MATH 728D: Machine Learning Lab #15: Norms, Condition Numbers, Interpolation

John Burkardt

January 14, 2019

Can we say a numerical error is small enough to be ignored?

Some computations are like a lion hunt. We are hunting for a value x, and we do so by laying a "trap" around an equation like $f(x) = 0$. If we discover a value $\hat{x}$ so that $|f(\hat{x})| \leq \epsilon$, we assume we've caught our lion...or really, that the difference between $\hat{x}$ and x is so small as to be unimportant.

It turns out that we can only say this confidently if we know that the equation $f(x) = 0$ is *well-conditioned*.

When our values are vectors, and our equations involve linear operators (matrices), then we need to be clear on how to measure the size of objects, the distance between two objects, the degree to which an operator can magnify size, and whether the solution to an approximately solved equation is close to the solution of the exact equation.

1 Vector Norms

In order to talk about the "size" and "distance between" n-dimensional vectors, we want a vector norm, a function similar to the absolute value. Three common vector norms are:

- $\ell 1$ vector norm: $||x||_1 = \sum_{i=1}^{n} |x_i|$
- $\ell 2$ or "Euclidean" vector norm: $||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$
- $\ell\infty$ or "max" vector norm: $||x||_\infty = \max\limits_{1 \leq i \leq n} |x_i|$

The subscript on the norm is usually omitted if it is clear which norm is being used; the $\ell 2$ norm is preferred.

A vector norm has the following properties:

- Positive: $0 \leq ||v||$, and 0 only if v is identically 0;
- Linear: $||\alpha v|| = |\alpha| \, ||v||$ for scalar α;
- Triangle inequality: $||v_1 + v_2|| \leq ||v_1|| + ||v_2||$

The MATLAB commands `norm(v,1)`, `norm(v,2)`, `norm(v,Inf)` will compute the $\ell 1, \ell 2$, and $\ell\infty$ norms of vector v respectively. The command `norm(v)` returns the value of the default $\ell 2$ norm.

Exercise 1:

1. Set spatial dimension $n = 3$;
2. Define vectors `x = [1;2;3], y=[1;1;1], z =[3;0;4];`
3. Evaluate $\ell\infty, \ell 2$, and $\ell 1$ norms for each vector;
4. For each of `v = x, y, z`, verify that $||v||_\infty \leq ||v||_2 \leq ||v||_1$
5. For each of `v = x, y, z`, verify that $||v||_1 \leq \sqrt{n} \, ||v||_2 \leq n \, ||v||_\infty$;
6. For each of `v = x, y, z`, verify that $||v||_2 \leq \sqrt{||v||_1 \, ||v||_\infty}$

7. For the $\ell 2$ norm, verify the triangle inequality for $||x+y||, ||y+z||, ||z+x||$
8. The $\ell 2$ norm also satisfies the Cauchy-Schwarz inequality $|v_1 \cdot v_2| <= ||v_1||_2 \, ||v_2||_2$; Verify this fact for $|x \cdot y|, |y \cdot z|, |z \cdot x|$;

2 Matrix Norms

A norm for $m \times n$ matrices is required to have the following properties:

- Positive: $0 \leq ||A||$, and 0 only if A is identically 0;
- Linear: $||\alpha A|| = |\alpha| \; ||A||$ for scalar α;
- Triangle inequality: $||A+B|| \leq ||A|| + ||B||$
- Submultiplicative: $||AB|| \leq ||A|| \; ||B||$

We need a matrix norm in order to measure the effect of a linear transformation on a vector. We want to make statements like this:

$$||A * x|| \leq ||A|| \, ||x||$$

but the right hand side, which involves both matrix and vector norms, is only valid if the matrix norm is "compatible" with the vector norm. If we are using the $\ell 1$, $\ell 2$, or $\ell\infty$ norm for vectors, then we typically use the corresponding $\ell 1$, $\ell 2$, or $\ell\infty$ norm for matrices, in order to make correct estimates.

- $\ell 1$ matrix norm: $||A||_1 = \max\limits_{1 \leq j \leq n} \sum_{i=1}^{m} |A_{i,j}|$
- $\ell 2$ matrix norm: $||A||_2 =$ square root of maximum eigenvalue of $A'A$;
- $\ell\infty$ matrix norm: $||A||_\infty = \max\limits_{1 \leq i \leq m} \sum_{j=1}^{n} |A_{i,j}|$

The MATLAB commands `norm(A,1)`, `norm(A,2)`, `norm(a,Inf)`, will compute the $\ell 1, \ell 2$, and $\ell\infty$ norms of matrix A respectively. The command `norm(A)` returns the value of the default $\ell 2$ norm.

Exercise 2:

1. Set the spatial dimension n to 5 ;
2. Define a random matrix `A=rand(n,n)`;
   - Loop 10 times:
   - Set `x = rand(n,1)`;
   - Compute `Ax = A*x`;
   - Print $||Ax||_1$ and $||A||_1 * ||x||_1$;
   - Print $||Ax||_2$ and $||A||_2 * ||x||_2$;
   - Print $||Ax||_\infty$ and $||A||_\infty * ||x||_\infty$;
   - end loop

3 The Operator Norm Controls Magnification

Suppose we are interested in what happens when a linear transformation ($m \times n$ matrix) A is applied to transform one column vector into another:

$$y = A * x$$

In particular, we would like to be able to say that if x is "small", then we can estimate that y will be "no bigger" than some limit. We measure the sizes of vectors using a vector norm, so really we want to say

$||x||$ *small* $\rightarrow ||A * x||$ *is small.*

If we define the norm of the matrix by:

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

then in fact we now have a guaranteed bound on $||A * x||$ for any x:

$$||Ax|| \leq ||Ax|| \, ||x||$$

Because there are many possible vector norms that can be used, it is necessary to keep in mind that the correct corresponding matrix norm must be used when making such estimates.

Exercise 3:

1. Set the spatial dimension $n = 2$;
2. Define an $n \times n$ matrix A with random entries;
3. Estimate the norm of this matrix by computing 100 random vectors x and taking the maximum ratio of $\frac{||A*x||}{||x||}$;
4. Ask MATLAB for the l2 norm of A and compare with your estimate:
   - Loop 20 times:
   - Compute a random n-vector x
   - Compute and print $\frac{||Ax||}{||x||}$;
   - end loop
5. Was it always true that $\frac{||Ax||}{||x||} \leq ||A||$?

Your results should show that $||A||$ limits how much multiplication by A can magnify any vector's size.

Why is this useful? Suppose your measured data includes a correct value plus a small error, so it has the form $x + \delta x$. If you apply the linear transform to your noisy data, you get $A * (x + \delta x)$. How far is this from what you would have gotten with the exact value?

$$||A * (x + \delta x) - A * x|| = ||A * \delta x|| \leq ||A|| \, ||\delta x||$$

4 The Condition Number

Define the condition number of a matrix A by:

$$\kappa(A) = ||A|| \, ||A^{-1}|| = \max_{||x|| \neq 0} \frac{||Ax||}{||x||} \min_{||x|| \neq 0} \frac{||x||}{||Ax||}$$

Suppose we are interested in the solving the linear system

$$A\,x = b$$

for which the exact solution is x^*, and we have arrived at a numerical estimated solution $\hat{x}$. We would like to know the norm of the *solution error*, defined by

$$e(\hat{x}) = ||\hat{x} - x^*||$$

but what we can compute is the norm of the *residual error*

$$r(\hat{x}) = ||A * \hat{x} - b||$$

It turns out that the condition number gives us a way of bounding the solution error in terms of the residual error. MATLAB can compute the condition number of a matrix by `cond(A)`.

Exercise 4:

1. Set spatial size n to 20 ;
2. Define the column vector `x=linspace(1,n,n);`;
3. Define a random diagonal matrix `D=diag(rand(n,1)`.
4. Define a random matrix `A=rand(n,n)`;
5. Define an orthogonal matrix $[Q, R] = qr(A)$;
6. Define the Hilbert matrix `H=hilb(n)`;
7. Compute b1=D*x, b2=A*x, b3=Q*x, b4=H*x;
8. Solve D*x1=b1, A*x2=b2, Q*x3=b3, H*x4=b4;
9. Compute $e1 = ||x1 - x||, \quad r1 = ||D * x1 - b1||, \quad \kappa(D)$;
10. Compute $e2 = ||x2 - x||, \quad r2 = ||A * x2 - b2||, \quad \kappa(A)$;
11. Compute $e3 = ||x3 - x||, \quad r3 = ||Q * x3 - b3||, \quad \kappa(Q)$;
12. Compute $e4 = ||x4 - x||, \quad r4 = ||H * x4 - b4||, \quad \kappa(H)$;

In what cases does a small residual not guarantee a small error? Your results should suggest that the condition number of the matrix associated with a linear algebra problem can be a warning sign. A very large condition number suggests that computed solutions may not be reliable, even though the residual norm is small.