10/3 Diagonalizability.

Def 6.1: $A \in \mathbb{R}^{n\times n}$ is diagonalizable if $\exists$ nonsingular $n\times n$ M $\ni$ $A = M\Lambda M^{-1}$, where Λ is a diagonal $n\times n$ matrix.

Suppose A is diagonalizable and we write $M = (v_1, \dots, v_n)$ and $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$. Then

$$A = M\Lambda M^{-1} \Rightarrow AM = \Lambda M \Rightarrow (Av_1, \dots, Av_n) = (\lambda_1 v_1, \dots, \lambda_n v_n),$$

so the columns of M are the eigenvectors of A and the diagonal entries of Λ are the eigenvalues. Since M is nonsingular, $\{v_1, \dots, v_n\}$ is linearly independent.

Conversely, if $\{v_1, \dots, v_n\}$ is a set of linearly independent eigenvectors, with $Av_i = \lambda_i$ for each i, then $M = (v_1, \dots, v_n)$ is invertible, and M and $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$ satisfy $AM = M\Lambda \Rightarrow A = M\Lambda M^{-1}$.

Note: Being careful to allow M and Λ to be complex, even though A is real.

Prop. 6.2: A is diagonalizable $\Leftrightarrow$ A has n linearly independent eigenvectors.

So when does A have n linearly independent eigenvectors? Explore.

Suppose $\lambda_1, \dots, \lambda_k$ are the distinct eigenvalues of A. For $i = 1, \dots, k$, set

$$k_i = \dim(\mathcal{N}(A - \lambda_i I)) = \text{geometric multiplicity of } \lambda_i.$$

Recall: In general, for $1 \le i \le n$,

$$k_i \le m_i = \underline{\text{algebraic}}\ \underline{\text{multiplicity}} \text{ of } \lambda_i. \quad (*)$$

(Prove this later.)

Then $\sum_{i=1}^{k} k_i \le \sum_{i=1}^{k} m_i = n$, and $\sum_{i=1}^{k} k_i = n$

$\Leftrightarrow k_i = m_i$ for $1 \le i \le n$!

Prop. 6.3: If $i \ne j$, then $\mathcal{N}(A-\lambda_i I) \cap \mathcal{N}(A-\lambda_j I) = \{0\}$.

Pf. Suppose $x \in \mathcal{N}(A-\lambda_i I) \cap \mathcal{N}(A-\lambda_j I)$ for $i \ne j$.
Then $\lambda_i \ne \lambda_j$, and

$$\begin{aligned} 0 &= (A-\lambda_i I)x = (A-\lambda_j)x \\ &= (\lambda_j - \lambda_i)x = (\lambda_i - \lambda_j)x \end{aligned}$$

Since $\lambda_i \ne \lambda_j$, it follows that $x = 0$.

Prop. 6.4: Suppose that, for $i = 1, \dots, k$, v_i is an eigenvector of A with eigenvalue λ_i.
Then $\{v_1, \dots, v_k\}$ is linearly independent.

Pf.: Since v_1 is an eigenvector and therefore nonzero, $\{v_1\}$ is linearly independent.
Inductive Hypothesis: Suppose $\{v_i\}_{i=1}^{l}$ is lin. ind. for $1 \le l < n$.
Claim: $\{v_i\}_{i=1}^{l+1}$ is lin. ind.
Why? Suppose $0 = \sum_{i=1}^{l+1} \alpha_i v_i$. If $\alpha_{l+1} = 0$, then $\alpha_1 = \dots = \alpha_l = 0$ since $v_1, \dots, v_l$ are lin. ind.
If $\alpha_{l+1} \ne 0$, then

$$v_{l+1} = \sum_{i=1}^{l} \beta_i v_i, \quad \beta_i = -\frac{\alpha_i}{\alpha_{l+1}}$$

Apply A to both sides: $\lambda_{l+1} v_{l+1} = \sum_{i=1}^{l} \beta_i \lambda_i v_i$.

Multiply 1st eqn by λ_{l+1} and subtract to get

$$0 = \sum_{i=1}^{l} \beta_i (\lambda_{l+1} - \lambda_i) v_i$$

Since $(\lambda_{l+1} - \lambda_i) \neq 0$ and $v_1, -, v_l$ are lin ind, conclude that $\beta_1 = ,,, = \beta_l = 0 \Rightarrow \alpha_1 = ,,, = \alpha_l = 0$. Then $0 = \alpha_{l+1} v_{l+1} \Rightarrow \alpha_{l+1} = 0$ (contradiction).

Cor. 6.5: If A has n distinct eigenvalues, then A is diagonalizable.

Pf.: If $v_1, -, v_n$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, -, \lambda_n$, then $\{v_1, -, v_n\}$ is linearly independent by Prop. 6.4.

Note that, in this case, we must have $k_i = m_i = 1$ for each i. (As before, k_i = geo. mult. and m_i = alg. mult.)

Term'y: $\lambda_i \in \sigma(A)$ is simple if $m_i = k_i = 1$.

Now consider a more general case where some k_i may be > 1.

For $i = 1, ,,, , k$, let $\{ v_{i1}, ,,, , v_{ik_i} \}$ be a basis of $\mathcal{N}(A - \lambda_i)$. Note that each $v_{ij} \neq 0$ and so is an eigenvector of A with eigenvalue λ_i.
Note also that $\mathcal{N}(A - \lambda_i)$ may be in $\mathbb{C}^n$ but not $\mathbb{R}^n$.

Lemma 6.6. $\bigcup_{i=1}^{k} \{v_{i1}, \dots, v_{ik_i}\}$ is linearly independent.

Pf: We have that $\{v_{11}, \dots, v_{1k_1}\}$ is a basis of $\mathcal{N}(A-\lambda_1 I)$ and so is linearly independent.

Ind. Hyp.: Suppose $\bigcup_{i=1}^{\ell} \{v_{i1}, \dots, v_{ik_i}\}$ is lin. ind. for some $\ell \geq 1$.

Claim: $\bigcup_{i=1}^{\ell+1} \{v_{i1}, \dots, v_{ik_i}\}$ is lin ind.

Why? Suppose

$$0 = \sum_{i=1}^{\ell+1} \sum_{j=1}^{k_i} \alpha_{ij} v_{ij}$$

$$= \sum_{i=0}^{\ell} \sum_{j=0}^{k_i} \alpha_{ij} v_{ij} + \sum_{j=0}^{k_{\ell+1}} \alpha_{\ell+1,j} v_{\ell+1,j}$$

Applying $(A-\lambda_{\ell+1} I)$ to both sides gives

$$0 = \sum_{i=0}^{\ell} \sum_{j=0}^{k_i} \alpha_{ij} (\lambda_i - \lambda_{\ell+1}) v_{ij} + \sum_{j=0}^{k_{\ell+1}} \alpha_{\ell+1,j} (\lambda_{\ell+1} - \lambda_{\ell+1}) v_{\ell+1,j}$$

$$= \sum_{i=0}^{\ell} \sum_{j=0}^{k_i} \alpha_{ij} (\lambda_i - \lambda_{\ell+1}) v_{ij}$$

Since $(\lambda_i - \lambda_{\ell+1}) \neq 0$ for each i and $\bigcup_{i=1}^{\ell} \{v_{i1}, \dots, v_{ik_i}\}$ is linearly independent, it follows that $\alpha_{ij} = 0$ for $0 \leq j \leq k_i$ and $0 \leq i \leq \ell$. Then

$$0 = \sum_{j=0}^{k_{\ell+1}} \alpha_{\ell+1,j} v_{\ell+1,j}$$

which implies $\alpha_{\ell+1,j} = 0$ for $1 \leq j \leq k_{\ell+1}$

Since $\{v_{\ell+1,1}, \dots, v_{\ell+1,k_{\ell+1}}\}$ is a basis of $\mathcal{N}(A-\lambda_{\ell+1} I)$ and so is linearly independent.

Note that span $\left(\bigcup_{i=1}^{k}\{v_{i_1},\dots,v_{ik_i}\}\right)$ includes all eigenvectors of A. Why? If v is an eigenvector of A, then $Av = \lambda_i v$ for some λ_i. It follows that $v \in \mathcal{N}(A-\lambda_i I)$ and so $v \in$ span $\{v_{i_1},\dots,v_{ik_i}\}$.

Th. 6.7: A is diagonalizable if and only if the geometric and algebraic multiplicities of each eigenvalue of A are equal.

Pf: By Proposition 6.2, A is diagonalizable $\Leftrightarrow$ A has n linearly independent eigenvectors. By Lemma 6.6, this holds $\Leftrightarrow$ $\sum_{i=1}^{k} k_i = n$, which by earlier remarks, holds $\Leftrightarrow$ $k_i = m_i$ for $1 \le i \le k$.

Term'y $\lambda_i \in \sigma(A)$ is semisimple if $m_i = k_i$.

So the theorem says A is diagonalizable $\Leftrightarrow$ all of its eigenvalues are semisimple.

Diagonalizability is a wonderful thing. Explore some uses in applications.

<u>Example</u>. Applications to ODE systems.
Suppose we have a first-order ODE IVP with constant coefficients:

$$y' = Ay, \quad y(0) = y_0$$

where $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$.
Recall: $n=1 \Rightarrow y(t) = y_0 e^{At}$. What about general n?

Suppose A is diagonalizable, i.e., $A = M \Lambda M^{-1}$. Then

$$y' = Ay,\ y(0) = y_0 \Leftrightarrow M^{-1}y' = M^{-1}AMM^{-1}y, \quad M^{-1}y(0) = M^{-1}y_0$$

$$\Leftrightarrow \quad z' = \Lambda z, \quad z(0) = z_0,$$

where $z = M^{-1}y$, $z_0 = M^{-1}y_0$. This is a <u>decoupled</u> system, i.e.,

$$z_i' = \lambda_i z_i, \quad z_i(0) = z_{0i} \quad \text{for } 1 \le i \le n.$$

Sol$\underline{\text{ns}}$ are $z_i(t) = z_{i0}\, e^{\lambda_i t}$ for $1 \le i \le n$, or

$$z(t) = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} z_0$$

Then

$$y(t) = M \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} M^{-1} y_0 .$$

Set $e^{At} = M \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} M^{-1}$.

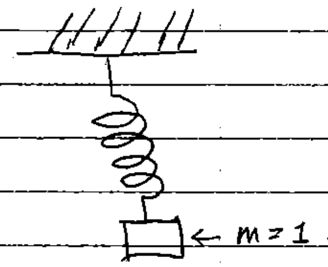
Makes sense: Know $e^x = \sum_0^\infty \frac{x^k}{k!}$, so

$$\sum_{k=0}^\infty \frac{(At)^k}{k!} = \sum_0^\infty \frac{t^k}{k!} \underbrace{(M\Lambda M^{-1})(M\Lambda M^{-1})\cdots(M\Lambda M^{-1})}_{k \text{ factors}}$$

$$= \sum_0^\infty \frac{t^k}{k!} M \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix} M^{-1} = M \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} M^{-1}$$

Then the solution is $y(t) = e^{At} y_0$. This is written without explicit reference to the eigenvectors or eigenvalues of A, but $\sigma(A)$ still governs the behavior of solutions (exponential growth, decay, etc.).

Don't forget: Λ and M may be complex, even if A is real. However, if A and y_0 are real, then things work out so that $y(t) = e^{At} y_0$ is real $\forall\, t$.



Illustrate with a specific ODE:
Suppose $x(t)$ = displacement from equilibrium of a unit mass suspended on a spring with spring constant k. Then

$$x''(t) = \text{restoring force} \quad (\text{Newton: } F = mA)$$
$$= -k\,x(t) \quad (\text{Hooke's law})$$

So we get the 2nd-order ODE IVP:

$$x'' + kx = 0, \quad x(0) = \alpha, \quad x'(0) = \beta.$$

Recast as a 1st-order system: $y_1(t) = x(t)$, $y_2(t) = x'(t)$, so

$$y' = \begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} x' \\ x'' \end{pmatrix} = \begin{pmatrix} y_2 \\ -ky_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} y$$

With $y_0 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, the IVP is

$$y' = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix} y, \quad y(0) = y_0.$$

Diagonalize $A = \begin{pmatrix} 0 & 1 \\ -k & 0 \end{pmatrix}$: Eigenvalues are solutions of
$0 = \det(A - \lambda I) = \det\begin{pmatrix} -\lambda & 1 \\ -k & -\lambda \end{pmatrix} = \lambda^2 + k$,
the characteristic equation.

Write $k = \mu^2$ for convenience. Then the characteristic eq π is

$$0 = \lambda^2 + \mu^2$$

and roots (eigenvalues) are $\lambda_+ = i\mu$, $\lambda_- = -i\mu$.
Corresponding eigenvectors are (setting first components of both equal to 1)

$$v_+ = \begin{pmatrix} 1 \\ i\mu \end{pmatrix}, \quad v_- = \begin{pmatrix} 1 \\ -i\mu \end{pmatrix}$$

and $M = \begin{pmatrix} 1 & 1 \\ i\mu & -i\mu \end{pmatrix}$, $M^{-1} = \frac{1}{-2i\mu}\begin{pmatrix} -i\mu & -1 \\ -i\mu & 1 \end{pmatrix} = \frac{i}{2\mu}\begin{pmatrix} -i\mu & -1 \\ -i\mu & 1 \end{pmatrix}$,

gives

$$M^{-1}AM = \Lambda = \begin{pmatrix} i\mu & 0 \\ 0 & -i\mu \end{pmatrix}.$$

Then $y(t) = e^{At}y_0 = M\begin{pmatrix} e^{i\mu t} & 0 \\ 0 & e^{-i\mu t} \end{pmatrix} M^{-1} y_0$

$$= \frac{i}{2\mu}\begin{pmatrix} 1 & 1 \\ i\mu & -i\mu \end{pmatrix} \underbrace{\begin{pmatrix} e^{i\mu t} & 0 \\ 0 & e^{-i\mu t} \end{pmatrix}\begin{pmatrix} -i\mu & -1 \\ -i\mu & 1 \end{pmatrix}}_{\begin{pmatrix} -i\mu e^{i\mu t} & -e^{i\mu t} \\ -i\mu e^{-i\mu t} & e^{-i\mu t} \end{pmatrix}} y_0$$

$$= \frac{i}{2\mu}\begin{pmatrix} -i\mu\left[e^{i\mu t}+e^{-i\mu t}\right] & -\left[e^{i\mu t}-e^{-i\mu t}\right] \\ \mu^2\left[e^{i\mu t}-e^{-i\mu t}\right] & -i\mu\left[e^{i\mu t}+e^{-i\mu t}\right] \end{pmatrix} y_0$$

Use Euler's formula ($e^{i\theta} = \cos\theta + i\sin\theta$) to obtain

$$y(t) = \begin{pmatrix} \frac{1}{2}\cdot 2\cos\mu t & -\frac{i}{2\mu}\cdot 2i\sin\mu t \\ \frac{i}{2\mu}2\mu^2 i\sin\mu t & \frac{1}{2}\cdot 2\cos\mu t \end{pmatrix} y_0$$

$$= \begin{pmatrix} \cos\mu t & \frac{1}{\mu}\sin\mu t \\ -\mu\sin\mu t & \cos\mu t \end{pmatrix} y_0$$

Extensions to general linear transformations.

What we've said about so far about eigenvalues and eigenvectors of matrices applies in particular to matrices that are representations of linear transformation on general finite-dimensional vector spaces and, by extension, to the transformations themselves.

In particular, our notions of geometric and algebraic multiplicity and simple and semi-simple eigenvalues naturally extends to general linear transformations.

Suppose we have a linear transformation $T: \mathcal{V} \to \mathcal{V}$ where $\dim \mathcal{V} = n$. If A is a matrix representation of T, then A is diagonalizable $\Leftrightarrow A = M \Lambda M^{-1}$ $\Leftrightarrow$ A has n linearly independent eigenvectors $\Leftrightarrow$ T has n linearly independent eigenvectors. (The eigenvectors of A are just the coordinate vectors of the eigenvectors of T with respect to the chosen basis.) Combining and extending our earlier results for matrices, we have

Th. 6.8: T has n linearly independent eigenvectors $\Leftrightarrow$ each eigenvalue of T is semi-simple.

In this case, if $\{v_1, \ldots, v_n\}$ is a set of n linearly independent eigenvectors with $T(v_i) = \lambda_i v_i$ for each i, then for $v = \sum_i \alpha_i v_i \in \mathcal{V}$,

$$T(v) = \sum_{i=1}^{n} \alpha_i T(v_i) = \sum_i \alpha_i \lambda_i v_i$$

Note: Carefully avoiding saying $\{v_1, \ldots, v_n\}$ is a basis of $\mathcal{V}$. (The v_i's might not all be in $\mathcal{V}$.) But the statements still hold.