Lecture 3, addendum

The demo of image compression with SVD runs as follows:
The black and white picture is an *m x n* array **A** of pixel values for the brightness. The SVD may be written

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{j=1}^{r}\Delta\mathbf{A}_j ; \Delta\mathbf{A}_j = \sigma_j \mathbf{u}_j \mathbf{v}_j^T ; \mathbf{A}_k = \sum_{j=1}^{k}\Delta\mathbf{A}_j$$

where $\mathbf{u}_k$, $\mathbf{v}_k$ are the *k*:th columns of **U** and **V** and *r* is the rank of **A**. $\mathbf{A}_k$ is a matrix of rank *k*:
We will show below that $\mathbf{A}_k$ is the *m x n* matrix of rank at most *k* which best approximates **A** in the "Frobenius" norm,

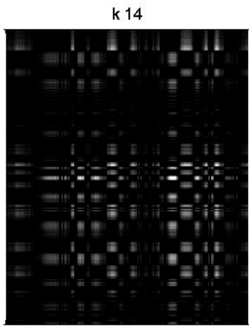
$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{tr(\mathbf{A}^T\mathbf{A})}$$

$$\mathbf{A}_k = \arg\min_{rank(\mathbf{X})\le k}\|\mathbf{A}-\mathbf{X}\|_F$$

which means that $\mathbf{A}_k$ is the rank-*k* matrix which minimizes the sum of squares of pixel differences to **A**. $\mathbf{A}_k$ requires only $k(m+n)$ to storage which may be a significant compression.
The demo shows **A**, $\mathbf{A}_k$, and the rank-1 matrix $\Delta\mathbf{A}_k$ (the latest term in the sum), for k = 1,2,… Here for $k = 14$:

A

$\mathbf{A}_{14}$



$\mathbf{B}_{14}$

The rank-1 character of $\Delta\mathbf{A}_{14}$ is revealed by its obvious row and column structure. The approximation error is in short wavelength features, it looks much better after blurring a little. This is easy to show with a projected image by turning the lens slightly out of focus … unless the projector has auto-focus like the one used in the class. I suspect that JPEG (lossy) compression does better than this.

Theorem
Let **A** be an *m x n* real matrix. The best approximant among rank-1 matrices in Frobenius norm is $\mathbf{A}_k$ defined above.
Proof

$$\|\mathbf{A}-\mathbf{X}\|_F^2 = tr((\mathbf{A}-\mathbf{X})^T(\mathbf{A}-\mathbf{X})) = tr(\mathbf{V}(\mathbf{S}-\mathbf{Y})^T \underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}}(\mathbf{S}-\mathbf{Y})\mathbf{V}^T), \mathbf{Y} = \mathbf{U}^T\mathbf{X}\mathbf{V}$$

Now the trace of a matrix equals the sum of its eigenvalues, so

$$tr(\mathbf{B}) = tr(\mathbf{S}\mathbf{B}\mathbf{S}^{-1})$$

since the latter matrix is a "similarity transformation" of B which has the same eigenvalues as B. There follows

$$\|\mathbf{A}-\mathbf{X}\|_F^2 = tr((\mathbf{S}-\mathbf{Y})^T(\mathbf{S}-\mathbf{Y})) = \sum_i(\sigma_i - y_{ii})^2 + \sum_{i\ne j} y_{ij}^2$$

and the minimum has $y_{ij} = 0$ for $i \ne j$. The SVD is defined to have

$$\sigma_1 > \sigma_2 > \sigma_3 > \ldots > \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \ldots = 0$$

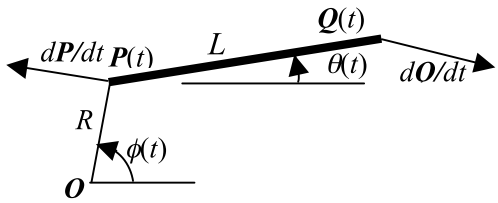
so the best approximant can annihilate the *k* first (the largest) and no more:

$$y_{ii} = \sigma_i, i = 1,2,\dots,k, y_{ii} = 0, i = k+1, k+2,\dots$$

Going back to **X** from **Y** we obtain the formula desired.

Prize problem III
is about "backing a trailer". As most of us can testify, it is hard to back up onto a driveway with a trailer hitched to the car. A test with a stick held lightly at one end, so it is free to rotate, the other resting on the ground shows the sensitivity. It needs careful control to *push* the stick and it seems to want to be *pulled*. The mathematical model will be given here. Your job is to simplify it by choice of coordinates to a stick of length L with one end moving in a circle of radius R. If $R << L$ one may think that the other end of the stick will experience very small net movement. As will be shown by physical experiment in class, this is not so: the other end progresses around a big circle by small zig-zag almost radial movements. Demonstrate that this simple model does predict such behaviour, and in particular compute what the average angular velocity of the stick is after long time.

Model
The relations are that



The distance between ***P*** and ***Q*** is L:

$$d/dt(\mathbf{Q}-\mathbf{P})\cdot(\mathbf{Q}-\mathbf{P}) = 0$$

The velocity vector at Q must point to P:

$$d\mathbf{Q}/dt = \alpha(\mathbf{P}-\mathbf{Q})$$

so $\alpha = 1/L^2 \frac{d\mathbf{P}}{dt}\cdot(\mathbf{P}-\mathbf{Q})$, or

$$d\mathbf{Q}/dt = \frac{d\mathbf{P}/dt\cdot(\mathbf{Q}-\mathbf{P})}{L^2}(\mathbf{Q}-\mathbf{P})$$

It *may* be advantageous to introduce $\mathbf{W} = \mathbf{Q} - \mathbf{P}$:

$$d\mathbf{W}/dt = -d\mathbf{P}/dt + \frac{d\mathbf{P}/dt\cdot\mathbf{W}}{L^2}\mathbf{W} = -\underbrace{(1-\frac{1}{L^2}\mathbf{W}\mathbf{W}^T)}_{\mathbf{S}(t)}d\mathbf{P}/dt$$

$\mathbf{W}\mathbf{W}^T/L^2$ is a projection matrix, so **S** picks out the component of d**P**/dt orthogonal to **W** itself.
A very non-linear equation!
Your job:
With $\mathbf{P}(t) = R(\cos\omega t, \sin\omega t)^T$, so $\mathbf{W}(t) = L(\cos\theta(t), \sin\theta(t))$

1. Derive the differential equation for θ
2. Solve it, and find

$$\lim_{t\to\infty} \frac{\omega}{2\pi}\left(\theta(t+2\pi/\omega) - \theta(t)\right)$$

which is the average angular velocity of the stick.