Math40003 Linear Algebra and Groups — Term 2 Unseen 2

1. Let F be a field, $A \in M_{n\times n}(F)$ and $\lambda \in F$. Prove that for all $\lambda \in F$, the sets $\{v \in F^n | Av = \lambda v\}$ and $\{v \in F^n | A^t v = \lambda v\}$ are subspaces of the same dimension. Conclude that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^t.

Definition 1. A matrix $A \in M_{n\times n}(\mathbb{R})$ is *column-stochastic* if all of its entries are non-negative and the sum of the entries in each column is equal to 1.

2. Prove that a column-stochastic matrix has $\lambda = 1$ as an eigenvalue.

3. (a) Let $a_1, \ldots, a_n \in \mathbb{R}$. Prove that $|\sum_{i=1}^n a_i| < \sum_{i=1}^n |a_i|$ if and only if there are i, j such that $a_i > 0$ and $a_j < 0$.

   (b) Prove that if $v, w \in \mathbb{R}^n$ are linearly independent, then there is some $u \in \operatorname{span}\{v, w\}$ such with at least one positive entry and one negative entry.

   (c) Let $A \in M_{n\times n}(\mathbb{R})$ be a column-stochastic matrix with all entries positive. Prove that $\{v \in \mathbb{R}^n | Av = v\}$ is a 1-dimensional vector space.

   (d) With A as in (c), conclude there is a unique vector v such that $Av = v$, all of its entries are positive and the sum of its entries is 1.

4. Let $n \in \mathbb{N}$. Let $\mathcal{P} := \{1, \ldots, n\}$ (elements of $\mathcal{P}$ are referred to as pages) and let $\mathcal{L} \subseteq \mathcal{P}^2$ be some set of ordered pairs of $\mathcal{L}$ (elements of $\mathcal{L}$ are called links). If $(a, b) \in \mathcal{L}$, we say there is a link from a to b. Fix some $0 < p < 1$. Consider the following random process :

   - In stage 0: we pick a page with uniform distribution (i.e., the probability of every element of $\mathcal{P}$ to be picked is $1/n$).
   - In stage $k+1$: Assuming in stage k we picked the page a, in stage $k+1$, with probability p we pick a new page out of $\mathcal{P}$ with uniform distribution, and with probability $1-p$ we pick, with uniform distribution, some page out of the pages with a link from a.

     Namely:

     – If $(a, b) \notin \mathcal{L}$, then the probability of picking b is $p \cdot 1/n$.

     – If $(a, b) \in \mathcal{L}$ and there are l many links from a, then the probability of picking b is $p \cdot 1/n + (1-p) \cdot 1/l$.

   Find a matrix $A \in M_{n\times n}(\mathbb{R})$ and a vector $v \in \mathbb{R}^n$ such that the probability of picking page i in stage k is the i-th coordinate of $A^k v$. Prove A is column-stochastic.

   In this sheet, you almost developed the basic PageRank, Google's ranking of web pages. For details, see the back of this page.

The process described in Question 4 describes a random surfer on the web. The surfer starts by surfing to a random web page picked at random, then with probability p (where p is quite small, e.g., 0.15) the surfer jumps to a completely random new page and with probability $(1 - p)$ the surfer follows one of the links on the page currently at.

It can be shown that the process $Av, A^2v, A^3v, \ldots$ converges on each coordinate. Moreover, the limit vector, call it v^*, is precisely the unique vector found in Quesion 3d. However, this requires some more analysis. The i-th coordinate of v^* is the PageRank of page i.

Of course, there is more to Google's algorithm, e.g., what's the magic p, giving some links more weight than others etc. More importantly, how to compute v^*? Having said that, still, these are the basic foundations of Google's PageRank.

As for the question of computing (or at least, approximating) v^*, this can be solved by taking powers of matrices efficiently, a matter that will be addressed later in this module.

For further information, see:
`http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/index.html`