Fast SVD

- Suppose we want a rank-k approximation.
- Best we can get $\left\|A - U\Sigma V^T\right\| = \sigma_{k+1}$
- Approximate algorithm: with high probability

$$\left\|A - U\Sigma V^T\right\| \leq 10\sqrt{l \cdot n}\,\sigma_{k+1}, \quad U \in \mathbb{R}^{m\times k},\ V \in \mathbb{R}^{n\times k},\ \Sigma \in \mathbb{R}^{k\times k}$$

- U, V have orthogonal columns.
- l is a parameter. For example, set $l = k + 20$.
- Complexity $C_{\text{SVD}} = l \cdot C_A + k \cdot C_{A^T} + O\left(l^2 m + k^2 n + k^2 m\right)$.
- C_A, C_{A^T} cost of applying A and A^T to a vector.
- Complexity for dense matrices and $l = O(k)$: $O(mnk)$ (compare to $O(m^2 n)$).
- Much faster in many cases.

Fast SVD – algorithm

Input: $A \in \mathbb{R}^{m \times n}$ and $k \leq m$.
Output: Matrices U, V, Σ s.t. $\left\|A - U\Sigma V^T\right\| \leq 10\sqrt{ln}\,\sigma_{k+1}(A)$, with probability at least $1 - 10^{-17}$.

1. Set $l = k + 20$, $G \in \mathbb{R}^{n \times l}$ with $G_{ij} \sim N(0,1)$ i.i.d.
2. Compute $B = AG$ ($B \in \mathbb{R}^{m \times l}$).
3. Compute the SVD of B, $B = X\Lambda Y^T$.
4. Set $Q = X(:, 1:k)$ ($Q \in \mathbb{R}^{m \times k}$).
5. Compute $T = Q^T A$ ($T \in \mathbb{R}^{k \times n}$).
6. Compute the T, $T = W\Sigma V^T$.
7. Compte $U = QW$.

Fast SVD – idea

1. Find $Q = \{q_1, q_2, ..., q_k\} \in \mathbb{R}^{m\times k}$ that span most of A's columns. That is, QQ^TA is very close to A.
2. Compute the SVD of Q^TA
$$Q^TA = W\Sigma V^T, \quad W \in \mathbb{R}^{k\times k}, \quad \Sigma \in \mathbb{R}^{k\times k}, \quad V \in \mathbb{R}^{n\times k}.$$
3. Compute the $m \times k$ matrix $U = QW$ and note that
$$U\Sigma V^T = QW\Sigma V^T = QQ^TA \approx A,$$
that is, $U\Sigma V^T$ is a rank-k approximation of A.

- How to chose Q whose columns span "most" of the range of A.

Fast SVD – randomization

- Take $G \in \mathbb{R}^{n \times l}$ be a matrix of i.i.d. $N(0, 1)$.

- There exists a matrix $F \in \mathbb{R}^{l \times n}$ such that $\|AGF - A\| \lesssim \sigma_{k+1}(A)$, that is, $AGF \approx A$.

- The range of A is approximately the same as the range of AG. However, $AG \in \mathbb{R}^{m \times l}$ is much smaller than $A \in \mathbb{R}^{m \times n}$.

- Choose Q to be a matrix whose columns form an orthonormal basis for the range of AG. E.g, compute $AG = X \Lambda Y^T$ and take $Q = X$.

Fast SVD – advantages

- Much faster for big data – $O(mnk)$ vs $O(m^2n)$.
- Only requires applying A to l vectors and A^T to k vectors.
- Explicit access to the entries of A is not required!
- A and A^T are not required explicitly – can be provided as functions for applying them to vectors.