Matrix decompositions.

Have already seen examples.

For square A:

$A = M \Lambda M^{-1}$ for diagonalizable A

$A = U \Lambda U^T$ for symmetric $A \in \mathbb{R}^{n\times n}$

$A = U \Lambda U^*$ for general normal $A \in \mathbb{C}^{n\times n}$

$A = M J M^{-1}$, $J = \begin{pmatrix} J_1 & 0 \\ 0 & J_\ell \end{pmatrix}$ (Jordan form)

For general $m \times n$ A:

$A = QR$ (QR decomposition), $Q^T Q = I \sim n\times n$, and upper triangular R.

Develop additional decompositions.

Singular value decomposition (SVD).

Suppose $A \in \mathbb{R}^{m\times n}$, with no particular relationship between m and n. (Development for $A \in \mathbb{C}^{m\times n}$ similar.)

Form: $A = U \Sigma V^T$, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{m\times n}$ has the form

$$\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & 0 & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & & & \end{pmatrix}, \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0$$

where $r = \text{rank } A$ and all entries not shown are zero.

Terminology: $\sigma_1, \dots, \sigma_r$ are the singular values of A.

The columns of V are right singular vectors of A.

The columns of U are left singular vectors of A.

The singular values of A are unique; the singular vectors are not.

Derivation. A^TA is symmetric, so $\sigma(A) \subseteq \mathbb{R}$.
Also, $\langle v, A^TAv\rangle = \langle Av, Av\rangle = \|Av\|^2 \geq 0$, so
A^TA is positive semidefinite. In particular,
if $A^TAv = \lambda v$, then $\lambda\langle v,v\rangle = \langle v, A^TAv\rangle \geq 0 \Rightarrow \lambda \geq 0$.

Recall from lecture 12 that $\mathcal{N}(A) = \mathcal{N}(A^TA)$. This implies that $\operatorname{rank}(A^TA) = \operatorname{rank}(A) = r$.
It follows that A^TA has r non-zero eigenvalues, and we can write

$$A = V\begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix} V^T$$

for orthogonal $V \in \mathbb{R}^{n\times n}$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$.
Set $\lambda_i = \sigma_i^2$ for $i = 1, \dots, r$.
For $i = 1, \dots, r$ we have (with $v_i = i^{th}$ column of V)

$$(A^TA)v_i = \sigma_i^2 v_i \Rightarrow AA^T(Av_i) = \sigma_i^2(Av_i),$$

so Av_i is an eigenvector of AA^T with eigenvalue σ_i^2.
Also, since $\|v_i\| = 1$,

$$\|Av_i\|^2 = \langle v_i, A^TAv_i\rangle = \sigma_i^2\|v_i\|^2 = \sigma_i^2$$
$$\Rightarrow \|Av_i\| = \sigma_i.$$

Set $u_i = Av_i/\sigma_i$. Then $\|u_i\| = 1$ and $Av_i = \sigma_i u_i$.

For $1 \leq i, j \leq r$ and $i \neq j$,

$$\langle u_i, u_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle \sigma_i u_i, \sigma_j u_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle Av_i, Av_j\rangle$$

$$= \frac{1}{\sigma_i\sigma_j}\langle v_i, A^TAv_j\rangle = \frac{\sigma_j}{\sigma_i}\langle v_i, v_j\rangle = 0$$

So we have orthonormal $(u_1, \ldots, u_r)$ such that

$$A(v_1, \ldots, v_r) = (u_1, \ldots, u_r)\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}$$

Note that if $r < n$, then $v_{r+1}, \ldots, v_n$ are null-vectors of A ($Av_j = 0$ for $j = r+1, \ldots, n$), so

$$AV = (u_1, \ldots, u_r)\begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \underbrace{0}_{r\times(n-r)} \\ & & \sigma_r & \end{pmatrix}$$

If $r < m$, then we can extend $\{u_1, \ldots, u_r\}$ to an orthonormal basis $\{u_1, \ldots, u_m\}$. Setting $U = (u_1, \ldots, u_m)$, we have

$$AV = U\begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \overbrace{0}^{r\times(n-r)} \\ & & \sigma_r & \\ & \underbrace{0}_{(m-r)\times n} & & \end{pmatrix} = U\Sigma$$

which gives the SVD

$$A = U\Sigma V^T.$$

<u>Note:</u> We began with $A^TA = V\underbrace{\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}}_{n\times n}V^T$.
We could just as well have begun with $AA^T = U\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}U^T$. Indeed, with $A = U\Sigma V^T$, we have

$$AA^T = U\Sigma V^TV\Sigma^TU^T = U\Sigma\Sigma^TU^T$$

$$= U\underbrace{\begin{pmatrix} \sigma_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_r^2 & & \\ & & & 0 & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}}_{m\times m}U^T$$

<u>Note:</u> $A = U\Sigma V^T \Rightarrow A^T = V\Sigma^TU^T$

Application to linear least-squares problems.

Suppose we want to minimize $\|b - Ax\|^2$, where $A \in \mathbb{R}^{m \times n}$, $m \geq n$. Have seen approaches based on the normal equations and on QR decomposition. The SVD offers still another approach.

Suppose $A = U \Sigma V^T$. Then

$$\|b - Ax\|^2 = \|b - U\Sigma V^T x\|^2 = \|U^T b - \Sigma V^T x\|^2$$

$$= \|c - \Sigma y\|^2$$

where $c = U^T b$ and $y = V^T x$. If $\Sigma = \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \; 0 \end{pmatrix}$ then

$$\|b - Ax\|^2 = \sum_{i=1}^{r} (c_i - \sigma_i y_i)^2 + \sum_{i=r+1}^{m} c_i^2$$

This minimized when $y_i = c_i / \sigma_i$, $1 \leq i \leq r$, and the minimal value is $\sum_{i=r+1}^{m} c_i^2$.

If A has full rank n $(r = n)$, then $y = \begin{pmatrix} c_1/\sigma_1 \\ \vdots \\ c_n/\sigma_n \end{pmatrix}$ is uniquely determined and so is

$$x = Vy = \sum_{i=1}^{n} (c_i/\sigma_i) v_i .$$

If A is rank-deficient $(r < n)$, then $y_i = c_i/\sigma_i$ for $1 \leq i \leq r$, but $y_{r+1}, \ldots, y_n$ can be arbitrarily specified. Then

$$x = \sum_{i=1}^{r} (c_i/\sigma_i) v_i + \sum_{i=r+1}^{n} y_i v_i$$

for arbitrary $y_{r+1}, \ldots, y_n$.

Note that $\|x\|^2 = \sum_{i=1}^{r} (c_i/\sigma_i)^2 + \sum_{i=r+1}^{n} y_i^2$,

so $x = \sum_{i=1}^{r} (c_i/\sigma_i) v_i$ is the unique minimum-norm solution, called the <u>pseudo-inverse</u> solution.

Define $\Sigma^+ = \begin{pmatrix} 1/\sigma_1 & & & & \\ & \ddots & & & \\ & & 1/\sigma_r & & \\ & & & 0 & \\ & & & & \ddots\ 0 \end{pmatrix} \in \mathbb{R}^{n \times m}$

and

$$A^+ = V \Sigma^+ U^T, \text{ the } \underline{\text{pseudo-inverse}} \text{ of } A.$$

Since $c = U^T b$, the pseudo-inverse solution is just $x = A^+ b$.

In practice, the SVD and QR approaches are the two main ways of solving linear least-squares problems. They provide about the same accuracy, provided the QR decomposition uses Householder transformations. SVD solution is somewhat more expensive. Its main advantage is that it makes rank deficiency easy to detect and to treat by using the pseudo-inverse solution. Sometimes, even in the full rank case, it is useful to reduce computation costs by using a <u>truncated SVD</u>, in which

$$A \approx \underset{m \times k}{U_k} \underset{k \times k}{\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{pmatrix}} \underset{n \times k}{V_k^T}$$

for some $k \le r$. This is often done in image compression.

Schur decomposition.

Return to the square case: $A \in \mathbb{C}^{n \times n}$.
(must consider complex matrices even if A is real, so just take $A \in \mathbb{C}^{n \times n}$.)

Form: $A = URU^*$, where $U \in \mathbb{C}^{n \times n}$ is unitary and $R \in \mathbb{C}^{n \times n}$ is upper (right) triangular.

Note: $A - \lambda I = A - \lambda UU^* = U(R - \lambda I)U^*$, so $\lambda \in \sigma(A) \Leftrightarrow \lambda \in \sigma(R)$. It follows that the eigenvalues of A are the diagonal entries of R.

Derivation. We will show $U^* A U = R$ by induction on n. Trivial for $n=1$. Suppose that for some $n > 1$, every $(n-1)\times(n-1)$ matrix has a Schur decomposition, and suppose $A \in \mathbb{C}^{n \times n}$. We know A has an eigenvalue λ, and there is an eigenvector u_1 such that $Au_1 = \lambda u_1$ and $\|u_1\| = 1$. Extend $\{u_1\}$ to an orthonormal basis $\{u_1, \ldots, u_n\}$ for $\mathbb{C}^n$. Set $\hat{U} = (u_1, \ldots, u_n)$. Then

$$A\hat{U} = (Au_1, \ldots, Au_n) = (\lambda u_1, Au_2, \ldots, Au_n)$$

$$= \hat{U} \begin{pmatrix} \lambda & \times & \cdots & \times \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix}, \quad B \sim (n-1)\times(n-1)$$

With the inductive hypothesis, we have $\tilde{U}^* B \tilde{U} = \tilde{R}$ for unitary $\tilde{U}$ and upper triangular $\tilde{R}$. Then $B = \tilde{U}\tilde{R}\tilde{U}^*$ and

$$\hat{U}^* A \hat{U} = \begin{pmatrix} \lambda & \times & \cdots & \times \\ 0 & & & \\ \vdots & & \tilde{U}\tilde{R}\tilde{U}^* & \\ 0 & & & \end{pmatrix}$$

which yields $U^* A U = R$ for

$$U = \hat{U}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U} & \\ 0 & & & \end{pmatrix} \text{ and } R = \begin{pmatrix} \lambda & x & \cdots & x \\ 0 & & & \\ \vdots & & \tilde{R} & \\ 0 & & & \end{pmatrix}.$$

The Schur decomposition is often useful in evaluating analytic functions of matrices. If $A = URU^*$, then $A^k = \underbrace{URU^* URU^* \ldots URU^*}_{k \text{ times}} = U R^k U^*$.

So if $f(z) = \sum^{\infty} a_k z^k$ is analytic in a domain including $\sigma(A)$, then we have convergent power series

$$f(A) = \sum_{k=1}^{\infty} a_k A^k = U\left(\sum_{k=1}^{\infty} a_k R^k\right) U^*.$$

Hessenberg decomposition.

Return to real case.

Suppose $A \in \mathbb{R}^{n\times n}$. (Similar for $A \in \mathbb{C}^{n\times n}$.)

Form: $A = QHQ^T$, where $Q \in \mathbb{R}^{n\times n}$ is orthogonal and H is upper Hessenberg.

Similar to Schur form. Disadvantage: H is upper-Hessenberg instead of upper triangular.
Advantage: Q and H are real if A is, and they can be computed directly, whereas the Schur decomposition can only be approximated iteratively.

The Hessenberg decomposition can still be used in many ways that the Schur decomposition can.

See: $A - \lambda I = Q H Q^T - \lambda Q Q^T = Q(H - \lambda I) Q^T$,
so $\sigma(A) = \sigma(H)$.

This is useful in computing eigenvalues: First determine H; then compute the eigenvalues of H (much cheaper).

Also, $A^k = Q H^k Q^T$, so if $f(z) = \sum_{k=0}^{\infty} a_k z^k$ is analytic on $\sigma(A)$ we have

$$f(A) = \sum_{k=0}^{\infty} a_k A^k = Q\left(\sum_{k=0}^{\infty} a_k H^k\right) Q^T$$

<u>Derivation</u>.

<u>First step:</u> Choose $H_1 = I - 2u_1u_1^T$, $\|u_1\| = 1$ and $u_1 = \begin{pmatrix} 0 \\ x \\ \vdots \\ x \end{pmatrix}$, so that

$$H_1 A = \begin{pmatrix} x & x & - \\ x & x & - \\ 0 & x & - \\ \vdots & \vdots & \\ 0 & x & - \end{pmatrix}$$

(same first row as A but modified rows 2 through n)

Have

$$H_1 A H_1 = \begin{pmatrix} x & x & = \\ x & x & = \\ 0 & x & - \\ \vdots & \vdots & - \\ 0 & x & - \end{pmatrix}$$

(same 1^{st} column as $H_1 A$ but modified columns 2 through n)

<u>Second step:</u> Choose $H_2 = I - 2u_2u_2^T$, $\|u_2\| = 1$ and $u_2 = \begin{pmatrix} 0 \\ 0 \\ x \\ \vdots \\ x \end{pmatrix}$ such that

$$H_2 H_1 A H_1 = \begin{pmatrix} x & x & x & \equiv \\ x & x & x & \equiv \\ 0 & x & x & \equiv \\ \vdots & 0 & x & \equiv \\ 0 & 0 & x & \equiv \end{pmatrix}$$

(same first two rows, modified rows 3 through n).

Then $$H_2 H_1 A H_1 H_2 = \begin{pmatrix} x & x & x & - & - \\ x & x & x & - & - \\ 0 & x & x & - & - \\ \vdots & 0 & x & - & - \\ \vdots & \vdots & \vdots & - & - \\ 0 & 0 & x & - & - \end{pmatrix}$$

(same first two columns, modified columns 3 through n).

Continue, obtain

$$H_{n-1} \cdots H_1 A H_1 \cdots H_{n-1} = \begin{pmatrix} x & - & - & - \\ x & - & - & - \\ & \ddots & - & - \\ 0 & & x & x \end{pmatrix} = H$$

or $A = Q H Q^T$, where $Q = H_1 \cdots H_{n-1}$.