ACCURATE SOLUTION OF STRUCTURED LEAST SQUARES PROBLEMS VIA RANK-REVEALING DECOMPOSITIONS [*]

NIEVES CASTRO-GONZÁLEZ[†], JOHAN CEBALLOS[‡], FROILÁN M. DOPICO[§], AND JUAN M. MOLERA[‡]

Abstract. Least squares problems $\min_x \|b - Ax\|_2$ where the matrix $A \in \mathbb{C}^{m\times n}$ $(m \geq n)$ has some particular structure arise frequently in applications. Polynomial data fitting is a well-known instance of problems that yield highly structured matrices, but many other examples exist. Very often, structured matrices have huge condition numbers $\kappa_2(A) = \|A\|_2 \, \|A^\dagger\|_2$ ($A^\dagger$ is the Moore-Penrose pseudo-inverse of A) and, therefore, standard algorithms fail to compute accurate minimum 2-norm solutions of least squares problems. In this work, we introduce a framework that allows us to compute minimum 2-norm solutions of many classes of structured least squares problems accurately, i.e., with errors $\|\widehat{x}_0 - x_0\|_2/\|x_0\|_2 = O(\mathtt{u})$, where $\mathtt{u}$ is the unit roundoff, independently of the magnitude of $\kappa_2(A)$ for most vectors b. The cost of these accurate computations is $O(n^2 m)$ flops, i.e., roughly the same cost as standard algorithms for least squares problems. The approach in this work relies in computing first an accurate rank-revealing decomposition of A, an idea that has been widely used in the last decades to compute, for structured ill-conditioned matrices, singular value decompositions, eigenvalues and eigenvectors in the Hermitian case, and solutions of linear systems with high relative accuracy. In order to prove that accurate solutions are computed, it is needed to develop a multiplicative perturbation theory of least squares problems. The results presented in this paper are valid in the case of both full rank and rank deficient problems and also in the case of underdetermined linear systems $(m < n)$. Among other types of matrices, the new method applies to rectangular Cauchy, Vandermonde, and graded matrices and detailed numerical tests for Cauchy matrices are presented.

Key words. accurate solutions, least squares problems, Moore-Penrose pseudo-inverse, multiplicative perturbation theory, rank revealing decompositions, structured matrices

AMS subject classifications. 65F20, 65F35, 15A09, 15A12, 15A23, 15B05.

1. Introduction.

Structured matrices arise frequently in theory and applications [44, 45]. As a consequence, the design and analysis of special algorithms for performing structured matrix computations is a classical area of Numerical Linear Algebra that attracts the attention of many researchers. Special algorithms for solving structured linear systems of equations or structured eigenvalue problems are included in many standard references [15, 24, 29, 35, 51], but special algorithms for solving structured least squares problems do not appear so often in the literature. The goal of special algorithms is to exploit the structure to increase the speed of computations, and/or to decrease storage requirements, and/or to improve the accuracy of the solutions with respect to standard algorithms. On this latter goal, let us mention that algorithms for solving structured linear systems of equations more accurately than standard methods have been developed from the early days of Numerical Linear Algebra [5] and many papers have been published on this topic since then (see the references in [17, 29]). Although some preliminary ideas on accurate structured eigenvalue computations date from the 60's [33], the systematic development of accurate algorithms for structured eigenvalue problems is much more recent, since it started in early 90's with the celebrated paper [13] and has also received considerable attention (see [3, 12, 14, 16, 18, 19, 20, 23, 34, 48, 53] among other references). The present paper focuses on a part of *"Accurate Numerical Linear Algebra"* for which there are not many references available in the literature: algorithms for solving structured least squares problems $\min_x \|b - Ax\|_2$, where $A \in \mathbb{C}^{m\times n}$ and $b \in \mathbb{C}^m$,

[*] This work was supported by the Ministerio de Economía y Competitividad of Spain through grants MTM-2009-09281, MTM-2012-32542 (Ceballos, Dopico, and Molera) and MTM2010-18057 (Castro-González).

[†] Facultad de Informática, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, Spain (nieves@fi.upm.es).

[‡] Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain (jceballo@math.uc3m.es, molera@math.uc3m.es).

[§] Instituto de Ciencias Matemáticas CSIC-UAM-UC3M-UCM and Departamento de Matemáticas, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Spain (dopico@math.uc3m.es).

with much more accuracy than the one provided by standard algorithms and roughly with the same computational cost, that is, $O(n^2 m)$ flops. These algorithms have received attention for diagonally scaled matrices [2, 10, 30, 47], but outside this class of matrices we only know reference [42].

The standard method for solving full column-rank least squares (LS) problems $\min_x \|b-Ax\|_2$ is via the QR factorization computed with the Householder algorithm [29, Chapters 19 and 20]. This method is backward stable, that is, the computed solution $\widehat{x}_0$ is the exact solution of a LS problem $\min_x \|(b+\Delta b)-(A+\Delta A)x\|_2$, where $\|\Delta b\|_2 \leq c\,\mathtt{u}\,mn\,\|b\|_2$, $\|\Delta A\|_2 \leq c\,\mathtt{u}\,m\,n^{3/2}\,\|A\|_2$, $\mathtt{u}$ is the unit roundoff of the computer, and c denotes a small integer constant [29, Theorem 20.3]. Backward error results of the same type hold for other methods of solution of LS problems based on orthogonal decompositions as, for instance, the singular value decomposition (SVD)*. This strong backward error result, together with classical normwise perturbation theory of LS problems [52, Theorem 5.1] (see also [4, Theorem 1.4.6, p. 30]), implies the following forward error bound in the computed solution $\widehat{x}_0$ with respect to the exact solution x_0

$$(1.1) \qquad \frac{\|\widehat{x}_0 - x_0\|_2}{\|x_0\|_2} \leq (c\,\mathtt{u}\,m\,n^{3/2}) \left(\kappa_2(A) + \frac{\|A^\dagger\|_2 \|b\|_2}{\|x_0\|_2} + \kappa_2(A)^2 \, \frac{\|b - Ax_0\|_2}{\|A\|_2 \, \|x_0\|_2} \right),$$

where $A^\dagger$ is the Moore-Penrose pseudo-inverse of A, $\|A\|_2$ denotes the spectral norm of A, and $\kappa_2(A) = \|A\|_2\,\|A^\dagger\|_2$ is the spectral condition number of A. The bound in (1.1) is larger than $\mathtt{u}\,\kappa_2(A)$ (in fact, it can be much larger) and, so, (1.1) does not guarantee any digit of accuracy in the computed solution if $\kappa_2(A) \gtrsim 1/\mathtt{u}$. Unfortunately, many structured matrices arising in applications are very ill-conditioned and standard algorithms for LS problems may compute solutions with huge relative errors. Two famous examples are Vandermonde matrices, which arise in polynomial data fitting, and Cauchy matrices [29, Chapters 22 and 28] (for more recent information on condition numbers of square and rectangular Vandermonde matrices see, respectively, [39, 40] and [41]). It should be noted that often ill-conditioned LS problems arise as a consequence of noisy data and that, in these cases, it may be more convenient to *regularize* first the problem, for instance by truncating the SVD of A [26], than to compute an accurate solution of the noisy problem. However, we emphasize that this is not the scenario studied in this work and that we assume that the data are not affected by noise and then it is of interest to compute the solution as accurately as possible.

Our goal in this work is to present a numerical framework for the solution of LS problems and to prove that it allows us to compute for many classes of structured matrices solutions with error bounds much smaller than (1.1). The framework we introduce relies on the concept of *rank-revealing decomposition* (RRD). Different versions of this idea have been studied as early as the 1960s [22, 27] (see also [36, 49] and references therein), but the definition we use of RRD is the one introduced recently in [12] for computing the SVD with high relative accuracy. An RRD of $A \in \mathbb{C}^{m\times n}$ is a factorization $A = XDY$, where $X \in \mathbb{C}^{m\times r}$, $D = \operatorname{diag}(d_1, d_2, \ldots, d_r) \in \mathbb{C}^{r\times r}$ is diagonal and nonsingular, and $Y \in \mathbb{C}^{r\times n}$, $\operatorname{rank}(X) = \operatorname{rank}(Y) = r$, and X and Y are well conditioned. Note that this means that the rank of A is r, and that if A is ill conditioned, then the diagonal factor D is also ill conditioned. We propose to compute the minimum 2-norm solution of $\min_x \|b - Ax\|_2$ in two main stages:

*It should be noted that the backward error in A committed by solving LS problems via the Householder QR factorization is columnwise, i.e., $\|\Delta A(:,j)\|_2 \leq c\,\mathtt{u}\,mn\,\|A(:,j)\|_2$ for $j = 1:n$ (MATLAB notation), and, therefore, it is stronger than the one mentioned above. This columnwise bound also holds for LS problems solved via the SVD computed with properly implemented one-sided or two-sided Jacobi methods [19, 20], but not if the SVD is computed by methods whose first step is to bidiagonalize the matrix, since in these methods orthogonal transformations are applied to A from both sides in a way that does not guarantee columnwise bounds.

1. First stage. Compute an RRD of $A = XDY$, accurately in the sense of [12] (we revise the precise meaning of "accuracy" in this context in Definition 2.3).
2. Second stage. It has three steps: (1) compute the unique solution x_1 of $\min_x \|b - Xx\|_2$ via Householder QR factorization; (2) compute the solution x_2 of the linear system $Dx_2 = x_1$ as $x_2(i) = x_1(i)/d_i$, $i = 1 : r$; and (3) compute the minimum 2-norm solution x_0 of the underdetermined linear system $Yx = x_2$ using the Q method [29, Chapter 21]. The vector x_0 is the minimum 2-norm solution of $\min_x \|b - Ax\|_2$.

The intuition behind why this procedure computes accurate solutions, even for extremely ill conditioned matrices A, is that each entry of x_2 is computed with a relative error less than $\mathtt{u}$ (given D and x_1), that is, the ill conditioned linear system $Dx_2 = x_1$ is solved very accurately, and that $\min_x \|b - Xx\|_2$ and $Yx = x_2$ are also solved accurately because X and Y are well conditioned. We will prove in Section 5 that the relative error for the minimum 2-norm solution $\widehat{x}_0$ computed by the proposed procedure is

$$\frac{\|\widehat{x}_0 - x_0\|_2}{\|x_0\|_2} \leq \mathtt{u}\, f(m,n) \left(\kappa_2(Y) + \kappa_2(X) \frac{\|A^\dagger\|_2 \, \|b\|_2}{\|x_0\|_2} \right), \tag{1.2}$$

where $f(m,n)$ is a modestly growing function of m and n. Note first that (1.2) improves (1.1), because X and Y are well conditioned and, so, the only potentially large factor in (1.2) is $\|A^\dagger\|_2 \, \|b\|_2/\|x_0\|_2$, which also appears in (1.1). But the really important point on the bound (1.2) is that if A is fixed, then $\|A^\dagger\|_2 \, \|b\|_2/\|x_0\|_2$ is small for most right-hand sides b, even for very ill conditioned matrices A. This is well-known if A is square and nonsingular (see [1, 9] and [17, Section 3.2]). It can be shown [7, Section 4.1] [8] that it also holds for general matrices in two senses: for most vectors b that are everywhere in the space, and for most vectors b with a fixed value of the relative residual $\|Ax_0 - b\|_2/\|b\|_2$ not too close to 1. In this paper the sentence "for most vectors b" may be understood in any of these two senses.

The framework and the results discussed above resemble those presented in [17] for computing accurate solutions of structured linear systems. However the analysis for LS problems requires completely different techniques for developing the new multiplicative perturbation theory that is needed to prove the error bound in (1.2). In addition, the results and algorithms we present are fully general, since they remain valid both for full rank and rank defective matrices A, and they can be also applied to solve accurately underdetermined linear systems.

The computation of an *accurate* RRD $A = XDY$ is the difficult part in the framework above. For most well-conditioned matrices an accurate RRD can be computed with standard Gaussian elimination with complete pivoting (GECP) or with the Householder QR algorithm with column-pivoting and when, very rarely, this fails to produce well conditioned X and Y factors, then other pivoting strategies can be used [25, 43, 46]. However, for ill conditioned matrices accurate RRDs can be computed only for particular classes of structured matrices through special implementations of GECP that exploit carefully the structure and, in the case of graded matrices, also via Householder QR factorization with *complete* pivoting [28].

Fortunately, as a by-product of the intense research performed in the last two decades on computing SVDs with high relative accuracy, there exist algorithms to compute accurate RRDs of many classes of $m \times n$ structured matrices in $O(m\,n^2)$ flops. See [12] and [17, Section 1] for a detailed list of these classes of matrices. Most of the algorithms that compute accurate RRDs determine exactly the rank of rank-deficient matrices and, even in this case, the framework introduced in this paper solves LS problems with relative errors bounded as in (1.2). This error bound is $O(\mathtt{u}\, f(m,n))$ for most right-hand sides independently of the traditional condition number of the matrices and so guarantees accurate solutions.

The paper is organized as follows. Preliminaries are introduced in Section 2. Section 3 provides an expression for the variation of the Moore-Penrose inverse under multiplicative perturbations, which is applied to LS problems in Subsection 4.1. Then, we get in Subsection 4.2 perturbation bounds for LS problems whose coefficient matrix is given as an RRD under perturbations of the factors. Section 5 presents a new algorithm for solving accurately LS problems via RRDs and its error analysis. The accuracy of this algorithm is checked via numerical tests in Section 6. Finally, conclusions and future work are discussed in Section 7.

2. Preliminaries and basic concepts. Since we consider LS problems, we will use the most natural norms for these problems: the Euclidean vector norm, i.e., given $x = [x_i]_i^n \in \mathbb{C}^n$, $\|x\|_2^2 := \sum_{i=1}^n |x_i|^2$, and for matrices $A \in \mathbb{C}^{m\times n}$ the corresponding subordinate matrix spectral norm $\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2$. The symbol I stands for the identity matrix, the size will be clear from the context, and A^* denotes the conjugate-transpose of A. We will use MATLAB notation for submatrices: $A(i : j, :)$ indicates the submatrix of A consisting of rows i through j and $A(:, k : l)$ indicates the submatrix of A consisting of columns k through l. Given $A \in \mathbb{C}^{m\times n}$, with $m \geq n$, its singular values are denoted as $\sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0$. Next Lemma 2.1 will be needed to derive some perturbation bounds.

LEMMA 2.1. *Let $B, C \in \mathbb{C}^{m\times n}$, let $\mathcal{S} \subseteq \mathbb{C}^m$ and $\mathcal{W} \subseteq \mathbb{C}^n$ be vector subspaces, and let $P_\mathcal{S} \in \mathbb{C}^{m\times m}$ and $P_\mathcal{W} \in \mathbb{C}^{n\times n}$ be the orthogonal projectors onto $\mathcal{S}$ and $\mathcal{W}$, respectively. Then the following statements hold:*

(a) $\|P_\mathcal{S} B + (I - P_\mathcal{S})C\|_2 \leq \sqrt{\|B\|_2^2 + \|C\|_2^2}\ .$

(b) $\|BP_\mathcal{W} + C(I - P_\mathcal{W})\|_2 \leq \sqrt{\|B\|_2^2 + \|C\|_2^2}\ .$

Proof. Part (a). Let $x \in \mathbb{C}^n$ with $\|x\|_2 = 1$. Since the vectors $P_\mathcal{S}Bx$ and $(I - P_\mathcal{S})Cx$ are orthogonal, then $\|(P_\mathcal{S}B + (I - P_\mathcal{S})C)x\|_2^2 = \|P_\mathcal{S}Bx\|_2^2 + \|(I - P_\mathcal{S})Cx\|_2^2 \leq \|Bx\|_2^2 + \|Cx\|_2^2 \leq \|B\|_2^2 + \|C\|_2^2$ and

$$\|P_\mathcal{S}B + (I - P_\mathcal{S})C\|_2 = \max_{\|x\|_2=1} \|(P_\mathcal{S}B + (I - P_\mathcal{S})C)x\|_2 \leq \sqrt{\|B\|_2^2 + \|C\|_2^2}\,.$$

Part (b) follows from part (a) applied to the conjugate-transpose matrix. □

In Sections 4.2 and 5, we will need the entrywise absolute value of a matrix. Given a matrix $G \in \mathbb{C}^{m\times n}$ with entries g_{ij}, we denote by $|G|$ the matrix with entries $|g_{ij}|$. Expressions like $|G| \leq |B|$, where $B \in \mathbb{C}^{m\times n}$, mean $|g_{ij}| \leq |b_{ij}|$ for $1 \leq i \leq m,\ 1 \leq j \leq n$.

We denote by $A^\dagger \in \mathbb{C}^{n\times m}$ the Moore-Penrose pseudo-inverse of $A \in \mathbb{C}^{m\times n}$ [6], which is the unique matrix that satisfies

$$\text{(2.1)} \qquad \text{(i) } AA^\dagger A = A,\ \text{(ii) } A^\dagger AA^\dagger = A^\dagger,\ \text{(iii) } (AA^\dagger)^* = AA^\dagger,\ \text{(iv) } (A^\dagger A)^* = A^\dagger A\,.$$

For any matrix Z, we denote by $\mathcal{R}(Z)$ its column space and by P_Z the orthogonal projector onto $\mathcal{R}(Z)$. It is easy to see that $\mathcal{R}(A^*) = \mathcal{R}(A^\dagger)$, $P_A = AA^\dagger$, and $P_{A^*} = P_{A^\dagger} = A^\dagger A$ [6].

The second stage of the computation of the minimum 2-norm solution of $\min_x \|b-Ax\|_2$ in the framework we propose is based on the following simple lemma.

LEMMA 2.2. *Let $A = XDY$ be an RRD of $A \in \mathbb{C}^{m\times n}$, then $A^\dagger = Y^\dagger D^{-1} X^\dagger$. Consequently, the minimum 2-norm solution of, both, a LS problem $\min_{x\in\mathbb{C}^n} \|b - XDYx\|_2$ and a consistent underdetermined linear system $XDYx = b$ is given by*

$$\text{(2.2)} \qquad x_0 = Y^\dagger D^{-1} X^\dagger b.$$

Proof. The minimum 2-norm solution of both problems is given by $x_0 = A^\dagger b$. A well-known property [6] of the Moore-Penrose pseudo-inverse states that if $F \in \mathbb{C}^{m\times r}$ and $G \in \mathbb{C}^{r\times n}$ and $\operatorname{rank}(F) = \operatorname{rank}(G) = r$, then $(FG)^\dagger = G^\dagger F^\dagger$. Lemma 2.2 follows from two successive applications of this property. □

We define, following [12], the precise meaning of an *accurate* computed RRD of a matrix A. We add, with respect to [12], the condition (2.5) that guarantees that the computed and exact "well conditioned" factors X and Y have condition numbers of similar magnitude.

DEFINITION 2.3. *Let $A = XDY$ be an RRD of $A \in \mathbb{C}^{m\times n}$, where $X \in \mathbb{C}^{m\times r}$, $D = \operatorname{diag}(d_1, \dots, d_r) \in \mathbb{C}^{r\times r}$, and $Y \in \mathbb{C}^{r\times n}$, and let $\widehat{X} \in \mathbb{C}^{m\times r}$, $\widehat{D} = \operatorname{diag}(\widehat{d}_1, \dots, \widehat{d}_r) \in \mathbb{C}^{r\times r}$, and $\widehat{Y} \in \mathbb{C}^{r\times n}$ be the factors computed by a certain algorithm in a computer with unit roundoff $\mathtt{u}$. We say that the factorization $\widehat{X}\widehat{D}\widehat{Y}$ has been accurately computed, or is an accurate RRD, if*

$$\frac{\|\widehat{X}-X\|_2}{\|X\|_2} \le \mathtt{u}\, p(m,n), \quad \frac{\|\widehat{Y}-Y\|_2}{\|Y\|_2} \le \mathtt{u}\, p(m,n), \quad \textit{and} \tag{2.3}$$

$$\frac{|\widehat{d}_i - d_i|}{|d_i|} \le \mathtt{u}\, p(m,n), \quad i = 1:r, \tag{2.4}$$

where $p(m,n)$ is a modestly growing function of m and n, i.e., a function bounded by a low degree polynomial in m and n, such that

$$\max\{\kappa_2(X)\,,\, \kappa_2(Y)\}\, \mathtt{u}\, p(m,n) < 1/2\,. \tag{2.5}$$

For example, the algorithm to compute an RRD of an $m \times n$ $(m \ge n)$ real Cauchy matrix presented[†] in [11, Section 4] computes the factors with an entrywise relative error bounded by $9n\mathtt{u}/(1-9n\mathtt{u})$.

Let us discuss the role of (2.5). Weyl perturbation theorem [50] for singular values and (2.5) imply that $\operatorname{rank}(X) = \operatorname{rank}(\widehat{X}) = r$, $\operatorname{rank}(Y) = \operatorname{rank}(\widehat{Y}) = r$, and that

$$\frac{\kappa_2(X)}{3} \le \kappa_2(\widehat{X}) \le 3\,\kappa_2(X) \quad \text{and} \quad \frac{\kappa_2(Y)}{3} \le \kappa_2(\widehat{Y}) \le 3\,\kappa_2(Y)\,. \tag{2.6}$$

This will allow us to use either $\kappa_2(X)$ and $\kappa_2(Y)$, or $\kappa_2(\widehat{X})$ and $\kappa_2(\widehat{Y})$ in the rounding error bounds obtained in Section 5 by modifying somewhat the constants involved in the bounds.

In Section 5 we will use the conventional error model for floating point arithmetic [29, Section 2.2]. In addition, we will assume that neither overflow nor underflow occurs.

3. Multiplicative perturbation results for the Moore-Penrose pseudo-inverse.

In this section and in Section 4.1, we consider a multiplicative perturbation of a general matrix $A \in \mathbb{C}^{m\times n}$, that is, a matrix $\widetilde{A} = (I+E)A(I+F)$, where $(I+E) \in \mathbb{C}^{m\times m}$ and $(I+F) \in \mathbb{C}^{n\times n}$ are nonsingular matrices. The final goal is to bound, in Section 4.1, $\|\widetilde{x}_0 - x_0\|_2/\|x_0\|_2$, where x_0 and $\widetilde{x}_0$ are the minimum 2-norm solutions of the LS problems $\min_{x\in\mathbb{C}^n} \|Ax-b\|_2$ and $\min_{x\in\mathbb{C}^n} \|\widetilde{A}x-\widetilde{b}\|_2$, respectively. This goal is achieved via Theorem 3.2, where we obtain two expressions for $\widetilde{A}^\dagger$ in terms of $A^\dagger$, $(I+E)^{-1}$, and $(I+F)^{-1}$. Multiplicative perturbation theory of matrices has received considerable attention in the literature in the context of accurate computations of eigenvalues and singular values [21, 31, 32, 37, 38] and also in the context of accurate solution of linear systems of equations [17, Lemma 3.1] but, as far as we know, it has not been studied yet in the context of accurate solution of LS problems. First we present a technical result that is used in the proof of Theorem 3.2.

LEMMA 3.1. *Let $A \in \mathbb{C}^{m\times n}$ and $\widetilde{A} = (I+E)A(I+F) \in \mathbb{C}^{m\times n}$, where $(I+E) \in \mathbb{C}^{m\times m}$ and $(I+F) \in \mathbb{C}^{n\times n}$ are nonsingular matrices. Then the following equalities hold:*

(a) $P_A(I+E^*)(I-P_{\widetilde{A}}) = 0$.

(b) $(I-P_{\widetilde{A}^*})(I+F^*)P_{A^*} = 0$.

[†] The algorithm in [11] covers only the square case, but it is trivial to modify it for rectangular Cauchy matrices.

Proof. (a) Since $\mathcal{R}(\widetilde{A}) = \mathcal{R}((I+E)A)$ then $(I - P_{\widetilde{A}})(I+E)A = 0$. Thus, $(I - P_{\widetilde{A}})(I+E)AA^\dagger = (I - P_{\widetilde{A}})(I+E)P_A = 0$, which is equivalent to $P_A(I+E^*)(I-P_{\widetilde{A}}) = 0$.

(b) Apply (a) to $\widetilde{A}^* = (I+F^*)A^*(I+E^*)$ and conjugate and transpose the equality. □

Theorem 3.2 is the main result in this section and is valid for matrices with any rank.

THEOREM 3.2. *Let $A \in \mathbb{C}^{m\times n}$ and $\widetilde{A} = (I+E)A(I+F) \in \mathbb{C}^{m\times n}$, where $(I+E) \in \mathbb{C}^{m\times m}$ and $(I+F) \in \mathbb{C}^{n\times n}$ are nonsingular matrices, and set $\widehat{E} = (I+E)^{-1}E$ and $\widehat{F} = (I+F)^{-1}F$. Then*

$$\widetilde{A}^\dagger = P_{\widetilde{A}^*}(I+F)^{-1}A^\dagger(I+E)^{-1}P_{\widetilde{A}} \quad \textit{and} \tag{3.1}$$

$$\widetilde{A}^\dagger = \left(I + (I - P_{\widetilde{A}^*})F^* - P_{\widetilde{A}^*}\widehat{F}\right) A^\dagger \left(I + E^*(I - P_{\widetilde{A}}) - \widehat{E}P_{\widetilde{A}}\right). \tag{3.2}$$

Proof. We prove first (3.1). To this purpose, we define $Z := P_{\widetilde{A}^*}(I+F)^{-1}A^\dagger(I+E)^{-1}P_{\widetilde{A}}$ as the right hand side of (3.1) and use (2.1)-(i) for A and (2.1)-(ii) for $\widetilde{A}$ as follows:

$$\begin{aligned} Z &= \widetilde{A}^\dagger\widetilde{A}\,(I+F)^{-1}A^\dagger(I+E)^{-1}\widetilde{A}\widetilde{A}^\dagger = \widetilde{A}^\dagger\,(I+E)\,AA^\dagger A\,(I+F)\widetilde{A}^\dagger \\ &= \widetilde{A}^\dagger\,(I+E)\,A\,(I+F)\widetilde{A}^\dagger = \widetilde{A}^\dagger\,\widetilde{A}\,\widetilde{A}^\dagger = \widetilde{A}^\dagger\,. \end{aligned}$$

Next, we use (3.1) to prove (3.2). First, we write $(I+E)^{-1} = I - (I+E)^{-1}E = I - \widehat{E}$ and $(I+F)^{-1} = I - (I+F)^{-1}F = I - \widehat{F}$. Substituting these expressions in (3.1), we get

$$\widetilde{A}^\dagger = P_{\widetilde{A}^*}(I - \widehat{F})A^\dagger(I - \widehat{E})P_{\widetilde{A}} = P_{\widetilde{A}^*}(P_{A^*} - \widehat{F})A^\dagger(P_A - \widehat{E})P_{\widetilde{A}}. \tag{3.3}$$

From Lemma 3.1-(a) it follows that $P_A(I + E^*(I - P_{\widetilde{A}})) = P_A P_{\widetilde{A}}$. Analogously, from Lemma 3.1-(b), $((I - P_{\widetilde{A}^*})F^* + I)P_{A^*} = P_{\widetilde{A}^*}P_{A^*}$. Finally, substitute these relations in (3.3), use $A^\dagger P_A = A^\dagger$ and $P_{A^*}A^\dagger = A^\dagger$, and get (3.2). □

We emphasize that expression (3.2) ensures that under “small" multiplicative perturbations of A, i.e., small E and F, we obtain “small" multiplicative perturbations of $A^\dagger$.

The assumptions of Theorem 3.2 guarantee that $\mathrm{rank}\,(A) = \mathrm{rank}\,(\widetilde{A})$. This has simplified considerably the analysis of the variation of the Moore-Penrose pseudo-inverse with respect to general “additive” perturbations $\widetilde{A} = A + \Delta A$ [50, 52].

Corollary 3.3 presents an expression for $\widetilde{A}^\dagger - A^\dagger$ that follows directly from (3.2).

COROLLARY 3.3. *Under the assumptions of Theorem 3.2, we have*

$$\widetilde{A}^\dagger - A^\dagger = A^\dagger\Theta_E + \Theta_F A^\dagger + \Theta_F A^\dagger \Theta_E, \tag{3.4}$$

where $\quad \Theta_E = E^*(I - P_{\widetilde{A}}) - \widehat{E}P_{\widetilde{A}} \quad$ *and* $\quad \Theta_F = (I - P_{\widetilde{A}^*})F^* - P_{\widetilde{A}^*}\widehat{F}\,.$

4. Perturbation results for least squares problems. The error analysis of the new algorithm outlined in the Introduction requires to find perturbation bounds for the variation of the minimum 2-norm solution of a LS problem only to first order. Finite bounds for the minimum 2-norm solution and for the residual have been also obtained. They appear in [7] and in more generality in [8]. We will derive the perturbation bounds in two steps. First, for multiplicative perturbations, and then for additive perturbations of the factors of an RRD.

4.1. Multiplicative perturbation results for least squares problems. First we consider the LS problem

$$\min_{x\in\mathbb{C}^n} \|Ax - b\|_2, \quad A \in \mathbb{C}^{m\times n}, \quad b \in \mathbb{C}^m, \tag{4.1}$$

and the multiplicatively perturbed LS problem

$$\min_{x\in\mathbb{C}^n} \|\widetilde{A}x-\widetilde{b}\|_2, \quad \widetilde{A}=(I+E)A(I+F)\in\mathbb{C}^{m\times n}, \quad \widetilde{b}=b+h\in\mathbb{C}^m\,, \tag{4.2}$$

where $(I+E)\in\mathbb{C}^{m\times m}$ and $(I+F)\in\mathbb{C}^{n\times n}$ are nonsingular matrices. An upper bound, to first order, for the relative difference between the minimum 2-norm solutions $x_0=A^\dagger b$ and $\widetilde{x}_0=\widetilde{A}^\dagger\widetilde{b}$ of (4.1) and (4.2) is derived in Theorem 4.1, where it is understood that $x_0\neq 0$.

THEOREM 4.1. *Let x_0 and $\widetilde{x}_0$ be the minimum 2-norm solutions of (4.1) and (4.2), respectively. Assume $\|E\|_2\leq\mu<1$, $\|F\|_2\leq\nu<1$, and $\|h\|_2\leq\epsilon\|b\|_2$. Then, to first order in ϵ,μ,ν,*

$$\frac{\|\widetilde{x}_0-x_0\|_2}{\|x_0\|_2}\leq\sqrt{2}\,\nu+\left(\epsilon+\sqrt{2}\,\mu\right)\frac{\|A^\dagger\|_2\|b\|_2}{\|x_0\|_2}+\,h.o.t\,, \tag{4.3}$$

where $h.o.t$ stands for "higher order terms" in ϵ,μ,ν.

Proof. From Corollary 3.3:

$$\begin{aligned}\widetilde{x}_0-x_0&=\widetilde{A}^\dagger(b+h)-A^\dagger b=\left(\widetilde{A}^\dagger-A^\dagger\right)(b+h)+A^\dagger h\\&=\left(A^\dagger\,\Theta_E+\Theta_F\,A^\dagger+\Theta_F\,A^\dagger\,\Theta_E\right)(b+h)+A^\dagger h\\&=\left(A^\dagger\,\Theta_E+\Theta_F\,A^\dagger\,\Theta_E\right)(b+h)+\Theta_F\,x_0+\Theta_F\,A^\dagger h+A^\dagger h\,.\end{aligned}$$

with Θ_E and Θ_F defined as in Corollary 3.3. Next, apply norm inequalities to get

$$\|\widetilde{x}_0-x_0\|_2\leq\|\Theta_F\|_2\|x_0\|_2+\left[\|\Theta_E\|_2\,(1+\|\Theta_F\|_2)\,(1+\epsilon)+\epsilon\,(1+\|\Theta_F\|_2)\right]\,\|A^\dagger\|_2\,\|b\|_2\,.$$

From Lemma 2.1, $\|\Theta_E\|_2\leq\sqrt{\|E\|_2^2+\|\widehat{E}\|_2^2}$ and $\|\Theta_F\|_2\leq\sqrt{\|F\|_2^2+\|\widehat{F}\|_2^2}$, where $\widehat{E}$ and $\widehat{F}$ are defined as in Theorem 3.2. To first order, $\|\widehat{E}\|_2=\|E\|_2$ and $\|\widehat{F}\|_2=\|F\|_2$, and therefore $\|\Theta_E\|_2\leq\sqrt{2}\,\mu$ and $\|\Theta_F\|_2\leq\sqrt{2}\,\nu$. Hence, (4.3) follows. □

The bound in Theorem 4.1 improves significatively the classical bound [52, Theorem 5.1] of LS problems under general additive perturbations $\widetilde{A}=A+\Delta A$. To describe this bound, let $\widetilde{x}_0$ be the minimum 2-norm solution of the LS problem $\min_{x\in\mathbb{C}^n}\|(b+\Delta b)-(A+\Delta A)x\|_2$ and define $\epsilon_A:=\|\Delta A\|_2/\|A\|_2$ and $\epsilon_b:=\|\Delta b\|_2/\|b\|_2$. Then a minor modification of [4, Theorem 1.4.6] states, if $\operatorname{rank}(A)=\operatorname{rank}(\widetilde{A})$ and to first order in ϵ_A and ϵ_b, that

$$\frac{\|\widetilde{x}_0-x_0\|_2}{\|x_0\|_2}\leq 2\,\kappa_2(A)\,\epsilon_A+\frac{\|A^\dagger\|_2\,\|b\|_2}{\|x_0\|_2}\,\epsilon_b+\,\kappa_2(A)^2\,\frac{\|b-Ax_0\|_2}{\|A\|_2\,\|x_0\|_2}\,\epsilon_A. \tag{4.4}$$

In (4.4), the relative variation of the minimum 2-norm solution depends on $\kappa_2(A)$ and $\kappa_2(A)^2$. However, the bound in Theorem 4.1 is independent of $\kappa_2(A)$ and $\kappa_2(A)^2$. The only potentially large factor in (4.3) is $\|A^\dagger\|_2\,\|b\|_2/\|x_0\|_2$, present also in (4.4). This can be even larger than $\kappa_2(A)$. But, it can be shown [7, 8] that $\|A^\dagger\|_2\,\|b\|_2/\|x_0\|_2$ is a moderate number except for very particular choices of b. This was previously studied when A is a nonsingular matrix in [17, Section 3.2]. Although the fact that $A\in\mathbb{C}^{m\times n}$ is rectangular forces nontrivial modifications, the main conclusions remain the same. Therefore, we can say that (4.3) always improves the classical bound (4.4) for general additive perturbations and that if $\|E\|_2$, $\|F\|_2$, and $\|h\|_2$ are tiny, then (4.3) produces tiny bounds for $\|\widetilde{x}_0-x_0\|_2/\|x_0\|_2$ for almost all b.

The bound (4.3) is essentially optimal, since $\|A^\dagger\|_2\|b\|_2/\|x_0\|_2$ is up to a moderate constant the condition number under multiplicative perturbations of LS problems [7, 8].

Finally, observe that Theorem 4.1 is valid both if $m\geq n$ or if $m<n$. Thus, it is valid also for multiplicative perturbations of solutions of underdetermined linear systems.

4.2. Perturbation of least squares problems through factors. The error analysis in Section 5 will require to use a perturbation bound for minimum 2-norm solutions of LS problems whose coefficient matrix is given as an RRD under additive perturbations of each factor of the RRD. Such perturbation bound is obtained in Theorem 4.2.

THEOREM 4.2. *Let* $X \in \mathbb{C}^{m\times r}$*,* $D \in \mathbb{C}^{r\times r}$*, and* $Y \in \mathbb{C}^{r\times n}$ *be such that* $\operatorname{rank}(X) = \operatorname{rank}(Y) = r$ *and* D *is diagonal and nonsingular, and let* $b \in \mathbb{C}^m$*. Let* x_0 *be the minimum* 2*-norm solution of* $\min_{x\in\mathbb{C}^n} \|b - XDY\,x\|_2$*, and* $\widetilde{x}_0$ *be the minimum* 2*-norm solution of* $\min_{x\in\mathbb{C}^n} \|(b+h) - (X+\delta X)(D+\delta D)(Y+\delta Y)\,x\|_2$*, where* $\|\delta X\|_2 \le \alpha\|X\|_2$*,* $\|\delta Y\|_2 \le \beta\|Y\|_2$*,* $|\delta D| \le \rho|D|$*, and* $\|h\|_2 \le \epsilon\|b\|_2$*. Assume that*

$$\mu := \alpha\,\kappa_2(X) < 1 \quad \textit{and} \quad \nu := [\beta + \rho(1+\beta)]\kappa_2(Y) < 1. \tag{4.5}$$

Then to first order in α, β, ρ*, and* ϵ

$$\frac{\|\widetilde{x}_0 - x_0\|_2}{\|x_0\|_2} \le \sqrt{2}\,(\beta+\rho)\,\kappa_2(Y) + \left(\epsilon + \sqrt{2}\,\alpha\,\kappa_2(X)\right)\frac{\|Y^\dagger D^{-1} X^\dagger\|_2\|b\|_2}{\|x_0\|_2} + h.o.t. \tag{4.6}$$

Proof. Let us call $A = XDY$ and $\widetilde{A} = (X+\delta X)(D+\delta D)(Y+\delta Y)$. Let us write $\widetilde{A}$ as a multiplicative perturbation of A as follows

$$\begin{aligned}\widetilde{A} &= (I + \delta X X^\dagger)\,XD\,(I + D^{-1}\,\delta D)Y(I + Y^\dagger\delta Y)\\ &= (I + \delta X X^\dagger)\,XDY\,(I + Y^\dagger\,D^{-1}\,\delta D\,Y)\,(I + Y^\dagger\delta Y)\\ &=: (I+E)A(I+F),\end{aligned}$$

where $E = \delta X X^\dagger$ and $F = Y^\dagger\delta Y + Y^\dagger D^{-1}\,\delta D\,Y + Y^\dagger D^{-1}\,\delta D\,\delta Y$. Next, taking into account that $\|D^{-1}\,\delta D\|_2 \le \rho$, we get

$$\|E\|_2 \le \alpha\,\kappa_2(X) = \mu < 1, \quad \|F\|_2 \le [\beta + \rho(1+\beta)]\,\kappa_2(Y) = \nu < 1,$$

and Theorem 4.2 follows immediately from Theorem 4.1 and Lemma 2.2. □

Since the factors X and Y of an RRD are well conditioned, we see from (4.6) that the sensitivity with respect to perturbations of the factors of the minimum 2-norm solution of the LS problem $\min_{x\in\mathbb{C}^n} \|b - XDYx\|_2$ is again controlled by $\|A^\dagger\|_2\|b\|_2/\|x_0\|_2$, where $A^\dagger = Y^\dagger\,D^{-1}\,X^\dagger$, which is a moderate number for most vectors b as discussed above.

5. Algorithm and error analysis.

We present in this section Algorithm 5.1 for solving a LS problem $\min_{x\in\mathbb{C}^n} \|b - A\,x\|_2$ and we prove that it computes the minimum 2-norm solution with relative error bounded by $O(\mathtt{u})\,\|A^\dagger\|_2\,\|b\|_2/\|x_0\|_2$, which is simply $O(\mathtt{u})$ for most vectors b according to the discussion in [7, 8, 17]. The first step of the algorithm computes an accurate RRD of $A = XDY \in \mathbb{C}^{m\times n}$ in the sense of Definition 2.3. Next steps of Algorithm 5.1 are based on Lemma 2.2 and the following observations: (1) $x_1 = X^\dagger b$ is the unique solution of the *full column rank* LS problem $\min_{x\in\mathbb{C}^r} \|b - X\,x\|_2$; (2) $x_2 = D^{-1}(X^\dagger b)$ is the unique solution of the linear system $Dx = x_1$; and (3) $Y^\dagger(D^{-1}(X^\dagger b))$ is the minimum 2-norm solution of the *full row rank* underdetermined linear system $Yx = x_2$. Observe that this procedure is valid both if $m \ge n$ and if $m < n$. Therefore, in the latter case and if $\operatorname{rank}(A) = m$, the procedure solves accurately the underdetermined linear system $Ax = b$. The minimum 2-norm solution x_0 of the underdetermined system $Yx = x_2$ is computed via the Q-method described in [29, Sec. 21.1]. We are now in position of stating Algorithm 5.1.

```
ALGORITHM 5.1. (Accurate solution of LS problems via RRD)
Input:  $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$
```

```
Output : $x_0$ minimum 2-norm solution of $\min_{x\in\mathbb{C}^n}\|b - A\,x\|_2$

Step 1 : Compute an accurate RRD of $A = XDY$ in the sense of Definition 2.3,
         where $X \in \mathbb{C}^{m\times r}$, $D \in \mathbb{C}^{r\times r}$ is diagonal, $Y \in \mathbb{C}^{r\times n}$, and
         $\mathrm{rank}\,(A) = \mathrm{rank}\,(X) = \mathrm{rank}\,(Y) = \mathrm{rank}\,(D) = r$.
Step 2 : Compute the unique solution $x_1$ of $\min_{x\in\mathbb{C}^r}\|b - X\,x\|_2$ using the
         Householder $QR$ factorization of $X$.
Step 3 : Compute the unique solution $x_2$ of the diagonal linear system $D\,x = x_1$ as
         $x_2(i) = x_1(i)/d_{ii}$, $i = 1, \ldots, r$.
Step 4 : Compute the minimum 2-norm solution $x_0$ of $Yx = x_2$ using the $Q$ method,
         i.e., via Householder $QR$ factorization of $Y^*$.
```

The cost of Step 1 of Algorithm 5.1 depends on the specific type of matrices and algorithm used, as discussed in the Introduction. Anyway all these algorithms cost $O(mn^2)$ flops if $m \geq n$ and $O(m^2n)$ flops if $m < n$. The leading terms of the costs of Steps 2, 3, and 4 are $2r^2(m - r/3)$, r, and $2r^2(n - r/3)$ flops, respectively. Since $r \leq \min\{m, n\}$, the total cost of Algorithm 5.1 is $O(mn^2)$ flops if $m \geq n$ and $O(m^2n)$ flops if $m < n$.

The backward rounding errors committed by Algorithm 5.1 are analyzed in Theorem 5.2. We will use the following notation introduced in [29, Secs. 3.1 and 3.4]

$$\gamma_n := \frac{n\mathtt{u}}{1 - n\mathtt{u}} \quad \text{and} \quad \widetilde{\gamma}_n := \frac{cn\mathtt{u}}{1 - cn\mathtt{u}}, \tag{5.1}$$

where c denotes a small integer constant whose exact value is not essential in the analysis. Before proving Theorem 5.2, let us comment the meaning of its assumptions. First, we assume that the factors $\widehat{X}$, $\widehat{D}$, $\widehat{Y}$ computed in Step 1 in floating point arithmetic satisfy (2.3), (2.4), and (2.5), which imply $\mathrm{rank}\,(X) = \mathrm{rank}\,(\widehat{X}) = r$, $\mathrm{rank}\,(D) = \mathrm{rank}\,(\widehat{D}) = r$, $\mathrm{rank}\,(Y) = \mathrm{rank}\,(\widehat{Y}) = r$, and (2.6). Therefore, we can use $\kappa_2(X)$ and $\kappa_2(Y)$ in the errors of Steps 2 and 4 instead of $\kappa_2(\widehat{X})$ and $\kappa_2(\widehat{Y})$ at the cost of not paying attention to the exact values of the numerical constants in the error bounds. The assumption (5.2) guarantees that the backward errors $\Delta\widehat{X}$ on $\widehat{X}$ in Step 2 preserve the full rank, i.e., $\mathrm{rank}\,(\widehat{X}) = \mathrm{rank}\,(\widehat{X} + \Delta\widehat{X}) = r$, and the same for the backward errors on $\widehat{Y}$ in Step 4. The technical assumption (5.3) is needed for applying [29, Theorem 21.4] in the error analysis of Step 4.

We present in Theorem 5.2 two statements for the backward errors of Algorithm 5.1, one with respect to the computed factors $\widehat{X}$, $\widehat{D}$, and $\widehat{Y}$ of A and another with respect to the exact ones, which is the result to be used in practice. The reason for presenting these two statements is that the former gives stronger column-wise and row-wise backward errors in $\widehat{X}$ and $\widehat{Y}$, respectively, than the latter. This may be used to give stronger final backward errors for some particular classes of matrices, as Cauchy matrices. We do not follow this line here.

THEOREM 5.2. *Let $\widehat{X} \in \mathbb{C}^{m\times r}$, $\widehat{D} \in \mathbb{C}^{r\times r}$, and $\widehat{Y} \in \mathbb{C}^{r\times n}$ be the factors of A computed in* Step 1 *of Algorithm 5.1 and assume that they satisfy the error bounds* (2.3) *and* (2.4) *with respect to the exact factors X, D, and Y of A. Assume also that* (2.5),

$$\max\{\kappa_2(X), \kappa_2(Y)\}\,\sqrt{r}\,\widetilde{\gamma}_{r\max\{m,n\}} < 1, \qquad \textit{and} \tag{5.2}$$

$$\kappa_2(Y)\,n\,r^2\,\widetilde{\gamma}_n < 1 \tag{5.3}$$

hold. Let $\widehat{x}_0$ be the computed minimum 2-norm solution of $\min_{x\in\mathbb{C}^n}\|b - A\,x\|_2$ *using Algorithm 5.1 in finite precision with unit roundoff* $\mathtt{u}$. *Then the following statements hold.*

(a) $\widehat{x}_0$ *is the exact minimum 2-norm solution of*

$$\min_{x\in\mathbb{C}^n}\|(b + \Delta b) - (\widehat{X} + \Delta\widehat{X})(\widehat{D} + \Delta\widehat{D})(\widehat{Y} + \Delta\widehat{Y})\,x\|_2, \tag{5.4}$$

where

$$\|\Delta\widehat{X}(:,j)\|_2 \leq \widetilde{\gamma}_{mr}\,\|\widehat{X}(:,j)\|_2, \quad \|\Delta\widehat{Y}(j,:)\|_2 \leq \widetilde{\gamma}_{nr}\,\|\widehat{Y}(j,:)\|_2, \ \textit{for}\ j=1,\dots,r$$
$$|\Delta\widehat{D}| \leq \widetilde{\gamma}_1\,|\widehat{D}|, \qquad \|\Delta b\|_2 \leq \widetilde{\gamma}_{mr}\,\|b\|_2\,.$$

(b) $\widehat{x}_0$ *is the exact minimum 2-norm solution of*

$$\min_{x\in\mathbb{C}^n} \|(b+\Delta b)-(X+\Delta X)(D+\Delta D)(Y+\Delta Y)\,x\|_2, \tag{5.5}$$

where

$$\|\Delta X\|_2 \leq\ (\mathtt{u}\,p(m,n)+\sqrt{r}\,\widetilde{\gamma}_{mr}+\sqrt{r}\,\widetilde{\gamma}_{mr}\,\mathtt{u}\,p(m,n))\,\|X\|_2,$$
$$\|\Delta Y\|_2 \leq\ (\mathtt{u}\,p(m,n)+\sqrt{r}\,\widetilde{\gamma}_{nr}+\sqrt{r}\,\widetilde{\gamma}_{nr}\,\mathtt{u}\,p(m,n))\,\|Y\|_2,$$
$$|\Delta D| \leq\ (\mathtt{u}\,p(m,n)+\widetilde{\gamma}_1+\widetilde{\gamma}_1\,\mathtt{u}\,p(m,n))\,|D|, \qquad \|\Delta b\|_2 \leq \widetilde{\gamma}_{mr}\,\|b\|_2\,.$$

(c) *If* x_0 *is the exact minimum 2-norm solution of* $\min_{x\in\mathbb{C}^n}\|b-A\,x\|_2$, *then* $\|\widehat{x}_0 - x_0\|_2/\|x_0\|_2$ *can be bounded as in Theorem 4.2 with* $\alpha = (\mathtt{u}\,p(m,n)+\sqrt{r}\,\widetilde{\gamma}_{mr}+\sqrt{r}\,\widetilde{\gamma}_{mr}\,\mathtt{u}\,p(m,n))$, $\beta = (\mathtt{u}\,p(m,n)+\sqrt{r}\,\widetilde{\gamma}_{nr}+\sqrt{r}\,\widetilde{\gamma}_{nr}\,\mathtt{u}\,p(m,n))$, $\rho = (\mathtt{u}\,p(m,n)+\widetilde{\gamma}_1+\widetilde{\gamma}_1\,\mathtt{u}\,p(m,n))$, *and* $\epsilon = \widetilde{\gamma}_{mr}$. *In particular, to first order in* $\mathtt{u}$, *and if* c *is an small integer constant, then*

$$\frac{\|\widehat{x}_0-x_0\|_2}{\|x_0\|_2} \leq c\,\mathtt{u}\left[p_y(m,n)\,\kappa_2(Y)+p_x(m,n)\,\kappa_2(X)\,\frac{\|A^\dagger\|_2\|b\|_2}{\|x_0\|_2}\right]+O(\mathtt{u}^2)\,,$$

where $p_y(m,n) := (p(m,n)+nr^{3/2})$ *and* $p_x(m,n) := (p(m,n)+mr^{3/2})$.

Proof. In order to prove part (a) let us write the backward errors in steps 2, 3, and 4.

1. The backward errors of `Step 2` are given in [29, Theorem 20.3]: the solution computed in `Step 2`, $\widehat{x}_1$, is the exact solution of the LS problem

$$\min_{x\in\mathbb{C}^r} \|(b+\Delta b)-(\widehat{X}+\Delta\widehat{X})\,x\|_2\,, \tag{5.6}$$

where $\|\Delta\widehat{X}(:,j)\|_2 \leq \widetilde{\gamma}_{mr}\|\widehat{X}(:,j)\|_2$, for $j=1,\dots,r$, and $\|\Delta b\|_2 \leq \widetilde{\gamma}_{mr}\,\|b\|_2$. Therefore, $\|\Delta\widehat{X}\|_2 \leq \sqrt{r}\widetilde{\gamma}_{mr}\|\widehat{X}\|_2$. Recall that (2.3) and (2.5) imply $\operatorname{rank}(X) = \operatorname{rank}(\widehat{X}) = r$, so Weyl perturbation theorem [50] for singular values and (5.2) imply $|\sigma_r(\widehat{X}+\Delta\widehat{X})-\sigma_r(\widehat{X})|/\sigma_r(\widehat{X}) \leq \|\Delta\widehat{X}\|_2/\sigma_r(\widehat{X}) \leq \sqrt{r}\widetilde{\gamma}_{mr}\kappa_2(\widehat{X}) < 1$, and $\operatorname{rank}(\widehat{X}) = \operatorname{rank}(\widehat{X}+\Delta\widehat{X}) = r$. So, $\widehat{x}_1$ satisfies

$$\widehat{x}_1 = (\widehat{X}+\Delta\widehat{X})^\dagger(b+\Delta b), \tag{5.7}$$

with $\widehat{X}+\Delta\widehat{X}\in\mathbb{C}^{m\times r}$ and $\operatorname{rank}(\widehat{X}+\Delta\widehat{X}) = r$.

2. As a consequence of [29, Lemma 3.5], the solution, $\widehat{x}_2$, computed in `Step 3` obeys

$$(\widehat{D}+\Delta\widehat{D})\,\widehat{x}_2 = \widehat{x}_1 \quad \text{with} \quad |\Delta\widehat{D}| \leq \widetilde{\gamma}_1|\widehat{D}|, \tag{5.8}$$

with $\widehat{D}+\Delta\widehat{D}\in\mathbb{C}^{r\times r}$ diagonal and nonsingular, by (2.4), (2.5), and (5.2).

3. The backward errors of `Step 4` are given in [29, Theorem 21.4] under the conditions $\operatorname{rank}(\widehat{Y}) = r$, which follows from (2.3) and (2.5), and $\|\,|\widehat{Y}^\dagger|\,|\widehat{Y}|\,\|_2\,r\,n\,\gamma_n < 1$, which is guaranteed by (5.3), since $\|\,|\widehat{Y}^\dagger|\,|\widehat{Y}|\,\|_2\,r\,n\,\gamma_n \leq \kappa_2(\widehat{Y})\,r^2\,n\,\gamma_n < 1$. Under these conditions, the minimum 2-norm solution computed in `Step 4`, $\widehat{x}_0$, is the exact minimum 2-norm solution of the underdetermined system $(\widehat{Y}+\Delta\widehat{Y})x =$

$\widehat{x}_2$, with $\|\Delta\widehat{Y}(j,:)\|_2 \leq \widetilde{\gamma}_{nr}\|\widehat{Y}(j,:)\|_2$, for $j = 1,\ldots,r$. Besides, we can prove $\operatorname{rank}(\widehat{Y}) = \operatorname{rank}(\widehat{Y}+\Delta\widehat{Y}) = r$ via an argument similar to the one we used to prove the same for $\widehat{X}+\Delta\widehat{X}$. Therefore, $\widehat{x}_0$ obeys

$$\widehat{x}_0 = (\widehat{Y}+\Delta\widehat{Y})^{\dagger}\widehat{x}_2, \tag{5.9}$$

with $\widehat{Y}+\Delta\widehat{Y} \in \mathbb{C}^{r\times n}$ and $\operatorname{rank}(\widehat{Y}+\Delta\widehat{Y}) = r$.

From (5.7), (5.8), (5.9), and Lemma 2.2 we have that

$$\begin{aligned}\widehat{x}_0 &= (\widehat{Y}+\Delta\widehat{Y})^{\dagger}(\widehat{D}+\Delta\widehat{D})^{-1}(\widehat{X}+\Delta\widehat{X})^{\dagger}(b+\Delta b)\\ &= \Big[(\widehat{X}+\Delta\widehat{X})\,(\widehat{D}+\Delta\widehat{D})\,(\widehat{Y}+\Delta\widehat{Y})\Big]^{\dagger}(b+\Delta b)\,.\end{aligned}$$

This and the bounds we have developed for $\Delta\widehat{X}$, $\Delta\widehat{D}$, and $\Delta\widehat{Y}$ prove Theorem 5.2-(a).

The proof of Theorem 5.2-(b) follows easily from part-(a). Equations (2.3) and (2.4) allow us to write $\widehat{X} = X + E_X$, $\widehat{D} = D + E_D$, and $\widehat{Y} = Y + E_Y$, where $\|E_X\|_2 \leq \mathtt{u}\,p(m,n)\,\|X\|_2$, $|E_D| \leq \mathtt{u}\,p(m,n)|D|$, and $\|E_Y\|_2 \leq \mathtt{u}\,p(m,n)\|Y\|_2$. Therefore,

$$\widehat{X}+\Delta\widehat{X} = X + E_X + \Delta\widehat{X} =: X + \Delta X\,, \tag{5.10}$$

where

$$\begin{aligned}\|\Delta X\|_2 &\leq \|E_X\|_2 + \|\Delta\widehat{X}\|_2 \leq \mathtt{u}\,p(m,n)\,\|X\|_2 + \sqrt{r}\,\widetilde{\gamma}_{mr}\|\widehat{X}\|_2\\ &\leq \mathtt{u}\,p(m,n)\,\|X\|_2 + \sqrt{r}\,\widetilde{\gamma}_{mr}\,(\|X\|_2 + \|E_X\|_2)\\ &\leq \left(\mathtt{u}\,p(m,n) + \sqrt{r}\,\widetilde{\gamma}_{mr} + \sqrt{r}\,\widetilde{\gamma}_{mr}\,\mathtt{u}\,p(m,n)\right)\|X\|_2\,.\end{aligned} \tag{5.11}$$

Analogously, we can write

$$\begin{aligned}&\widehat{D}+\Delta\widehat{D} =: D+\Delta D\,, \quad \text{with } |\Delta D| \leq \left(\mathtt{u}\,p(m,n) + \widetilde{\gamma}_1 + \widetilde{\gamma}_1\,\mathtt{u}\,p(m,n)\right)|D|\,,\\ &\widehat{Y}+\Delta\widehat{Y} =: Y+\Delta Y\,, \quad \text{with } \|\Delta Y\|_2 \leq \left(\mathtt{u}\,p(m,n) + \sqrt{r}\,\widetilde{\gamma}_{nr} + \sqrt{r}\,\widetilde{\gamma}_{nr}\,\mathtt{u}\,p(m,n)\right)\|Y\|_2\,.\end{aligned}$$

If these equations and (5.10)-(5.11) are inserted into (5.4), then (5.5) is obtained and part (b) is proved. Finally, part (c) is an immediate consequence of part (b) and Theorem 4.2. □

Observe that, since in an RRD the factors X and Y are well conditioned, Theorem 5.2-(c) guarantees that the forward error in the solution computed by Algorithm 5.1 is bounded by $O(\mathtt{u})\|A^{\dagger}\|_2\|b\|_2/\|x_0\|_2$. Theorem 5.2-(c) proves the bound (1.2) in the Introduction.

6. Numerical experiments. Our numerical tests were done using MATLAB™ and they illustrate how well the errors committed by Algorithm 5.1 compare with the theoretical predictions and with the errors committed by the usual method to solve LS problems using the QR factorization computed with the Householder algorithm as implemented in MATLAB™ [29, Section 20.2]. We have done tests for three important classes of rectangular structured matrices that may have huge condition numbers: Cauchy, Vandermonde, and Graded matrices. Here we present only (to avoid a very lengthly article) results for Cauchy matrices, being the results for the other classes similar [7]. For matrices in these classes, accurate RRDs in the sense of Definition 2.3 can be computed using the algorithms in [11] and [28].

The entries of a Cauchy matrix $C \in \mathbb{R}^{m\times n}$, $m \geq n$, are defined in terms of two vectors $z = [z_1,\ldots,z_m]^T \in \mathbb{R}^m$, $y = [y_1,\ldots,y_n]^T \in \mathbb{R}^n$ as

$$c_{ij} = \frac{1}{z_i + y_j}, \qquad i = 1,\ldots,m,\ j = 1,\ldots,n. \tag{6.1}$$

Cauchy matrices have full column rank if $z_i \neq z_j$ for any $i \neq j$, $y_k \neq y_l$ for any $k \neq l$, and $z_i \neq -y_j$ for all i, j. We will choose z and y with these properties and, so, we consider only LS problems with unique solutions. Algorithm 3 in [11] uses a structured version of GECP to compute an accurate RRD of any *square* Cauchy matrix. This algorithm can be very easily extended to deal with rectangular matrices, and this version is the one used in the tests of this section to compute the RRD in `Step 1` of Algorithm 5.1. The overall cost of this step is $2mn^2 - 2n^3/3 + O(n^2 + mn)$ operations plus $mn^2/2 - n^3/6 + O(n^2 + mn)$ comparisons.

In order to make easy references, let us summarize and give names to the two algorithms that are used in this section for solving $\min_{x\in\mathbb{R}^n} \|Cx - b\|_2$, with C a Cauchy matrix:

- `LS-QR`: given vectors z and y, the entries of C are computed as in (6.1) and the LS problem is solved via the Householder QR factorization implemented in `MATLAB`™.
- `LS-RRD`: the LS problem is solved using Algorithm 5.1 with the RRD in `Step 1` computed with the rectangular version discussed above of Algorithm 3 in [11].

If $\widehat{x}_0$ is the unique solution of $\min_{x\in\mathbb{R}^n} \|Cx - b\|_2$ computed by any of the two algorithms, and x_0 is the exact solution, we know that the forward relative error committed by Algorithm 5.1 satisfies (1.2) (see Theorem 5.2-(c)). We also know that the computed solution by the QR algorithm in `MATLAB`™ has a forward relative error given in (1.1). This bound can be easily upper bounded to get [7, eq. (7.1)]

$$\frac{\|\widehat{x}_0 - x_0\|_2}{\|x_0\|_2} \leq c\, m\, n^{3/2}\, \mathrm{u}\, \kappa_2(A) \frac{\|A^\dagger\|_2 \|b\|_2}{\|x_0\|_2}, \tag{6.2}$$

which is a bound larger than (1.1) but simpler and reliable in most situations.

In our tests, we have generated Cauchy matrices with random z and y vectors, and random right-hand side vectors b. They have been chosen either from the uniform distribution in $[0, 1]$, or from the standard normal distribution. In all experiments we have tested the eight resulting possibilities in the choice of the random distributions for z, y, and b. We take as "exact" solution x_0 the one computed via the `svd` command of `MATLAB`™run in variable precision arithmetic. In each test we have set the precision to $2\log_{10}|D_1/D_n| + 30$ decimal digits, where D_1 and D_n are, respectively, the largest and the smallest (in absolute value) diagonal entries of the matrix D in the RRD of C computed in `Step 1` of Algorithm 5.1.

Two kind of experiments have been done. In the first group, we have fixed the size of the matrix: $m \times n = 100 \times 50, 50 \times 30$, or 25×10. For each size we have generated 50×8 different sets of the random vectors z, y, and b, therefore, generating a total of 400 different LS problems for each size. Figure 6.1 shows some of the results for the size 100×50. We observe that the relative error in the `LS-RRD` algorithm is always of order u times a small constant, as predicted, while the error for `LS-QR` scales almost linearly with $\kappa_2(C)$ until it saturates. The linear dependence on $\kappa_2(C)$ of the relative error in `LS-QR` is the predicted by (6.2) since $\|C^\dagger\|_2\|b\|_2/\|x_0\|_2$ has been always moderate in these tests. Besides this, we have computed the right hand sides of (1.2) and (6.2) suppressing the dimensional constants. In all the experiments both bounds are rather sharp and do not overestimate the actual errors. For other sizes and other ways to generate z, y, and b the results have been similar.

We have performed a second group of tests where we have fixed the number of rows of the matrix and varied the number of the columns. We have tested matrices of sizes $m \times n$ with $m = 100,\ n = 10:10:90$ (5×8 sets of random vectors z, y, and b for each size), $m = 50,\ n = 10:2:40$ (10×8 sets of random vectors z, y, and b for each size), and $m = 25,\ n = 5{:}5{:}20$ (20×8 sets of random vectors z, y, and b for each size). This makes a total of 2280 matrices. Our results also agree here with (1.2) and (6.2).

For all our experiments with Cauchy matrices, the range of the condition numbers has been $10^0 \lesssim \kappa_2(C) \lesssim 10^{100}$, the maximum value of the term $\|C^\dagger\|_2\|b\|_2/\|x_0\|_2$ has been 1376, $8 \leq \kappa_2(X) \leq 72$, and $13 \leq \kappa_2(Y) \leq 58$.

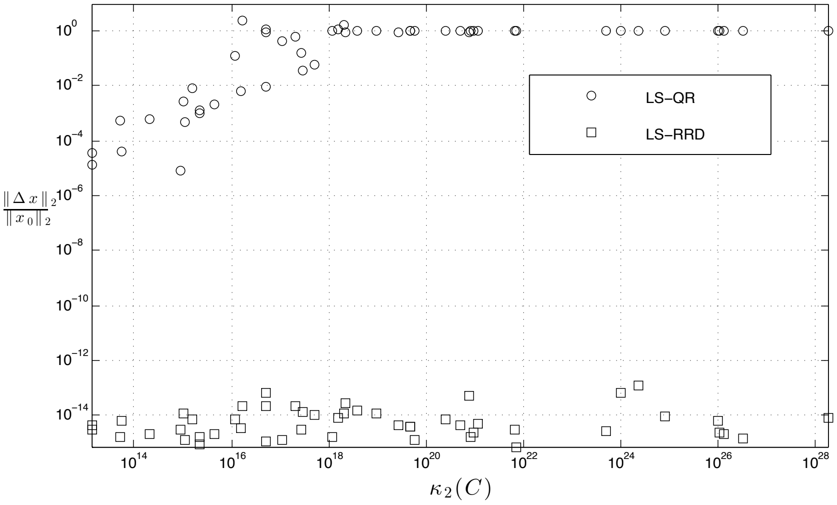

FIG. 6.1. *Forward relative error* $\|\widehat{x}_0 - x_0\|_2/\|x_0\|_2$ *against* $\kappa_2(C)$. C *are random* 100×50 *Cauchy matrices. The vectors* z *and* b *are selected from the standard normal distribution and the vector* y *from the uniform distribution in* $[0,1]$. *In these tests,* $\kappa_2(C)$ *has been computed via high precision arithmetic in* MATLAB*TM*.

7. Conclusions and future work. In this paper we have introduced, and carefully analyzed, a new algorithm to compute accurate solutions of those least squares problems $\min_{x\in\mathbb{C}^n} \|Ax - b\|_2$ such that an accurate rank-revealing decomposition of the coefficient matrix A can be computed. This is nowadays possible for many classes of structured matrices [12, 17, 7] that may have extremely large traditional condition numbers, and, probably, it will be possible for more classes in the future. In addition, the new algorithm can be also applied to compute accurate minimum 2-norm solutions of underdetermined linear systems. This work together with the previous papers [12, 16, 17] show that, for those matrices for which accurate rank-revealing decompositions can be computed, we can perform accurately and efficiently almost all basic tasks of Numerical Linear Algebra, i.e., solution of linear systems, solution of least squares problems, computation of eigenvalues and eigenvectors of symmetric matrices, and computation of the singular value decomposition, and to obtain relative errors of order $\mathbf{u}$ for very ill-conditioned problems where standard algorithms fail to provide even a single correct digit of accuracy. The only basic problem that is excluded from this framework is the nonsymmetric eigenvalue problem. To investigate at which extent rank-revealing decompositions allow us to solve accurately nonsymmetric eigenvalue problems will be the subject of our future research.

Acknowledgements. The authors thank Nick Higham for providing the codes of the algorithms and numerical tests presented in [28]. The authors also thank the Associate Editor of the paper and two anonymous Referees for many suggestions and valuable comments that have improved the paper considerably. Froilán M. Dopico wishes to especially thank Rodolfo Cuerno and Rocío Ortega for all the encouragement provided during the completion of this paper. Froilán finishes this paper *“with a big help from his Friends”*.

REFERENCES

[1] J. M. BANOCZI, N.-C. CHIU, G. E. CHO, AND I. C. F. IPSEN, *The lack of influence of the right-hand side on the accuracy of linear system solution*, SIAM J. Sci. Comput., 20 (1998), pp. 203–227.

[2] J. BARLOW, *Error analysis and implementation aspects of deferred correction for equality constrained least squares problems*, SIAM J. Num. Anal., 25 (1988), pp. 1340–1358.

[3] J. BARLOW AND J. DEMMEL, *Computing accurate eigensystems of scaled diagonally dominant matrices*, SIAM J. Num. Anal., 27 (1990), pp. 762–791.

[4] Å. BJÖRCK, *Numerical methods for least squares problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[5] Å. BJÖRCK AND V. PEREYRA, *Solution of Vandermonde systems of equations*, Math. Comp., 24 (1970), pp. 893–903.

[6] S. L. CAMPBELL AND C. D. MEYER, JR., *Generalized inverses of linear transformations*, Dover Publications Inc., New York, 1991. Corrected reprint of the 1979 original.

[7] N. CASTRO-GONZÁLEZ, J. CEBALLOS, F. M. DOPICO, AND J. M. MOLERA, *Multiplicative perturbation theory and accurate solution of least squares problems*, technical report, `http://gauss.uc3m.es/web/personal_web/fdopico/index_sp.html` (2013).

[8] ———, *Multiplicative perturbation theory of the Moore-Penrose inverse and the least squares problem*, in preparation, (2013).

[9] T. F. CHAN AND D. E. FOULSER, *Effectively well-conditioned linear systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 963–969.

[10] A. J. COX AND N. J. HIGHAM, *Stability of Householder QR factorization for weighted least squares problems*, in Numerical Analysis 1997, Proceedings of the 17th Dundee Conference, D. Griffiths, D. Higham, and G. Watson, eds., Harlow, Essex, UK, 1998, Addison-Wesley-Longman, pp. 57–73.

[11] J. DEMMEL, *Accurate singular value decompositions of structured matrices*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 562–580.

[12] J. DEMMEL, M. GU, S. EISENSTAT, I. SLAPNIČAR, K. VESELIĆ, AND Z. DRMAČ, *Computing the singular value decomposition with high relative accuracy*, Linear Algebra Appl., 299 (1999), pp. 21–80.

[13] J. DEMMEL AND W. KAHAN, *Accurate singular values of bidiagonal matrices*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 873–912.

[14] J. DEMMEL AND K. VESELIĆ, *Jacobi's method is more accurate than QR*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 1204–1246.

[15] J. W. DEMMEL, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[16] F. M. DOPICO, P. KOEV, AND J. M. MOLERA, *Implicit standard Jacobi gives high relative accuracy*, Numer. Math., 113 (2009), pp. 519–553.

[17] F. M. DOPICO AND J. M. MOLERA, *Accurate solution of structured linear systems via rank-revealing decompositions*, IMA J. Numer. Anal., 32 (2012), pp. 1096–1116.

[18] F. M. DOPICO, J. M. MOLERA, AND J. MORO, *An orthogonal high relative accuracy algorithm for the symmetric eigenproblem*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 301–351.

[19] Z. DRMAČ AND K. VESELIĆ, *New fast and accurate Jacobi SVD algorithm. I*, SIAM Journal on Matrix Analysis and Applications, 29 (2008), pp. 1322–1342.

[20] ———, *New fast and accurate Jacobi SVD algorithm. II*, SIAM Journal on Matrix Analysis and Applications, 29 (2008), pp. 1343–1362.

[21] S. EISENSTAT AND I. IPSEN, *Relative perturbation techniques for singular value problems*, SIAM J. Numer. Anal., 32 (1995), pp. 1972–1988.

[22] D. FADDEEV, V. KUBLANOVSKAYA, AND V. FADDEEVA, *Solution of linear algebraic systems with rectangular matrices*, Proc. Steklov Inst. Math, 96 (1968), pp. 93–111.

[23] K. FERNANDO AND B. PARLETT, *Accurate singular values and differential qd algorithms*, Numer. Math., 67 (1994), pp. 191–229.

[24] G. GOLUB AND C. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 3rd ed., 1996.

[25] M. GU AND S. C. EISENSTAT, *Efficient algorithms for computing a strong rank-revealing QR factorization*, SIAM J. Sci. Comput., 17 (1996), pp. 848–869.

[26] P. C. HANSEN, *Rank-deficient and discrete ill-posed problems*, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[27] R. HANSON AND C. LAWSON, *Extensions and applications of the Householder algorithm for solving linear least squares problems*, Math. Comp., 23 (1969), pp. 787–812.

[28] N. J. HIGHAM, *QR factorization with complete pivoting and accurate computation of the SVD*, Linear Algebra Appl., 309 (2000), pp. 153–174.

[29] ———, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2002.

[30] P. D. HOUGH AND S. A. VAVASIS, *Complete orthogonal decomposition for weighted least squares*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 369–392.

[31] I. C. F. IPSEN, *Relative perturbation results for matrix eigenvalues and singular values*, in Acta numerica, 1998, vol. 7 of Acta Numer., Cambridge Univ. Press, Cambridge, 1998, pp. 151–201.

[32] ———, *An overview of relative* $\sin\Theta$ *theorems for invariant subspaces of complex matrices*, J. Comput. Appl. Math., 123 (2000), pp. 131–153. Numerical analysis 2000, Vol. III. Linear algebra.

[33] W. KAHAN, *Accurate eigenvalues of a symmetric tridiagonal matrix*, Computer Science Dept. Technical Report CS41, Stanford University, Stanford, CA, July 1966 (revised June 1968).

[34] P. KOEV, *Accurate eigenvalues and SVDs of totally nonnegative matrices*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 1–23.

[35] D. KRESSNER, *Numerical methods for general and structured eigenvalue problems*, vol. 46 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2005.

[36] C. LAWSON AND R. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliff, NJ, 1974.

[37] R.-C. LI, *Relative perturbation theory. I. Eigenvalue and singular value variations*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 956–982.

[38] ———, *Relative perturbation theory. II. Eigenspace and singular subspace variations*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 471–492.

[39] ———, *Asymptotically optimal lower bounds for the condition number of a real Vandermonde matrix*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 829–844.

[40] ———, *Lower bounds for the condition number of a real confluent Vandermonde matrix*, Math. Comp., 75 (2006), pp. 1987–1995.

[41] ———, *Vandermonde matrices with Chebyshev nodes*, Linear Algebra Appl., 428 (2008), pp. 1803–1832.

[42] A. MARCO AND J.-J. MARTÍNEZ, *Polynomial least squares fitting in the Bernstein basis*, Linear Algebra Appl., 433 (2010), pp. 1254–1264.

[43] L. MIRANIAN AND M. GU, *Strong rank revealing LU factorizations*, Linear Algebra Appl., 367 (2003), pp. 1–16.

[44] V. OLSHEVSKY, ed., *Structured matrices in mathematics, computer science, and engineering. I*, vol. 280 of Contemporary Mathematics, Providence, RI, 2001, American Mathematical Society.

[45] ———, ed., *Structured matrices in mathematics, computer science, and engineering. II*, vol. 281 of Contemporary Mathematics, Providence, RI, 2001, American Mathematical Society.

[46] C.-T. PAN, *On the existence and computation of rank-revealing LU factorizations*, Linear Algebra Appl., 316 (2000), pp. 199–222.

[47] M. POWELL AND J. REID, *On applying Householder transformations to linear least squares problems*, in Information Processing 68, Proc. International Federation of Information Processing Congress, Edinburgh, 1968, North Holland, Amsterdam, 1969, pp. 122–126.

[48] I. SLAPNIČAR, *Highly accurate symmetric eigenvalue decomposition and hyperbolic SVD*, Linear Algebra Appl., 358 (2003), pp. 387–424.

[49] G. W. STEWART, *Matrix algorithms. Vol. I. Basic decompositions*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998.

[50] G. W. STEWART AND J.-G. SUN, *Matrix Perturbation Theory*, Academic Press, New York, 1990.

[51] D. S. WATKINS, *The matrix eigenvalue problem: GR and Krylov subspace methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2007.

[52] P.-Å. WEDIN, *Perturbation theory for pseudo-inverses*, BIT, 13 (1973), pp. 217–232.

[53] Q. YE, *Computing singular values of diagonally dominant matrices to high relative accuracy*, Math. Comp., 77 (2008), pp. 2195–2230.