Goal: singular value decomposition!

Recall the (fundamental) Spectral Theorem, which says that for any symmetric $n \times n$ matrix A, there is an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathbb{R}^n$ in terms of which the effect of A looks very simple: A simply "stretches" each $\mathbf{v}_i$ by the corresponding eigenvalue. Applications of eigenvalues and eigenvectors are pervasive throughout the natural sciences, data science, computer science, economics, statistics, and so on. These concepts are truly one of the most important things you learn in Math 51.

There are efficient and stable algorithms that compute the λ_i's and $\mathbf{v}_i$'s for a given symmetric matrix A, and we can use this to easily describe how A acts on any vector $\mathbf{v}$: writing $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ for scalars c_i given by the Fourier formula, we have

$$A\mathbf{v} = c_1 A\mathbf{v}_1 + \cdots + c_n A\mathbf{v}_n = c_1\lambda_1\mathbf{v}_1 + \cdots + c_n\lambda_n\mathbf{v}_n.$$

If Q is the matrix whose columns are the $\mathbf{v}_i$'s, then Q is orthogonal, and the Spectral Theorem tells us that

$$A = QDQ^{\intercal} = QDQ^{-1},$$

where D is the diagonal matrix whose ith diagonal entry is λ_i.

Unfortunately, an orthogonal eigenvector decomposition of $\mathbb{R}^n$ to describe the effect of A is only available when A is symmetric. It turns out that for a typical non-symmetric $n \times n$ matrix A, there is no basis of $\mathbb{R}^n$ consisting of eigenvectors. Moreover, even when such a basis exists, the algorithms for finding eigenvectors are numerically stable. Furthermore, this circle of ideas applies to square matrices only.

There is a remarkable, and closely related, idea called the **singular value decomposition (SVD)**. This works for ***ALL*** matrices A of any size and shape, and there are good stable algorithms for computation. Furthermore, if A is a symmetric matrix, the SVD is the matrix decomposition above up to some sign issues when there are negative eigenvalues. For this reason, SVD is regarded, by almost everyone, who uses applied linear algebra as ***THE*** primary method for doing computations with a matrix of interest.

The basic idea is a remarkably simple tweak of what we have already done. If A is an $m \times n$ matrix, unless $m = n$ it is impossible to find a common basis in the source and target spaces as the dimensions are different. Once we realize this, it is not a huge leap to wonder if we might find potentially quite different orthonormal bases for the input and output in terms of which the effect of A looks nice; even when $m = n$, this is a new idea!

To be precise, the main idea behind SVD is that we look for an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbb{R}^n$ and an orthonormal basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ of $\mathbb{R}^m$ so that for each $1 \le i \le n$,

$$A\mathbf{v}_i = \sigma_i\mathbf{w}_i$$

for some scalar $\sigma_i \ge 0$. This does not quite make sense if $m < n$, since $\mathbf{w}_i$ has no meaning when $i > m$, so in such cases we make the convention that $\mathbf{w}_i = \mathbf{0}$. In particular, $\{\mathbf{v}_1, \ldots \mathbf{v}_n\}$ and $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ would *both* be orthogonal collections of vectors, which would be rather astounding to find for a general A. The scalars σ_i are called the **singular values** of A.

An $m \times n$ matrix $D = (d_{ij})$ is called **diagonal** if the entries d_{ij} vanish whenever $i \neq j$. For example,

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \end{bmatrix}$$

are both diagonal.

Theorem 27.3.3. **(Singular Value Decomposition)** For ***every*** $m \times n$ matrix A, we can find:

- an $m \times n$ diagonal matrix D which has diagonal entries $d_{ii} = \sigma_i$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$;
- an $m \times m$ orthogonal matrix Q and an $n \times n$ orthogonal matrix Q' for which

$$A = QDQ'^{\mathsf{T}}.$$

The diagonal matrix D is uniquely determined; the numbers σ_i are called the **singular values** of A.

Example 1. Consider the matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix}$. The SVD of A is

$$A = QDQ'^{\mathsf{T}} = \begin{array}{c} \vec{w}_1 \quad \vec{w}_2 \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \sigma_1 \quad \sigma_2 \\ \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{6} & 0 \end{bmatrix} \end{array} \begin{array}{c} \begin{bmatrix} -\sqrt{2/7} & 1/\sqrt{14} & 3/\sqrt{14} \\ \sqrt{2/3} & 1/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{21} & -4/\sqrt{21} & 2/\sqrt{21} \end{bmatrix} \end{array} \begin{array}{c} \vec{v}_1 \\ \vec{v}_2 \\ \vec{v}_3 \end{array}$$

$$A\vec{v}_1 = \begin{bmatrix} 2 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \end{bmatrix} = \sqrt{14} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \sigma_1 \vec{w}_1$$

$$A\vec{v}_2 = \begin{bmatrix} 2 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix} = \sqrt{6} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \sigma_2 \vec{w}_2$$

$$A\vec{v}_3 = \begin{bmatrix} 2 & 1 & 1 \\ -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{21}} \\ -\frac{4}{\sqrt{21}} \\ \frac{2}{\sqrt{21}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vec{0}$$

Example 2. Consider the matrix $A = \begin{bmatrix} 2 & -1 \\ -1 & 3 \\ -1 & 1 \end{bmatrix}$. The SVD of A is

$$A = QDQ'^{\mathsf{T}} = \begin{array}{c} \vec{w}_1 \qquad \vec{w}_2 \qquad \vec{w}_3 \\ \begin{bmatrix} -7/\sqrt{195} & 2\sqrt{2/13} & \sqrt{2/15} \\ 11/\sqrt{195} & 3/\sqrt{26} & -1/\sqrt{30} \\ \sqrt{5/39} & -1/\sqrt{26} & \sqrt{5/6} \end{bmatrix} \end{array} \begin{array}{c} \sigma_1 \quad \sigma_2 \\ \begin{bmatrix} \sqrt{15} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} -2/\sqrt{13} & 3/\sqrt{13} \\ 3/\sqrt{13} & 2/\sqrt{13} \end{bmatrix} \begin{array}{c} \vec{v}_1 \\ \vec{v}_2 \end{array}$$

$$A\vec{v}_1 = \begin{bmatrix} 2 & -1 \\ -1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{bmatrix} = \begin{bmatrix} -\frac{7}{\sqrt{13}} \\ \frac{11}{\sqrt{13}} \\ \frac{5}{\sqrt{13}} \end{bmatrix} = \sqrt{15} \begin{bmatrix} -\frac{7}{\sqrt{195}} \\ \frac{11}{\sqrt{195}} \\ \frac{5}{\sqrt{195}} \end{bmatrix} = \sqrt{15}\, \vec{w}_1$$

$$A\vec{v}_2 = \begin{bmatrix} 2 & -1 \\ -1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{3}{\sqrt{13}} \\ \frac{2}{\sqrt{13}} \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \\ -\frac{1}{\sqrt{13}} \end{bmatrix} = \sqrt{2} \begin{bmatrix} \frac{4}{\sqrt{26}} \\ \frac{3}{\sqrt{26}} \\ -\frac{1}{\sqrt{26}} \end{bmatrix} = \sqrt{2}\, \vec{w}_2$$

<u>Note</u> If we replace Q with $[\vec{w}_1 \;\; \vec{w}_2 \;\; -\vec{w}_3]$, it still holds that $A = QDQ'^{\mathsf{T}}$.

Example 3. (Rank reduction) The following problem is highly relevant for data compression.

Suppose that you have an $m \times n$ matrix A, where $m, n \sim 10^6$ and the individual entries a_{ij}'s correspond to some data. It takes a lot of memory to store and possibly manipulate $mn \approx 10^{12}$ numbers. Is there a way to avoid such storage problems?

The basic idea of data compression is that it would be very nice to somehow reduce this information to a much smaller set of numbers which still captures almost all of the features of the original data set (which itself is perhaps only a sampling of the true image). This sounds like a fantasy, but remarkably there are a number of very ingenious ideas which do it very effectively. We now explain the key underlying mathematical idea.

To motivate this, let us examine another way of interpreting the SVD $QDQ'^{\intercal}$ of an $m \times n$ matrix A. Suppose we write the columns of the "prefactor" orthogonal matrix $m \times m$ matrix Q as $\mathbf{w}_1, \dots, \mathbf{w}_m$ and the columns of the "postfactor" orthogonal $n \times n$ matrix Q' as $\mathbf{v}_1, \dots, \mathbf{v}_n$. Then, in the $m \leq n$ case,

$$A = QDQ'^{\intercal} = \begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_m \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \sigma_m & 0 \end{bmatrix} \begin{bmatrix} - & \mathbf{v}_1^{\intercal} & - \\ - & \mathbf{v}_2^{\intercal} & - \\ & \vdots & \\ - & \mathbf{v}_n^{\intercal} & - \end{bmatrix}$$

can be written as

$$A = \sigma_1 \mathbf{w}_1 \mathbf{v}_1^{\intercal} + \cdots + \sigma_m \mathbf{w}_m \mathbf{v}_m^{\intercal}.$$

Note that the vectors $\mathbf{v}_{m+1}, \dots, \mathbf{v}_n \in \mathbb{R}^n$ do not appear here. For similar reasons, if $m > n$, then this sum has n terms, and the vectors $\mathbf{w}_{n+1}, \dots, \mathbf{w}_m \in \mathbb{R}^m$ do not appear.

Now, what do these $m \times n$ matrices $\mathbf{w}_i \mathbf{v}_i^{\intercal}$'s look like? These are examples of **rank-1 matrices,** and they look like (assuming $m \leq n$)

$$\mathbf{w}_i \mathbf{v}_i^{\intercal} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \vdots & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

if $\mathbf{w}_i = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$ and $\mathbf{v}_i = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$. The special feature here is that *every* column is a multiple of the *same* vector $\mathbf{w}_i$ – the jth column is $b_j \mathbf{w}_i$. Now, let us look at $A\mathbf{x}$ – one of the matrix algebra properties we know is

$$(\mathbf{w}_i \mathbf{v}_i^{\intercal})\mathbf{x} = (\mathbf{v}_i \cdot \mathbf{x})\mathbf{w}_i.$$

Thus,

$$A\mathbf{x} = (\sigma_1 \mathbf{w}_1 \mathbf{v}_1^{\intercal} + \cdots + \sigma_m \mathbf{w}_m \mathbf{v}_m^{\intercal})\mathbf{x} = (\sigma_1 \mathbf{v}_1 \cdot \mathbf{x})\mathbf{w}_1 + \cdots + (\sigma_m \mathbf{v}_m \cdot \mathbf{x})\mathbf{w}_m.$$

Since the scalar $\mathbf{v}_i \cdots \mathbf{x}$ is the length of the orthogonal projection of $\mathbf{x}$ onto the *unit* vector $\mathbf{v}_i$, we are simply taking the lengths of the orthogonal projections of $\mathbf{x}$ along the m orthonormal unit vectors $\mathbf{v}_1, \dots, \mathbf{v}_m$ in $\mathbb{R}^n$ (these can be thought of as "coordinates" of the n-vector $\mathbf{x}$ in those m directions), multiplying these lengths by the scalars σ_i, and then using these as coefficients of an expansion in terms of the orthonormal basis $\mathbf{w}_1, \dots, \mathbf{w}_m$ of $\mathbb{R}^m$.

Suppose that only one of these terms appears in the SVD of A, i.e. $\sigma_2 = \cdots = \sigma_m = 0$. Then, $A = \sigma_1 \mathbf{w}_1 \mathbf{v}_1^{\intercal}$. How much "data" is on the right-hand side? Just $m + n + 1$ numbers – We have encoded the mn entries of A by using only $m + n + 1$ numbers! If $m, n \sim 10^9$, we only need to store around 10^9 numbers instead of 10^{18}.

It might seem weird to consider a situation where all but the first singular value vanish. However, for many matrices which *arise in practice*, most of the singular values are *relatively or very small*! For example, if the singular values $\sigma_{k+1}, \dots, \sigma_m$ are all quite small, then it is reasonable to just drop these to get a good approximation:

$$A = \sigma_1 \mathbf{w}_1 \mathbf{v}_1^{\intercal} + \cdots + \sigma_m \mathbf{w}_m \mathbf{v}_m^{\intercal} \approx \sigma_1 \mathbf{w}_1 \mathbf{v}_1^{\intercal} + \cdots + \sigma_k \mathbf{w}_k \mathbf{v}_k^{\intercal}.$$

Hence, we only need $k(m + n + 1)$ numbers instead of mn!

Example 4. Find the singular value decomposition of $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

① Compute $A^T A$.

$$A^T A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

② Find the eigenvalues and eigenvectors of $A^T A$.

We see that $p_{A^TA}(\lambda) = \lambda^2 - 4\lambda + 4$, and the 2-eigenspace is $\mathbb{R}^2$.

We can pick $\lambda_1 = 2$, $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\lambda_2 = 2$, $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. So, $Q' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

③ The singular values are the square roots of the eigenvalues.

So, $\sigma_1 = \sqrt{2}$ and $\sigma_2 = \sqrt{2}$.

④ Compute $\vec{w}_1, \vec{w}_2$.

$$A\vec{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \sqrt{2}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow \vec{w}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \quad \text{and} \quad A\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sqrt{2}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \Rightarrow \vec{w}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Hence, $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$, and a singular value decomposition for A is

$$\underbrace{\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{Q'^T}.$$

<u>Note</u> If we switch $\vec{v}_1$ and $\vec{v}_2$, we get another SVD

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Example 5. Find the singular value decomposition of $A = \begin{bmatrix} 5 & 1 \\ 5 & 1 \end{bmatrix}$.

$A^T A = \begin{bmatrix} 50 & 10 \\ 10 & 2 \end{bmatrix}$ has characteristic polynomial $p_{A^TA}(\lambda) = \lambda^2 - 52\lambda$, and so, the eigenvalues are $\lambda_1 = 52$ and $\lambda_2 = 0$ with corresponding eigenvectors $\vec{v}_1{}' = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ and $\vec{v}_2{}' = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, respectively. Hence, $Q' = \begin{bmatrix} \frac{5}{\sqrt{26}} & \frac{1}{\sqrt{26}} \\ \frac{1}{\sqrt{26}} & -\frac{5}{\sqrt{26}} \end{bmatrix}$ and $D = \begin{bmatrix} \sqrt{52} & 0 \\ 0 & 0 \end{bmatrix}$.

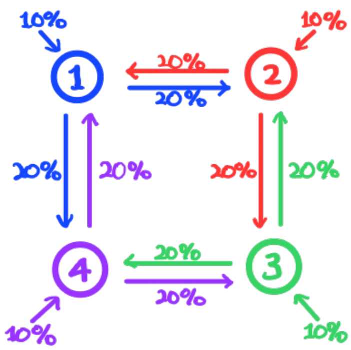
Next, we compute $A\vec{v}_1 = \sqrt{52}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and $A\vec{v}_2 = 0\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$, and so, $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$.

Thus, a SVD of A is

$$\begin{bmatrix} 5 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \sqrt{52} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \frac{5}{\sqrt{26}} & \frac{1}{\sqrt{26}} \\ \frac{1}{\sqrt{26}} & -\frac{5}{\sqrt{26}} \end{bmatrix}.$$

This column can be replaced by $\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

Example 6. Suppose that are several colonies of ants, one at each corner of a square. Initially there are 10,000 ants at one corner, and 1,000 ants at each of the other three corners.
Each colony grows by 10% each week and at each corner, there is also "diffusion": 20% of the original ant population (from the start of the week) leaves and moves clockwise to the next corner; another 20% of that original population leaves and moves counterclockwise to the next corner.
How many ants are there at each corner after 10 weeks? What can we say about the long-term behavior?



$$\text{Ant vector } \vec{a}_n = \begin{bmatrix} \#\text{ants at colony 1 after } n \text{ weeks} \\ \#\text{ants at colony 2 after } n \text{ weeks} \\ \#\text{ants at colony 3 after } n \text{ weeks} \\ \#\text{ants at colony 4 after } n \text{ weeks} \end{bmatrix}$$

$$\text{Then, } \vec{a}_{n+1} = \underbrace{\begin{bmatrix} 0.7 & 0.2 & 0 & 0.2 \\ 0.2 & 0.7 & 0.2 & 0 \\ 0 & 0.2 & 0.7 & 0.2 \\ 0.2 & 0 & 0.2 & 0.7 \end{bmatrix}}_{M} \vec{a}_n.$$

We are looking for $\vec{a}_{10} = M^{10}\vec{a}_0 = M^{10}(10000, 1000, 1000, 1000)$. It turns out that M has eigenvectors

$$\vec{w}_1 = \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \quad \vec{w}_2 = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{w}_3 = \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \quad \vec{w}_4 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

with eigenvalues $\lambda_1 = 0.3$, $\lambda_2 = 0.7$, $\lambda_3 = 0.7$, $\lambda_4 = 1.1$, respectively. By Fourier,

$$\vec{a}_0 = -4500\vec{w}_1 - 4500\vec{w}_2 + 4500\vec{w}_3 + 6500\vec{w}_4.$$

Thus,

$$\begin{aligned} \vec{a}_{10} &= -4500(0.3)^{10}\vec{w}_1 - 4500(0.7)^{10}\vec{w}_2 + 4500(0.7)^{10}\vec{w}_3 + 6500(1.1)^{10}\vec{w}_4 \\ &= (8556, 8430, 8303, 8430). \end{aligned}$$

A good estimate is

$$\vec{a}_{10} \approx 6500(1.1)^{10}\vec{w}_4 = (8429.66, 8429.66, 8429.66, 8429.66).$$

In situations involving raising a matrix to a power, *the eigenvalue that is largest* λ (in the sense of absolute value) *and a corresponding unit eigenvector* control the long-term behavior. The eigenvalue μ that is *second largest* (in absolute value) controls how *long* it takes before the eigenvalue λ "takes over." The closer $|\mu|$ is to $|\lambda|$, the longer it takes for the behavior of λ and a corresponding unit eigenvector to dominate.

Example 7. In a friendly soccer tournament between Argentina, Brazil, and Colombia, the results were

- A vs. B: $2-0$
- A vs. C: $3-1$
- B vs. C: $0-1$

We form a matrix

$$S = \begin{bmatrix} 0 & 1 & 3/4 \\ 0 & 0 & 0 \\ 1/4 & 1 & 0 \end{bmatrix}$$

with the ratios of team score over total score, where s_{ij} is the ratio of goals scored by team i in a match against team j. The eigenvalues of S are 0 and $\pm\sqrt{3}/4$.

Suppose $\mathbf{v} = \begin{bmatrix} v_A \\ v_B \\ v_C \end{bmatrix}$ is the eigenvector corresponding to the largest eigenvalue $\sqrt{3}/4$. Can you infer a ranking of the teams from the components of $\mathbf{v}$?

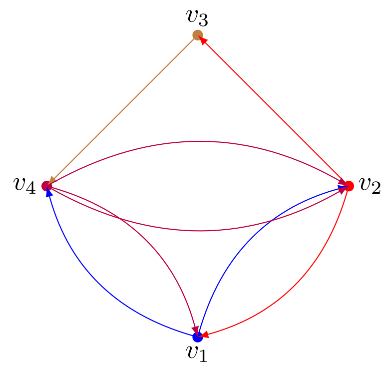
Computing $S - \frac{\sqrt{3}}{4} I_3 = \begin{bmatrix} -\frac{\sqrt{3}}{4} & 1 & \frac{3}{4} \\ 0 & -\frac{\sqrt{3}}{4} & 0 \\ \frac{1}{4} & 1 & -\frac{\sqrt{3}}{4} \end{bmatrix}$, we see that if $\begin{bmatrix} v_A \\ v_B \\ v_C \end{bmatrix} \in N\left(S - \frac{\sqrt{3}}{4} I_3\right)$,

$$\begin{aligned} -\tfrac{\sqrt{3}}{4} v_A + v_B + \tfrac{3}{4} v_C &= 0 \\ -\tfrac{\sqrt{3}}{4} v_B &= 0 \\ \tfrac{1}{4} v_A + v_B - \tfrac{\sqrt{3}}{4} v_C &= 0 \end{aligned}$$

which leads to $v_B = 0$ and $v_A = \sqrt{3}\, v_C$. Hence,

$$\vec{v} = v_C \begin{bmatrix} \sqrt{3} \\ 0 \\ 1 \end{bmatrix} \begin{matrix} \leftarrow \text{rank } 1: A \\ \leftarrow \text{rank } 3: B \\ \leftarrow \text{rank } 2: C \end{matrix}$$

Example 8. Consider the following graph, where the vertices are cities (resp. websites), and edges are flights (resp. links).



The adjacency matrix A has entries a_{ij}, which counts the number of edges from j to i.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}, \qquad W = \begin{bmatrix} 0 & 1/2 & 0 & 1/3 \\ 1/2 & 0 & 0 & 2/3 \\ 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1 & 0 \end{bmatrix}$$

W is obtained from A by dividing each column of A by its corresponding column sum.

(a) What does the ij-entry of A^2 signify?

$$A^2 = \begin{bmatrix} 2 & 0 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 2 & 0 & 1 \end{bmatrix}$$

The ij-entry of A^2 counts the number of paths of length 2 from v_j to v_i.

(b) In the Google PageRank algorithm (Appendix D), v_i is the value (or importance) of website i.

$$v_i = \sum_{j=1}^{n} w_{ij} v_j, \qquad \mathbf{v} = W\mathbf{v},$$

where $n = 4$ here. W is Markov and $\mathbf{v}$ is a 1-eigenvector of W.

$$W - I_4 = \begin{bmatrix} -1 & \frac{1}{2} & 0 & \frac{1}{3} \\ \frac{1}{2} & -1 & 0 & \frac{2}{3} \\ 0 & \frac{1}{2} & -1 & 0 \\ \frac{1}{2} & 0 & 1 & -1 \end{bmatrix} \Rightarrow \begin{bmatrix} 8 \\ 10 \\ 5 \\ 9 \end{bmatrix} \in N(W - I_4).$$

Value in decreasing order: v_2, v_4, v_1, v_3.