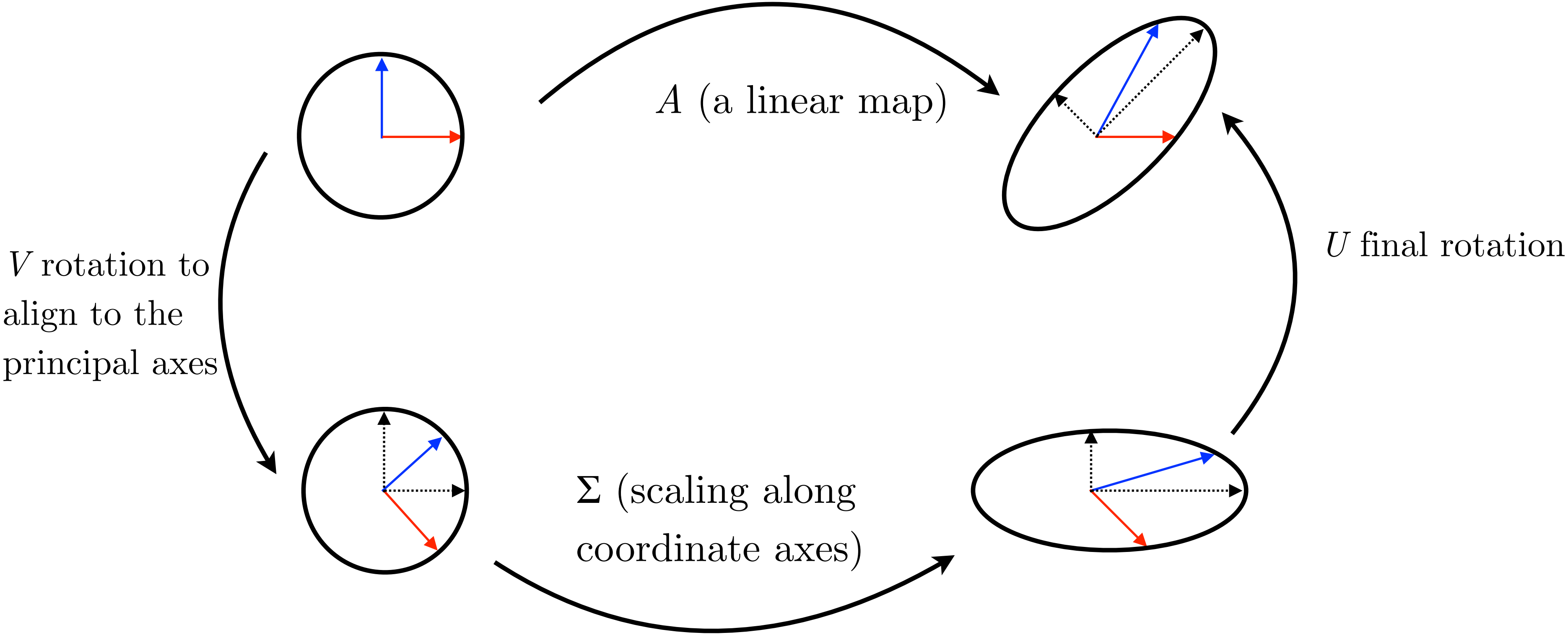
EPFL

Singular Value Decomposition (SVD)
EE-312

Prof. Pierre Vandergheynst

The Geometry of a Linear Map



The SVD!

Let $A \in \mathbb{R}_r^{m\times n}$

There exists orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that

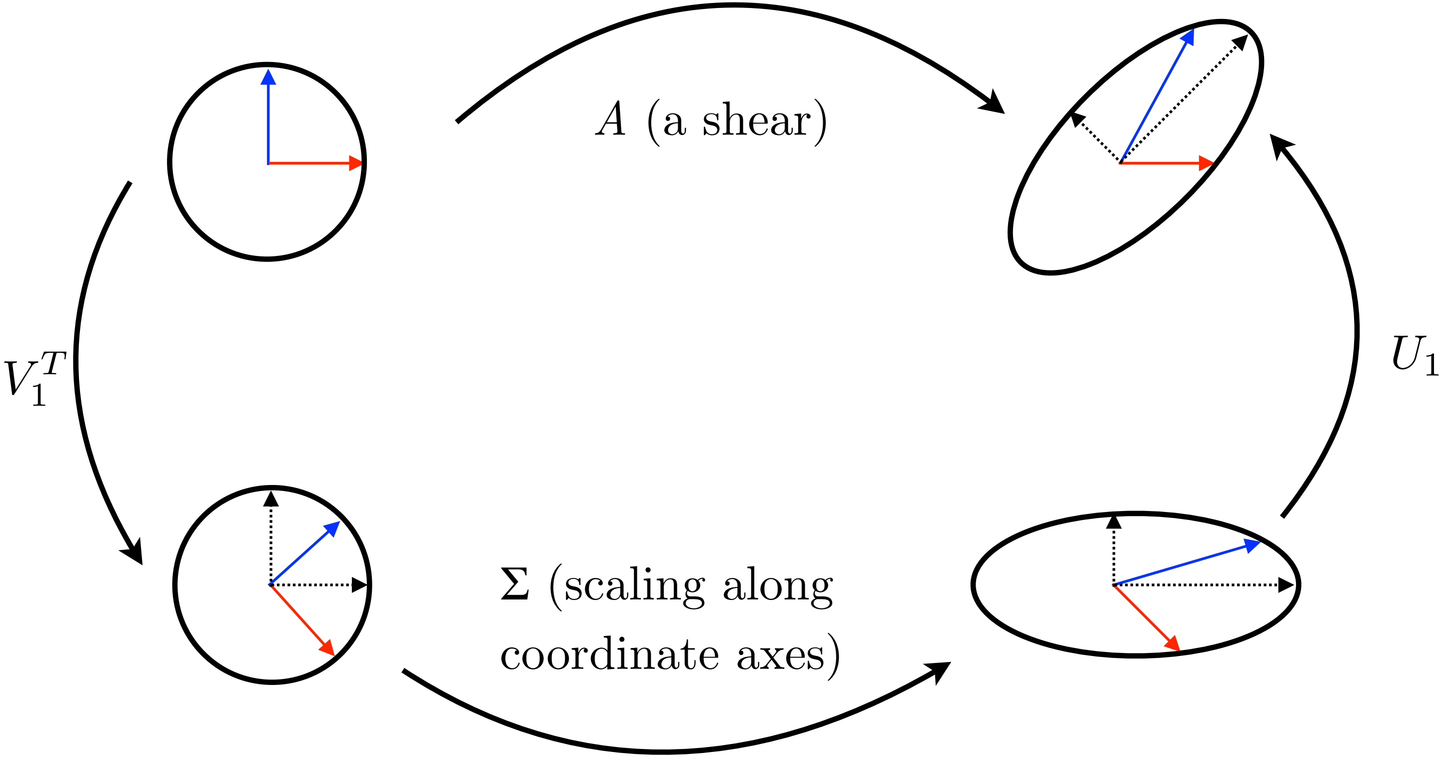
$$A = U\Sigma V^T$$

where $\Sigma = \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix}$, $\quad S = \text{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$ and $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \sigma_r \geqslant 0$

$$A = [U_1\, U_2] \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$
$$= U_1 S V_1^T$$

$U_1 \in \mathbb{R}^{m\times r}$ $\quad V_1 \in \mathbb{R}^{n\times r}$ rotations/reflections in r-dimensional subspaces of $\mathbb{R}^m$ and $\mathbb{R}^n$

Geometric Interpretation



Some SVD terminology

$\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ are the singular values of A.

$$\sigma_i = \sqrt{\lambda_i(AA^T)} = \sqrt{\lambda_i(A^TA)}$$

The columns of U are the left singular vectors and orthonormal eigenvectors of AA^T

The columns of V are the right singular vectors and orthonormal eigenvectors of A^TA

Singular values are uniquely defined by A

But not singular vectors ! (although their span is)

Singular Vectors Basis

The left and right singular vectors provide useful basis of $\mathbb{R}^n$ and $\mathbb{R}^m$

$A \in \mathbb{R}^{m \times n}$ with SVD $A = U\Sigma V^T$

$$x \in \mathbb{R}^n \text{ and } x = V\alpha$$

$$y \in \mathbb{R}^m \text{ and } y = U\beta$$

Suppose $y = Ax$ then $\beta = \Sigma\alpha$

Expressed on these bases, the matrix is diagonal !

SVD and the fundamental subspaces

Let $A \in \mathbb{R}_r^{m\times n}$ and $A = U\Sigma V^T$

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \quad \text{and } A \text{ has rank } r \text{ (\# non-zero singular values)}$$

sum of rank 1 matrices

Using this decomposition we easily see:

$$Av_i = \sigma_i u_i$$

$$A^T u_i = \sigma_i v_i$$

Using the Penrose conditions we see:

$$A^+ = V\Sigma^+ U^T, \text{ where } \Sigma^+ = \begin{bmatrix} S^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

$$A^+ = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$$

SVD and the fundamental subspaces

Let $A \in \mathbb{R}_r^{m \times n}$ and $A = U\Sigma V^T$ $\quad U_1 = [u_1, \ldots, u_r] \quad U_2 = [u_{r+1}, \ldots, u_m]$

$V_1 = [v_1, \ldots, v_r] \quad V_2 = [v_{r+1}, \ldots, v_n]$

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \qquad A^T = \sum_{i=1}^{r} \sigma_i v_i u_i^T$$

$$(a) \quad \mathcal{R}(U_1) = \mathcal{R}(A) = \mathcal{N}(A^T)^\perp$$

$$(b) \quad \mathcal{R}(U_2) = \mathcal{R}(A)^\perp = \mathcal{N}(A^T)$$

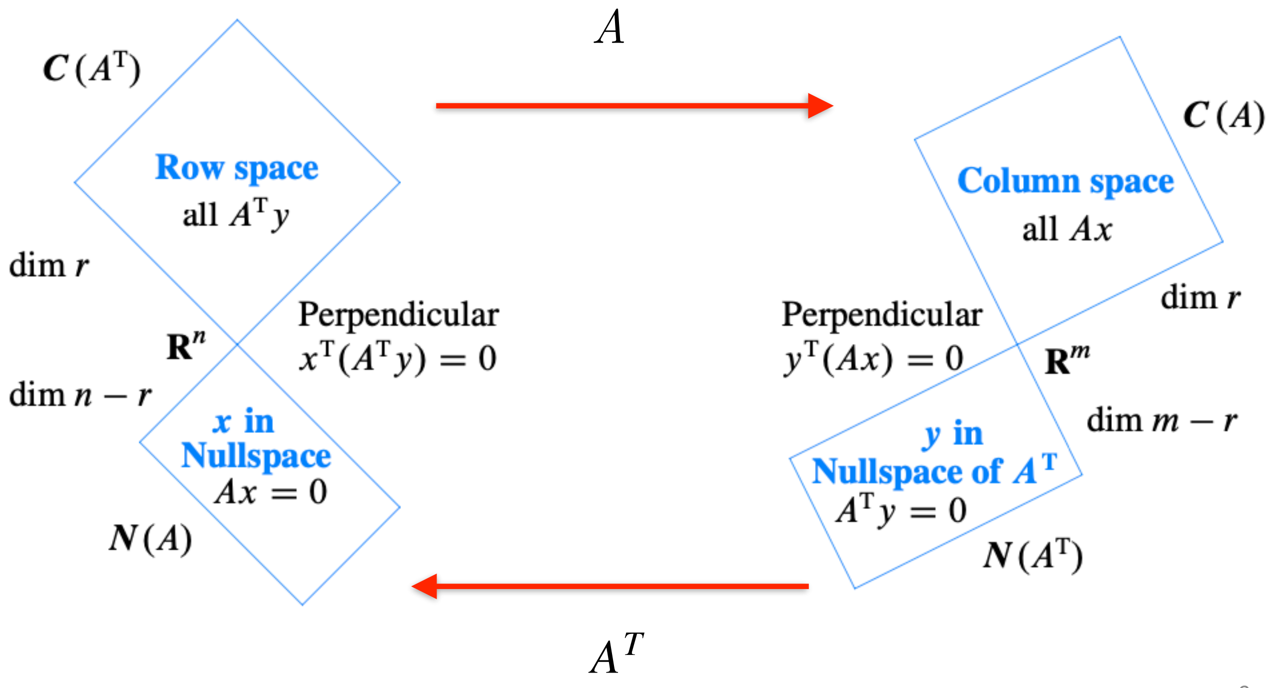
$$(c) \quad \mathcal{R}(V_1) = \mathcal{R}(A^T) = \mathcal{N}(A)^\perp$$

$$(d) \quad \mathcal{R}(V_2) = \mathcal{R}(A^T)^\perp = \mathcal{N}(A)$$

provides orthobases
for the fundamental subspaces

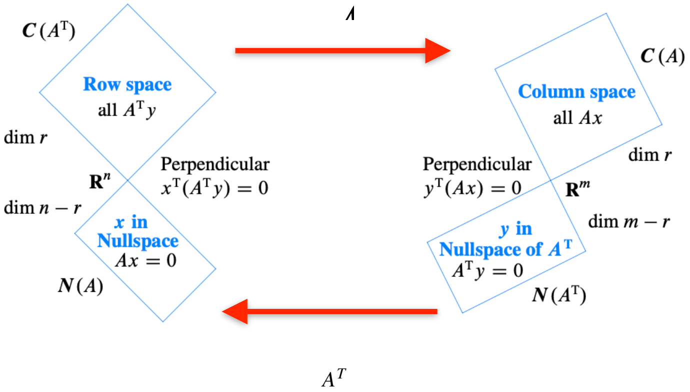
SVD and the fundamental subspaces

Wrapping it all together: the full picture



SVD and the fundamental subspaces

Wrapping it all together: the full picture



SVD gives us ortho-bases for the fundamental subspaces

the right singular vectors V are a basis of $\mathbb{R}^n$

the left singular vectors U are a basis of $\mathbb{R}^m$

But more importantly: Restriction T of A from row space to column space is bijective

$V_1 = [v_1, \ldots v_r]$ is a an orthobasis of the row space

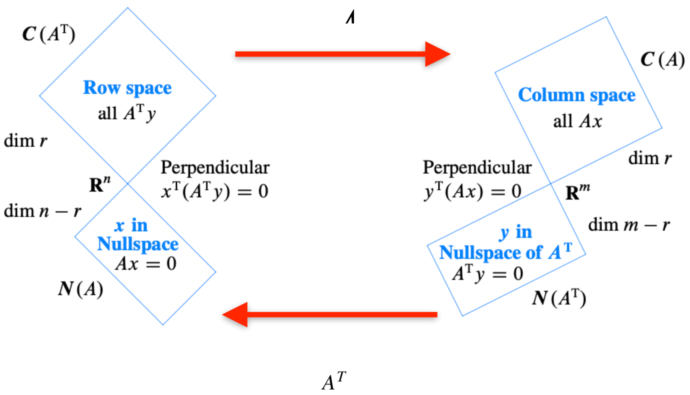
$V_2 = [v_{r+1}, \ldots v_n]$ is a an orthobasis of the null space of A

$U_1 = [u_1, \ldots u_r]$ is a an orthobasis of the col. space

$U_2 = [u_{r+1}, \ldots u_m]$ is a an orthobasis of the null space of A^T

SVD and the fundamental subspaces

Wrapping it all together: the full picture



Restriction T of A from row space to column space is bijective

$V_1 = [v_1, \ldots v_r]$ is a an orthobasis of the row space

$V_2 = [v_{r+1}, \ldots v_n]$ is a an orthobasis of the null space of A

$U_1 = [u_1, \ldots u_r]$ is a an orthobasis of the col. space

$U_2 = [u_{r+1}, \ldots u_m]$ is a an orthobasis of the null space of A^T

The action of T is specified by its action in the corresponding bases

$Tv_i = Av_i = \sigma_i u_i \ \forall i = 1,\ldots,r$ and the matrix is therefore diagonal in these bases!

T is a bijection therefore invertible and clearly $T^{-1}u_i = \sigma_i^{-1} v_i$

and this defines the pseudo-inverse: $A^+ u_i = \sigma_i^{-1} v_i \ \forall i = 1,\ldots r$

Low-Rank Approximation

Consider a full rank matrix $A \in \mathbb{R}^{m \times n}$ $A = U\Sigma V^T$

Now construct the following low-rank matrix

$$\mathcal{T}_k(A) = U_k \Sigma_k V_k^T \qquad U_k = [u_1, \ldots, u_k] \qquad \Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k) \qquad V_k = [v_1, \ldots, v_k]$$

$\mathcal{T}_k(A)$ has rank at most k

$$\mathcal{T}_k(A) = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

How well does it **approximate** A?

We will measure the error of approximation with a well-chosen matrix norm

Low-Rank Approximation

Consider a *unitarily invariant* matrix norm: $\|UAV^T\| = \|A\|$

$$\|\mathcal{T}_k(A) - A\| = \|\mathrm{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_n)\|$$

Frobenius norm:

$$\|\mathcal{T}_k(A) - A\|_F = \sqrt{\sigma_{k+1}^2 + \ldots + \sigma_n^2}$$

Operator norm:

$$\|\mathcal{T}_k(A) - A\|_2 = \sigma_{k+1}$$

Good approximation when singular values decay fast

We have just solved $\arg\min_B \|A - B\|_{F \text{ or } 2}$ subject to $\mathrm{rank}(B) \leqslant r$

See Eckart-Young-Mirsky theorem

Low-Rank Approximation

The same idea allows finding an $m \times k$ orthogonal matrix s.t. $\text{range}(Q) \approx \text{range}(A)$

minimising (for unitarily invariant norm) $\|(\mathbb{I} - QQ^T)A\|$

orthogonal projector on $\text{range}(Q)^{\perp}$

Leads to $Q = U_k$

SVD and Least Squares

$$\min_x \|Ax - b\|_2, \; A \in \mathbb{R}^{m \times n} \; b \in \mathbb{R}^m$$

Assume A has SVD $A = U\Sigma V^T = U_1 S V_1^T$

We can write the error term as: $\|Ax - b\|_2^2 = \|Sz_1 - c_1\|_2^2 + \|c_2\|_2^2$

With: $z = V^T x$ and $c = U^T b$

Since S is invertible, the minimum error is reached for $z_1 = S^{-1} c_1$

$$\min_x \|Ax - b\|_2^2 = \|c_2\|_2^2$$

SVD and Least Squares

By switching back to original coordinates:

$$x = Vz$$
$$= \underbrace{V_1 S^{-1} U_1^T}_{A^+} b + \underbrace{V_2 z_2}_{\text{an arbitrary vector in } \mathcal{N}(A)}$$

And so we recover the general solution:

$$x = A^+ b + (I - A^+ A)y, \text{ with } y \in \mathbb{R}^m \text{ arbitrary}$$

SVD and Least Squares

Moreover, minimum error (least squares residual): $\|c_2\|_2 = \|U_2^T b\|_2$

$$\begin{aligned} \text{minimum residual } &\Leftrightarrow b \text{ is orthogonal to all vectors in } U_2 \\ &\Leftrightarrow b \text{ is orthogonal to all vectors in } \mathcal{R}(A)^\perp \\ &\Leftrightarrow b \in \mathcal{R}(A) \end{aligned}$$

SVD allows us to quickly recover all our results about Least Squares !

An aside: QR for Least Squares

Let $A \in \mathbb{R}^{m \times n}$ be a full column rank matrix (n linearly indep. columns)

$a_j = A[:, j] \in \mathbb{R}^m$ j-th column

Gram-Schmidt to the columns:

$$u_1 = a_1 \to e_1 = \frac{u_1}{\|u_1\|_2}$$

$$u_2 = a_2 - (e_1^T a_2) e_1 \to e_2 = \frac{u_2}{\|u_2\|_2}$$

$$\vdots$$

$$u_j = a_j - \sum_{k=1}^{j-1} (e_k^T a_j) e_k \to e_j = \frac{u_j}{\|u_j\|_2}$$

Express in the GS basis

$$a_j = \sum_{k=1}^{j} (e_k^T a_j) e_k$$

An aside: QR for Least Squares

$$a_j = \sum_{k=1}^{j} e_k^T a_j e_k \qquad \text{In matrix form:} \quad a_j = A[;,j] = \sum_{k=1}^{j} \underset{e_k}{Q[:,k]} \underset{e_k^T a_j}{R[k,j]}$$

The QR decomposition $A = QR$

$Q \in \mathbb{R}^{m \times n}$ with orthonormal columns

$R \in \mathbb{R}^{n \times n}$ upper triangular

Complete Q into an orthonormal basis of $\mathbb{R}^m$

$U \in \mathbb{R}^{m \times m} \, U = [QQ_2]$

An aside: QR for Least Squares

We can now easily express the LS residual:

$$\begin{aligned}\|Ax - y\|_2^2 &= \|U^T Ax - U^T y\|_2^2 \\ &= \|Rx - c_1\|_2^2 + \|c_2\|_2^2\end{aligned}$$

where we used $U^T y = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ $\quad U^T A = [QQ_2]^T A = \begin{bmatrix} R \\ 0 \end{bmatrix}$ Q_2 is orthogonal to $\mathcal{R}(A)$

And we immediately get minimum residual at:

$$x = R^{-1} c_1 = R^{-1} Q^T y$$

check that $A^+ = R^{-1} Q^T$

The minimum value of the residual: $\|c_2\|_2^2 = \|Q_2^T y\|_2^2$