Norms, Sensitivity and Conditioning

CS 205A:
Mathematical Methods for Robotics, Vision, and Graphics

Doug James (and Justin Solomon)

Announcements

- Reminder: HW0 due Thursday 11:59pm.
- `http://cs205a.stanford.edu` now works
- JuliaBox fine for online. Try Jupyter, or JuliPro+Atom, or JUNO, etc. for offline.

Questions

Gaussian elimination works in theory, but what about floating point precision?

How much can we trust $\vec{x}_0$ if $0 < \|A\vec{x}_0 - \vec{b}\| \ll 1$?

Recall: Backward Error

Backward Error

The amount a problem statement would have to change to realize a given approximation of its solution

Example 1: $\sqrt{x}$

Example 2: $A\vec{x} = \vec{b}$

Perturbation Analysis

How does $\vec{x}$ change if we solve
$$(A + \delta A)\vec{x} = \vec{b} + \delta\vec{b}?$$

Two viewpoints:

- Thanks to floating point precision, A and $\vec{b}$ are approximate
- If $\vec{x}_0$ isn't the exact solution, what is the backward error?

What is “Small?”

What does it mean for a statement to hold for *small* $\delta\vec{x}$?

What is "Small?"

What does it mean for a statement to hold for *small* $\delta\vec{x}$?

Vector norm

A function $\|\cdot\| : \mathbb{R}^n \to [0, \infty)$ satisfying:

1. $\|\vec{x}\| = 0$ iff $\vec{x} = 0$
2. $\|c\vec{x}\| = |c|\|\vec{x}\| \ \forall c \in R, \vec{x} \in \mathbb{R}^n$
3. $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\| \ \forall \vec{x}, \vec{y} \in \mathbb{R}^n$

Our Favorite Norm

$$\|\vec{x}\|_2 \equiv \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

p-Norms

For $p \geq 1$,

$$\|\vec{x}\|_p \equiv \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$$

Taxicab norm: $\|\vec{x}\|_1$

∞ Norm

$$\|\vec{x}\|_\infty \equiv \max\left(|x_1|, |x_2|, \ldots, |x_n|\right)$$

How are Norms Different?

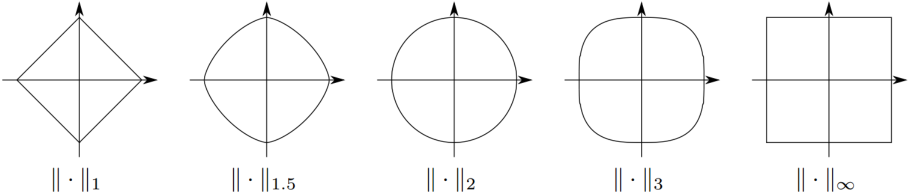

Figure 4.7 The set $\{\vec{x} \in \mathbb{R}^2 : \|\vec{x}\| = 1\}$ for different vector norms $\|\cdot\|$.

How are Norms the Same?

Equivalent norms

Two norms $||\cdot||$ and $||\cdot||'$ are *equivalent* if there exist constants c_{low} and c_{high} such that $c_{low}||\vec{x}|| \leq ||\vec{x}||' \leq c_{high}||\vec{x}||$ for all $\vec{x} \in \mathbb{R}^n$.

How are Norms the Same?

Equivalent norms

Two norms $\|\cdot\|$ and $\|\cdot\|'$ are *equivalent* if there exist constants c_{low} and c_{high} such that $c_{low}\|\vec{x}\| \leq \|\vec{x}\|' \leq c_{high}\|\vec{x}\|$ for all $\vec{x} \in \mathbb{R}^n$.

Theorem

All norms on $\mathbb{R}^n$ are equivalent.

How are Norms the Same?

Equivalent norms

Two norms $\|\cdot\|$ and $\|\cdot\|'$ are *equivalent* if there exist constants c_{low} and c_{high} such that $c_{low}\|\vec{x}\| \leq \|\vec{x}\|' \leq c_{high}\|\vec{x}\|$ for all $\vec{x} \in \mathbb{R}^n$.

Theorem

All norms on $\mathbb{R}^n$ are equivalent.

$(10000, 1000, 1000)$ vs. $(10000, 0, 0)$?

Matrix Norms: "Unrolled" Construction

Convert to vector, and use vector p-norm:

$$A \in \mathbb{R}^{m \times n} \leftrightarrow \texttt{a[:]} \in \mathbb{R}^{mn}$$

- Achieved by `vecnorm(A, p)` in Julia.

Special Case: Frobenius norm (p=2):

$$\|A\|_{\text{Fro}} \equiv \sqrt{\sum_{ij} a_{ij}^2}$$

Matrix Norms: "Induced" Construction

Maximum stretching of a unit vector by A:

$$\|A\| \equiv \max\{\|A\vec{x}\| : \|\vec{x}\| = 1\}$$

Matrix Norms: "Induced" Construction

Maximum stretching of a unit vector by A:

$$\|A\| \equiv \max\{\|A\vec{x}\| : \|\vec{x}\| = 1\}$$

Different matrix norms induced by different vector p-norms.

Case p=2: What is the norm induced by $\|\cdot\|_2$?

Matrix Norms: $\|A\|_2$ Visualization

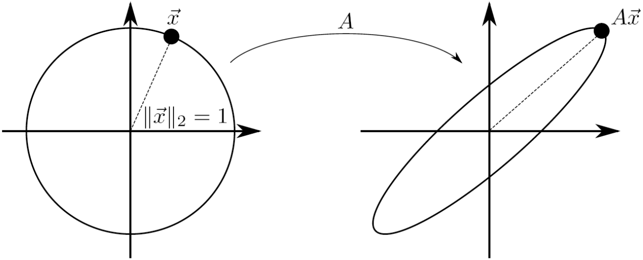

Figure 4.8 The norm $\|\cdot\|_2$ induces a matrix norm measuring the largest distortion of any point on the unit circle after applying A.

Induced two-norm, or *spectral norm*, of $A \in \mathbb{R}^{n\times n}$ is the square root of the largest eigenvalue of A^TA:

$$\|A\|_2^2 = \max\{\lambda : \text{there exists } \vec{x} \in \mathbb{R}^n \text{ with } A^TA\vec{x} = \lambda\vec{x}\}$$

Other Induced Norms

$$\|A\|_1 \equiv \max_j \sum_i |a_{ij}|$$

$$\|A\|_\infty \equiv \max_i \sum_j |a_{ij}|$$

Question

Are all matrix norms equivalent?

Recall: Condition Number

Condition number

Ratio of forward to backward error

Root-finding example:

$$\frac{1}{f'(x^*)}$$

Model Problem

$$(A + \varepsilon\, \delta A)\, \vec{x}(\varepsilon) = \vec{b} + \varepsilon\, \delta\vec{b}$$

Simplification (on the board!)

$$\left.\frac{d\vec{x}}{d\varepsilon}\right|_{\varepsilon=0} = A^{-1}(\delta\vec{b} - \delta A\,\vec{x}(0))$$

$$\frac{\|\vec{x}(\varepsilon) - \vec{x}(0)\|}{\|\vec{x}(0)\|} \leq |\varepsilon|\|A^{-1}\|\|A\|\left(\frac{\|\delta\vec{b}\|}{\|\vec{b}\|} + \frac{\|\delta A\|}{\|A\|}\right) + O(\varepsilon^2)$$

Condition Number

Condition number

The condition number of $A \in \mathbb{R}^{n\times n}$ for a given matrix norm $\|\cdot\|$ is $\operatorname{cond} A \equiv \kappa \equiv \|A^{-1}\|\|A\|$.

Condition Number

Condition number

The condition number of $A \in \mathbb{R}^{n \times n}$ for a given matrix norm $\|\cdot\|$ is $\operatorname{cond} A \equiv \kappa \equiv \|A^{-1}\|\|A\|$.

$$\text{Relative change: } D \equiv \frac{\delta \vec{b}}{\|\vec{b}\|} + \frac{\|\delta A\|}{\|A\|}$$

$$\frac{\|\vec{x}(\varepsilon) - \vec{x}(0)\|}{\|\vec{x}(0)\|} \leq \varepsilon \cdot D \cdot \kappa + O(\varepsilon^2)$$

Condition Number

Condition number

The condition number of $A \in \mathbb{R}^{n\times n}$ for a given matrix norm $\|\cdot\|$ is $\mathrm{cond}\, A \equiv \kappa \equiv \|A^{-1}\|\|A\|$.

Relative change: $D \equiv \frac{\delta\vec{b}}{\|\vec{b}\|} + \frac{\|\delta A\|}{\|A\|}$

$$\frac{\|\vec{x}(\varepsilon) - \vec{x}(0)\|}{\|\vec{x}(0)\|} \leq \varepsilon \cdot D \cdot \kappa + O(\varepsilon^2)$$

Invariant to scaling (unlike determinant!); equals one for the identity.

Condition Number of Induced Norm

$$\operatorname{cond} A = \frac{\max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|}{\|\vec{x}\|}}{\min_{\vec{y}\neq\vec{0}} \frac{\|A\vec{y}\|}{\|\vec{y}\|}} = \frac{\max_{\|\vec{x}\|=1} \|A\vec{x}\|}{\min_{\|\vec{y}\|=1} \|A\vec{y}\|}$$

Condition Number: Visualization

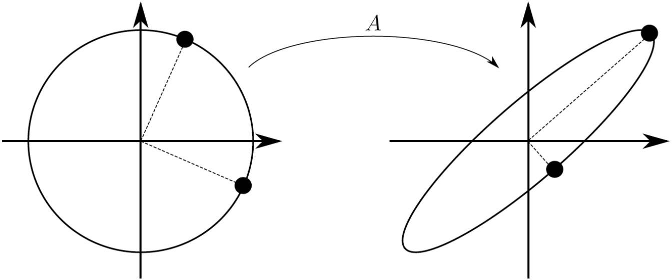

Figure 4.9 The condition number of A measures the ratio of the largest to smallest distortion of any two points on the unit circle mapped under A.

Experiments with an ill-conditioned Vandermonde matrix

Chicken $\Longleftrightarrow$ Egg

$$\operatorname{cond} A \equiv \|A\| \boxed{\|A^{-1}\|}$$

Computing $\|A^{-1}\|$ typically requires solving $A\vec{x} = \vec{b}$, but how do we know the reliability of $\vec{x}$?

To Avoid...

What is the condition number of computing the condition number of A?

To Avoid...

What is the condition number of computing the condition number of A?

What is the condition number of computing what the condition number is of computing the condition number of A?

$\vdots$

Instead

Bound the condition number.

- Below: Problem is at *least* this hard
- Above: Problem is at *most* this hard

Potential for Approximation

$$\|A^{-1}\vec{x}\| \leq \|A^{-1}\|\|\vec{x}\|$$
$$\Downarrow$$
$$\operatorname{cond} A = \|A\|\|A^{-1}\| \geq \frac{\|A\|\|A^{-1}\vec{x}\|}{\|\vec{x}\|}$$

Next