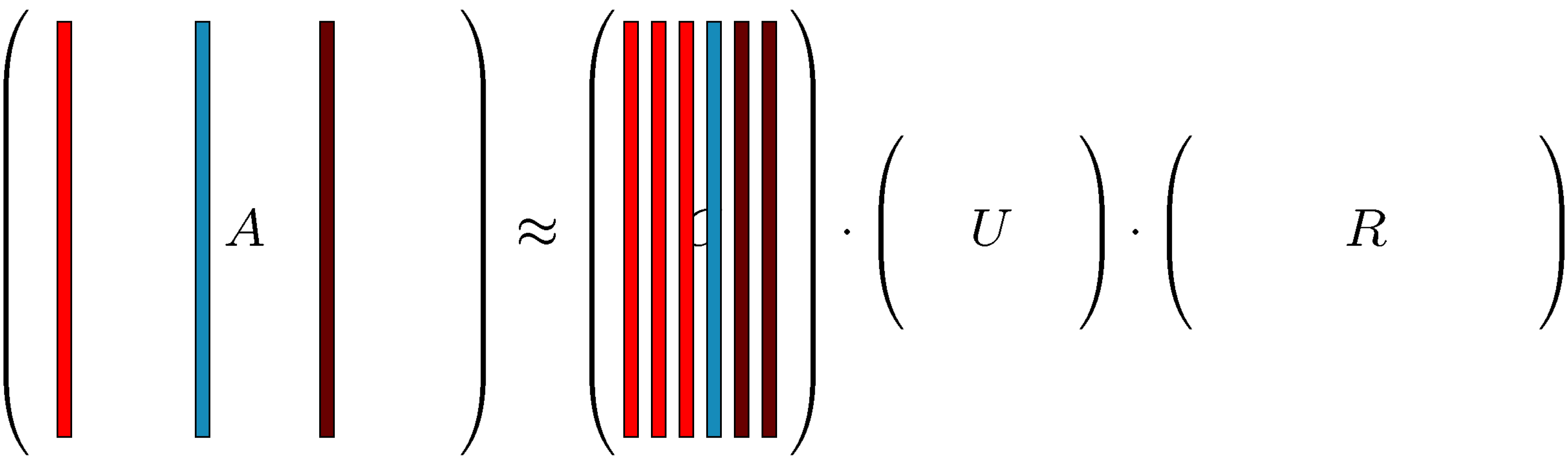
Dimensionality Reduction: CUR Decomposition

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

CUR Decomposition

Frobenius norm:
$\|X\|_F = \sqrt{\Sigma_{ij} X_{ij}^2}$

- **Goal: Express A as a product of matrices C,U,R**
  Make $\|A - C \cdot U \cdot R\|_F$ small
- **"Constraints" on C and R:**

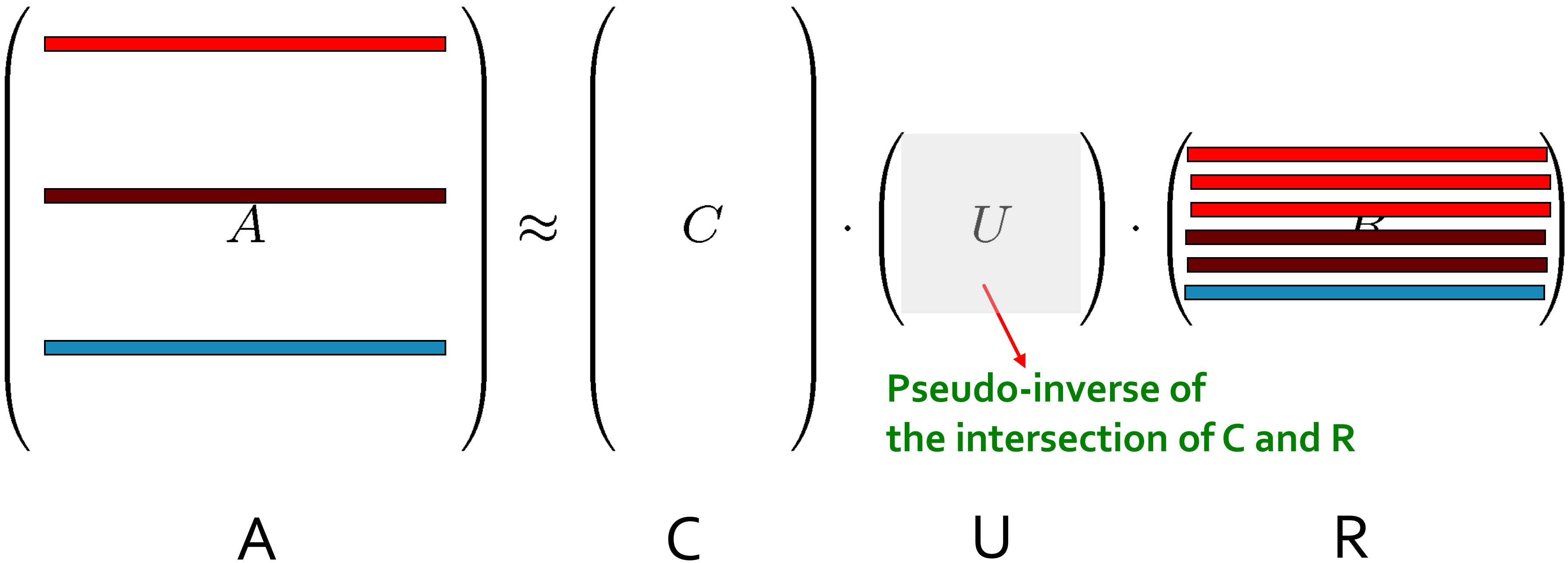


A C U R

CUR Decomposition

Frobenius norm:
$\|X\|_F = \sqrt{\Sigma_{ij} X_{ij}^2}$

- **Goal: Express A as a product of matrices C,U,R**
  Make $\|A - C \cdot U \cdot R\|_F$ small
- **"Constraints" on C and R:**



CUR: Provably good approx. to SVD

- **Let:**
  $\mathbf{A}_k$ be the "best" rank $\boldsymbol{k}$ approximation to $\mathbf{A}$ (that is, $\mathbf{A}_k$ is SVD of $\mathbf{A}$)

Theorem [Mahoney & Drineas]

CUR in O(**m·n**) time achieves

- $\|A\text{-}CUR\|_F \leq \|A\text{-}A_k\|_F + \varepsilon\|A\|_F$

with probability at least **1-δ**, by picking

- $O(k \log(1/\delta)/\varepsilon^2)$ columns, and
- $O(k^2 \log^3(1/\delta)/\varepsilon^6)$ rows

In practice:
Pick 4*k* cols/rows

CUR: How it Works

- **Sampling columns (similarly for rows):**
- **ColumnSelect algorithm:**

Input: matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, sample size c

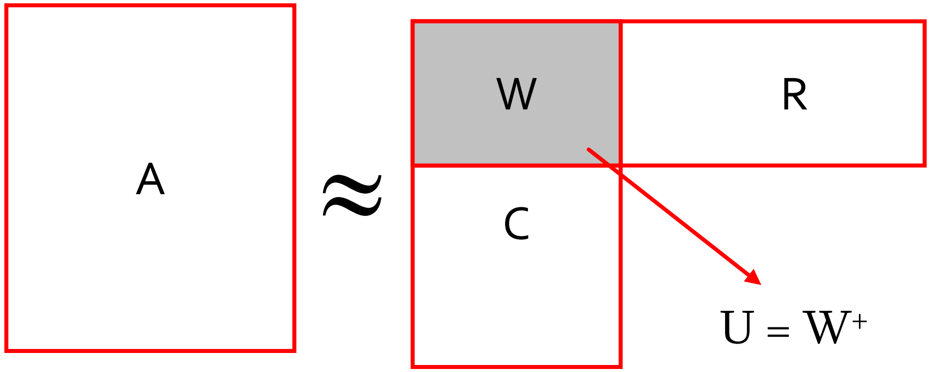
Output: $\mathbf{C}_d \in \mathbb{R}^{m \times c}$

1. for $x = 1 : n$ [column distribution]
2. $P(x) = \sum_i \mathbf{A}(i,x)^2 / \sum_{i,j} \mathbf{A}(i,j)^2$
3. for $i = 1 : c$ [sample columns]
4. Pick $j \in 1 : n$ based on distribution $P(x)$
5. Compute $\mathbf{C}_d(:,i) = \mathbf{A}(:,j) / \sqrt{cP(j)}$

Note this is a randomized algorithm, same column can be sampled more than once

Computing U

- Let **W** be the "intersection" of sampled columns **C** and rows **R**
  - Let SVD of **W** = **X Z Y**^{T}
- **Then: U = W$^+$ = Y Z$^+$ X^T**
  - Z^+: **reciprocals of non-zero singular values:** $Z^+_{ii} = 1/Z_{ii}$
  - W$^+$ is the "**pseudoinverse**"



Why pseudoinverse works?
W = X Z Y then W^{-1} = X^{-1} Z^{-1} Y^{-1}
Due to orthonomality
X^{-1}=X^T and Y^{-1}=Y^T
Since Z is diagonal Z^{-1} = 1/Z$_{ii}$
Thus, if **W** is non-singular, pseudoinverse is the true inverse

CUR: Provably good approx. to SVD

- For example:
  - Select $c = O\left(\frac{k \log k}{\varepsilon^2}\right)$ columns of A using ColumnSelect algorithm
  - Select $r = O\left(\frac{k \log k}{\varepsilon^2}\right)$ rows of A using ColumnSelect algorithm
  - Set $U = W^+$
- **Then:** $\|A - CUR\|_F \leq (2 + \epsilon)\,\|A - A_k\|_F$ with probability 98%