CS 331, Fall 2026
Lecture 17 (10/26)

Today: - Low-rank approx
- SVD/PCA
- Power method

Low-rank approximation (Part VI, Section 4.1)

Quiz: What are missing entries?

$$\begin{pmatrix} 7 & ? & ? & 14 & 21 \\ ? & 8 & 12 & ? & 6 \\ ? & ? & 6 & 2 & ? \end{pmatrix}$$

Obviously, an unfair question.

Nonetheless, we believe data "in the wild" is structured... maybe guessable?

Ans: $\begin{pmatrix} 7 & 28 & 42 & 14 & 21 \\ 2 & 8 & 12 & 4 & 6 \\ 1 & 4 & 6 & 2 & 3 \end{pmatrix}$

Highly-structured. Every col = multiple of $\begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}$

Rank-1 matrix: $A = \underbrace{u}_{n\times 1}\underbrace{v^T}_{d\times 1} \in \mathbb{R}^{n\times d}$

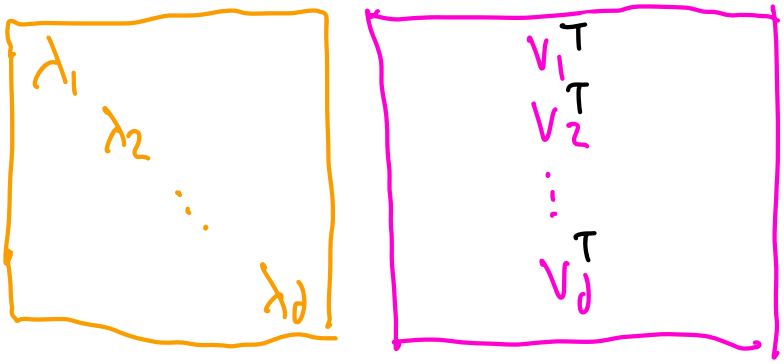
$$A = \begin{pmatrix} A_{:1} & A_{:2} & \dots & A_{:d} \end{pmatrix} = \begin{pmatrix} (v)_1 u & (v)_2 u & \dots & (v)_d u \end{pmatrix}$$

More generally, rank-r

$$A = \left(u_1 v_1^T\right) + \left(u_2 v_2^T\right) + \dots + \left(u_r v_r^T\right)$$

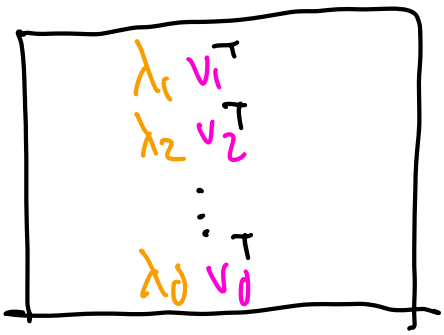
Matrix factorization crash course

1) $\Lambda V^T =$

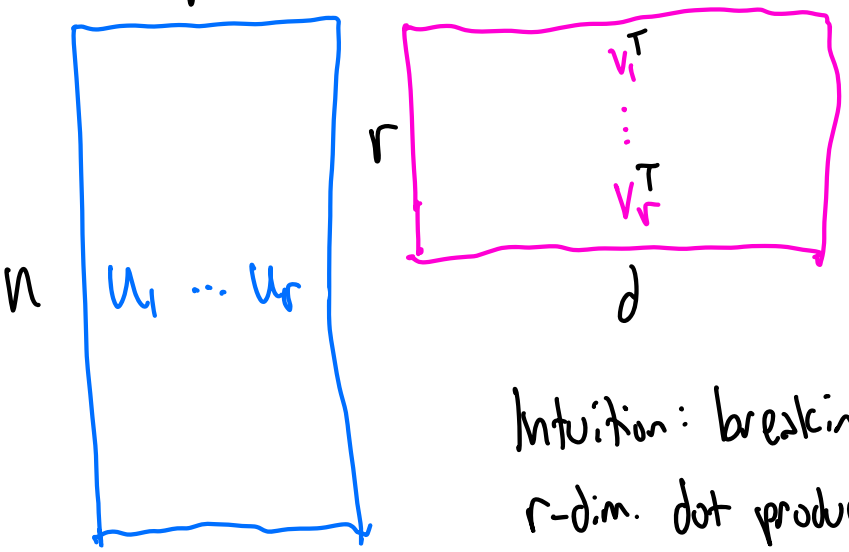


Λ: $\text{diag}(\lambda)$

$=$

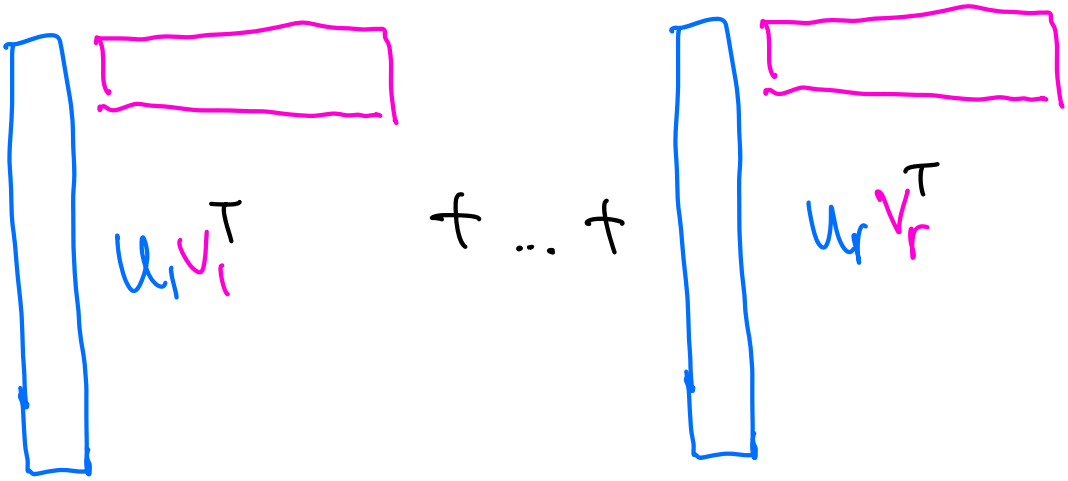


2) $UV^T =$



Intuition: breaking up r-dim. dot products

$=$



Compactly, rank-r decomposition

$$A = UV^T = \sum_{k \in [r]} u_k v_k^T \in \mathbb{R}^{n \times d}$$

$$U = \begin{pmatrix} u_1 & u_2 & \cdots & u_r \end{pmatrix} \in \mathbb{R}^{n \times r}$$

$$V = \begin{pmatrix} v_1 & v_2 & \cdots & v_r \end{pmatrix} \in \mathbb{R}^{d \times r}$$

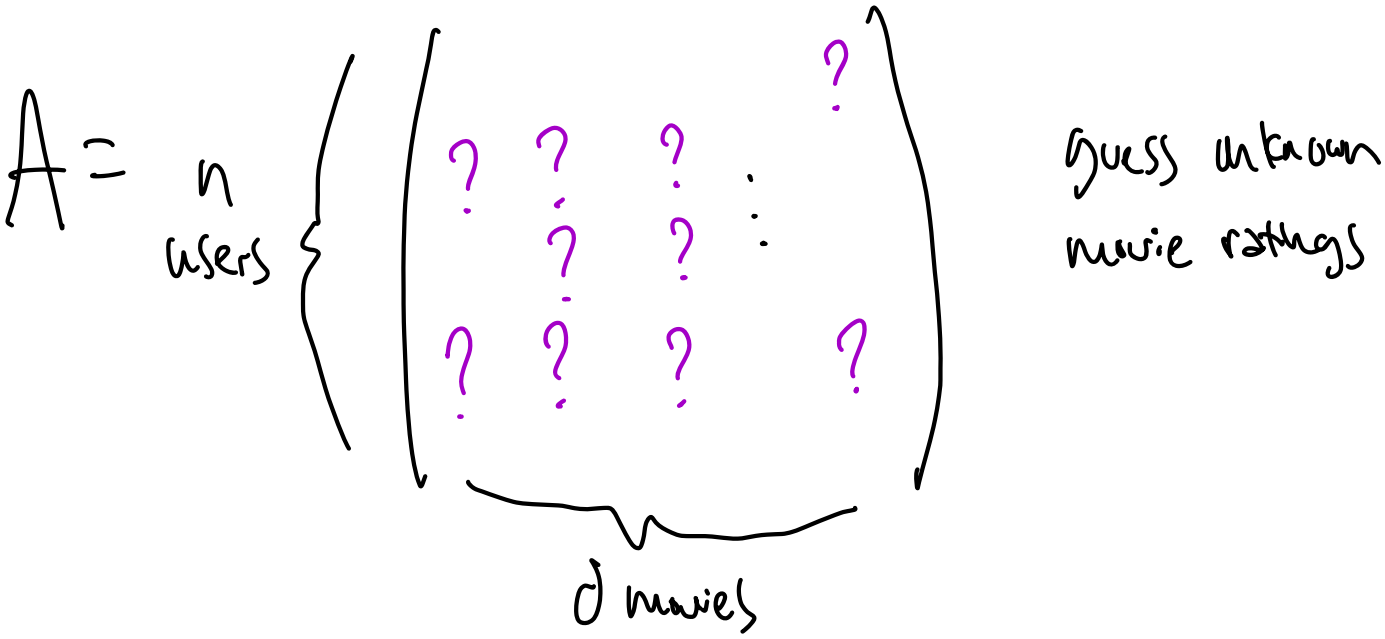
Most interesting when $r \ll \min(n, d)$

We can always achieve $r = \min(n, d)$ by choosing $U = I$ or $V = I$

I = "identity matrix" $= \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix}$

Who cares? One motivation:

Netflix prize (matrix completion)



Simplified model: everyone agrees

$$V^T = \begin{pmatrix} 9.3 & 9.2 & \cdots & 1.9 & 1.5 \end{pmatrix}$$

e.g. Shawshank Redemption, Godfather, Dipsy Doodle movie, Superbabies

$$A_{i:} \approx V \cdot [u]_i \qquad A \approx UV^T$$

$[u]_i$: how much does user i appreciate movies?

More sophisticated model:

$$A \approx \underset{n \times r}{U} \underset{d \times r}{V^T} = \left(u_1 v_1^T \right) + \dots + \left(u_r v_r^T \right)$$

$v_1^T = (\dots \underset{\text{Silence of the lambs}}{9.5} \dots \underset{\text{Get out}}{9.3} \dots \dots)$ "horror enjoyer"

$v_r^T = (\dots \dots \underset{\text{John Wick}}{8.7} \dots \underset{\text{Kill Bill}}{9.9} \dots)$ "action enjoyer"

$$\underbrace{A_{i:}}_{\text{ratings of user } i} = \sum_{k \in [r]} \underbrace{[u_k]_i}_{\text{explained by topic } k} v_k \in \mathbb{R}^d$$

In real life, not <u>exactly</u> low-rank

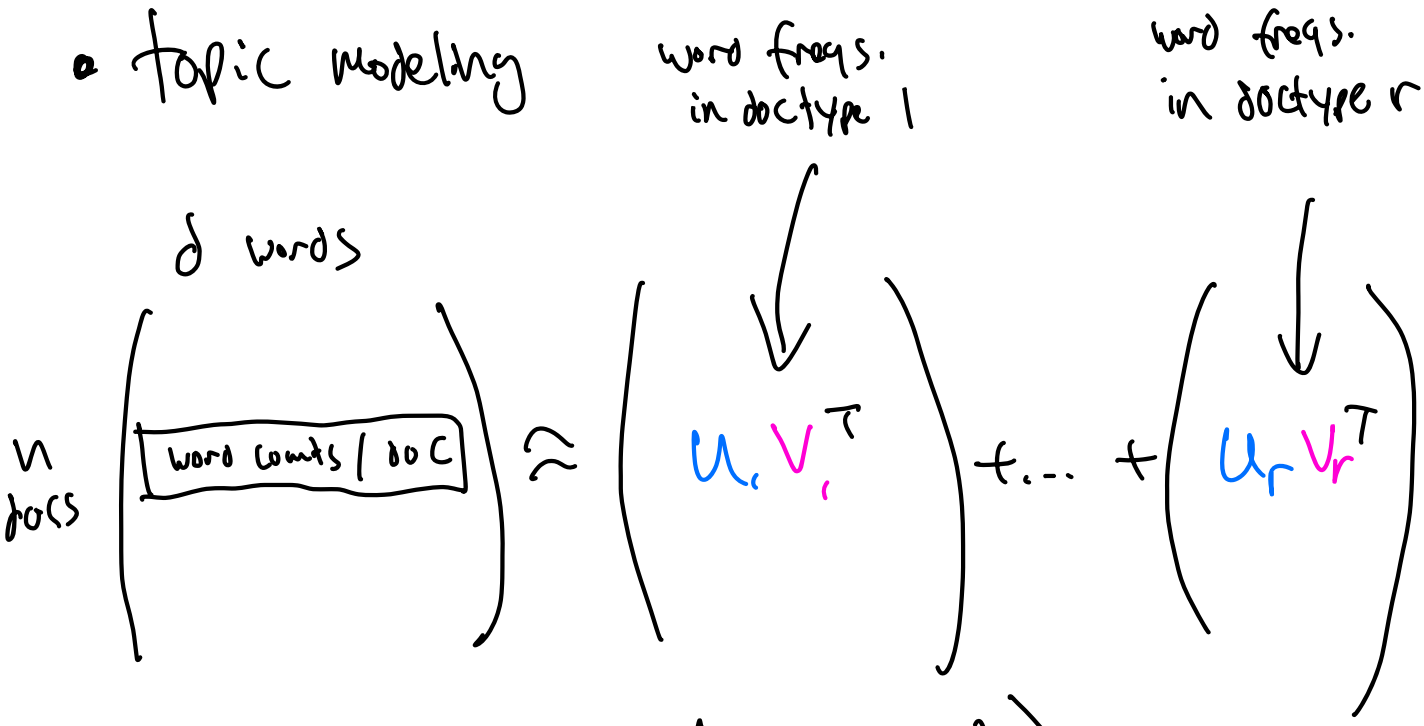
$$A = \underbrace{UV^T}_{r \text{ "explanatory factors"}} + \underbrace{N}_{\text{noise}} = \underset{\text{"simple" explanation}}{n \begin{bmatrix} r \\ \vdots \\ \end{bmatrix} \begin{bmatrix} d \\ r \end{bmatrix}} + \underbrace{\begin{bmatrix} \text{small?} \end{bmatrix}}_{\text{Sampling error, rounding etc.}}$$

Goal in low-rank approximation (LRA):

Make "unexplained" $A - UV^T$ as small as possible

Other motivations:

- topic modeling



- word embeddings (see HW!)
- Save on computation, e.g. if $r = O(1)$

$$Ax \approx UV^T x = U(V^T x)$$

takes $O(nd)$ time — takes $O(d)$ time — takes $O(n)$ time

Singular value decomposition (Part VI, Section 4.2)

We have a complete characterization of

$$\operatorname{argmin}_{U \in \mathbb{R}^{n \times r},\ V \in \mathbb{R}^{d \times r}} \underbrace{\|A - UV^T\|}$$

"unitarily invariant" norm. incl. basically all common size bounds...

Frobenius, operator, nuclear norms

Detour to SVD needed.

Say that $U \in \mathbb{R}^{d \times d}$ is orthonormal if

$$U^T U = I = \begin{pmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 1 \end{pmatrix}$$

Interpretation: let $U = \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix}$

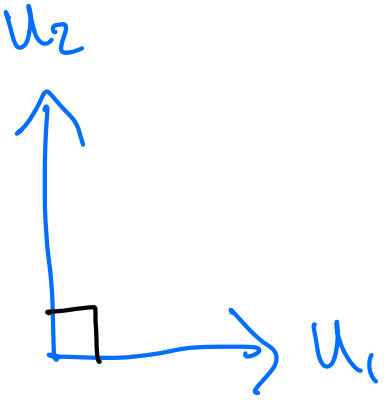
Then, $\forall (i,j) \in [d] \times [d]$

$$\left[U^T U\right]_{ij} = \underbrace{u_i^T u_j}_{\cos(\theta_{ij})} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

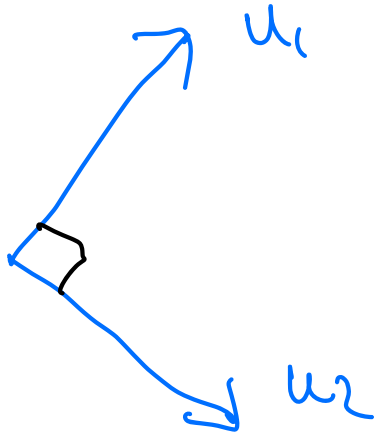
Recall $u^T u = \sum u_i^2 = \underbrace{\|u\|_2^2}_{\text{length of } u}$ (Pythagoras)

Hence columns of U
- unit length
- pairwise perpendicular

e.g.



"Standard basis" $U = I$



Orthonormal basis

Also possible for rectangular "tall" = orthonormal

$$U \in \mathbb{R}^{n \times d}, \quad n \geq d \qquad U^T U = I$$

$$U = \begin{pmatrix} u_1 & u_2 & \cdots & u_d \end{pmatrix}$$

($u_1 \ldots u_d$: unit length, pairwise perp)

Subset of full basis $\{u_i\}_{i \in [n]}$

e.g.

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

$e_1 \quad e_2 \quad e_4$

is orthonormal,

Span(U) is "subspace" of $\mathbb{R}^5$ with dimension 3

$$Ux = x_1 e_1 + x_2 e_2 + x_4 e_4$$

(no mass on e_3, e_5 allowed)

SVD: All matrices $A \in \mathbb{R}^{n \times d}$, $n \geq d$

Can be decomposed as

$$A = U \Sigma V^T = \sum_{i \in (d)} \sigma_i u_i v_i^T$$

$$U = \begin{pmatrix} u_1 & \cdots & u_d \end{pmatrix} \in \mathbb{R}^{n \times d}$$

$$V = \begin{pmatrix} v_1 & \cdots & v_d \end{pmatrix} \in \mathbb{R}^{d \times d}$$

orthonormal matrices

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_d \end{pmatrix} \in \mathbb{R}^{d \times d}_{\geq 0}$$

nonneg. diagonal matrix

$$\#\text{nonzero} = \text{rank}(A) = \dim(\text{span}(A))$$

Interpretation: SVD sends $x \to Ax$

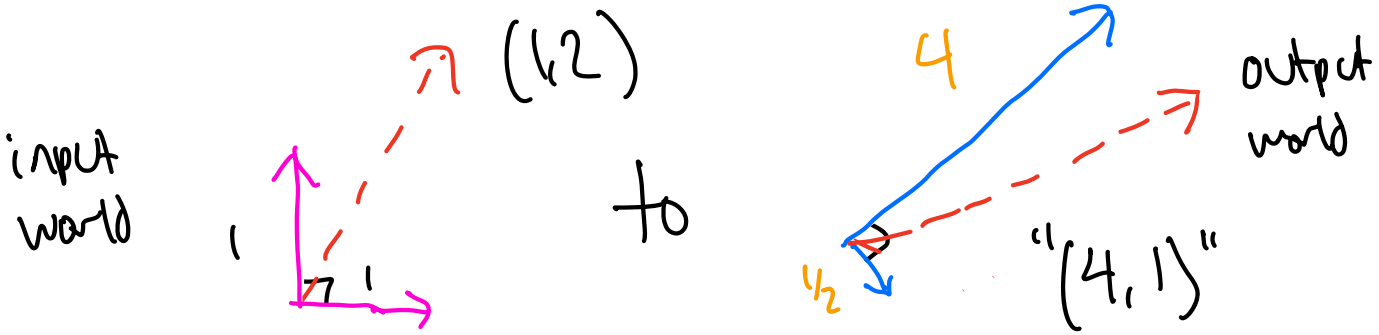
v_i ($\in \mathbb{R}^d$) = "input world" to $\sigma_i u_i$ ($\in \mathbb{R}^n$) = "output world"

$$x = \sum_{i \in [d]} c_i v_i = V c \quad (\text{coeffs. } c \in \mathbb{R}^d)$$

$$\to Ax = U \Sigma \underbrace{V^T V}_{= I} c = \sum_{i \in [d]} \sigma_i c_i u_i$$

e.g.

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



Eckart-Young-Mirsky:

For any $A \in \mathbb{R}^{n \times d}$, $n \geq d$, unitarily-invariant $\|\cdot\|$

$$\underset{\text{rank-}r\ M}{\operatorname{argmin}} \|A - M\|$$

achieved by $M = \sum_{k \in [r]} \sigma_k u_k v_k^T$

where $A = U \Sigma V^T$ (SVD)

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$$

PCA, optimal low-rank approx.

= do SVD, keep top r components,

all else $\Rightarrow 0$.

Special case: Symmetric matrix $M \in \mathbb{R}^{d \times d}$

$M = U \Lambda U^T$, but $\Lambda \in \mathbb{R}^d$

"eigendecomposition"

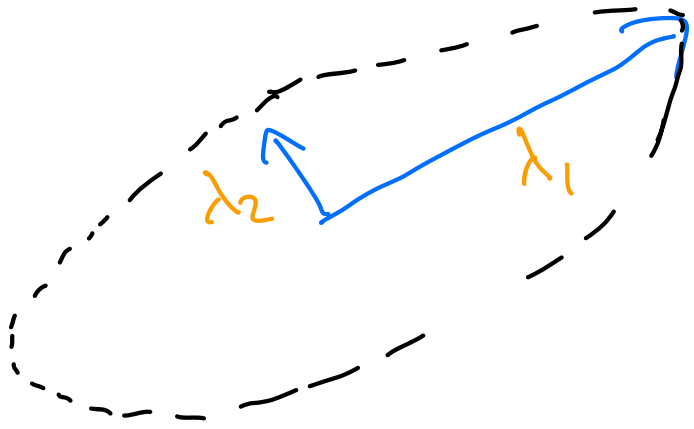
diagonal, could be neg.

Even more special: if $\Lambda \in \mathbb{R}^{d \times d}_{\geq 0}$ already

M is "positive semidefinite" (PSD)

(Definition: symmetric matrix, all eigvals $\lambda_i \geq 0$)

Input world = output world!



$M u_i = \lambda_i u_i$

geometric interpretation of eigvecs/eigvals

"every PSD matrix is an ellipse"

If $A = U\Sigma V^T$, we have

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

$$AA^T = U\Sigma V^T V\Sigma U^T = U\Sigma^2 U^T$$

Both PSD, eigvectors give SVD of A.

Enough to give algos to recover:

top eigvec of PSD matrix M (PCA)

Power method (Part VI, Section 4.3)

Let PSD $M \in \mathbb{R}^{d\times d}$

$$M = U\Lambda U^T = \sum_{i\in[d]} \lambda_i u_i u_i^T$$

Λ: $\mathrm{diag}(\lambda)$

$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$

Goal: recover u_1 to high accuracy.

Idea: suppose $U = I$, $\lambda_1 \geq 2000 d^2 \lambda_2$

Then $Mg \approx$ multiple of e_1

random normal vector

In fact, $Ug \sim g$ (still works for $U \neq I$)

But $\lambda_1 \geq 2000 d^2 \lambda_2$ really strong...

What if only $\lambda_1 \geq 1.1 \lambda_2$?

No problem: for $p = O(\log d)$

$$M^p g = U\Lambda U^T U\Lambda U^T U\Lambda U^T \cdots$$

$$= U\Lambda^p U^T g$$

gap now $1.1^p \geq 2000 d^2$

normalize $M^p g$

Runtime: $O(d^2 \log(d))$ gives $\cos\theta(u, u_1) \geq 0.99$ w.p. ≥ 0.99