always tightly satisfied.

For these examples, constraint matrix C is dense.

θ is threshold pivoting parameter; *density* is density of $Z^T A^T A Z$.

		$\theta = 1$		$\theta = 0.1$	
Identifier	p	*density*	$\|r_c\|$	*density*	$\|r_c\|$
scagr7-2b	7	0.03	1.27×10^{-8}	0.0007	2.79×10^{-8}
south31	5	0.20	‡	0.02	3.26×10^{-7}
testbig	8	0.03	2.53×10^{-11}	0.0002	2.90×10^{-11}
lp_fit2p	25	0.47	4.14×10^{-5}	0.11	3.07×10^{-5}

‡ not enough memory to run sparse direct solver on normal equs.

Null-space approach: results

Also results can deteriorate if number of constraints increases.

Here the constraints are **not dense**.

		$\theta = 1$		$\theta = 0.1$	
Identifier	p	*density*	$\lVert r_c \rVert$	*density*	$\lVert r_c \rVert$
fxm4_6_5	5	0.0005	7.80×10^{-11}	0.0005	1.10×10^{-11}
fxm4_6_20	20	0.0006	5.43×10^{-6}	0.0006	6.67×10^{-7}
stormg2-8_5	5	0.002	1.13×10^{-10}	0.002	8.16×10^{-11}
stormg2-8_20	20	0.003	7.44×10^{-9}	0.002	7.23×10^{-9}

Note: if a sequence of LSE problems is to be solved with **same** set of constraints but different A, null-space basis can be reused, substantially reducing work required.

But if C changes then Z will also change (an updating strategy may be possible to reduce the work).

Method of direct elimination (Björck and Golub 1967)

Find permutation P s.t

$$C\,x = CP\,y = \begin{pmatrix} C_1 & C_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = d,$$

with $p \times p$ matrix C_1 is nonsingular.

Let $AP = \begin{pmatrix} A_1 & A_2 \end{pmatrix}$ be a conformal partitioning.

Substituting $y_1 = C_1^{-1}(d - C_2\,y_2)$ gives transformed LS problem

$$\min_{y_2} \left\| A_T y_2 - (b - A_1\,C_1^{-1} d) \right\|_2^2,$$

with the transformed matrix

$$A_T = A_2 - A_1\,C_1^{-1} C_2 \in \mathbb{R}^{m \times (n-p)}.$$

Method of direct elimination

- If both A and $C_1^{-1} C_2$ are sparse then solve sparse LS problem.
- If C has dense rows then A_T has more dense rows than A.
- In this case, employ solution schemes for sparse-dense LS problems (recent work of Scott and Tůma).
- For example, Cholesky factorization of sparse part of A_T and use as preconditioner within a CG method.
- Construct C_1 using a QR factorization with a threshold parameter τ to balance stability and limiting fill-in in A_T.

Method of direct elimination: results

ndense is number of dense rows in A_T

		$\tau = 1$		$\tau = 0.1$	
Identifier	p	*ndense*	$\|r_c\|$	*ndense*	$\|r_c\|$
scagr7-2b	7	14	4.03×10^{-13}	6	1.48×10^{-13}
scrs8-2c	22	16	6.80×10^{-14}	16	1.25×10^{-13}
sctap1-2b	34	72	7.60×10^{-13}	63	1.31×10^{-12}
gemat1_20	20	284	8.95×10^{-15}	147	1.64×10^{-14}
gemat1_5	5	142	9.77×10^{-14}	17	4.61×10^{-14}
stormg2-8_20	20	136	5.30×10^{-14}	94	2.29×10^{-15}
stormg2-8_5	5	61	3.28×10^{-15}	35	1.50×10^{-14}

- *ndense* increases more rapidly than p.
- Reducing τ can significantly reduce *ndense*, making transformed problem cheaper to solve.
- Small τ does not adversely affect $\|r_c\|$.

Method of direct elimination: weaknesses

- *ndense* can be relatively large $\Longrightarrow$ transformed problem expensive to solve.
- Suppose need to solve a sequence of problems in which A or C is fixed. Coupling of the two blocks in the solution process $\Longrightarrow$ it must be restarted for each problem.

Null-space and direct elimination approaches both face potential issues $\Longrightarrow$ other possibilities?

Consider weighted least squares problem (WLS)

$$\min_x \|A_\gamma\, x_\gamma - b_\gamma\|^2 \quad \text{with} \quad A_\gamma = \begin{pmatrix} A \\ \gamma\, C \end{pmatrix}, \quad b_\gamma = \begin{pmatrix} b \\ \gamma\, d \end{pmatrix},$$

for some large $\gamma \gg 1$.

$$\lim_{\gamma \to \infty} x_\gamma = x_{LSE}$$

where x_{LSE} is the solution of the LSE problem.

Problems: for very large values of γ, normal matrix $A_\gamma^T A_\gamma$ is extremely ill-conditioned.

And if C has dense rows, $A_\gamma^T A_\gamma$ is dense.

Regularized normal equations

$$(A_\gamma^T A_\gamma + \omega^2 I)\, x = A_\gamma^T\, b_\gamma, \qquad \omega > 0.$$

Equivalent to 3-block augmented system

$$\begin{pmatrix} \omega I & 0 & A \\ 0 & \omega I & \gamma C \\ A^T & \gamma C^T & -\omega I \end{pmatrix} \begin{pmatrix} y_s \\ y_c \\ x \end{pmatrix} = \begin{pmatrix} b \\ \gamma d \\ 0 \end{pmatrix}$$

where $\begin{pmatrix} y_s \\ y_c \end{pmatrix} = \omega^{-1} \left[\begin{pmatrix} b \\ \gamma\, d \end{pmatrix} - \begin{pmatrix} A \\ \gamma\, C \end{pmatrix} x \right].$

$$\begin{pmatrix} \omega I & 0 & A \\ 0 & \omega I & \gamma C \\ A^T & \gamma C^T & -\omega I \end{pmatrix} \begin{pmatrix} y_s \\ y_c \\ x \end{pmatrix} = \begin{pmatrix} b \\ \gamma d \\ 0 \end{pmatrix}.$$

Eliminating y_s and setting $\omega\gamma = 1$ gives $(n+p) \times (n+p)$ system

$$\begin{pmatrix} -H(\omega) & C^T \\ C & \omega^2 I \end{pmatrix} \begin{pmatrix} x \\ y_c \end{pmatrix} = \begin{pmatrix} -A^T b \\ d \end{pmatrix}, \quad H(\omega) = A^T A + \omega^2 I.$$

Block signed (incomplete) Cholesky factorization

$$\begin{pmatrix} -H(\omega) & C^T \\ C & \omega^2 I \end{pmatrix} \approx \begin{pmatrix} L & \\ B & L_\omega \end{pmatrix} \begin{pmatrix} -I & \\ & I \end{pmatrix} \begin{pmatrix} L^T & B^T \\ & L_\omega^T \end{pmatrix},$$

$$H(\omega) \approx LL^T, \; LB^T = -C^T, \; S = \omega^2 I + BB^T = L_\omega L_\omega^T.$$

Regularized normal equations: results

We use $\omega\gamma = 1$. GMRES convergence tolerance 10^{-11}

			Direct solver	GMRES	
Identifier	ω	$\|r\|$	$\|r_c\|$	iters	$\|r_c\|$
sctap1-2r	1.0×10^{-3}	2.07×10^{2}	7.38×10^{-3}	6	7.38×10^{-3}
	1.0×10^{-5}	2.07×10^{2}	7.71×10^{-7}	6	7.42×10^{-7}
	1.0×10^{-9}	2.07×10^{2}	1.29×10^{-7}	6	4.09×10^{-13}
south31	1.0×10^{-3}	1.88×10^{2}	8.34×10^{-7}	337	7.34×10^{-7}
	1.0×10^{-5}	1.88×10^{2}	8.31×10^{-11}	354	8.85×10^{-11}
	1.0×10^{-9}	1.88×10^{2}	6.77×10^{-14}	354	1.08×10^{-11}
deter3_20	1.0×10^{-3}	1.22×10^{2}	6.83×10^{-6}	36	6.83×10^{-6}
	1.0×10^{-5}	1.22×10^{2}	6.83×10^{-10}	36	6.83×10^{-10}
	1.0×10^{-9}	1.22×10^{2}	1.35×10^{-12}	36	1.37×10^{-12}

Comments on using augmented system

- Several options for using an augmented system formulation (just one considered here).
- Allows use of "black box" sparse solvers $\Longrightarrow$ greatly reduces work to develop robust and efficient implementations.
- Can be generalised to handle dense rows in A and offers the potential for mixed-precision computation.
- Significant amount of work can be reused when solving sequence of problems in which only C changes.

Concluding remarks

- Large-scale LSEs can be tough to solve, particularly if constraints are dense.
- We need to bring together different techniques and experience with LS problems (especially for mixed sparse-dense case).
- General lack of iterative methods and preconditioners that can be used to extend the size of problems that can be solved.
- Software packages?

Thank you for listening!

Scott and Tůma (2022) *Solving large linear least squares problems with linear equality constraints.*
To appear in BIT. Also `https://arxiv.org/abs/2106.13142`

Papers on relating to our work on sparse LS problems available from `http://www.numerical.rl.ac.uk/people/j_scott/`

My work was partially supported by EPSRC grant EP/I013067/1

Questions or comments?