Linear Least-Squares

$$\underset{m\times n}{A}\,\underset{n\times 1}{x} = \underset{m\times 1}{b} \iff b = x_1A_1 + x_2A_2 + \dots + x_nA_n$$

solvable precisely when b is a linear combination of the columns of A

Even if a "genuine" (exact) solution does not exist, we can find approximate solutions, which minimise the distance (in some norm) between Ax and b. Using the Euclidean norm leads to a linear least-squares problem:

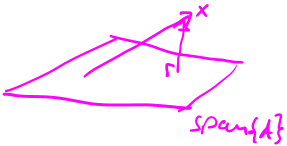
$$\underset{x}{\text{minimise}}\ \|Ax-b\|_2 \iff \underset{x}{\text{minimise}}\ \|Ax-b\|_2^2$$

Define cost function $\mathcal{E}(x) = \|Ax-b\|^2$

If $r = Ax-b$ (residuals), then

$$\mathcal{E}(x) = r^Tr$$

Linear Least-Squares



Minimising $\mathcal{E}(x) = \|Ax-b\|^2 = (Ax-b)^T(Ax-b) = x^TA^TAx - \underbrace{x^TA^Tb}_{=b^TAx} - b^TAx + b^Tb =$

$$= x^TA^TAx - 2b^TAx + b^Tb$$

Differentiate wrt x:

$\frac{\partial \mathcal{E}}{\partial x} = 2x^TA^TA - 2b^TA$, now set equal to zero and transpose $\Rightarrow$ $2\underbrace{A^T}_{n\times m}\underbrace{Ax}_{m\times 1} - 2A^Tb = 0$ (n equations)

$\Leftrightarrow \underline{A^TAx = A^Tb}$ normal equations

If A has $\underline{\text{full rank}}$ (A^TA will have full rank), then $x = \boxed{(A^TA)^{-1}A^T}b$

pseudo inverse: A^+

rank A = min(m,n)

If A is square and invertible: $A^+ = (A^TA)^{-1}A^T = A^{-1}(A^T)^{-1}A^T = A^{-1}$

Approximating the Null Space

The inhomogeneous method:

$Ax=0$ (homogeneous), always solvable, $x=0$ works!

Sometimes $x=0$ is not "valid" for the problem under consideration!

Force $x\neq 0$ by letting $\overset{m\times n}{A} = (\overset{m\times(n-1)}{A_0}\ \ \overset{m\times 1}{b})$ and $\overset{n\times 1}{x} = \begin{pmatrix}\overset{(n-1)\times 1}{x_0} \\ 1\end{pmatrix}$,

Now, $Ax=0 \iff (A_0\ b)\begin{pmatrix}x_0\\1\end{pmatrix}=0 \iff A_0x_0+b=0 \iff A_0x_0=-b.$

Solving in least-squares sense,

$$x_0 = A_0^{+}(-b) = -(A_0^TA_0)^{-1}A_0b \Rightarrow x = \begin{pmatrix}-(A_0^TA_0)^{-1}A_0b \\ 1\end{pmatrix}.$$

Singular Value Decomposition (SVD)

diagonal

$$\underbrace{A}_{m\times n} = \underbrace{U}_{m\times m}\,\underbrace{S}_{m\times n}\,\underbrace{V^*}_{n\times n} = \underbrace{\begin{pmatrix} | & | & | & | \end{pmatrix}}_{m\times m}\underbrace{\begin{pmatrix} \ddots \end{pmatrix}}_{m\times n}\underbrace{\begin{pmatrix} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{pmatrix}}_{n\times n} = USV^T$$

unitary (orthogonal)

if real valued A

$$S = \begin{pmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_p \\ & & & \\ & & & \end{pmatrix}$$

where $p = \min(m,n)$, and

$$s_1 \geq s_2 \geq \dots \geq s_r > s_{r+1} = \dots = s_p = 0$$

r is the number of nonzero singular values of A $= \operatorname{rank} A$

SVD and Linear Least-Squares

$Ax=b$ if $A=USV^T$, then

$$E(x)=\|Ax-b\|^2=\|USV^Tx-b\|^2=\|\underbrace{U^TU}_{=I}SV^Tx-U^Tb\|^2=\|S\underbrace{V^Tx}_{y}-\underbrace{U^Tb}_{c}\|^2=$$

$$=\|(s_1y_1,s_2y_2,\dots,s_ry_r,0,\dots,0)-(c_1,\dots,c_m)\|^2$$

This is minimized when $y_1=\frac{c_1}{s_1}$, $y_2=\frac{c_2}{s_2}$, …, $y_r=\frac{c_r}{s_r}$

$$S^+=\begin{pmatrix}\frac{1}{s_1}&&&\\&\frac{1}{s_2}&&\\&&\ddots&\\&&&0\end{pmatrix}$$

$$\Rightarrow y=\Big(\frac{c_1}{s_1},\frac{c_2}{s_2},\dots,\frac{c_r}{s_r},t_{r+1},\dots,t_m\Big)\Rightarrow x=Vy=\underbrace{VS^+U^Tb}_{\text{pseudo-inverse of }A}+\underbrace{t_{r+1}v_{r+1}+\dots+t_mv_m}_{\text{parametrization of the null space of }A}$$

arbitrary!
(they do not affect $E(x)$)

$\|x\|$ will be smallest if $t_{r+1}=\dots=t_m=0$

Approximating the Null Space

The homogeneous method:

$$Ax=0 \Rightarrow x = \cancel{A^+0} + t_{r+1}v_{r+1} + \dots + t_m v_m$$

if A has full rank, we won't get this part (m=r)

rightmost columns in V.

The best (as measured in the Euclidean norm) q-dimensional approximation of the null space of A is spanned by the q rightmost vectors in V:

$$V_{m-q+1}, V_{m-q+2}, \dots, V_{m-1}, V_m$$

Rank vs Numerical Rank, SVD Profile

In theory: rank A = number of non-zero singular values of A

$$A = USV^T = U\begin{pmatrix} s_1 & & & & & \\ & \ddots & & & & \\ & & s_r & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{pmatrix} V^T$$

In practice: what is non-zero?

We must consider relative sizes of the singular values!

One alternative is to consider the ratios $\dfrac{s_{k+1}}{s_k}$ and say

that s_{k+1} is "zero" if the quotient is really small

Hartley Normalisation

Homogeneous coordinates for typical points in a digital image: $y = \begin{pmatrix} 2000 \\ 1500 \\ 1 \end{pmatrix}$

- Very sensitive to changes in the final coordinate! $y = \begin{pmatrix} 2000 \\ 1500 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 2000 \\ 1500 \\ 2 \end{pmatrix} \sim \begin{pmatrix} 1000 \\ 750 \\ 1 \end{pmatrix}$

  This can have a bad (and large) influence on estimation problems such as $y_i' \sim H y_i$!

- To preserve as much precision as possible, Hartley suggests making all numbers of approximately the same magnitude before doing any estimation.

- Centre the data round the origin (subtract mean), apply a uniform scaling to the x and y coordinates so the Euclidean point has an average distance of $\sqrt{2}$ to the origin.

- Hartley normalise -> estimate -> Hartley de-normalise

Geometric vs Algebraic Cost Functions

Geometric

$\mathcal{E}_G(\beta) =$ "actual distances"

"Easy" to interpret, "difficult" to minimise in practice (often requires non-linear optimisation).

Example: Line estimation,

$$\mathcal{E}_G(l) = \sum_{j=1}^{N} d_{PL}(x_j, l)^2 = \sum_{j=1}^{N} \left((\text{norm}_p\, x_j)^T (\text{norm}_D\, l)\right)^2$$

Algebraic

$\mathcal{E}_A(\beta) =$ "algebraic expression that is small when things fit well"

Difficult to interpret, "easy" to minimise.

Example: Line estimation,

$$\mathcal{E}_A(l) = \sum_{j=1}^{N} (x_j^T l)^2$$

residuals $r = \underbrace{\begin{pmatrix} x_1^T \\ \vdots \\ x_N^T \end{pmatrix}}_{X} l$

$$X l = 0$$