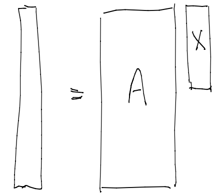
lecture 18

least squares, SVD

Least Squares: version 1

Given an $m \times n$ matrix A, find an n-vector $\vec{x}$ that minimizes $\|A\vec{x}\|^2$, subject to $\|\vec{x}\| = 1$.



$m \geq n$

Use method of Lagrange multipliers:

Minimize $\|A\vec{x}\|^2 + \lambda(\vec{x}^T\vec{x} - 1)$

Idea: The expression to be minimized is quadratic in $\vec{x}$ and has a unique minimum when for any $\lambda \geq 0$.

Take derivatives with respect to each x_i and set to 0, gives a set of equations:

$$0 = \frac{\partial}{\partial \vec{x}}\left[\vec{x}^T A^T A \vec{x} + \lambda(\vec{x}^T\vec{x} - 1)\right]$$

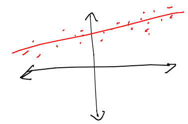
$$= 2A^T A\vec{x} + 2\lambda\vec{x}$$

$$\Rightarrow A^T A\vec{x} = -\lambda\vec{x}$$

$\Rightarrow$ $\vec{x}$ is an eigenvector of $A^T A$

But minimizing $\|\vec{x}^T A^T A\vec{x}\|$

$\Rightarrow$ $\vec{x}$ is the eigenvector of $A^T A$ with smallest eigenvalue

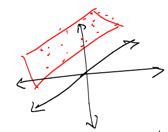


Fit line to a set of points in 2D.

Minimize the sum of squares of:

$$\begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} \\ \vdots & \\ x_m - \bar{x} & y_m - \bar{y} \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix}$$

subject to $|(a,b)| = 1$



Fit plane to a set of points in 3D.

Minimize the sum of squares of:

$$\begin{bmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_n - \bar{x} & y_n - \bar{y} & z_n - \bar{z} \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

subject to $\|(a,b,c)\| = 1$

Least Squares: Version 2:

Given $m \times n$ matrix A and m-vector $\vec{b} \neq 0$, find $\vec{x}$ that minimizes

$$\|A\vec{x} - \vec{b}\|^2.$$

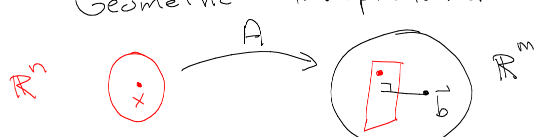
$$0 = \frac{\partial}{\partial \vec{x}}(A\vec{x} - \vec{b})^T(A\vec{x} - b)$$

$$\Rightarrow 0 = 2A^T A\vec{x} - 2A^T b$$

$$\Rightarrow \underbrace{A^T A}_{n \times n}\vec{x} = \underbrace{A^T b}_{n \times 1}$$

and solve for x using basic linear algebra methods, assuming that A has rank n i.e. invertible.

Geometric Interpretation



$\vec{b}$ can be written as the sum of a vector in the column space of A and a vector perpendicular to column space of A. The solution is the former, i.e. $A^T(A\vec{x} - \vec{b}) = \vec{0}$.

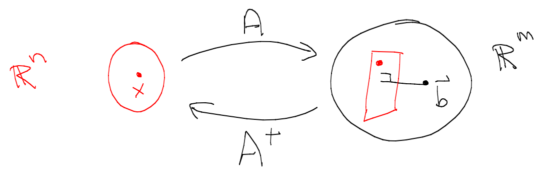
Suppose the columns of A are linearly independent. Then $A^T A$ is invertible (prove that on your own):

$$\therefore A^T A\vec{x} = A^T\vec{b}$$

$$\Rightarrow \vec{x} = \underbrace{(A^T A)^{-1} A^T}_{\text{called the "pseudo inverse"}}\vec{b}$$

$$= A^+ b$$

i.e. gives the least squares solution.



$A^+ A = I$ $\quad$ But $A A^+ = I$

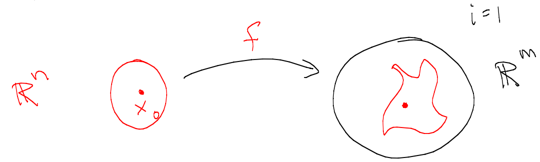
$n \times n$

always. $\quad$ only if A is invertible ($m = n$, in particular)

Non-linear least squares

$\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$

Given $\vec{x}_0 \in \mathbb{R}^n$, we want a nearby x that minimizes $\| \vec{f}(\vec{x}) \|$ i.e. minimize $\sum_{i=1}^{m} f_i(\vec{x})^2$.



Look at $\vec{f}(\vec{x})$ in local neighborhood of $\vec{x}_0$.

"Jacobian"

$$\vec{f}(\vec{x}) \approx \vec{f}(\vec{x}_0) + \left.\frac{\partial \vec{f}}{\partial x}\right|_{x = x_0} (\vec{x} - \vec{x}_0)$$

Minimize

$$\| f(\vec{x}) \|^2 \approx \left\| \vec{f}(\vec{x}_0) + \left.\frac{\partial \vec{f}}{\partial \vec{x}}\right|_{\vec{x}=\vec{x}_0} (\vec{x} - \vec{x}_0) \right\|^2$$

$$\vec{x}^{(k+1)} \leftarrow \vec{x}^{(k)} + \Delta x$$

i.e. $\vec{x} - \vec{x}_0$

"Gauss-Newton" method

Example : minimize over $\vec{h}$

$$\sum_{x \in Nbd(x_0, y_0)} \left(I(\vec{x} + \vec{h}) - J(\vec{x}) \right)^2$$

interpretation: $f_i(\vec{h}) = I(\vec{x}_i + \vec{h}) - J(\vec{x}_i)$

In our solution, we linearized $I(\vec{x} + \vec{h})$ at $h = 0$, solved for $\vec{h}$, then iterated.

$$h^{(k+1)} \leftarrow h^{(k)} + h$$

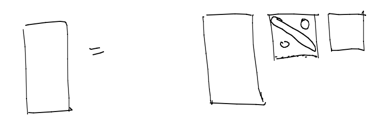
Singular Value Decomposition (SVD)

Any $m \times n$ matrix A can be written

assume $m > n$ here

$$A = U \Sigma V^T$$

$m \times \quad m \times n \quad n \times n$



i.e. any linear transformation is composed of rotate/reflect, scale/embed, rotate/reflect.

$A^T A$ is symmetric, $n \times n$

$\therefore$ eigenvalues are non-negative, σ_i^2

Without loss of generality, eigenvectors are orthonormal.

Define n eigenvectors V and n eigenvalues $\sigma_i = \Sigma_{ii}$.

$$A^T A V = V \Sigma^2$$

σ_i are the "singular values".

$$A^T A V = V \Sigma^2$$

Define $\tilde{U} \equiv A V$

$$\therefore \quad \tilde{U}^T \tilde{U} = V^T A^T A V = V^T V \Sigma^2 = \Sigma^2$$

$\therefore$ columns of $\tilde{U}$ are orthogonal and have length σ_i

Define U to have normalized columns of $\tilde{U}$, i.e. $\tilde{U} = U \Sigma$.

Then $\tilde{U} = A V \Rightarrow U \Sigma = A V$

$$\Rightarrow U \Sigma V^T = A$$

Note: $\quad$ Matlab

$$[U, \Sigma, V] = \mathrm{svd}(A)$$