FINM 331: DATA ANALYSIS FOR FINANCE AND STATISTICS
FALL 2015
MATRIX THEORY BACKGROUND

1. NORMS

- a *norm* is a real-valued function on a vector space (over $\mathbb{R}$), denoted $\|\cdot\| : V \to \mathbb{R}$ satisfying
  1. $\|\mathbf{x}\| \geq 0$ for all $\mathbf{x} \in V$
  2. $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
  3. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x} \in V$
  4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in V$
- we will be interested in two specific choices of V
  - $V = \mathbb{R}^n$
  - $V = \mathbb{R}^{m \times n}$

2. VECTOR NORMS

- if $V = \mathbb{R}^n$, we call a norm on V a *vector norm*
- example: consider $\|\cdot\|_1 : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

  for $\mathbf{x} = [x_1, \dots, x_n]^\mathsf{T} \in \mathbb{R}^n$ and where $|x|$ denotes the modulus/absolute value of $x \in \mathbb{R}$
  - check that this is a norm:
    1. clearly $\|\mathbf{x}\|_1 \geq 0$
    2. the only way a sum nonnegative entries $\|\mathbf{x}\|_1 = 0$ is if all entries $|x_i| = 0$ and so $\mathbf{x} = [0, \dots, 0]^\mathsf{T} = \mathbf{0}$
    3. we have

$$\|\alpha\mathbf{x}\|_1 = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha|\|\mathbf{x}\|_1$$

       since complex modulus satisfies $|\alpha x| = |\alpha||x|$
    4. using the triangle inequality for complex numbers, we obtain

$$\|\mathbf{x} + \mathbf{y}\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} |x_i| + |y_i| \leq \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$$

  - therefore the function defines a norm, called the 1-*norm* or *Manhattan norm*
- example: more generally, for $p \geq 1$ (can be any real number, not necessarily an integer), we define the p-*norm* $\|\mathbf{x}\|_p$ by

$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$

  - most commonly used p-norms is the 2-*norm* or *Euclidean norm*:

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

Date: October 10, 2015, version 1.1. Comments, bug reports: `lekheng@galton.uchicago.edu`.

– easy to see that for any p, we have

$$\left(\max_{i=1,\dots,n}|x_i|^p\right)^{1/p} \le \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n}|x_i|^p\right)^{1/p} \le \left(n\max_{i=1,\dots,n}|x_i|^p\right)^{1/p}$$

– from which it follows that

$$\max_{i=1,\dots,n}|x_i| \le \|\mathbf{x}\|_p \le n^{1/p}\max_{i=1,\dots,n}|x_i|$$

– as $p \to \infty$, we obtain the *infinity norm*

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty}\|\mathbf{x}\|_p = \max_{i=1,\dots,n}|x_i|$$

which is also known as the *Chebyshev norm*

– easy to verify that p-norms for any $p \in [1,\infty]$ are indeed norms

3. more vector norms

- example: variation of the p-norm is the *weighted p-norm*, defined by

$$\|\mathbf{x}\|_{p,\mathbf{w}} = \left(\sum_{i=1}^{n} w_i|x_i|^p\right)^{1/p}$$

  – again it can be shown that this is a norm as long as the *weights* w_i, $i = 1,\dots,n$, are strictly positive real numbers
- example: a vast generalization of all of the above is the *A-norm* or *Mahalanobis norm*, defined in terms of a matrix A by

$$\|\mathbf{x}\|_A = (\mathbf{x}^\mathsf{T}A\mathbf{x})^{1/2} = \left(\sum_{i,j=1}^{n} a_{ij}\overline{x}_i x_j\right)^{1/2}$$

  – this defines a norm provided that the matrix A is positive definite

  – note that if $W = \operatorname{diag}(\mathbf{w})$, then

$$\|\mathbf{x}\|_W = \|\mathbf{x}\|_{2,\mathbf{w}}$$

- we now highlight some additional, and useful, relationships for a norm

  – first of all, the triangle inequality generalizes directly to sums of more than two vectors:

$$\|\mathbf{x}+\mathbf{y}+\mathbf{z}\| \le \|\mathbf{x}+\mathbf{y}\| + \|\mathbf{z}\| \le \|\mathbf{x}\| + \|\mathbf{y}\| + \|\mathbf{z}\|$$

  – more generally,

$$\left\|\sum_{i=1}^{m}\mathbf{x}_i\right\| \le \sum_{i=1}^{m}\|\mathbf{x}_i\|$$

  – secondly, what can we say about the norm of the difference of two vectors? we know that $\|\mathbf{x}-\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$ but we can obtain a more useful relationship as follows:

$$\|\mathbf{x}\| = \|(\mathbf{x}-\mathbf{y})+\mathbf{y}\| \le \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}\|$$

  we obtain

$$\|\mathbf{x}-\mathbf{y}\| \ge \|\mathbf{x}\| - \|\mathbf{y}\|$$

  – thirdly, from

$$\|\mathbf{y}\| = \|\mathbf{y}-\mathbf{x}+\mathbf{x}\| \le \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{x}\|$$

  it follows that

$$\|\mathbf{x}-\mathbf{y}\| \ge \|\mathbf{y}\| - \|\mathbf{x}\|$$

and therefore

$$||\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\| \tag{3.1}$$

– the inequality (3.1) yields a very important property of norms, namely, they are all (uniformly) continuous functions of the entries of their arguments

- there is a relationship that applies to products of norms, the *Hölder inequality*

$$|\mathbf{x}^{\mathsf{T}}\mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

– a well-known corollary arises when $p = q = 2$, the *Cauchy–Schwarz inequality*

$$|\mathbf{x}^{\mathsf{T}}\mathbf{y}| \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

– by setting $\mathbf{x} = [1, 1, \ldots, 1]^{\mathsf{T}}$, the Hölder inequality yields the relationships

$$\left|\sum_{i=1}^{n} y_i\right| \leq \sum_{i=1}^{n} |y_i|$$

and

$$\left|\sum_{i=1}^{n} y_i\right| \leq n \max_{i=1,\ldots,n} |y_i|$$

and

$$\left|\sum_{i=1}^{n} y_i\right| \leq \sqrt{n} \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2}$$

4. MATRIX NORMS

- note that the space of complex $m \times n$ matrices $\mathbb{R}^{m \times n}$ is a vector space over $\mathbb{R}$ of dimension mn
- we write O for the $m \times n$ zero matrix, i.e., all entries are 0
- a norm on either $\mathbb{R}^{m \times n}$ or $\mathbb{R}^{m \times n}$ is called a *matrix norm*
- recall that these means $\|\cdot\| : \mathbb{R}^{m \times n} \to \mathbb{R}$ satisfies
  (1) $\|A\| \geq 0$ for all $A \in \mathbb{R}^{m \times n}$
  (2) $\|A\| = 0$ if and only if $A = O$
  (3) $\|\alpha A\| = |\alpha| \|A\|$
  (4) $\|A + B\| \leq \|A\| + \|B\|$
- often we add a fifth condition that $\|\cdot\|$ satisfies the *submultiplicative property*

$$\|AB\| \leq \|A\| \|B\|$$

- submultiplicative is some also called *consistent*
- example: *Frobenius norm*

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

which is submultiplicative since

$$\|AB\|_F^2 = \sum_{i=1}^{m} \sum_{k=1}^{p} \left|\sum_{j=1}^{n} a_{ij} b_{jk}\right|^2 \leq \sum_{i=1}^{m} \sum_{k=1}^{p} \left[\left(\sum_{j=1}^{n} |a_{ij}|^2\right) \left(\sum_{j=1}^{n} |b_{jk}|^2\right)\right]$$

by the Cauchy–Schwarz inequality and the last expression is equal to

$$\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{k=1}^{p}\sum_{j=1}^{n}|b_{jk}|^2\right) = \|A\|_F^2\|B\|_F^2$$

- a very important class of matrix norms are the so called *operator* or *induced* or *natural norms* defined as
$$\|A\|_{p,q} := \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_q} \tag{4.1}$$
for any $A \in \mathbb{R}^{m\times n}$ and any vector norms $\|\cdot\|_p : \mathbb{R}^m \to \mathbb{R}$ and $\|\cdot\|_q : \mathbb{R}^n \to \mathbb{R}$ defined on the domain and codomain of A respectively
- the induced norm may also be written as
$$\|A\|_{p,q} = \max\{\|A\mathbf{x}\|_p : \|\mathbf{x}\|_q \le 1\} \tag{4.2}$$
or as
$$\|A\|_{p,q} = \max\{\|A\mathbf{x}\|_p : \|\mathbf{x}\|_q = 1\} \tag{4.3}$$
- in other words, the induced norm measures how far the operator A sends points in the unit disc (or the unit circle)
- the matrix 2-norm is obtained when $p = q = 2$
$$\|A\|_2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$
- it is very widely used and has its own special name, *spectral norm*, because of its relation to the spectrum of a matrix (i.e., the eigenvalues); we will discuss it later
- the matrix 1-norm and ∞-norm are obtained by setting $p = q = 1$ and $p = q = \infty$ respectively
- they also very widely used, largely because, unlike the matrix 2-norm, we have closed form expressions for these
- let $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{R}^{m\times n}$, then
$$\|A\|_1 = \max_{j=1,\dots,n}\left[\sum_{i=1}^{m}|a_{ij}|\right] \tag{4.4}$$
and
$$\|A\|_\infty = \max_{i=1,\dots,m}\left[\sum_{j=1}^{n}|a_{ij}|\right] \tag{4.5}$$
- an easy way to remember these is that $\|A\|_1$ is the maximum column sum and $\|A\|_\infty$ is the maximum row sum of A

5. Eigenvalue decomposition

- recall our two fundamental problems:
$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad A\mathbf{x} = \lambda\mathbf{x}$$
- even if we are just interested to solve $A\mathbf{x} = \mathbf{b}$ and its variants, we will need to understand eigenvalues and eigenvectors
- we will use properties of eigenvalues and eigenvectors but will only briefly describe its computation (towards the last few lectures)
- recall: $A \in \mathbb{R}^{n\times n}$, if there exists $\lambda \in \mathbb{C}$ and $\mathbf{0} \neq \mathbf{x} \in \mathbb{C}^n$ such that
$$A\mathbf{x} = \lambda\mathbf{x},$$
we call λ an *eigenvalue* of A and $\mathbf{x}$ an *eigenvector* of A corresponding to λ or λ-eigenvector

- note that real matrices can have complex eigenvalues and eigenvectors
- some basic properties
  - eigenvector is a scale invariant notion, if $\mathbf{x}$ is a λ-eigenvector, than so is $c\mathbf{x}$ for any $c \in \mathbb{C}^\times$
  - we usually, but not always, require that $\mathbf{x}$ be a unit vector, i.e., $\|\mathbf{x}\|_2 = 1$
  - note that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are λ-eigenvectors, then so is $\mathbf{x}_1 + \mathbf{x}_2$
  - for an eigenvalue λ, the subspace $V_\lambda = \{\mathbf{x} \in \mathbb{C}^n : A\mathbf{x} = \lambda\mathbf{x}\}$ is called the λ-*eigenspace* of A and is the set of all λ-eigenvectors of A together with $\mathbf{0}$
  - the set of all eigenvalues of A is called its *spectrum* and often denoted $\lambda(A)$
  - an $n \times n$ matrix always have n eigenvalues in $\mathbb{C}$ counted with multiplicty
  - however an $n \times n$ matrix may not have n linear independent eigenvectors
  - an example is
$$J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{5.1}$$
    which has eigenvalue 0 with multiplicity 2 but only one eigenvector (up to scaling) $\mathbf{x} = [1, 0]^\mathsf{T}$
- an $n \times n$ matrix A that has n linear independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is called a *diagonalizable matrix* since if we write these as columns of a matrix $X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, then X is necessarily nonsingular and
$$AX = [A\mathbf{x}_1, \ldots, A\mathbf{x}_n] = [\lambda_1\mathbf{x}_1, \ldots, \lambda_n\mathbf{x}_n] = X \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} =: X\Lambda \tag{5.2}$$
  and so
$$A = X\Lambda X^{-1} \tag{5.3}$$
  where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues
- the decomposition (5.3) is called the *eigenvalue decomposition* (EVD) of A
- not every matrix has an EVD, an example is the J in (5.1)
- summary: a matrix has an EVD iff it has n linearly independent eigenvectors iff it is diagonalizable
- normally we will sort the eigenvalues in descending order of magnitude
$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$$
- λ_1, also denoted $\lambda_{\max}$, is called the *principle eigenvalue* of A and a $\lambda_{\max}$-eigenvector is called a *principal eigenvector*
- since $\mathbf{x}_1, \ldots, \mathbf{x}_n$ form a basis for the domain of A, we call this an *eigenbasis*
- note that the matrix of eigenvectors X in (5.3) is only required to be non-singular (a.k.a. invertible)

6. Spectral theorem for symmetric matrices

- in general it is difficult to check whether a matrix is diagonalizable
- however there is a special class of matrices for which we check diagonalizability easily, namely, the symmetric matrices

Theorem 1 (Spectral Theorem for symmetric Matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$. If A is a symmetric matrix, i.e.*
$$A^\mathsf{T} = A,$$
iff A has an EVD *of the form*
$$A = V\Lambda V^\mathsf{T}$$

where $V \in \mathbb{R}^{n\times n}$ is orthogonal and $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal.

- note that saying the column vectors of $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ are mutually orthonormal is the same as saying $V^\mathsf{T} V = I = V V^\mathsf{T}$ and is the same as saying that V is orthogonal
- a special class of normal matrices are the ones that are equal to their adjoint, i.e. $A^\mathsf{T} = A$,and these are called *symmetric* or self-adjoint matrices
- exercise: show that an eigenvalue of a real symmetric matrix is always real

7. Jordan canonical form

- if A is not diagonalizable and we want something like a diagonalization, then the best we could do is a *Jordan canonical form* or Jordan normal form where we get
$$A = XJX^{-1} \tag{7.1}$$
  - the matrix J has the following characteristics
    * it is not diagonal but it is the next best thing to diagonal, namely, *bidiagonal*, i.e. only the entries a_{ii} and $a_{i,i+1}$ can be non-zero, every other entry in J is 0
    * the diagonal entries of J are precisely the eigenvalues of A, counted with multiplicity
    * the superdiagonal entries $a_{i,i+1}$ are as simple as they can be — they can take one of two possible values $a_{i,i+1} = 0$ or 1
    * if $a_{i,i+1} = 0$ for all i, then J is in fact diagonal and (7.1) reduces to the eigenvalue decomposition
  - the matrix J is more commonly viewed as a block diagonal matrix
$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{bmatrix}$$
    * each block J_r, for $r = 1, \dots, k$, has the form
$$J_r = \begin{bmatrix} \lambda_r & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_r \end{bmatrix}$$
where J_r is $n_r \times n_r$
    * clearly $\sum_{r=1}^{k} n_r = n$
  - the set of column vectors of X are called a *Jordan basis* of A
- in general the Jordan basis X include all eigenvectors of A but also additional vectors that are not eigenvectors of A
- the Jordan canonical form provides valuable information about the eigenvalues of A
- the values λ_j, for $j = 1, \dots, k$, are the eigenvalues of A
- for each distinct eigenvalue λ, the number of Jordan blocks having λ as a diagonal element is equal to the number of linearly independent eigenvectors associated with λ, this number is called the *geometric multiplicity* of the eigenvalue λ
- the sum of the sizes of all of these blocks is called the *algebraic multiplicity* of λ
- Jordan canonical form suffers however from one major defect that makes them useless in practice: they cannot be computed in finite precision or in the presence of rounding errors in general, a result of Golub and Wilkinson
- that is why you won't find a Matlab function for Jordan canonical form