Lecture 19

The Power Method

(& FRIENDS)

* PSET 3 is out! $\boxed{\text{Due 10/30 9PM.}}$
(has Julia component, so start early!)

* NEW checkup policy!

Last time:

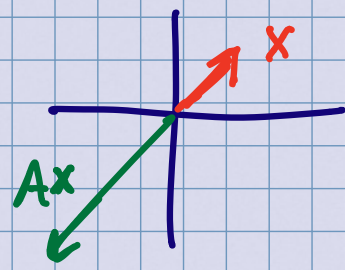
$A \in \mathbb{R}^{n \times n}$

eigenvalue equation

$$Ax = \lambda x$$

matrix

scalar (eigenvalue)



$$p(\lambda) = \det(A - \lambda I) = \prod_{i=1}^{n} (\lambda_i - \lambda)$$

Sometimes

A is diagonalizable $\boxed{A = TDT^{-1}}$

- E.g., when all eigenvalues are distinct.

- More generally, if for every eigenvalue λ_i:

algebraic mult.(λ_i) = geometric mult.(λ_i)

$p(\lambda)$ has a factor $(\lambda_i - \lambda)^K$

$\dim N(A - \lambda_i I) = K$

Assume $A \in \mathbb{R}^{n \times n}$ is diagonalizable,

i.e., $\boxed{A = TDT^{-1}}$ $= T \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} T^{-1}$

• Then $A^k = T D^k T^{-1}$

Polynomials applied to matrices

Let $\boxed{q(t) = q_k t^k + q_{k-1} t^{k-1} + \cdots \; q_1 t + q_0}$

be a univariate polynomial of degree k.

Then, we can define $q(A)$, where A is a matrix.

$$q(A) = q_k A^k + q_{k-1} A^{k-1} + \cdots + q_1 A + q_0 I$$

In fact, this also makes sense for power series, if they converge.

Useful, for instance, to define the exponential of a matrix.

$$e^{x} = \sum_{k=0}^{\infty} \frac{x^k}{k!} \quad (x \in \mathbb{R}) \quad \rightsquigarrow \quad e^{A} = \sum_{k=0}^{\infty} \frac{A^k}{k!} \quad (A \in \mathbb{R}^{n \times n}).$$

(or sin/cos, log, ...)

here's a nice (and useful) application:

Recall that: (if $|x|<1$) $\frac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots\cdots$ (infinite geometric series)

Then: $$\boxed{(I-A)^{-1} = I + A + A^2 + A^3 + \cdots}$$

(this converges if A is "small").

An interesting observation:

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

Let $q(t)$ be a polynomial, and

$A = TDT^{-1}$ a diagonalizable matrix.

Then
$$\boxed{q(A) = T \begin{bmatrix} q(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & q(\lambda_n) \end{bmatrix} T^{-1}}$$

characteristic polynomial

Q: what happens if we choose $q = p$, i.e., what is

$$p(A) = \, ?$$

Example: (Lecture 17)

$A = \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$ $p(\lambda) = \lambda^2 - 3\lambda - 4$

characteristic polynomial of A.

$$p(A) = A^2 - 3A - 4I$$

$$= \begin{bmatrix} 7 & 6 \\ 9 & 10 \end{bmatrix} - \begin{bmatrix} 3 & 6 \\ 9 & 6 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} =$$

In fact, this true for all matrices (even if they are not diagonalizable).

CAYLEY-HAMILTON THEOREM:

"Every square matrix satisfies its own characteristic polynomial", i.e.,

if $\underset{\text{scalar}}{p(\lambda)} = \det(A - \lambda I) \implies \underset{\text{matrix}}{p(A)} = 0$.

DIAGONALIZATION (AND $A^k = TD^kT^{-1}$)

also allows us to understand:

- the long-term behavior of linear dynamical systems.
- Asympotic growth of recurrences
- Convergence (and design) of algorithms

(and much more!)

EXAMPLE:

LINEAR DYNAMICAL SYSTEMS

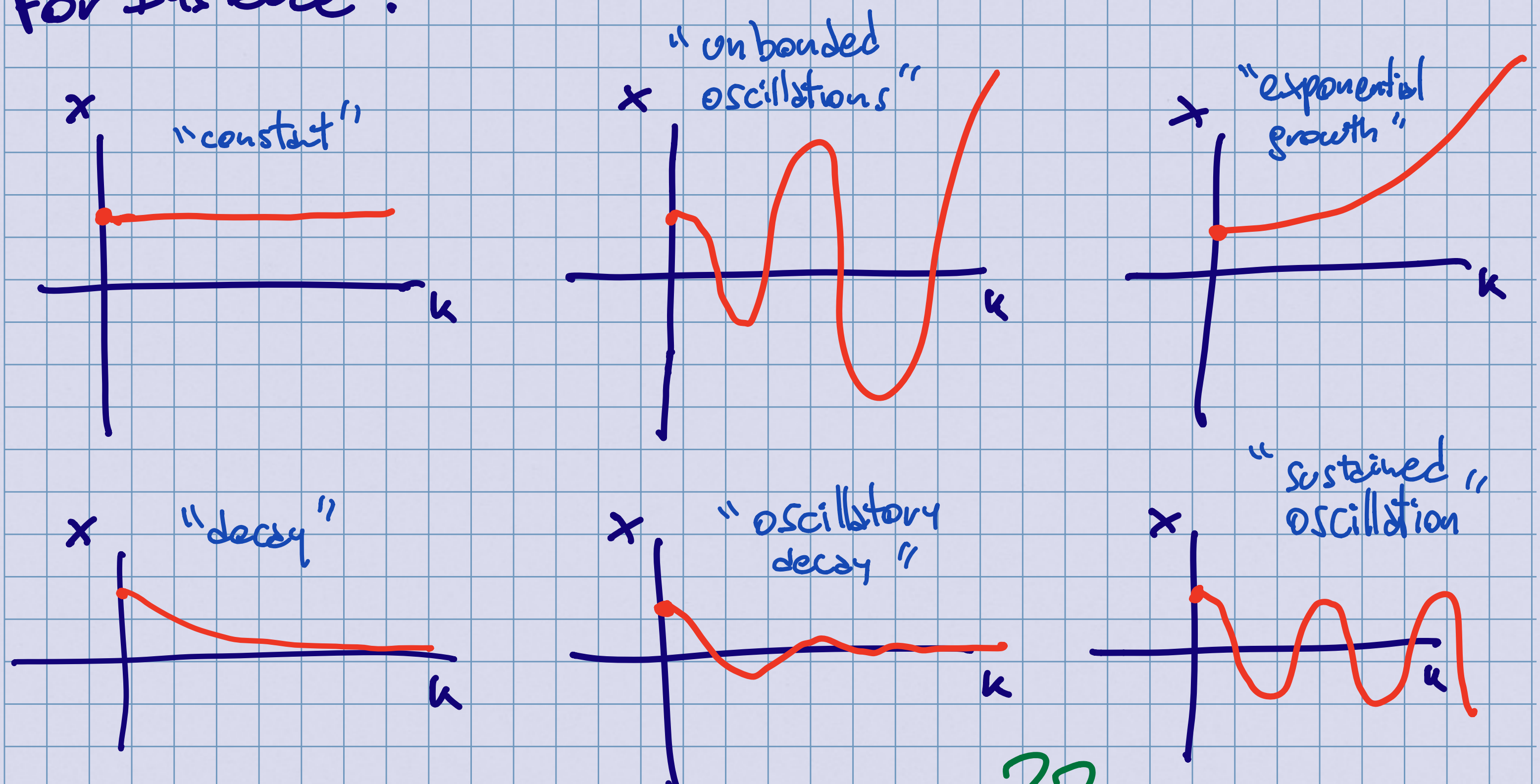
$$\boxed{x_{k+1} = A x_k}$$

x_k is the **state** at time k

(e.g., vehicle, epidemic, population, algorithm, etc)

Q: What is the **long-term behavior** of x_k, i.e., what happens as $k \to \infty$?

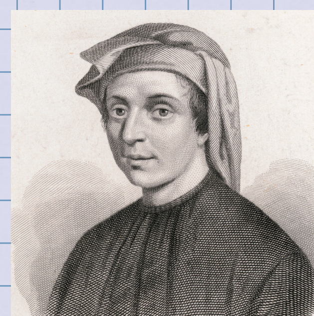
For Instance:



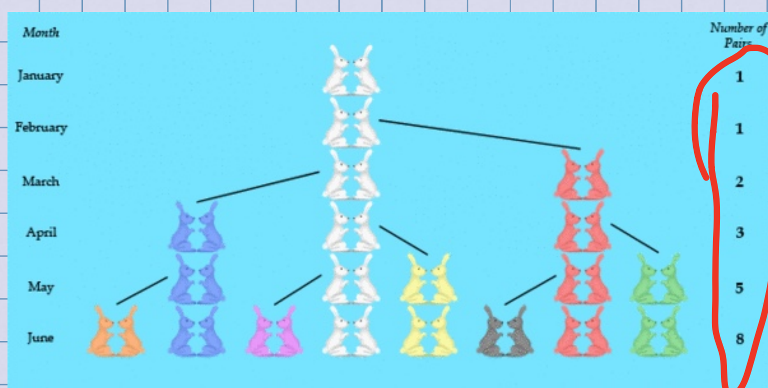
Q: HOW TO DECIDE??

FIBONACCI'S RABBITS

(1202 AD)



$$F_0 = 0$$
$$F_1 = 1$$
$$F_2 = 1$$
$$F_3 = 2$$
$$F_4 = 3$$



$$\boxed{F_{k+1} = F_k + F_{k-1}}$$

$\Leftarrow$ linear recurrence sequence

(also, a dynamical system)

(convert to matrices!)

Define

$$X_k = \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix} \Rightarrow \underbrace{\begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}}_{X_{k+1}} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}}_{X_k}$$

Q: How does the sequence F_k behave, for large k? (ie, how many rabbits will we have in month k?)

<u>Diagonalization!</u>

$$P(\lambda) = \det(A - \lambda I) = \lambda^2 - \lambda - 1$$

Eigenvalues of A

$$\Rightarrow \begin{cases} \lambda_1 = \phi \\ \lambda_2 = 1 - \phi \end{cases}$$

$$\boxed{\phi = \frac{1+\sqrt{5}}{2} = 1.618\ldots}$$

"golden ratio"

$$x_k = A\,x_{k-1} \quad \Rightarrow \quad \boxed{x_k = A^k x_0}$$

powers of A

if $A = T \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} T^{-1}$

$$A^k = T \begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix} T^{-1}$$

$\lambda_1 = \phi \sim 1.61$

$\lambda_2 = 1-\phi \sim -0.61$

as $k \to \infty$

$\lambda_1^k \to \infty$

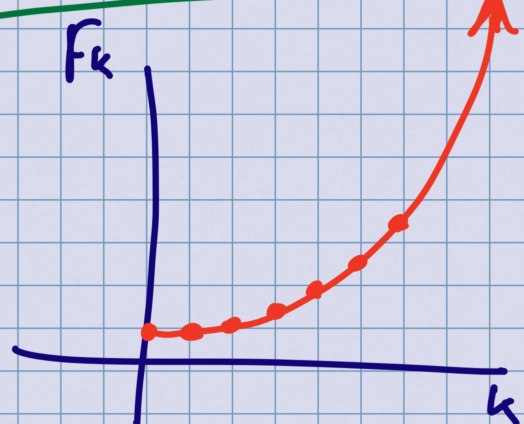
$\lambda_2^k \to 0$

$\Rightarrow$ for large k,

$$\boxed{F_k \approx \phi^k = (1.618\ldots)^k}$$

"exponential growth"

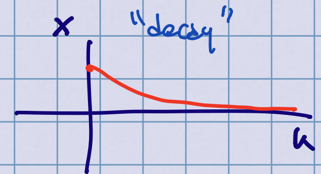
rate is given by largest eigenvalue (ϕ)

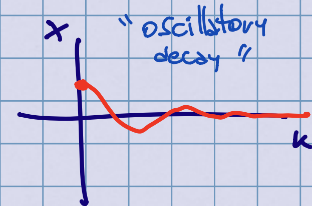


Stability : $x_{k+1} = A x_k$, fixed initial condition x_0

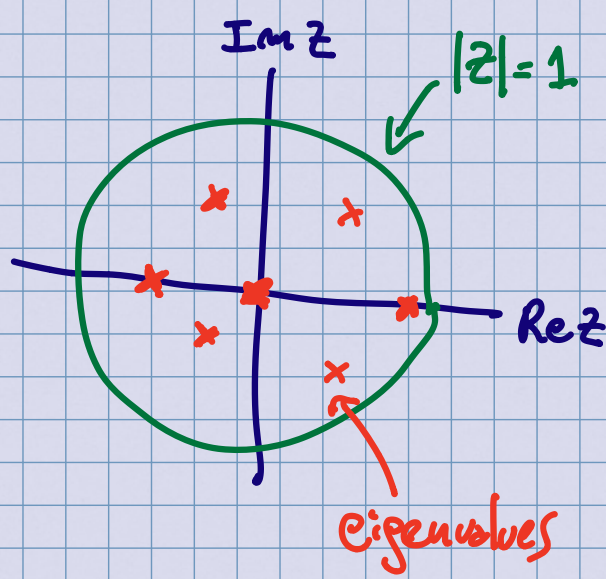
Does the state x_k goes to zero, as $k \to \infty$?

$$A^k = T \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} T^{-1}$$





Thm: If all eigenvalues satisfy $|\lambda_i| < 1$, then $x_k \to 0$ as $k \to \infty$.



"system is stable if all eigenvalues are strictly inside the unit disk"

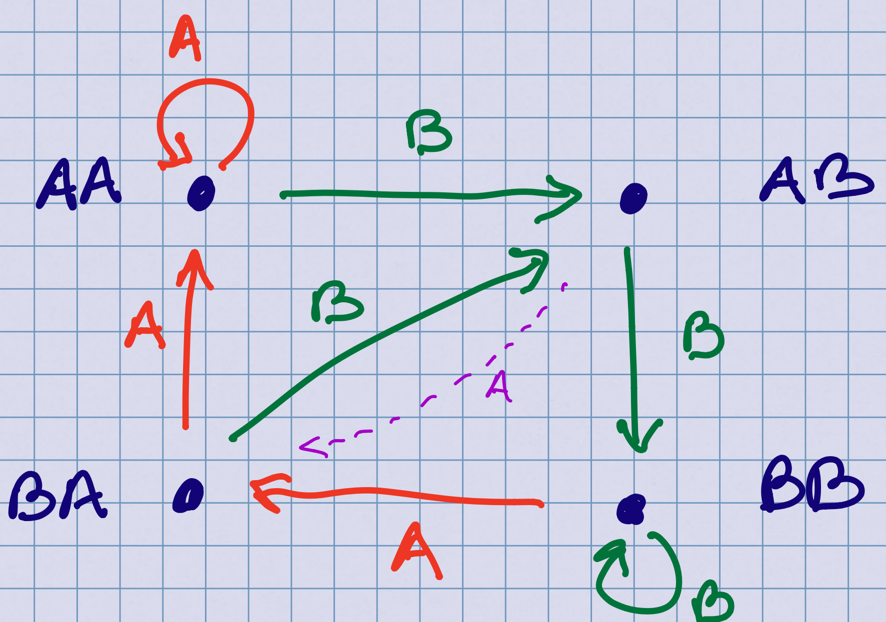
Word asymptotics

ABBA

How many "words" of length n (in the letters A, B) are there, that do not contain the string "ABA"?

E.g.: AABBBAABBBA vs BABABAAAB

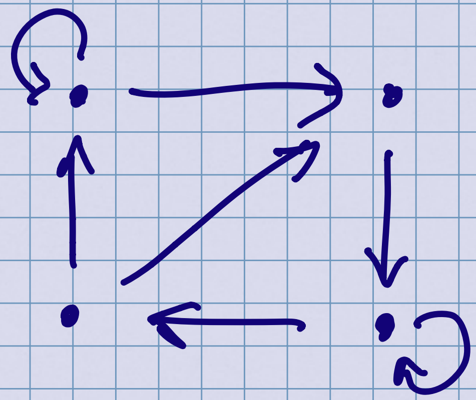
Hey, this about counting paths on a graph!



last two letters.

ABA

FORBIDDEN



$$G = \begin{array}{cccc} AA & AB & BA & BB \\ \end{array}$$
$$G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

words : compute G^k !

Q1: IF I DO NOT HAVE THE CONSTRAINT, HOW MANY WORDS OF LENGTH N ARE THERE?

$$w(N) \sim 2^N$$

Q2: WITH THE CONSTRAINT (NO ABA), HOW MANY ?

$$w(N) \sim (1.7548)^N \quad \left(\text{for large } N\right)$$

Largest eigenvalue (in absolute value) of G

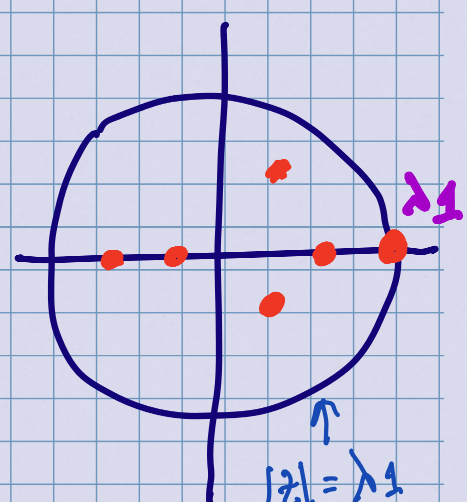
POWER ITERATION

Natural method for computing one eigenvalue / eigenvector. λ_1, v_1

Many variations exist, here we'll present the simplest version.

- Matrix $A \in \mathbb{R}^{n \times n}$, eigs $\lambda_1 \dots \lambda_n$
- Dominant eigenvalue : $|\lambda_1| > |\lambda_i| \quad i=2 \dots n$
- Initial vector $x_0 \in \mathbb{R}^n$ is "random"

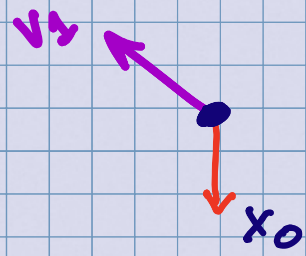
(more precisely, x_0 has a nonzero component along v_1)



Iteration:

$$\boxed{x_{k+1} = \frac{A x_k}{\| A x_k \|}}$$

(multiply by A, then normalize)



Then, as $k \to \infty$

- x_k converges* to eigenvector v_1
- $x_k^T A x_k$ converges to eigenvalue λ_1

Why?

$$T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

$$\begin{aligned} x_k \sim A^k x_0 &= T D^k T^{-1} x_0 \\ &= T \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} (T^{-1} x_0) \\ &\approx T \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} * \\ \vdots \end{pmatrix} \\ &\approx v_1 \end{aligned}$$

first component is nonzero (A2)

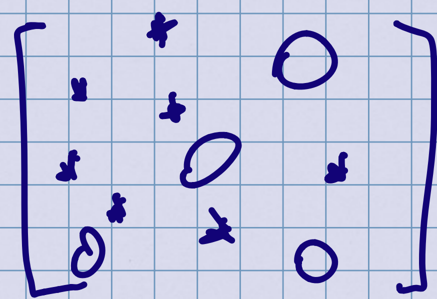
$|\lambda_1^k| \gg |\lambda_i^k|$ (A2)

What determines convergence rate? $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \ldots$
(fast, slow)

spectral gap $\dfrac{|\lambda_1|}{|\lambda_2|}$

Power iteration is very useful for large, sparse matrices.

(these typically arise, e.g., from graphs or PDEs)



lots of zeros
few nonzeros.

Variations

- Specific eigenvalues (inverse iteration, ...)
- Faster convergence (Rayleigh quotient, ...)