Chapter 6

# Modified Least Squares Problems and Methods



In this chapter we discuss an assortment of least square problems that can be solved using QR and SVD. We also introduce a generalization of the SVD that can be used to simultaneously diagonalize a pair of matrices, a maneuver that is useful in certain applications.

The first three sections deal with variations of the ordinary least squares problem that we treated in Chapter 5. The unconstrained minimization of $\| Ax - b \|_2$ does not always make a great deal of sense. How do we balance the importance of each equation in $Ax = b$? How might we control the size of $x$ if $A$ is ill-conditioned? How might we minimize $\| Ax - b \|_2$ over a proper subspace of $\mathbb{R}^n$? What if there are errors in the "data matrix" $A$ in addition to the usual errors in the "vector of observations" $b$?

In §6.4 we consider a number of multidimensional subspace computations including the problem of determining the principal angles between a pair of given subspaces. The SVD plays a prominent role.

The final section is concerned with the updating of matrix factorizations. In many applications, one is confronted with a succession of least squares (or linear equation) problems where the matrix associated with the current step is highly related to the matrix associated with the previous step. This opens the door to updating strategies that can reduce factorization overheads by an order of magnitude.

*Reading Notes*

Knowledge of Chapter 5 is assumed. The sections in this chapter are independent of each other except that §6.1 should be read before §6.2. Excellent global references include Björck (NMLS) and Lawson and Hansen (SLS).

## 6.1 Weighting and Regularization

We consider two basic modifications to the linear least squares problem. The first concerns how much each equation "counts" in the $\| Ax - b \|_2$ minimization. Some equations may be more important than others and there are ways to produce approximate minimzers that reflect this. Another situation arises when $A$ is ill-conditioned. Instead of minimizing $\| Ax - b \|_2$ with a possibly wild, large norm $x$-vector, we settle for a predictor $Ax$ in which $x$ is "nice" according to some regularizing metric.

### 6.1.1 Row Weighting

In ordinary least squares, the minimization of $\| Ax - b \|_2$ amounts to minimizing the sum of the squared discrepancies in each equation:

$$\| Ax - b \|^2 = \sum_{i=1}^{m} \left( a_i^T x - b_i \right)^2 .$$

We assume that $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $a_i^T = A(i,:)$. In the *weighted least squares problem* the discrepancies are scaled and we solve

$$\min_{x\in\mathbb{R}^n} \| D(Ax - b) \|^2 = \min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} d_i^2 \left( a_i^T x - b_i \right)^2 \tag{6.1.1}$$

where $D = \mathrm{diag}(d_1, \ldots, d_m)$ is nonsingular. Note that if $x_D$ minimizes this summation, then it minimizes $\| \widetilde{A}x - \tilde{b} \|_2$ where $\widetilde{A} = DA$ and $\tilde{b} = Db$. Although there can be numerical issues associated with disparate weight values, it is generally possible to solve the weighted least squares problem by applying any Chapter 5 method to the "tilde problem." For example, if $A$ has full column rank and we apply the method of normal equations, then we are led to the following positive definite system:

$$(A^T D^2 A)x_D = A^T D^2 b. \tag{6.1.2}$$

Subtracting the unweighted system $A^T A x_{LS} = A^T b$ we see that

$$x_D - x_{LS} = (A^T D^2 A)^{-1} A^T (D^2 - I)(b - Ax_{LS}). \tag{6.1.3}$$

Note that weighting has less effect if $b$ is almost in the range of $A$.

At the component level, increasing $d_k$ relative to the other weights stresses the importance of the $k$th equation and the resulting residual $r = b - Ax_D$ tends to be smaller in that component. To make this precise, define

$$D(\delta) = \mathrm{diag}(d_1, \ldots, d_{k-1}, d_k \sqrt{1+\delta}\,, d_{k+1}, \ldots, d_m)$$

where $\delta > -1$. Assume that $x(\delta)$ minimizes $\| D(\delta)(Ax - b) \|_2$ and set

$$r_k(\delta) = e_k^T (b - Ax(\delta)) = b_k - a_k^T (A^T D(\delta)^2 A)^{-1} A^T D(\delta)^2 b$$

where $e_k = I_m(:,k)$. We show that the penalty for disagreement between $a_k^T x$ and $b_k$ increases with $\delta$. Since

$$\frac{d}{d\delta}\left[ D(\delta)^2 \right] = d_k^2 e_k e_k^T$$

and

$$\frac{d}{d\delta}\left[(A^TD(\delta)^2A)^{-1}\right] = -(A^TD(\delta)^2A)^{-1}(A^T(d_k^2e_ke_k^T)A)(A^TD(\delta)^2A)^{-1},$$

it can be shown that

$$\frac{d}{d\delta}r_k(\delta) = -d_k^2\left(a_k^T(A^TD(\delta)^2A)^{-1}a_k\right)r_k(\delta). \tag{6.1.4}$$

Assuming that $A$ has full rank, the matrix $(A^TD(\delta)A)^{-1}$ is positive definite and so

$$\frac{d}{d\delta}[r_k(\delta)^2] = 2\,r_k(\delta)\cdot\frac{d}{d\delta}r_k(\delta) = -2d_k^2\left(a_k^T(A^TD(\delta)^2A)^{-1}a_k\right)r_k(\delta)^2 < 0.$$

It follows that $|r_k(\delta)|$ is a monotone decreasing function of $\delta$. Of course, the change in $r_k$ when all the weights are varied at the same time is much more complicated.

Before we move on to a more general type of row weighting, we mention that (6.1.1) can be framed as a symmetric indefinite linear system. In particular, if

$$\begin{bmatrix} D^{-2} & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \tag{6.1.5}$$

then $x$ minimizes (6.1.1). Compare with (5.3.20).

### 6.1.2 Generalized Least Squares

In statistical data-fitting applications, the weights in (6.1.1) are often chosen to increase the relative importance of accurate measurements. For example, suppose the vector of observations $b$ has the form $b_{\text{true}} + \Delta$ where $\Delta_i$ is normally distributed with mean zero and standard deviation $\sigma_i$. If the errors are uncorrelated, then it makes statistical sense to minimize (6.1.1) with $d_i = 1/\sigma_i$.

In more general estimation problems, the vector $b$ is related to $x$ through the equation

$$b = Ax + w \tag{6.1.6}$$

where the *noise vector* $w$ has zero mean and a symmetric positive definite *covariance matrix* $\sigma^2W$. Assume that $W$ is known and that $W = BB^T$ for some $B \in \mathbb{R}^{m\times m}$. The matrix $B$ might be given or it might be $W$'s Cholesky triangle. In order that all the equations in (6.1.6) contribute equally to the determination of $x$, statisticians frequently solve the LS problem

$$\min_{x\in\mathbb{R}^n} \| B^{-1}(Ax-b) \|_2 . \tag{6.1.7}$$

An obvious computational approach to this problem is to form $\widetilde{A} = B^{-1}A$ and $\tilde{b} = B^{-1}b$ and then apply any of our previous techniques to minimize $\| \widetilde{A}x - \tilde{b} \|_2$. Unfortunately, if $B$ is ill-conditioned, then $x$ will be poorly determined by such a procedure.

A more stable way of solving (6.1.7) using orthogonal transformations has been suggested by Paige (1979a, 1979b). It is based on the idea that (6.1.7) is equivalent to the *generalized least squares problem*,

$$\min_{b=Ax+Bv} v^Tv . \tag{6.1.8}$$

Notice that this problem is defined even if $A$ and $B$ are rank deficient. Although in the Paige technique can be applied when this is the case, we shall describe it under the assumption that both these matrices have full rank.

The first step is to compute the QR factorization of $A$:

$$Q^T A = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}, \qquad Q = \underset{n}{[\, Q_1 \,|} \; \underset{m-n}{Q_2 \,]} .$$

Next, an orthogonal matrix $Z \in \mathbb{R}^{m\times m}$ is determined such that

$$(Q_2^T B)Z = \underset{n}{[\, 0 \,|} \; \underset{m-n}{S \,]}, \qquad Z = \underset{n}{[\, Z_1 \,|} \; \underset{m-n}{Z_2 \,]}$$

where $S$ is upper triangular. With the use of these orthogonal matrices, the constraint in (6.1.8) transforms to

$$\begin{bmatrix} Q_1^T b \\ Q_2^T b \end{bmatrix} = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x + \begin{bmatrix} Q_1^T B Z_1 & Q_1^T B Z_2 \\ 0 & S \end{bmatrix} \begin{bmatrix} Z_1^T v \\ Z_2^T v \end{bmatrix} .$$

The bottom half of this equation determines $v$ while the top half prescribes $x$:

$$Su = Q_2^T b, \qquad v = Z_2 u, \tag{6.1.9}$$

$$R_1 x = Q_1^T b - (Q_1^T B Z_1 Z_1^T + Q_1^T B Z_2 Z_2^T)v = Q_1^T b - Q_1^T B Z_2 u. \tag{6.1.10}$$

The attractiveness of this method is that all potential ill-conditioning is concentrated in the triangular systems (6.1.9) and (6.1.10). Moreover, Paige (1979b) shows that the above procedure is numerically stable, something that is not true of any method that explicitly forms $B^{-1}A$.

### 6.1.3 A Note on Column Weighting

Suppose $G \in \mathbb{R}^{n\times n}$ is nonsingular and define the $G$-norm $\|\cdot\|_G$ on $\mathbb{R}^n$ by

$$\| z \|_G = \| Gz \|_2 .$$

If $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and we compute the minimum 2-norm solution $y_{LS}$ to

$$\min_{x\in\mathbb{R}^n} \| (AG^{-1})y - b \|_2 ,$$

then $x_G = G^{-1}y_{LS}$ is a minimizer of $\| Ax - b \|_2$. If $\mathsf{rank}(A) < n$, then within the set of minimizers, $x_G$ has the smallest $G$-norm.

The choice of $G$ is important. Sometimes its selection can be based upon a priori knowledge of the uncertainties in $A$. On other occasions, it may be desirable to normalize the columns of $A$ by setting

$$G = G_0 \equiv \mathrm{diag}(\| A(:,1) \|_2, \ldots, \| A(:,n) \|_2).$$

Van der Sluis (1969) has shown that with this choice, $\kappa_2(AG^{-1})$ is approximately minimized. Since the computed accuracy of $y_{LS}$ depends on $\kappa_2(AG^{-1})$, a case can be made for setting $G = G_0$.

We remark that column weighting affects singular values. Consequently, a scheme for determining numerical rank may not return the same estimate when applied to $A$ and $AG^{-1}$. See Stewart (1984).

### 6.1.4 Ridge Regression

In the *ridge regression problem* we are given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ and proceed to solve

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\lambda} I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2 = \min_x \| Ax - b \|_2^2 + \lambda \| x \|_2^2 . \tag{6.1.11}$$

where the value of the *ridge parameter* $\lambda$ is chosen to "shape" the solution $x = x(\lambda)$ in some meaningful way. Notice that the normal equation system for this problem is given by

$$(A^TA + \lambda I)x = A^Tb. \tag{6.1.12}$$

It follows that if

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{6.1.13}$$

is the SVD of $A$, then (6.1.12) converts to

$$(\Sigma^T\Sigma + \lambda I_n)(V^Tx) = \Sigma^T U^T b$$

and so

$$x(\lambda) = \sum_{i=1}^{r} \frac{\sigma_i\, u_i^T b}{\sigma_i^2 + \lambda} v_i. \tag{6.1.14}$$

By inspection, it is clear that

$$\lim_{\lambda\to 0} x(\lambda) = x_{LS}$$

and $\| x(\lambda) \|_2$ is a monotone decreasing function of $\lambda$. These two facts show how an ill-conditioned least squares solution can be *regularized* by judiciously choosing $\lambda$. The idea is to get sufficiently close to $x_{LS}$ subject to the constraint that the norm of the ridge regression minimzer $x(\lambda)$ is sufficiently modest. Regularization in this context is all about the intelligent balancing of these two tensions.

The ridge parameter can also be chosen with an eye toward balancing the "impact" of each equation in the overdetermined system $Ax = b$. We describe a $\lambda$-selection procedure due to Golub, Heath, and Wahba (1979). Set

$$D_k = I - e_k e_k^T = \text{diag}(1,\ldots,1,0,1,\ldots,1) \in \mathbb{R}^{m\times m}$$

and let $x_k(\lambda)$ solve

$$\min_{x\in\mathbb{R}^n} \| D_k(Ax - b) \|_2^2 + \lambda \| x \|_2^2 . \tag{6.1.15}$$

Thus, $x_k(\lambda)$ is the solution to the ridge regression problem with the $k$th row of $A$ and $k$th component of $b$ deleted, i.e., the $k$th equation in the overdetermined system $Ax = b$ is deleted. Now consider choosing $\lambda$ so as to minimize the *cross-validation weighted square error* $C(\lambda)$ defined by

$$C(\lambda) = \frac{1}{m}\sum_{k=1}^{m} w_k(a_k^T x_k(\lambda) - b_k)^2 .$$

Here, $w_1, \ldots, w_m$ are nonnegative weights and $a_k^T$ is the $k$th row of $A$. Noting that

$$\| Ax_k(\lambda) - b \|_2^2 = \| D_k(Ax_k(\lambda) - b) \|_2^2 + (a_k^T x_k(\lambda) - b_k)^2,$$

we see that $\left(a_k^T x_k(\lambda) - b_k\right)^2$ is the increase in the sum of squares that results when the $k$th row is "reinstated." Minimizing $C(\lambda)$ is tantamount to choosing $\lambda$ such that the final model is not overly dependent on any one experiment.

A more rigorous analysis can make this statement precise and also suggest a method for minimizing $C(\lambda)$. Assuming that $\lambda > 0$, an algebraic manipulation shows that

$$x_k(\lambda) = x(\lambda) + \frac{a_k^T x(\lambda) - b_k}{1 - z_k^T a_k} z_k \tag{6.1.16}$$

where $z_k = (A^T A + \lambda I)^{-1} a_k$ and $x(\lambda) = (A^T A + \lambda I)^{-1} A^T b$. Applying $-a_k^T$ to (6.1.16) and then adding $b_k$ to each side of the resulting equation gives

$$r_k = b_k - a_k^T x_k(\lambda) = \frac{e_k^T (I - A(A^T A + \lambda I)^{-1} A^T) b}{e_k^T (I - A(A^T A + \lambda I)^{-1} A^T) e_k} . \tag{6.1.17}$$

Noting that the residual $r = [\, r_1, \ldots, r_m \,]^T = b - Ax(\lambda)$ is given by the formula

$$r = [I - A(A^T A + \lambda I)^{-1} A^T] b,$$

we see that

$$C(\lambda) = \frac{1}{m}\sum_{k=1}^{m} w_k \left( \frac{r_k}{\partial r_k / \partial b_k} \right)^2 . \tag{6.1.18}$$

The quotient $r_k/(\partial r_k/\partial b_k)$ may be regarded as an inverse measure of the "impact" of the $k$th observation $b_k$ on the model. If $\partial r_k/\partial b_k$ is small, then this says that the error in the model's prediction of $b_k$ is somewhat independent of $b_k$. The tendency for this to be true is lessened by basing the model on the $\lambda^*$ that minimizes $C(\lambda)$.

The actual determination of $\lambda^*$ is simplified by computing the SVD of $A$. Using the SVD (6.1.13) and Equations (6.1.17) and (6.1.18), it can be shown that

$$C(\lambda) = \frac{1}{m}\sum_{k=1}^{m} w_k \left[ \frac{\tilde{b}_k - \sum_{j=1}^{r} u_{kj}\tilde{b}_j \left( \frac{\sigma_j^2}{\sigma_j^2 + \lambda} \right)}{1 - \sum_{j=1}^{r} u_{kj}^2 \left( \frac{\sigma_j^2}{\sigma_j^2 + \lambda} \right)} \right]^2 \tag{6.1.19}$$

where $\tilde{b} = U^T b$. The minimization of this expression is discussed in Golub, Heath, and Wahba (1979).

### 6.1.5 Tikhonov Regularization

In the *Tikhonov regularization problem*, we are given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times n}$, and $b \in \mathbb{R}^m$ and solve

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\lambda}B \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2 = \min_x \| Ax - b \|_2^2 + \lambda \| Bx \|_2^2. \tag{6.1.20}$$

The normal equations for this problem have the form

$$(A^TA + \lambda B^TB)x = A^Tb. \tag{6.1.21}$$

This system is nonsingular if $\mathsf{null}(A) \cap \mathsf{null}(B) = \{0\}$. The matrix $B$ can be chosen in several ways. For example, in certain data-fitting applications second derivative smoothness can be promoted by setting $B = \mathcal{T}_{DD}$, the second difference matrix defined in Equation 4.8.7.

To analyze how $A$ and $B$ interact in the Tikhonov problem, it would be handy to transform (6.1.21) into an equivalent diagonal problem. For the ridge regression problem ($B = I_n$) the SVD accomplishes this task. For the Tikhonov problem, we need a generalization of the SVD that simultaneously diagonalizes both $A$ and $B$.

### 6.1.6 The Generalized Singular Value Decomposition

The generalized singular value decomposition (GSVD) set forth in Van Loan (1974) provides a useful way to simplify certain two-matrix problems such as the Tychanov regularization problem.

**Theorem 6.1.1 (Generalized Singular Value Decomposition).** *Assume that $A \in \mathbb{R}^{m_1\times n_1}$ and $B \in \mathbb{R}^{m_2\times n_1}$ with $m_1 \geq n_1$ and*

$$r = \mathsf{rank}\left( \begin{bmatrix} A \\ B \end{bmatrix} \right).$$

*There exist orthogonal $U_1 \in \mathbb{R}^{m_1\times m_1}$ and $U_2 \in \mathbb{R}^{m_2\times m_2}$ and invertible $X \in \mathbb{R}^{n_1\times n_1}$ such that*

$$U_1^TAX = D_A = \begin{bmatrix} I & 0 & 0 \\ 0 & \mathrm{diag}(\alpha_{p+1},\ldots,\alpha_r) & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} p \\ r-p \\ m_1-r \end{matrix}, \tag{6.1.22}$$

$$\begin{matrix} & p & r-p & n_1-r \end{matrix}$$

$$U_2^TBX = D_B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mathrm{diag}(\beta_{p+1},\ldots,\beta_r) & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{matrix} p \\ r-p \\ m_2-r \end{matrix}, \tag{6.1.23}$$

$$\begin{matrix} & p & r-p & n_1-r \end{matrix}$$

*where $p = \max\{r - m_2, 0\}$.*

***Proof.*** The proof makes use of the SVD and the CS decomposition (Theorem 2.5.3). Let

$$\begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} Z^T \tag{6.1.24}$$

be the SVD where $\Sigma_r \in \mathbb{R}^{r\times r}$ is nonsingular, $Q_{11} \in \mathbb{R}^{m_1\times r}$, and $Q_{21} \in \mathbb{R}^{m_2\times r}$. Using the CS decomposition, there exist orthogonal matrices $U_1$ ($m_1$-by-$m_1$), $U_2$ ($m_2$-by-$m_2$), and $V_1$ ($r$-by-$r$) such that

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix} V_1 = \begin{bmatrix} D_A(:,1{:}r) \\ D_B(:,1{:}r) \end{bmatrix} \tag{6.1.25}$$

where $D_A$ and $D_B$ have the forms specified by (6.1.21) and (6.1.22). It follows from (6.1.24) and (6.1.25) that

$$\begin{aligned} \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T \begin{bmatrix} A \\ B \end{bmatrix} Z &= \begin{bmatrix} D_A(:,1{:}r) & U_1Q_{12} \\ D_B(:,1{:}r) & U_2Q_{22} \end{bmatrix} \begin{bmatrix} V_1^T\Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} D_A(:,1{:}r) & 0 \\ D_B(:,1{:}r) & 0 \end{bmatrix} \begin{bmatrix} V_1^T\Sigma_r & 0 \\ 0 & I_{n_1-r} \end{bmatrix} \\ &= \begin{bmatrix} D_A \\ D_B \end{bmatrix} \begin{bmatrix} V_1^T\Sigma_r & 0 \\ 0 & I_{n_1-r} \end{bmatrix}. \end{aligned}$$

By setting

$$X = Z \begin{bmatrix} V_1^T\Sigma_r & 0 \\ 0 & I_{n_1-r} \end{bmatrix}^{-1}$$

the proof is complete. □

Note that if $B = I_{n_1}$ and we set $X = U_2$, then we obtain the SVD of $A$. The GSVD is related to the generalized eigenvalue problem

$$A^TAx = \mu^2B^TBx$$

which is considered in §8.7.4. As with the SVD, algorithmic issues cannot be addressed until we develop procedures for the symmetric eigenvalue problem in Chapter 8.

To illustrate the insight that can be provided by the GSVD, we return to the Tikhonov regularization problem (6.1.20). If $B$ is square and nonsingular, then the GSVD defined by (6.1.22) and (6.1.23) transforms the system (6.1.21) to

$$(D_A^TD_A + \lambda D_B^TD_B)y = D_A^T\tilde{b}$$

where $x = Xy$, $\tilde{b} = U_1^Tb$, and

$$(D_A^TD_A + \lambda D_B^TD_B) = \operatorname{diag}(\alpha_1^2 + \lambda\beta_1^2, \ldots, \alpha_n^2 + \lambda\beta_n^2).$$

Thus, if

$$X = [\, x_1 \,|\, \cdots \,|\, x_n \,]$$

is a column partitioning, then

$$x(\lambda) = \sum_{k=1}^{n} \left( \frac{\alpha_k \tilde{b}_k}{\alpha_k^2 + \lambda \beta_k^2} \right) x_k \tag{6.1.26}$$

solves (6.1.20). The "calming influence" of the regularization is revealed through this representation. Use of $\lambda$ to manage "trouble" in the direction of $x_k$ depends on the values of $\alpha_k$ and $\beta_k$.

### Problems

**P6.1.1** Verify (6.1.4).

**P6.1.2** What is the inverse of the matrix in (6.1.5)?

**P6.1.3** Show how the SVD can be used to solve the generalized LS problem (6.1.8) if the matrices $A$ and $B$ are rank deficient.

**P6.1.4** Suppose $A$ is the $m$-by-1 matrix of 1's and let $b \in \mathbb{R}^m$. Show that the cross-validation technique with unit weights prescribes an optimal $\lambda$ given by

$$\lambda = \left( \left( \frac{\tilde{b}}{s} \right)^2 - \frac{1}{m} \right)^{-1}$$

where $\tilde{b} = (b_1 + \cdots + b_m)/m$ and

$$s = \sum_{i=1}^{m} (b_i - \tilde{b})^2/(m-1).$$

**P6.1.5** Using the GSVD, give bounds for $\| x(\lambda) - x(0) \|$ and $\| Ax(\lambda) - b \|_2^2 - \| Ax(0) - b \|_2^2$ where $x(\lambda)$ is defined by (6.1.26).

### Notes and References for §6.1

Row and column weighting in the LS problem is discussed in Lawson and Hanson (SLS, pp. 180-88). Other analyses include:

A. van der Sluis (1969). "Condition Numbers and Equilibration of Matrices," *Numer. Math. 14,* 14–23.

G.W. Stewart (1984). "On the Asymptotic Behavior of Scaled Singular Value and QR Decompositions," *Math. Comput. 43,* 483–490.

A. Forsgren (1996). "On Linear Least-Squares Problems with Diagonally Dominant Weight Matrices," *SIAM J. Matrix Anal. Applic. 17*, 763–788.

P.D. Hough and S.A. Vavasis (1997). "Complete Orthogonal Decomposition for Weighted Least Squares," *SIAM J. Matrix Anal. Applic. 18*, 551–555.

J.K. Reid (2000). "Implicit Scaling of Linear Least Squares Problems," *BIT 40*, 146–157.

For a discussion of cross-validation issues, see:

G.H. Golub, M. Heath, and G. Wahba (1979). "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics 21,* 215–23.

L. Eldén (1985). "A Note on the Computation of the Generalized Cross-Validation Function for Ill-Conditioned Least Squares Problems," *BIT 24,* 467–472.

Early references concerned with the generalized singular value decomposition include:

C.F. Van Loan (1976). "Generalizing the Singular Value Decomposition," *SIAM J. Numer. Anal. 13,* 76–83.

C.C. Paige and M.A. Saunders (1981). "Towards A Generalized Singular Value Decomposition," *SIAM J. Numer. Anal. 18,* 398–405.

The theoretical and computational aspects of the generalized least squares problem appear in:

C.C. Paige (1979). "Fast Numerically Stable Computations for Generalized Linear Least Squares Problems," *SIAM J. Numer. Anal. 16,* 165–171.
C.C. Paige (1979b). "Computer Solution and Perturbation Analysis of Generalized Least Squares Problems," *Math. Comput. 33,* 171–84.
S. Kourouklis and C.C. Paige (1981). "A Constrained Least Squares Approach to the General Gauss-Markov Linear Model," *J. Amer. Stat. Assoc. 76,* 620–625.
C.C. Paige (1985). "The General Limit Model and the Generalized Singular Value Decomposition," *Lin. Alg. Applic. 70,* 269–284.

Generalized factorizations have an important bearing on generalized least squares problems, see:

C.C. Paige (1990). "Some Aspects of Generalized QR Factorization," in *Reliable Numerical Computations*, M. Cox and S. Hammarling (eds.), Clarendon Press, Oxford.
E. Anderson, Z. Bai, and J. Dongarra (1992). "Generalized QR Factorization and Its Applications," *Lin. Alg. Applic. 162/163/164*, 243–271.

The development of regularization techniques has a long history, see:

L. Eldén (1977). "Algorithms for the Regularization of Ill-Conditioned Least Squares Problems," *BIT 17,* 134–45.
D.P. O'Leary and J.A. Simmons (1981). "A Bidiagonalization-Regularization Procedure for Large Scale Discretizations of Ill-Posed Problems," *SIAM J. Sci. Stat. Comput. 2*, 474–489.
L. Eldén (1984). "An Algorithm for the Regularization of Ill-Conditioned, Banded Least Squares Problems," *SIAM J. Sci. Stat. Comput. 5*, 237–254.
P.C. Hansen (1990). "Relations Between SVD and GSVD of Discrete Regularization Problems in Standard and General Form," *Lin.Alg. Applic. 141*, 165–176.
P.C. Hansen (1995). "Test Matrices for Regularization Methods," *SIAM J. Sci. Comput. 16*, 506–512.
A. Neumaier (1998). "Solving Ill–Conditioned and Singular Linear Systems: A Tutorial on Regularization," *SIAM Review 40*, 636–666.
P.C. Hansen (1998). *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM Publications, Philadelphia, PA.
M.E. Gulliksson and P.-A. Wedin (2000). "The Use and Properties of Tikhonov Filter Matrices," *SIAM J. Matrix Anal. Applic. 22*, 276–281.
M.E. Gulliksson, P.-A. Wedin, and Y. Wei (2000). "Perturbation Identities for Regularized Tikhonov Inverses and Weighted Pseudoinverses," *BIT 40*, 513–523.
T. Kitagawa, S. Nakata, and Y. Hosoda (2001). "Regularization Using QR Factorization and the Estimation of the Optimal Parameter," *BIT 41*, 1049–1058.
M.E. Kilmer and D.P. O'Leary. (2001). "Choosing Regularization Parameters in Iterative Methods for Ill-Posed Problems," *SIAM J. Matrix Anal. Applic. 22*, 1204–1221.
A. N. Malyshev (2003). "A Unified Theory of Conditioning for Linear Least Squares and Tikhonov Regularization Solutions," *SIAM J. Matrix Anal. Applic. 24*, 1186–1196.
M. Hanke (2006). "A Note on Tikhonov Regularization of Large Linear Problems," *BIT 43*, 449–451.
P.C. Hansen, J.G. Nagy, and D.P. OLeary (2006). *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM Publications, Philadelphia, PA.
M.E. Kilmer, P.C. Hansen, and M.I. Español (2007). "A Projection-Based Approach to General-Form Tikhonov Regularization," *SIAM J. Sci. Comput. 29*, 315–330.
T. Elfving and I. Skoglund (2009). "A Direct Method for a Regularized Least-Squares Problem," *Num. Lin. Alg. Applic. 16*, 649–675.
I. Hnětynková and M. Plešinger (2009). "The Regularizing Effect of the Golub-Kahan Iterative Bidiagonalization and revealing the Noise level in Data," *BIT 49*, 669–696.
P.C. Hansen (2010). *Discrete Inverse Problems: Insight and Algorithms*, SIAM Publications, Philadelphia, PA.

## 6.2 Constrained Least Squares

In the least squares setting it is sometimes natural to minimize $\| Ax - b \|_2$ over a proper subset of $\mathbb{R}^n$. For example, we may wish to predict $b$ as best we can with $Ax$ subject to the constraint that $x$ is a unit vector. Or perhaps the solution defines a fitting function $f(t)$ which is to have prescribed values at certain points. This can lead to an equality-constrained least squares problem. In this section we show how these problems can be solved using the QR factorization, the SVD, and the GSVD.

### 6.2.1 Least Squares Minimization Over a Sphere

Given $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and a positive $\alpha \in \mathbb{R}$, we consider the problem

$$\min_{\|x\|_2 \le \alpha} \| Ax - b \|_2 . \tag{6.2.1}$$

This is an example of the *LSQI* (least squares with quadratic inequality constraint) problem. This problem arises in nonlinear optimization and other application areas. As we are soon to observe, the LSQI problem is related to the ridge regression problem discussed in §6.1.4.

Suppose

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{6.2.2}$$

is the SVD of $A$ which we assume to have rank $r$. If the unconstrained minimum norm solution

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i$$

satisfies $\| x_{LS} \|_2 \le \alpha$, then it obviously solves (6.2.1). Otherwise,

$$\| x_{LS} \|_2^2 = \sum_{i=1}^{r} \left( \frac{u_i^T b}{\sigma_i} \right)^2 > \alpha^2, \tag{6.2.3}$$

and it follows that the solution to (6.2.1) is on the boundary of the constraint sphere. Thus, we can approach this constrained optimization problem using the method of Lagrange multipliers. Define the parameterized objective function $\phi$ by

$$\phi(x,\lambda) = \frac{1}{2}\| Ax - b \|_2^2 + \frac{\lambda}{2}\left( \| x \|_2^2 - \alpha^2 \right)$$

and equate its gradient to zero. This gives a shifted normal equation system:

$$(A^T A + \lambda I)\cdot x(\lambda) = A^T b.$$

The goal is to choose $\lambda$ so that $\| x(\lambda) \|_2 = \alpha$. Using the SVD (6.2.2), this leads to the problem of finding a zero of the function

$$f(\lambda) = \| x(\lambda) \|_2^2 - \alpha^2 = \sum_{k=1}^{n} \left( \frac{\sigma_k u_k^T b}{\sigma_k^2 + \lambda} \right)^2 - \alpha^2.$$

This is an example of a *secular equation* problem. From (6.2.3), $f(0) > 0$. Since $f'(\lambda) < 0$ for $\lambda \geq 0$, it follows that $f$ has a unique positive root $\lambda_+$. It can be shown that

$$\rho(\lambda) = \| Ax(\lambda) - b \|_2^2 = \| Ax_{LS} - b \|_2^2 + \sum_{i=1}^{r} \left( \frac{\lambda u_i^T b}{\sigma_i^2 + \lambda} \right)^2 . \tag{6.2.4}$$

It follows that $x(\lambda_+)$ solves (6.2.1).

**Algorithm 6.2.1** Given $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, $b \in \mathbb{R}^m$, and $\alpha > 0$, the following algorithm computes a vector $x \in \mathbb{R}^n$ such that $\| Ax - b \|_2$ is minimum subject to the constraint that $\| x \|_2 \leq \alpha$.

Compute the SVD $A = U\Sigma V^T$, save $V = [\, v_1 \,|\, \cdots \,|\, v_n \,]$, form $\tilde{b} = U^T b$,
and determine $r = \mathsf{rank}(A)$.

**if** $\displaystyle\sum_{i=1}^{r} \left( \frac{\tilde{b}_i}{\sigma_i} \right)^2 > \alpha^2$

Find $\lambda_+ > 0$ such that $\displaystyle\sum_{i=1}^{r} \left( \frac{\sigma_i \tilde{b}_i}{\sigma_i^2 + \lambda_+} \right)^2 = \alpha^2$.

$\displaystyle x = \sum_{i=1}^{r} \left( \frac{\sigma_i \tilde{b}_i}{\sigma_i^2 + \lambda_+} \right) v_i$

**else**

$\displaystyle x = \sum_{i=1}^{r} \left( \frac{\tilde{b}_i}{\sigma_i} \right) v_i$

**end**

The SVD is the dominant computation in this algorithm.

### 6.2.2 More General Quadratic Constraints

A more general version of (6.2.1) results if we minimize $\| Ax - b \|_2$ over an arbitrary hyperellipsoid:

$$\text{minimize } \| Ax - b \|_2 \qquad \text{subject to } \| Bx - d \|_2 \leq \alpha. \tag{6.2.5}$$

Here we are assuming that $A \in \mathbb{R}^{m_1 \times n_1}$, $b \in \mathbb{R}^{m_1}$, $B \in \mathbb{R}^{m_2 \times n_1}$, $d \in \mathbb{R}^{m_2}$, and $\alpha \geq 0$. Just as the SVD turns (6.2.1) into an equivalent diagonal problem, we can use the GSVD to transform (6.2.5) into a diagonal problem. In particular, if the GSVD of $A$ and $B$ is given by (6.1.22) and (6.2.23), then (6.2.5) is equivalent to

$$\text{minimize } \| D_A y - \tilde{b} \|_2 \qquad \text{subject to } \| D_B y - \tilde{d} \|_2 \leq \alpha \tag{6.2.6}$$

where

$$\tilde{b} = U_1^T b, \qquad \tilde{d} = U_2^T d, \qquad y = X^{-1} x.$$

The simple form of the objective function and the constraint equation facilitate the analysis. For example, if $\mathsf{rank}(B) = m_2 < n_1$, then

$$\| D_A y - \tilde{b} \|_2^2 = \sum_{i=1}^{n_1} (\alpha_i y_i - \tilde{b}_i)^2 + \sum_{i=n_1+1}^{m_1} \tilde{b}_i^2 \tag{6.2.7}$$

and

$$\| D_B y - \tilde{d} \|_2^2 = \sum_{i=1}^{m_2} (\beta_i y_i - \tilde{d}_i)^2 + \sum_{i=m_2+1}^{n_1} \tilde{d}_i^2 \le \alpha^2. \tag{6.2.8}$$

A Lagrange multiplier argument can be used to determine the solution to this transformed problem (if it exists).

### 6.2.3 Least Squares With Equality Constraints

We consider next the constrained least squares problem

$$\min_{Bx=d} \| Ax - b \|_2 \tag{6.2.9}$$

where $A \in \mathbb{R}^{m_1 \times n_1}$ with $m_1 \ge n_1$, $B \in \mathbb{R}^{m_2 \times n_1}$ with $m_2 < n_1$, $b \in \mathbb{R}^{m_1}$, and $d \in \mathbb{R}^{m_2}$. We refer to this as the *LSE problem* (least squares with equality constraints). By setting $\alpha = 0$ in (6.2.5) we see that the LSE problem is a special case of the LSQI problem. However, it is simpler to approach the LSE problem directly rather than through Lagrange multipliers.

For clarity, we assume that both $A$ and $B$ have full rank. Let

$$Q^T B^T = \begin{bmatrix} R \\ 0 \end{bmatrix} \begin{matrix} n_1 \\ n_1 - m_2 \end{matrix}$$

be the QR factorization of $B^T$ and set

$$AQ = \underset{m_2}{[\, A_1 \,|} \; \underset{n_1-m_2}{A_2 \,]}, \qquad Q^T x = \begin{bmatrix} y \\ z \end{bmatrix} \begin{matrix} m_2 \\ n_1 - m_2 \end{matrix}.$$

It is clear that with these transformations (6.2.9) becomes

$$\min_{R^T y = d} \| A_1 y + A_2 z - b \|_2.$$

Thus, $y$ is determined from the constraint equation $R^T y = d$ and the vector $z$ is obtained by solving the unconstrained LS problem

$$\min_{z \in \mathbb{R}^{n_1 - m_2}} \| A_2 z - (b - A_1 y) \|_2.$$

Combining the above, we see that the following vector solves the LSE problem:

$$x = Q \begin{bmatrix} y \\ z \end{bmatrix}.$$

**Algorithm 6.2.2** Suppose $A \in \mathbb{R}^{m_1 \times n_1}$, $B \in \mathbb{R}^{m_2 \times n_1}$, $b \in \mathbb{R}^{m_1}$, and $d \in \mathbb{R}^{m_2}$. If $\mathsf{rank}(A) = n_1$ and $\mathsf{rank}(B) = m_2 < n_1$, then the following algorithm minimizes $\| Ax - b \|_2$ subject to the constraint $Bx = d$.

Compute the QR factorization $B^T = QR$.
Solve $R(1{:}m_2, 1{:}m_2)^T \cdot y = d$ for $y$.
$A = AQ$
Find $z$ so $\| A(:, m_2 + 1{:}n_1)z - (b - A(:, 1{:}m_2)\cdot y) \|_2$ is minimized.
$x = Q(:, 1{:}m_2)\cdot y + Q(:, m_2 + 1{:}n_1)\cdot z$ .

Note that this approach to the LSE problem involves two QR factorizations and a matrix multiplication. If $A$ and/or $B$ are rank deficient, then it is possible to devise a similar solution procedure using the SVD instead of QR. Note that there may not be a solution if $\mathsf{rank}(B) < m_2$. Also, if $\mathsf{null}(A) \cap \mathsf{null}(B) \neq \{0\}$ and $d \in \mathsf{ran}(B)$, then the LSE solution is not unique.

### 6.2.4 LSE Solution Using the Augmented System

The LSE problem can also be approached through the method of Lagrange multipliers. Define the augmented objective function

$$f(x, \lambda) = \frac{1}{2}\| Ax - b \|_2^2 + \lambda^T(d - Bx), \qquad \lambda \in \mathbb{R}^{m_2},$$

and set to zero its gradient with respect to $x$:

$$A^T Ax - A^T b - B^T \lambda = 0.$$

Combining this with the equations $r = b - Ax$ and $Bx = d$ we obtain the symmetric indefinite linear system

$$\begin{bmatrix} 0 & A^T & B^T \\ A & I & 0 \\ B & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ r \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ d \end{bmatrix}. \tag{6.2.10}$$

This system is nonsingular if both $A$ and $B$ have full rank. The augmented system presents a solution framework for the sparse LSE problem.

### 6.2.5 LSE Solution Using the GSVD

Using the GSVD given by (6.1.22) and (6.1.23), we see that the LSE problem transforms to

$$\min_{D_B y = \tilde{d}} \| D_A y - \tilde{b} \|_2 \tag{6.2.11}$$

where $\tilde{b} = U_1^T b$, $\tilde{d} = U_2^T d$, and $y = X^{-1}x$. It follows that if $\mathsf{null}(A) \cap \mathsf{null}(B) = \{0\}$ and $X = [\, x_1 \,|\, \cdots \,|\, x_n \,]$, then

$$x = \sum_{i=1}^{m_2} \left( \frac{\tilde{d}_i}{\beta_i} \right) x_i + \sum_{i=m_2+1}^{n_1} \left( \frac{\tilde{b}_i}{\alpha_i} \right) x_i \tag{6.2.12}$$

solves the LSE problem.

### 6.2.6 LSE Solution Using Weights

An interesting way to obtain an approximate LSE solution is to solve the unconstrained LS problem

$$\min_x \left\| \begin{bmatrix} A \\ \sqrt{\lambda}B \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\lambda}d \end{bmatrix} \right\|_2 \tag{6.2.13}$$

for large $\lambda$. (Compare with the Tychanov regularization problem (6.1.21).) Since

$$\left\| \begin{bmatrix} A \\ \sqrt{\lambda}B \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\lambda}d \end{bmatrix} \right\|_2^2 = \| Ax - b \|_2^2 + \lambda \| Bx - d \|^2,$$

we see that there is a penalty for discrepancies among the constraint equations. To quantify this, assume that both $A$ and $B$ have full rank and substitute the GSVD defined by (6.1.22) and (6.1.23) into the normal equation system

$$(A^TA + \lambda B^TB)x = A^Tb + \lambda B^Td.$$

This shows that the solution $x(\lambda)$ is given by $x(\lambda) = Xy(\lambda)$ where $y(\lambda)$ solves

$$(D_A^TD_A + \lambda D_B^TD_B)y = D_A^T\tilde{b} + \lambda D_B^T\tilde{d}$$

with $\tilde{b} = U_1^Tb$ and $\tilde{d} = U_2^Td$. It follows that

$$x(\lambda) = \sum_{i=1}^{m_2} \left( \frac{\alpha_i\tilde{b}_i + \lambda\beta_i\tilde{d}_i}{\alpha_i^2 + \lambda\beta_i^2} \right) x_i + \sum_{i=m_2+1}^{n_1} \left( \frac{\tilde{b}_i}{\alpha_i} \right) x_i$$

and so from (6.2.13) we have

$$x(\lambda) - x = \sum_{i=1}^{p} \frac{\alpha_i}{\beta_i} \left( \frac{\beta_i u_i^Tb - \alpha_i v_i^Td}{\alpha_i^2 + \lambda^2\beta_i^2)} \right) x_i. \tag{6.2.14}$$

This shows that $x(\lambda) \to x$ as $\lambda \to \infty$. The appeal of this approach to the LSE problem is that it can be implemented with unconstrained LS problem software. However, for large values of $\lambda$ numerical problems can arise and it is necessary to take precautions. See Powell and Reid (1968) and Van Loan (1982).

#### Problems

**P6.2.1** Is the solution to (6.2.1) always unique?

**P6.2.2** Let $p_0(x), \ldots, p_n(x)$ be given polynomials and $(x_0, y_0), \ldots, (x_m, y_m)$ be a given set of coordinate pairs with $x_i \in [a, b]$. It is desired to find a polynomial $p(x) = \sum_{k=0}^{n} \alpha_k p_k(x)$ such that

$$\phi(\alpha) = \sum_{i=0}^{m} (p(x_i) - y_i)^2$$

is minimized subject to the constraint that

$$\int_a^b [p''(x)]^2 dx \approx h \sum_{i=0}^{N} \left( \frac{p(z_{i-1}) - 2p(z_i) + p(z_{i+1})}{h^2} \right)^2 \leq \alpha^2$$

where $z_i = a + ih$ and $b = a + Nh$. Show that this leads to an LSQI problem of the form (6.2.5) with $d = 0$.

**P6.2.3** Suppose $Y = [\, y_1 \,|\, \cdots \,|\, y_k \,] \in \mathbb{R}^{m \times k}$ has the property that

$$Y^T Y = \text{diag}(d_1^2, \ldots, d_k^2), \qquad d_1 \geq d_2 \geq \cdots \geq d_k > 0.$$

Show that if $Y = QR$ is the QR factorization of $Y$, then $R$ is diagonal with $|r_{ii}| = d_i$.

**P6.2.4** (a) Show that if $(A^T A + \lambda I)x = A^T b$, $\lambda > 0$, and $\| x \|_2 = \alpha$, then $z = (Ax - b)/\lambda$ solves the *dual equations* $(AA^T + \lambda I)z = -b$ with $\| A^T z \|_2 = \alpha$. (b) Show that if $(AA^T + \lambda I)z = -b$, $\| A^T z \|_2 = \alpha$, then $x = -A^T z$ satisfies $(A^T A + \lambda I)x = A^T b$, $\| x \|_2 = \alpha$.

**P6.2.5** Show how to compute $y$ (if it exists) so that both (6.2.7) and (6.2.8) are satisfied.

**P6.2.6** Develop an SVD version of Algorithm 6.2.2 that can handle the situation when $A$ and/or $B$ are rank deficient.

**P6.2.7** Suppose

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

where $A_1 \in \mathbb{R}^{n \times n}$ is nonsingular and $A_2 \in \mathbb{R}^{(m-n) \times n}$. Show that

$$\sigma_{\min}(A) \geq \sqrt{1 + \sigma_{\min}(A_2 A_1^{-1})^2}\ \sigma_{\min}(A_1)\,.$$

**P6.2.8** Suppose $p \geq m \geq n$ and that $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times p}$ Show how to compute orthogonal $Q \in \mathbb{R}^{m \times m}$ and orthogonal $V \in \mathbb{R}^{n \times n}$ so that

$$Q^T A = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad Q^T B V = [\, 0 \,|\, S \,]$$

where $R \in \mathbb{R}^{n \times n}$ and $S \in \mathbb{R}^{m \times m}$ are upper triangular.

**P6.2.9** Suppose $r \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, and $\delta > 0$. Show how to solve the problem

$$\min_{E \in \mathbb{R}^{m \times n},\, \|E\|_F \leq \delta} \|Ey - r\|_2$$

Repeat with "min" replaced by "max."

**P6.2.10** Show how the constrained least squares problem

$$\min_{Bx = d} \| Ax - b \|_2 \qquad A \in \mathbb{R}^{m \times n},\ B \in \mathbb{R}^{p \times n},\ \text{rank}(B) = p$$

can be reduced to an unconstrained least square problem by performing $p$ steps of Gaussian elimination on the matrix

$$\begin{bmatrix} B \\ A \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ A_1 & A_2 \end{bmatrix}, \qquad B_1 \in \mathbb{R}^{p \times p},\ \text{rank}(B_1) = p.$$

Explain. Hint: The Schur complement is of interest.

### Notes and References for §6.2

The LSQI problem is discussed in:

G.E. Forsythe and G.H. Golub (1965). "On the Stationary Values of a Second-Degree Polynomial on the Unit Sphere," *SIAM J. App. Math. 14,* 1050–1068.

L. Eldén (1980). "Perturbation Theory for the Least Squares Problem with Linear Equality Constraints," *SIAM J. Numer. Anal. 17,* 338–350.

W. Gander (1981). "Least Squares with a Quadratic Constraint," *Numer. Math. 36,* 291–307.

L. Eldén (1983). "A Weighted Pseudoinverse, Generalized Singular Values, and Constrained Least Squares Problems," *BIT 22* , 487–502.

G.W. Stewart (1984). "On the Asymptotic Behavior of Scaled Singular Value and QR Decompositions," *Math. Comput. 43,* 483–490.
G.H. Golub and U. von Matt (1991). "Quadratically Constrained Least Squares and Quadratic Problems," *Numer. Math. 59*, 561–580.
T.F. Chan, J.A. Olkin, and D. Cooley (1992). "Solving Quadratically Constrained Least Squares Using Black Box Solvers," *BIT 32*, 481–495.

Secular equation root-finding comes up in many numerical linear algebra settings. For an algorithmic overview, see:

O.E. Livne and A. Brandt (2002). "N Roots of the Secular Equation in O(N) Operations," *SIAM J. Matrix Anal. Applic. 24*, 439–453.

For a discussion of the augmented systems approach to least squares problems, see:

Å. Björck (1992). "Pivoting and Stability in the Augmented System Method," *Proceedings of the 14th Dundee Conference*, D.F. Griffiths and G.A. Watson (eds.), Longman Scientific and Technical, Essex, U.K.
Å. Björck and C.C. Paige (1994). "Solution of Augmented Linear Systems Using Orthogonal Factorizations," *BIT 34,* 1–24.

References that are concerned with the method of weighting for the LSE problem include:

M.J.D. Powell and J.K. Reid (1968). "On Applying Householder's Method to Linear Least Squares Problems," *Proc. IFIP Congress*, pp. 122–26.
C. Van Loan (1985). "On the Method of Weighting for Equality Constrained Least Squares Problems," *SIAM J. Numer. Anal. 22,* 851–864.
J.L. Barlow and S.L. Handy (1988). "The Direct Solution of Weighted and Equality Constrained Least-Squares Problems," *SIAM J. Sci. Stat. Comput. 9*, 704–716.
J.L. Barlow, N.K. Nichols, and R.J. Plemmons (1988). "Iterative Methods for Equality Constrained Least Squares Problems," SIAM J. Sci. Stat. Comput. 9, 892–906.
J.L. Barlow (1988). "Error Analysis and Implementation Aspects of Deferred Correction for Equality Constrained Least-Squares Problems," *SIAM J. Numer. Anal. 25*, 1340–1358.
J.L. Barlow and U.B. Vemulapati (1992). "A Note on Deferred Correction for Equality Constrained Least Squares Problems," *SIAM J. Numer. Anal. 29*, 249–256.
M. Gulliksson and P.-Å. Wedin (1992). "Modifying the QR-Decomposition to Constrained and Weighted Linear Least Squares," *SIAM J. Matrix Anal. Applic. 13*, 1298–1313.
M. Gulliksson (1994). "Iterative Refinement for Constrained and Weighted Linear Least Squares," *BIT 34*, 239–253.
G. W. Stewart (1997). "On the Weighting Method for Least Squares Problems with Linear Equality Constraints," *BIT 37*, 961–967.

For the analysis of the LSE problem and related methods, see:

M. Wei (1992). "Perturbation Theory for the Rank-Deficient Equality Constrained Least Squares Problem," *SIAM J. Numer. Anal. 29*, 1462–1481.
M. Wei (1992). "Algebraic Properties of the Rank-Deficient Equality-Constrained and Weighted Least Squares Problems," *Lin. Alg. Applic. 161*, 27–44.
M. Gulliksson (1995). "Backward Error Analysis for the Constrained and Weighted Linear Least Squares Problem When Using the Weighted QR Factorization," *SIAM J. Matrix. Anal. Applic. 13*, 675–687.
M. Gulliksson (1995). "Backward Error Analysis for the Constrained and Weighted Linear Least Squares Problem When Using the Weighted *QR* Factorization," *SIAM J. Matrix Anal. Applic. 16*, 675–687.
J. Ding and W. Hang (1998). "New Perturbation Results for Equality-Constrained Least Squares Problems," *Lin. Alg. Applic. 272,* 181–192.
A.J. Cox and N.J. Higham (1999). "Accuracy and Stability of the Null Space Method for Solving the Equality Constrained Least Squares Problem," *BIT 39*, 34–50.
A.J. Cox and N.J. Higham (1999). "Row-Wise Backward Stable Elimination Methods for the Equality Constrained Least Squares Problem," *SIAM J. Matrix Anal. Applic. 21*, 313–326.
A.J. Cox and Nicholas J. Higham (1999). "Backward Error Bounds for Constrained Least Squares Problems," *BIT 39*, 210–227.

M. Gulliksson and P-A. Wedin (2000). "Perturbation Theory for Generalized and Constrained Linear Least Squares," *Num. Lin. Alg. 7*, 181–195.
M. Wei and A.R. De Pierro (2000). "Upper Perturbation Bounds of Weighted Projections, Weighted and Constrained Least Squares Problems," *SIAM J. Matrix Anal. Applic. 21*, 931–951.
E.Y. Bobrovnikova and S.A. Vavasis (2001). "Accurate Solution of Weighted Least Squares by Iterative Methods *SIAM. J. Matrix Anal. Applic. 22*, 1153–1174.
M. Gulliksson, X-Q.Jin, and Y-M. Wei (2002). "Perturbation Bounds for Constrained and Weighted Least Squares Problems," *Lin. Alg. Applic. 349,* 221–232.

## 6.3 Total Least Squares

The problem of minimizing $\| Ax - b \|_2$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ can be recast as follows:

$$\min_{b+r \in \mathsf{ran}(A)} \| r \|_2 \ . \tag{6.3.1}$$

In this problem, there is a tacit assumption that the errors are confined to the *vector of observations* $b$. If error is also present in the *data matrix* $A$, then it may be more natural to consider the problem

$$\min_{b+r \in \mathsf{ran}(A+E)} \| [\, E \,|\, r \,] \|_F \ . \tag{6.3.2}$$

This problem, discussed by Golub and Van Loan (1980), is referred to as the *total least squares* (TLS) problem. If a minimizing $[\, E_0 \,|\, r_0 \,]$ can be found for (6.3.2), then any $x$ satisfying $(A + E_0)x = b + r_0$ is called a TLS solution. However, it should be realized that (6.3.2) may fail to have a solution altogether. For example, if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad E_\epsilon = \begin{bmatrix} 0 & 0 \\ 0 & \epsilon \\ 0 & \epsilon \end{bmatrix},$$

then for all $\epsilon > 0$, $b \in \mathsf{ran}(A + E_\epsilon)$. However, there is no smallest value of $\| [\, E \,,\, r \,] \|_F$ for which $b + r \in \mathsf{ran}(A + E)$.

A generalization of (6.3.2) results if we allow multiple right-hand sides and use a weighted Frobenius norm. In particular, if $B \in \mathbb{R}^{m\times k}$ and the matrices

$$D = \mathrm{diag}(d_1, \ldots, d_m),$$
$$T = \mathrm{diag}(t_1, \ldots, t_{n+k})$$

are nonsingular, then we are led to an optimization problem of the form

$$\min_{B+R \in \mathsf{ran}(A+E)} \| D\,[\, E \,|\, R \,]\, T \|_F \tag{6.3.3}$$

where $E \in \mathbb{R}^{m\times n}$ and $R \in \mathbb{R}^{m\times k}$. If $[\, E_0 \,|\, R_0 \,]$ solves (6.3.3), then any $X \in \mathbb{R}^{n\times k}$ that satisfies

$$(A + E_0)X \ = \ (B + R_0)$$

is said to be a TLS solution to (6.3.3).

In this section we discuss some of the mathematical properties of the total least squares problem and show how it can be solved using the SVD. For a more detailed introduction, see Van Huffel and Vanderwalle (1991).

### 6.3.1 Mathematical Background

The following theorem gives conditions for the uniqueness and existence of a TLS solution to the multiple-right-hand-side problem.

**Theorem 6.3.1.** *Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times k}$ and that $D = \mathrm{diag}(d_1,\ldots,d_m)$ and $T = \mathrm{diag}(t_1,\ldots,t_{n+k})$ are nonsingular. Assume $m \geq n+k$ and let the SVD of*

$$C = D[\,A \mid B\,]T = [\underset{n}{\,C_1\,} \mid \underset{k}{C_2\,}]$$

*be specified by $U^TCV = \mathrm{diag}(\sigma_1,\ldots,\sigma_{n+k}) = \Sigma$ where $U$, $V$, and $\Sigma$ are partitioned as follows:*

$$U = [\underset{n}{\,U_1\,} \mid \underset{k}{U_2\,}], \qquad V = \begin{bmatrix} \underset{n}{V_{11}} & \underset{k}{V_{12}} \\ V_{21} & V_{22} \end{bmatrix}\begin{matrix} n \\ k \end{matrix}, \qquad \Sigma = \begin{bmatrix} \underset{n}{\Sigma_1} & \underset{k}{0} \\ 0 & \Sigma_2 \end{bmatrix}\begin{matrix} n \\ k \end{matrix}.$$

*If $\sigma_n(C_1) > \sigma_{n+1}(C)$, then the matrix $[\,E_0 \mid R_0\,]$ defined by*

$$D[\,E_0 \mid R_0\,]T = -U_2\Sigma_2[\,V_{12}^T \mid V_{22}^T\,] \tag{6.3.4}$$

*solves (6.3.3). If $T_1 = \mathrm{diag}(t_1,\ldots,t_n)$ and $T_2 = \mathrm{diag}(t_{n+1},\ldots,t_{n+k})$, then the matrix*

$$X_{TLS} = -T_1V_{12}V_{22}^{-1}T_2^{-1}$$

*exists and is the unique TLS solution to $(A+E_0)X = B+R_0$.*

***Proof.*** We first establish two results that follow from the assumption $\sigma_n(C_1) > \sigma_{n+1}(C)$. From the equation $CV = U\Sigma$ we have

$$C_1V_{12} + C_2V_{22} = U_2\Sigma_2.$$

We wish to show that $V_{22}$ is nonsingular. Suppose $V_{22}x = 0$ for some unit 2-norm $x$. It follows from

$$V_{12}^TV_{12} + V_{22}^TV_{22} = I$$

that $\| V_{12}x \|_2 = 1$. But then

$$\sigma_{n+1}(C) \geq \| U_2\Sigma_2x \|_2 = \| C_1V_{12}x \|_2 \geq \sigma_n(C_1)\,,$$

a contradiction. Thus, the submatrix $V_{22}$ is nonsingular. The second fact concerns the strict separation of $\sigma_n(C)$ and $\sigma_{n+1}(C)$. From Corollary 2.4.5, we have $\sigma_n(C) \geq \sigma_n(C_1)$ and so

$$\sigma_n(C) \geq \sigma_n(C_1) > \sigma_{n+1}(C).$$

We are now set to prove the theorem. If $\mathsf{ran}(B+R) \subset \mathsf{ran}(A+E)$, then there is an $X$ ($n$-by-$k$) so $(A+E)X = B+R$, i.e.,

$$\{\, D[\,A \mid B\,]T + D[\,E \mid R\,]T\,\}\,T^{-1}\begin{bmatrix} X \\ -I_k \end{bmatrix} = 0\,. \tag{6.3.5}$$

Thus, the rank of the matrix in curly brackets is at most equal to $n$. By following the argument in the proof of Theorem 2.4.8, it can be shown that

$$\| D [ E \mid R ] T \|_F^2 \geq \sum_{i=n+1}^{n+k} \sigma_i(C)^2.$$

Moreover, the lower bound is realized by setting $[ E \mid R ] = [ E_0 \mid R_0 ]$. Using the inequality $\sigma_n(C) > \sigma_{n+1}(C)$, we may infer that $[ E_0 \mid R_0 ]$ is the unique minimizer.

To identify the TLS solution $X_{TLS}$, we observe that the nullspace of

$$\{ D [ A \mid B ] T + D [ E_0 \mid R_0 ] T \} = U_1 \Sigma_1 [ V_{11}^T \mid V_{21}^T ]$$

is the range of $\begin{bmatrix} V_{12} \\ V_{22} \end{bmatrix}$. Thus, from (6.3.5)

$$T^{-1} \begin{bmatrix} X \\ -I_k \end{bmatrix} = \begin{bmatrix} V_{12} \\ V_{22} \end{bmatrix} S$$

for some $k$-by-$k$ matrix $S$. From the equations $T_1^{-1} X = V_{12} S$ and $-T_2^{-1} = V_{22} S$ we see that $S = -V_{22}^{-1} T_2^{-1}$ and so

$$X = T_1 V_{12} S = -T_1 V_{12} V_{22}^{-1} T_2^{-1} = X_{\text{TLS}}. \quad \square$$

Note from the thin CS decomposition (Theorem 2.5.2) that

$$\| X \|_\tau^2 = \| V_{12} V_{22}^{-1} \|_2^2 = \frac{1 - \sigma_k(V_{22})^2}{\sigma_k(V_{22})^2}$$

where we define the "$\tau$-norm" on $\mathbb{R}^{n \times k}$ by $\| Z \|_\tau = \| T_1^{-1} Z T_2 \|_2$.

If $\sigma_n(C_1) = \sigma_{n+1}(C)$, then the solution procedure implicit in the above proof is problematic. The TLS problem may have no solution or an infinite number of solutions. See §6.3.4 for suggestions as to how one might proceed.

### 6.3.2 Solving the Single Right Hand Side Case

We show how to maximize $\sigma_k(V_{22})$ in the important $k = 1$ case. Suppose the singular values of $C$ satisfy $\sigma_{n-p} > \sigma_{n-p+1} = \cdots = \sigma_{n+1}$ and let $V = [ v_1 \mid \cdots \mid v_{n+1} ]$ be a column partitioning of $V$. If $\widetilde{Q}$ is a Householder matrix such that

$$V(:, n+1-p{:}n+1) \widetilde{Q} = \begin{bmatrix} W & z \\ 0 & \alpha \end{bmatrix} \begin{matrix} n \\ 1 \end{matrix},$$
$$\qquad\qquad\qquad\quad p \quad 1$$

then the last column of this matrix has the largest $(n+1)$st component of all the vectors in $\mathsf{span}\{v_{n+1-p}, \ldots, v_{n+1}\}$. If $\alpha = 0$, then the TLS problem has no solution. Otherwise

$$x_{\text{TLS}} = -T_1 z / (t_{n+1} \alpha).$$

Moreover,

$$\begin{bmatrix} I_{n-1} & 0 \\ 0 & \widetilde{Q} \end{bmatrix} U^T (D[\,A \,|\, b\,]T)V \begin{bmatrix} I_{n-p} & 0 \\ 0 & \widetilde{Q} \end{bmatrix} = \Sigma$$

and so

$$D\,[\,E_0 \,|\, r_0\,]\,T \;=\; -D\,[\,A \,|\, b\,]\,T \begin{bmatrix} z \\ \alpha \end{bmatrix} [\,z^T \,|\, \alpha\,].$$

Overall, we have the following algorithm:

**Algorithm 6.3.1** Given $A \in \mathbb{R}^{m\times n}$ $(m > n)$, $b \in \mathbb{R}^m$, nonsingular $D = \mathrm{diag}(d_1, \ldots, d_m)$, and nonsingular $T = \mathrm{diag}(t_1, \ldots, t_{n+1})$, the following algorithm computes (if possible) a vector $x_{\mathrm{TLS}} \in \mathbb{R}^n$ such that $(A+E_0)x_{\mathrm{TLS}} = (b+r_0)$ and $\| D[\,E_0 \,|\, r_0\,]T \|_F$ is minimal.

Compute the SVD $U^T(D[\,A \,|\, b\,]T)V \;=\; \mathrm{diag}(\sigma_1, \ldots, \sigma_{n+1})$ and save $V$.

Determine $p$ such that $\sigma_1 \geq \cdots \geq \sigma_{n-p} > \sigma_{n-p+1} = \cdots = \sigma_{n+1}$.

Compute a Householder $P$ such that if $\tilde{V} = VP$, then $\tilde{V}(n+1, n-p+1{:}n) = 0$.

**if** $\tilde{v}_{n+1,n+1} \neq 0$

  **for** $i = 1{:}n$

    $x_i = -t_i\tilde{v}_{i,n+1}/(t_{n+1}\tilde{v}_{n+1,n+1})$

  **end**

  $x_{\mathrm{TLS}} = x$

**end**

This algorithm requires about $2mn^2 + 12n^3$ flops and most of these are associated with the SVD computation.

### 6.3.3 A Geometric Interpretation

It can be shown that the TLS solution $x_{TLS}$ minimizes

$$\psi(x) \;=\; \sum_{i=1}^{m} d_i^2 \left( \frac{|a_i^T x - b_i|^2}{x^T T_1^{-2} x + t_{n+1}^{-2}} \right) \tag{6.3.6}$$

where $a_i^T$ is the $i$th row of $A$ and $b_i$ is the $i$th component of $b$. A geometrical interpretation of the TLS problem is made possible by this observation. Indeed,

$$\delta_i \;=\; \frac{|a_i^T x - b_i|^2}{x^T T_1^{-2} x + t_{n+1}^{-2}}$$

is the square of the distance from

$$\begin{bmatrix} a_i \\ b_i \end{bmatrix} \in \mathbb{R}^{n+1}$$

to the nearest point in the subspace

$$P_x \;=\; \left\{ \begin{bmatrix} a \\ b \end{bmatrix} : a \in \mathbb{R}^n,\; b \in \mathbb{R},\; b = x^T a \right\}$$

where the distance in $\mathbb{R}^{n+1}$ is measured by the norm $\| z \| = \| Tz \|_2$. The TLS problem is essentially the problem of *orthogonal regression*, a topic with a long history. See Pearson (1901) and Madansky (1959).

### 6.3.4 Variations of the Basic TLS Problem

We briefly mention some modified TLS problems that address situations when additional constraints are imposed on the optimizing $E$ and $R$ and the associated TLS solution.

In the *restricted TLS problem*, we are given $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times k}$, $P_1 \in \mathbb{R}^{m\times q}$, and $P_2 \in \mathbb{R}^{n+k\times r}$, and solve

$$\min_{B+R \subset \mathsf{ran}(A+E)} \| P_1^T [\, E \mid R \,] P_2 \|_F \ . \tag{6.3.7}$$

We assume that $q \le m$ and $r \le n + k$. An important application arises if some of the columns of $A$ are error-free. For example, if the first $s$ columns of $A$ are error-free, then it makes sense to force the optimizing $E$ to satisfy $E(:, 1{:}s) = 0$. This goal is achieved by setting $P_1 = I_m$ and $P_2 = I_{m+k}(:, s + 1{:}n + k)$ in the restricted TLS problem.

If a particular TLS problem has no solution, then it is referred to as a *nongeneric TLS problem*. By adding a constraint it is possible to produce a meaningful solution. For example, let $U^T [\, A \mid b \,] V = \Sigma$ be the SVD and let $p$ be the largest index so $V(n + 1, p) \neq 0$. It can be shown that the problem

$$\min_{\substack{(A+E)x=b+r \\ [\, E \mid r \,]V(:,p+1:n+1)=0}} \| [\, E \mid r \,] \|_F \tag{6.3.8}$$

has a solution $[\, E_0 \mid r_0 \,]$ and the nongeneric TLS solution satisfies $(A + E_0)x + b + r_0$. See Van Huffel (1992).

In the *regularized TLS problem* additional constraints are imposed to ensure that the solution $x$ is properly constrained/smoothed:

$$\min_{\substack{(A+E)x=b+r \\ \|Lx\|_2 \le \delta}} \| [\, E \mid r \,] \|_F \ . \tag{6.3.9}$$

The matrix $L \in \mathbb{R}^{n\times n}$ could be the identity or a discretized second-derivative operator. The regularized TLS problem leads to a Lagrange multiplier system of the form

$$(A^T A + \lambda_1 I + \lambda_2 L^T L)x \ = \ A^T b.$$

See Golub, Hansen, and O'Leary (1999) for more details. Another regularization approach involves setting the small singular values of $[A \mid b]$ to zero. This is the *truncated TLS problem* discussed in Fierro, Golub, Hansen, and O'Leary (1997).

#### Problems

**P6.3.1** Consider the TLS problem (6.3.2) with nonsingular $D$ and $T$. (a) Show that if $\mathsf{rank}(A) < n$, then (6.3.2) has a solution if and only if $b \in \mathsf{ran}(A)$. (b) Show that if $\mathsf{rank}(A) = n$, then (6.3.2) has no

solution if $A^T D^2 b = 0$ and $|t_{n+1}|\| Db \|_2 \geq \sigma_n(DAT_1)$ where $T_1 = \text{diag}(t_1, \ldots, t_n)$.

**P6.3.2** Show that if $C = D[\, A \mid b\,]T = [\, A_1 \mid d\,]$ and $\sigma_n(C) > \sigma_{n+1}(C)$, then $x_{TLS}$ satisfies

$$(A_1^T A_1 - \sigma_{n+1}(C)^2 I)x_{\text{TLS}} = A_1^T d.$$

Appreciate this as a "negatively shifted" system of normal equations.

**P6.3.3** Show how to solve (6.3.2) with the added constraint that the first $p$ columns of the minimizing $E$ are zero. Hint: Compute the QR factorization of $A(:, 1{:}p)$.

**P6.3.4** Show how to solve (6.3.3) given that $D$ and $T$ are general nonsingular matrices.

**P6.3.5** Verify Equation (6.3.6).

**P6.3.6** If $A \in \mathbb{R}^{m\times n}$ has full column rank and $B \in \mathbb{R}^{p\times n}$ has full row rank, show how to minimize

$$f(x) = \frac{\| Ax - b \|_2^2}{1 + x^T x}$$

subject to the constraint that $Bx = 0$.

**P6.3.7** In the *data least squares problem*, we are given $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ and minimize $\| E \|_F$ subject to the constraint that $b \in \mathsf{ran}(A + E)$. Show how to solve this problem. See Paige and Strakoš (2002b).

**Notes and References for §6.3**

Much of this section is based on:

G.H. Golub and C.F. Van Loan (1980). "An Analysis of the Total Least Squares Problem," *SIAM J. Numer. Anal. 17,* 883–93.

The idea of using the SVD to solve the TLS problem is set forth in:

G.H. Golub and C. Reinsch (1970). "Singular Value Decomposition and Least Squares Solutions," *Numer. Math. 14*, 403–420.
G.H. Golub (1973). "Some Modified Matrix Eigenvalue Problems," *SIAM Review 15*, 318–334.

The most comprehensive treatment of the TLS problem is:

S. Van Huffel and J. Vandewalle (1991). *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM Publications, Philadelphia, PA.

There are two excellent conference proceedings that cover just about everything you would like to know about TLS algorithms, generalizations, applications, and the associated statistical foundations:

S. Van Huffel (ed.) (1996). *Recent Advances in Total Least Squares Techniques and Errors in Variables Modeling,* SIAM Publications, Philadelphia, PA.
S. Van Huffel and P. Lemmerling (eds.) (2002) *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms, and Applications*, Kluwer Academic, Dordrecht, The Netherlands.

TLS is but one approach to the errors-in-variables problem, a subject that has a long and important history in statistics:

K. Pearson (1901). "On Lines and Planes of Closest Fit to Points in Space," *Phil. Mag. 2,* 559–72.
A. Wald (1940). "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics 11,* 284–300.
G.W. Stewart (2002). "Errors in Variables for Numerical Analysts," in *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling*, S. Van Huffel (ed.), SIAM Publications, Philadelphia PA, pp. 3–10,

In certain settings there are more economical ways to solve the TLS problem than the Golub-Kahan-Reinsch SVD algorithm:

S. Van Huffel and H. Zha (1993). "An Efficient Total Least Squares Algorithm Based On a Rank-Revealing Two-Sided Orthogonal Decomposition," *Numer. Alg. 4,* 101–133.
Å. Björck, P. Heggernes, and P. Matstoms (2000). "Methods for Large Scale Total Least Squares Problems," *SIAM J. Matrix Anal. Applic. 22*, 413–429.

H. Guo and R.A. Renaut (2005). "Parallel Variable Distribution for Total Least Squares," *Num. Lin. Alg. 12*, 859–876.

The condition of the TLS problem is analyzed in:

M. Baboulin and S. Gratton (2011). "A Contribution to the Conditioning of the Total Least-Squares Problem," *SIAM J. Matrix Anal. Applic. 32*, 685–699.

Efforts to connect the LS and TLS paradigms have lead to nice treatments that unify the presentation of both approaches:

B.D. Rao (1997). "Unified Treatment of LS, TLS, and Truncated SVD Methods Using a Weighted TLS Framework," in *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling*, S. Van Huffel (ed.), SIAM Publications, Philadelphia, PA., pp. 11–20.
C.C. Paige and Z. Strakoš (2002a). "Bounds for the Least Squares Distance Using Scaled Total Least Squares," *Numer. Math. 91*, 93–115.
C.C. Paige and Z. Strakoš (2002b). "Scaled Total Least Squares Fundamentals," *Numer. Math. 91*, 117–146.
X.-W. Chang, G.H. Golub, and C.C. Paige (2008). "Towards a Backward Perturbation Analysis for Data Least Squares Problems," *SIAM J. Matrix Anal. Applic. 30*, 1281–1301.
X.-W. Chang and D. Titley-Peloquin (2009). "Backward Perturbation Analysis for Scaled Total Least-Squares," *Num. Lin. Alg. Applic. 16*, 627–648.

For a discussion of the situation when there is no TLS solution or when there are multiple solutions, see:

S. Van Huffel and J. Vandewalle (1988). "Analysis and Solution of the Nongeneric Total Least Squares Problem," *SIAM J. Matrix Anal. Appl. 9*, 360–372.
S. Van Huffel (1992). "On the Significance of Nongeneric Total Least Squares Problems," *SIAM J. Matrix Anal. Appl. 13*, 20–35.
M. Wei (1992). "The Analysis for the Total Least Squares Problem with More than One Solution," *SIAM J. Matrix Anal. Appl. 13*, 746–763.

For a treatment of the multiple right hand side TLS problem, see:

I. Hnětynkovã, M. Plešinger, D.M. Sima, Z. Strakoš, and S. Van Huffel (2011). "The Total Least Squares Problem in AX ≈ B: A New Classification with the Relationship to the Classical Works," *SIAM J. Matrix Anal. Applic. 32*, 748–770.

If some of the columns of $A$ are known exactly then it is sensible to force the TLS perturbation matrix $E$ to be zero in the same columns. Aspects of this constrained TLS problem are discussed in:

J.W. Demmel (1987). "The Smallest Perturbation of a Submatrix which Lowers the Rank and Constrained Total Least Squares Problems," *SIAM J. Numer. Anal.* 24, 199–206.
S. Van Huffel and J. Vandewalle (1988). "The Partial Total Least Squares Algorithm," *J. Comput. App. Math. 21,* 333–342.
S. Van Huffel and J. Vandewalle (1989). "Analysis and Properties of the Generalized Total Least Squares Problem $AX \approx B$ When Some or All Columns in $A$ are Subject to Error," *SIAM J. Matrix Anal. Applic. 10*, 294–315.
S. Van Huffel and H. Zha (1991). "The Restricted Total Least Squares Problem: Formulation, Algorithm, and Properties," *SIAM J. Matrix Anal. Applic. 12*, 292–309.
C.C. Paige and M. Wei (1993). "Analysis of the Generalized Total Least Squares Problem $AX = B$ when Some of the Columns are Free of Error," *Numer. Math. 65*, 177–202.

Another type of constraint that can be imposed in the TLS setting is to insist that the optimum perturbation of $A$ have the same structure as $A$. For examples and related strategies, see:

J. Kamm and J.G. Nagy (1998). "A Total Least Squares Method for Toeplitz Systems of Equations," *BIT 38*, 560–582.
P. Lemmerling, S. Van Huffel, and B. De Moor (2002). "The Structured Total Least Squares Approach for Nonlinearly Structured Matrices," *Num. Lin. Alg. 9*, 321–332.
P. Lemmerling, N. Mastronardi, and S. Van Huffel (2003). "Efficient Implementation of a Structured Total Least Squares Based Speech Compression Method," *Lin. Alg. Applic. 366*, 295–315.
N. Mastronardi, P. Lemmerling, and S. Van Huffel (2004). "Fast Regularized Structured Total Least Squares Algorithm for Solving the Basic Deconvolution Problem," *Num. Lin. Alg. 12*, 201–209.

I. Markovsky, S. Van Huffel, and R. Pintelon (2005). "Block-Toeplitz/Hankel Structured Total Least Squares," *SIAM J. Matrix Anal. Applic. 26*, 1083–1099.
A. Beck and A. Ben-Tal (2005). "A Global Solution for the Structured Total Least Squares Problem with Block Circulant Matrices," *SIAM J. Matrix Anal. Applic. 27*, 238–255.
H. Fu, M.K. Ng, and J.L. Barlow (2006). "Structured Total Least Squares for Color Image Restoration," *SIAM J. Sci. Comput. 28*, 1100–1119.

As in the least squares problem, there are techniques that can be used to regularlize an otherwise "wild" TLS solution:

R.D. Fierro and J.R. Bunch (1994). "Collinearity and Total Least Squares," *SIAM J. Matrix Anal. Applic. 15*, 1167–1181.
R.D. Fierro, G.H. Golub, P.C. Hansen and D.P. O'Leary (1997). "Regularization by Truncated Total Least Squares," *SIAM J. Sci. Comput. 18*, 1223–1241.
G.H. Golub, P.C. Hansen, and D.P. O'Leary (1999). "Tikhonov Regularization and Total Least Squares," *SIAM J. Matrix Anal. Applic. 21*, 185–194.
R.A. Renaut and H. Guo (2004). "Efficient Algorithms for Solution of Regularized Total Least Squares," *SIAM J. Matrix Anal. Applic. 26*, 457–476.
D.M. Sima, S. Van Huffel, and G.H. Golub (2004). "Regularized Total Least Squares Based on Quadratic Eigenvalue Problem Solvers," *BIT 44*, 793–812.
N. Mastronardi, P. Lemmerling, and S. Van Huffel (2005). "Fast Regularized Structured Total Least Squares Algorithm for Solving the Basic Deconvolution Problem," *Num. Lin. Alg. Applic. 12*, 201–209.
S. Lu, S.V. Pereverzev, and U. Tautenhahn (2009). "Regularized Total Least Squares: Computational Aspects and Error Bounds," *SIAM J. Matrix Anal. Applic. 31*, 918–941.

Finally, we mention an interesting TLS problem where the solution is subject to a unitary constraint:

K.S. Arun (1992). "A Unitarily Constrained Total Least Squares Problem in Signal Processing," *SIAM J. Matrix Anal. Applic. 13*, 729–745.

## 6.4 Subspace Computations with the SVD

It is sometimes necessary to investigate the relationship between two given subspaces. How close are they? Do they intersect? Can one be "rotated" into the other? And so on. In this section we show how questions like these can be answered using the singular value decomposition.

### 6.4.1 Rotation of Subspaces

Suppose $A \in \mathbb{R}^{m\times p}$ is a data matrix obtained by performing a certain set of experiments. If the same set of experiments is performed again, then a different data matrix, $B \in \mathbb{R}^{m\times p}$, is obtained. In the *orthogonal Procrustes problem* the possibility that $B$ can be rotated into $A$ is explored by solving the following problem:

$$\text{minimize} \parallel A - BQ \parallel_F , \qquad \text{subject to } Q^TQ = I_p . \tag{6.4.1}$$

We show that optimizing $Q$ can be specified in terms of the SVD of $B^TA$. The *matrix trace* is critical to the derivation. The trace of a matrix is the sum of its diagonal entries:

$$\mathsf{tr}(C) = \sum_{i=1}^{n} c_{ii} , \qquad C \in \mathbb{R}^{n\times n}.$$

It is easy to show that if $C_1$ and $C_2$ have the same row and column dimension, then

$$\mathsf{tr}(C_1^TC_2) = \mathsf{tr}(C_2^TC_1) . \tag{6.4.2}$$

Returning to the Procrustes problem (6.4.1), if $Q \in \mathbb{R}^{p\times p}$ is orthogonal, then

$$\begin{aligned}
\| A - BQ \|_F^2 &= \sum_{k=1}^{p} \| A(:,k) - B\cdot Q(:,k) \|_2^2 \\
&= \sum_{k=1}^{p} \| A(:,k) \|_2^2 + \| BQ(:,k) \|_2^2 \; - \; 2Q(:,k)^T B^T A(:,k) \\
&= \| A \|_F^2 \; + \; \| BQ \|_F^2 \; - \; 2\sum_{k=1}^{p} \left[ Q^T (B^T A) \right]_{kk} \\
&= \| A \|_F^2 \; + \; \| B \|_F^2 \; - \; 2\mathsf{tr}(Q^T (B^T A)).
\end{aligned}$$

Thus, (6.4.1) is equivalent to the problem

$$\max_{Q^TQ=I_p} \mathsf{tr}(Q^T B^T A)\,.$$

If $U^T(B^TA)V = \Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_p)$ is the SVD of $B^TA$ and we define the orthogonal matrix $Z$ by $Z = V^TQ^TU$, then by using (6.4.2) we have

$$\mathsf{tr}(Q^T B^T A) \; = \; \mathsf{tr}(Q^T U \Sigma V^T) \; = \; \mathsf{tr}(Z\Sigma) \; = \; \sum_{i=1}^{p} z_{ii}\sigma_i \; \leq \; \sum_{i=1}^{p} \sigma_i\,.$$

The upper bound is clearly attained by setting $Z = I_p$, i.e., $Q = UV^T$.

**Algorithm 6.4.1** Given $A$ and $B$ in $\mathbb{R}^{m\times p}$, the following algorithm finds an orthogonal $Q \in \mathbb{R}^{p\times p}$ such that $\| A - BQ \|_F$ is minimum.

$C = B^TA$
Compute the SVD $U^TCV = \Sigma$ and save $U$ and $V$.
$Q = UV^T$

We mention that if $B = I_p$, then the problem (6.4.1) is related to the *polar decomposition*. This decomposition states that any square matrix $A$ has a factorization of the form $A = QP$ where $Q$ is orthogonal and $P$ is symmetric and positive semidefinite. Note that if $A = U\Sigma V^T$ is the SVD of $A$, then $A = (UV^T)(V\Sigma V^T)$ is its polar decomposition. For further discussion, see §9.4.3.

### 6.4.2 Intersection of Nullspaces

Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times n}$ be given, and consider the problem of finding an orthonormal basis for $\mathsf{null}(A) \cap \mathsf{null}(B)$. One approach is to compute the nullspace of the matrix

$$C \; = \; \begin{bmatrix} A \\ B \end{bmatrix}$$

since this is just what we want: $Cx = 0 \Leftrightarrow x \in \mathsf{null}(A) \cap \mathsf{null}(B)$. However, a more economical procedure results if we exploit the following theorem.

**Theorem 6.4.1.** *Suppose $A \in \mathbb{R}^{m\times n}$ and let $\{z_1,\ldots,z_t\}$ be an orthonormal basis for* $\mathsf{null}(A)$. *Define $Z = [\, z_1 \,|\cdots|\, z_t \,]$ and let $\{w_1,\ldots,w_q\}$ be an orthonormal basis for* $\mathsf{null}(BZ)$ *where $B \in \mathbb{R}^{p\times n}$. If $W = [\, w_1 \,|\cdots|\, w_q \,]$, then the columns of $ZW$ form an orthonormal basis for* $\mathsf{null}(A) \cap \mathsf{null}(B)$.

***Proof.*** Since $AZ = 0$ and $(BZ)W = 0$, we clearly have $\mathsf{ran}(ZW) \subset \mathsf{null}(A) \cap \mathsf{null}(B)$. Now suppose $x$ is in both $\mathsf{null}(A)$ and $\mathsf{null}(B)$. It follows that $x = Za$ for some $0 \neq a \in \mathbb{R}^t$. But since $0 = Bx = BZa$, we must have $a = Wb$ for some $b \in \mathbb{R}^q$. Thus, $x = ZWb \in \mathsf{ran}(ZW)$. □

If the SVD is used to compute the orthonormal bases in this theorem, then we obtain the following procedure:

**Algorithm 6.4.2** Given $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times n}$, the following algorithm computes and integer $s$ and a matrix $Y = [\, y_1 \,|\cdots|\, y_s \,]$ having orthonormal columns which span $\mathsf{null}(A) \cap \mathsf{null}(B)$. If the intersection is trivial, then $s = 0$.

```
Compute the SVD U_A^T A V_A = diag(σ_i), save V_A, and set r = rank(A).
if r < n
    C = B V_A(:, r+1:n)
    Compute the SVD U_C^T C V_C = diag(γ_i), save V_C, and set q = rank(C).
    if q < n − r
        s = n − r − q
        Y = V_A(:, r+1:n) V_C(:, q+1:n−r)
    else
        s = 0
    end
else
    s = 0
end
```

The practical implementation of this algorithm requires an ability to reason about numerical rank. See §5.4.1.

### 6.4.3 Angles Between Subspaces

Let $F$ and $G$ be subspaces in $\mathbb{R}^m$ whose dimensions satisfy

$$p = \mathsf{dim}(F) \geq \mathsf{dim}(G) = q \geq 1.$$

The *principal angles* $\{\theta_i\}_{i=1}^q$ between these two subspaces and the associated *principal vectors* $\{f_1, g_i\}_{i=1}^q$ are defined recursively by

$$\cos(\theta_k) = f_k^T g_k = \max_{\substack{f\in F,\ \|f\|_2=1 \\ f^T[f_1,\ldots,f_{k-1}]=0}} \ \max_{\substack{g\in G,\ \|g\|_2=1 \\ g^T[g_1,\ldots,g_{k-1}]=0}} f^T g \ . \tag{6.4.3}$$

Note that the principal angles satisfy $0 \le \theta_1 \le \cdots \le \theta_q \le \pi/2$.. The problem of computing principal angles and vectors is oftentimes referred to as the *canonical correlation problem.*

Typically, the subspaces $F$ and $G$ are matrix ranges, e.g.,

$$F = \mathsf{ran}(A), \qquad A \in \mathbb{R}^{n\times p},$$
$$G = \mathsf{ran}(B), \qquad B \in \mathbb{R}^{n\times q}.$$

The principal vectors and angles can be computed using the QR factorization and the SVD. Let $A = Q_A R_A$ and $B = Q_B R_B$ be thin QR factorizations and assume that

$$Q_A^T Q_B \;=\; Y\Sigma Z^T \;=\; \sum_{i=1}^{q} \sigma_i y_i z_i^T$$

is the SVD of $Q_A^T Q_B \in \mathbb{R}^{p\times q}$. Since $\| Q_A^T Q_B \|_2 \le 1$, all the singular values are between 0 and 1 and we may write $\sigma_i = \cos(\theta_i)$, $i = 1{:}q$. Let

$$Q_A Y = [\, f_1 \,|\, \cdots \,|\, f_p \,], \tag{6.4.4}$$
$$Q_B Z = [\, g_1 \,|\, \cdots \,|\, g_q \,] \tag{6.4.5}$$

be column partitionings of the matrices $Q_A Y \in \mathbb{R}^{n\times p}$ and $Q_B Z \in \mathbb{R}^{n\times q}$. These matrices have orthonormal columns. If $f \in F$ and $g \in G$ are unit vectors, then there exist unit vectors $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$ so that $f = Q_A u$ and $g = Q_B v$. Thus,

$$\begin{aligned} f^T g = (Q_A u)^T (Q_B v) \;&=\; u^T (Q_A^T Q_B) v \;=\; u^T (Y\Sigma Z^T) v \\ &= (Y^T u)^T \Sigma (Z^T v) \;=\; \sum_{i=1}^{q} \sigma_i (y_i^T u)(z_i^T v)\,. \end{aligned} \tag{6.4.6}$$

This expression attains its maximal value of $\sigma_1 = \cos(\theta_1)$ by setting $u = y_1$ and $v = z_1$. It follows that $f = Q_A y_1 = f_1$ and $v = Q_B z_1 = g_1$.

Now assume that $k > 1$ and that the first $k-1$ columns of the matrices in (6.4.4) and (6.4.5) are known, i.e., $f_1, \ldots, f_{k-1}$ and $g_1, \ldots, g_{k-1}$. Consider the problem of maximizing $f^T g$ given that $f = Q_A u$ and $g = Q_B v$ are unit vectors that satisfy

$$f^T [\, f_1 \,|\, \cdots \,|\, f_{k-1} \,] = 0,$$
$$g^T [\, g_1 \,|\, \cdots \,|\, g_{k-1} \,] = 0.$$

It follows from (6.4.6) that

$$f^T g \;=\; \sum_{i=k}^{q} \sigma_i (y_i^T u)(z_i^T v) \;\le\; \sigma_k \sum_{i=k}^{q} |y_i^T u| \cdot |z_i^T v|.$$

This expression attains its maximal value of $\sigma_k = \cos(\theta_k)$ by setting $u = y_k$ and $v = z_k$. It follows from (6.4.4) and (6.4.5) that $f = Q_A y_k = f_k$ and $g = Q_B z_k = g_k$. Combining these observations we obtain

**Algorithm 6.4.3 (Principal Angles and Vectors)** Given $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{m\times q}$ $(p \geq q)$ each with linearly independent columns, the following algorithm computes the cosines of the principal angles $\theta_1 \geq \cdots \geq \theta_q$ between $\mathsf{ran}(A)$ and $\mathsf{ran}(B)$. The vectors $f_1, \ldots, f_q$ and $g_1, \ldots, g_q$ are the associated principal vectors.

Compute the thin QR factorizations $A = Q_A R_A$ and $B = Q_B R_B$.
$C = Q_A^T Q_B$
Compute the SVD $Y^T C Z = \mathrm{diag}(\cos(\theta_k))$ .
$Q_A Y(:, 1{:}q) = [\, f_1 \,|\, \cdots \,|\, f_q \,]$
$Q_B Z(:, 1{:}q) = [\, g_1 \,|\, \cdots \,|\, g_q \,]$

The idea of using the SVD to compute the principal angles and vectors is due to Björck and Golub (1973). The problem of rank deficiency in $A$ and $B$ is also treated in this paper. Principal angles and vectors arise in many important statistical applications. The largest principal angle is related to the notion of distance between equidimensional subspaces that we discussed in §2.5.3. If $p = q$, then

$$\mathsf{dist}(F, G) \;=\; \sqrt{1 - \cos(\theta_p)^2} \;=\; \sin(\theta_p).$$

### 6.4.4 Intersection of Subspaces

In light of the following theorem, Algorithm 6.4.3 can also be used to compute an orthonormal basis for $\mathsf{ran}(A) \cap \mathsf{ran}(B)$ where $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{m\times q}$

**Theorem 6.4.2.** *Let $\{\cos(\theta_i)\}_{i=1}^q$ and $\{f_i, g_i\}_{i=1}^q$ be defined by Algorithm 6.4.3. If the index $s$ is defined by $1 = \cos(\theta_1) = \cdots = \cos(\theta_s) > \cos(\theta_{s+1})$, then*

$$\mathsf{ran}(A) \cap \mathsf{ran}(B) \;=\; \mathsf{span}\{f_1, \ldots, f_s\} \;=\; \mathsf{span}\{g_1, \ldots, g_s\}.$$

***Proof.*** The proof follows from the observation that if $\cos(\theta_i) = 1$, then $f_i = g_i$. □

The practical determination of the intersection dimension $s$ requires a definition of what it means for a computed singular value to equal 1. For example, a computed singular value $\hat{\sigma}_i = \cos(\hat{\theta}_i)$ could be regarded as a unit singular value if $\hat{\sigma}_i \geq 1 - \delta$ for some intelligently chosen small parameter $\delta$.

#### Problems

**P6.4.1** Show that if $A$ and $B$ are $m$-by-$p$ matrices, with $p \leq m$, then

$$\min_{Q^T Q = I_p} \| A - BQ \|_F^2 \;=\; \sum_{i=1}^{p} (\sigma_i(A)^2 - 2\sigma_i(B^T A) + \sigma_i(B)^2).$$

**P6.4.2** Extend Algorithm 6.4.2 so that it computes an orthonormal basis for $\mathsf{null}(A_1) \cap \cdots \cap \mathsf{null}(A_s)$ where each matrix $A_i$ has $n$ columns.

**P6.4.3** Extend Algorithm 6.4.3 so that it can handle the case when $A$ and $B$ are rank deficient.

**P6.4.4** Verify Equation (6.4.2).

**P6.4.5** Suppose $A, B \in \mathbb{R}^{m\times n}$ and that $A$ has full column rank. Show how to compute a symmetric matrix $X \in \mathbb{R}^{n\times n}$ that minimizes $\| AX - B \|_F$. Hint: Compute the SVD of $A$.

**P6.4.6** This problem is an exercise in F-norm optimization. (a) Show that if $C \in \mathbb{R}^{m\times n}$ and $e \in \mathbb{R}^m$ is a vector of ones, then $v = C^T e/m$ minimizes $\| C - ev^T \|_F$. (b) Suppose $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$ and that we wish to solve

$$\min_{Q^TQ=I_n\,,\,v\in\mathbb{R}^n} \| A - (B + ev^T)Q \|_F$$

Show that $v_{\rm opt} = (A-B)^T e/m$ and $Q_{\rm opt} = U\Sigma V^T$ solve this problem where $B^T(I-ee^T/m)A = U\Sigma V^T$ is the SVD.

**P6.4.7** A 3-by-3 matrix $H$ is *ROPR matrix* if $H = Q + xy^T$ where $Q \in \mathbb{R}^{3\times 3}$ rotation and $x, y \in \mathbb{R}^3$. (A rotation matrix is an orthogonal matrix with unit determinant. "ROPR" stands for "rank-1 perturbation of a rotation.") ROPR matrices arise in computational photography and this problem highlights some of their properties. (a) If $H$ is a ROPR matrix, then there exist rotations $U, V \in \mathbb{R}^{3\times 3}$, such that $U^THV = \mathrm{diag}(\sigma_1, \sigma_2, \sigma_3)$ satisfies $\sigma_1 \geq \sigma_2 \geq |\sigma_3|$. (b) Show that if $Q \in \mathbb{R}^{3\times 3}$ is a rotation, then there exist cosine-sine pairs $(c_i, s_i) = (\cos(\theta_i), \sin(\theta_i))$, $i = 1{:}3$ such that $Q = Q(\theta_1, \theta_2, \theta_3)$ where

$$Q(\theta_1,\theta_2,\theta_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_1 & s_1 \\ 0 & -s_1 & c_1 \end{bmatrix} \begin{bmatrix} c_2 & s_2 & 0 \\ -s_2 & c_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_3 & s_3 \\ 0 & -s_3 & c_3 \end{bmatrix}$$

$$= \begin{bmatrix} c_2 & s_2c_3 & s_2s_3 \\ -c_1s_2 & c_1c_2c_3 - s_1s_3 & c_1c_2s_3 + s_1c_3 \\ s_1s_2 & -s_1c_2c_3 - c_1s_3 & -s_1c_2s_3 + c_1c_3 \end{bmatrix}.$$

Hint: The Givens QR factorization involves three rotations. (c) Show that if

$$\begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix} = Q(\theta_1,\theta_2,\theta_3) - xy^T, \qquad x, y \in \mathbb{R}^3$$

then $xy^T$ must have the form

$$xy^T = \begin{bmatrix} s_2 \\ \mu c_1 \\ -\mu s_1 \end{bmatrix} \begin{bmatrix} -s_2/\mu \\ c_3 \\ s_3 \end{bmatrix}^T$$

for some $\mu \geq 0$ and

$$\begin{bmatrix} c_2 - \mu & 1 \\ 1 & c_2 - \mu \end{bmatrix} \begin{bmatrix} c_1s_3 \\ s_1c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

(d) Show that the second singular value of a ROPR matrix is 1.

**P6.4.8** Let $U_* \in \mathbb{R}^{n\times d}$ be a matrix with orthonormal columns whose span is a subspace $S$ that we wish to estimate. Assume that $U_c \in \mathbb{R}^{n\times d}$ is a given matrix with orthonormal columns and regard $\mathsf{ran}(U_c)$ as the "current" estimate of $S$. This problem examines what is required to get an improved estimate of $S$ given the availability of a vector $v \in S$. (a) Define the vectors

$$w = U_c^T v, \qquad v_1 = U_cU_c^T v, \qquad v_2 = (I_n - U_cU_c^T)v,$$

and assume that each is nonzero. (a) Show that if

$$z_\theta = \left(\frac{\cos(\theta) - 1}{\| v_1 \| \| w \|}\right) v_1 + \left(\frac{\sin(\theta)}{\| v_2 \| \| w \|}\right) v_2$$

and

$$U_\theta = (I_n + z_\theta v^T)U_c,$$

then $U_\theta^T U_\theta = I_d$. Thus, $U_\theta U_\theta^T$ is an orthogonal projection. (b) Define the distance function

$$\mathsf{dist}_F(\mathsf{ran}(V), \mathsf{ran}(W)) = \| VV^T - WW^T \|_F$$

where $V, W \in \mathbb{R}^{n\times d}$ have orthonormal columns and show

$$\mathsf{dist}_F(\mathsf{ran}(V),\mathsf{ran}(W))^2 \;=\; 2(d - \| W^TV \|_F^2) \;=\; 2\sum_{i=1}^{d}(1-\sigma_i(W^TV)^2).$$

Note that $\mathsf{dist}(\mathsf{ran}(V),\mathsf{ran}(W))^2 \;=\; 1-\sigma_1(W^TV)^2$. (c) Show that

$$d_\theta^2 \;=\; d_c^2 \;-\; 2\cdot\mathsf{tr}(U_*U_*^T(U_\theta U_\theta^T - U_cU_c^T))$$

where $d_\theta = \mathsf{dist}_F(\mathsf{ran}(U_*),\mathsf{ran}(U_\theta))$ and $d_c \;=\; \mathsf{dist}_F(\mathsf{ran}(U_*),\mathsf{ran}(U_c))$. (d) Show that if

$$y_\theta \;=\; \cos(\theta)\frac{v_1}{\| v_1 \|} \;+\; \sin(\theta)\frac{v_2}{\| v_2 \|},$$

then

$$U_\theta U_\theta^T - U_cU_c^T \;=\; y_\theta y_\theta^T - \frac{v_1v_1^T}{v_1^Tv_1}$$

and

$$d_\theta^2 \;=\; d_c^2 \;+\; 2\left(\frac{\| U_*^Tv_1 \|_2^2}{\| v_1 \|_2^2} \;-\; \| U_*^Ty_\theta \|_2^2\right).$$

(e) Show that if $\theta$ minimizes this quantity, then

$$\sin(2\theta)\left(\frac{\| P_Sv_2 \|^2}{\| v_2 \|_2^2} \;-\; \frac{\| P_Sv_1 \|^2}{\| v_1 \|_2^2}\right) \;+\; \cos(2\theta)\frac{v_1^TP_Sv_2}{\| v_1 \|_2\| v_2 \|_2} \;=\; 0, \qquad P_S = U_*U_*^T.$$

**Notes and References for §6.4**

References for the Procrustes problem include:


B. Green (1952). "The Orthogonal Approximation of an Oblique Structure in Factor Analysis," *Psychometrika 17,* 429–40.

P. Schonemann (1966). "A Generalized Solution of the Orthogonal Procrustes Problem," *Psychometrika 31,* 1–10.

R.J. Hanson and M.J. Norris (1981). "Analysis of Measurements Based on the Singular Value Decomposition," *SIAM J. Sci. Stat. Comput. 2,* 363–374.

N.J. Higham (1988). "The Symmetric Procrustes Problem," *BIT 28,* 133–43.

H. Park (1991). "A Parallel Algorithm for the Unbalanced Orthogonal Procrustes Problem," *Parallel Comput. 17*, 913–923.

L.E. Andersson and T. Elfving (1997). "A Constrained Procrustes Problem," *SIAM J. Matrix Anal. Applic. 18*, 124–139.

L. Eldén and H. Park (1999). "A Procrustes Problem on the Stiefel Manifold," *Numer. Math. 82*, 599–619.

A.W. Bojanczyk and A. Lutoborski (1999). "The Procrustes Problem for Orthogonal Stiefel Matrices," *SIAM J. Sci. Comput. 21*, 1291–1304.


If $B = I$, then the Procrustes problem amounts to finding the closest orthogonal matrix. This computation is related to the polar decomposition problem that we consider in §9.4.3. Here are some basic references:


Å. Björck and C. Bowie (1971). "An Iterative Algorithm for Computing the Best Estimate of an Orthogonal Matrix," *SIAM J. Numer. Anal. 8,* 358–64.

N.J. Higham (1986). "Computing the Polar Decomposition with Applications," *SIAM J. Sci. Stat. Comput. 7*, 1160–1174.


Using the SVD to solve the angles-between-subspaces problem is discussed in:


Å. Björck and G.H. Golub (1973). "Numerical Methods for Computing Angles Between Linear Subspaces," *Math. Comput. 27,* 579–94.

L.M. Ewerbring and F.T. Luk (1989). "Canonical Correlations and Generalized SVD: Applications and New Algorithms," *J. Comput. Appl. Math. 27*, 37–52.

G.H. Golub and H. Zha (1994). "Perturbation Analysis of the Canonical Correlations of Matrix Pairs," *Lin. Alg. Applic. 210,* 3–28.

Z. Drmac (2000). "On Principal Angles between Subspaces of Euclidean Space," *SIAM J. Matrix Anal. Applic. 22*, 173–194.

A.V. Knyazev and M.E. Argentati (2002). "Principal Angles between Subspaces in an A–Based Scalar Product: Algorithms and Perturbation Estimates," *SIAM J. Sci. Comput. 23*, 2008–2040.

P. Strobach (2008). "Updating the Principal Angle Decomposition," *Numer. Math. 110*, 83–112.

In reduced-rank regression the object is to connect a matrix of signals to a matrix of noisey observations through a matrix that has specified low rank. An svd-based computational procedure that involves principal angles is discussed in:

L. Eldén and B. Savas (2005). "The Maximum Likelihood Estimate in Reduced-Rank Regression," *Num. Lin. Alg. Applic. 12*, 731–741,

The SVD has many roles to play in statistical computation, see:

S.J. Hammarling (1985). "The Singular Value Decomposition in Multivariate Statistics," *ACM SIGNUM Newsletter 20,* 2-25.

An algorithm for computing the rotation and rank-one matrix in P6.4.7 that define a given ROPR matrix is discussed in:

R. Schreiber, Z. Li, and H. Baker (2009). "Robust Software for Computing Camera Motion Parameters," *J. Math. Imaging Vision 33*, 1–9.

For a more details about the estimation problem associated with P6.4.8, see:

L. Balzano, R. Nowak, and B. Recht (2010). "Online Identification and Tracking of Subspaces from Highly Incomplete Information," *Proceedings of the Allerton Conference on Communication, Control, and Computing 2010.*

## 6.5 Updating Matrix Factorizations

In many applications it is necessary to refactor a given matrix $A \in \mathbb{R}^{m\times n}$ after it has undergone a small modification. For example, given that we have the QR factorization of a matrix $A$, we may require the QR factorization of the matrix $\widetilde{A}$ obtained from $A$ by appending a row or column or deleting a row or column. In this section we show that in situations like these, it is much more efficient to "update" $A$'s QR factorization than to generate the required QR factorization of $\widetilde{A}$ from scratch. Givens rotations have a prominent role to play. In addition to discussing various update-QR strategies, we show how to downdate a Cholesky factorization using hyperbolic rotations and how to update a rank-revealing ULV decomposition.

### 6.5.1 Rank-1 Changes

Suppose we have the QR factorization $QR = A \in \mathbb{R}^{n\times n}$ and that we need to compute the QR factorization $\widetilde{A} = A + uv^T = Q_1R_1$ where $u, v \in \mathbb{R}^n$ are given. Observe that

$$\widetilde{A} \;=\; A + uv^T \;=\; Q(R + wv^T) \tag{6.5.1}$$

where $w = Q^Tu$. Suppose rotations $J_{n-1}, \ldots, J_2, J_1$ are computed such that

$$J_1^T \cdots J_{n-1}^T w \;=\; \pm \| \, w \, \|_2 \, e_1.$$

where each $J_k$ is a Givens rotation in planes $k$ and $k+1$. If these same rotations are applied to $R$, then

$$H = J_1^T \cdots J_{n-1}^T R \tag{6.5.2}$$

is upper Hessenberg. For example, in the $n = 4$ case we start with

$$w \leftarrow \begin{bmatrix} \times \\ \times \\ \times \\ \times \end{bmatrix}, \qquad R \leftarrow \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix},$$

and then update as follows:

$$w \leftarrow J_3^T w = \begin{bmatrix} \times \\ \times \\ \times \\ 0 \end{bmatrix}, \qquad R \leftarrow J_3^T R = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix},$$

$$w \leftarrow J_2^T w = \begin{bmatrix} \times \\ \times \\ 0 \\ 0 \end{bmatrix}, \qquad R \leftarrow J_2^T R = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix},$$

$$w \leftarrow J_1^T w = \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad H \leftarrow J_1^T R = \begin{bmatrix} \times & \times & \times & \times \\ \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}.$$

Consequently,

$$(J_1^T \cdots J_{n-1}^T)(R + wv^T) = H \pm \| w \|_2 e_1 v^T = H_1 \tag{6.5.3}$$

is also upper Hessenberg. Following Algorithm 5.2.4, we compute Givens rotations $G_k$, $k = 1{:}n-1$ such that $G_{n-1}^T \cdots G_1^T H_1 = R_1$ is upper triangular. Combining everything we obtain the QR factorization $\widetilde{A} = A + uv^T = Q_1 R_1$ where

$$Q_1 = Q J_{n-1} \cdots J_1 G_1 \cdots G_{n-1}.$$

A careful assessment of the work reveals that about $26n^2$ flops are required.

The technique readily extends to the case when $A$ is rectangular. It can also be generalized to compute the QR factorization of $A + UV^T$ where $U \in \mathbb{R}^{m\times p}$ and $V \in \mathbb{R}^{n\times p}$.

### 6.5.2 Appending or Deleting a Column

Assume that we have the QR factorization

$$QR = A = [\, a_1 \,|\, \cdots \,|\, a_n \,], \qquad a_i \in \mathbb{R}^m, \tag{6.5.4}$$

and for some $k$, $1 \le k \le n$, partition the upper triangular matrix $R \in \mathbb{R}^{m\times n}$ as follows:

$$R = \begin{bmatrix} R_{11} & v & R_{13} \\ 0 & r_{kk} & w^T \\ 0 & 0 & R_{33} \end{bmatrix} \begin{matrix} k-1 \\ 1 \\ m-k \end{matrix}.$$

$$\begin{matrix} k-1 & 1 & n-k \end{matrix}$$

Now suppose that we want to compute the QR factorization of

$$\widetilde{A} = [\, a_1 \,|\, \cdots \,|\, a_{k-1} \,|\, a_{k+1} \,|\, \cdots \,|\, a_n \,] \in \mathbb{R}^{m\times(n-1)} .$$

Note that $\widetilde{A}$ is just $A$ with its $k$th column deleted and that

$$Q^T\widetilde{A} = \begin{bmatrix} R_{11} & R_{13} \\ 0 & w^T \\ 0 & R_{33} \end{bmatrix} = H$$

is upper Hessenberg, e.g.,

$$H = \begin{bmatrix} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad m = 7,\ n = 6,\ k = 3.$$

Clearly, the unwanted subdiagonal elements $h_{k+1,k}, \ldots, h_{n,n-1}$ can be zeroed by a sequence of Givens rotations: $G_{n-1}^T \cdots G_k^T H = R_1$. Here, $G_i$ is a rotation in planes $i$ and $i+1$ for $i = k{:}n-1$. Thus, if $Q_1 = QG_k \cdots G_{n-1}$ then $\widetilde{A} = Q_1 R_1$ is the QR factorization of $\widetilde{A}$.

The above update procedure can be executed in $O(n^2)$ flops and is very useful in certain least squares problems. For example, one may wish to examine the significance of the $k$th factor in the underlying model by deleting the $k$th column of the corresponding data matrix and solving the resulting LS problem.

Analogously, it is possible to update efficiently the QR factorization of a matrix after a column has been added. Assume that we have (6.5.4) but now want the QR factorization of

$$\widetilde{A} = [\, a_1 \,|\, \ldots \,|\, a_k \,|\, z \,|\, a_{k+1} \,|\, \ldots \,|\, a_n \,]$$

where $z \in \mathbb{R}^m$ is given. Note that if $w = Q^T z$ then

$$Q^T\widetilde{A} = \left[\, Q^T a_1 \,|\, \cdots \,|\, Q^T a_k \,|\, w \,|\, Q^T a_{k+1} \,|\, \cdots \,|\, Q^T a_n \,\right]$$

is upper triangular except for the presence of a "spike" in its $(k+1)$st column, e.g.,

$$\widetilde{A} \leftarrow Q^T\widetilde{A} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & 0 & \times \\ 0 & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & \times & 0 & 0 \end{bmatrix}, \qquad m = 7,\ n = 5,\ k = 3.$$

It is possible to determine a sequence of Givens rotations that restores the triangular form:

$$\widetilde{A} \leftarrow J_6^T \widetilde{A} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & 0 & \times \\ 0 & 0 & 0 & \times & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \widetilde{A} \leftarrow J_5^T \widetilde{A} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\widetilde{A} \leftarrow J_4^T \widetilde{A} = \begin{bmatrix} \times & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \times \\ 0 & 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This update requires $O(mn)$ flops.

### 6.5.3 Appending or Deleting a Row

Suppose we have the QR factorization $QR = A \in \mathbb{R}^{m\times n}$ and now wish to obtain the QR factorization of

$$\widetilde{A} = \begin{bmatrix} w^T \\ A \end{bmatrix}$$

where $w \in \mathbb{R}^n$. Note that

$$\mathrm{diag}(1, Q^T)\widetilde{A} = \begin{bmatrix} w^T \\ R \end{bmatrix} = H$$

is upper Hessenberg. Thus, rotations $J_1, \ldots, J_n$ can be determined so $J_n^T \cdots J_1^T H = R_1$ is upper triangular. It follows that $\widetilde{A} = Q_1 R_1$ is the desired QR factorization, where $Q_1 = \mathrm{diag}(1, Q)J_1 \cdots J_n$. See Algorithm 5.2.5.

No essential complications result if the new row is added between rows $k$ and $k+1$ of $A$. Indeed, if

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = QR, \qquad A_1 \in \mathbb{R}^{k\times n},\ A_2 \in \mathbb{R}^{(m-k)\times n},$$

and

$$P = \begin{bmatrix} 0 & 1 & 0 \\ I_k & 0 & 0 \\ 0 & 0 & I_{m-k} \end{bmatrix},$$

then

$$\mathrm{diag}(1, Q^T)P \begin{bmatrix} A_1 \\ w^T \\ A_2 \end{bmatrix} = \begin{bmatrix} w^T \\ R \end{bmatrix} = H$$

is upper Hessenberg and we proceed as before.

Lastly, we consider how to update the QR factorization $QR = A \in \mathbb{R}^{m\times n}$ when the first row of $A$ is deleted. In particular, we wish to compute the QR factorization of the submatrix $A_1$ in

$$A = \begin{bmatrix} z^T \\ A_1 \end{bmatrix} \begin{matrix} 1 \\ m-1 \end{matrix}.$$

(The procedure is similar when an arbitrary row is deleted.) Let $q^T$ be the first row of $Q$ and compute Givens rotations $G_1, \ldots, G_{m-1}$ such that $G_1^T \cdots G_{m-1}^T q = \alpha e_1$ where $\alpha = \pm 1$. Note that

$$H = G_1^T \cdots G_{m-1}^T R = \begin{bmatrix} v^T \\ R_1 \end{bmatrix} \begin{matrix} 1 \\ m-1 \end{matrix}$$

is upper Hessenberg and that

$$QG_{m-1} \cdots G_1 = \begin{bmatrix} \alpha & 0 \\ 0 & Q_1 \end{bmatrix}$$

where $Q_1 \in \mathbb{R}^{(m-1)\times(m-1)}$ is orthogonal. Thus,

$$A = \begin{bmatrix} z^T \\ A_1 \end{bmatrix} = (QG_{m-1} \cdots G_1)(G_1^T \cdots G_{m-1}^T R) = \begin{bmatrix} \alpha & 0 \\ 0 & Q_1 \end{bmatrix} \begin{bmatrix} v^T \\ R_1 \end{bmatrix}$$

from which we conclude that $A_1 = Q_1 R_1$ is the desired QR factorization.

### 6.5.4 Cholesky Updating and Downdating

Suppose we are given a symmtetric positive definite matrix $A \in \mathbb{R}^{n\times n}$ and its Cholesky factor $G$. In the *Cholesky updating problem*, the challenge is to compute the Cholesky factorization $\widetilde{A} = \widetilde{G}\widetilde{G}^T$ where

$$\widetilde{A} = A + zz^T, \qquad z \in \mathbb{R}^n. \tag{6.5.5}$$

Noting that

$$\widetilde{A} = \begin{bmatrix} G^T \\ z^T \end{bmatrix}^T \begin{bmatrix} G^T \\ z^T \end{bmatrix}, \tag{6.5.6}$$

we can solve this problem by computing a product of Givens rotations $Q = Q_1 \cdots Q_n$ so that

$$Q^T \begin{bmatrix} G^T \\ z^T \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad R \in \mathbb{R}^{n\times n} \tag{6.5.7}$$

is upper triangular. It follows that $\widetilde{A} = RR^T$ and so the updated Cholesky factor is given by $\widetilde{G} = R^T$. The zeroing sequence that produces $R$ is straight forward, e.g.,

$$\begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{Q_1} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{Q_2} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{Q_3} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & 0 \end{bmatrix}.$$

The $Q_k$ update involves only rows $k$ and $n+1$. The overall process is essentially the same as the strategy we outlined in the previous subsection for updating the QR factorization of a matrix when a row is appended.

The *Cholesky downdating problem* involves a different set of tools and a new set of numerical concerns. We are again given a Cholesky factorization $A = GG^T$ and a vector $z \in \mathbb{R}^n$. However, now the challenge is to compute the Cholesky factorization $\widetilde{A} = \widetilde{G}\widetilde{G}^T$ where

$$\widetilde{A} = A - zz^T \tag{6.5.8}$$

is presumed to be positive definite. By introducing the notion of a *hyperbolic rotation* we can develop a downdating framework that corresponds to the Givens-based updating framework. Define the matrix $S$ as follows

$$S = \begin{bmatrix} I_n & 0 \\ 0 & -1 \end{bmatrix} \tag{6.5.9}$$

and note that

$$\widetilde{A} = GG^T - zz^T = \begin{bmatrix} G^T \\ z^T \end{bmatrix}^T S \begin{bmatrix} G^T \\ z^T \end{bmatrix}. \tag{6.5.10}$$

This corresponds to (6.5.6), but instead of computing the QR factorization (6.5.7), we seek a matrix $H \in \mathbb{R}^{(n+1)\times(n+1)}$ that satisfies two properties:

$$HSH^T = S, \tag{6.5.11}$$

$$H^T \begin{bmatrix} G^T \\ z^T \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad R \in \mathbb{R}^{n\times n} \text{ (upper triangular)}. \tag{6.5.12}$$

If this can be accomplished, then it follows from

$$\widetilde{A} = \left( H^T \begin{bmatrix} G^T \\ z^T \end{bmatrix} \right)^T \begin{bmatrix} I_n & 0 \\ 0 & -1 \end{bmatrix} \left( H^T \begin{bmatrix} G^T \\ z^T \end{bmatrix} \right) = R^T R$$

that the Cholesky factor of $\widetilde{A} = A - zz^T$ is given by $\widetilde{G} = R^T$. A matrix $H$ that satisfies (6.5.11) is said to be *$S$-orthogonal*. Note that the product of $S$-orthogonal matrices is also $S$-orthogonal.

An important subset of the $S$-orthogonal matrices are the *hyperbolic rotations* and here is a 4-by-4 example:

$$H_2(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c & 0 & -s \\ 0 & 0 & 1 & 0 \\ 0 & -s & 0 & c \end{bmatrix}, \qquad c = \cosh(\theta)\,,\ s = \sinh(\theta).$$

The $S$-orthogonality of this matrix follows from $\cosh(\theta)^2 - \sinh(\theta)^2 = 1$. In general, $H_k \in \mathbb{R}^{(n+1)\times(n+1)}$ is a hyperbolic rotation if it agrees with $I_{n+1}$ except in four locations:

$$\begin{bmatrix} [H_k]_{k,k} & [H_k]_{k,n+1} \\ [H_k]_{n+1,k} & [H_k]_{n+1,n+1} \end{bmatrix} = \begin{bmatrix} \cosh(\theta) & -\sinh(\theta) \\ -\sinh(\theta) & \cosh(\theta) \end{bmatrix}.$$

Hyperbolic rotations look like Givens rotations and, not surprisingly, can be used to introduce zeros into a vector or matrix. However, upon consideration of the equation

$$\begin{bmatrix} c & -s \\ -s & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}, \qquad c^2 - s^2 = 1$$

we see that the required cosh-sinh pair may not exist. Since we always have $|\cosh(\theta)| > |\sinh(\theta)|$, there is no real solution to $-sx_1 + cx_2 = 0$ if $|x_2| > |x_1|$. On the other hand, if $|x_1| > |x_2|$, then $\{c, s\} = \{\cosh(\theta), \sinh(\theta)\}$ can be computed as follows:

$$\tau = \frac{x_2}{x_1}, \qquad c = \frac{1}{\sqrt{1-\tau^2}}, \qquad s = c \cdot \tau. \tag{6.5.13}$$

There are clearly numerical issues if $|x_1|$ is just slightly greater than $|x_2|$. However, it is possible to organize hyperbolic rotation computations successfully, see Alexander, Pan, and Plemmons (1988).

Putting these concerns aside, we show how the matrix $H$ in (6.5.12) can be computed as a product of hyperbolic rotations $H = H_1 \cdots H_n$ just as the transforming $Q$ in the updating problem is a product of Givens rotations. Consider the role of $H_1$ in the $n = 3$ case:

$$\begin{bmatrix} c & 0 & 0 & -s \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -s & 0 & 0 & c \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{21} & g_{31} \\ 0 & g_{22} & g_{32} \\ 0 & 0 & g_{33} \\ z_1 & z_2 & z_3 \end{bmatrix} = \begin{bmatrix} \tilde{g}_{11} & \tilde{g}_{21} & \tilde{g}_{31} \\ 0 & g_{22} & g_{32} \\ 0 & 0 & g_{33} \\ 0 & z_2' & z_3' \end{bmatrix}.$$

Since $\widetilde{A} = GG^T - zz^T$ is positive definite, $[\widetilde{A}]_{11} = g_{11}^2 - z_1^2 > 0$. It follows that $|g_{11}| > |z_1|$ which guarantees that the cosh-sinh computations (6.5.13) go through. For the overall process to be defined, we have to guarantee that hyperbolic rotations $H_2, \ldots, H_n$ can be found to zero out the bottom row in the matrix $[\, G^T \;\; z \,]^T$. The following theorem ensures that this is the case.

**Theorem 6.5.1.** *If*

$$A = \begin{bmatrix} \alpha & v^T \\ v & B \end{bmatrix} = \begin{bmatrix} g_{11} & 0 \\ g_1 & G_1 \end{bmatrix} \begin{bmatrix} g_{11} & g_1^T \\ 0 & G_1^T \end{bmatrix}$$

*and*

$$\widetilde{A} = A - zz^T = A - \begin{bmatrix} \mu \\ w \end{bmatrix} \begin{bmatrix} \mu \\ w \end{bmatrix}^T$$

*are positive definite, then it is possible to determine $c = \cosh(\theta)$ and $s = \sinh(\theta)$ so*

$$\begin{bmatrix} c & 0 & -s \\ 0 & I_{n-1} & 0 \\ -s & 0 & c \end{bmatrix} \begin{bmatrix} g_{11} & g_1^T \\ 0 & G_1^T \\ \mu & w^T \end{bmatrix} = \begin{bmatrix} \tilde{g}_{11} & \tilde{g}_1^T \\ 0 & G_1^T \\ 0 & w_1^T \end{bmatrix}.$$

*Moreover, the matrix $\widetilde{A}_1 = G_1 G_1^T - w_1 w_1^T$ is positive definite.*

***Proof.*** The blocks in $A$'s Cholesky factor are given by

$$g_{11} = \sqrt{\alpha}, \qquad g_1 = v/g_{11}, \qquad G_1 G_1^T = B - \frac{1}{\alpha} v v^T. \tag{6.5.14}$$

Since $A - zz^T$ is positive definite, $a_{11} - z_1^2 = g_{11}^2 - \mu^2 > 0$ and so from (6.5.13) with $\tau = \mu/g_{11}$ we see that

$$c = \frac{\sqrt{\alpha}}{\sqrt{\alpha - \mu^2}}, \qquad s = \frac{\mu}{\sqrt{\alpha - \mu^2}}. \tag{6.5.15}$$

Since $w_1 = -sg_1 + c\,w$ it follows from (6.5.14) and (6.5.15) that

$$\begin{aligned} \widetilde{A}_1 = G_1 G_1^T - w_1 w_1^T &= B - \frac{1}{\alpha} v v^T - (-sg_1 + cw)(-sg_1 + cw)^T \\ &= B - \frac{c^2}{\alpha} v v^T - c^2 w w^T + \frac{sc}{\sqrt{\alpha}}(v w^T + w v^T) \\ &= B - \frac{1}{\alpha - \mu^2} v v^T - \frac{\alpha}{\alpha - \mu^2} w w^T + \frac{\mu}{\alpha - \mu^2}(v w^T + w v^T). \end{aligned}$$

It is easy to verify that this matrix is precisely the Schur complement of $\alpha$ in

$$\widetilde{A} = A - zz^T = \begin{bmatrix} \alpha - \mu^2 & v^T - \mu w^T \\ v - \mu w & B - w w^T \end{bmatrix}$$

and is therefore positive definite. □

The theorem provides the key step in an induction proof that the factorization (6.5.12) exists.

### 6.5.5 Updating a Rank-Revealing ULV Decomposition

We close with a discussion about updating a nullspace basis after one or more rows have been appended to the underlying matrix. We work with the ULV decomposition which is much more tractable than the SVD from the updating point of view. We pattern our remarks after Stewart(1993).

A *rank -revealing ULV decomposition* of a matrix $A \in \mathbb{R}^{m\times n}$ has the form

$$U^T A V = \begin{bmatrix} L \\ 0 \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \\ 0 & 0 \end{bmatrix}, \qquad U^T U = I_m,\ V^T V = I_n \tag{6.5.16}$$

where $L_{11} \in \mathbb{R}^{r\times r}$ and $L_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$ are lower triangular and $\| L_{21} \|_2$ and $\| L_{22} \|_2$ are small compared to $\sigma_{\min}(L_{11})$. Such a decomposition can be obtained by applying QR with column pivoting

$$U^T A \Pi = \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad R \in \mathbb{R}^{n\times n}$$

followed by a QR factorization $V_1^T R^T = L^T$. In this case the matrix $V$ in (6.5.16) is given by $V = \Pi V_1$. The parameter $r$ is the estimated rank. Note that if

$$V = \underset{r}{[\, V_1 \,|\,} \underset{n-r}{V_2\,]}, \qquad U = \underset{r}{[\, U_1 \,|\,} \underset{m-r}{U_2\,]},$$

then the columns of $V_2$ define an approximate nullspace:

$$\| AV_2 \|_2 = \| U_2 L_{22} \|_2 = \| L_{22} \|_2.$$

Our goal is to produce cheaply a rank-revealing ULV decomposition for the row-appended matrix

$$\tilde{A} = \begin{bmatrix} A \\ z^T \end{bmatrix},$$

In particular, we show how to revise $L$, $V$, and possibly $r$ in $O(n^2)$ flops. Note that

$$\begin{bmatrix} U & 0 \\ 0 & 1 \end{bmatrix}^T \begin{bmatrix} A \\ z^T \end{bmatrix} V = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \\ 0 & 0 \\ w^T & y^T \end{bmatrix}.$$

We illustrate the key ideas through an example. Suppose $n = 7$ and $r = 4$. By permuting the rows so that the bottom row is just underneath $L$, we obtain

$$\begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \\ w^T & y^T \end{bmatrix} = \left[\begin{array}{ccccccc} \ell & 0 & 0 & 0 & 0 & 0 & 0 \\ \ell & \ell & 0 & 0 & 0 & 0 & 0 \\ \ell & \ell & \ell & 0 & 0 & 0 & 0 \\ \ell & \ell & \ell & \ell & 0 & 0 & 0 \\ \hline \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & 0 & 0 \\ \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & 0 \\ \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon \\ \hline w & w & w & w & y & y & y \end{array}\right].$$

The $\epsilon$ entries are small while the $\ell$, $w$, and $y$ entries are not. Next, a sequence of Givens rotations $G_7, \ldots, G_1$ are applied from the left to zero out the bottom row:

$$\begin{bmatrix} \tilde{L} \\ \hline 0 \end{bmatrix} = \left[\begin{array}{ccccccc} \times & 0 & 0 & 0 & 0 & 0 & 0 \\ \times & \times & 0 & 0 & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & \times & \times & 0 \\ \times & \times & \times & \times & \times & \times & \times \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] = G_{17} \cdots G_{57} G_{67} \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \\ w^T & y^T \end{bmatrix}.$$

Because this zeroing process intermingles the (presumably large) entries of the bottom row with the entries from each of the other rows, the lower triangular form is typically *not* rank revealing. However, and this is key, we can restore the rank-revealing structure with a combination of condition estimation and Givens zero chasing.

Let us assume that with the added row, the new nullspace has dimension 2. With a reliable condition estimator we produce a unit 2-norm vector $p$ such that

$$\| p^T \widetilde{L} \|_2 \approx \sigma_{\min}(\widetilde{L}).$$

(See §3.5.4). Rotations $\{U_{i,i+1}\}_{i=1}^{6}$ can be found such that

$$U_{67}^T U_{56}^T U_{45}^T U_{34}^T U_{23}^T U_{12}^T p \;=\; e_7 = I_7(:,7).$$

Applying these rotations to $\widetilde{L}$ produces a lower Hessenberg matrix

$$H \;=\; U_{67}^T U_{56}^T U_{45}^T U_{34}^T U_{23}^T U_{12}^T \tilde{L}.$$

Applying more rotations from the right restores $H$ to a lower triangular form:

$$L_+ \;=\; H V_{12} V_{23} V_{34} V_{45} V_{56} V_{67}.$$

It follows that

$$e_7^T L_+ = \left(e_8^T H\right) V_{12} V_{23} V_{34} V_{45} V_{56} V_{67} \;=\; \left(p^T \tilde{L}\right) V_{12} V_{23} V_{34} V_{45} V_{56} V_{67}$$

has approximate norm $\sigma_{\min}(\widetilde{L})$. Thus, we obtain a lower triangular matrix of the form

$$L_+ \;=\; \left[\begin{array}{cccccc|c} \times & 0 & 0 & 0 & 0 & 0 & 0 \\ \times & \times & 0 & 0 & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & \times & \times & \times & 0 \\ \hline \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon \end{array}\right]$$

We can repeat the condition estimation and zero chasing on the leading 6-by-6 portion. Assuming that the nullspace of the augmented matrix has dimension two, this produces another row of small numbers:

$$\left[\begin{array}{ccccc|cc} \times & 0 & 0 & 0 & 0 & 0 & 0 \\ \times & \times & 0 & 0 & 0 & 0 & 0 \\ \times & \times & \times & 0 & 0 & 0 & 0 \\ \times & \times & \times & \times & 0 & 0 & 0 \\ \times & \times & \times & \times & \times & 0 & 0 \\ \hline \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & 0 \\ \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon & \epsilon \end{array}\right].$$

This illustrates how we can restore any lower triangular matrix to rank-revealing form.

**Problems**

**P6.5.1** Suppose we have the QR factorization for $A \in \mathbb{R}^{m\times n}$ and now wish to solve

$$\min_{x\in\mathbf{R}^n} \| (A + uv^T)x - b \|_2$$

where $u, b \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ are given. Give an algorithm for solving this problem that requires $O(mn)$ flops. Assume that $Q$ must be updated.

**P6.5.2** Suppose

$$A = \begin{bmatrix} c^T \\ B \end{bmatrix}, \qquad c \in \mathbb{R}^n,\ B \in \mathbb{R}^{(m-1)\times n}$$

has full column rank and $m > n$. Using the Sherman-Morrison-Woodbury formula show that

$$\frac{1}{\sigma_{\min}(B)} \leq \frac{1}{\sigma_{\min}(A)} + \frac{\| (A^TA)^{-1}c \|_2^2}{1 - c^T(A^TA)^{-1}c}.$$

**P6.5.3** As a function of $x_1$ and $x_2$, what is the 2-norm of the hyperbolic rotation produced by (6.5.13)?

**P6.5.4** Assume that

$$A = \begin{bmatrix} R & H \\ 0 & E \end{bmatrix}, \qquad \rho = \frac{\| E \|_2}{\sigma_{\min}(R)} < 1,$$

where $R$ and $E$ are square. Show that if

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

is orthogonal and

$$\begin{bmatrix} R & H \\ 0 & E \end{bmatrix} \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} = \begin{bmatrix} R_1 & 0 \\ H_1 & E_1 \end{bmatrix},$$

then $\| H_1 \|_2 \leq \rho \| H \|_2$.

**P6.5.5** Suppose $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ with $m \geq n$. In the *indefinite least squares* (ILS) problem, the goal is to minimize

$$\phi(x) = (b - Ax)^T J (b - Ax),$$

where

$$S = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}, \qquad p + q = m.$$

It is assumed that $p \geq 1$ and $q \geq 1$. (a) By taking the gradient of $\phi$, show that the ILS problem has a unique solution if and only if $A^TSA$ is positive definite. (b) Assume that the ILS problem has a unique solution. Show how it can be found by computing the Cholesky factorization of $Q_1^TQ_1 - Q_2^TQ_2$ where

$$A = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}, \qquad Q_1 \in \mathbb{R}^{p\times n},\ Q_2 \in \mathbb{R}^{q\times n}$$

is the thin QR factorization. (c) A matrix $Q \in \mathbb{R}^{m\times m}$ is *S-orthogonal* if $QSQ^T = S$ If

$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{matrix} p \\ q \end{matrix}$$

$$\qquad\quad p \qquad q$$

is $S$-orthogonal, then by comparing blocks in the equation $Q^TSQ = S$ we have

$$Q_{11}^TQ_{11} = I_p + Q_{21}^TQ_{21}, \qquad Q_{11}^TQ_{12} = Q_{21}^TQ_{22}, \qquad Q_{22}^TQ_{22} = I_q + Q_{12}^TQ_{12}.$$

Thus, the singular values of $Q_{11}$ and $Q_{22}$ are never smaller than 1. Assume that $p \geq q$. By analogy with how the CS decomposition is established in §2.5.4, show that there exist orthogonal matrices $U_1$, $U_2$, $V_1$ and $V_2$ such that

$$\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^T Q \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix} = \left[\begin{array}{ccc} D & 0 & (D^2 - I)^{1/2} \\ 0 & I_{p-q} & 0 \\ \hline (D^2 - I_p)^{1/2} & 0 & D \end{array}\right]$$

where $D = \mathrm{diag}(d_1, \ldots, d_p)$ with $d_i \geq 1$, $i = 1{:}p$. This is the *hyperbolic CS decomposition* and details can be found in Stewart and Van Dooren (2006).

**Notes and References for §6.5**

The seminal matrix factorization update paper is:

P.E. Gill, G.H. Golub, W. Murray, and M.A. Saunders (1974). "Methods for Modifying Matrix Factorizations," *Math. Comput. 28,* 505–535.

Initial research into the factorization update problem was prompted by the development of quasi-Newton methods and the simplex method for linear programming. In these venues, a linear system must be solved in step $k$ that is a low-rank perturbation of the linear system solved in step $k-1$, see:

R.H. Bartels (1971). "A Stabilization of the Simplex Method," *Numer. Math. 16,* 414–434.
P.E. Gill, W. Murray, and M.A. Saunders (1975). "Methods for Computing and Modifying the LDV Factors of a Matrix," *Math. Comput. 29,* 1051–1077.
D. Goldfarb (1976). "Factored Variable Metric Methods for Unconstrained Optimization," *Math. Comput. 30,* 796–811.
J.E. Dennis and R.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall, Englewood Cliffs, NJ.
W.W. Hager (1989). "Updating the Inverse of a Matrix," *SIAM Review 31,* 221–239.
S.K. Eldersveld and M.A. Saunders (1992). "A Block-LU Update for Large-Scale Linear Programming," *SIAM J. Matrix Anal. Applic. 13*, 191–201.

Updating issues in the least squares setting are discussed in:

J. Daniel, W.B. Gragg, L. Kaufman, and G.W. Stewart (1976). "Reorthogonaization and Stable Algorithms for Updating the Gram-Schmidt QR Factorization," *Math. Comput. 30,* 772–795.
S. Qiao (1988). "Recursive Least Squares Algorithm for Linear Prediction Problems," *SIAM J. Matrix Anal. Applic. 9*, 323–328.
Å. Björck, H. Park, and L. Eldén (1994). "Accurate Downdating of Least Squares Solutions," *SIAM J. Matrix Anal. Applic. 15*, 549–568.
S.J. Olszanskyj, J.M. Lebak, and A.W. Bojanczyk (1994). "Rank-k Modification Methods for Recursive Least Squares Problems," *Numer. Alg. 7,* 325–354.
L. Eldén and H. Park (1994). "Block Downdating of Least Squares Solutions," *SIAM J. Matrix Anal. Applic. 15*, 1018–1034.

Kalman filtering is a very important tool for estimating the state of a linear dynamic system in the presence of noise. An illuminating, stable implementation that involves updating the QR factorization of an evolving block banded matrix is given in:

C.C. Paige and M.A. Saunders (1977). "Least Squares Estimation of Discrete Linear Dynamic Systems Using Orthogonal Transformations," *SIAM J. Numer. Anal. 14*, 180–193.

The Cholesky downdating literature includes:

G.W. Stewart (1979). "The Effects of Rounding Error on an Algorithm for Downdating a Cholesky Factorization," *J. Inst. Math. Applic. 23,* 203–213.
A.W. Bojanczyk, R.P. Brent, P. Van Dooren, and F.R. de Hoog (1987). "A Note on Downdating the Cholesky Factorization," *SIAM J. Sci. Stat. Comput. 8,* 210–221.
C.-T. Pan (1993). "A Perturbation Analysis of the Problem of Downdating a Cholesky Factorization," *Lin. Alg. Applic. 183*, 103–115.
L. Eldén and H. Park (1994). "Perturbation Analysis for Block Downdating of a Cholesky Decomposition," *Numer. Math. 68*, 457–468.
M.R. Osborne and L. Sun (1999). "A New Approach to Symmetric Rank-One Updating," *IMA J. Numer. Anal. 19*, 497–507.
E.S. Quintana-Orti and R.A. Van Geijn (2008). "Updating an LU Factorization with Pivoting," *ACM Trans. Math. Softw. 35(2)*, Article 11.

Hyperbolic tranformations have been successfully used in a number of settings:

G.H. Golub (1969). "Matrix Decompositions and Statistical Computation," in *Statistical Computation*, ed., R.C. Milton and J.A. Nelder, Academic Press, New York, pp. 365–397.
C.M. Rader and A.O. Steinhardt (1988). "Hyperbolic Householder Transforms," *SIAM J. Matrix Anal. Applic. 9*, 269–290.

S.T. Alexander, C.T. Pan, and R.J. Plemmons (1988). "Analysis of a Recursive Least Squares Hyperbolic Rotation Algorithm for Signal Processing," *Lin. Alg. and Its Applic. 98,* 3–40.
G. Cybenko and M. Berry (1990). "Hyperbolic Householder Algorithms for Factoring Structured Matrices," *SIAM J. Matrix Anal. Applic. 11*, 499–520.
A.W. Bojanczyk, R. Onn, and A.O. Steinhardt (1993). "Existence of the Hyperbolic Singular Value Decomposition," *Lin. Alg. Applic. 185*, 21–30.
S. Chandrasekaran, M. Gu, and A.H. Sayad (1998). "A Stable and Efficient Algorithm for the Indefinite Linear Least Squares Problem," *SIAM J. Matrix Anal. Applic. 20*, 354–362.
A.J. Bojanczyk, N.J. Higham, and H. Patel (2003a). "Solving the Indefinite Least Squares Problem by Hyperbolic QR Factorization," *SIAM J. Matrix Anal. Applic. 24*, 914–931.
A. Bojanczyk, N.J. Higham, and H. Patel (2003b). "The Equality Constrained Indefinite Least Squares Problem: Theory and Algorithms," *BIT 43*, 505–517.
M. Stewart and P. Van Dooren (2006). "On the Factorization of Hyperbolic and Unitary Transformations into Rotations," *SIAM J. Matrix Anal. Applic. 27*, 876–890.
N.J. Higham (2003). "J-Orthogonal Matrices: Properties and Generation," *SIAM Review 45*, 504–519.

High-performance issues associated with QR updating are discussed in:

B.C. Gunter and R.A. Van De Geijn (2005). "Parallel Out-of-Core Computation and Updating of the QR Factorization," *ACM Trans. Math. Softw. 31*, 60–78.

Updating and downdating the ULV and URV decompositions and related topics are covered in:

C.H. Bischof and G.M. Shroff (1992). "On Updating Signal Subspaces," *IEEE Trans. Signal Proc. 40*, 96–105.
G.W. Stewart (1992). "An Updating Algorithm for Subspace Tracking," *IEEE Trans. Signal Proc. 40*, 1535–1541.
G.W. Stewart (1993). "Updating a Rank-Revealing ULV Decomposition," *SIAM J. Matrix Anal. Applic. 14*, 494–499.
G.W. Stewart (1994). "Updating URV Decompositions in Parallel," *Parallel Comp. 20*, 151–172.
H. Park and L. Eldén (1995). "Downdating the Rank-Revealing URV Decomposition," *SIAM J. Matrix Anal. Applic. 16*, 138–155.
J.L. Barlow and H. Erbay (2009). "Modifiable Low-Rank Approximation of a Matrix," *Num. Lin. Alg. Applic. 16*, 833–860.

Other interesting update-related topics include the updating of condition estimates, see:

W.R. Ferng, G.H. Golub, and R.J. Plemmons (1991). "Adaptive Lanczos Methods for Recursive Condition Estimation," *Numerical Algorithms 1*, 1-20.
G. Shroff and C.H. Bischof (1992). "Adaptive Condition Estimation for Rank-One Updates of QR Factorizations," *SIAM J. Matrix Anal. Applic. 13*, 1264–1278.
D.J. Pierce and R.J. Plemmons (1992). "Fast Adaptive Condition Estimation," *SIAM J. Matrix Anal. Applic. 13*, 274–291.

and the updating of solutions to constrained least squares problems:

K. Schittkowski and J. Stoer (1979). "A Factorization Method for the Solution of Constrained Linear Least Squares Problems Allowing for Subsequent Data changes," *Numer. Math. 31,* 431–463.
Å. Björck (1984). "A General Updating Algorithm for Constrained Linear Least Squares Problems," *SIAM J. Sci. Stat. Comput. 5,* 394–402.

Finally, we mention the following paper concerned with SVD updating:

M. Moonen, P. Van Dooren, and J. Vandewalle (1992). "A Singular Value Decomposition Updating Algorithm," *SIAM J. Matrix Anal. Applic. 13*, 1015–1038.