Deviation Maximization for rank-deficient problems

Monica Dessole[1,2], Fabio Marcuzzi[1]

[1]Dipartimento di Matematica “Tullio Levi Civita”, Università di Padova
[2]Leonardo Labs, Leonardo Company, Genova

Due Giorni di Algebra Lineare Numerica e Applicazioni

14–15 Febbraio 2021

The problem

Consider A matrix of size $m \times n$

- possibly overdetermined $m \geq n$, numerical rank $r < n$
- underdetermined $m < n$, no assumption on the rank

Find $\mathbf{x}^\star$ that solves

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2. \tag{1}$$

The problem

Consider A matrix of size $m \times n$

- possibly overdetermined $m \geq n$, numerical rank $r < n$
- underdetermined $m < n$, no assumption on the rank

Find $\mathbf{x}^\star$ that solves

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2. \tag{1}$$

- Infinitely many solutions: if $\mathbf{x}^\star$ solves (1), then

$$\|A(\mathbf{x}^\star + \mathbf{y}) - \mathbf{b}\|^2 = \|A\mathbf{x}^\star - \mathbf{b}\|^2$$

  for any $\mathbf{y} \in \mathcal{N}(A) = \{\mathbf{x} : \; A\mathbf{x} = 0\} \neq \emptyset$.
- The standard QR may not lead a solution.

Gold standard is the SVD, but it is expensive.

Find $r = \mathrm{rank}(A)$ linearly independent columns of A, namely $\left\{\mathbf{a}_{j_1} \ldots, \mathbf{a}_{j_r}\right\}$, then

$$A\Pi = (Q_1\ Q_2)\begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}, \tag{2}$$

where R_{11} is upper triangular of order r, and Π permutes $\left\{\mathbf{a}_{j_1} \ldots, \mathbf{a}_{j_r}\right\}$ to the left-most positions.

Find $r = \text{rank}(A)$ linearly independent columns of A, namely $\left\{\mathbf{a}_{j_1} \dots, \mathbf{a}_{j_r}\right\}$, then

$$A\Pi = (Q_1 \; Q_2) \begin{pmatrix} R_{11} & R_{12} \\ 0 & 0 \end{pmatrix}, \tag{2}$$

where R_{11} is upper triangular of order r, and Π permutes $\left\{\mathbf{a}_{j_1} \dots, \mathbf{a}_{j_r}\right\}$ to the left-most positions. The associated basic solution is given by

$$\mathbf{x}^\star = \Pi \begin{pmatrix} R_{11}^{-1} Q_1^T \mathbf{b} \\ 0 \end{pmatrix} \tag{3}$$

- it has at most r nonzero entries;
- it depends on the choice of the basis $\left\{\mathbf{a}_{j_1} \dots, \mathbf{a}_{j_r}\right\}$ of $\mathcal{R}(A)$;
- it is not the minimum ℓ_2 solution in general.

A Rank-Revealing QR (RRQR) factorisation is

$$A\Pi = QR = Q\begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{4}$$

- A has numerical rank $r = \text{rank}(A, \varepsilon)$;
- Q is an orthogonal, R_{11} is upper triangular of order r;
- $\sigma_{\min}(R_{11}) \gg \|R_{22}\| = \mathcal{O}(\varepsilon)$.

A Rank-Revealing QR (RRQR) factorisation is

$$A\Pi = QR = Q\begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{4}$$

- A has numerical rank $r = \operatorname{rank}(A, \varepsilon)$;
- Q is an orthogonal, R_{11} is upper triangular of order r;
- $\sigma_{\min}(R_{11}) \gg \|R_{22}\| = \mathcal{O}(\varepsilon)$.

Best we can do is to find a column pivoting Π such that

$$\max_{\Pi} \sigma_{\min}(R_{11}), \tag{5}$$

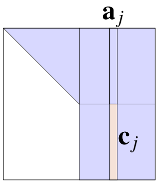
which is NP-hard. Therefore, we solve (5) approximately and we are happy with

$$\sigma_{\min}(R_{11}) \geq \frac{\sigma_r(A)}{p(n)}. \tag{6}$$

First greedy algorithm[1] for approximate solving $\max_\Pi \sigma_{\min}(R_{11})$.

QR with column pivoting (QRP)

1: **for** $s = 1, \ldots, n-1$ **do**
2: Choose j such that $\|\mathbf{c}_j\|$ is maximum
3: Swap columns $s+1$ and $s+j$
4: Compute and apply the Householder reflector
5: **end for**



- Column pivoting is a performance killer
- QP3[2], block version implemented in LAPACK

[1] Peter Businger and Gene H. Golub. "Linear Least Squares Solutions by Householder Transformations". In: *Numer. Math.* 7.3 (June 1965), pp. 269–276. ISSN: 0029-599X.

[2] G. Quintana-Ortí, X. Sun, and C. H. Bischof. "A BLAS-3 Version of the QR Factorization with Column Pivoting". In: *SIAM Journal on Scientific Computing* 19.5 (1998), pp. 1486–1494.

Problem: How to pick $k > 1$ columns at once?

Lemma

$C = (\mathbf{c}_1 \dots \mathbf{c}_k)$. *If there exists* $1 > \tau > 0$ *such that*

- $||\mathbf{c}_j|| \geq \tau ||\mathbf{c}_1|| = \tau \max_i ||\mathbf{c}_i||$, *for all* $1 \leq j \leq k$,
- $C^T C$ *is* τ*-scaled diagonally dominant w.r.t. the* ∞*-norm, i.e.*

$$C^T C = D \Theta D = D(I + N)D, \quad ||N||_\infty < \tau, \tag{7}$$

where D *is diagonal and* Θ *is the correlation matrix,*

then

$$\sigma_{\min}(C) \geq \tau \sqrt{1 - \tau}\, ||\mathbf{c}_1|| > 0. \tag{8}$$

Deviation Maximization (DM)[3]: Pick k indices such that the corresponding columns have

- a large norm w.r.t. to τ, i.e. $\|\mathbf{c}_j\| \geq \tau \max_i \|\mathbf{c}_i\|$,
- large deviations, i.e. pairwise orthogonal columns up to δ:

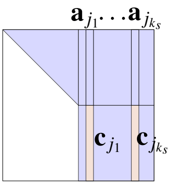
$$\left|\theta_{ij}\right| = \left|\frac{\mathbf{c}_i^T \mathbf{c}_j}{\|\mathbf{c}_i\|\|\mathbf{c}_j\|}\right| < \delta, \quad i \neq j, \tag{9}$$

where θ_{ij} is the cosine of the angle (mod π) between $\mathbf{c}_i$ and $\mathbf{c}_j$ and it is the (i, j)–th entry of the correlation matrix Θ.

[3] M. D. and F. Marcuzzi. "Deviation Maximization for Rank-Revealing QR Factorizations". To appear in Numerical Algorithms. 2021.

QR with Deviation Maximization (QRDM(τ, δ))

while $n_s < n$ **do**
    Choose k_s columns within $\{c_i : ||c_i|| \geq \tau \max_i ||c_i||\}$ pairwise orthogonal up to δ
    Move selected columns in the first k_s leading positions
    Compute and apply the block Householder reflectors
    $s = s + 1$, $n_s = n_s + k_s$
end while



- Naturally based on BLAS–3 operations for efficiency
- Communication avoiding: if a column is already within the the first k_s leading positions, then it is not moved

Worst-case bounds on σ_{min}

Let $\bar{\sigma}^{(s)}$ be the smallest singular value of the computed R_{11} block at step s.

Worst-case bounds on σ_{min}

Let $\bar{\sigma}^{(s)}$ be the smallest singular value of the computed R_{11} block at step s.

Theorem

The standard pivoting guarantees

$$\bar{\sigma}^{(s+1)} \geq \sigma_{s+1}(A) \frac{\bar{\sigma}^{(s)}}{\sigma_1(A)} \frac{1}{\sqrt{2(n-s)(s+1)}}. \tag{10}$$

The DM pivoting guarantees

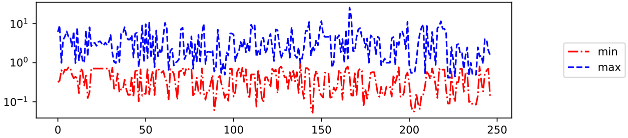
$$\bar{\sigma}^{(s+1)} \geq \sigma_{n_{s+1}}(A) \frac{\bar{\sigma}^{(s)}}{\sigma_1(A)} \frac{1}{\sqrt{2(n-n_{s+1})n_{s+1}}} \frac{\sqrt{\delta+\tau^2-1}}{k^2 n_s}. \tag{11}$$

- Theoretically, the quality of the two RRQR factorizations is similar.

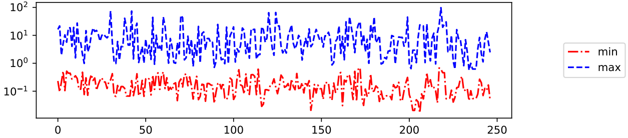
- Subset of San Jose State University singular matrices dataset, $m, n = \mathcal{O}(10^3) - \mathcal{O}(10^4)$
- Double precision codes QRDM vs QP3 (LAPACK)

- Subset of San Jose State University singular matrices dataset, $m, n = \mathcal{O}(10^3) - \mathcal{O}(10^4)$
- Double precision codes QRDM vs QP3 (LAPACK)

Minimum (red) and maximum (blue) values of ratio $\frac{|\text{diag}(R_{11})_i|}{\sigma_i(A)}$ for each A:

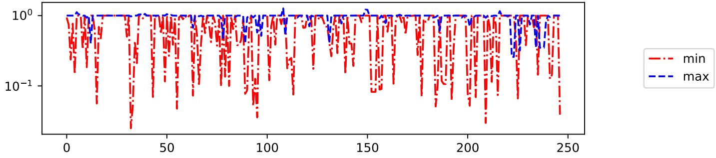


(a) QP3

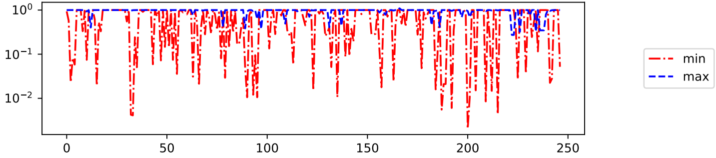


(b) QRDM

Minimum (red) and maximum (blue) values of ratio $\frac{\sigma_i(R_{11})}{\sigma_i(A)}$ for each A:

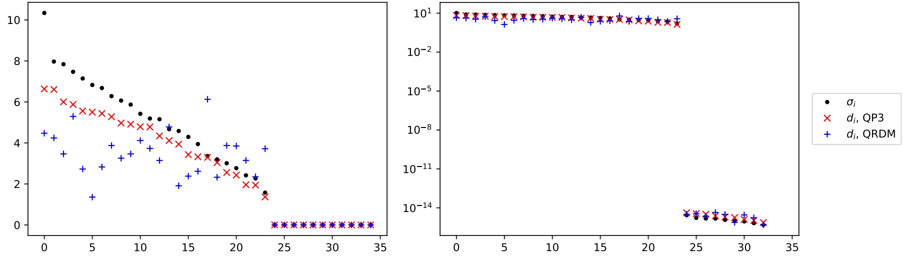


(a) QP3



(b) QRDM

Singular values σ_i $(\cdot)$ and diagonal values $d_i = |\,\mathrm{diag}(R_{11})_i|$ computed with QP3 $(\times)$ and QRDM $(+)$:

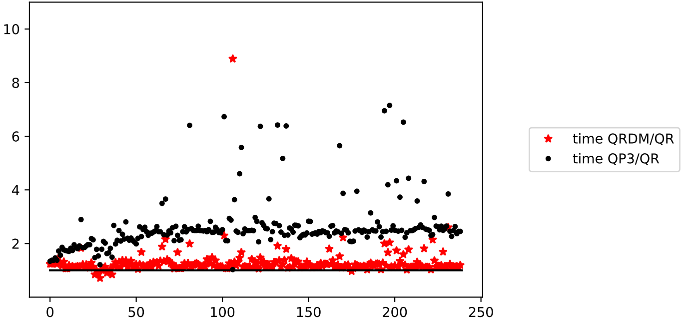


(a) Natural scale

(b) Logarithmic scale

- QP3 d_i's are monotonically non increasing
- QRDM does not show this property

Execution times of QRDM (⋆) and QP3 (·) over QR without pivoting:



- QR is 3× faster than QP3;
- QR is only 1.3× faster than QRDM;
- QRDM is 2.1× faster then QP3.

The proposed Deviation Maximization block pivoting

- is naturally based on BLAS–3 kernels for efficiency;
- can substantially decrease the amount of communication due to column permutation;
- has been successfully extended to NonNegative Least Squares (NNLS) problems[4].

Future perspectives

- investigate the benefits on parallel environments;
- extension to least squares problems with general linear constraints.

[4] M. D., F. Marcuzzi, and M. Vianello. "Accelerating the Lawson-Hanson NNLS solver for large-scale Tchakaloff regression designs". In: *Dolomites Research Notes on Approximation* 13 (1 2020), pp. 20–29.

Thank you for your attention!

Questions?

References I

[1] Peter Businger and Gene H. Golub. "Linear Least Squares Solutions by Householder Transformations". In: *Numer. Math.* 7.3 (June 1965), pp. 269–276. ISSN: 0029-599X.

[2] M. D. and F. Marcuzzi. "Deviation Maximization for Rank-Revealing QR Factorizations". To appear in Numerical Algorithms. 2021.

[3] M. D., F. Marcuzzi, and M. Vianello. "Accelerating the Lawson-Hanson NNLS solver for large-scale Tchakaloff regression designs". In: *Dolomites Research Notes on Approximation* 13 (1 2020), pp. 20–29.

[4] G. Quintana-Ortí, X. Sun, and C. H. Bischof. "A BLAS-3 Version of the QR Factorization with Column Pivoting". In: *SIAM Journal on Scientific Computing* 19.5 (1998), pp. 1486–1494.