Math 2270-002 Week 10 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material in 4.9 (google page rank), 5.1-5.4

Mon Oct 29

- 4.9-5.1 google page rank as the stationary vector for a Markov Chain; introduction to eigenvectors and eigenvalues

Announcements:

Warm-up Exercise:

Monday Review!

We've been studying section 4.9 Markov Chains, which are *discrete dynamical systems* with initial state $\underline{\boldsymbol{x}}_0 \in \mathbb{R}^n$, and with *transition matrix* P, so that

$$\underline{\boldsymbol{x}}_k = P\,\underline{\boldsymbol{x}}_{k-1}, \quad k = 1, 2, 3, ...$$
$$\Rightarrow \underline{\boldsymbol{x}}_k = P^k\,\underline{\boldsymbol{x}}_0 \quad k \in \mathbb{N}.$$

For a *Markov Chain* the transition matrix P is required to be *stochastic*, i.e. each column of P is a *probability vector* having non-negative components which sum to 1. For a Markov Chain we also usually take the initial vector $\underline{\boldsymbol{x}}_0$ to be a probability vector, in which case the successive vectors $\underline{\boldsymbol{x}}_k$ are as well.

A stochastic matrix P is called *regular* if some power of P has all positive entries (as opposed to just non-negative).

A probability vector $\underline{\boldsymbol{q}}$ is called a *steady state* vector for a Markov Chain with transition matrix P if

$$P\,\underline{\boldsymbol{q}} = \underline{\boldsymbol{q}}.$$

(Notice that in this case, if $\underline{\boldsymbol{x}}_0 = \underline{\boldsymbol{q}}$ then each $\underline{\boldsymbol{x}}_k = \underline{\boldsymbol{q}}$ as well.)

Long-time behavior of Markov chains:

<u>Theorem</u> (Perron-Frobenius Theorem) If P is an $n \times n$ regular stochastic matrix, then P has a unique steady state vector $\underline{\boldsymbol{q}}$. Furthermore, if $\underline{\boldsymbol{x}}_0$ is any initial state (probability vector) for the Markov chain

$$\underline{\boldsymbol{x}}_{k+1} = P\,\underline{\boldsymbol{x}}_k \qquad k = 0, 1, 2,....$$

then the Markov chain $\{\underline{\boldsymbol{x}}_k\}$ converges to the steady state $\underline{\boldsymbol{q}}$ as $k \to \infty$. In particular, since the j^{th} column of P^k is $P^k\,\underline{\boldsymbol{e}}_j$ and $\underline{\boldsymbol{e}}_j$ is an admissable initial state probability vector, each column of P^k converges to $\underline{\boldsymbol{q}}$.

On Friday we discussed two examples of Markov Chains from section 4.9, and then moved into the google page rank notes....

The Giving Game: Google Page Rank

University of Utah Teachers' Math Circle

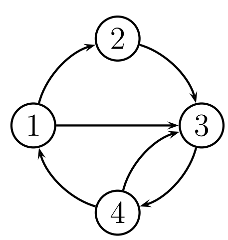
Nick Korevaar

March 24, 2009

Stage 1: The Game

Imagine a game in which you repeatedly distribute something desirable to your friends, according to a fixed template. For example, maybe you're giving away "play–doh" or pennies! (Or it could be you're a web site, and you're voting for the sites you link to. Or maybe, you're a football team, and you're voting for yourself, along with any teams that have beaten you.)

Let's play a small–sized game. Maybe there are four friends in your group, and at each stage you split your material into equal sized lumps, and pass it along to your friends, according to this template:

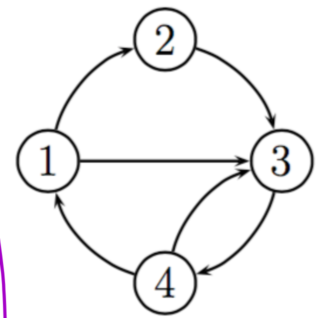


The question at the heart of the basic Google page rank algorithm is: in a voting game like this, with billions of linked web sites and some initial vote distribution, does the way the votes are distributed settle down in the limit? If so, sites with more limiting votes must ultimately be receiving a lot of votes, so must be considered important by a lot of sites, or at least by sites which themselves are receiving a lot of votes. Let's play!

1. Decide on your initial material allocations. I recommend giving it all to one person at the start, even though that doesn't seem fair. If you're using pennies, 33 is a nice number for this template. At each stage, split your current amount into equal portions and distribute it to your friends, according to the template above. If you have remainder pennies, distribute them randomly. Play the game many (20?) times, and see what ultimately happens to the amounts of material each person controls. Compare results from different groups, with different initial allocations.

2. While you're playing the giving game, figure out a way to model and explain this process algebraically!

Play the google game! each vertex splits its current vote fraction into n equal pieces, if it links to n other sites, and sends those vote fractions according to the digraph ("directed graph")



initial vote fraction

$$\begin{bmatrix} x_{0,1} \\ x_{0,2} \\ x_{0,3} \\ x_{0,4} \end{bmatrix}$$

after one play of the game

$$x_{0,1}\begin{bmatrix} 0 \\ .5 \\ .5 \\ 0 \end{bmatrix} + x_{0,2}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{0,3}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_{0,4}\begin{bmatrix} .5 \\ 0 \\ .5 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & .5 \\ .5 & 0 & 0 & 0 \\ .5 & 1 & 0 & .5 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_{0,1} \\ x_{0,2} \\ x_{0,3} \\ x_{0,4} \end{bmatrix}$$

↑ P

Transition matrix for problem **1**, to a large power:

{{0,0,0,.5},{.5,0,0,0},{.5,1,0,.5},{0,0,1,0}}^30

Browse Examples Surprise Me

Input:

$$\begin{pmatrix} 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{pmatrix}^{30}$$

Open code

Result:

$$\begin{pmatrix} 0.181842 & 0.181658 & 0.181942 & 0.181723 \\ 0.0908937 & 0.091013 & 0.0908289 & 0.090971 \\ 0.363665 & 0.363445 & 0.363784 & 0.363523 \\ 0.3636 & 0.363884 & 0.363445 & 0.363784 \end{pmatrix}$$

← 3rd
← 4th

Stage 2: Modeling the game algebraically

The game we just played is an example of a *discrete dynamical system*, with constant *transition matrix*. Let the initial fraction of play dough distributed to the four players be given by

$$\mathbf{x}_0 = \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ x_{0,3} \\ x_{0,4} \end{bmatrix}, \quad \sum_{i=1}^{4} x_{0,i} = 1$$

Then for our game template on page 1, we get the fractions at later stages by

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \\ x_{k+1,3} \\ x_{k+1,4} \end{bmatrix} = x_{k,1} \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \end{bmatrix} + x_{k,2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{k,3} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_{k,4} \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \\ x_{k+1,3} \\ x_{k+1,4} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 \\ 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ x_{k,3} \\ x_{k,4} \end{bmatrix}$$

P

So in matrix form, $\mathbf{x}_k = A^k \mathbf{x}_0$ for the transition matrix A given above.

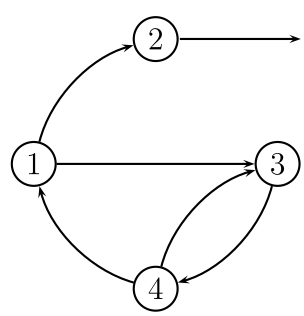
3. Compute a large power of A. What do you notice, and how is this related to the page 1 experiment?

4. The limiting "fractions" in this problem really are fractions (and not irrational numbers). What are they? Is there a matrix equation you could solve to find them, for this small problem? Hint: the limiting fractions should remain fixed when you play the game.

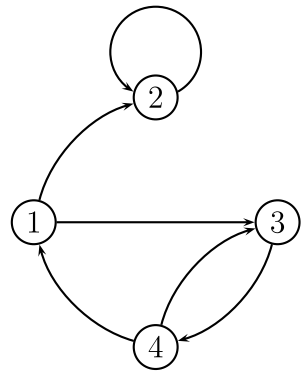
   $P\vec{q} = \vec{q}$ $\quad (P-I)\vec{q} = \vec{0}$

5. Not all giving games have happy endings. What happens for the following templates?

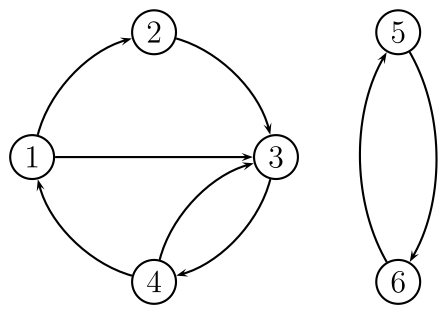
   (a)



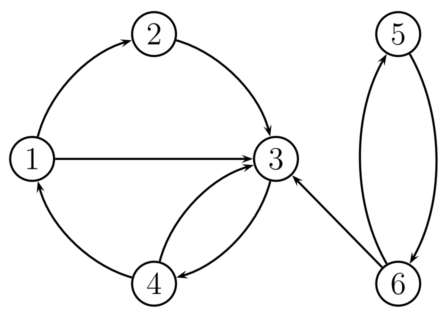
(b)



(c)



(d)



Here's what separates good giving–game templates, like the page 1 example, from the bad examples 5a,b,c,d.

Definition: A square matrix S is called *stochastic* if all its entries are positive, and the entries in each column add up to exactly one.

Definition: A square matrix A is *almost stochastic* if all its entries are non–negative, the entries in each column add up to one, and if there is a positive power k so that A^k is stochastic.

6. What do these definitions mean *vis-à-vis* play–doh distribution? Hint: if it all starts at position j, then the inital fraction vector $\mathbf{x}_0 = \mathbf{e_j}$, i.e. has a 1 in position j and zeroes elsewhere. After k steps, the material is distributed according to $A^k\mathbf{e_j}$, which is the j^{th} column of A^k.

Stage 3: Theoretical basis for Google page rank

Theorem. *(Perron–Frobenius) Let A be almost stochastic. Let $\mathbf{x}_0$ be any "fraction vector" i.e. all its entries are non–negative and their sum is one. Then the discrete dynamical system*

$$\mathbf{x}_k = A^k \mathbf{x}_0$$

has a unique limiting fraction vector $\mathbf{z}$, and each entry of $\mathbf{z}$ is positive. Furthermore, the matrix powers A^k converge to a limit matrix, each of whose columns are equal to $\mathbf{z}$.

proof: Let $A = [a_{ij}]$ be almost stochastic. We know, by "conservation of play–doh", that if $\mathbf{v}$ is a fraction vector, then so is $A\mathbf{v}$. As a warm–up for the full proof of the P.F. theorem, let's check this fact algebraically:

$$\begin{aligned}\sum_{i=1}^{n} (A\mathbf{v})_i &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} v_j = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij} v_j \\ &= \sum_{j=1}^{n} v_j \left(\sum_{i=1}^{n} a_{ij}\right) = \sum_{j=1}^{n} v_j = 1\end{aligned}$$

Thus as long as $\mathbf{x}_0$ is a fraction vector, so is each iterate $A^N\mathbf{x}_0$.

Since A is almost stochastic, there is a power l so that $S = A^l$ is stochastic. For any (large) N, write $N = kl + r$, where $N/l = k$ with remainder r, $0 \le r < l$. Then

$$A^N \mathbf{x}_0 = A^{kl+r}\mathbf{x}_0 = \left(A^l\right)^k A^r \mathbf{x}_0 = S^k A^r \mathbf{x}_0$$

As $N \to \infty$ so does k, and there are only l choices for $A^r\mathbf{x}_0$, $0 \le r \le l-1$. Thus if we prove the P.F. theorem for stochastic matrices S, i.e. $S^k\mathbf{y}_0$ has a unique limit independent of $\mathbf{y}_0$, then the more general result for almost stochastic A follows.

So let $S = [s_{ij}]$ be an $n \times n$ stochastic matrix, with each $s_{ij} \ge \varepsilon > 0$. Let 1 be the matrix for which each entry is 1. Then we may write:

$$B = S - \varepsilon 1; \quad S = B + \varepsilon 1. \tag{1}$$

Here $B = [b_{ij}]$ has non–negative entries, and each column of B sums to

$$1 - n\varepsilon := \mu < 1. \tag{2}$$

We prove the P.F. theorem in a way which reflects your page 1 experiment: we'll show that whenever $\mathbf{v}$ and $\mathbf{w}$ are fraction vectors, then $S\mathbf{v}$ and $S\mathbf{w}$ are geometrically closer to each other than were $\mathbf{v}$ and $\mathbf{w}$. Precisely, our "metric" for measuring the distance "d" between two fraction vectors is

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) := \sum_{i=1}^{n} |v_i - w_i|. \tag{3}$$

Here's the magic: if $\mathbf{v}$ is any fraction vector, then for the matrix 1, of ones,

$$(1\mathbf{v})_i = \sum_{j=1}^{n} 1 v_j = 1.$$

So if $\mathbf{v}, \mathbf{w}$ are both fraction vectors, then $1\mathbf{v} = 1\mathbf{w}$. Using matrix and vector algebra, we compute using equations (1), (2):

$$\begin{aligned} S\mathbf{v} - S\mathbf{w} &= (B + \varepsilon 1)\mathbf{v} - (B + \varepsilon 1)\mathbf{w} \\ &= B(\mathbf{v} - \mathbf{w}) \end{aligned} \tag{4}$$

So by equation (3),

$$\begin{aligned} \mathrm{d}(S\mathbf{v}, S\mathbf{w}) &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} b_{ij}(v_j - w_j) \right| \\ &\leq \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} |v_j - w_j| \\ &= \sum_{j=1}^{n} |v_j - w_j| \sum_{i=1}^{n} b_{ij} \\ &= \mu \sum_{j=1}^{n} |v_j - w_j| \\ &= \mu \mathrm{d}(\mathbf{v}, \mathbf{w}) \end{aligned} \tag{5}$$

Iterating inequality (5) yields

$$\mathrm{d}(S^k \mathbf{v}, S^k \mathbf{w}) \leq \mu^k \mathrm{d}(\mathbf{v}, \mathbf{w}). \tag{6}$$

Since fraction vectors have non–negative entries which sum to 1, the greatest distance between any two fraction vectors is 2:

$$\mathrm{d}(\mathbf{v}, \mathbf{w}) = \sum_{i=1}^{n} |v_i - w_i] \leq \sum_{i=1}^{n} v_i + w_i = 2$$

So, no matter what different initial fraction vectors experimenters begin with, after k iterations the resulting fraction vectors are within $2\mu^k$ of each other, and by choosing k large enough, we can deduce the existence of, and estimate the common limit $\mathbf{z}$ with as much precision as desired. Furthermore, if all initial material is allotted to node j, then the initial fraction vector $\mathbf{e_j}$ has a 1 in position j and zeroes elsewhere. $S^k\mathbf{e_j}$, (or $A^N\mathbf{e_j}$) is on one hand the j^{th} column of S^k (or A^N), but on the other hand is converging to $\mathbf{z}$. So each column of the limit matrix for S^k and A^N equals $\mathbf{z}$. Finally, if $\mathbf{x}_0$ is any initial fraction vector, then $S(S^k\mathbf{x}_0) = S^{k+1}(\mathbf{x}_0)$ is converging to $S(\mathbf{z})$ and also to $\mathbf{z}$, so $S(\mathbf{z}) = \mathbf{z}$ (and $A\mathbf{z} = \mathbf{z}$). Since the entries of $\mathbf{z}$ are non–negative (and sum to 1) and the entries of S are all positive, the entries of $S\mathbf{z}$ $(= \mathbf{z})$ are all positive. ■

Stage 4: The Google fudge factor

Sergey Brin and Larry Page realized that the world wide web is not almost stochastic. However, in addition to realizing that the Perron–Frobenius theorem was potentially useful for ranking URLs, they figured out a simple way to guarantee stochasticity—the "Google fudge factor."

Rather than using the voting matrix A described in the previous stages, they take a combination of A with the matrix of 1s we called $\mathbb{1}$. For (Brin an Pages' choice of) $\varepsilon = .15$ and n equal the number of nodes, consider the Google matrix

$$G = (1 - \varepsilon)A + \frac{\varepsilon}{n}\mathbb{1}.$$

(See [Austin, 2008]).

If A is almost stochastic, then each column of G also sums to 1 and each entry is at least ε/n. This G is stochastic! In other words, if you use this transition matrix everyone gets a piece of your play–doh, but you still get to give more to your friends.

7. Consider the giving game from 5c. Its transition matrix

$$A = \left[\begin{array}{cccc|cc} 0 & 0 & 0 & .5 & 0 & 0 \\ .5 & 0 & 0 & 0 & 0 & 0 \\ .5 & 1 & 0 & .5 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

is not almost stochastic. For $\varepsilon = .3$ and $\varepsilon/n = .05$, work out the Google matrix G, along with the limit rankings for the six sites. If you were upset that site 4 was ranked as equal to site 3 in the game you played for stage 1, you may be happier now.

Historical notes

The Perron–Frobenius theorem had historical applications to input–output economic modeling. The idea of using it for ranking seems to have originated with Joseph B. Keller, a Stanford University emeritus mathematics professor. According to a December 2008 article in the Stanford Math Newsletter [Keller, 2008], Professor Keller originally explained his team ranking algorithm in the 1978 Courant Institute Christmas Lecture, and later submitted an article to Sports Illustrated in which he used his algorithm to deduce unbiased rankings for the National League baseball teams at the end of the 1984 season. His article was rejected. Utah professor James Keener visited Stanford in the early 1990s, learned of Joe Keller's idea, and wrote a SIAM article in which he ranked football teams [Keener, 1993].

Keener's ideas seem to have found their way into some of the current BCS college football ranking schemes which often cause boosters a certain amount of heartburn. I know of no claim that there is any direct path from Keller's original insights, through Keener's paper, to Brin and Pages' amazing Google success story. Still it is interesting to look back and notice that the seminal idea had been floating “in the air” for a number of years before it occurred to anyone to apply it to Internet searches.

Acknowledgement: Thanks to Jason Underdown for creating the graph diagrams and for typesetting this document in LaTeX.

References

David D. Austin. How Google Finds Your Needle in the Web’s Haystack. 2008. URL http://www.ams.org/featurecolumn/archive/pagerank.html.

Sergey Brin and Lawrence Page. The anatomy of a large–scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 33:107–117, 1998. URL http://infolab.stanford.edu/pub/papers/google.pdf.

James Keener. The Perron–Frobenius Theorem and the ranking of football teams. *SIAM Rev.*, 35:80–93, 1993.

Joseph B. Keller. Stanford University Mathematics Department newsletter, 2008.

Pac 12 football rankings as of last week and based only on games played between two Pac 12 teams:

P:

	1	2	3	4	5	6	7	8	9	10	11	12	
	Az	Az St.	Cal	Col	Or	Or St.	Stan	UCLA	USC	UT	Wa	Wa St	
1	0	0	0.3333	0	0	0.2500	0	0	0	0	0	0	Az
2	0	0	0	0	0	0.2500	0	0	0	0	0	0	Az St
3	0	0	0	0	0	0.2500	0	0	0	0	0	0	Cal
4	0	0.3333	0	0	0	0	0	0.5000	0	0	0	0	Col
5	0	0	0.3333	0	0	0	0	0	0	0	1	0	Or
6	0	0	0	0	0	0	0	0	0	0	0	0	Or St
7	0	0.3333	0	0	0.5000	0	0	0	0.5000	0	0	0	Stan
8	0.3333	0	0.3333	0	0	0	0	0	0	0	0	0	UCLA
9	0.3333	0	0	0.5000	0	0	0	0	0	0	0	1	USC
10	0.3333	0	0	0	0	0	1	0	0.5000	0	0	0	UT
11	0	0.3333	0	0.5000	0	0	0	0.5000	0	0.5000	0	0	Wa
12	0	0	0	0	0.5000	0.2500	0	0	0	0.5000	0	0	Wa St

$$S := \left(\frac{.15}{12}\right)[1] + .85\,P:$$

12x12 double

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0125	0.0125	0.2958	0.0125	0.0125	0.2250	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
2	0.0125	0.0125	0.0125	0.0125	0.0125	0.2250	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
3	0.0125	0.0125	0.0125	0.0125	0.0125	0.2250	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
4	0.0125	0.2958	0.0125	0.0125	0.0125	0.0125	0.0125	0.4375	0.0125	0.0125	0.0125	0.0125
5	0.0125	0.0125	0.2958	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.8625	0.0125
6	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
7	0.0125	0.2958	0.0125	0.0125	0.4375	0.0125	0.0125	0.0125	0.4375	0.0125	0.0125	0.0125
8	0.2958	0.0125	0.2958	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125
9	0.2958	0.0125	0.0125	0.4375	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.8625
10	0.2958	0.0125	0.0125	0.0125	0.0125	0.0125	0.8625	0.0125	0.4375	0.0125	0.0125	0.0125
11	0.0125	0.2958	0.0125	0.4375	0.0125	0.0125	0.0125	0.4375	0.0125	0.4375	0.0125	0.0125
12	0.0125	0.0125	0.0125	0.0125	0.4375	0.2250	0.0125	0.0125	0.0125	0.4375	0.0125	0.0125

S^{20}:

	1	2	3	4	5	6	7	8	9	10	11	12	
1	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	0.0195	Az
2	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	Az St
3	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	0.0152	Cal
4	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	0.0263	Col
5	0.1210	0.1210	0.1210	0.1210	0.1210	0.1210	0.1209	0.1210	0.1210	0.1210	0.1209	0.1210	Or
6	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	0.0125	Or St
7	0.1355	0.1354	0.1354	0.1354	0.1353	0.1354	0.1355	0.1354	0.1354	0.1353	0.1355	0.1354	Stan
8	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	0.0223	UCLA
9	0.1581	0.1582	0.1581	0.1582	0.1582	0.1581	0.1581	0.1582	0.1581	0.1582	0.1581	0.1581	USC
10	0.2003	0.2003	0.2003	0.2003	0.2004	0.2003	0.2003	0.2003	0.2003	0.2004	0.2003	0.2003	UT
11	0.1226	0.1225	0.1226	0.1226	0.1226	0.1226	0.1226	0.1226	0.1225	0.1226	0.1226	0.1226	Wa
12	0.1517	0.1517	0.1517	0.1517	0.1517	0.1517	0.1518	0.1517	0.1517	0.1517	0.1518	0.1517	Wa St

Part 2 Monday
Eigenvalues and eigenvectors for square matrices, sections 5.1-5.2

The steady state vectors for stochastic matrices in section 4.9, i.e. the vectors $\underline{\boldsymbol{x}}$ with $P(\underline{\boldsymbol{x}}) = \underline{\boldsymbol{x}}$ when P is stochastic, are a special case of the concept of eigenvectors and eigenvalues for general square matrices, as we'll see below.

To introduce the general idea of eigenvalues and eigenvectors we'll first think geometrically.

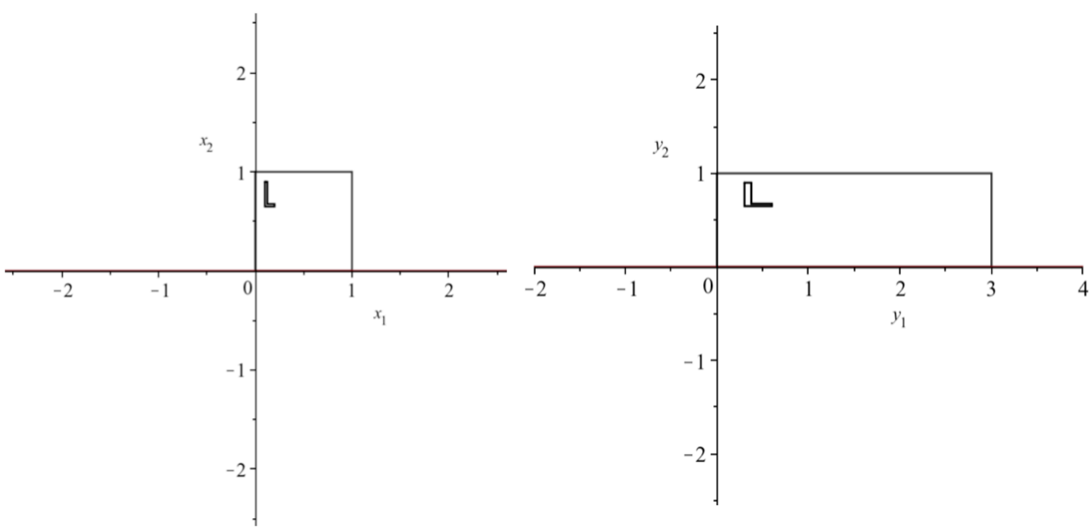
<u>Example</u> Consider the matrix transformation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ with formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1\begin{bmatrix} 3 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Notice that for the standard basis vectors $\underline{\boldsymbol{e}}_1 = [1, 0]^T, \underline{\boldsymbol{e}}_2 = [0, 1]^T$

$$T(\underline{\boldsymbol{e}}_1) = 3\underline{\boldsymbol{e}}_1$$
$$T(\underline{\boldsymbol{e}}_2) = \underline{\boldsymbol{e}}_2 \ .$$

In other words, T stretches by a factor of 3 in the $\underline{\boldsymbol{e}}_1$ direction, and by a factor of 1 in the $\underline{\boldsymbol{e}}_2$ direction, transforming a square grid in the domain into a parallel rectangular grid in the image:



Definition: If $A_{n \times n}$ and if $A\,\underline{\boldsymbol{v}} = \lambda\,\underline{\boldsymbol{v}}$ for a scalar λ and a vector $\underline{\boldsymbol{v}} \neq \underline{\boldsymbol{0}}$ then $\underline{\boldsymbol{v}}$ is called an eigenvector of A, and λ is called the eigenvalue of $\underline{\boldsymbol{v}}$. (In some texts the words characteristic vector and characteristic value are used as synonyms for these words.)

• In the example above, the standard basis vectors (or multiples of them) are eigenvectors, and the corresponding eigenvalues are the diagonal matrix entries. A non-diagonal matrix may still have eigenvectors and eigenvalues, and this geometric information can still be important to find. (For example, a stochastic matrix P always has eigenvectors with eigenvalue 1, namely the steady-state vector and its multiples. But how do you find eigenvectors and eigenvalues for general non-diagonal matrices?

Exercise 1) Try to find eigenvectors and eigenvalues for the non-diagonal matrix, by just trying random input vectors $\underline{\boldsymbol{x}}$ and computing $A\,\underline{\boldsymbol{x}}$.

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}.$$

How to find eigenvalues and eigenvectors (including eigenspaces) systematically:

If

$$A\,\underline{\boldsymbol{v}} = \lambda\,\underline{\boldsymbol{v}}$$

$$\Leftrightarrow A\,\underline{\boldsymbol{v}} - \lambda\,\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

$$\Leftrightarrow A\,\underline{\boldsymbol{v}} - \lambda\,I\,\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

where I is the identity matrix.

$$\Leftrightarrow (A - \lambda\,I)\underline{\boldsymbol{v}} = \underline{\mathbf{0}}\,.$$

Unlike in section 4.9 where the stationary vector was an eigenvector with eigenvalue 1, we don't have a clue as to what the eigenvalues of A are, in general. But we can figure that out using what we know about determinants! As we know, this last equation can have non-zero solutions $\underline{\boldsymbol{v}}$ if and only if the matrix $(A - \lambda\,I)$ is not invertible, i.e.

$$\Leftrightarrow det(A - \lambda\,I) = 0\,.$$

So, to find the eigenvalues and eigenvectors of matrix you can proceed as follows:

- Compute the polynomial in λ

$$p(\lambda) = det(A - \lambda\,I)\,.$$

If $A_{n \times n}$ then $p(\lambda)$ will be degree n. This polynomial is called the characteristic polynomial of the matrix A.

- λ_j can be an eigenvalue for some non-zero eigenvector $\underline{\boldsymbol{v}}$ if and only if it's a root of the characteristic polynomial, i.e. $p(\lambda_j) = 0$. For each such root, the homogeneous solution space of vectors $\underline{\boldsymbol{v}}$ solving

$$(A - \lambda_j\,I)\underline{\boldsymbol{v}} = \underline{\mathbf{0}}$$

i.e. by finding

$$Nul\,(A - \lambda_j\,I).$$

This subspace of eigenvectors will be at least one dimensional, since $(A - \lambda_j\,I)$ does not reduce to the identity. Find a basis of eigenvectors for this subspace. Follow this procedure for each eigenvalue, i.e. for each root of the characteristic polynomial.

Notation: The subspace of eigenvectors for eigenvalue λ_j is called the λ_j eigenspace, and we'll denote it by $E_{\lambda = \lambda_j}$. The basis of eigenvectors is called an eigenbasis for E_{λ_j}.

Exercise 2) a) Use the systematic algorithm to find the eigenvalues and eigenbases for the non-diagonal matrix of Exercise 2.

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}.$$

b) Use your work to describe the geometry of the linear transformation in terms of directions that get scaled:

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Tues Oct 30

- 5.1-5.2 finding matrix eigenvalues and eigenvectors via the characteristic equation

Announcements:

Warm-up Exercise:

Exercise 1) Find the eigenvalues and eigenspace bases for

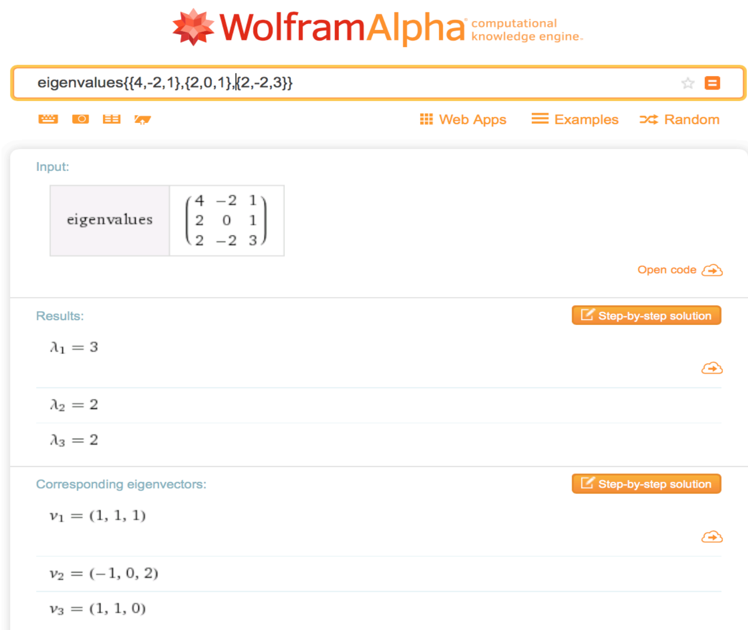
$$B := \begin{bmatrix} 4 & -2 & 1 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix}.$$

(i) Find the characteristic polynomial and factor it to find the eigenvalues. ($p(\lambda) = -(\lambda - 2)^2(\lambda - 1)$)

(ii) for each eigenvalue, find bases for the corresponding eigenspaces.

(iii) Can you describe the transformation $T(\underline{\mathbf{x}}) = B\underline{\mathbf{x}}$ geometrically using the eigenbases? Does $det(B)$ have anything to do with the geometry of this transformation?

Your solution will be related to the output below:



It often turns out that by collecting bases from each eigenspace for the matrix $A_{n \times n}$, and putting them together, we get a basis for $\mathbb{R}^n$. This lets us understand the <u>geometry</u> of the transformation

$$T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$$

almost as well as if A is a diagonal matrix. It does not always happen that the matrix A an basis of $\mathbb{R}^n$ made consisting of eigenvectors for A. (Even when all the eigenvalues are real.) When it does happen, we say that A is <u>diagonalizable.</u>

There are situations where we are guaranteed a basis of $\mathbb{R}^n$ made out eigenvectors of A:

<u>Theorem 1</u>: Let A be an $n \times n$ matrix with distinct real eigenvalues $\lambda_1, \lambda_2, ... \lambda_n$. Let $\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, ... \underline{\boldsymbol{v}}_n$ be corresponding (non-zero) eigenvectors, $A\,\underline{\boldsymbol{v}}_j = \lambda_j\,\underline{\boldsymbol{v}}_j$. Then the set

$$\{\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \cdots \underline{\boldsymbol{v}}_n\}$$

is linearly independent, and so is a basis for $\mathbb{R}^n$.....this is one theorem we can prove!

Exercise 2) Find the eigenvalues and eigenspace bases for the matrix below, and explain why there is no basis for $\mathbb{R}^2$ consisting of eigenvectors for this matrix:

$$C = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix}.$$

Theorem 2

Let $A_{n \times n}$ have factored characteristic polynomial

$$p(\lambda) = (-1)^n (\lambda - \lambda_1)^{k_1} (\lambda - \lambda_2)^{k_2} \ldots (\lambda - \lambda_m)^{k_m}$$

where like terms have been collected so that each λ_j is distinct (i.e different). Notice that

$$k_1 + k_2 + \ldots + k_m = n$$

because the degree of $p(\lambda)$ is n.

- Then $1 \leq dim\left(E_{\lambda = \lambda_j}\right) \leq k_j$. If $dim\left(E_{\lambda = \lambda_j}\right) < k_j$ then the λ_j eigenspace is called defective.
- The matrix A is diagonalizable if and only if each $dim\left(E_{\lambda = \lambda_j}\right) = k_j$. In this case, one obtains an $\mathbb{R}^n$

eigenbasis simply by combining bases for each eigenspace into one collection of n vectors. (The same definitions and reasoning can apply to complex eigenvalues and eigenvectors, and a basis of $\mathbb{C}^n$.)

(The proof of this theorem is fairly involved. It is illustrated in a positive way by Exercise 1, and in a negative way by Exercise 2.)

Wed Oct 31

- 5.3 diagonalizable matrices

Announcements:

Warm-up Exercise:

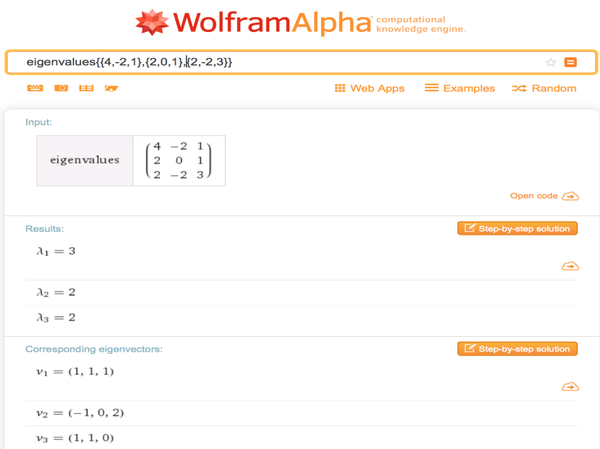
Continuing with the example from yesterday ...

If, for the matrix $A_{n \times n}$, there is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A, then we can understand the geometry of the transformation

$$T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$$

almost as well as if A is a diagonal matrix, and so we call such matrices *diagonalizable*. Having such a basis of eigenvectors for a given matrix is also extremely useful for algebraic computations, and will give another reason for the word *diagonalizable* to describe such matrices.

Use an $\mathbb{R}^3$ basis made of out eigenvectors of the matrix B in Exercise 1, yesterday, and put them into the columns of a matrix we will call P. We could order the eigenvectors however we want, but we'll put the $E_{\lambda=2}$ basis vectors in the first two columns, and the $E_{\lambda=3}$ basis vector in the third column:



$$P := \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & -2 & 1 \end{bmatrix}.$$

Now do algebra (check these steps and discuss what's going on!)

$$\begin{bmatrix} 4 & -2 & 1 \\ 2 & 0 & 1 \\ 2 & -2 & 3 \end{bmatrix} \left[\begin{array}{c|c|c} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & -2 & 1 \end{array} \right]$$

$$= \left[\begin{array}{c|c|c} 0 & 2 & 3 \\ 2 & 0 & 3 \\ 4 & -4 & 3 \end{array} \right]$$

$$= \left[\begin{array}{c|c|c} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2 & -2 & 1 \end{array} \right] \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

In other words,

$$B\,P = P\,\mathrm{D}\,,$$

where D is the diagonal matrix of eigenvalues (for the corresponding columns of eigenvectors in P).
Equivalently (multiply on the right by P^{-1} or on the left by P^{-1}):

$$B = P\,\mathrm{D}\,P^{-1} \text{ and } P^{-1}BP = \mathrm{D}.$$

Exercise 1) Use one of the the identities above to show how B^{100} can be computed with only two matrix multiplications!

Definition: Let $A_{n \times n}$. If there is an $\mathbb{R}^n$ (or $\mathbb{C}^n$) basis $\underline{\boldsymbol{v}}_1, \underline{\boldsymbol{v}}_2, \ldots, \underline{\boldsymbol{v}}_n$ consisting of eigenvectors of A, then A is called diagonalizable. This is precisely why:

Write $A\,\underline{\boldsymbol{v}}_j = \lambda_j\,\underline{\boldsymbol{v}}_j$ (some of these λ_j may be the same, as in the previous example). Let P be the matrix

$$P = \left[\underline{\boldsymbol{v}}_1 | \underline{\boldsymbol{v}}_2 | \cdots | \underline{\boldsymbol{v}}_n\right].$$

Then, using the various ways of understanding matrix multiplication, we see

$$A\,P = A\left[\underline{\boldsymbol{v}}_1 | \underline{\boldsymbol{v}}_2 | \cdots | \underline{\boldsymbol{v}}_n\right] = \left[\lambda_1 \underline{\boldsymbol{v}}_1 | \lambda_2 \underline{\boldsymbol{v}}_2 | \cdots | \lambda_n \underline{\boldsymbol{v}}_n\right]$$

$$= \left[\underline{\boldsymbol{v}}_1 | \underline{\boldsymbol{v}}_2 | \cdots | \underline{\boldsymbol{v}}_n\right] \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ : & : & \ldots & : \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}.$$

$$A\,P = P\,\mathrm{D}$$

$$A = P\,\mathrm{D}\,P^{-1}$$

$$P^{-1} A\,P = \mathrm{D}\,.$$

Unfortunately, as we've already seen, not all matrices are diagonalizable:

Exercise 2) Show that

$$C := \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is not diagonalizable. (Even though it has the same characteristic polynomial as B, which was diagonalizable.

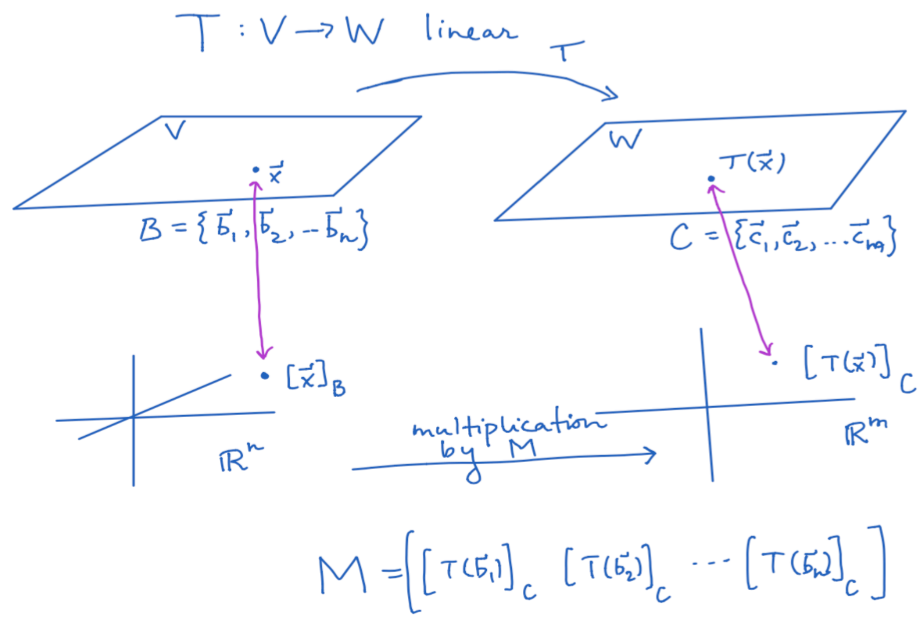
Fri Nov 2

- 5.4 eigenvalues, eigenvectors and linear transformations

Announcements:

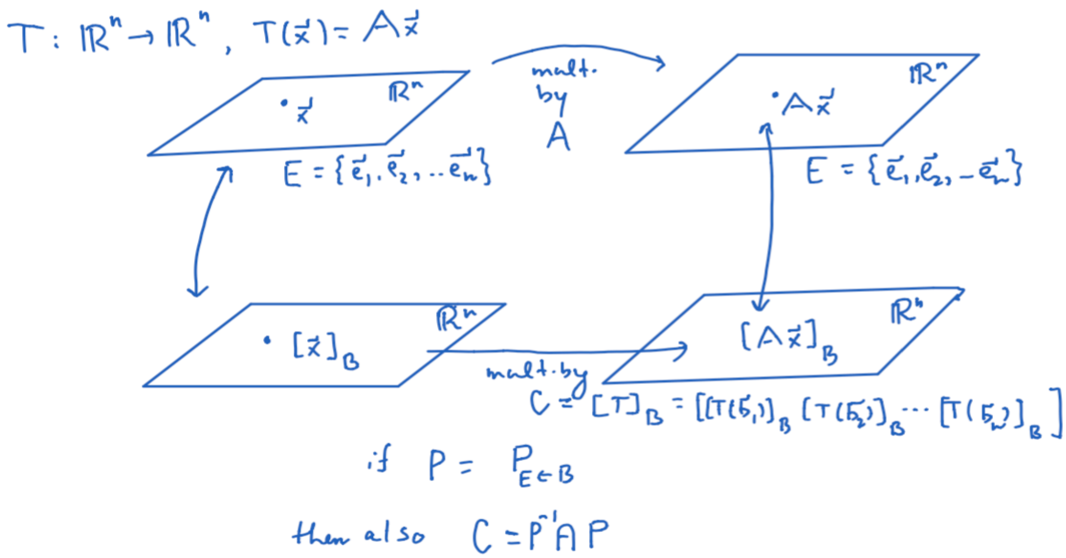
Warm-up Exercise:

If we have a linear transformation $T : V \rightarrow W$ and bases $B = \{\underline{b}_1, \underline{b}_2, \ldots \underline{b}_n\}$ in V, $C = \{\underline{c}_1, \underline{c}_2, \ldots \underline{c}_m\}$ in W, then the matrix of T with respect to these two bases transforms the B coordinates of vectors $\underline{v} \in V$ to the C coordinates of $T(\underline{v})$ in a straightforward way, although it takes a while to get used to:



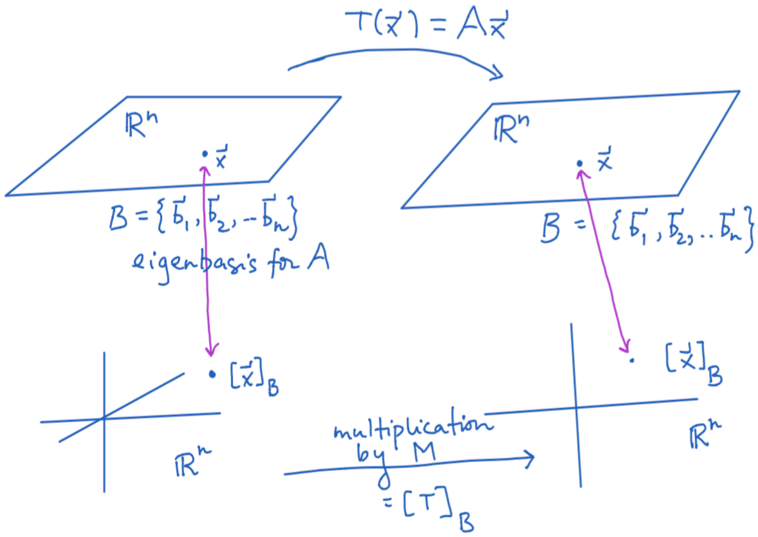
Exercise 1) Let $V = P_3 = span\{1, t, t^2, t^3\}$, $W = P_2 = span\{1, t, t^2\}$, and let $D : V \rightarrow W$ be the derivative operator. Find the matrix of D with respect to the bases $\{1, t, t^2, t^3\}$ in V and $\{1, t, t^2\}$ in W. Test your result.

A special case of the previous page is when $T : \mathbb{R}^n \to \mathbb{R}^n$ is a matrix transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$, and we find the matrix of T with respect to a non-standard basis (the same non-standard basis in the domain and in the codomain).



<u>Definition</u> Two matrices A, B are called *similar* if there is an invertible matrix P with $B = P^{-1}AP$. As the diagram above shows, similar matrices arise when one is describing the same linear transformation, but with respect to different bases.

Exercise 2) What if a matrix A is diagonalizable? What is the matrix of $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$ with respect to the eigenbasis? How does this connect to our matrix identities for diagonalization? Fill in the matrix M below, and then compute another way to express it, as a triple product using the diagram.

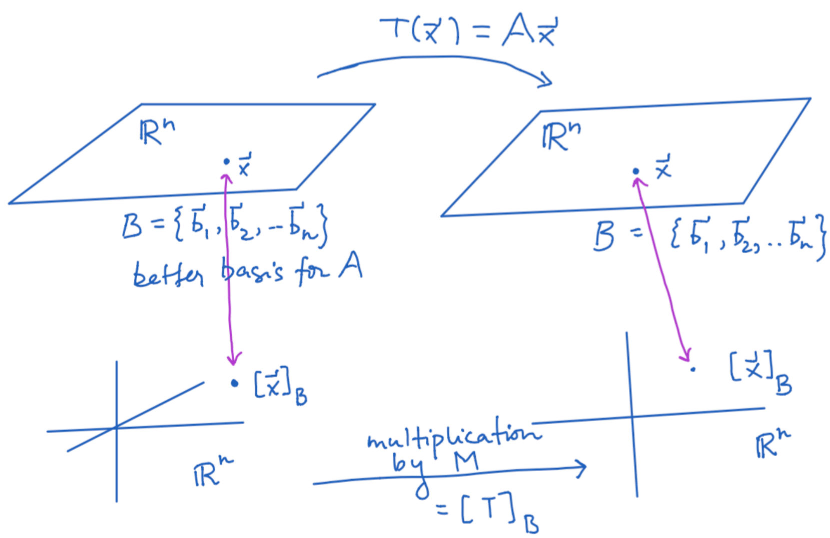


Example, from earlier this week:

$$A = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix} \qquad E_{\lambda=4} = span\left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\} \qquad E_{\lambda=1} = span\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\} \qquad B = \left\{\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$$

Write the various matrices corresponding to the diagram above.

Even if the matrix A is not diagonalizable, there may be a better basis to help understand the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$. The diagram on the previous page didn't require that B be a basis of eigenvectors....maybe it was just a "better" basis than the standard basis, to understand T.



<u>Exercise 3</u> (If we have time - this one is not essential.) Try to pick a better basis to understand the matrix transformation $T(\underline{\mathbf{x}}) = C\,\underline{\mathbf{x}}$, even though the matrix C is not diagonalizable. Compute $M = P^{-1} A\, P$ or compute M directly, to see if it really is a "better" matrix.

$$C = \begin{bmatrix} 4 & 4 \\ -1 & 0 \end{bmatrix}$$