CSC338 Numerical Methods Lecture 5

Today: Special cases of $A\underline{x} = \underline{b}$; other methods

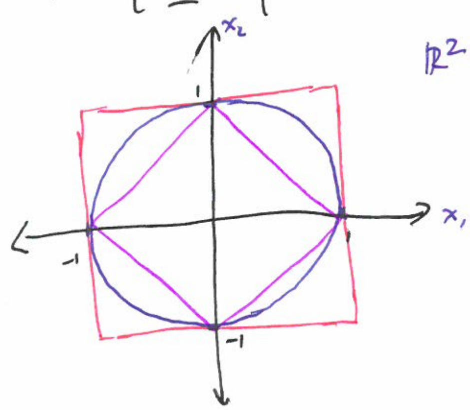
Systems of Linear Equations. where $A \in \mathbb{R}^{m \times n}$ $m > n$

(overdetermined system)

Geometry of vector & matrix norms

Def: The unit sphere in $\mathbb{R}^n$ with respect to a vector norm $\|\cdot\|$ is the set of points $\underline{x} \in \mathbb{R}^n$ distance 1 from a fixed central point, e.g. the origin.

$\hookrightarrow \{\underline{x} \mid \|\underline{x}\| = 1\}$



$L_1 : \{\|\underline{x}\|_1 = 1\}$

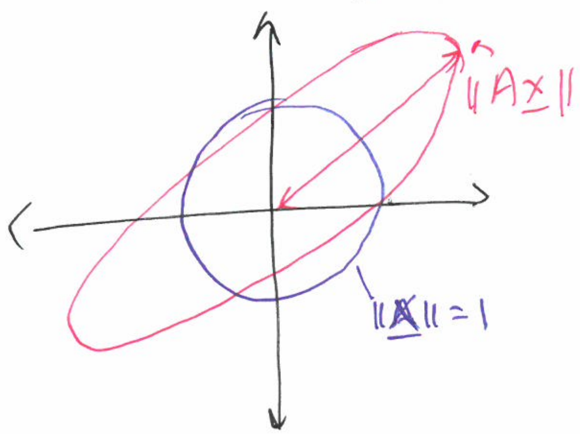
$L_2 : \{\|\underline{x}\|_2 = 1\}$

$L_\infty : \{\|\underline{x}\|_\infty = 1\}$

The induced matrix norm is defined based on how A acts on the unit sphere

$$\|\underline{A}\| = \max_{\|\underline{x}\|=1} \|A\underline{x}\|$$

~ "largest stretch" when A is applied to the unit sphere



Gauss-Jordan Elimination

When performing Gauss Elimination, we reduce a matrix (system) to an upper triangular form:

eg//
$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

With Gauss-Jordan Elimination, we reduce the matrix to a diagonal form:

eg//.

$$R_2 \leftarrow R_2 - R_3 \qquad \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \underline{x} = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$$

$$R_1 \leftarrow R_1 + \tfrac{1}{2}R_2 - 2R_3 \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \underline{x} = \begin{bmatrix} 5.5 \\ -1 \\ 3 \end{bmatrix}$$

- LU factorization takes $\frac{n^3}{3}$ multiplications (similar # adds).
- forward / back prop $\sim n^2$.
- LU factorization can be done in place.

$$\left[\; A \;\right] \Rightarrow \left[\; L \;\; U \;\right]$$

- G.J. requires ~50% more flops than LU factorization $\left(\frac{n^3}{2}\right)$
- Last step requires n divisions $O(n)$

Iterative Refinement

Suppose $\underline{x_0}$ is a computed solution to $A\underline{x} = \underline{b}$.

with a residual $\underline{r_0} = \underline{b} - A\underline{x_0} \neq \underline{0}$.

We can obtain a better solution by solving $A\underline{z_0} = \underline{r_0}$

So if we let $\underline{x_1} := \underline{x_0} + \underline{z_0}$

$$
\begin{aligned}
A\underline{x_1} &= A(\underline{x_0} + \underline{z_0}) \\
&= A\underline{x_0} + A\underline{z_0} \\
&= (\underline{b} - \underline{r_0}) + \underline{r_0} \\
&= \underline{b}.
\end{aligned}
$$

If $\underline{z_0}$ is not exact, we can repeat this process.

Iterative refinement is not practical for this problem, but is useful idea in general.

Modified Problems.

If we have a solution to $A\underline{x} = \underline{b}$, and ~~either~~ want a solution to

① $A\underline{x} = \underline{b'}$ (different $\underline{b}$)

Keep the LU factorization, do one more fwd/backwd. pass

② $A'\underline{x} = \underline{b}$ (different A)

In general, we have to start over, but there are some special cases.

eg//

$$A = \begin{bmatrix} 1 & 2 & 5 \\ 3 & -1 & 2 \\ -1 & -5 & 0 \end{bmatrix} \qquad A' = \begin{bmatrix} 4 & 2 & 5 \\ 3 & -1 & 2 \\ -1 & -5 & 0 \end{bmatrix}$$

only this elt is different.

<u>Def</u> A rank one update to a matrix A can be expressed as $A' = A - \underline{u}\,\underline{v}^T$ where $\underline{u}, \underline{v} \in \mathbb{R}^n$

eg//

$$\begin{bmatrix} 4 & 2 & 5 \\ 3 & -1 & 2 \\ -1 & -5 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 5 \\ 3 & -1 & 2 \\ -1 & -5 & 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}^T$$

Sherman Morrison Formula

$$(\underset{n\times n}{A} - \underset{n\times 1}{\underline{u}}\,\underset{1\times n}{\underline{v}^T})^{-1} = \underset{n\times n}{A^{-1}} + \underset{n\times n}{A^{-1}}\,\underset{n\times 1}{\underline{u}}\,\underbrace{(1 - \underset{1\times n}{\underline{v}^T}\,\underset{n\times n}{A^{-1}}\,\underset{n\times 1}{\underline{u}})^{-1}}_{\text{Scalar}}\,\underset{1\times n}{\underline{v}^T}\,\underset{n\times n}{A^{-1}}$$

If we have a system $A'\underline{x} = \underline{b}$

$$\Rightarrow \quad \underline{x} = (A')^{-1}\underline{b}$$

$$\underline{x} = (A - \underline{u}\,\underline{v}^T)^{-1}\underline{b}$$

$$= \underbrace{A^{-1}\underline{b}}_{\underline{y}} + \underbrace{A^{-1}\underline{u}}_{\underline{z}}\,(1 - \underline{v}^T \underbrace{A^{-1}\underline{u}}_{\underline{z}})^{-1}\,\underline{v}^T \underbrace{A^{-1}\underline{b}}_{\underline{y}}$$

Strategy:

1. Factorize $A = LU$
2. Solve $A\underline{y} = \underline{b}$
3. Solve $A\underline{z} = \underline{u}$
4. Compute: $\underline{x} = \underline{y} + \underline{z}\,(1 - \underline{v}^T\underline{z})^{-1}\,\underline{v}^T\underline{y}$.

Cholesky Factorization

Def An $n \times n$ matrix A is symmetric if $A^T = A$
(or $a_{ij} = a_{ji}$ for all i, j)

Def An $n \times n$ matrix A is positive definite if $\underline{x}^T A \underline{x} > 0$
for all $\underline{x} \neq \underline{0}$.

Intuition Let $\underline{y} = A\underline{x}$ (matrix A applied to $\underline{x}$)

$$\underline{x}^T A \underline{x} = \underline{x}^T \underline{y} = \|\underline{x}\|_2 \|\underline{y}\|_2 \cos(\theta)$$

where θ is the angle between $\underline{x}$ and $\underline{y}$

So $\underline{x}^T \underline{y} > 0$ then $\cos(\theta) > 0$.

$$\Rightarrow \theta < \frac{\pi}{2} \quad (90^\circ)$$

Fact Let M be a $m \times n$ matrix with $m > n$, $\text{rank}(M) = n$
Then $M^T M$ is positive definite. (HW5)

If $A \in \mathbb{R}^{n \times n}$ is positive definite, then the LU factorization $A = LU$ can be arranged so that $U = L^T$ so $A = LL^T$.
The factorization $A = LL^T$ is called the Cholesky Factorization.

Computation

$$A = L \times L^T$$

$$\begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots \\ a_{21} & a_{22} & a_{32} & \cdots \\ a_{31} & a_{32} & \ddots & \\ \vdots & \vdots & & \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & \cdots \\ l_{21} & l_{22} & 0 & \cdots \\ l_{31} & l_{32} & l_{33} & \cdots \\ \vdots & & & \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} & \cdots \\ 0 & l_{22} & l_{32} & \cdots \\ 0 & 0 & l_{33} & \cdots \\ \vdots & \vdots & & \ddots \end{bmatrix}$$

Can directly compute.

$$a_{11} = l_{11} \cdot l_{11} \Rightarrow l_{11} = \sqrt{a_{11}}$$

$$a_{21} = l_{11} \cdot l_{21} \Rightarrow l_{21} = \frac{a_{21}}{l_{11}}$$

$$a_{31} = l_{11} \cdot l_{31} \Rightarrow l_{31} = \frac{a_{31}}{l_{11}}$$

$$\vdots \qquad\qquad \vdots$$

$$a_{22} = l_{21}^2 + l_{22}^2 \Rightarrow l_{22} = \sqrt{a_{22} - l_{21}^2}$$

$$a_{23} = l_{21} l_{31} + l_{22} \cdot l_{32} \Rightarrow l_{32} = \frac{a_{23} - l_{21} \cdot l_{31}}{l_{22}}.$$

$$\vdots \qquad\qquad \vdots$$

Properties

1. Since A is positive definite, the square roots are all positive values
2. We don't need pivoting for numerical stability.
3. Half as much work as LU factorization $\left(\sim \frac{n^3}{6} \text{ x's, similar +'s}\right)$ and half as much storage.

Linear Least Squares Problem (New problem, §3)

We have a system of linear equations $A\underline{x} = \underline{b}$ but with $A \in \mathbb{R}^{m \times n}$ with $m > n$. This system is overdetermined.

we generally don't have exact solutions $A\underline{x} = \underline{b}$.

we can choose $\underline{x}$ to minimize $\| A\underline{x} - \underline{b} \|_2$. — the L_2 norm of the residual.

eg// We want to predict a student's hw3 grade given their hw1 and hw2 grades. We want to build a simple model:

$$x_1 a_1 + x_2 a_2 = b$$

Coefficients: x_1, x_2 — a_1: hw1 grade — a_2: hw2 grade — b: hw3 grade

How can we find good values of x_1 and x_2?

If we have data from students who submitted all 3 homeworks, then we have data in the form:

$a_1^{(1)}, a_2^{(1)}, b^{(1)}$ ~ hw grade for student 1

$a_1^{(2)}, a_2^{(2)}, b^{(2)}$ ~ - - - 2.

$\vdots$

$a_1^{(73)}\ a_2^{(73)}, b^{(73)}$

Ideally we would find x_1, x_2 so that

$$x_1 a_1^{(i)} + x_2 a_2^{(i)} = b^{(i)} \quad \text{for all } i \in \{1, 2, \ldots, 73\}$$

writing in matrix form: $A\underline{x} = \underline{b}$ where:

$$A = \begin{bmatrix} a_1^{(1)} & a_2^{(1)} \\ a_1^{(2)} & a_2^{(2)} \\ \vdots & \\ a_1^{(73)} & a_2^{(73)} \end{bmatrix} \qquad \underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} b^{(1)} \\ b^{(2)} \\ \vdots \\ b^{(73)} \end{bmatrix}$$

73×2 $\qquad$ 2×1 $\qquad$ 73×1

$m = 73$
$n = 2.$

$\cancel{A\underline{x} = \underline{b}}$

But the system is overdetermined, so we minimize

$$\| A\underline{x} - \underline{b} \|_2$$

<u>Fact</u> The existence of $\underline{x}$ that minimizes $\|A\underline{x}-\underline{b}\|_2$ is guaranteed, but $\underline{x}$ might not be unique if $\text{rank}(A) < n$.

Normal Equation

Since $\|A\underline{x}-\underline{b}\|_2 = \sqrt{(A\underline{x}-\underline{b})^T(A\underline{x}-\underline{b})}$, minimizing $\|A\underline{x}-\underline{b}\|_2$ is equivalent to minimizing the function

$$\phi(\underline{x}) = (A\underline{x}-\underline{b})^T(A\underline{x}-\underline{b}) \qquad \phi: \mathbb{R}^n \to \mathbb{R}.$$

Method: Set the gradient $\nabla\phi(x)$ to $\underline{0}$:

$$\underbrace{\nabla\phi(\underline{x})}_{n\times 1} = \frac{\partial\phi}{\partial\underline{x}} = 2\underbrace{A^T}_{n\times m}\underbrace{A}_{m\times n}\underbrace{\underline{x}}_{n\times 1} - 2\underbrace{A^T}_{n\times m}\underbrace{\underline{b}}_{m\times 1} = 2A^T(A\underline{x}-\underline{b})$$

$$\nabla\phi(x) = 0 \iff \boxed{\underbrace{A^TA}_{n\times n}\underline{x} = \underbrace{A^T\underline{b}}_{n\times 1}}$$

The normal equation

If A is full rank ($\text{rank}(A)=n$) then A^TA is positive definite.

We can solve the system using Cholesky factorization.