Jan 31, 2019

Last time:
regularization for LLS

Today:
sparse LLS

Next week: linear dimensionality reduction
+ latent factor models

Direct solvers

$m\ \overset{n}{\boxed{A}}$

$A \to$ factorize

$\to$ solution $\quad O(mn^2)$

(normal equations)

"Iterative solvers"

LSQR, LSMR, CG, MINRES

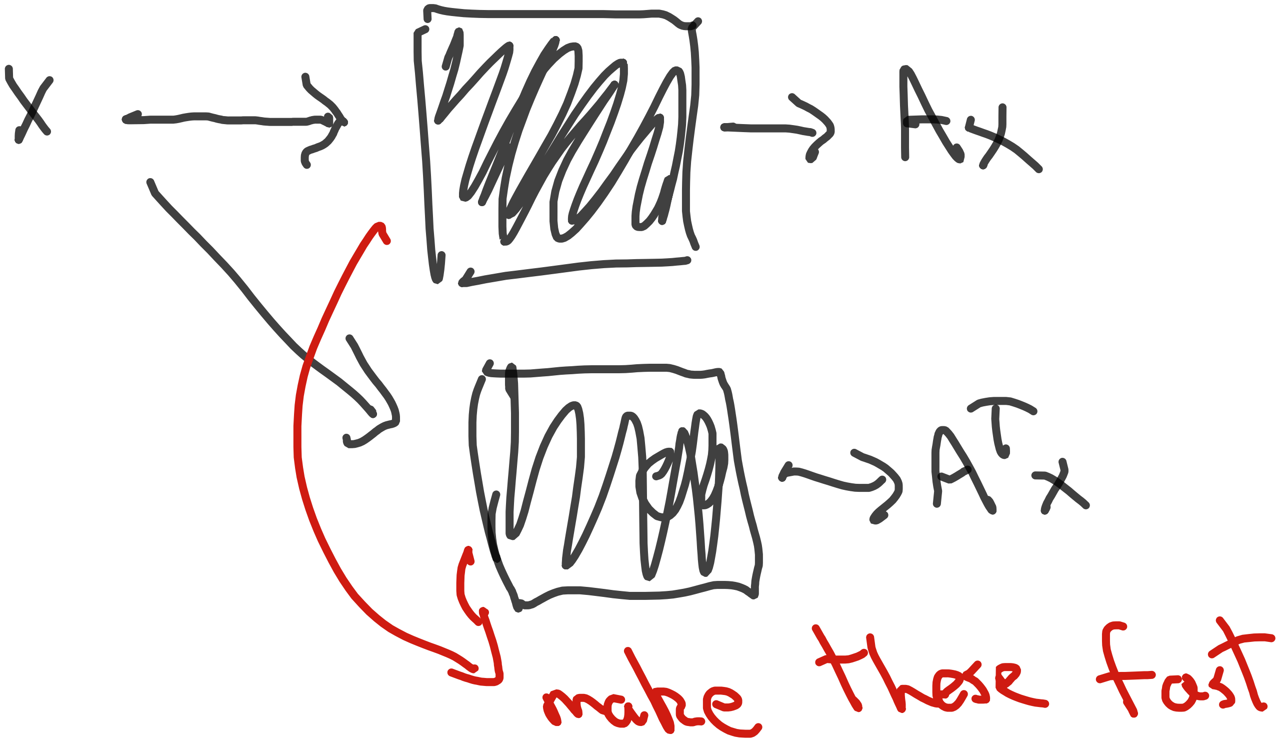
Krylov (1931)

Krylov methods

Often, many features are sparse

on_Rx1	on_Rx2	on_Rx3	...
0	0	1	

Krylov



1. sparsity: $O(\mathrm{nnz}(A))$
2. Toeplitz / Hankel $\begin{pmatrix} a & b & & \\ & a & b & \\ & & \ddots & \ddots \\ & & & b \\ & & & a \end{pmatrix}$
3. Triangular $O(mk^2)$
4. low-rank

5. low-rank + sparse
6. n-body physics (fast multipole method)

Krylov for $Ax = b$

$$\begin{array}{cc} & n \\ n & \boxed{\quad A \quad} \end{array} \boxed{x} = \boxed{b}$$

$b, \; Ab, \; A \cdot Ab = A^2 b, \; A^3 b, \ldots$

Krylov matrix:

$$K_r = [b, Ab, \ldots, A^{r-1} b]$$

1 $A = A^T \quad z^T A z \geq 0$ (SPD)

Method: Conjugate Gradients (CG)

for $r = 1, 2, \ldots$

solve $\displaystyle \min_{x_r} \overbrace{\tfrac{1}{2} x_r^T A x_r - b^T x_r}^{f(x)}$

s.t. $x_r \in K_r y$

$\nabla f(x) = Ax - b$

$= 0 \iff Ax = b$

Key: can compute x_{r+1} from x_r with $A \cdot x_r$

(2) $A = A^T$ (not nec. SPD)

Method: MINRES

$$K_r = [b \; Ab \; \cdots \; A^{r-1}b]$$

V_r = orthogonal basis

$$V_r^T V_r = I$$

for $r = 1, 2, \ldots$

$$\min_x \|x - b\|^2$$

s.t. $x = V_r y$

$$\min_y \|V_r y - b\|^2$$

LLS problem!

Key : can compute x_{r+1} from x_r quickly

Az

Advantages of Krylov:

1. Only need $y = Ax$
2. Sequence of solutions $x_1, x_2, \dots$
3. works if you can ~~approx~~ apply $y = Ax$

Back to LLS

$$\min_x \|Ax - b\|^2$$

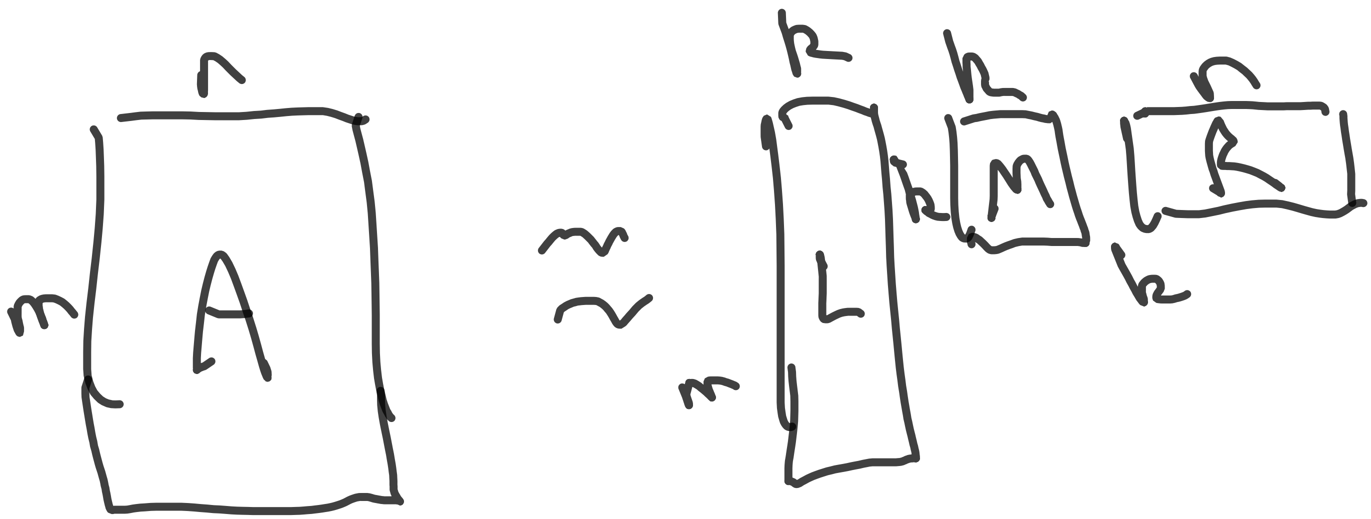
Solution from normal equations:

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

1. CG on normal equations
   Method: LSQR
2. MINRES on normal equations
   Method: LSMR

Next: dimensionality reduction
(linear)

$A \approx LMR$ $\quad k < n$



helps us:

1. Denoise (truncated SVD)
2. Find latent information
(latent factors)

Truncated SVD
(through opt)

$$\min_{Z} \|A - Z\|_F^2$$

s.t. $\operatorname{rank}(Z) = k$

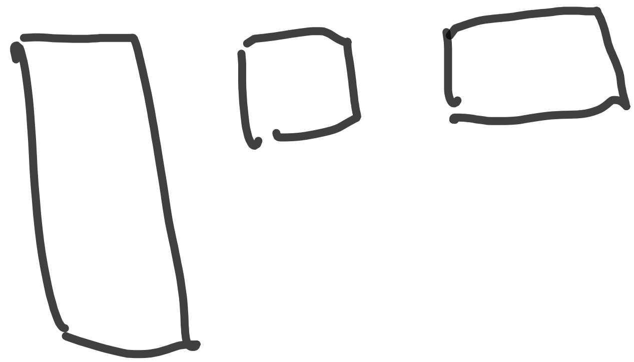
$$Z^* = U_k \Sigma_k V_k^T$$

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

$$f(\Sigma) = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

$$Z^* = U \cdot f(\Sigma) V^T$$

$$Z^* = U_k \Sigma_k V_k^T$$



Eckart-Young theorem

---

k-means clustering

points $a_1, \dots, a_m \in \mathbb{R}^n$

Lloyd's alg:

1. Assign each point to nearest cluster representative

② Recompute representative of each cluster

$$A = \begin{array}{cc} & n \\ m & \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \end{array}$$

$$L_{ij} = \begin{cases} 1 & \text{if point } i \text{ is in the } j^{th} \text{ cluster} \\ 0 & \text{o/w} \end{cases}$$

$$R = \begin{bmatrix} r_1^T \\ \vdots \\ r_k^T \end{bmatrix}$$

r_i = representative of the i^{th} cluster

k-means!

$$\min_{L,R} \|A - LR\|_F^2$$

$e_i^T A = a_i^T$ (i^{th} point)

$\approx r_j^T$ (i in j^{th} cluster)

$= (e_i^T L) R$

$= e_i^T (LR)$

Alternating minimization

1. Fix L

$$R = \min_S \|A - LS\|_F^2$$

$R \Rightarrow$ means

(2) Fix R

$$L = \min_{W} \|A - WR\|_F^2$$

each row of W sums to 1

. $W \in \{0,1\}^{n \times k}$

$L \Rightarrow$ assignment to nearest rep