1. Statistics and SVD

See Section 7.3 of Strang's book.

Suppose we do n trials of an experiment, and each trial results in a measurement $\boldsymbol{v}_i \in \mathbb{R}^m$. We obtain a data set which is an n-tuple of m-vectors: $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$.

- *Example.* On n different days, we measure the temperature and humidity. The measurement for day i is a two-vector $\boldsymbol{v}_i = (\text{temp on day } i, \text{humidity on day } i) \in \mathbb{R}^2$.

The *average* measurement is

$$\boldsymbol{\mu} = \frac{\boldsymbol{v}_1 + \cdots + \boldsymbol{v}_n}{n}.$$

Let A be the $m \times n$ matrix whose i-th column is $\boldsymbol{v}_i - \boldsymbol{\mu}$. (So each row of A sums to zero, by an earlier homework problem.) Then

$$\begin{aligned} S &= \frac{AA^\top}{n-1} \\ &= \frac{(\boldsymbol{v}_1 - \boldsymbol{\mu})(\boldsymbol{v}_1 - \boldsymbol{\mu})^\top + \cdots + (\boldsymbol{v}_n - \boldsymbol{\mu})(\boldsymbol{v}_n - \boldsymbol{\mu})^\top}{n-1} \end{aligned}$$

is the *sample covariance matrix.*[1] The entry $S_{j_1 j_2}$ is large if the j_1 and j_2 coordinates tend to have the same sign, and it is negative if these coordinates tend to have the opposite sign. The diagonal entry S_{jj} is a sum of squares, hence positive; it measures how much the j-th coordinate tends to vary.

- In the previous example, S would be a 2×2 matrix. The entries S_{11} and S_{22} tell you the variance of the temperature and humidity, respectively. The entry S_{12} tells you how temperature correlates with humidity.

The 'correlation' between two measured variables, such as temperature and humidity, is more frequently expressed by drawing a 'line of best fit' on a scatter plot. The SVD of A will tell us how to do this.

Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots \in \mathbb{R}^m$ be the (orthonormal) singular vectors of A (the columns of U), and let $\sigma_1^2, \sigma_2^2, \ldots$ be their singular values (the entries of Σ). Assume $\sigma_1 > \sigma_2 > \cdots$.

- *The statistical meaning of the SVD.* The 'main theorem of linear algebra in statistics' is that our data looks as if it's generated in this way:

  (1) On step i, choose new values for the scalars $z_1, z_2, \ldots$ based on a bell curve centered at zero with variance 1.

  (2) Set $\boldsymbol{v}_i = \boldsymbol{\mu} + z_1\sigma_1\boldsymbol{u}_1 + z_2\sigma_2\boldsymbol{u}_2 + \cdots$.

  The 'total variance' is $\sigma_1^2 + \sigma_2^2 + \cdots$, and the 'amount of variance explained by $\boldsymbol{u}_i$' is σ_i^2.

Since σ_1 is relatively large, $\boldsymbol{u}_1$ is the direction in which the data fluctuate away from the mean $\boldsymbol{\mu}$ most wildly. The line through μ parallel to $\boldsymbol{u}_1$ is the line of best fit (which minimizes the *perpendicular* distances to the points in the data set).

We will look at the following example:

$$A = \begin{pmatrix} 6 & 5 & -4 & -3 \\ 3 & 4 & -5 & -6 \end{pmatrix}.$$

[1] We will briefly discuss why the denominator is $n-1$ and not n, as would usually be the case for a statistical average.

2. Fourier series

See Section 10.5 in Strang's book.

We focus on Example 3 from that section.

- *Problem.* Express the square wave function

$$f(x) = \begin{cases} 1 & \text{if } \lfloor \frac{x}{\pi} \rfloor \text{ is even} \\ -1 & \text{otherwise} \end{cases}$$

  as an infinite linear combination of the functions $\sin(x), \cos(x), \sin(2x), \cos(2x), \ldots$.

The crux of the computation is the integral

$$\begin{aligned} \int_0^{2\pi} f(x)\, \sin(nx)\, dx &= \int_0^{\pi} \sin(nx)\, dx - \int_{\pi}^{2\pi} \sin(nx)\, dx \\ &= \left[-\tfrac{1}{n}\cos(nx)\right]_{x=0}^{\pi} - \left[-\tfrac{1}{n}\cos(nx)\right]_{x=\pi}^{2\pi} \\ &= \begin{cases} \frac{4}{n} & \text{if } n \text{ is odd} \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

We also need the 'orthonormality' relations

$$\begin{aligned} \int_0^{2\pi} \sin(nx)\sin(mx)\, dx &= \tfrac{1}{2}\left(\int_0^{2\pi} \cos((n-m)x)\, dx - \int_0^{2\pi} \cos((n+m)x)\, dx\right) \\ &= \begin{cases} \pi & \text{if } n = m \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

assuming that n and m are positive integers.

Once we find the answer for $f(x)$, taking the derivatives gives the Taylor series for $f'(x)$, which is an infinite sum of δ-functions.

Another phenomenon to ponder: (pointwise) multiplication by $\sin(x)$ yields a linear map

$$(\text{functions with period } 2\pi) \to (\text{functions with period } 2\pi)$$

which is not invertible, but has nullspace $= \{0\}$. Hence, some of the facts we learned in the finite-dimensional setting break down in this infinite-dimensional vector space.