DS-GA 3001.03: Extended Syllabus Lecture 2

Optimization and Computational Linear Algebra for Data Science (Fall 2016)

Afonso S. Bandeira
bandeira@cims.nyu.edu
http://www.cims.nyu.edu/~bandeira

September 20, 2016

These are not meant to be Lecture Notes. They are simply extended syllabi with the most important definitions and results from the lecture. As such, they lack the intuition and motivation and so they are not a good place to learn the material the first time, just to briefly review it. These extended syllabi will also have references.

There are many amazing books about linear algebra and virtually all of them will contain the material for this particular lecture, examples include the book suggested for the course [2]. Another place you can read about some of this is the Lecture Notes from last years DSGA1002 [1].

Please let me know if you find any typos!

- Given a matrix $L \in \mathbb{R}^{n\times m}$ (which means it is a linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$) the rank of L is
$$\mathrm{rank}(L) = \dim(\mathrm{Im}(L)).$$
- A matrix $L \in \mathbb{R}^{n\times n}$ has an inverse L^{-1} if and only if $\mathrm{rank}(L) = n$.
- Rank is a very important concept in recommendation systems. As we will see later, data matrices naturally arising in recommendation systems tend to be (approximately) low rank, and this can be leveraged to make meaningful recommendations.
- Given a matrix $L \in \mathbb{R}^{n\times m}$ we define its transpose $L_T \in \mathbb{R}^{m\times n}$ as
$$\left(L^T\right)_{ij} = L_{ji}.$$
- Given a matrix $L \in \mathbb{R}^{n\times m}$ we have
$$\mathrm{rank}(L) = \mathrm{rank}(L^T).$$
- Given two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times p}$ we have
$$(AB)^T = B^T A^T.$$
- Given two matrices $A, B \in \mathbb{R}^{n\times n}$ both invertible (meaning that A^{-1} and B^{-1} exist) we have
$$(AB)^{-1} = B^{-1}A^{-1}.$$

- A matrix $L \in \mathbb{R}^{n \times n}$ is said to be symmetric if $L^T = L$. They arise very naturally in Data Science if, for example, entry i, j corresponds to some (symmetric) similarity measure (or, say, distance) between i and j.

- A matrix $L \in \mathbb{R}^{n \times n}$ is said to be anti-symmetric if $L^T = -L$.

- Systems of Linear Equations, matrix representation, Inverse matrix, Triangular Systems, Gaussian Elimination, LU factorization, See (for example) Chapter 2 of [2].

- Given a vector $x \in \mathbb{R}^n$ we define it's norm $\|x\|$ (distance to 0) as
$$\|x\|^2 = x^T x = \sum_{i=1}^n x_i^2.$$

- Given two vectors $x, y \in \mathbb{R}^n$ we define their inner product as
$$\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i.$$

- We have
$$\begin{array}{rc} (i) & \forall_{x,y \in \mathbb{R}^n} \langle x, y \rangle = \langle y, x \rangle \\ (ii) & \forall_{x,y,z \in \mathbb{R}^n} \langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle \\ (iii) & \forall_{x,y \in \mathbb{R}^n} \forall_{\alpha \in \mathbb{R}} \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \end{array}$$

- Given a matrix $L \in \mathbb{R}^{n \times m}$, the matrix $L^T \in \mathbb{R}^{m \times n}$ is the matrix such that, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ we have:
$$\langle x, Ly \rangle = \langle L^T x, y \rangle$$

- Interesting example: If we take the vector space of smooth functions from $[0, 1]$ to $\mathbb{R}$ that satisfy $f(0) = 1$ and $f(1) = 1$ and define the inner product
$$\angle f, g \rangle = \int_0^1 f(t) g(t) dt,$$
then integration by parts shows that the derivative (as a linear transformation) is anti-symmetric. You can see example in, for example, Section 2.7. of [2].

- Cauchy-Schwartz inequality: $\forall_{x,y \in \mathbb{R}^n}$ we have
$$|x^T y| \leq \|x\| \|y\|.$$
To prove this, notice that $\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\|^2 \geq 0$ and $\left\| \frac{x}{\|x\|} + \frac{y}{\|y\|} \right\|^2 \geq 0$ and expand the norm square of the sum...

- Triangular inequality: $\|x + y\| \leq \|x\| + \|y\|$

- The Projection of y onto span of x is given by

$$P_{\langle x \rangle}(y) = \frac{x^T y}{\|x\|^2} x.$$

- If θ is the angle between x and y we have

$$\cos(\theta) = \frac{x^T y}{\|x\| \|y\|}$$

  Note that, since $-1 \leq \cos(\theta) \leq 1$, this is an alternative proof of the Cauchy-Schwartz inequality.

- If $x^T y = 0$ then the angle between them is $\frac{\pi}{2}$ and we say that the vectors are orthogonal. The observation that in that case $\|x - y\|^2 = \|x\|^2 + \|y\|^2$ is Pythagoras theorem.

References

[1] Carlos Fernandez-Granda, *Lecture Notes of DSGA1002*, available at `http://www.cims.nyu.edu/~cfgranda/pages/DSGA1002_fall15/notes.html`, 2015

[2] Gilbert Strang, *Introduction to Linear Algebra*, Fifth Edition, 2016