Efficient low rank approximation with affine embeddings

Kadin Zhang

15-851, Spring 2025

Let A be a large $n \times d$ matrix, like a customer-product matrix. Typically these are well approximated by lower rank matrices, which take much less parameters to store (we can store a $n \times k$ matrix and a $k \times d$ matrix), and have the added benefit of denoising.

We can solve for the best rank k approximation to A in closed form with SVD:

Theorem (Truncated SVD is optimal low rank approximation): Let $A_k = U\Sigma_k V^\top$, where Σ_k zeros all singular values outside of the top k. Then

$$A_k = \underset{\mathrm{rank}(B)=k}{\arg\min} \|B - A\|_F$$

The problem is that SVD will cost $O(nd^2)$, which is intractible when n and d are large. We will show that we can get a rank k approximation A' with

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

with constant probability of failure in $O(nnz(A) + (n+d)\,\mathrm{poly}(d/\varepsilon))$ time. We will make use of affine embeddings:

Definition (Affine embedding): Let A be a $n \times d$ and B be $n \times m$. We wish to solve the regression problem

$$\min_X \|AX - B\|_F^2.$$

An *affine embedding* is a short matrix S such that for all X,

$$\|S(AX - B)\|_F \leq (1+\varepsilon)\|AX - B\|_F$$

holds with constant probability. CountSketch matrices of dimension $O(d^2/\varepsilon^2) \times n$ are one such family that satisfy the affine embedding property.

1 Motivation

We would like to output a rank matrix A' such that

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

As motivation, consider the regression problem $\min_X \|A_k X - A\|_F$ with optimum $X^* = I$, and an affine embedding S. The property of S tells us for all X:

$$\|SA_k X - SA\|_F \leq (1+\varepsilon)\|A_k X - A\|_F.$$

The optimal X for the LHS is $X' = (SA_k)^- SA$, which is in the rowspan of SA. Since S is an affine embedding, the sketched optimum is a good approximation:

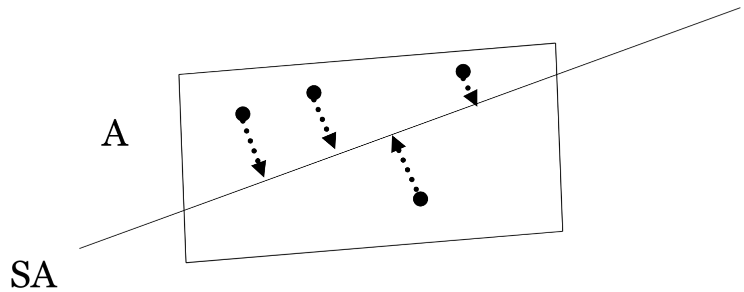
$$\begin{aligned}(1-\varepsilon)\|A_k X' - A\|_F &\leq \|SA_k X' - SA\|_F \\ &\leq \|SA_k X^* - SA\|_F \\ &\leq (1+\varepsilon)\|A_k X^* - A\|_F \\ &= (1+\varepsilon)\|A_k - A\|_F.\end{aligned}$$

In conclusion, $A_k(SA_k)^{-}SA$ is a good rank k approximation to A_k in the rowspan of SA.

This outlines an initial strategy:

(a) Choose a $k/\varepsilon \times n$ sketching matrix S. We assume CountSketch henceforth, since this allows us to compute SA in $O(nnz(A))$.

(b) Find a rank k approximation with SVD within the rowspan of SA rather than A. We can think of SA through rows: each row is a random linear combination of rows of A, which live in $\mathbb{R}^d$. But the rows of SA are now in a k/ε-dimensional subspace, so we have "projected" the rows of A onto SA:



2 First attempt: finding best approximation in rowspan SA

We solve want the minimum of $\|XSA - A\|_F$ over rank k X. Using the normal equations and Pythagorean theorem (think of this as over rows, where we can represent the distance from X_iSA as the distance from X_iSA to the optimum projection of A_i onto SA, and the distance from the optimum to A_i):

$$\|XSA - A\|_F^2 = \left\|XSA - A(SA)^{-}SA\right\|_F^2 + \left\|A(SA)^{-}SA - A\right\|_F^2.$$

So the minimizer of X is

$$\arg\min_X \left\|XSA - A(SA)^{-}SA\right\|_F^2.$$

Write $SA = U\Sigma V^\top$ in sparse form (so that $U\Sigma$ is square) then

$$\begin{aligned}\arg\min_X \left\|XSA - A(SA)^{-}SA\right\|_F^2 &= \arg\min_X \left\|XU\Sigma - A(SA)^{-}U\Sigma\right\|_F \\ &= \arg\min_Y \left\|Y - A(SA)^{-}U\Sigma\right\|_F,\end{aligned}$$

where the first equality comes from the fact that $V^\top$ has orthonormal rows, and the second equality comes from the fact that $U\Sigma$ is invertible.

We can solve for the minimum now by taking SVD of $A(SA)^{-}U\Sigma$, but the problem is left multiplying by A, which takes $O(nnz(A)\,\mathrm{poly}(k/\varepsilon))$ (each row A_i sums $nnz(A_i)$ rows of size k/ε). We will address this by approximating the projection of A onto SA using affine embeddings.

3 Faster by approximating projection

Let's revisit

$$\min_X \|XSA - A\|_F^2.$$

Let R be a transposed CountSketch matrix with k/ε columns so that we can compute AR, SAR in $O(nnz(A))$ time. We have the sketched problem guarantee:

$$\|XSAR - AR\|_F \leq (1+\varepsilon)\|XSA - A\|_F.$$

Mirroring the earlier approach with Pythagorean theorem and a change of variables:

$$\begin{aligned}
&\min_X \|XSAR - AR\|_F \\
&= \min_X \left\|XSAR - AR(SAR)^- SAR\right\|_F^2 + \min_X \left\|AR(SAR)^- SAR - AR\right\|_F^2 \\
&= \min_X \left\|XSAR - AR(SAR)^- SAR\right\|_F^2 \\
&= \min_Y \left\|Y - AR(SAR)^- SAR\right\|_F^2.
\end{aligned}$$

Now, since AR is $n \times k/\varepsilon$ and SAR is $k/\varepsilon \times k/\varepsilon$, we can compute $AR(SAR)^- SAR$ in $O(n(k/\varepsilon)^2)$, this time in bounds. The failure probability is constant by a union bound over constant failure probabilities of S and R. In summary, the algorithm

(a) Compute SA .

(b) Compute $\min_Y \|Y - AR(SAR)^- SAR\|_F^2$ with truncated SVD.

(c) Output $Y(SAR)^- SA$ (in factored form to stay in complexity bounds).

achieves with constant failure probability a rank k approximation of A in time $O(nnz(A) + (n + d)\,\mathrm{poly}(d/\varepsilon))$.