SIAM J. MATRIX ANAL. APPL.
Vol. 20, No. 2, pp. 354–362

A STABLE AND EFFICIENT ALGORITHM FOR THE INDEFINITE LINEAR LEAST-SQUARES PROBLEM*

S. CHANDRASEKARAN†, M. GU‡, AND A. H. SAYED§

Abstract. We develop an algorithm for the solution of indefinite least-squares problems. Such problems arise in robust estimation, filtering, and control, and numerically stable solutions have been lacking. The algorithm developed herein involves the QR factorization of the coefficient matrix and is provably numerically stable.

Key words. backward stability, error analysis, indefinite least-squares problems

AMS subject classifications. 15A06, 65F05, 65G05

PII. S0895479896302229

1. Introduction. Many optimization criteria have been used for parameter estimation, starting with the standard least-squares formulation of Gauss (ca. 1795) and moving to more recent works on total least-squares (TLS) and robust (or H^∞) estimation (see, e.g., [4, 5, 7, 8, 10, 11]). The latter formulations have been motivated by an increasing interest in estimators that are less sensitive to data uncertainties and measurement errors. They can both be shown to require the minimization of indefinite quadratic forms, where the standard inner product of two vectors, say $a^T b$, is replaced by an indefinite inner product of the form $a^T Jb$ for a given *signature matrix*

$$J = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix},$$

where I_p and I_q are the identity matrices of dimensions p and q, respectively.

In this paper, we consider the *indefinite least-squares problems* of the form

$$\min_x \ (A\,x - b)^T \ J \ (A\,x - b)\,, \tag{1.1}$$

where $A \in R^{m \times n}$ is a given matrix with $m \geq n$; $b \in R^m$ is a given vector; and $p + q = m$. This problem reduces to the standard linear least-squares problem when $q = 0$. This is a characteristic of the so-called Krein spaces [5, 10].

Contrary to standard least-squares problems that always have solutions, the introduction of J with both positive and negative inertia can lead to minimization problems that are not necessarily solvable. Under certain solvability conditions, however, they lead to normal equations with positive-definite coefficient matrices. In this paper, we propose an algorithm for the solution of (1.1). We show that it is backward stable.

In section 2 we discuss situations where problem (1.1) might arise. In section 3 we solve problem (1.1). In section 4 we perform an error analysis.

*Received by the editors April 17, 1996; accepted for publication (in revised form) by N. J. Higham July 8, 1997; published electronically October 20, 1998.

http://www.siam.org/journals/simax/20-2/30222.html

†Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106 (shiv@ece.ucsb.edu).

‡Department of Mathematics, University of California, Los Angeles, CA 90095 (mgu@math.ucla.edu).

§Department of Electrical Engineering, University of California, Los Angeles, CA 90095 (sayed@ee.ucla.edu).

Throughout this paper, a *flop* is a real floating-point operation $\alpha \circ \beta$, where α and β are real floating-point numbers and $\circ$ is one of the operations $+$, $-$, $\times$, or $\div$. In our error analysis, we assume the following model for floating-point arithmetic:

$$\mathbf{fl}(\alpha \circ \beta) = (\alpha\,(1+\eta_1)) \circ (\beta\,(1+\eta_2)),$$

where $\mathbf{fl}(\alpha \circ \beta)$ is the floating-point result of the operation $\circ$, and $|\eta_i| \le \epsilon$ with ϵ being the machine precision. For simplicity, we ignore the possibility of overflow and underflow.

2. Motivation of indefinite quadratic forms. We briefly indicate in this section how indefinite quadratic forms arise in the context of TLS and robust estimation methods.

Let $A \in R^{m\times n}$ be a given matrix with $m \ge n$, and let $b \in R^m$ be a given vector, which are assumed to be linearly related via an unknown vector of parameters $x \in R^n$,

$$b = A\,x + v. \tag{2.1}$$

The vector $v \in R^m$ explains the mismatch between Ax and the given vector (or observation) b.

2.1. The TLS problem. The TLS method has been devised to deal with data errors in both A and b; it incorporates possible errors in the matrix A into the problem formulation. More specifically, given (A, b) and assuming that both data quantities are noisy, the TLS problem seeks a matrix $\widehat{A}$ and a vector $\widehat{x}$ that minimize the following optimization problem (defined in terms of the Frobenius norm):

$$\min_{\widehat{A},\widehat{x}} \left\| \begin{bmatrix} \widehat{A} - A & \widehat{A}\,\widehat{x} - b \end{bmatrix} \right\|_F^2 \iff \min_{\widehat{A},\widehat{b}\in\mathcal{R}(\widehat{A})} \left\| \begin{bmatrix} A & b \end{bmatrix} - \begin{bmatrix} \widehat{A} & \widehat{b} \end{bmatrix} \right\|_F^2. \tag{2.2}$$

The optimal solution $\widehat{A}$ is regarded as an approximation for A, which in turn is used to determine an $\widehat{x}$ that guarantees $\widehat{b} \in \mathcal{R}(\widehat{A})$. The solution of the TLS problem is known to be given by the following construction [7, p. 36].

Assume A is $m \times n$ with $m > n$ (i.e., A is a nonsquare matrix). Let $\{\sigma_1, \ldots, \sigma_n\}$ denote the singular values of A, with $\sigma_1 \ge \cdots \ge \sigma_n \ge 0$. Also, let $\{\bar{\sigma}_1, \ldots, \bar{\sigma}_n, \bar{\sigma}_{n+1}\}$ denote the singular values of the extended matrix $\begin{bmatrix} A & b \end{bmatrix}$. If $\bar{\sigma}_{n+1} < \sigma_n$, then the unique solution $\widehat{x}$ of (2.2) is given by

$$\widehat{x} = \left(A^T\,A - \bar{\sigma}_{n+1}^2 I_n\right)^{-1}\,A^T\,b.$$

This form of the solution shows that the TLS solution can also be obtained by minimizing the *indefinite* quadratic cost function

$$\min_x \left[\,\|b - A\,x\|_2^2 \;-\; \bar{\sigma}_{n+1}^2 \|\,x\,\|_2^2\,\right].$$

The cost function can be rewritten in the form

$$\min_x \left(\begin{bmatrix} b \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ \bar{\sigma}_{n+1} \end{bmatrix} x \right)^T \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} \left(\begin{bmatrix} b \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ \bar{\sigma}_{n+1} \end{bmatrix} x \right),$$

where I_n denotes the identity matrix of size $n \times n$. This is a special case of the indefinite quadratic cost function to be studied in this paper (see (1.1)). A similar cost function also arises in the solution of a least-squares problem with bounded errors-in-variables [3].

2.2. Robust or H$^\infty$-smoothing. In recent years there has been an interest in (suboptimal) min-max estimation, with the belief that the resulting so-called robust or H$^\infty$ algorithms will be more robust and less sensitive to modeling assumptions (e.g., [8, 11]). In this section, we review the H$^\infty$-smoothing formulation, which can be shown to include as a special case the standard least-squares solution. The application to parameter estimation given in this section follows [5, 10].

Consider again the model (2.1). Assume that an arbitrary vector $\widehat{x}$ is picked as an estimate for the unknown x. Then, no matter what the given (A, b) are, it is always possible to find a vector $\widehat{v}$ that matches (2.1), i.e., that satisfies

$$b = A\,\widehat{x} + \widehat{v}\,.$$

The particular choice $\widehat{x}$ induces an error norm $\|x - \widehat{x}\|_2$ and a noise norm $\|\widehat{v}\|_2$. But since $\widehat{x}$ has been picked arbitrarily, these norms may be arbitrarily large or small. That is, the estimate may be good or bad, and one would like to develop a procedure that picks an estimate that always guarantees a certain level of performance.

To clarify this point even further, consider the case when the norm of the original perturbation v in (2.1) is small. In this case, the data vector b is only a slight perturbation apart from Ax. So one expects in this situation to be able to come up with a better estimate for x than in the case when the noise v is large. In other words, one would like to define a procedure that picks an $\widehat{x}$ in such a way that if the original perturbation v is small, then so will be the resulting error $(x - \widehat{x})$.

This idea can be formalized and leads to a so-called robust estimation problem. In this context, one seeks an estimate $\widehat{x}$ (affine in b, say $\widehat{x} = Kb + k$ for some $K \in R^{n\times m}, k \in R^n$) in order to guarantee that the following bound holds irrespective of the nature of the noise component v:

$$\text{find } \widehat{x} \text{ such that} \qquad \max_{v\neq 0} \frac{\| x - \widehat{x} \|_2^2}{\| v \|_2^2} \;\leq\; \gamma^2 \tag{2.3}$$

for a specified value of γ (say $\gamma = 1$ or some other value). The resulting estimate $\widehat{x}$, when it exists (and this depends on the value of γ), will guarantee that the maximum 2-norm gain from the disturbance v to the estimation error $(x - \widehat{x})$ will always be less than γ^2; hence the qualification "robust" estimate since it guarantees that if the disturbance v is small, then so will be the estimation error.

It is not difficult to see that, since $v = b - Ax$, an alternative way of requiring expression (2.3) to hold is to equivalently require the *indefinite* quadratic cost function

$$\mathcal{J} = \| A\,x - b \|_2^2 \; - \gamma^{-2} \,\| x - \widehat{x} \|_2^2$$

to be nonnegative for all x. That is, the optimization problem (2.3) over v is now an optimization problem over x. We shall not pursue in detail the complete solution of the robust smoothing problem here (instead, see [2, 5, 10]). We only note that $\mathcal{J}$ can be rewritten in the form

$$\mathcal{J} = \left(\begin{bmatrix} c \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ \gamma^{-1} \end{bmatrix} z \right)^T \begin{bmatrix} I_n & 0 \\ 0 & -I_n \end{bmatrix} \left(\begin{bmatrix} c \\ 0 \end{bmatrix} - \begin{bmatrix} A \\ \gamma^{-1} \end{bmatrix} z \right),$$

where $z = x - \widehat{x}$ and $c = b - A\widehat{x}$. This is again a special case of (1.1).

3. Solution of the indefinite least-squares problem. It is known that (see [5]) problem (1.1) has a unique solution if and only if

$$A^T JA \quad \text{is symmetric positive-definite.} \tag{3.1}$$

If this condition does not hold, then (1.1) can either have no solution or infinitely many solutions. We shall assume throughout this paper that condition (3.1) holds and, therefore, that problem (1.1) has a unique solution. In particular, condition (3.1) implies that $p \geq n$.

To solve (1.1), we first note that the quadratic cost function can be rewritten as

$$\begin{aligned}(A\,x-b)^T\;J\;(A\,x-b) &= x^T\;\left(A^T\;J\;A\right)\;x-2\left(A^T\;J\;b\right)^T\;x+b^T\;J\;b\\ &= (x-x_s)^T\;\left(A^T\;J\;A\right)\;(x-x_s)\\ &\quad +b^T\;J\;b-\left(A^T\;J\;b\right)^T\;\left(A^T\;J\;A\right)^{-1}\;\left(A^T\;J\;b\right),\end{aligned}$$

where x_s is the unique solution of the linear system of equations

$$\left(A^T\;J\;A\right)\;x_s=\left(A^T\;J\;b\right). \tag{3.2}$$

It follows from condition (3.1) that x_s is the unique solution to (1.1). Parallel to the usual least-squares problem, we refer to (3.2) as the *normal equation* associated with (1.1).

One straightforward approach to solving (3.2) is to directly form the coefficient matrix and the right-hand side, and then solve the equation by computing the Cholesky factorization of the coefficient matrix. However, this approach is in general *not* backward stable even for the usual least-squares problem, where J is the identity matrix (see [4, Chap. 5]).

In the following, we derive a new stable algorithm for computing x_s. We first compute the QR factorization of the matrix A, say

$$A=\begin{pmatrix} Q_1\\ Q_2\end{pmatrix} R = Q\;R\,,$$

where $R \in R^{n\times n}$ is upper triangular; $Q_1 \in R^{p\times n}$ and $Q_2 \in R^{q\times n}$; and $Q = (Q_1^T \quad Q_2^T)^T$ is column orthogonal, i.e.,

$$Q^T\;Q=Q_1^T\;Q_1+Q_2^T\;Q_2=I_n.$$

It follows from condition (3.1) that R is nonsingular. For our purposes, we explicitly compute the matrix Q as well. The cost of this decomposition is about $4n^2m-\frac{4}{3}n^3$ flops.

Now substituting the QR factorization of A into (3.2) and simplifying, we obtain

$$\left(Q^T\;J\;Q\right)\;R\,x_s=\left(Q_1^T\;Q_1-Q_2^T\;Q_2\right)R\,x_s=Q^T\;J\;b. \tag{3.3}$$

We remark that the fact that Q is orthogonal is *not* needed to get (3.3). In fact, this equation is equivalent to (3.2) for any factorization $A = QR$ as long as $R \in R^{n\times n}$ is nonsingular. This fact will be very important for our error analysis in section 4.

We also remark that condition (3.1) implies that the matrix

$$2\,Q_1^T\;Q_1-I_n=Q_1^T\;Q_1-Q_2^T\;Q_2$$

is symmetric positive-definite. Hence the singular values of Q_1 are all between $1/\sqrt{2}$ and 1. In other words, Q_1 is very well-conditioned, even though the matrix $Q_1^T Q_1 - Q_2^T Q_2$ itself could be very ill-conditioned.

To solve the linear system of equations (3.3), we form the matrix $Q_1^T Q_1 - Q_2^T Q_2$ explicitly and then compute its Cholesky factorization

$$Q_1^T \, Q_1 - Q_2^T \, Q_2 = L \, L^T,$$

where $L \in R^{n \times n}$ is lower triangular. The cost of this decomposition is about $n^2 m + \frac{1}{3} n^3$ flops. To compute x_s, we then compute the right-hand side in (3.3) and perform one forward and two backward substitutions. These computations cost about $O(mn)$ flops. Hence the total cost for computing x_s is about $(5m - n)n^2$ flops. We shall establish in section 4.2 that the proposed algorithm is backward stable.

Remark. A referee pointed out that an alternative method to solve the indefinite least-squares problem (1.1) can be derived by using hyperbolic Householder transforms (see Berry and Cybenko [1] and Rader and Steinhardt [9]). This alternative is potentially less expensive than the one in section 3, although it might not be backward stable.

4. Error analysis. We now perform an error analysis for the proposed solution to the indefinite least-squares problem. We begin with some definitions and well-known results.

4.1. Preliminaries. We use the definition of stability in [12, pp. 75–76]. Let $\mathcal{F}(\mathcal{X})$ be a function of the input data $\mathcal{X}$. We say that an algorithm for computing $\mathcal{F}(\mathcal{X})$ is *backward stable* if its output is exactly $\mathcal{F}(\bar{\mathcal{X}})$, where $\bar{\mathcal{X}}$ is a small perturbation of $\mathcal{X}$.

Let $A \in R^{m \times n}$ and $B \in R^{n \times l}$. When the matrix–matrix product $A \cdot B$ is computed in the straightforward way, the computed product $\mathbf{fl}\,(A \cdot B)$ satisfies (see [4, pp. 66–68])

$$\mathbf{fl}\,(A \cdot B) = A \cdot B + O\left(\epsilon \cdot \|A\|_2 \, \|B\|_2\right).$$

Let $b \in R^m$. When the matrix–vector product $A \cdot b$ is computed in the straightforward way, the computed product satisfies (see [6, Chap. 3])

$$\mathbf{fl}\,(A \cdot b) = (A + \delta A) \cdot b,$$

where $\|\delta A\|_2 = O(\epsilon \cdot \|A\|_2)$. Let

$$A + \delta \widehat{A} = \widehat{Q} \, \widehat{R}$$

be the computed QR factorization of A (say, by Householder transformations) with $\delta \widehat{A} \in R^{m \times n}$, $\widehat{Q} \in R^{m \times n}$, and $\widehat{R} \in R^{n \times n}$. Then $\widehat{R}$ is upper triangular. The computed QR factorization is stable in that (see [6, Chap. 18])

$$\widehat{Q}^T \, \widehat{Q} = I_n + \Delta_1 \quad \text{with} \quad \Delta_1 = \Delta_1^T = O(\epsilon) \in R^{n \times n} \quad \text{and} \quad \|\delta \widehat{A}\|_2 = O(\epsilon \cdot \|A\|_2).$$

Let $M \in R^{n \times n}$ be a symmetric positive-definite matrix, and assume that a numerical Cholesky factorization of M can be successfully computed

$$M + \delta M = \widehat{L} \widehat{L}^T,$$

where $\widehat{L} \in R^{n\times n}$ is lower triangular. Then δM is symmetric and satisfies $\|\delta M\|_2 = O(\epsilon \cdot \|M\|_2)$. For details see [6, Chap. 10].

Let $R \in R^{n\times n}$ be a nonsingular upper (or lower) triangular matrix; let $b \in R^n$ be a vector; and let $\widehat{x}$ be the computed solution via backward (or forward) substitution to the linear system of equations

$$Rx = b.$$

Then $\widehat{x}$ satisfies

$$(R + \delta R)\,\widehat{x} = b, \tag{4.1}$$

where (see [6, Chap. 8])

$$|\delta R| \leq \frac{n\,\epsilon}{1 - n\,\epsilon} \cdot |R|.$$

Here $|\delta R|$ and $|R|$ are matrices of moduli of δR and R, respectively, and the inequality is meant entrywise. It follows that $\delta R = O(\epsilon \cdot \|R\|_2)$.

4.2. Analysis of the indefinite least-squares solution. Let

$$A + \delta\widehat{A} = \begin{pmatrix} \widehat{Q}_1 \\ \widehat{Q}_2 \end{pmatrix} \widehat{R} = \widehat{Q}\,\widehat{R}$$

be the *numerical* QR factorization of A, with $\widehat{R} \in R^{n\times n}$ upper triangular. It follows from section 4.1 that $\|\delta\widehat{A}\|_2 = O(\epsilon \cdot \|A\|_2)$ and that $\widehat{Q}$ is numerically column orthogonal, i.e.,

$$\widehat{Q}^T\,\widehat{Q} = \widehat{Q}_1^T\,\widehat{Q}_1 + \widehat{Q}_2^T\,\widehat{Q}_2 = I_n + \Delta_1, \tag{4.2}$$

where $\Delta_1 = \Delta_1^T = O(\epsilon) \in R^{n\times n}$.

We first assume that the matrix $\widehat{Q}_1^T\widehat{Q}_1 - \widehat{Q}_2^T\widehat{Q}_2$ has been computed and successfully Cholesky factorized, so that

$$\widehat{Q}_1^T\,\widehat{Q}_1 - \widehat{Q}_2^T\,\widehat{Q}_2 + \Delta_2 = \widehat{L}\,\widehat{L}^T. \tag{4.3}$$

It follows from section 4.1 that $\Delta_2 \in R^{n\times n}$ is symmetric and $\|\Delta_2\|_2 = O(\epsilon)$.

Let $\widehat{x}_s$ be the computed solution to (3.3). For simplicity we assume that $\widehat{R}$ is nonsingular and $\widehat{x}_s \neq 0$. According to section 4.1, $\widehat{x}_s$ satisfies

$$\left(\widehat{L} + \delta\widehat{L}_1\right)\,\left(\widehat{L} + \delta\widehat{L}_2\right)^T\,\left(\widehat{R} + \delta\widehat{R}\right)\,\widehat{x}_s = \left(\widehat{Q} + \delta\widehat{Q}\right)^T\,J\,b, \tag{4.4}$$

where

$$\|\delta\widehat{R}\|_2 = O(\epsilon \cdot \|\widehat{R}\|_2)\,, \quad \|\delta\widehat{Q}\|_2 = O(\epsilon), \quad \text{and} \quad \|\delta\widehat{L}_i\|_2 = O(\epsilon \cdot \|\widehat{L}\|_2) \quad \text{for} \quad i = 1, 2.$$

Since $\widehat{R}$ is a nonsingular upper triangular matrix, it follows from (4.1) that $\widehat{R} + \delta\widehat{R}$ is also nonsingular, and hence

$$\widehat{y} \stackrel{\text{def}}{=} \left(\widehat{R} + \delta\widehat{R}\right)\,\widehat{x}_s \neq 0.$$

We will write some of the round-off errors in (4.4) into $\widehat{Q}_1$ and b. To this end, define

$$\Delta_3 = v\,\widehat{y}^T + \widehat{y}\,v^T, \quad \text{where} \quad v = \frac{I - \dfrac{\widehat{y}\,\widehat{y}^T}{2\,\|\widehat{y}\|_2^2}}{\|\widehat{y}\|_2^2} \cdot \left(\delta\widehat{L}_1\,\widehat{L}^T + \widehat{L}\,\delta\widehat{L}_2^T + \delta\widehat{L}_1\,\delta\widehat{L}_2^T\right)\,\widehat{y}.$$

It is easy to verify that $\Delta_3 \in R^{n\times n}$ is symmetric and satisfies

$$\left(\widehat{L} + \delta\widehat{L}_1\right)\,\left(\widehat{L} + \delta\widehat{L}_2\right)^T\,\left(\widehat{R} + \delta\widehat{R}\right)\widehat{x}_s = \left(\widehat{L}\widehat{L}^T + \Delta_3\right)\,\left(\widehat{R} + \delta\widehat{R}\right)\widehat{x}_s.$$

Furthermore,

$$\begin{aligned}
\|\Delta_3\|_2 &\le 2\,\|v\|_2\,\|\widehat{y}\|_2 \\
&\le 2\,\frac{\left\|I - \dfrac{\widehat{y}\,\widehat{y}^T}{2\,\|\widehat{y}\|_2^2}\right\|_2}{\|\widehat{y}\|_2^2} \cdot \left\|\delta\widehat{L}_1\,\widehat{L}^T + \widehat{L}\,\delta\widehat{L}_2^T + \delta\widehat{L}_1\,\delta\widehat{L}_2^T\right\|_2\,\|\widehat{y}\|_2\,\|\widehat{y}\|_2 \\
&= O\left(\epsilon\cdot\|\widehat{L}\|_2^2\cdot\left\|I - \frac{\widehat{y}\,\widehat{y}^T}{2\,\|\widehat{y}\|_2^2}\right\|_2\right) = O\left(\epsilon\right),
\end{aligned}$$

where we have used the fact that

$$\|\widehat{L}\|_2^2 = \left\|\widehat{Q}_1^T\,\widehat{Q}_1 - \widehat{Q}_2^T\,\widehat{Q}_2 + \Delta_2\right\|_2 \le 1 + O(\epsilon) \quad \text{and} \quad \left\|I - \frac{\widehat{y}\,\widehat{y}^T}{2\,\|\widehat{y}\|_2^2}\right\|_2 \le 1.$$

Combining the above with equations (4.3) and (4.4), we have

$$\left(\widehat{Q}_1^T\widehat{Q}_1 - \widehat{Q}_2^T\widehat{Q}_2 + \Delta_2 + \Delta_3\right)\left(\widehat{R} + \delta\widehat{R}\right)\widehat{x}_s = \left(\widehat{Q} + \delta\widehat{Q}\right)^T\,J\,b.$$

Similar to section 3, relations (4.2) and (4.3) imply that the singular values of $\widehat{Q}_1$ are all between $1/\sqrt{2} + O(\epsilon)$ and $1 + O(\epsilon)$. Hence $\widehat{Q}_1$ is very well-conditioned.

In the following we shall rewrite $\Delta_2 + \Delta_3$ as a perturbation to $\widehat{Q}_1$. Let $\widehat{P} \in R^{p\times p}$ be the unique symmetric positive-definite matrix such that

$$\widehat{P}^2 = I_p + \widehat{Q}_1\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,(\Delta_2 + \Delta_3)\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,\widehat{Q}_1^T,$$

and let $(\widehat{P} + I)^{\frac{1}{2}}$ be the unique symmetric positive-definite square root of $\widehat{P} + I$. It follows that

$$\widehat{P} - I_p = \left(\widehat{P} + I\right)^{-\frac{1}{2}}\cdot\widehat{Q}_1\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,(\Delta_2 + \Delta_3)\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,\widehat{Q}_1^T\cdot\left(\widehat{P} + I\right)^{-\frac{1}{2}},$$

and that

$$\begin{aligned}
\|\widehat{P} - I_p\|_2 &\le \left\|\left(\widehat{P} + I\right)^{-\frac{1}{2}}\right\|_2\cdot\left\|\widehat{Q}_1\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,(\Delta_2 + \Delta_3)\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,\widehat{Q}_1^T\right\|_2 \\
&\quad\cdot\left\|\left(\widehat{P} + I\right)^{-\frac{1}{2}}\right\|_2 \\
&\le \left\|\widehat{Q}_1\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,(\Delta_2 + \Delta_3)\,\left(\widehat{Q}_1^T\,\widehat{Q}_1\right)^{-1}\,\widehat{Q}_1^T\right\|_2 = O(\epsilon)\,.
\end{aligned}$$

Now define

$$\widetilde{Q} = \begin{pmatrix} \widehat{P}\,\widehat{Q}_1 \\ \widehat{Q}_2 \end{pmatrix} = \widehat{Q} + O(\epsilon).$$

Since both $\widehat{P}$ and $\widehat{Q}_1$ are very well-conditioned, it follows that $\widetilde{Q}$ is itself very well-conditioned. Hence

$$\begin{aligned}\widetilde{Q}^T\;J\;\widetilde{Q} &= \left(\widehat{Q}_1^T\;\widehat{P}\right)\left(\widehat{Q}_1^T\;\widehat{P}\right)^T - \widehat{Q}_2^T\;\widehat{Q}_2 \\ &= \widehat{Q}_1^T\;\left(I_p + \widehat{Q}_1\;\left(\widehat{Q}_1^T\;\widehat{Q}_1\right)^{-1}\;(\Delta_2+\Delta_3)\;\left(\widehat{Q}_1^T\;\widehat{Q}_1\right)^{-1}\widehat{Q}_1^T\right)\;\widehat{Q}_1 - \widehat{Q}_2^T\;\widehat{Q}_2 \\ &= \widehat{Q}_1^T\;\widehat{Q}_1 - \widehat{Q}_2^T\;\widehat{Q}_2 + \Delta_2 + \Delta_3.\end{aligned}$$

Hence we can now rewrite equation (4.4) simply as

$$\left(\widetilde{Q}^T\;J\;\widetilde{Q}\right)\;\left(\widehat{R}+\delta\widehat{R}\right)\;\widehat{x}_s = \left(\widehat{Q}+\delta\widehat{Q}\right)^T\;J\;b. \tag{4.5}$$

In the following, we write the round-off errors on the right-hand side as an error in b.

$$\begin{aligned}\left(\widehat{Q}+\delta\widehat{Q}\right)^T\;J\;b &= \left(\widetilde{Q}^T + \widetilde{Q}^T\;\widetilde{Q}\;\left(\widetilde{Q}^T\;\widetilde{Q}\right)^{-1}\;\left(\delta\widehat{Q}+\widehat{Q}-\widetilde{Q}\right)^T\right)\;J\;b \\ &= \widetilde{Q}^T\;J\;\left(b+\delta b\right),\end{aligned}$$

where

$$\delta b = J\;\widetilde{Q}\;\left(\widetilde{Q}^T\;\widetilde{Q}\right)^{-1}\;\left(\delta\widehat{Q}+\widehat{Q}-\widetilde{Q}\right)^T\;J\;b,$$

and hence $\|\delta b\|_2 = O(\epsilon\cdot\|b\|_2)$. Equation (4.5) can now be rewritten as

$$\left(\widetilde{Q}^T\;J\;\widetilde{Q}\right)\;\left(\widehat{R}+\delta\widehat{R}\right)\;\widehat{x}_s = \widetilde{Q}^T\;J\;\left(b+\delta b\right). \tag{4.6}$$

Finally we define the perturbation in A as

$$\delta A = \delta\widehat{A} + \left(\widetilde{Q}-\widehat{Q}\right)\;\widehat{R} + \widetilde{Q}\;\delta\widehat{R}.$$

Then it follows that $\|\delta A\|_2 = O(\epsilon\cdot\|A\|_2)$. It can be easily checked that

$$A + \delta A = \widetilde{Q}\left(\widehat{R}+\delta\widehat{R}\right). \tag{4.7}$$

With these backward errors, we note that (4.6) is exactly the equation (3.3) for the perturbed indefinite least-squares problem

$$\min_x\;\left(\left(A+\delta A\right)x - \left(b+\delta b\right)\right)^T\;J\;\left(\left(A+\delta A\right)x - \left(b+\delta b\right)\right).$$

Hence the new algorithm in section 3 is *backward stable.* Note that the matrix $\widetilde{Q}$ is in general not orthogonal, and hence the factorization (4.7) is in general *not* a QR factorization.

Now we consider the case where one fails to numerically compute the Cholesky factorization (4.3). This can happen only if the matrix

$$\widehat{Q}_1^T \widehat{Q}_1 - \widehat{Q}_2^T \widehat{Q}_2 + \Delta_2$$

is not symmetric positive-definite for a symmetric $\Delta_2 \in R^{n\times n}$ with a small 2-norm. With the techniques developed above, it is straightforward to show that this implies that there exists a $\delta A \in R^{m\times n}$ such that the matrix $(A + \delta A)^T J(A + \delta A)$ is not symmetric positive-definite. In other words, the indefinite least-squares problem (1.1) does not have a unique solution for a slightly perturbed A. Such a problem cannot be expected to have a numerically meaningful solution in general.

5. Conclusion. In this paper we proposed a stable and efficient algorithm for solving the indefinite least-squares problem. Our error analysis shows that this algorithm is backward stable.

Acknowledgments. The authors thank the anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] M. Berry and G. Cybenko, *Hyperbolic Householder algorithms for factoring structured matrices*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 499–520.

[2] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Parameter estimation in the presence of bounded data uncertainties*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 235–252.

[3] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *An efficient algorithm for a bounded errors-in-variables model*, SIAM J. Matrix Anal. Appl., to appear.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[5] B. Hassibi, A. H. Sayed, and T. Kailath, *Linear estimation in Krein spaces - Part* I*: Theory*, IEEE Trans. Automat. Control, 41 (1996), pp. 18–33.

[6] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, SIAM, Philadelphia, PA, 1996.

[7] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, PA, 1991.

[8] P. Khargonekar and K. M. Nagpal, *Filtering and smoothing in an H^∞-setting*, IEEE Trans. Automat. Control, AC-36 (1991), pp. 151–166.

[9] C. M. Rader and A. O. Steinhardt, *Hyperbolic Householder transforms*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 269–290.

[10] A. H. Sayed, B. Hassibi, and T. Kailath, *Fundamental Inertia Conditions for the Minimization of Quadratic Forms in Indefinite Metric Spaces*, Oper. Theory: Adv. Appl., Birkhauser, Cambridge, MA, 1996.

[11] U. Shaked and Y. Theodor, *H^∞-optimal estimation: A tutorial*, in Proc. IEEE Conference on Decision and Control, Tucson, AZ, 1992, pp. 2278–2286.

[12] G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York, 1973.