CONDITIONING OF LEVERAGE SCORES AND COMPUTATION BY QR DECOMPOSITION[*]

JOHN T. HOLODNAK[†], ILSE C. F. IPSEN[‡], AND THOMAS WENTWORTH[§]

Abstract. The leverage scores of a full-column rank matrix A are the squared row norms of any orthonormal basis for range (A). We show that corresponding leverage scores of two matrices A and $A + \Delta A$ are close in the relative sense if they have large magnitude and if all principal angles between the column spaces of A and $A + \Delta A$ are small. We also show three classes of bounds that are based on perturbation results of QR decompositions. They demonstrate that relative differences between individual leverage scores strongly depend on the particular type of perturbation ΔA. The bounds imply that the relative accuracy of an individual leverage score depends on its magnitude and the two-norm condition of A if ΔA is a general perturbation; the two-norm condition number of A if ΔA is a perturbation with the same normwise row-scaling as A; (to first order) neither condition number nor leverage score magnitude if ΔA is a componentwise row-scaled perturbation. Numerical experiments confirm the qualitative and quantitative accuracy of our bounds.

Key words. principal angles, stable rank, condition number, row-scaling, componentwise perturbations

AMS subject classifications. 65F25, 65F35, 62J20, 68W20, 15A12, 15A23

DOI. 10.1137/140988541

1. Introduction. Leverage scores are scalar quantities associated with the column space of a matrix and can be computed from the rows of *any* orthonormal basis for this space.

Leverage scores. We restrict our discussion here to leverage scores of full-column rank matrices.

DEFINITION 1.1. *Let A be a real $m \times n$ matrix with $\text{rank}(A) = n$. If Q is any $m \times n$ matrix whose columns form an orthonormal basis for* range (A), *then the leverage scores of A are*

$$\ell_j \equiv \left\| e_j^T Q \right\|_2^2, \qquad 1 \leq j \leq m.$$

Here e_j denotes the jth column of the $m \times m$ identity matrix, and $e_j^T Q$ denotes the jth row of Q.

[†]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (jtholodn@ncsu.edu, http://www4.ncsu.edu/~jtholodn/).
[‡]Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (ipsen@ncsu.edu, http://www4.ncsu.edu/~ipsen/).
[§]Koch Institute, Massachusetts Institute of Cambridge, MA 02139 (tawentwo@mit.edu).

Note that leverage scores are independent of the orthonormal basis, since

$$\left\|e_j^T Q\right\|_2^2 = e_j^T Q Q^T e_j = (Q Q^T)_{jj}, \quad 1 \le j \le m,$$

and $Q Q^T$ is the unique orthogonal projector onto range (A).

The basic properties of leverage scores are

$$0 \le \ell_j \le 1, \quad 1 \le j \le m, \quad \text{and} \quad \sum_{j=1}^m \ell_j = n.$$

Hoaglin and Welsch introduced statistical leverage scores in 1978 to detect outliers in regression problems [16, section 2], [17, section 5.1], [28, section 2.2]. About 30 years later, Mahoney, Drineas, and their coauthors started to advocate the use of leverage scores in randomized matrix algorithms [10, 11, 20]. More specifically, leverage scores are the basis for importance sampling strategies in the context of low rank approximations [11], CUR decompositions [12], subset selection [2], Nyström approximations [27], least squares problems [10], and matrix completion [3], to name just a few. Leverage scores also play a crucial role in the analysis of randomized algorithms [17], and fast algorithms have been developed for their approximation [9, 18, 19].

Motivation. Since leverage scores depend only on the column space and are not tied to any particular orthonormal basis, the question is how to compute them. Many existing papers, among them the survey monograph [20, Definition 1], define leverage scores as row norms of a thin left singular vector matrix. However, the sensitivity of singular vectors is determined by the corresponding singular value gaps.

This, and the fact that QR decompositions, when implemented via Householder transformations or Givens rotations, are numerically stable [15, sections 19.1–19.7], motivated us to investigate QR decompositions for the computation of leverage scores. In this paper, we derive bounds on the difference between the leverage scores of a matrix A and a perturbation $A + \Delta A$, when the leverage scores are computed from a QR decomposition. Note that we do not assume a particular implementation of the QR decomposition and assume that quantities are computed in exact arithmetic. We consider our results to be a first step toward determining whether computing leverage scores with a QR decomposition is numerically stable. Since most of our bounds do not exploit the zero structure of the upper triangular factor, they can be readily extended to polar decompositions.

1.1. Overview. We present a short overview of the contents of the paper and the main results. For brevity, we display only the first-order terms in the bounds and omit the technical assumptions.

Notation. The $m \times m$ identity matrix is I_m, with columns e_j and rows e_j^T, $1 \le j \le m$.

For a real $m \times n$ matrix A with rank$(A) = n$, the two-norm condition number with respect to left inversion is $\kappa_2(A) \equiv \|A\|_2 \|A^\dagger\|_2$, where $A^\dagger$ is the Moore–Penrose inverse. The stable rank is sr$(A) \equiv \|A\|_F^2 / \|A\|_2^2$, where sr$(A) \le$ rank(A).

We denote the leverage scores of a perturbed matrix $A + \Delta A$ by $\widetilde{\ell}_j$ and refer to the quantities $|\widetilde{\ell}_j - \ell_j|$ and $|\widetilde{\ell}_j - \ell_j|/\ell_j$ as the the absolute leverage score difference and relative leverage score difference, respectively. We assume, tacitly, that relative leverage score difference bounds $|\widetilde{\ell}_j - \ell_j|/\ell_j$ apply only for $\ell_j > 0$.

Conditioning of leverage scores (section 2). Before even thinking about computation, we need to determine the conditioning of individual leverage scores. To this end, let A and $A + \Delta A$ be real $m \times n$ matrices with $\text{rank}(A) = \text{rank}(A + \Delta A) = n$, and leverage scores ℓ_j and $\widetilde{\ell}_j$, $1 \leq j \leq m$, respectively. We show (Corollary 2.3) that the relative sensitivity of individual leverage scores to subspace rotations is determined by their magnitude. That is, if θ_n is the largest principal angle between $\text{range}(A)$ and $\text{range}(A + \Delta A)$, then

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 2 \sqrt{\frac{1 - \ell_j}{\ell_j}} \, \sin \theta_n + \mathcal{O}\left((\sin \theta_n)^2\right), \qquad 1 \leq j \leq m.$$

Thus, large leverage scores tend to be better conditioned, in the relative sense, to subspace rotations than small leverage scores.

The same holds for general perturbations in the two-norm (Theorem 2.4). If $\epsilon = \|\Delta A\|_2 / \|A\|_2$ is the two-norm of the perturbation, then

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 2 \sqrt{\frac{1 - \ell_j}{\ell_j}} \, \kappa_2(A) \, \epsilon + \mathcal{O}(\epsilon^2), \qquad 1 \leq j \leq m.$$

Therefore, all leverage scores are ill-conditioned under general normwise perturbations if A is ill-conditioned with respect to inversion; in addition, larger leverage scores are better conditioned than smaller ones.

A bound similar to the one above holds also for projected perturbations $\epsilon^\perp = \|\mathcal{P}^\perp \Delta A\|_2 / \|A\|_2$, where $\mathcal{P}^\perp = I_m - A A^\dagger$ is the orthogonal projector onto the orthogonal complement of $\text{range}(A)$. The projected perturbations remove the contribution of ΔA that lies in $\text{range}(A)$. This is important when ϵ is large, but ΔA has only a small contribution in $\text{range}(A)^\perp$. Note that ΔA does not change the leverage scores if $\text{range}(A + \Delta A) = \text{range}(A)$.

Leverage scores computed with a QR decomposition (section 3). With the conditioning of individual leverage scores now established, we present perturbation bounds that represent the first step in assessing the numerical stability of the QR decomposition for computing leverage scores.

Section 3.1. Our first result is a bound derived from existing QR perturbation results that make no reference to a particular implementation. If $\epsilon_F = \|\Delta A\|_F / \|A\|_F$ is the total mass of the perturbation, then the leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 12 \sqrt{\frac{1 - \ell_j}{\ell_j}} \, \text{sr}(A)^{1/2} \, \kappa_2(A) \, \epsilon_F + \mathcal{O}(\epsilon_F^2), \qquad 1 \leq j \leq m.$$

Therefore, if ΔA is a general matrix perturbation, then leverage scores computed from a QR decomposition of $A + \Delta A$ are well-conditioned in the normwise sense, provided they have large magnitude and A is well-conditioned.

Section 3.2. The next bound is derived from scratch and does not rely on existing QR perturbation results. Again, it makes no assumptions on the matrix perturbation ΔA but is able to recognize normwise row-scaling in ΔA. If $\epsilon_j = \|e_j^T \Delta A\|_2 / \|e_j^T A\|_2$, $1 \leq j \leq m$, are normwise perturbations of the rows of A, then the leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy

$$\frac{\left|\widetilde{\ell}_j - \ell_j\right|}{\ell_j} \leq 2 \left(\epsilon_j + \sqrt{2} \, \text{sr}(A)^{1/2} \, \epsilon_F\right) \kappa_2(A) + \mathcal{O}(\epsilon_F^2), \qquad 1 \leq j \leq m.$$

The perturbation ϵ_j represents the *local* effect of ΔA, because it indicates how the jth relative leverage score difference depends on the perturbation in row j of A. In contrast, ϵ_F, containing the total mass of the perturbation, represents the *global* effect on all leverage scores.

A similar bound holds for projected perturbations $\epsilon_F^\perp = \|\mathcal{P}^\perp A\|_F/\|A\|_F$ and $\epsilon_j^\perp = \|e_j^T \mathcal{P}^\perp \Delta A\|_2/\|e_j^T A\|_2$, $1 \le j \le m$.

Section 3.3. The natural follow-up question is: What if ΔA does indeed represent a row-scaling of A? Can we get tighter bounds? The answer is yes. If $|e_j^T \Delta A| \le \eta_j \, |e_j^T A|$, $1 \le j \le m$, with $\eta = \max_{1 \le j \le m} \eta_j$, are componentwise row-scaled perturbations, then the leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy

$$\frac{\left|\widetilde{\ell}_j - \ell_j\right|}{\ell_j} \le 2 \left(\eta_j + \sqrt{2}\, n\, \eta\right) + \mathcal{O}(\eta^2), \qquad 1 \le j \le m.$$

Thus, under componentwise row-scaled perturbations, leverage scores computed with a QR decomposition have relative leverage score differences that depend, to first order, neither on the condition number nor on the magnitudes of the leverage scores.

Numerical experiments (sections 2 and 3). After each of the bounds presented in sections 2 and 3, we perform numerical experiments that illustrate that the bounds correctly capture the relative leverage score differences under different types of perturbations.

Summary (section 4). We summarize the results in this paper and describe a few directions for future research.

Appendix (section A). We present the proofs for all results in sections 2 and 3.

2. Conditioning of leverage scores. We determine the absolute and relative sensitivity of leverage scores to rotations of the column space (section 2.1) and to general matrix perturbations in the two-norm (section 2.2).

2.1. Principal angles. We show that the leverage scores of two matrices are close in the absolute sense if the all angles between their column spaces are small (Theorem 2.2). Larger leverage scores tend to be better conditioned in the relative sense (Corollary 2.3).

Principal angles between two subspaces quantify the distance between the spaces in "every dimension."

DEFINITION 2.1 (section 6.4.3 in [13]). *Let A and ΔA be real $m \times n$ matrices with* $\mathrm{rank}(A) = \mathrm{rank}(A + \Delta A) = n$, *and let Q and $\widetilde{Q}$ be orthonormal bases for* $\mathrm{range}\,(A)$ *and* $\mathrm{range}\,(A + \Delta A)$, *respectively.*

Let $Q^T \widetilde{Q} = U \Sigma V^T$ be an SVD, where U and V are $n \times n$ orthogonal matrices, and $\Sigma = \mathrm{diag}\left(\cos\theta_1 \quad \cdots \quad \cos\theta_n\right)$ is an $n \times n$ diagonal matrix with $1 \ge \cos\theta_1 \ge \cdots \ge \cos\theta_n \ge 0$. Then, $0 \le \theta_1 \le \cdots \le \theta_n \le \pi/2$ are the principal angles between the column spaces of $\mathrm{range}\,(A)$ *and* $\mathrm{range}\,(A + \Delta A)$.

Below we bound the absolute leverage score difference in terms of the largest and smallest principal angles.

THEOREM 2.2 (absolute leverage score difference). *Let A and ΔA be real $m \times n$ matrices, with* $\mathrm{rank}(A) = \mathrm{rank}(A + \Delta A) = n$. *Then,*

$$|\widetilde{\ell}_j - \ell_j| \le 2\sqrt{\ell_j(1 - \ell_j)}\, \cos\theta_1 \sin\theta_n + (\sin\theta_n)^2, \qquad 1 \le j \le m.$$

If, in addition, $m = 2n$, then also

$$1 - \left(\sin \theta_n \sqrt{\ell_j} + \cos \theta_1 \sqrt{1 - \ell_j} \right)^2 \leq \widetilde{\ell}_j \leq \left(\cos \theta_1 \sqrt{\ell_j} + \sin \theta_n \sqrt{1 - \ell_j} \right)^2.$$

Proof. See section A.1.   □

Theorem 2.2 implies that the leverage scores of A and $A + \Delta A$ are close in the absolute sense if the principal angles between their column spaces are small. Theorem 2.2 holds with equality if A and $A + \Delta A$ have the same column space, because then the smallest angle θ_n is zero, and so is the bound.

In the special case $m = 2n$, better bounds are possible because range (A) and its orthogonal complement range $(A)^{\perp}$ have the same dimension. In addition to implying $\widetilde{\ell}_j = \ell_j$ for range $(A) = $ range $(A + \Delta A)$, Theorem 2.2 also implies $\widetilde{\ell}_j = 1 - \ell_j$ for range $(A + \Delta A) = $ range $(A)^{\perp}$, $1 \leq j \leq m$.

Next is a bound for the relative leverage score difference in terms of principal angles.

COROLLARY 2.3 (relative leverage score difference). *Under the conditions of Theorem 2.2,*

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 2 \sqrt{\frac{1 - \ell_j}{\ell_j}} \cos \theta_1 \sin \theta_n + \frac{(\sin \theta_n)^2}{\ell_j}, \qquad 1 \leq j \leq m.$$

Corollary 2.3 implies that the relative sensitivity of leverage scores to rotations of range (A) depends on the magnitude of the leverage scores. In particular, larger leverage scores tend to be better conditioned.

Numerical experiments: Figure 1. We illustrate the effect of subspace rotations on the relative leverage score differences $|\widetilde{\ell}_j - \ell_j|/\ell_j$. We compute the leverage scores of a matrix A and a perturbation $A + \Delta A$. The matrix A has dimension 1000×25, $\kappa_2(A) = 1$, and leverage scores that increase in four steps, from 10^{-10} to about 10^{-1}; see Figure 1(a). It is generated with the MATLAB commands

$$(2.1) \qquad A1 = \mathrm{diag}\left(I_{250} \quad 10^2\, I_{250} \quad 10^3\, I_{250} \quad 10^4\, I_{250} \right) \; \mathsf{randn}(1000, 25)$$
$$[A, \sim] = \mathsf{qr}(A1, 0).$$

The leverage scores of the perturbed matrix $A + \Delta A$ are computed with the MATLAB QR decomposition $\mathsf{qr}(A + \Delta A, 0)$.

The perturbations in the following sections are chosen so that they are large enough to dominate the round-off errors.

Figure 1(b)–(d) shows the relative sensitivities and the bound from Corollary 2.3 for perturbations due to increasing principal angles $\sin \theta_n \approx 10^{-8}, 10^{-6}, 10^{-4}$. The relative leverage score differences decrease with the same step size with which the leverage score magnitude increases.

In Figure 1(b), where $\sin \theta_n \approx 10^{-8}$, the relative leverage score differences decrease from 10^{-5} for the smallest leverage scores to about 10^{-9} for the largest leverage scores. The differences are larger by a factor of 100 in Figure 1(c), and again in Figure 1(d), where the 250 smallest leverage scores have lost all accuracy because they are dominated by the perturbation $\sin \theta_n$. Thus, the relative changes in leverage scores are proportional to $\sin \theta_n$.

Corollary 2.3 shows the same qualitative behavior as the leverage score differences. The bound decreases with the leverage score magnitude and overestimates the worst-case differences by a factor of about 100. Thus, Corollary 2.3 represents a realistic
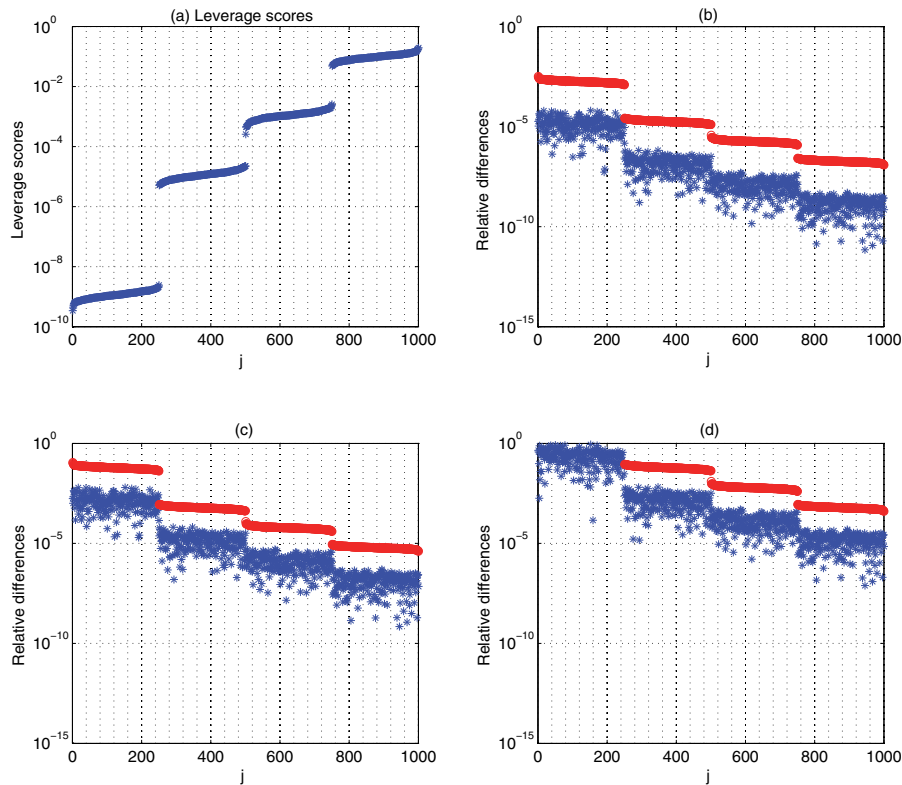
FIG. 1. (a) *Leverage scores* ℓ_j *for matrices in* (2.1); (b)–(d) *relative leverage score differences* $|\widetilde{\ell}_j - \ell_j|/\ell_j$ *(blue stars) and bound for Corollary* 2.3 *(red line above the stars) versus index* j *for* $\sin\theta_n = 10^{-8}$ (b), 10^{-6} (c), *and* 10^{-4} (d) *(Color figures are available online).*

estimate for the relative conditioning of the leverage scores to changes in principal angles.

2.2. General matrix perturbations in the two-norm. From the bounds for principal angles in section 2.1, we derive bounds for the relative leverage score differences in terms of general perturbations in the two-norm,

$$\epsilon \equiv \frac{\|\Delta A\|_2}{\|A\|_2}, \qquad \epsilon^\perp \equiv \frac{\|(I_m - AA^\dagger)\,\Delta A\|_2}{\|A\|_2}.$$

The second perturbation removes the contribution of ΔA that lies in range (A). This is important when ϵ is large, but ΔA has only a small contribution in range (A). Note that ΔA does not change the leverage scores if range $(A + \Delta A) = $ range (A).

THEOREM 2.4. *Let A and ΔA be real $m \times n$ matrices, with* $\mathrm{rank}(A) = n$ *and* $\|\Delta A\|_2 \left\|A^\dagger\right\|_2 \leq 1/2$. *Then*

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 4 \left(\sqrt{\frac{1 - \ell_j}{\ell_j}} + \frac{\kappa_2(A)}{\ell_j}\, \epsilon^\perp \right) \kappa_2(A)\, \epsilon^\perp,$$

and

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq \left(2\sqrt{\frac{1-\ell_j}{\ell_j}} + \frac{\kappa_2(A)}{\ell_j}\,\epsilon\right)\kappa_2(A)\,\epsilon, \qquad 1 \leq j \leq m.$$

Proof. See section A.2. □

Theorem 2.4 implies that relative leverage score differences are bounded by the condition number $\kappa_2(A)$, a normwise perturbation, and a function that depends on the size of the leverage scores. Thus, an individual leverage score is well-conditioned if it has large magnitude and if A is well-conditioned with respect to left inversion.

The first bound in Theorem 2.4 recognizes, through the use of the projected perturbation $\epsilon^\perp$, when the column spaces of A and $A + \Delta A$ are close. In particular, $\epsilon^\perp = 0$ for range $(A) =$ range $(A + \Delta A)$. The second bound does not do this but has the advantage of being simpler. Although the first bound contains a smaller perturbation, $\epsilon^\perp \leq \epsilon$, it also has an additional factor of 2. Therefore, it is not clear that, in general, the first bound is tighter than the second one.

Numerical experiments: Figure 2. We demonstrate that both bounds capture the qualitative behavior of the leverage score sensitivities but that the first bound appears more accurate when the perturbation has a substantial contribution in range (A).

Figure 2 shows the relative leverage score sensitivities and the bounds from Theorem 2.4 for perturbations $\epsilon = \|\Delta A\|_2/\|A\|_2$ and their projections $\epsilon^\perp = \|(I - AA^\dagger)\Delta A\|_2/\|A\|_2$. Panels (a) and (b) in Figure 2 show the leverage scores for perfectly conditioned matrices A constructed as in (2.1), and for matrices B with $\kappa_2(B) \approx 10^5$ and leverage scores like those of A,

(2.2) $\quad B = \mathrm{diag}\begin{pmatrix} I_{250} & 10^2\,I_{250} & 10^3\,I_{250} & 10^4\,I_{250} \end{pmatrix}$ gallery$('$randsvd$', [m, n], 10^6, 3)$.

Panels (c)–(f) in Figure 2 show the relative leverage score differences for A and B under two-norm perturbations $\epsilon = 10^{-8}$.

Panels (c) and (e) show that the relative leverage score differences from the well-conditioned matrix A reflect the leverage score distribution. That is, the smallest leverage scores have relative differences of about 10^{-10}, while the largest leverage scores have relative differences that are several magnitude lower. The relative leverage score differences of the worse conditioned matrix B in panels (d) and (f) can be as high as 10^{-7} and do not follow the leverage distribution quite as clearly. Therefore, the relative leverage score differences from normwise perturbations increase with the condition number of the matrix.

Panels (c) and (d) show the bound with ϵ in Theorem 2.4, while (e) and (f) show the bound for the projected perturbation $\epsilon^\perp \approx 6 \cdot 10^{-14}$ for both matrices. Note that, as explained above, the bounds with the projected perturbations are not *guaranteed* to be tighter, although in this case they are several orders of magnitude more accurate.

3. Leverage scores computed with a QR decomposition. We derive bounds for relative leverage score differences for leverage scores that are computed with a QR decomposition. The bounds assume exact arithmetic and are based on perturbation results for QR decompositions; they make no reference to particular QR implementations.

Specifically, our bounds include normwise bounds for general matrix perturbations (section 3.1), bounds for general perturbations that recognize row-scaling in the
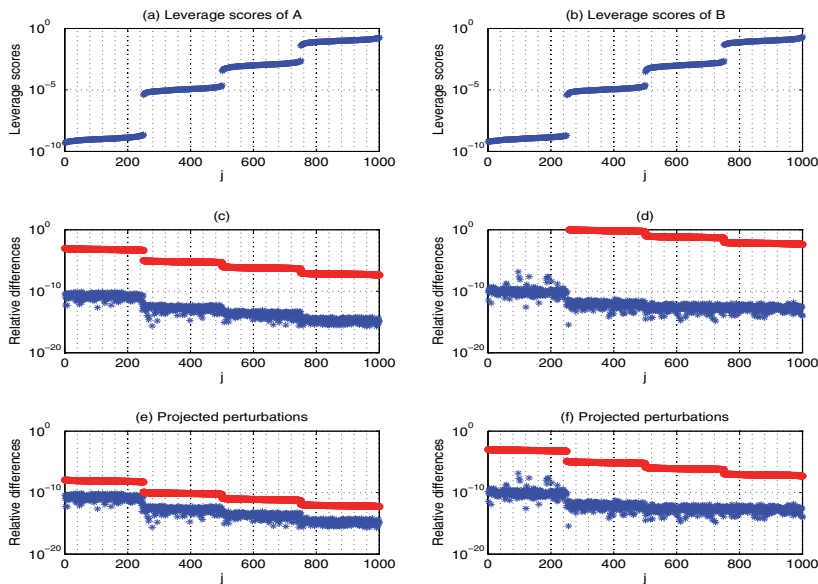
FIG. 2. (a)–(b) *Leverage scores for matrices A in* (2.1) *and B in* (2.2); (c)–(f) *relative leverage score differences* $|\widetilde{\ell}_j - \ell_j|/\ell_j$ *(blue stars) and bounds from Theorem* 2.4 *(red line above the stars) versus index j, for perturbations* $\epsilon = 10^{-8}$ *in* (c) *and* (d)*, and* $\epsilon^\perp \approx 10^{-14}$ *in* (e) *and* (f) *(Color figures are available online).*

perturbations (section 3.2), and bounds for componentwise row-scaled perturbations (section 3.3). Since the bounds do not exploit the zero structure of the triangular factor in the QR decomposition, they can be readily extended to the polar decomposition as well.

3.1. General normwise perturbations. The first bound is derived from a normwise perturbation result for QR decompositions [24, Theorem 1.6]. Among the existing and sometimes tighter QR perturbation bounds [1, 4, 5, 6, 7, 21, 22, 25, 26, 31], we chose [24, Theorem 1.6] because it is simple and has the required key ingredients.

THEOREM 3.1. *Let A and* $A + \Delta A$ *be real* $m \times n$ *matrices with* $\mathrm{rank}(A) = n$ *and* $\|\Delta A\|_2 \|A^\dagger\| \leq 1/2$. *The leverage scores* $\widetilde{\ell}_j$ *computed from a QR decomposition of* $A + \Delta A$ *satisfy*

$$\frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 12 \left(\sqrt{\frac{1 - \ell_j}{\ell_j}} + 3\, \frac{\kappa_2(A)\, \mathrm{sr}\,(A)^{1/2}}{\ell_j}\, \epsilon_F \right) \kappa_2(A)\, \mathrm{sr}\,(A)^{1/2}\, \epsilon_F, \quad 1 \leq j \leq m.$$

Proof. See section A.3. ☐

The perturbation bound in Theorem 3.1 sends the message that if ΔA is a general perturbation, then leverage scores computed from a QR decomposition of $A + \Delta A$ are well-conditioned in the normwise relative sense if they have large magnitude and if A is well-conditioned. We demonstrate that this conclusion is valid in the following experiment.
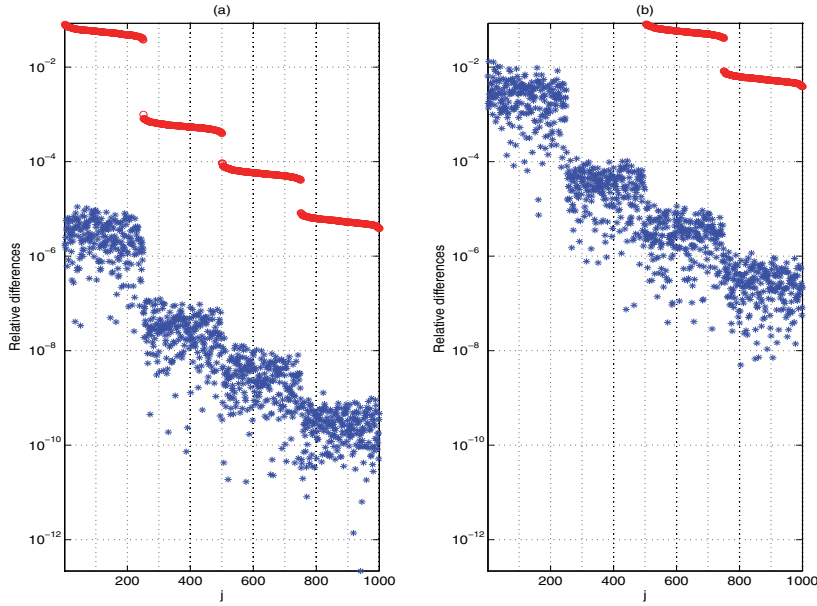
FIG. 3. *Relative leverage score differences $|\widetilde{\ell}_j - \ell_j|/\ell_j$ (blue stars) and bound from Theorem 3.1 (red line above the stars) versus index j for $\epsilon_F = 10^{-8}$ (a) and $\epsilon_F = 10^{-5}$ (b) (Color figures are available online).*

Numerical experiments: Figure 3. For matrices A in (2.1), Figure 3 shows the relative leverage score differences $|\widetilde{\ell}_j - \ell_j|/\ell_j$ from normwise perturbations $\epsilon_F = \|\Delta A\|_F / \|A\|_F$ and the bound from Theorem 3.1 for two different perturbations: $\epsilon_F = 10^{-8}$ and $\epsilon_F = 10^{-5}$.

Figure 3 illustrates that the relative leverage score differences decrease with the same step size with which the leverage score magnitude increases. In particular, for $\epsilon_F = 10^{-8}$ in panel (a), the relative leverage score differences decrease from 10^{-5} for the smallest leverage scores to about 10^{-9} for the largest leverage scores. The differences for $\epsilon_F = 10^{-5}$ in panel (b) are larger by a factor of 1000; the 250 smallest leverage scores have lost all accuracy because they are smaller than the perturbation ϵ_F.

The bound in Theorem 3.1 differs from the actual differences by several orders of magnitude but reflects the qualitative behavior of the relative leverage score differences.

3.2. General normwise perturbation bounds that detect row-scaling in the perturbations. The two first-order bounds presented here are based on a perturbation of the QR decomposition. Although the bounds make no assumptions on the perturbations ΔA, they are able to recognize row-scaling in ΔA of the form

$$\epsilon_j \equiv \frac{\|e_j^T \Delta A\|_2}{\|e_j^T A\|_2}, \qquad 1 \le j \le m.$$

THEOREM 3.2. *Let A and $A + \Delta A$ be real $m \times n$ matrices with $\mathrm{rank}(A) = n$ and $\|\Delta A\|_2 \|A^\dagger\|_2 < 1$. The leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy*

$$\frac{\left|\widetilde{\ell}_j - \ell_j\right|}{\ell_j} \leq 2 \left(\epsilon_j + \sqrt{2}\operatorname{sr}(A)^{1/2}\epsilon_F\right)\kappa_2(A) + \mathcal{O}(\epsilon_F^2), \qquad 1 \leq j \leq m.$$

Proof. See section A.4. ∎

The relative leverage score difference bound for the jth leverage score in Theorem 3.2 contains three main ingredients:

1. The two-norm condition number of A with respect to left inversion, $\kappa_2(A)$. This indicates that leverage scores computed from matrices with smaller condition numbers have smaller relative leverage score differences.
2. The relative normwise perturbation in the jth row of A, ϵ_j. This perturbation represents the *local* effect of ΔA, because it shows how the jth relative leverage score difference depends on the perturbation in row j of A.
3. The total normwise perturbation ϵ_F. This is the total relative mass of the perturbation, since

$$\epsilon_F^2 = \sum_{i=1}^m \left\|e_i^T \Delta A\right\|_2^2 / \|A\|_F^2$$

represents the *global* effect of ΔA.

Numerical experiments: Figure 4. We illustrate that the local effect described above is real by examining the effect of row-scaled perturbations on the relative accuracy of leverage scores computed with a QR decomposition.

Figure 4 shows the relative leverage score difference $|\widetilde{\ell}_j - \ell_j|/\ell_j$ from normwise perturbations $\epsilon_F = \|\Delta A\|_F/\|A\|_F = 10^{-8}$ and the bound from Theorem 3.2. In panel (a), only rows 501–750 of A are perturbed, while in panel (b) the perturbation has the same row-scaling as A, that is, $\Delta A = 10^{-8}A1/\|A1\|_F$, where $A1$ is of the form (2.1).

In panel (a), the leverage scores corresponding to rows 1–500 and 751–1000 have relative leverage score differences between 10^{-12} and 10^{-10}, which illustrates that the local perturbation in rows 501–750 has a global effect on all leverage scores. However, the leverage scores corresponding to the perturbed rows 501–750 have larger relative differences of 10^{-8} or more, which illustrates the strong effect of local perturbations. The bound from Theorem 3.2 hovers around 10^{-7} but is slightly larger for the leverage scores corresponding to the perturbed rows. Thus, Theorem 3.2 is able to detect strongly local row-scaling in normwise perturbations.

In panel (b), almost all leverage scores have relative differences between 10^{-10} and 10^{-9}, and the bound from Theorem 3.2 is flat at 10^{-7}. Thus, the relative leverage score differences tend to be more uniform when the normwise perturbations have the same row-scaling as the matrix. This effect is recognized by Theorem 3.2.

Therefore, although Theorem 3.2 makes no assumptions about the perturbations ΔA, it is able to detect row-scaling in normwise perturbations and correctly predicts the qualitative behavior of relative leverage score differences.

Projected perturbations. The following bound is a refinement of Theorem 3.2 that projects out the part of the perturbation that lies in range(A) and does not contribute to a change in leverage scores:

$$\epsilon_F^\perp \equiv \frac{\|(I_m - AA^\dagger)\Delta A\|_F}{\|A\|_F}, \qquad \epsilon_j^\perp \equiv \frac{\|e_j^T(I - AA^\dagger)\Delta A\|_2}{\|e_j^T A\|_2}, \qquad 1 \leq j \leq m.$$
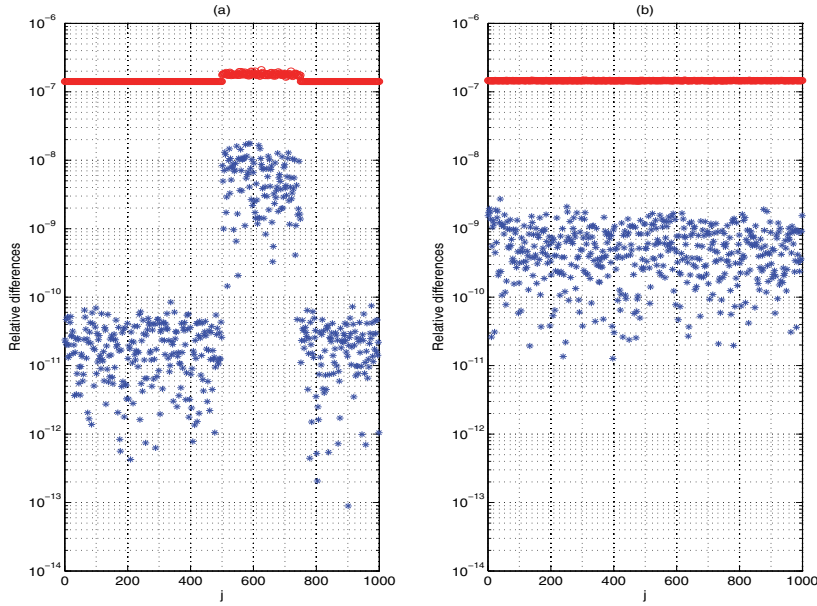
Fig. 4. *Relative leverage score difference $|\widetilde{\ell}_j - \ell_j|/\ell_j$ (blue stars) and bound from Theorem 3.2 (red line above the stars) versus index j for rowwise scaled perturbations with $\epsilon_F = 10^{-8}$. In (a) only rows 501–750 of A are perturbed, while in (b) the perturbation has the same row-scaling as A (Color figures are available online).*

THEOREM 3.3 (projected perturbations). *Let A and $A + \Delta A$ be real $m \times n$ matrices with $\mathrm{rank}(A) = n$ and $\|\Delta A\|_2 \|A^\dagger\|_2 \le 1/2$. The leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy*

$$\frac{\left|\widetilde{\ell}_j - \ell_j\right|}{\ell_j} \le 4 \left(\epsilon_j^\perp + \sqrt{2}\,\mathrm{sr}\,(A)^{1/2}\,\epsilon_F^\perp\right) \kappa_2(A) + \mathcal{O}\left((\epsilon_F^\perp)^2\right), \qquad 1 \le j \le m.$$

Proof. See section A.5.   □

It is not clear that Theorem 3.3 is tighter than Theorem 3.2. First, Theorem 3.3 contains an additional factor of 2 in the bound. Second, although the total projected perturbation is smaller, i.e., $\epsilon_F^\perp \le \epsilon_F$, this is not necessarily true for $\epsilon_j^\perp$ and ϵ_j. For instance, if

$$A = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad \Delta A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

then

$$(I - AA^\dagger)\,\Delta A = (I - AA^T)\,\Delta A = \tfrac{1}{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Here we have $\epsilon_3 = \left\| e_3^T \Delta A \right\|_2 / \left\| e_3^T A \right\|_2 = 0$ and $\epsilon_3^{\perp} = \left\| e_3^T (I - AA^{\dagger}) \Delta A \right\|_2 / \left\| e_3^T A \right\|_2 = 1$, so that $\epsilon_3^{\perp} > \epsilon_3$.

3.3. Componentwise row-scaled perturbations. Motivated by section 3.2, where bounds for general perturbations ΔA can recognize row-scaling in ΔA, we ask the natural follow-up question: What if ΔA does indeed represent a row-scaling of A? Can we get tighter bounds? To this end, we consider componentwise row perturbations of the form $|e_j^T \Delta A| \leq \eta_j \, |e_j^T A|$, where $\eta_j \geq 0$, $1 \leq j \leq m$, and model them as

$$(3.1) \qquad e_j^T \Delta A = \zeta_j \, \eta_j \, e_j^T A, \qquad 1 \leq j \leq m, \qquad \eta \equiv \max_{1 \leq j \leq m} \eta_j,$$

where ζ_j are uniform random variables in $[-1, 1]$, $1 \leq j \leq m$. We show that, under componentwise row-scaled perturbations (3.1), leverage scores computed with a QR decomposition have relative differences that do not depend, to first order, on the condition number or the magnitudes of the leverage scores.

THEOREM 3.4. *Let A be a real $m \times n$ matrix with $\mathrm{rank}(A) = n$, and let the perturbations ΔA be of the form* (3.1) *with $\eta \, \kappa_2(A) < 1$. The leverage scores $\widetilde{\ell}_j$ computed from a QR decomposition of $A + \Delta A$ satisfy*

$$\frac{\left| \widetilde{\ell}_j - \ell_j \right|}{\ell_j} \leq 2 \left(\eta_j + \sqrt{2} \, n \, \eta \right) + \mathcal{O}(\eta^2), \qquad 1 \leq j \leq m.$$

Proof. See section A.6. $\square$

The quantities η_j represent the local effects of the individual rowwise perturbations, while the factor $n \, \eta$ represents the global effect of all perturbations. In contrast to our previous results, the bound depends on neither the condition number nor the leverage score magnitude.

Numerical experiments: Figure 5. We illustrate the effect of componentwise row-scaled perturbations on the relative accuracy of leverage scores that are computed with a QR decomposition.

Figure 5 shows the relative leverage score differences from a well-conditioned matrix A with $\kappa_2(A) = 1$ in (a), and from a worse conditioned matrix B in (2.2) with $\kappa_2(B) \approx 10^5$ in (b). The componentwise row-scaled perturbations from (3.1) are $\eta = \eta_j = 10^{-8}$ for $1 \leq j \leq m$. The leverage scores for these types of matrices are shown in Figure 2.

Figure 5 shows that the relative leverage score differences for both matrices look almost the same, hovering around 10^{-8}, except for a few outliers. Thus, the relative accuracy of most leverage scores does not depend on the condition number, but a few small leverage scores do show a slight effect. Note that Theorem 3.3 is based only on a perturbation analysis, not a round-off error analysis of the QR decomposition, and that we did not take into account errors arising in the computation of the two-norm.

Furthermore, Figure 5 shows that the relative leverage score differences do not depend on the leverage score magnitude. Hence Theorem 3.3 captures the relative leverage score accuracy under componentwise row-scaled perturbations.

4. Summary. We analyzed the conditioning of individual leverage scores (section 2) and took the first steps in assessing the numerical stability of QR decompositions for computing leverage scores (section 3). To this end, we derived several bounds for the relative accuracy of individual leverage scores. The bounds are expressed in
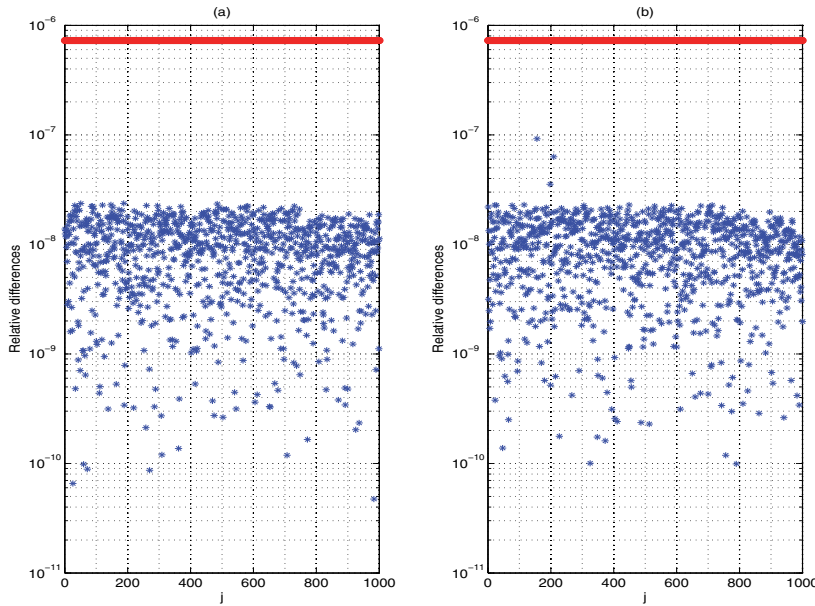
FIG. 5. *Relative leverage score differences $|\widetilde{\ell}_j - \ell_j|/\ell_j$ (blue stars) and the bound from Theorem 3.3 (red line above the stars) versus index j for componentwise rowwise scaled perturbations with $\eta_j = 10^{-8}$, $1 \le j \le m$ (Color figures are available online).*

terms of principal angles between column spaces and three classes of matrix perturbations: General normwise, normwise row-scaled, and componentwise row-scaled.

Since most of the bounds in section 3 do not exploit the zero structure of the upper triangular factor, they are readily extended to polar decompositions as well.

Future research. The next step is to extend the results in section 3.3 to componentwise perturbations $|\Delta A_{jk}| \le \eta_{jk}|A_{jk}|$, $1 \le j \le m$, $1 \le k \le n$. Numerical experiments strongly suggest that leverage scores computed from QR decompositions of such perturbed matrices have relative leverage score differences that do not depend on the magnitude of the leverage scores.

The most popular method for computing leverage scores is the singular value decomposition. The numerical stability of the SVD in this context needs to be investigated, and whether the sensitivity of the singular vectors to singular value gaps matters for leverage score computations.

Another issue is the numerically stable computation of "k-leverage scores." These are leverage scores of the best rank k approximation to A in the two-norm. Determining leverage scores from a truncated SVD is necessary when A is (numerically) rank deficient, or when noisy data are well represented, as in the case of PCA, by only a few dominant singular vectors.

Appendix A. Proofs. We present proofs for the results in sections 2 and 3.

A.1. Proof of Theorem 2.2. The full-column rank of A and $A + \Delta A$ assures that the leverage scores are well defined according to Definition 1.1.

The proof proceeds in three stages: Expressing the perturbed leverage scores $\widetilde{\ell}_j$ in terms of the exact leverage scores ℓ_j; an upper bound for $\widetilde{\ell}_j - \ell_j$; and a lower bound for $\widetilde{\ell}_j - \ell_j$.

1. *Expressing the perturbed leverage scores in terms of the exact ones.* Consider any orthonormal basis for the column spaces: Let $A = QX$, where X is nonsingular and $Q^T Q = I_n$. Similarly, let $A + \Delta A = \widetilde{Q}\widetilde{X}$, where $\widetilde{X}$ is nonsingular and $\widetilde{Q}^T\widetilde{Q} = I_n$. The leverage scores are $\ell_j = \|e_j^T Q\|_2^2$ and $\widetilde{\ell}_j = \|e_j^T \widetilde{Q}\|_2^2$, $1 \le j \le m$.

With Definition 2.1, rotate to the basis of principal vectors $Q_1 \equiv QU$ and $\widetilde{Q}_1 \equiv \widetilde{Q}V$, which satisfy $Q_1^T \widetilde{Q}_1 = \Sigma$. Since the leverage scores are basis independent, we can write $\ell_j = \|e_j^T Q_1\|_2^2$ and $\widetilde{\ell}_j = \|e_j^T \widetilde{Q}_1\|_2^2$, $1 \le j \le m$.

The goal is to express $\widetilde{Q}_1$ in terms of Q_1. To this end choose Q_2 so that $\mathcal{Q} \equiv \begin{pmatrix} Q_1 & Q_2 \end{pmatrix}$ is an $m \times m$ orthogonal matrix. Then $I_m = \mathcal{Q}\mathcal{Q}^T = Q_1 Q_1^T + Q_2 Q_2^T$ implies

$$(A.1) \quad \|e_j^T Q_2\|_2^2 = e_j^T Q_2 Q_2^T e_j = e_j^T e_j - e_j^T Q_1 Q_1^T e_j = 1 - \|e_j^T Q_1\|_2^2 = 1 - \ell_j.$$

Write

$$\mathcal{Q}^T \widetilde{Q}_1 = \begin{pmatrix} Q_1^T \widetilde{Q}_1 \\ Q_2^T \widetilde{Q}_1 \end{pmatrix} = \begin{pmatrix} \Sigma \\ Z \end{pmatrix}, \qquad \text{where} \quad Z \equiv Q_2^T \widetilde{Q}_1.$$

This gives the desired expressions $\widetilde{Q}_1 = Q_1 \Sigma + Q_2 Z$ and

$$\widetilde{Q}_1 \widetilde{Q}_1^T = Q_1 \Sigma^2 Q_1^T + Q_2 Z \Sigma Q_1^T + Q_1 \Sigma Z^T Q_2^T + Q_2 Z Z^T Q_2^T.$$

From $\mathcal{Q}^T \widetilde{Q}_1$ having orthonormal columns follows $Z^T Z = I_n - \Sigma^2$ and

$$(A.2) \qquad \qquad \|Z\|_2 = \sqrt{\|I_n - \Sigma^2\|_2} = \sqrt{1 - (\cos\theta_n)^2} = \sin\theta_n.$$

At last, we can express the perturbed leverage scores in terms of the exact ones,

$$(A.3) \;\; \widetilde{\ell}_j = e_j^T \widetilde{Q}_1 \widetilde{Q}_1^T e_j = e_j^T Q_1 \, \Sigma^2 \, Q_1^T e_j + 2 e_j^T Q_2 \, Z\Sigma \, Q_1^T e_j + e_j^T Q_2 \, Z Z^T \, Q_2^T e_j.$$

2. *Upper bound.* Applying (A.1) and (A.2) in (A.3) gives

$$\widetilde{\ell}_j \le \|\Sigma\|_2^2 \, \ell_j + 2\|e_j^T Q_2\| \, \|Z\|_2 \|\Sigma\|_2 \, \sqrt{\ell_j} + \|e_j^T Q_2\|_2^2 \|Z\|_2^2$$

$$(A.4) \qquad \le (\cos\theta_1)^2 \ell_j + 2\cos\theta_1 \sin\theta_n \, \sqrt{\ell_j(1 - \ell_j)} + (\sin\theta_n)^2 \, (1 - \ell_j)$$

$$= \left((\cos\theta_1)^2 - (\sin\theta_n)^2\right) \ell_j + 2\cos\theta_1 \sin\theta_n \, \sqrt{\ell_j(1 - \ell_j)} + (\sin\theta_n)^2.$$

Subtracting ℓ_j from both sides and omitting the summand with the negative sign gives

$$\widetilde{\ell}_j - \ell_j \le -\left((\sin\theta_1)^2 + (\sin\theta_n)^2\right) \ell_j + 2\cos\theta_1 \sin\theta_n \, \sqrt{\ell_j(1 - \ell_j)} + (\sin\theta_n)^2$$

$$\le 2\cos\theta_1 \sin\theta_n \, \sqrt{\ell_j(1 - \ell_j)} + (\sin\theta_n)^2.$$

If $m = 2n$, write (A.4) as

$$\widetilde{\ell}_j \le \left(\cos\theta_1 \sqrt{\ell_j} + \sin\theta_j \sqrt{1 - \ell_j}\right)^2.$$

3. *Lower bound.* Applying (A.1) and (A.2) in (A.3) gives

$$\widetilde{\ell}_j \geq (\cos\theta_n)^2\,\ell_j - 2\|e_j^T Q_2\|\,\|Z\|_2\|\Sigma\|_2\,\sqrt{\ell_j}$$
$$= (\cos\theta_n)^2\,\ell_j - 2\,\cos\theta_1\sin\theta_n\sqrt{\ell_j(1-\ell_j)}.$$

Subtracting ℓ_j from both sides and using $\ell_j \leq 1$ gives

$$\widetilde{\ell}_j - \ell_j \geq -(\sin\theta_n)^2\,\ell_j - 2\,\cos\theta_1\sin\theta_n\sqrt{\ell_j(1-\ell_j)}$$
$$\geq -(\sin\theta_n)^2 - 2\,\cos\theta_1\sin\theta_n\sqrt{\ell_j(1-\ell_j)}.$$

If $m = 2n$, then Z is a square matrix, and the smallest eigenvalue of ZZ^T is equal to

$$\lambda_n(ZZ^T) = \lambda_n(Z^T Z) = \lambda_n(I_n - \Sigma^2) = 1 - (\cos\theta_1)^2.$$

Applying this in (A.3) gives

$$\widetilde{\ell}_j \geq (\cos\theta_n)^2\,\ell_j - 2\|e_j^T Q_2\|\,\|Z\|_2\|\Sigma\|_2\,\sqrt{\ell_j} + \left(1 - (\cos\theta_1)^2\right)\,\|e_j^T Q_2\|_2^2$$
$$= (\cos\theta_n)^2\,\ell_j - 2\,\cos\theta_1\sin\theta_n\sqrt{\ell_j(1-\ell_j)} + \left(1 - (\cos\theta_1)^2\right)\,(1-\ell_j)$$
$$= 1 - \left(\sin\theta_n\,\sqrt{\ell_j} + \cos\theta_1\,\sqrt{1-\ell_j}\right)^2.$$

A.2. Proof of Theorem 2.4. The assumption $\|\Delta A\|_2\,\|A^\dagger\|_2 < 1$ implies that $\mathrm{rank}(A + \Delta A) = \mathrm{rank}(A) = n$; hence the perturbed leverage scores $\widetilde{\ell}_j$ are well defined.

We express $\sin\theta_n$ in terms of orthogonal projectors onto the column spaces. Let $\mathcal{P}$ be the orthogonal projector onto range (A), and let $\widetilde{\mathcal{P}}$ be the orthogonal projector onto range $(A + \Delta A)$. Then [23, Theorem 5.5] and [29, (5.6)] imply

$$\sin\theta_n = \|(I_m - \mathcal{P})\widetilde{\mathcal{P}}\|_2.$$

From $\mathcal{P} = AA^\dagger$ and $\widetilde{\mathcal{P}} = (A + \Delta A)(A + \Delta A)^\dagger$ follows

$$\sin\theta_n = \|(I_m - AA^\dagger)\,(A + \Delta A)(A + \Delta A)^\dagger\|_2 = \|(I_m - AA^\dagger)\,\Delta A\,(A + \Delta A)^\dagger\|_2$$
$$\leq \|(I_m - AA^\dagger)\,\Delta A\|_2\|(A + \Delta A)^\dagger\|_2 = \|A\|_2\,\|(A + \Delta A)^\perp\|_2\,\epsilon^\perp.$$

It remains to express $\|(A + \Delta A)^\dagger\|_2$ in terms of $\|A^\dagger\|_2$. The well-conditioning of singular values [13, Corollary 8.6.2] implies

$$\|(A + \Delta A)^\dagger\|_2 \leq \frac{\|A^\dagger\|_2}{1 - \|\Delta A\|_2\|A^\dagger\|_2} \leq 2\,\|A^\dagger\|_2,$$

where the last inequality is due to the assumption $\|\Delta A\|_2\|A^\dagger\|_2 \leq 1/2$. Substituting this into the bound for $\sin\theta_n$ yields $\sin\theta_n \leq 2\,\kappa_2(A)\,\epsilon^\perp$. In turn now, inserting this into Corollary 2.3 and bounding $\cos\theta_1 \leq 1$ gives the first bound in Theorem 2.4. The second one follows more easily from $\sin\theta_n \leq \kappa_2(A)\,\epsilon$; see [30, (4.4)].

A.3. Proof of Theorem 3.1. We start with a special case of Theorem 2.4 applied to $m \times n$ matrices Q and $Q + \Delta Q$ with orthonormal columns and leverage scores $\ell_j = \|e_j^T Q\|_2^2$ and $\widetilde{\ell}_j = \|e_j^T (Q + \Delta Q)\|_2^2$. Since $\|Q\|_2 = \kappa_2(Q) = 1$, we obtain

$$(A.5) \qquad \frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq \left(2 \sqrt{\frac{1 - \ell_j}{\ell_j}} + \frac{\|\Delta Q\|_2}{\ell_j} \right) \|\Delta Q\|_2, \qquad 1 \leq j \leq m.$$

The bound for $\|\Delta Q\|_2 \leq \|\Delta Q\|_F$ is obtained from a simpler version of the lemma below.

LEMMA A.1 (Theorem 1.6 in [24]). *Let A and ΔA be real $m \times n$ matrices with* $\mathrm{rank}(A) = n$, *and* $\|A^\dagger\|_2 \|\Delta A\|_2 < 1$. *If* $A + \Delta A = (Q + \Delta Q) \widetilde{R}$ *is the thin QR decomposition, then*

$$\|\Delta Q\|_F \leq \frac{1 + \sqrt{2}}{1 - \|A^\dagger\|_2 \|\Delta A\|_2} \|A^\dagger\|_2 \|\Delta A\|_F .$$

Below is a simpler but not much more restrictive version of Lemma A.1. If $\|\Delta A\|_2 \|A^\dagger\|_2 \leq 1/2$, then

$$\|\Delta Q\|_F \leq 6 \|A^\dagger\|_2 \|\Delta A\|_F = 6 \, \mathrm{sr} \, (A)^{1/2} \, \kappa_2(A) \, \epsilon_F.$$

Substituting this into (A.5) gives Theorem 3.1.

A.4. Proof of Theorem 3.2. We start with a simplified version of Theorem 2.4. Let Q and $Q + \Delta Q$ be $m \times n$ matrices with orthonormal columns, and let $\ell_j = \|e_j^T Q\|_2^2$ and $\widetilde{\ell}_j = \|e_j^T (Q + \Delta Q)\|_2^2$, $1 \leq j \leq m$, be their leverage scores. Multiplying out the inner product in $\widetilde{\ell}_j$ and using triangle and submultiplicative inequalities gives

$$(A.6) \qquad \frac{|\widetilde{\ell}_j - \ell_j|}{\ell_j} \leq 2 \frac{\|e_j^T \Delta Q\|_2}{\sqrt{\ell_j}} + \frac{\|e_j^T \Delta Q\|_2^2}{\ell_j}, \qquad 1 \leq j \leq m.$$

Next we derive bounds for $\|e_j^T \Delta Q\|_2$ in terms of ΔA. To this end we represent the perturbed matrix by a function $A(t)$, with a smooth decomposition $A(t) = Q(t) R(t)$.

This is a very common approach; see, for instance, [4, section 3], [5, section 4], [6, section 3], [7, section 5], [8, section 2.1], [15, section 2.4], [21, section 3], [24, section 2], [25, section 4], [26, section 5], and [31, section 2].

Define the function

$$A(t) \equiv A + \frac{t}{\epsilon_F} \Delta A, \qquad 0 \leq t \leq \epsilon_F \equiv \frac{\|\Delta A\|_F}{\|A\|_F}.$$

Let $A(t) = Q(t) R(t)$ be a thin QR decomposition, where we set $Q \equiv Q(0)$, $R \equiv R(0)$, $Q + \Delta Q \equiv Q(\epsilon_F)$, and $R + \Delta R \equiv R(\epsilon_F)$. The derivative of R with regard to t is $\dot{R}$.

THEOREM A.2. *Let A and $A + \Delta A$ be real $m \times n$ matrices with* $\mathrm{rank}(A) = n$ *and* $\|\Delta A\|_2 \|A^\dagger\|_2 < 1$. *Then*

$$\Delta Q = \Delta A \, R^{-1} - \epsilon_F \, Q \, \dot{R} \, R^{-1} + \mathcal{O}(\epsilon_F^2),$$

where $\|\dot{R} \, R^{-1}\|_F \leq \sqrt{2} \, \mathrm{sr} \, (A)^{1/2} \, \kappa_2(A)$.

Proof. The proof is inspired in particular by [5, section 4] and [14, section 2.4].

Smooth decomposition. From $\text{rank}(A) = n$, $\|\frac{t}{\epsilon_F}\Delta A\|_2 \leq \|\Delta A\|_2$ for $0 \leq t \leq \epsilon_F$, and $\|\Delta A\|_2 \|A^\dagger\|_2 < 1$ follows $\text{rank}(A(t)) = n$. Furthermore, since $A(t)$ has at least two continuous derivatives, so do $Q(t)$ and $R(t)$ [8, Proposition 2.3].

Expression for ΔQ. The existence of two derivatives allows us to take a Taylor expansion of $Q(t)$ around $t = 0$ and get $Q(t) - Q(0) = t\,\dot{Q}(0) + \mathcal{O}(t^2)$. Evaluating at $t = \epsilon_F$ gives

$$(A.7) \qquad \Delta Q = (Q + \Delta Q) - Q = Q(\epsilon_F) - Q(0) = \epsilon_F\,\dot{Q} + \mathcal{O}(\epsilon_F^2).$$

To get an expression for $\dot{Q}$, differentiate $A(t) = Q(t)R(t)$,

$$\frac{\Delta A}{\epsilon_F} = \dot{Q}(t)\,R(t) + Q(t)\,\dot{R}(t),$$

and evaluate at $t = 0$,

$$\dot{Q} = \frac{\Delta A}{\epsilon_F}\,R^{-1} - Q\,\dot{R}\,R^{-1}.$$

Inserting the above into (A.7) gives the expression for ΔQ in Theorem 3.2.

Bound for $\|\dot{R}R^{-1}\|_F$. Differentiating $A(t)^T A(t) = R(t)^T R(t)$ gives

$$\frac{1}{\epsilon_F}\left((\Delta A)^T A + A^T \Delta A + \frac{2t}{\epsilon}(\Delta A)^T \Delta A\right) = \dot{R}(t)^T R(t) + R(t)^T \dot{R}(t),$$

and evaluating at $t = 0$ yields

$$\frac{1}{\epsilon_F}\left((\Delta A)^T A + A^T \Delta A\right) = \dot{R}^T R + R^T \dot{R}.$$

Multiplying by R^{-T} on the left and by R^{-1} on the right gives

$$(A.8) \qquad \dot{R}R^{-1} + \left(\dot{R}R^{-1}\right)^T = \frac{1}{\epsilon_F}\left(Q^T \Delta A R^{-1} + (Q^T \Delta A R^{-1})^T\right).$$

Now we take advantage of the fact that $\dot{R}R^{-1}$ is upper triangular and define a function that extracts the upper triangular part of a square matrix Z via

$$\text{up}(Z) \equiv \frac{1}{2}\text{diagonal}(Z) + \text{strictly upper triangular part}(Z).$$

Applying the function to (A.8) gives

$$\dot{R}R^{-1} = \frac{1}{\epsilon_F}\,\text{up}\left(Q^T \Delta A R^{-1} + (Q^T \Delta A R^{-1})^T\right).$$

Taking norms yields [5, equation (3.5)]

$$(A.9) \qquad \left\|\dot{R}R^{-1}\right\|_F \leq \frac{\sqrt{2}}{\epsilon_F}\left\|Q^T \Delta A R^{-1}\right\|_F$$

$$\leq \frac{\sqrt{2}}{\epsilon_F}\,\|\Delta A\|_F\,\|R^{-1}\|_2 = \sqrt{2}\,\text{sr}\,(A)^{1/2}\,\kappa_2(A). \qquad \square$$

Now we are ready to derive a bound for the row norms of ΔQ. Combining the two bounds from Theorem A.2, that is, inserting $\|\dot{R}\, R^{-1}\|_F \leq \sqrt{2}\, \mathrm{sr}\,(A)^{1/2}\, \kappa_2(A)$ into

$$\left\|e_j^T \Delta Q\right\|_2 \leq \left\|e_j^T \Delta A\right\|_2 \left\|A^\dagger\right\|_2 + \epsilon_F\, \sqrt{\ell_j}\, \left\|\dot{R}R^{-1}\right\|_2 + \mathcal{O}(\epsilon^2), \qquad 1 \leq j \leq m,$$

gives

$$\left\|e_j^T \Delta Q\right\|_2 \leq \left\|e_j^T \Delta A\right\|_2 \|A^\dagger\|_2 + \sqrt{2\,\ell_j}\, \sqrt{\mathrm{sr}\,(A)}\, \epsilon_F\, \kappa_2(A) + \mathcal{O}(\epsilon_F^2).$$

Into the first summand substitute

$$\|e_j^T \Delta A\|_2 = \epsilon_j\, \|e_j^T A\|_2 \leq \epsilon_j\, \|e_j^T Q\|_2\, \|R\|_2 = \epsilon_j\, \sqrt{\ell_j}\, \|A\|_2,$$

and obtain

$$\|e_j^T \Delta Q\|_2 \leq \sqrt{\ell_j}\, \left(\epsilon_j + \sqrt{2}\, \sqrt{\mathrm{sr}\,(A)}\, \epsilon_F\right)\, \kappa_2(A) + \mathcal{O}(\epsilon_F^2), \qquad 1 \leq j \leq m.$$

Inserting the above into (A.6) and focusing on the first-order terms in ϵ_F gives Theorem 3.2.

A.5. Proof of Theorem 3.3. To remove the contribution of ΔA in range (A), let $\mathcal{P} \equiv AA^\dagger$ be the orthogonal projector onto range (A), and let $\mathcal{P}^\perp \equiv I_m - \mathcal{P}$ be the orthogonal projector onto range $(A)^\perp$. Extracting the contribution in range (A) gives

$$A + \Delta A = A + \mathcal{P}\, \Delta A + \mathcal{P}^\perp\, \Delta A = (A + \mathcal{P}\, \Delta A) + \mathcal{P}^\perp\, \Delta A = M + \Delta M,$$

where $M \equiv A + \mathcal{P}\, \Delta A$ and $\Delta M \equiv \mathcal{P}^\perp\, \Delta A$.

Leverage scores. Here $\mathrm{rank}(M) = n$, because $\mathcal{P}$ is an orthogonal projector, so that $\|\mathcal{P}\, \Delta A\|_2 \|A^\dagger\|_2 \leq \|\Delta A\|_2 \|A^\dagger\|_2 < 1$. With $M = \mathcal{P}\,(A + \Delta A)$ this implies range $(M) = $ range (A). Furthermore, $\mathrm{rank}(M + \Delta M) = \mathrm{rank}(A + \Delta A) = n$. Thus M and $M + \Delta M$ have thin QR decompositions, $M = QX$ and $M + \Delta M = (Q + \Delta Q)\widetilde{X}$, and have the same leverage scores ℓ_j and $\widetilde{\ell}_j$, respectively, as A and $A + \Delta A$.

Ultimately, we want to apply Theorem 3.2 to M and $M + \Delta M$, but the perturbation $\Delta M = \mathcal{P}^\perp\, \Delta A$ is to be related to A rather than to M, and the bound is to be expressed in terms of $\kappa_2(A)$ rather than $\kappa_2(M)$.

Applying Theorem A.2. With

$$M(t) \equiv M + \tfrac{t}{\mu}\Delta M, \qquad 0 \leq t \leq \mu \equiv \frac{\|\Delta M\|_F}{\|A\|_F} = \epsilon_F^\perp,$$

Theorem A.2 implies

$$(A.10) \qquad \Delta Q = \Delta M\, X^{-1} - \mu\, Q\, \dot{X}\, X^{-1} + \mathcal{O}(\mu^2).$$

To bound $\|\dot{X}X^{-1}\|_F$, we apply (A.9) and obtain

$$(A.11) \qquad \|\dot{X}\, X^{-1}\|_F \leq \frac{\sqrt{2}}{\mu}\, \|\Delta M\|_F\, \|X^{-1}\|_2.$$

Bounding $\|e_j^T \Delta Q\|_2$. Combining (A.10) and (A.11) gives

$$\|e_j^T \Delta Q\|_2 \leq \left(\|e_j^T \Delta M\|_2 + \sqrt{2}\, \|e_j^T Q\|_2 \|\Delta M\|_F \right) \|X^{-1}\|_2 + \mathcal{O}(\mu^2), \qquad 1 \leq j \leq m,$$

$$= \left(\epsilon_j^{\perp} \frac{\|e_j^T A\|_2}{\|A\|_2} + \sqrt{2}\, \sqrt{\ell_j}\, \mu \operatorname{sr}(A)^{1/2} \right) \|A\|_2 \|M^{\dagger}\|_2 + \mathcal{O}(\mu^2).$$

From $\|e_j^T A\|_2 \leq \|e_j^T Q\|_2 \|A\|_2 = \sqrt{\ell_j}\, \|A\|_2$ follows

$$(A.12) \qquad \|e_j^T \Delta Q\|_2 \leq \sqrt{\ell_j} \left(\epsilon_j^{\perp} + \sqrt{2}\, \mu \operatorname{sr}(A)^{1/2} \right) \|A\|_2 \|M^{\dagger}\|_2 + \mathcal{O}(\mu^2).$$

It remains to express $\|M^{\dagger}\|_2$ in terms of $\|A^{\dagger}\|_2$. The well-conditioning of singular values [13, Corollary 8.6.2] applied to $M = A + Z$, where $Z \equiv \mathcal{P} \Delta A$, implies

$$\|M^{\dagger}\|_2 = \|(A + Z)^{\dagger}\|_2 \leq \frac{\|A^{\dagger}\|_2}{1 - \|Z\|_2 \|A^{\dagger}\|_2} \leq 2 \|A^{\dagger}\|_2,$$

where the last inequality is due to the assumption $\|Z\|_2 \|A^{\dagger}\|_2 \leq \|\Delta A\|_2 \|A^{\dagger}\|_2 \leq 1/2$. Inserting this bound for $\|M^{\dagger}\|_2$ into (A.12) yields

$$\|e_j^T \Delta Q\|_2 \leq 2 \sqrt{\ell_j} \left(\epsilon_j^{\perp} + \sqrt{2}\, \mu \operatorname{sr}(A)^{1/2} \right) \kappa_2(A) + \mathcal{O}(\mu^2), \qquad 1 \leq j \leq m.$$

At last, substituting the above into (A.6) and focusing on the first-order terms in $\mu = \epsilon_F^{\perp}$ gives Theorem 3.3.

A.6. Proof of Theorem 3.4. Write the perturbations (3.1) as $\Delta A = DA$, where D is a diagonal matrix with diagonal elements $D_{jj} = \zeta_j \eta_j$, $1 \leq j \leq m$. By assumption $\|\Delta A\|_2 \|A^{\dagger}\|_2 \leq \eta\, \kappa_2(A) < 1$, so that $\operatorname{rank}(A + \Delta A) = n$.

As in the proof of Theorem 3.2, we start with (A.6). To derive bounds for $\|e_j^T \Delta Q\|_2$ in terms of η_j and η, represent the perturbed matrix by

$$A(t) \equiv A + \frac{t}{\eta} \Delta A, \qquad 0 \leq t \leq \eta.$$

Let $A(t) = Q(t)R(t)$ be a thin QR decomposition, where $Q \equiv Q(0)$, $R \equiv R(0)$, $Q + \Delta Q \equiv Q(\eta)$, and $R + \Delta R \equiv R(\eta)$. The derivative of R with respect to t is $\dot{R}$.

Theorem A.2 implies

$$\Delta Q = \Delta A\, R^{-1} - \epsilon\, Q\, \dot{R}\, R^{-1} + \mathcal{O}(\eta^2) = DQ - Q\, \dot{R}\, R^{-1} + \mathcal{O}(\eta^2).$$

With $\Delta A = DA$ this gives

$$e_j^T \Delta Q = \eta_j\, e_j^T Q + \eta\, e_j^T Q\, \dot{R}\, R^{-1} + \mathcal{O}(\eta^2), \qquad 1 \leq j \leq m.$$

Taking norms gives

$$(A.13) \qquad \|e_j^T \Delta Q\|_2 \leq \sqrt{\ell_j} \left(\eta_j + \eta\, \|\dot{R}\, R^{-1}\|_2 \right) + \mathcal{O}(\eta^2), \qquad 1 \leq j \leq m.$$

From (A.9) follows

$$\|\dot{R}\, R^{-1}\|_2 \leq \left\| \dot{R} R^{-1} \right\|_F \leq \frac{\sqrt{2}}{\eta}\, \left\| Q^T \Delta A R^{-1} \right\|_F = \frac{\sqrt{2}}{\eta}\, \left\| Q^T DQ \right\|_F$$

$$\leq \frac{\sqrt{2}}{\eta}\, \left\| Q^T \right\|_F\, \|DQ\|_2 \leq \sqrt{2}n.$$

Combining this with (A.13) yields

$$\|e_j^T \, \Delta Q\|_2 \le \sqrt{\ell_j} \, \left(\eta_j + \sqrt{2}\,n\,\eta\right) + \mathcal{O}(\eta^2), \qquad 1 \le j \le m.$$

Inserting the above into (A.6) and focusing on the first-order terms in η gives Theorem 3.4.

REFERENCES

[1] R. Bhatia and K. Mukherjea, *Variation of the unitary part of a matrix*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1007–1014.

[2] C. Boutsidis, M. W. Mahoney, and P. Drineas, *An Improved Approximation Algorithm for the Column Subset Selection Problem*, preprint, arXiv:0812.4293, 2010.

[3] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.

[4] X.-W. Chang, *On the perturbation of the Q-factor of the QR factorization*, Numer. Linear Algebra Appl., 19 (2012), pp. 607–619.

[5] X.-W. Chang and C. C. Paige, *Componentwise perturbation analyses for the QR factorization*, Numer. Math., 88 (2001), pp. 319–345.

[6] X.-W. Chang, C. C. Paige, and G. W. Stewart, *Perturbation analyses for the QR factorization*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 775–791.

[7] X.-W. Chang and D. Stehlé, *Rigorous perturbation bounds of some matrix factorizations*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2841–2859.

[8] L. Dieci and T. Eirola, *On smooth decompositions of matrices*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 800–819.

[9] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff, *Fast approximation of matrix coherence and statistical leverage*, J. Mach. Learn. Res., 13 (2012), pp. 3475–3506.

[10] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Sampling algorithms for l_2 regression and applications*, in Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '06 (New York, 2006), ACM, Providence, RT, SIAM, Philadelphia, 2006, pp. 1127–1136.

[11] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Subspace sampling and relative-error matrix approximation: Column-based methods*, in Approximation, Randomization and Combinatorial Optimization, Lecture Notes in Comput. Sci. 4110, Springer, Berlin, 2006, pp. 316–326.

[12] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, *Relative-error CUR matrix decompositions*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 844–881.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.

[14] N. J. Higham, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1160–1174.

[15] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[16] D. C. Hoaglin and R. E. Welsch, *The Hat matrix in regression and ANOVA*, Amer. Statist., 32 (1978), pp. 17–22.

[17] I. C. F. Ipsen and T. Wentworth, *The effect of coherence on sampling from matrices with orthonormal columns, and preconditioned least squares problems*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1490–1520.

[18] M. Li, G. L. Miller, and R. Peng, *Iterative row sampling*, in Proceedings of the 54th IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society, Los Alamitos, CA, 2013, pp. 127–136.

[19] M. Magdon-Ismail, *Row Sampling For Matrix Algorithms Via A Non-Commutative Bernstein Bound*, preprint, arXiv:1008.0587, 2010.

[20] M. W. Mahoney, *Randomized algorithms for matrices and data*, Found. Trends Machine Learning, 3 (2011), pp. 123–224.

[21] G. W. Stewart, *Perturbation bounds for the QR factorization of a matrix*, SIAM J. Numer. Anal., 14 (1977), pp. 509–518.

[22] G. W. Stewart, *On the perturbation of LU, Cholesky, and QR factorizations*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1141–1145.

[23] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, CA, 1990.

[24] J.-G. SUN, *Perturbation bounds for the Cholesky and QR factorizations*, BIT, 31 (1991), pp. 341–352.

[25] J.-G. SUN, *Componentwise perturbation bounds for some matrix decompositions*, BIT, 32 (1992), pp. 702–714.

[26] J.-G. SUN, *On perturbation bounds for the QR factorization*, Linear Algebra Appl., 215 (1995), pp. 95–111.

[27] A. TALWALKAR AND A. ROSTAMIZADEH, *Matrix coherence and the Nyström method*, in Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence (UAI-10), AUAI Press, Corvalis, OR, 2010, pp. 572–579.

[28] P. F. VELLEMAN AND R. E. WELSCH, *Efficient computing of regression diagnostics*, Amer. Statist., 35 (1981), pp. 234–242.

[29] P.-Æ. WEDIN, *Perturbation bounds in connection with singular value decomposition*, Nordisk Tidskr. Informationsbehandling (BIT), 12 (1972), pp. 99–111.

[30] P.-Æ. WEDIN, *On angles between subspaces of a finite dimensional inner product space*, in Matrix Pencils, B. Kågström and A. Ruhe, eds., Lecture Notes in Math. 973, Springer, Berlin, 1983, pp. 263–285.

[31] H. Y. ZHA, *A componentwise perturbation analysis of the QR decomposition*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1124–1131.