$$
\begin{aligned}
&\text{Given } A \in \mathbb{C}^{m\times n},\ m \ge n,\ b \in \mathbb{C}^m, \\
&\text{find } x \in \mathbb{C}^n \text{ such that } \|b - Ax\|_2 \text{ is minimized.}
\end{aligned}
\tag{11.2}
$$

SVD

In Lecture 31 we shall describe an algorithm for computing the reduced singular value decomposition $A = \hat{U}\hat{\Sigma}V^*$. This suggests another method for solving least squares problems. Now P is represented in the form $P = \hat{U}\hat{U}^*$, giving

$$y = Pb = \hat{U}\hat{U}^*b, \tag{11.19}$$

and the analogues of (11.16) and (11.17) are

$$\hat{U}\hat{\Sigma}V^*x = \hat{U}\hat{U}^*b \tag{11.20}$$

and

$$\hat{\Sigma}V^*x = \hat{U}^*b. \tag{11.21}$$

(Multiplying by $V\hat{\Sigma}^{-1}$ gives $A^+ = V\hat{\Sigma}^{-1}\hat{U}^*$.) The algorithm looks like this.

Algorithm 11.3. Least Squares via SVD

1. Compute the reduced SVD $A = \hat{U}\hat{\Sigma}V^*$.
2. Compute the vector $\hat{U}^*b$.
3. Solve the diagonal system $\hat{\Sigma}w = \hat{U}^*b$ for w.
4. Set $x = Vw$.

Note that whereas QR factorization reduces the least squares problem to a triangular system of equations, the SVD reduces it to a diagonal system of equations, which is of course trivially solved. If A has full rank, the diagonal system is nonsingular.

As before, (11.21) can be derived from the normal equations. If $A^*Ax = A^*b$, then $V\hat{\Sigma}^*\hat{U}^*\hat{U}\hat{\Sigma}V^*x = V\hat{\Sigma}^*\hat{U}^*b$, implying $\hat{\Sigma}V^*x = \hat{U}^*b$.

The operation count for Algorithm 11.3 is dominated by the computation of the SVD. As we shall see in Lecture 31, for $m \gg n$ this cost is approximately the same as for QR factorization, but for $m \approx n$ the SVD is more expensive. A typical estimate is

$$\text{Work for Algorithm 11.3:} \quad \sim 2mn^2 + 11n^3 \text{ flops}, \tag{11.22}$$

6.6.5. The Singular Value Decomposition

Finally, the minimum-length solution of a least-squares problem can be obtained from the singular value decomposition of A,

$$A = USV^T = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^T,$$

where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and Σ is an $r \times r$ diagonal matrix of positive singular values, $\text{diag}(\sigma_i)$, $i = 1, \ldots, r$.

As in Section 5.6, we designate the first r columns of the orthogonal matrices U and V as U_r and V_r. The rank-retaining form $A = GH$ of the singular value decomposition is

$$A = U_r \Sigma V_r^T, \quad \text{with} \quad G = U_r \;\text{ and }\; H = \Sigma V_r^T. \tag{6.6.9}$$

Substituting in expression (6.6.4) for x_+, the special properties of U_r, Σ and V_r imply that $G^TG = I_r$ and $HH^T = \Sigma^2$. The resulting form for x_+ is

$$x_+ = V_r \Sigma^{-1} U_r^T b.$$

The definition and uniqueness of the pseudoinverse A^+ (Section 5.3.3.1) imply that the minimum-length least-squares solution can be written as

$$x_+ = A^+ b, \quad \text{where} \quad A^+ = V_r \Sigma^{-1} U_r^T, \tag{6.6.10}$$

which is (again) the identical form obtained in the compatible case (see (5.6.1)), this time because the columns of U_r are orthonormal. This formula is particularly useful in cases of near rank-deficiency in which a decision about numerical rank is based on treating certain singular values as negligible (see Section 5.8.2).

To find the minimum-length solution of a least-squares problem involving the *transposed* matrix A^T, recall that the pseudoinverse of A^T is simply the transpose of A^+, i.e., $(A^T)^+ = (A^+)^T$. Since we have imposed no dimensionality restrictions in deriving the expression $x_+ = A^+ b$, it follows from (6.6.10) that the minimum-length solution y_+ of

$$\underset{y \in \Re^m}{\text{minimize}} \; \|c - A^T y\|_2^2 \qquad \text{is} \qquad y_+ = (A^T)^+ c = U_r \Sigma^{-1} V_r^T c.$$