March 19th
Tuesday, March 19, 2024 1:59 PM

Linear least Square Problems

$Ax=b$ we know A and b

The goal is to find an n by 1 vector x minimizing $\|Ax-b\|_2$

A is an m by n matrix

If $m=n$, and A is nonsingular, then $x=A^{-1}b$

If $m>n$, that is, there are more equations than unknowns, then the system is overdetermined, generally there is no <u>one</u> solution to an overdetermined,

If $m<n$ the system is undetermined.

we will focus on "overdetermined cases"

Three applications of least squares problems

- Curve fitting
- statistical modeling of "noisy" data
- geodetic modeling,

Three standard ways to solve the least squares problem:

- The normal equations
- QR decomposition
- Singular value decomposition (SVD)

Example "Curve fitting"

Suppose we have m pairs of numbers $(y_1,b_1),(y_2,b_2),\ldots(y_m,b_m)$. We want to find the "best" cubic polynomial fit to b_i as a function of y_i.

$$f(y)=x_1y^3+x_2y^2+x_3y+x_4$$

Cubic polynomial fit to ...

Cubic function $f(y) = x_1 y^3 + x_2 y^2 + x_3 y + x_4$

Where we need identify the x_i that minimize the residual

$r_i \equiv f(y_i) - b_i$ for $i = 1$ to m.

$$r \equiv \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} = \begin{bmatrix} f(y_1) \\ f(y_2) \\ \vdots \\ f(y_m) \end{bmatrix} - \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} 1 & y_1 & y_1^2 & y_1^3 \\ 1 & y_2 & y_2^2 & y_2^3 \\ \vdots & & & \\ 1 & y_m & y_m^2 & y_m^3 \end{bmatrix} \begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$= A \cdot x - b$$

We want to minimize the residual.

We would want to minimize $\|r\|$ for some norm.

We could $\|r\|_\infty$, $\|r\|_1$, or $\|r\|_2$.

The last one $\sqrt{\sum r_i^2}$ to minimize this is equivalent to minimizing $\sum r_i^2$ is a linear least squares problem.

$b = \sin(\pi y/5) + y/5$ at the points $y = -5, -4.5, -4, \ldots 5.5, 6$.

<u>Note</u>: As the degree increases from 1 to 17 the residual norm decreases.

But when we reach degree 18, the residual norm increases dramatically.

Typically, polynomial is kept to lower degrees.

<u>Example</u> linear regression in statistics is very commonly used.

Get a best fit line to the data provided.

See textbook for details and graphs (chapter 3 pages 102 to 105)

See textbook ...
(chapter 3 pages 102 to 105)

Hopefully we will have the opportunity to show
Image compression can be interpreted using least squares problems

$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ $\quad y^T y = y_1^2 + y_2^2 + \dots y_m^2$

Normal Equations

To derive the normal equations looking for x
where the gradient of $\|Ax-b\|_2^2$ vanishes.

$$\|Ax-b\|_2^2 = (Ax-b)^T(Ax-b).$$

We want to find x such that the gradient of $\|Ax-b\|_2^2$ is zero

$$0 = \lim_{h \to 0} \frac{[A(x+h)-b]^T[A(x+h)-b] - [Ax-b]^T[Ax-b]}{\|h\|_2}$$

$$= \lim_{h \to 0} \frac{2h^T(A^TAx - A^Tb) + h^TA^TAh}{\|h\|_2}$$

The second term $\dfrac{|h^TA^TAh|}{\|h\|_2} \le \dfrac{\|A\|_2^2\|h\|_2^2}{\|h\|_2} = \|A\|_2^2\|h\|_2$

We have $\|h\|_2 \to 0$ thus the second term approaches zero.

Thus, $A^TAx - A^Tb = 0$ or $\underline{\underline{A^TAx = A^Tb}}$

This is a system of n linear equations in n unknowns

Example Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}$ with $b = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$

Because A is a 3 by 2 matrix the system $Ax = b$ is overdetermined.

Because A is a ...

$Ax = b$ is overdetermined.

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Let's calculate $\|Ax-b\|_2^2$

$$Ax - b = \begin{bmatrix} 2x_1 \\ x_2 \\ x_1+2x_2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 1 \\ x_2 \\ x_1+2x_2-3 \end{bmatrix}$$

$$\|Ax-b\|_2^2 = (2x_1-1)^2 + x_2^2 + (x_1+2x_2-3)^2$$

$$D(\|Ax-b\|_2^2) = \left\langle \frac{\partial}{\partial x_1}\left((2x_1-1)^2 + x_2^2 + (x_1+2x_2-3)^2\right), \frac{\partial}{\partial x_2}\left((2x_1-1)^2 + x_2^2 + (x_1+2x_2-3)^2\right) \right\rangle$$

$$= \langle 4(2x_1-1) + 2(x_1+2x_2-3), 2x_2 + 4(x_1+2x_2-3) \rangle$$

$$= \langle 10x_1 + 4x_2 - 10, 4x_1 + 10x_2 - 12 \rangle$$

and
$$\begin{cases} 10x_1 + 4x_2 - 10 = 0 \\ (4x_1 + 10x_2 - 12 = 0) \leftarrow -\frac{10}{4} \end{cases}$$

$$\begin{cases} 10x_1 + 4x_2 - 10 = 0 \\ -10x_1 - 25x_2 + 30 = 0 \end{cases} \Rightarrow \quad -21x_2 + 20 = 0$$

$$-21x_2 = -20$$

$$x_2 = \frac{20}{21}$$

$-\frac{210}{80}$

$$10x_1 + 4\left(\frac{20}{21}\right) - 10 = 0$$

$$10x_1 + \frac{80}{21} - \frac{210}{21} = 0$$

$$10x_1 = \frac{130}{21}$$

$$x_1 = \frac{13}{21}$$

$$x = \begin{bmatrix} 13/21 \\ 20/21 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ \cdots \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

$$x = \begin{bmatrix} \cdot/21 \\ \frac{20}{21} \end{bmatrix} \qquad A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} \cdot \\ 0 \\ 3 \end{bmatrix}$$

$$\frac{13}{21} + \frac{40}{21} \qquad Ax = \begin{bmatrix} \frac{26}{21} \\ \frac{20}{21} \\ \frac{53}{21} \end{bmatrix} \neq b$$

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

$$\boxed{A^T A x = A^T b}$$

$$A^T = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \qquad A^T A = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \qquad = \begin{bmatrix} 4+1 & 2 \\ 2 & 4+1 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix}$$

$$A^T A x = A^T b$$

$$\begin{bmatrix} 5 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\begin{cases} 5x_1 + 2x_2 = 5 \\ 2x_1 + 5x_2 = 6 \ast \left(-\frac{5}{2}\right) \end{cases}$$

$$\begin{cases} 5x_1 + 2x_2 = 5 \\ -5x_1 - \frac{25}{2} x_2 = -15 \end{cases}$$

$$-\frac{21}{2} x_2 = -10$$

$$x_2 = \frac{20}{21}$$

$$5x_1 + 2\left(\frac{20}{21}\right) = 5$$

$$x_1 + \frac{8}{21} = 1$$

$$x_1 + \frac{8}{21} = 1$$

$$x_1 = \frac{13}{21}$$

we also get $x = \begin{bmatrix} 13/21 \\ 20/21 \end{bmatrix}$.

$$A^T A x = A^T b$$

$x = (A^T A)^{-1} A^T b$ why does this minimize $\|Ax-b\|_2^2$?

Note $A^T A$ is positive definite

the function is strictly convex

thus, any critical point is a global minimum.

or

Let $x' = x + h$

$$\|Ax'-b\|_2^2 = (Ax'-b)^T(Ax'-b) = (Ah + Ax - b)^T(Ah + Ax - b)$$

$$= (Ah)^T(Ah) + (Ax-b)^T(Ax-b) + 2(Ah)^T(Ax-b)$$

$$= \|Ah\|_2^2 + \|Ax-b\|_2^2 + 2h^T\underbrace{(A^TAx - A^Tb)}_{0}$$

$A^TAx = A^Tb$

$$= \|Ah\|_2^2 + \|Ax-b\|_2^2$$

Since A^TA is symmetric and positive definite we can Cholesky decomposition to solve the normal equations.

The total cost of computing A^TA, A^Tb and the Cholesky decomposition is $n^2m + \frac{1}{3}n^3 + O(n^2)$ flops.

Note $m > n$, the term n^2m dominates the cost.

flops:

Note $m \geq n$, the term $n^2 m$ dominates the cost.

QR decomposition

Theorem Let A be m by n matrix with $m \geq n$. Suppose that A has full column rank, then there exists a unique m by n orthogonal matrix Q ($Q^T Q = I_n$) and a unique n by n upper triangular matrix R with positive diagonals $r_{ii} > 0$ such that

$A = QR$.

Orthogonal Matrices

A matrix $Q \in \mathbb{R}^{n \times n}$ is said to be orthogonal if $QQ^T = I$.

This equation says that Q has an inverse and $Q^{-1} = Q^T$.

Theorem If $Q \in \mathbb{R}^{n \times n}$ is orthogonal then for all $x, y \in \mathbb{R}^n$

inner product

(a) $\langle Qx, Qy \rangle = \langle x, y \rangle$ (b) $\|Qx\|_2 = \|x\|_2$