Low-Rank PSD Approximation in Input-Sparsity Time

Kenneth L. Clarkson and David P. Woodruff
IBM Research Almaden

Low Rank Approximation

- A is an n x n matrix
- A is typically well-approximated by low rank matrix
  - E.g., high rank because of noise
- Goal: find a low rank matrix approximating A
  - Easy to store in factored form
  - Data more interpretable

What is a Good Low Rank Approximation?

Singular Value Decomposition (SVD)

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

- U has orthonormal columns
- Σ is diagonal with non-increasing positive entries down the diagonal
- V^T has orthonormal rows

- Rank-k approximation: $A_k = U_k \Sigma_k V_k^T$

$A_k = \text{argmin}_{\text{rank k matrices B}} |A-B|_F$

$(|C|_F = (\Sigma_{i,j} C_{i,j}^2)^{1/2})$

Computing A_k exactly is expensive

$$\left(\mathbf{A} \right) = \left(\mathbf{U}_k \right) \left(\Sigma_k \right) \left(\mathbf{V}_k^T \right) + \left(\mathbf{E} \right)$$

Approximate Low Rank Approximation

- [CW13] output a rank k matrix A', so that with probability > 2/3,

$$|A-A'|_F \leq (1+\varepsilon)\ |A-A_k|_F,$$

in nnz(A) + n · poly(k/ε) time

Structure-Preserving Low Rank Approximation

- Let A be an arbitrary n x n matrix
- Instead of just finding a rank-k matrix A' for which $|A-A'|_F$ is small, suppose we also require A' to be positive semidefinite (PSD)
  - A' is symmetric and all eigenvalues are non-negative
- Covariance matrices, kernel matrices, Laplacians are PSD
  - Approximate them for efficiency
- Roundoff errors may make a PSD matrix non-PSD
  - We do not assume A is PSD but want A' to be PSD

Structure-Preserving Low Rank Approximation

- Goal: output a PSD rank-k matrix A' for which $|A-A'|_F$ is small
- Can assume A is symmetric
  - $A = A^{sym} + A^{asym}$, where $A^{sym}_{i,j} = \frac{A_{i,j}+A_{j,i}}{2}$ and $A^{asym}_{i,j} = \frac{A_{i,j}-A_{j,i}}{2}$
  - $|A - A'|^2{}_F = |A^{sym} - A'|_F^2 + |A^{asym}|_F^2$
  - Compute A^{sym} in nnz(A) time
- What is the best PSD rank-k approximation $A_{k,+}$ to A?
- Lemma: $A_{k,+}$ is obtained by zeroing out all but the top k *positive* eigenvalues in eigendecomposition of A
  - If $A = U\,D\,U^T$, then $A_{k,+} = UD_{k,+}U^T$
  - If A has fewer than k positive eigenvalues, zero out all except these eigenvalues

Our Result for PSD Low Rank Approximation

- (PSD low rank approximation result): In nnz(A) + n poly(k/ ε) time, can find a PSD rank-k A' so that

$$|A-A'|_F \leq (1+\varepsilon)\, |A-A_{k,+}|_F$$

- Previous work
  - "Nystrom method" based on uniform sampling requires incoherence assumptions on A
  - [GM] Weaker bound $|A-A'|_F \leq |A-A_{k,+}|_F + \epsilon |A - A_{k,+}|_*$ where $|.|_*$ is nuclear norm
  - [WLZ] Running time at least $n^2 k/\epsilon$ and A' has a larger rank k/ϵ

Our Result for PSD Column Subset Selection

- (PSD Column Subset Selection): In nnz(A)log n + n poly(k(log n)/ ε) time, find a subset C of O(k/ϵ) columns of A so that A'= CUC^T is rank-k, PSD, and

$$|A-A'|_F \leq (1+\varepsilon)\ |A-A_{k,+}|_F$$

- Column subsets preserve sparsity, interpretability

- Most previous results require incoherence assumptions or achieve weaker guarantees in terms of the nuclear norm

- [WLZ] Get $O\left(\frac{k}{\epsilon}\right)$ columns but running time at least n^2k/ϵ and rank(A') = k/ϵ

Talk Outline

- Low Rank PSD Approximation
  - Pitfalls of Usual Approach
  - $SF(\epsilon, k)$ Property
  - Error Term Lemmas
  - Solving a Small Problem Quickly

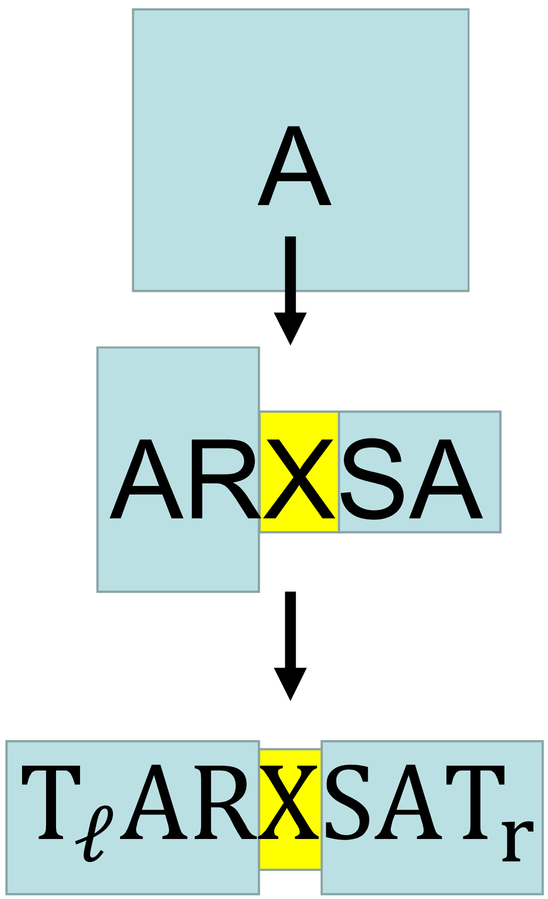
- Low Rank PSD Column Subset Selection

Pitfalls of Usual Approach

- [CW09,HMT,BWZ] Choose random matrices R, S
  - R has k/ϵ columns and S has k/ϵ rows
  - $A' = \underset{\text{rank}-k\ ARXSA}{\operatorname{argmin}} |ARXSA - A|_F^2$ and $|A-A'|_F \leq (1+\varepsilon)\, |A-A_k|_F$
  - Compute AR and SA in nnz(A) time

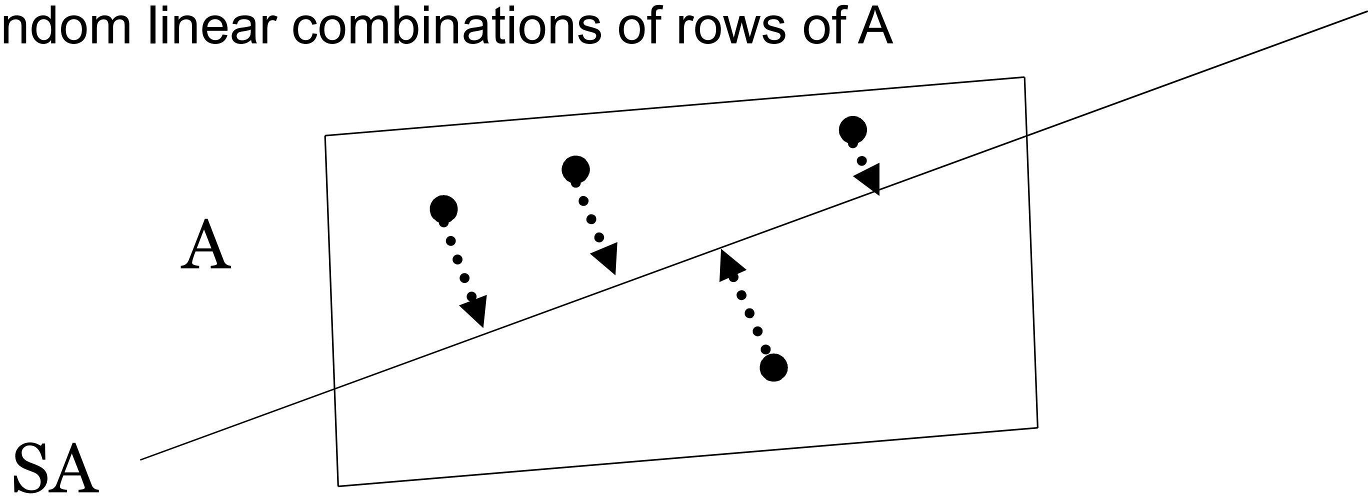
- To solve for A', solve
  - $\min_{\text{rank}-k\ X} |T_\ell ARXSAT_r - T_\ell AT_r|_F^2$, where T_ℓ, T_r are random
  - [FT] poly(k/ϵ) time

- Need $R = S^T$ for A' to be symmetric, but analysis requires $|x^T SAR|_2 = \Theta\left(|x^T SA|_2\right)$ for all x !



$SF(\epsilon, k)$ Property

- Given n x n input matrix A
- Compute S*A using a sketching matrix S with k/ε << n rows. S·A takes random linear combinations of rows of A



- Let P be the n x n projection matrix onto the row span of SA
- [S,CW,MM,NN,…] $|A(I-P)|_F^2 \le (1+\epsilon)|A-A_k|_F^2$
- P is $SF(\epsilon, k)$ if $|A(I-P)|_2^2 \le \frac{\epsilon}{k}|A-A_k|_F^2$ (recall $|B|_2 = \sup_x |Bx|_2/|x|_2$)

$SF(\epsilon, k)$ Property

- $SF(\epsilon, k)$ property implies usual theory

$|A(I - P)|_F^2$

$= |A_k(I - P)|_F^2 + |(A - A_k)(I - P)|_F^2$ by Pythagorean theorem

$\leq k \cdot |A_k(I - P)|_2^2 + |(A - A_k)(I - P)|_F^2$ since A_k has rank k

$\leq k \cdot |A(I - P)|_2^2 + |(A - A_k)(I - P)|_F^2$ since $|A(I - P)x|_2 \geq |A_k(I - P)x|_2$ for all x

$\leq k \left(\frac{\epsilon}{k}\right) |A - A_k|_F^2 + |(A - A_k)(I - P)|_F^2$ by $SF(\epsilon, k)$ property

$\leq \epsilon \, |A - A_k|_F^2 + |A - A_k|_F^2$ since projections don't increase norms

A Basic Lemma

- Lemma: For symmetric A, B with (A-B)B = 0, and projection matrix P,

$$|A - PBP|_F^2 = |A - B|_F^2 + |B - PBP|_F^2 + 2\mathrm{Tr}(A - B)(I - P)BP$$

$SF(\epsilon, k)$ Projections Give Good Solutions

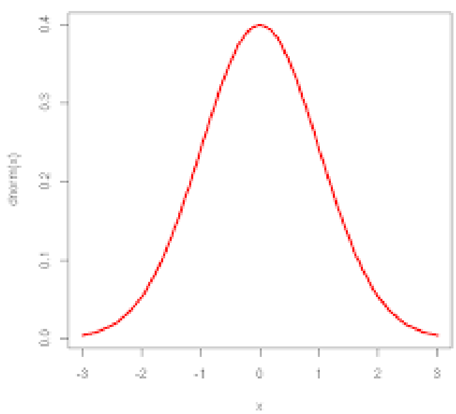
- Lemma: If P is $SF(\epsilon, k)$ for A, then $|A-PA_{k,+}P|_F \leq (1+O(\varepsilon)) |A-A_{k,+}|_F$
- Proof: Since $(A - A_{k,+})A_{k,+} = 0$, can apply the basic lemma:

$$|A - PA_{k,+}P|_F^2 = |A - A_{k,+}|_F^2 + \underbrace{|A_{k,+} - PA_{k,+}P|_F^2 + 2\mathrm{Tr}(A - A_{k,+})(I - P)A_{k,+}P}$$

Use $SF(\epsilon, k)$ property of P to show these terms are $O(\epsilon)|A - A_{k,+}|_F^2$

Finding an SF(ϵ,k) Projection

- [CEMMP] implies if S is poly(k/ϵ) x n i.i.d. Gaussian, Fast Hadamard Transform, or a Sparse Embedding Matrix, then P = $(SA)^{-}SA$ is $SF(\epsilon, k)$



$$, P \cdot H \cdot D, \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

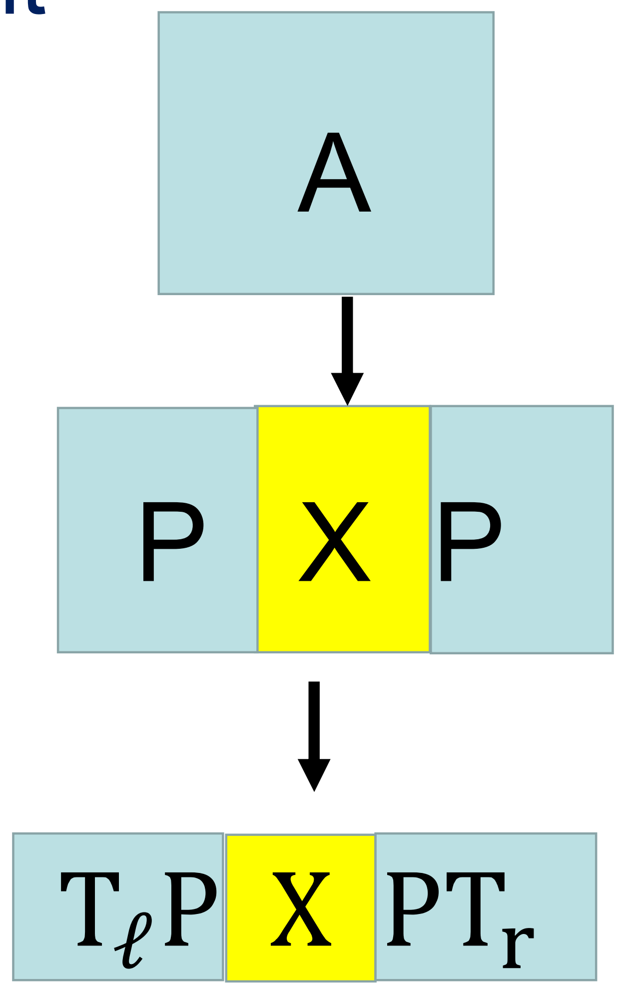
Completing the Argument

- If P is $\mathrm{SF}(\epsilon, \mathrm{k})$ for A, then $\mathrm{A}' = \mathrm{argmin}_{\mathrm{rank-k\ PSD\ PXP}} |\mathrm{A} - \mathrm{PXP}|_\mathrm{F}$ satisfies
$$|\mathrm{A}' - \mathrm{A}|_\mathrm{F} \leq (1+\epsilon)\left|\mathrm{A} - \mathrm{A}_{\mathrm{k},+}\right|_\mathrm{F}$$

- We can also find $\mathrm{P} = (\mathrm{SA})^{-}\mathrm{SA}$ in factored form in nnz(A) time

- How to solve for A'?
  - Multiply by small random matrices $\mathrm{T}_\ell, \mathrm{T}_\mathrm{r}$, solve $\mathrm{A}' = \mathrm{argmin}_{\mathrm{rank-k\ PSD\ PXP}} |\mathrm{T}_\ell \mathrm{A} \mathrm{T}_\mathrm{r} - \mathrm{T}_\ell \mathrm{PXPT}_\mathrm{r}|_\mathrm{F}$

- Tiny problem but is it in polynomial time?
  - We show how to solve it up to a $(1+\epsilon)$ −factor quickly



Talk Outline

- Low Rank PSD Approximation
  - Pitfalls of Usual Approach
  - $\mathrm{SF}(\epsilon, \mathrm{k})$ Property
  - Error Term Lemmas
  - Solving a Small Problem Quickly
- Low Rank PSD Column Subset Selection

Low Rank Column Subset Selection

- [CEMMP], together with a composition lemma, gives a P = $(SA)^{-}SA$ where S is a sampling and rescaling matrix, such that P is $SF(\epsilon, k)$ for A and SA can be found in nnz(A)log n + n poly$\left(\frac{k}{\epsilon}\right)$ time

- S samples $O\left(\frac{k}{\epsilon}\right)$ rows of A

- Our earlier lemma implies there is a rank-k PSD solution A' = PXP

- Our earlier procedure finds X

Other Results

- Symmetric Matrices
  - Analogous results for nnz(A) time for finding a rank-k symmetric approximation to A, and for column/row subset selection

- Low Rank Approximation with Tail Guarantee
  - Let t = $2k/\epsilon$
  - A PSD rank-k matrix A' can be found in nnz(A) + n $\mathrm{poly}\left(\frac{k}{\epsilon}\right)$ time with

$$|A - A'|_F^2 \leq \left|A - A_{k,+}\right|_F^2 + |A_{t+k} - A_t|_F^2$$

Conclusion and Open Questions

- First rank-k PSD approximation of an arbitrary matrix A in nnz(A) time
- Optimal $O\left(\frac{k}{\epsilon}\right)$ columns/rows for rank-k PSD subset selection, in $\mathrm{nnz}(A)\log n$ time
- Similar results for symmetric approximations
- Should be able to improve the time for column/row subset selection to nnz(A) using known techniques
- High-level question – quickly find low rank approximations with additional structure, such as being PSD

What is $A_{k,+}$?

- $A = U D U^T$ is eigendecomposition
- $|A - A_{k,+}|_F^2 = |D - U^T A_{k,+} U|_F^2$, where $U^T A_{k,+} U$ is PSD and rank k
- Let $D_{k,+}$ be the best rank-k PSD approximation to D
- $|D - U^T A_{k,+} U|_F^2 \geq |D - D_{k,+}|_F^2 = |UDU^T - UD_{k,+}U^T|_F^2 = |A - UD_{k,+}U^T|_F^2$
- $A_{k,+} = UD_{k,+}U^T$

But what is $D_{k,+}$, the best rank k PSD approximation to diagonal matrix D?

$$\left|\left[\begin{matrix} D_+ & 0 \\ 0 & D_- \end{matrix}\right] - \left[\begin{matrix} Y_t & Y_r \\ Y_l & Y_b \end{matrix}\right]\right|_F^2 \geq |D_+ - Y_t|_F^2 + |D_- - Y_b|_F^2$$

D $\quad$ $D_{k,+}$

$D_{k,+}$ is diagonal matrix with top k non-negative eigenvalues of A

Solving the Small Problem

- How to solve for $A' = \text{argmin}_{\text{rank}-k \text{ PSD PXP}} |T_\ell A T_r - T_\ell PXPT_r|_F$?
- Can find A' minimizing this up to a 1+ϵ factor
- Write SA = $R\Sigma W^T$ in its SVD, so $P = (SA)^- SA = WW^T$
- $A' = \text{argmin}_{\text{rank}-k \text{ PSD PXP}} \left|T_\ell A T_r - (T_\ell W)W^T XW\, (W^T T_r)\,\right|_F$
- Write $(T_\ell W) = U_\ell \Sigma_\ell V_\ell^T$ and $\left(W^T T_r\right) = U_r \Sigma_r V_r^T$ in their SVD, use that all their singular values are $1 \pm \epsilon$, and after some algebra,

  $X = \left[\frac{M+M^T}{2}\right]_{k,+}$ where M = $V_\ell \Sigma_\ell^{-1} U_\ell^T T_\ell A T_r V_r \Sigma_r^{-1} U_r^T$