Lecture 7

Norms and Condition Numbers

To discuss the errors in numerical problems involving vectors, it is useful to employ norms.

Vector Norm

On a vector space V, a norm is a function $\|\cdot\|$ from V to the set of non-negative reals that obeys three postulates:

$$\|x\| > 0 \qquad if \quad x \neq 0, C$$
$$\|\lambda x\| = |\lambda| \|x\| \qquad if \quad \lambda \in R, x \in V$$
$$\|x + y\| \leq \|x\| + \|y\| \qquad if \quad x, y \in V \qquad (Trinagular \quad Inequality)$$

we can think of $\|x\|$ as the length or magnitude of the vector x.

The most familiar norm on R^n is the Euclidean

ℓ_2-norm defined by $\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$

ℓ_∞-norm defined by $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

ℓ_1-norm defined by $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

In general *p*-norm, defined by

ℓ_p-norm defined by $\|x\|_p = \left(\sum_{i=1}^{n} x_i^p \right)^{1/p}$ for *p* > 0 and *n*-vector x

Example 1: Using the norm $\|\cdot\|_1$, compare the lengths of the following three vectors in R^4. Repeat it for other norms

$$x = (4, 4, -4, 4)^T, v = (0, 5, 5, 5)^T, w = (6, 0, 0, 0)^T$$

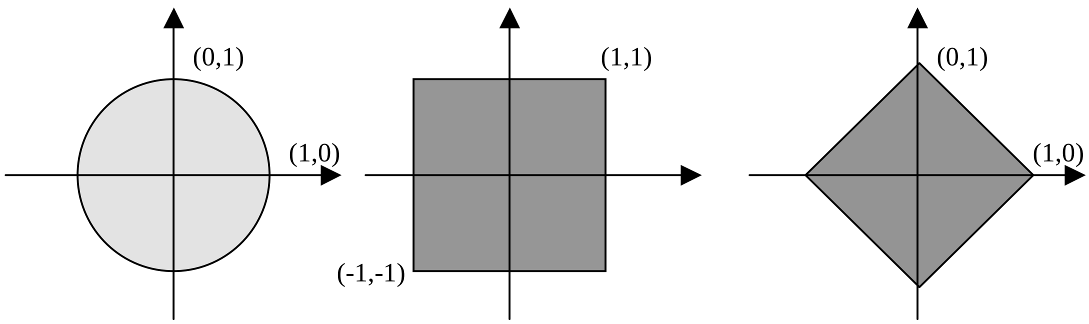
Solution:

	$\|\cdot\|_1$	$\|\cdot\|_2$	$\|\cdot\|_\infty$
X	16	8	4
v	15	8.66	5
w	6	6	6

To understand these norms better, it is instructive to consider R2. For the three norms give above, we sketches in Figure 1 of the set

$$\{x : x \in R^2, \|x\| \leq 1\}$$

This set is called the unit cell or the unit ball in two-dimensional vector space.



$\|x\|_2 \leq 1$ $\qquad$ $\|x\|_\infty \leq 1$ $\qquad$ $\|x\|_1 \leq 1$

Figure 1: Unit cells in R^2 for three norms

In general, for any vector in *x* in R^n, $\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty$

Matrix Norm

Matrix norm corresponding to given vector norm defined by

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

Norm of matrix measures maximum stretching matrix does to any vector in given vector norm.

Matrix norm corresponding to vector 1-norm is maximum absolute column sum

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

Matrix norm corresponding to vector ∞- norm is maximum absolute row sum,

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

Properties of Matrix Norm

Any matrix norm satisfies:

1. $\|\boldsymbol{A}\| > \boldsymbol{O}$ if $\boldsymbol{A} \neq \boldsymbol{O}$
2. $\|\gamma\boldsymbol{A}\| = |\gamma| \cdot \|\boldsymbol{A}\|$ for any scalar value γ

3. $\|\boldsymbol{A}+\boldsymbol{B}\| \le \|\boldsymbol{A}\|+\|\boldsymbol{B}\|$

Matrix norm also satisfies

4. $\|\boldsymbol{AB}\| \le \|\boldsymbol{A}\| \cdot \|\boldsymbol{B}\|$
5. $\|\boldsymbol{Ax}\| \le \|\boldsymbol{A}\| \cdot \|\boldsymbol{x}\|$ for any vector $\boldsymbol{x}$

Matrix Condition Number

Condition number of square nonsingular matrix A defined by

$$\boldsymbol{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\| \cdot \|\boldsymbol{A}\|^{-1}$$

By convention, cond ($\boldsymbol{A}$) = ∞ if A singular

Example: $\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 3 & -1 & 4 \end{bmatrix} \quad \|\boldsymbol{A}\|_1 = 6 \qquad \|\boldsymbol{A}\|_\infty = 8$

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 2.5 & -0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix} \quad \|\boldsymbol{A}^{-1}\|_1 = 4.5 \qquad \|\boldsymbol{A}^{-1}\|_\infty = 3.5$$

cond_1 (A) = 6 x 4.5 = 27
cond_∞ (A) = 8 x 3.5 = 28

The numerical value of the condition number of an *n x n* matrix depends on the particular norm used (indicated by the corresponding subscript), but because of the equivalence of the underlying vector norms, these values can differ by at most a fixed constant (which depends on n), and hence they are equally useful as quantitative measure of conditioning.

$$\text{Since} \quad \|A\| \cdot \|A^{-1}\| = \left(\max_{x\neq 0} \frac{\|Ax\|}{\|x\|}\right) \cdot \left(\min_{x\neq 0} \frac{\|Ax\|}{\|x\|}\right)^{-1}$$

The condition number of the matrix measures the ratio of the maximum relative stretching to the maximum relative shrinking that matrix does to any non zero vectors.

Another way to say that the condition number of a matrix measures the amount of distortion of the unit sphere (in the corresponding vector norm) under the transformation by the matrix. The larger the condition number, the more distorted (relatively long and thin) the unit sphere becomes when transformed by the matrix.

In two dimensions, for example, the unit circle in the 2-norm becomes and increasingly cigar shaped ellipse, and with the 1-norm or ∞- norm, the unit sphere is transformed from a square into increasingly skewed parallelogram as the condition number increases.

The condition number a measure of how close a matrix is to being singular: a matrix with large condition number is nearly singular, whereas a matrix with condition number close to 1 is far from being singular.

It is obvious from the definition that a nonsingular matrix and its inverse have the same condition number.

Note: Large condition number of $\boldsymbol{A}$ mean $\boldsymbol{A}$ is nearly singular.

Properties of the condition number

1. For any matrix $\boldsymbol{A}$, cond ($\boldsymbol{A}$) ≥ 1
2. For identity matrix, cond (I) = 1
3. For any matrix A and scalar γ, cond (γ A) = cond (A)
4. For any diagonal matrix D = Diag(d_i), cond (D) = (max $|d_i|$)/(min$|d_i|$)

Computing Condition number

Definition of condition number involves matrix inverse, so nontrivial to compute

Computing condition number from the definition would require much more work than computing solution whose accuracy to be assessed.

In practice, condition number estimated inexpensively as byproduct of solution process

Matrix norm $\|\boldsymbol{A}\|$ can be easily computed as maximum absolute column sum (or row sum, depending on norm used)

But, estimating $\|\boldsymbol{A}^{-1}\|$ at low cost more challenging

Computing Condition Number

We will now see the usefulness of the condition number in assessing the accuracy of the solution to linear system. In fact, to compute the condition number directly from definition would require substantially more work than solving the linear system whose accuracy is to be assessed.

In practice, therefore the condition number is merely estimated, to perhaps within an order of magnitude, as a relatively inexpensive byproduct of the solution procedure.

From the properties of norm, we know that if **z** is the solution of $\boldsymbol{Ax = y}$ then

$\|\boldsymbol{z}\| = \|\boldsymbol{A}^{-1}\boldsymbol{y}\| \leq \|\boldsymbol{A}^{-1}\| \cdot \|\boldsymbol{y}\|$ So that $\dfrac{\|\boldsymbol{z}\|}{\|\boldsymbol{y}\|} \leq \|\boldsymbol{A}^{-1}\|$

and this bound is associate for some optimally chosen vector ***y.***

Thus, if we can choose a vector y such that the ratio $\dfrac{\|\boldsymbol{z}\|}{\|\boldsymbol{y}\|}$ is as large as possible, then we will have reasonable estimate for $\|\boldsymbol{A}^{-1}\|$.

Example: $\boldsymbol{A} = \begin{bmatrix} 0.913 & 0.659 \\ 0.457 & 0.330 \end{bmatrix}$

If we choose y = [0, 1.5]^T, then z = [-7780, 10780]^T

So that $\|\boldsymbol{A}^{-1}\|_1 \approx \dfrac{\|\boldsymbol{z}\|_1}{\|\boldsymbol{y}\|_1} \approx 1.238 \times 10^4$

and hence $\boldsymbol{cond}_1(\boldsymbol{A}) = \|\boldsymbol{A}\|_1 \cdot \|\boldsymbol{A}^{-1}\|_1 \approx 1.370 \times 1.238 \times 10^4 = 1.696 \times 10^4$

which turns out to be exact to the number of digits shown.

The vector in this example was chosen to produce the maximum possible ratio $\dfrac{\|\boldsymbol{z}\|}{\|\boldsymbol{y}\|}$, and hence the correct value for $\|\boldsymbol{A}^{-1}\|$. Finding such an optimum value ***y*** would be prohibitively expensive. In general, but a useful approximation can be obtained much more cheaply.

One heuristic is to choose ***y*** as the solution to the system $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{y} = \boldsymbol{c}$. ***c*** is a vector whose components are ± 1 with sign chosen successively to make y as large as possible.