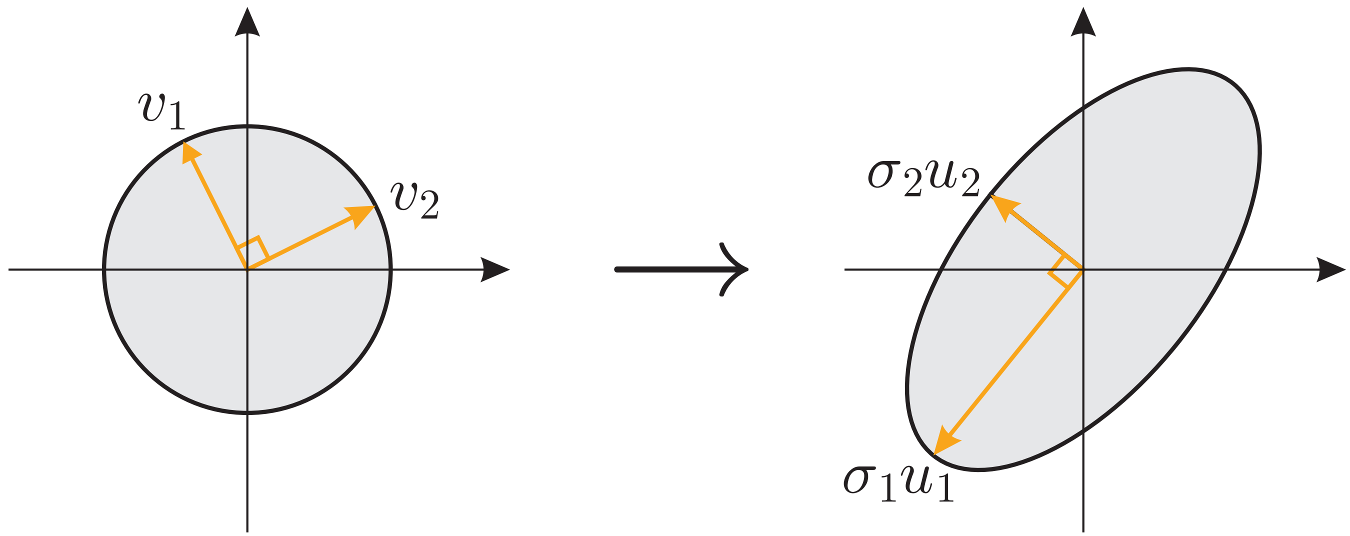
7. Singular Values and Matrix Norms

- Geometry of linear maps
- The singular value decomposition (SVD)
- Interpretation of the SVD
- Matrix properties via the SVD, rank, range and null space
- Example: testing achievability of desired outputs for control
- The SVD and the control ellipsoid
- Example: forces applied to a hovercraft
- Summary: svd, control and estimation
- The matrix norm, and inequalities
- Example: matrix norm and estimation
- Minimal-rank approximation
- Example: minimal-rank approximation
- Example: image compression

The key points of this section

- there is an ellipsoid for control problems, which tells which directions have strong and weak actuator authority
- ellipsoids in both control and estimation problems arise because of the geometry of linear transformations
- the *singular value decomposition* gives a way to both *compute* and *understand* this geometry
- the svd also gives us a way to pick out the *essential features* of any linear map, and simplify it by remove the unessentials

Geometry of Linear Maps



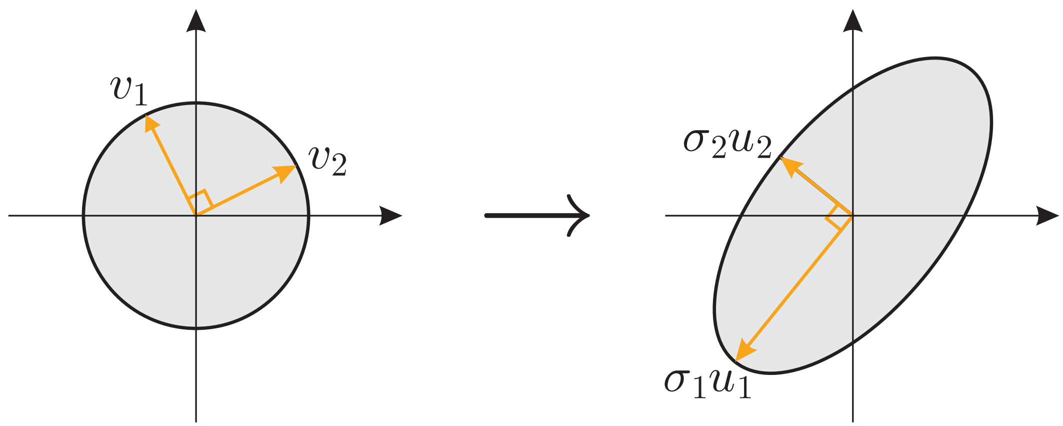
an *extremely* important fact:

every matrix $A \in \mathbb{R}^{m \times n}$ maps the unit ball in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^m$

$$S = \left\{ x \in \mathbb{R}^n \mid \|x\| \leq 1 \right\} \qquad AS = \left\{ Ax \mid x \in S \right\}$$

singular values and singular vectors

first, assume $A \in \mathbb{R}^{m \times n}$ is skinny and full rank



- the numbers $\sigma_1, \ldots, \sigma_n$ are called the *singular values* of A

  by convention, $\sigma_i > 0$

- the vectors $u_1, \ldots, u_n$ are called the *left singular vectors* of A

  these are *unit vectors* along the principal semiaxes of AS

- the vectors $v_1, \ldots, v_n$ are called the *right singular vectors* of A

  these are the *preimages* of the principal semiaxes, defined so that

$$Av_i = \sigma_i u_i$$

The Thin Singular Value Decomposition

we have $A \in \mathbb{R}^{m \times n}$, skinny and full rank, and

$$Av_i = \sigma_i u_i \qquad \text{for } 1 \le i \le n$$

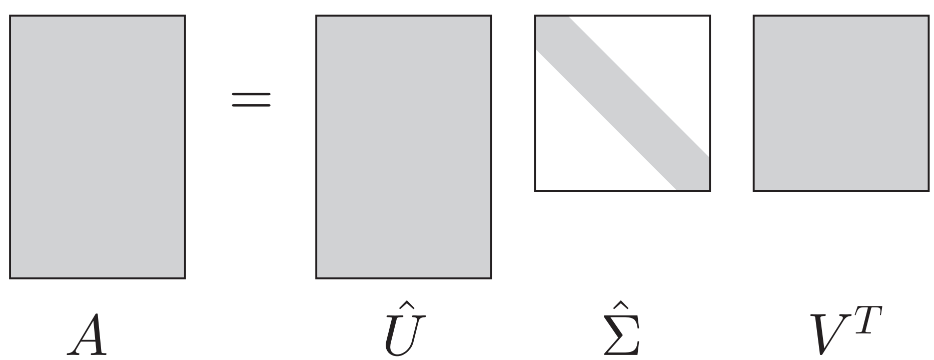
let

$$\hat{U} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \qquad \hat{\Sigma} = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix} \qquad V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

in matrix form, the above equation is $AV = \hat{U}\hat{\Sigma}$ and since V is orthogonal

$$A = \hat{U}\hat{\Sigma}V^T$$

called the *thin (or reduced) SVD* of A



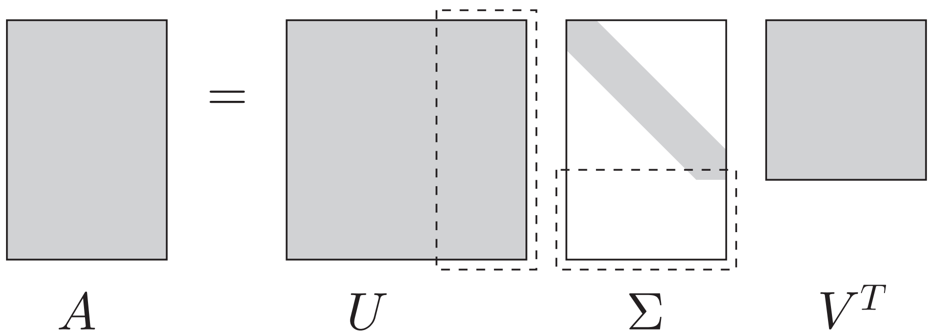
The Full Singular Value Decomposition

we can add extra orthonormal columns to $\hat{U}$ to make

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

an orthogonal matrix; we also add extra rows of zeros to $\hat{\Sigma}$, so

$$\boxed{A = U \Sigma V^T}$$



this is the *(full) singular value decomposition* of A

every matrix A has a singular value decomposition; if A is not full rank, then some of the diagonal entries of Σ will be zero

example: SVD

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 4 & 2 \end{bmatrix}$$

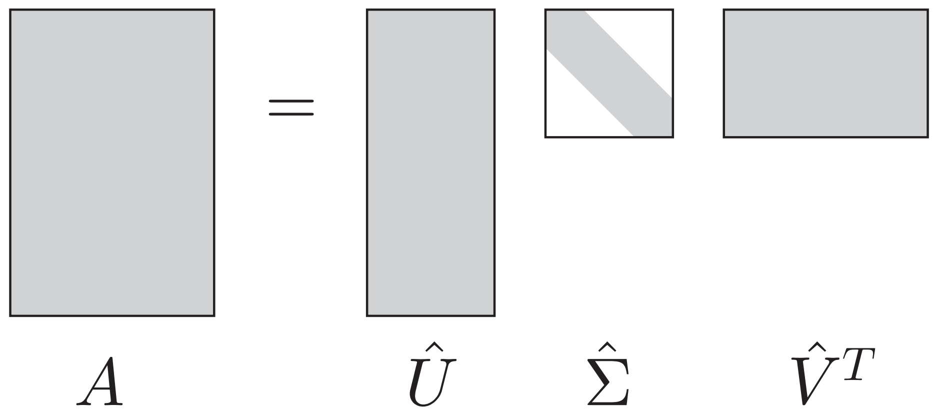
use `[U,S,V]=svd(A)` in Matlab

$$A = \begin{bmatrix} -0.319 & 0.915 & -0.248 \\ -0.542 & -0.391 & -0.744 \\ -0.778 & -0.103 & 0.620 \end{bmatrix} \begin{bmatrix} 5.747 & 0 \\ 0 & 1.403 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.880 & -0.476 \\ -0.476 & 0.880 \end{bmatrix}$$

thin SVD for matrices without full rank

when $A \in \mathbb{R}^{m \times n}$, skinny, not full rank, the *thin SVD* is

$$A = \hat{U}\hat{\Sigma}\hat{V}^T$$

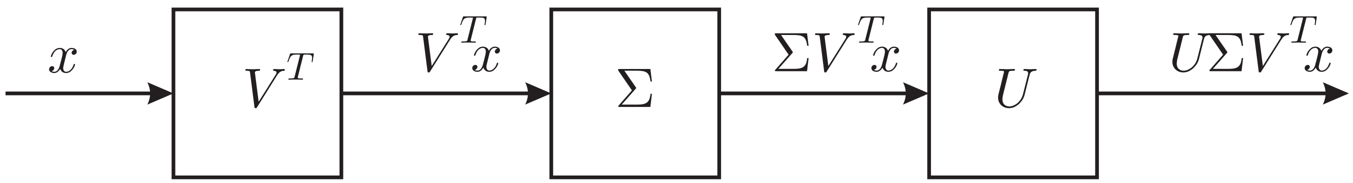


here

- $\hat{U} \in \mathbb{R}^{m \times r}$ has orthonormal columns
- $\hat{\Sigma} \in \mathbb{R}^{r \times r}$ is diagonal
- $\hat{V} \in \mathbb{R}^{n \times r}$ has orthonormal columns

if A is fat and not full rank, there is a similar *thin SVD*

interpretation of SVD



the SVD decomposes the linear map into

- *rotate* by V^T
- *diagonal scaling* by σ_i
- pad with zeros (if $m > n$) or truncate (if $m < n$) to make m-vector
- *rotate* by U

note that, unlike the eigen-decomposition, input and output directions are different

rank

the SVD tells us the rank of a matrix:

> if A has rank r, then A has r non-zero singular values

this is because $A = U\Sigma V^T$, and U and V are rotations, so

$$\operatorname{rank}(A) = \dim \operatorname{range}(A) = \dim \operatorname{range}(\Sigma)$$

range and null space

if $r = \operatorname{rank}(A)$, then

- $\{u_1, \dots, u_r\}$ are an orthonormal basis for $\operatorname{range}(A)$
- $\{v_{r+1}, \dots, v_n\}$ are an orthonormal basis for $\operatorname{null}(A)$

example: testing achievability of desired outputs for control

we want to find x so that $y_{\text{des}} = Ax$

question: is there such a x? i.e., is $y_{\text{des}} \in \operatorname{range}(A)$?

bad approach: (numerically unreliable)

check if $\operatorname{rank} \begin{bmatrix} y_{\text{des}} & A \end{bmatrix} = \operatorname{rank}(A)$

good approach:

use svd: $A = U\Sigma V^T$

component of y_{des} in $\operatorname{range}(A)$ is $\displaystyle\sum_{i=1}^{r} u_i u_i^T y_{\text{des}}$

residual z is

$$z = y_{\text{des}} - \sum_{i=1}^{r} u_i u_i^T y_{\text{des}} = (I - \hat{U}\hat{U}^T) y_{\text{des}}$$

algebraic interpretation of SVD

the SVD

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

can be written as

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

where $r = \operatorname{rank} A$

so every matrix A is the sum of *rank one* matrices

Computing the SVD

the singular values of A are
the square roots of the nonzero eigenvalues of A^TA or AA^T.

because

$$\begin{aligned} AA^T &= U\Sigma V^T V \Sigma^T U^T \\ &= U\Sigma\Sigma^T U^T \end{aligned}$$

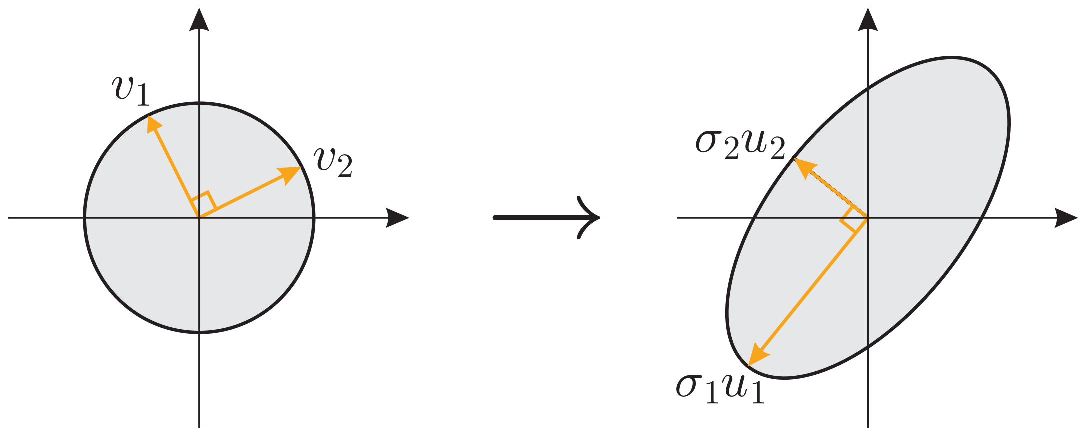
this also implies that

- left singular vectors u_i are the eigenvectors of AA^T
- similarly, right singular vectors v_i are the eigenvectors of A^TA

this gives one way to compute the SVD; it is never used in practice

The SVD and the Control Ellipsoid

we want to choose x so that $Ax = y_{\text{des}}$.



- input right singular vector v_i is mapped to left singular vector u_i, amplified by σ_i
- σ_i measures the *actuator authority* in the direction $u_i \in \mathbb{R}^m$
- $r < m \quad \Longrightarrow$ no control authority in directions $u_{r+1}, \ldots, u_m$ (outside $\operatorname{range}(A)$)
- if A is fat and full rank, then the ellipsoid is

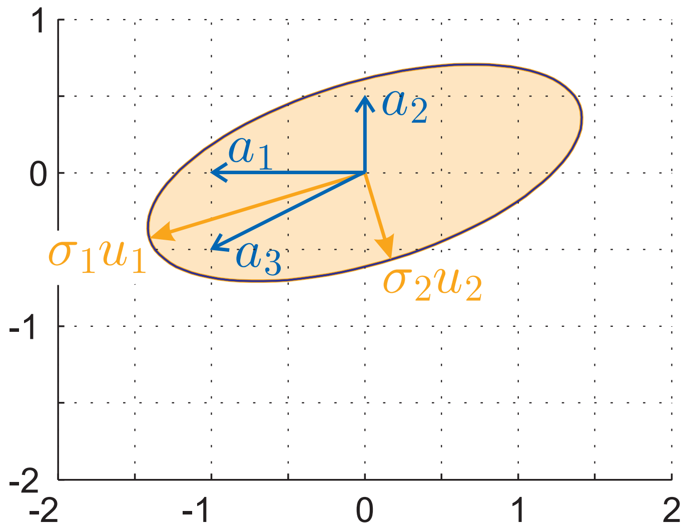
$$E = \left\{ y \in \mathbb{R}^m \mid y^T \left(AA^T\right)^{-1} y \leq 1 \right\}$$

because

$$AA^T = U\hat{\Sigma}\hat{V}^T\hat{V}\hat{\Sigma}U^T = U\hat{\Sigma}^2U^T$$

example: forces applied to a hovercraft

apply forces via thrusters x_i in specific directions on a hovercraft



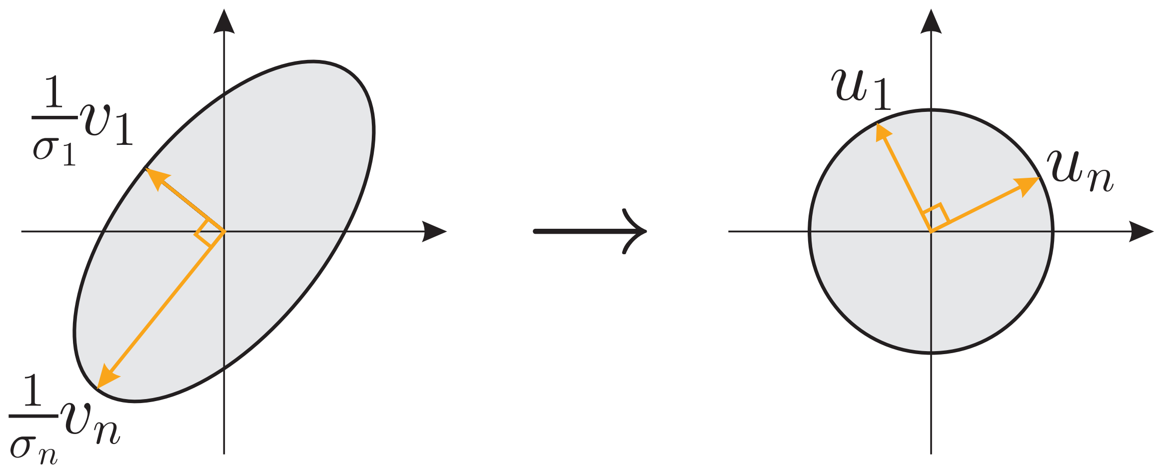
$$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 0.5 & -0.5 \end{bmatrix}$$

- total force on body $y = Ax$,
- x_i is power (in W) supplied to thruster i
- $\|a_i\|$ is *efficiency* of thruster
- most efficient direction we can apply thrust is given by long axis
- $\sigma_1 = 1.4668$, $\sigma_2 = 0.5904$

The Preimage

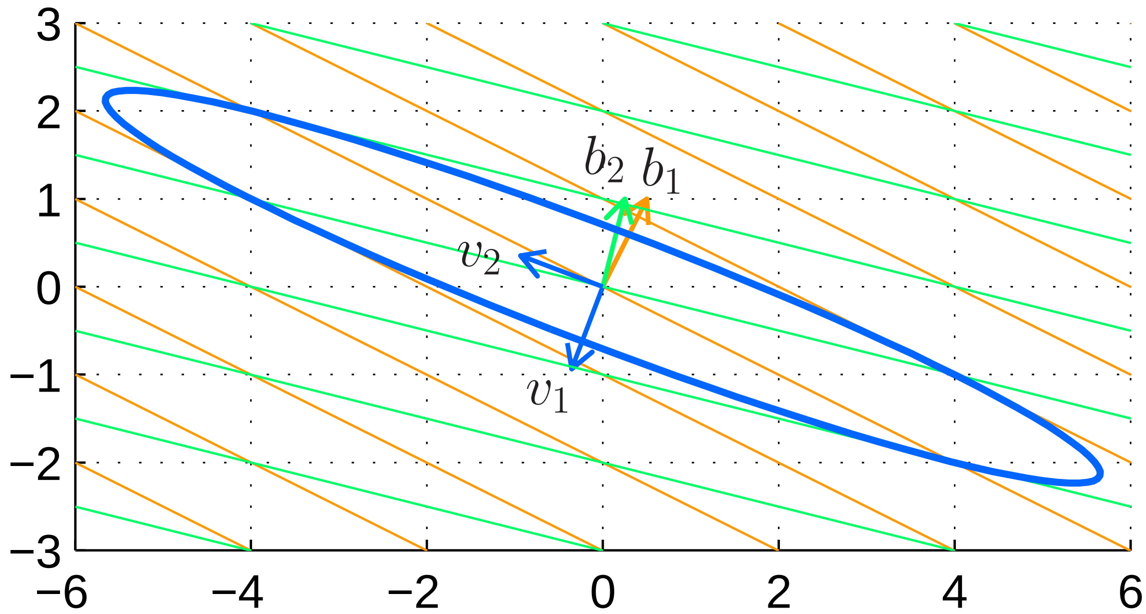
- For control, uses SVD in forwards direction; If $\|x\| \leq 1$, where is y?
- For estimation, consider backward direction; if $\|y\| \leq 1$, where was x?

  Answer: in the ellipsoid $\{\, x \in \mathbb{R}^n \mid x^T A^T A x \leq 1 \,\}$



- short axis of ellipsoid is *stretched most* by sensing. Small changes to x in the direction v_1 cause large changes in sensor readings y; sensors are *highly sensitive*
- long axis of ellipsoid is *stretched least* by sensing. Small changes to x in the direction v_n cause small changes in sensor readings y; sensors are *insensitive*

example: row interpretation



$$A = \begin{bmatrix} b_1^T \\ b_2^T \end{bmatrix} = \begin{bmatrix} 0.5 & 1 \\ 0.25 & 1 \end{bmatrix}$$

v_1 and v_2 are eigenvectors of $A^T A$ with corresponding singular values

$$\sigma_1 \approx 1.5117 \qquad \sigma_2 \approx 0.1654$$

sensors are 10 times more sensitive to changes in x in the v_1 direction than in v_2 direction.

summary: svd, control and estimation

$A \in \mathbb{R}^{m \times n}$; singular values of A are the square-roots of the eigenvalues of AA^T, *the same as* those of $A^T A$

control

- *control ellipsoid* $\subset \mathbb{R}^m$ (output space)
- semiaxes are *left* singular vectors (they are in the output space)
- singular values measure actuator authority/efficiency

estimation

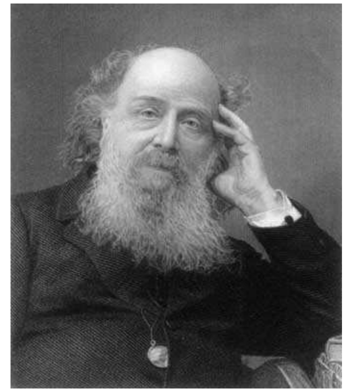
- *estimation ellipsoid* $\subset \mathbb{R}^n$ (space of unknowns)
- semiaxes are *right* singular vectors (they are in the space of inputs/unknowns)
- singular values measure sensor sensitivity

History

the SVD was invented/discovered by

Eugenio Beltrami
(1835–1900) (Italy)

Marie E. Camille Jordan
(1838–1922) (France)



James Sylvester
(1814–1897) (England)

coined words *graph, group, Hessian, hyper-plane, invariant, jacobian, matrix. . .*

wrote one book, on poetry

Erhard Schmidt
(1876–1959) (Germany)

Hermann Weyl
(1885–1955) (Switzerland)

at ETH with Einstein; moved to Princeton in 1933

The Matrix Norm

the *norm* of a matrix A is

$$\|A\| = \max \left\{ \frac{\|Ax\|}{\|x\|} \;\middle|\; x \in \mathbb{R}^n, x \neq 0 \right\}$$

also called the *operator norm*, *spectral norm* or *induced norm*.

- gives the maximum *gain* or *amplification* of A
- the maximizing x is given by v_1 and the output is $Av_1 = \sigma_1 u_1$
- so the norm of A is the *maximum singular value* of A:

$$\|A\| = \sigma_1(A)$$

- similarly the minimum (nonzero) gain is achieved by v_r, with an output of $\sigma_r u_r$

properties of the matrix norm

satisfies the usual properties of a norm:

- *scaling:* $\|cA\| = |c|\|A\|$ for $c \in \mathbb{R}$.
- *triangle inequality:* $\|A + B\| \leq \|A\| + \|B\|$.
- *definiteness:* $\|A\| = 0 \iff A = 0$.

also

- $\|A\| = \|A^T\|$
- if $A \in \mathbb{R}^{n \times 1}$ (so A is just a vector) then

  matrix norm of A = vector norm of A

- $\|A\| \leq \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \|a_{ij}\|^2 \right)^{\frac{1}{2}}$
- $\|A\| \geq \max_i \max_j |a_{ij}|$

triangle inequality

$$\|A + B\| \leq \|A\| + \|B\|$$

holds because

$$\begin{aligned}
\|A + B\| &= \max_{x \neq 0} \frac{\big\|(A + B)x\big\|}{\|x\|} \\
&\leq \max_{x \neq 0} \frac{\|Ax\| + \|Bx\|}{\|x\|} \\
&\leq \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} + \max_{x \neq 0} \frac{\|Bx\|}{\|x\|}
\end{aligned}$$

matrix-vector composition

$$\|Ax\| \leq \|A\| \|x\|$$

matrix-matrix composition

$$\|AB\| \leq \|A\| \|B\|$$

because

$$\begin{aligned}
\|AB\| &= \max_{x \neq 0} \frac{\|ABx\|}{\|x\|} \\
&\leq \max_{x \neq 0} \frac{\|A\| \|Bx\|}{\|x\|} \\
&\leq \|A\| \|B\|
\end{aligned}$$

this inequality is also called the *submultiplicative inequality*.

example: matrix norm and estimation

example: suppose A is square and invertible

we have

$$y_{\mathsf{meas}} = Ax + w$$

where w is noise

estimator $x_{\mathsf{est}} = A^{-1} y_{\mathsf{meas}}$ gives

$$x_{\mathsf{est}} = x + A^{-1} w$$

so the estimation error satisfies

$$\|x - x_{\mathsf{est}}\| \leq \|A^{-1}\| \|w\|$$

so if we know $\|w\| \leq 1$ then we know $\|x - x_{\mathsf{est}}\| \leq \|A^{-1}\|$

The SVD and Rank

SVD captures the *numerical rank* of a matrix $A \in \mathbb{R}^{m \times n}$.

$$\min\left\{ \|A - B\| \mid B \in \mathbb{R}^{m \times n}, \operatorname{rank}(B) \le k \right\} = \sigma_{k+1}$$

the optimal B is given by

$$B = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

this is not obvious; we will not prove it

example

if a matrix $A \in \mathbb{R}^{10 \times 10}$ has singular values

$$\sigma_1 = 100, \quad \sigma_2 = 35, \quad \sigma_3 = 10, \quad \sigma_4 = 2$$

and $\sigma_5 \le 0.00001$, then we might say its *numerical rank* is 4

example:low rank approximation

$$A = \begin{bmatrix} 11.08 & 6.82 & 1.76 & -6.82 \\ 2.50 & -1.01 & -2.60 & 1.19 \\ -4.88 & -5.07 & -3.21 & 5.20 \\ -0.49 & 1.52 & 2.07 & -1.66 \\ -14.04 & -12.40 & -6.66 & 12.65 \\ 0.27 & -8.51 & -10.19 & 9.15 \\ 9.53 & -9.84 & -17.00 & 11.00 \\ -12.01 & 3.64 & 11.10 & -4.48 \end{bmatrix}$$

$$\approx \begin{bmatrix} -0.25 & 0.45 & 0.62 & 0.33 & 0.46 & 0.05 & -0.19 & 0.01 \\ 0.07 & 0.11 & 0.28 & -0.78 & -0.10 & 0.33 & -0.42 & 0.05 \\ 0.21 & -0.19 & 0.49 & 0.11 & -0.47 & -0.61 & -0.24 & -0.01 \\ -0.08 & -0.02 & 0.20 & 0.06 & -0.27 & 0.30 & 0.20 & -0.86 \\ 0.50 & -0.55 & 0.14 & -0.02 & 0.61 & 0.02 & -0.08 & -0.20 \\ 0.44 & 0.03 & -0.05 & 0.50 & -0.30 & 0.55 & -0.36 & 0.18 \\ 0.59 & 0.43 & 0.21 & -0.14 & -0.03 & -0.00 & 0.62 & 0.13 \\ -0.30 & -0.51 & 0.43 & 0.02 & -0.14 & 0.34 & 0.41 & 0.40 \end{bmatrix} \begin{bmatrix} 36.83 & 0 & 0 & 0 \\ 0 & 26.24 & 0 & 0 \\ 0 & 0 & 0.02 & 0 \\ 0 & 0 & 0 & 0.01 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.04 & -0.54 & -0.61 & 0.58 \\ 0.92 & 0.17 & -0.33 & -0.14 \\ -0.14 & -0.49 & -0.31 & -0.80 \\ -0.36 & 0.66 & -0.65 & -0.09 \end{bmatrix}$$

$$A_{\text{approx}} \approx \begin{bmatrix} -0.25 & 0.45 \\ 0.07 & 0.11 \\ 0.21 & -0.19 \\ -0.08 & -0.02 \\ 0.50 & -0.55 \\ 0.44 & 0.03 \\ 0.59 & 0.43 \\ -0.30 & -0.51 \end{bmatrix} \begin{bmatrix} 36.83 & 0 \\ 0 & 26.24 \end{bmatrix} \begin{bmatrix} -0.04 & -0.54 & -0.61 & 0.58 \\ 0.92 & 0.17 & -0.33 & -0.14 \end{bmatrix} \approx \begin{bmatrix} 11.08 & 6.83 & 1.77 & -6.81 \\ 2.50 & -1.00 & -2.60 & 1.19 \\ -4.88 & -5.07 & -3.21 & 5.21 \\ -0.49 & 1.52 & 2.07 & -1.66 \\ -14.04 & -12.40 & -6.66 & 12.65 \\ 0.27 & -8.51 & -10.19 & 9.15 \\ 9.53 & -9.84 & -17.00 & 11.00 \\ -12.01 & 3.64 & 11.10 & -4.47 \end{bmatrix}$$

here $\|A - A_{\text{approx}}\| \le \sigma_3 \approx 0.02$

example: application of low rank approximants

suppose $A \in \mathbb{R}^{10000 \times 10000}$ is a dense matrix, so computing the matrix-vector product Ax is computationally expensive; 10^8 multiplications

if A has singular values $\sigma_1 = 100, \quad \sigma_2 = 35, \quad \sigma_3 = 10, \quad \sigma_4 = 2$, and $\sigma_k \le 0.001$ for $k \ge 5$, we can compute Ax very efficiently

the optimal rank 4 approximant to A is $A_{\text{approx}} = \displaystyle\sum_{i=1}^{4} \sigma_i u_i v_i^T$

so let $y_{\text{approx}} = A_{\text{approx}} x = 100(v_1^T x)u_1 + 35(v_2^T x)u_2 + 10(v_3^T x)u_3 + 2(v_4^T x)u_4$

this is a *simplified model* for $y = Ax$

error analysis

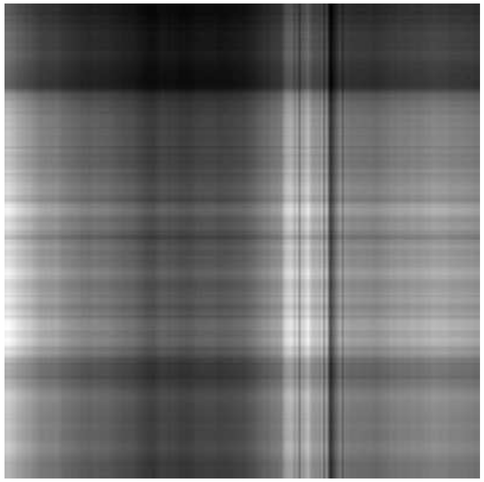
$$\begin{aligned}\|Ax - y_{\text{approx}}\| &= \|(A - A_{\text{approx}})x\| \\ &\le \|A - A_{\text{approx}}\| \|x\| \le 0.001\|x\|\end{aligned}$$

which gives a relative error of 0.1% in 4×10^4 multiplications

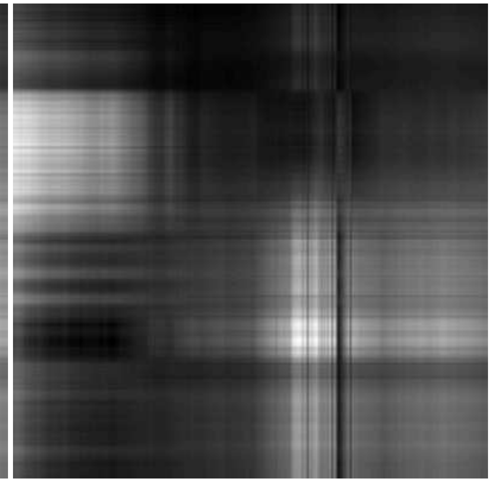
example: image compression

we can represent a (grayscale) image by a matrix; each pixel is an entry between 0 and 1.
image size: 400×400

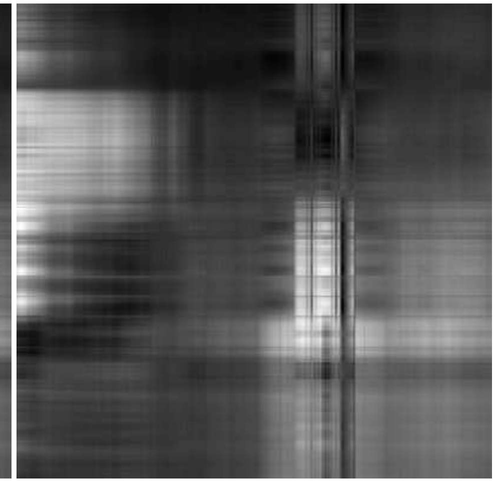
rank 400



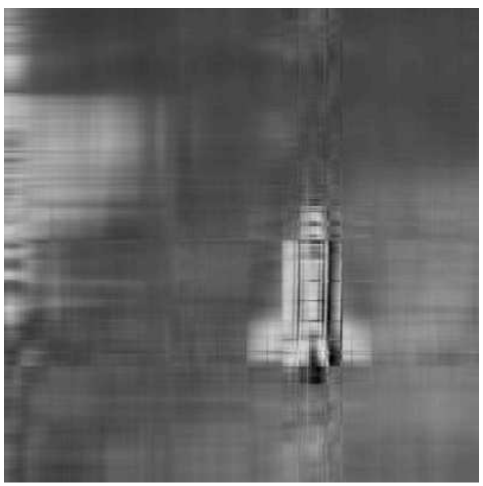
rank 1



rank 2

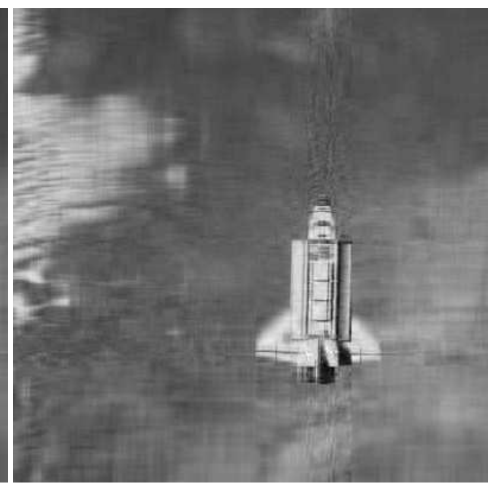


rank 3

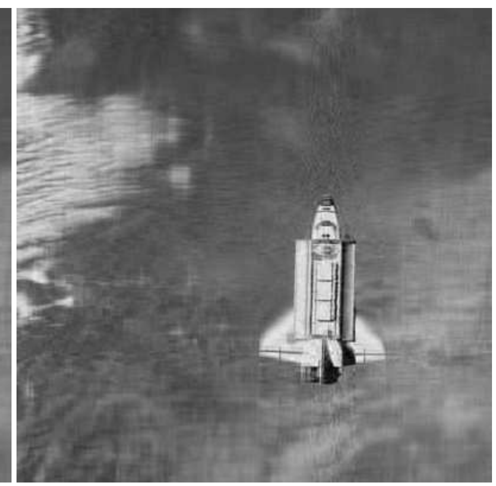


rank 8

rank 10



rank 20



rank 30

Summary: the SVD

- svd is readily computed
- gives a numerically reliable way to compute with e.g. rank, null space, range of A
- gives the lengths and directions of the semi-axes of the ellipsoids for control and estimation problems
- gives low-rank approximations, which make models simpler and computations cheaper