Overview

- The singular value decomposition (SVD)
  - Motivation
  - Theorem
- Another essential diagram: SVD finds orthonormal bases for $C(\boldsymbol{A}), N(\boldsymbol{A}^T), C(\boldsymbol{A}^T), N(\boldsymbol{A})$
- SVD computation
- Rank-1 expansion of a matrix
- Matrix norm
- SVD in image compression, analysis
- The pseudo-inverse

- Motivation: FTLA does not specify bases for $C(\boldsymbol{A}), N(\boldsymbol{A}^T), C(\boldsymbol{A}^T), N(\boldsymbol{A})$.
- Question: Is there some "natural" basis for the fundamental matrix subspaces?
- Consider linear mapping: $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
  - The input $\boldsymbol{x}$ is given in the identity matrix basis, $\boldsymbol{x} = \boldsymbol{I}_n \boldsymbol{x}$
  - The output $\boldsymbol{b} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ is also obtained in the identity matrix basis, $\boldsymbol{y} = \boldsymbol{I}_m \boldsymbol{y}$
  - In these basis the effect of $\boldsymbol{A}$ might be costly to compute

$$\boldsymbol{b} = x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + \cdots + x_n \boldsymbol{a}_n \Rightarrow b_i = x_1 a_{i1} + x_2 a_{i2} + \cdots + x_n a_{i,n}.$$

  - What would be simpler? One possibility: a simple scaling of each component suggested by

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \boldsymbol{D}\boldsymbol{x} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)\boldsymbol{x} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \Rightarrow b_i = \lambda_i x_i$$

- Seek different bases for domain, codomain of $\boldsymbol{f}:\mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$
  - an orthonormal basis $\boldsymbol{V}$ in $\mathbb{R}^n$, $\boldsymbol{V} \in \mathbb{R}^{n\times n}$, $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$

$$\boldsymbol{I}\boldsymbol{x} = \boldsymbol{V}\boldsymbol{y} \Rightarrow \boldsymbol{y} = \boldsymbol{V}^T\boldsymbol{x}$$

  - an orthonormal basis $\boldsymbol{U}$ in $\mathbb{R}^m$, $\boldsymbol{U} \in \mathbb{R}^{m\times m}$, $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$

$$\boldsymbol{I}\boldsymbol{b} = \boldsymbol{U}\boldsymbol{c} \Rightarrow \boldsymbol{c} = \boldsymbol{U}^T\boldsymbol{b}$$

  - impose that the effect of $\boldsymbol{A}$ in the new bases is a simple component scaling

$$\boldsymbol{c} = \boldsymbol{\Sigma}\boldsymbol{y} \Rightarrow \boldsymbol{U}^T\boldsymbol{b} = \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} \Rightarrow \boldsymbol{b} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} \Rightarrow$$

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

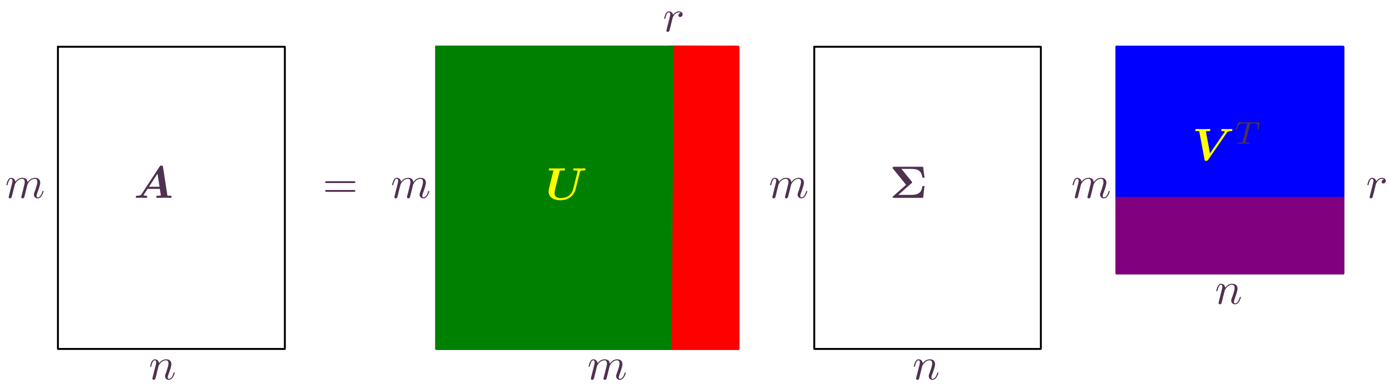
  - Note that $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$

Theorem. *(SVD) For any* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, *with* $\boldsymbol{U} \in \mathbb{R}^{m \times m}$, $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ *orthogonal,* $\boldsymbol{\Sigma} \in \mathbb{R}_+^{m \times n}$ *pseudo-diagonal* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, \ldots, 0)$, $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, $r \leqslant \min(m, n)$

The SVD is determined by eigendecomposition of $\boldsymbol{A}^T\boldsymbol{A}$, and $\boldsymbol{A}\boldsymbol{A}^T$

- $\boldsymbol{A}^T\boldsymbol{A} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T) = \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})\boldsymbol{V}^T$, an eigendecomposition of $\boldsymbol{A}^T\boldsymbol{A}$. The columns of $\boldsymbol{V}$ are eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ and called *right singular vectors* of $\boldsymbol{A}$
- $\boldsymbol{A}\boldsymbol{A}^T = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)(\boldsymbol{U}\boldsymbol{\Sigma}^T\boldsymbol{V}^T)^T = \boldsymbol{U}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T)\boldsymbol{U}^T$, an eigendecomposition of $\boldsymbol{A}\boldsymbol{A}^T$. The columns of $\boldsymbol{U}$ are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ and called *left singular vectors* of $\boldsymbol{A}$
- The matrix $\boldsymbol{\Sigma}$ has zero elements except for the diagonal that contains σ_i, the *singular values* of $\boldsymbol{A}$, computed as the square roots of the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ (or $\boldsymbol{A}\boldsymbol{A}^T$)

- SVD of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ reveals: $\mathrm{rank}(\boldsymbol{A})$, bases for $C(\boldsymbol{A}), N(\boldsymbol{A}^T), C(\boldsymbol{A}^T), N(\boldsymbol{A})$

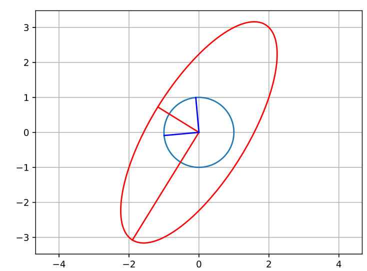


$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \dots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_r^T \\ \boldsymbol{v}_{r+1}^T \\ \vdots \\ \boldsymbol{v}_n^T \end{bmatrix}$$

- From $A = U\Sigma V^T$ deduce $AA^T = U\Sigma^2 U^T$, $A^TA = V\Sigma^2 V^T$, hence U is the eigenvector matrix of AA^T, and V is the eigenvector matrix of A^TA
- SVD computation is available in Julia, Octave, Matlab, Mathematica ...

```
∴ short(x) = round(x,digits=6);
∴ A=[2 -1; -3 1]; F=svd(A); U=F.U; Σ=Diagonal(F.S); Vt=F.Vt; short.([A U*Σ*Vt])
∴ short.([U Σ Vt'])
∴
```

- SVD of $A = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}$ diagram, $f(x) = Ax$



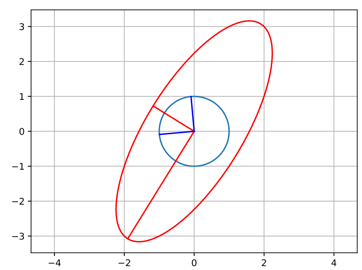
- From $A = U\Sigma V^T$ deduce $AA^T = U\Sigma^2 U^T$, $A^T A = V\Sigma^2 V^T$, hence U is the eigenvector matrix of AA^T, and V is the eigenvector matrix of $A^T A$
- SVD computation is available in Julia, Octave, Matlab, Mathematica ...

```
∴ short(x) = round(x,digits=6);
∴ A=[2 -1; -3 1]; F=svd(A); U=F.U; Σ=Diagonal(F.S); Vt=F.Vt; short.([A U*Σ*Vt])
```

$$\begin{bmatrix} 2.0 & -1.0 & 2.0 & -1.0 \\ -3.0 & 1.0 & -3.0 & 1.0 \end{bmatrix} \tag{1}$$

```
∴ short.([U Σ Vt'])
∴
```

- SVD of $A = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}$ diagram, $f(x) = Ax$



- From $A = U\Sigma V^T$ deduce $AA^T = U\Sigma^2 U^T$, $A^TA = V\Sigma^2 V^T$, hence U is the eigenvector matrix of AA^T, and V is the eigenvector matrix of A^TA
- SVD computation is available in Julia, Octave, Matlab, Mathematica …

```
∴ short(x) = round(x,digits=6);
```

```
∴ A=[2 -1; -3 1]; F=svd(A); U=F.U; Σ=Diagonal(F.S); Vt=F.Vt; short.([A U*Σ*Vt])
```

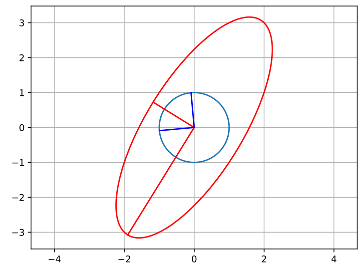
$$\begin{bmatrix} 2.0 & -1.0 & 2.0 & -1.0 \\ -3.0 & 1.0 & -3.0 & 1.0 \end{bmatrix} \quad (2)$$

```
∴ short.([U Σ Vt'])
```

$$\begin{bmatrix} -0.576048 & 0.817416 & 3.864328 & 0.0 & -0.932722 & -0.360597 \\ 0.817416 & 0.576048 & 0.0 & 0.258777 & 0.360597 & -0.932722 \end{bmatrix} \quad (3)$$

```
∴
```

- SVD of $A = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}$ diagram, $f(x) = Ax$



- $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, carry out block multiplication

$$\boldsymbol{A} = \left[\begin{array}{cccccc} \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \ldots & \boldsymbol{u}_m \end{array}\right] \left[\begin{array}{ccccc} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \end{array}\right] \left[\begin{array}{c} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_r^T \\ \boldsymbol{v}_{r+1}^T \\ \vdots \\ \boldsymbol{v}_n^T \end{array}\right] \Rightarrow$$

$$\boldsymbol{A} = \left[\begin{array}{cccccc} \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_r & \boldsymbol{u}_{r+1} & \ldots & \boldsymbol{u}_m \end{array}\right] \left[\begin{array}{c} \sigma_1 \boldsymbol{v}_1^T \\ \vdots \\ \sigma_r \boldsymbol{v}_r^T \\ \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{array}\right] = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \ldots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^T$$

- The above is known as a "rank-one" expansion since $\operatorname{rank}(\boldsymbol{u}_k \boldsymbol{v}_k^T) = 1$. Note that $\boldsymbol{u}_k \boldsymbol{v}_k^T \in \mathbb{R}^{m \times n}$ and is a matrix whose columns are n scalings of $\boldsymbol{u}_k$
- SVD theorem: $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$, Often $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_k \gg \sigma_{k+1} \geqslant \cdots \geqslant \sigma_r > 0$

- $\boldsymbol{U}, \boldsymbol{V}$ specify intrinsic directions within $\mathbb{R}^m, \mathbb{R}^n$ along which $\boldsymbol{A}$ acts as scaling transformation
- Applying linear mapping to the $\boldsymbol{v}_1$ vector, $\boldsymbol{f}(\boldsymbol{v}_1) = \boldsymbol{A}\boldsymbol{v}_1$

$$\boldsymbol{A}\boldsymbol{v}_1 = \left(\sum_{i=1}^{p} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T \right) \boldsymbol{v}_1 = \sum_{i=1}^{p} \sigma_i \boldsymbol{u}_i (\boldsymbol{v}_i^T \boldsymbol{v}_1) = \sigma_1 \boldsymbol{u}_1$$

- Direction most amplified by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ is $\boldsymbol{v}_1$ and the result is the vector $\sigma_1 \boldsymbol{u}_1$
- Define a matrix norm as the largest amplification factor

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2.$$

- The largest singular value σ_1 is the matrix 2-norm

$$\sigma_1 = \max_{\|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2 .$$

- Full SVD

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T, r \leqslant \min(m, n).$$

- Truncated SVD

$$\boldsymbol{A} \cong \boldsymbol{A}_p = \sum_{i=1}^{p} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T.$$

- Many applications, e.g., image compression

Figure 1. Successive SVD approximations of Andy Warhol's painting, *Marilyn Diptych* (~1960), with $k = 10, 20, 40$ rank-one updates.

- Consider $x_1, x_2 \colon \mathbb{R} \to \mathbb{R}$, data streams in time of inputs $x_1(t)$ and outputs $x_2(t)$
- Is there some function f linking outputs to inputs? $f(x_1(t)) = x_2(t)$
- Seek answer by first asking: is x_2 *correlated* to x_1?
- Introduce mean values

$$\mu_1 \cong \bar{x}_1 = \frac{1}{N}\sum_{i=1}^{N} x_1(t_i) = E[x_1], \mu_2 \cong \bar{x}_2 = \frac{1}{N}\sum_{i=1}^{N} x_2(t_i) = E[x_2].$$

- E is the *expectation*, a linear mapping, $E \colon \mathbb{R}^N \to \mathbb{R}$ whose associated matrix is

$$\boldsymbol{E} = \frac{1}{N}[\ 1\ \ 1\ \ \ldots\ \ 1\ ].$$

- Shift data such that $\bar{x}_1 = \bar{x}_2 = 0$. Define correlation coefficient

$$\rho(x_1, x_2) = \frac{E[x_1 x_2]}{\sigma_1\, \sigma_2} = \frac{E[x_1 x_2]}{\sqrt{E[x_1^2]\, E[x_2^2]}} = \frac{\boldsymbol{x}_1^T \boldsymbol{x}_2}{\|\boldsymbol{x}_1\|\, \|\boldsymbol{x}_2\|}.$$

- *uncorrelated*, if $\rho = 0$; *correlated*, if $\rho = 1$; *anti-correlated*, if $\rho = -1$.

- Correlated signals

```
∴ t=0:0.01:1; x1=1.0*t; x2=t.^2; rho=transpose(x1)*x2/norm(x1)/norm(x2)
∴
```

- Uncorrelated signals

```
∴ m=size(x1)[1]; x3=2*(rand(m,1).-0.5)[:,1]; rho=transpose(x1)*x3/norm(x1)/norm(x3)
∴
```

- Anticorrelated signals

```
∴ x4=-t.^2; rho=transpose(x1)*x4/norm(x1)/norm(x4)
```

- Correlated signals

```
∴ t=0:0.01:1; x1=1.0*t; x2=t.^2; rho=transpose(x1)*x2/norm(x1)/norm(x2)
```

0.968249831385581

```
∴
```

- Uncorrelated signals

```
∴ m=size(x1)[1]; x3=2*(rand(m,1).-0.5)[:,1]; rho=transpose(x1)*x3/norm(x1)/norm(x3)
∴
```

- Anticorrelated signals

```
∴ x4=-t.^2; rho=transpose(x1)*x4/norm(x1)/norm(x4)
```

- Correlated signals

```
∴ t=0:0.01:1; x1=1.0*t; x2=t.^2; rho=transpose(x1)*x2/norm(x1)/norm(x2)
```

0.968249831385581

```
∴
```

- Uncorrelated signals

```
∴ m=size(x1)[1]; x3=2*(rand(m,1).-0.5)[:,1]; rho=transpose(x1)*x3/norm(x1)/norm(x3)
```

0.040354121344516436

```
∴
```

- Anticorrelated signals

```
∴ x4=-t.^2; rho=transpose(x1)*x4/norm(x1)/norm(x4)
```

- Are input and output parameters $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$ well chosen?
- Perhaps components are redundant, a more economical description might be

$$\boldsymbol{u} \in \mathbb{R}^q, \boldsymbol{v} \in \mathbb{R}^p, p < m, q < n$$

- Extend idea from correlation coefficient: take N measurements

$$\boldsymbol{X} = [\begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{array}] \in \mathbb{R}^{N \times n}, \boldsymbol{Y} = [\begin{array}{cccc} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \dots & \boldsymbol{y}_n \end{array}] \in \mathbb{R}^{N \times m}.$$

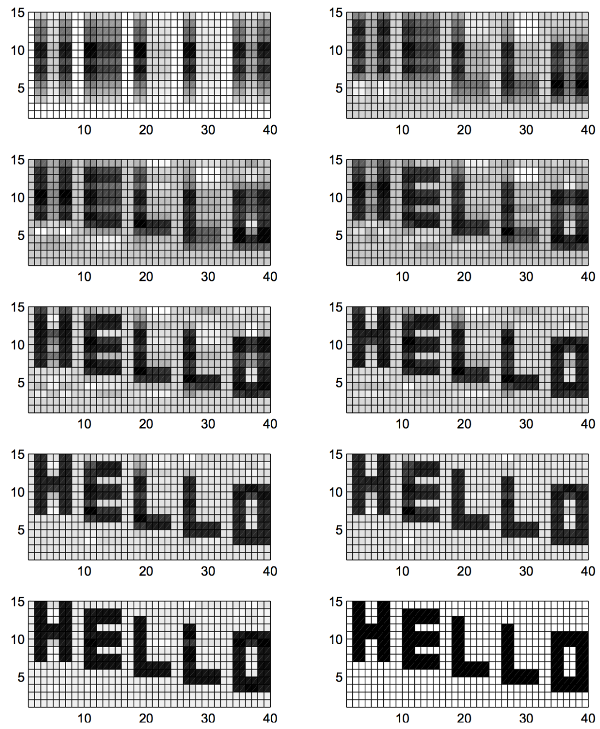
- Choose origin such that $E[\boldsymbol{x}] = \boldsymbol{0}, E[\boldsymbol{y}] = \boldsymbol{0}$.
- *Covariance matrix* (generalization of single variable *variance*)

$$\boldsymbol{C}_{\boldsymbol{X}} = \boldsymbol{X}^T \boldsymbol{X} = \left[\begin{array}{c} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_n^T \end{array}\right] [\begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{array}] = \left[\begin{array}{cccc} \boldsymbol{x}_1^T \boldsymbol{x}_1 & \boldsymbol{x}_1^T \boldsymbol{x}_2 & \dots & \boldsymbol{x}_1^T \boldsymbol{x}_n \\ \boldsymbol{x}_2^T \boldsymbol{x}_1 & \boldsymbol{x}_2^T \boldsymbol{x}_2 & \dots & \boldsymbol{x}_2^T \boldsymbol{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_n^T \boldsymbol{x}_1 & \boldsymbol{x}_n^T \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n^T \boldsymbol{x}_n \end{array}\right]$$

- SVDs: $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, $C_{\boldsymbol{X}} = \boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$
- Take first q column vectors of $\boldsymbol{V}$, $\boldsymbol{V}_q = [\ \boldsymbol{v}_1\ \ \dots\ \ \boldsymbol{v}_q\ ]$, $q < n$

$$\boldsymbol{x} = \boldsymbol{V}_q\boldsymbol{u} = [\ \boldsymbol{v}_1\ \ \boldsymbol{v}_2\ \ \dots\ \ \boldsymbol{v}_q\ ]\,\boldsymbol{u}.$$

- System description in terms of $\boldsymbol{u} \in \mathbb{R}^q$ is more economical than that in terms of $\boldsymbol{x} \in \mathbb{R}^n$
- In image compression, successive pixel columns are correlated and reduced descriptions are possible



- Consider linear system $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, $\boldsymbol{A}\in\mathbb{R}^{m\times m}$, $\operatorname{rank}(\boldsymbol{A})=m$. SVD solution steps:

1. Compute the SVD, $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{A}$;
2. Find the coordinates of $\boldsymbol{b}$ in the orthogonal basis $\boldsymbol{U}$, $\boldsymbol{c}=\boldsymbol{U}^T\boldsymbol{b}$;
3. Scale the coordinates of $\boldsymbol{c}$ by the inverse of the singular values $y_i=c_i/\sigma_i$, $i=1,\ldots,m$, such that $\Sigma\boldsymbol{y}=\boldsymbol{c}$ is satisfied;
4. Find the coordinates of $\boldsymbol{y}$ in basis $\boldsymbol{V}^T$, $\boldsymbol{x}=\boldsymbol{V}\boldsymbol{y}$.

- What if $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\operatorname{rank}(\boldsymbol{A})=r<m$. If $\boldsymbol{b}\in C(\boldsymbol{A})$ above procedure still works with a simple modification of step 3 with i going now from 1 to r

1. $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{A}$;
2. $\boldsymbol{c}=\boldsymbol{U}^T\boldsymbol{b}$;
3. $y_i=c_i/\sigma_i$, $i=1,\ldots,r$
4. $\boldsymbol{x}=\boldsymbol{V}\boldsymbol{y}$.

- If $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\operatorname{rank}(\boldsymbol{A})=r<m$. If $\boldsymbol{b}\notin C(\boldsymbol{A})$, the above steps give the best approximation of $\boldsymbol{b}$ by linear combination of columns of $\boldsymbol{A}$ in the 2-norm

- Since the steps to solve a linear system or find best approximation are identical define a matrix $\boldsymbol{A}^+$ that carries out all steps:

$$(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)\boldsymbol{x}=\boldsymbol{b}\Leftrightarrow\boldsymbol{U}(\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x})=\boldsymbol{b}$$

$$(\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x})=\boldsymbol{U}^T\boldsymbol{b}\Leftrightarrow\boldsymbol{\Sigma}(\boldsymbol{V}^T\boldsymbol{x})=\boldsymbol{U}^T\boldsymbol{b}$$

- Recall $\boldsymbol{\Sigma}=\mathrm{diag}(\sigma_1,\sigma_2,\ldots,\sigma_r,0,\ldots 0)$. Define $\boldsymbol{\Sigma}^+=\mathrm{diag}(1/\sigma_1,1/\sigma_2,\ldots,1/\sigma_r,0,\ldots 0)$

$$\boldsymbol{V}^T\boldsymbol{x}=\boldsymbol{\Sigma}^+\boldsymbol{U}^T\boldsymbol{b}$$

$$\boldsymbol{x}=\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T\boldsymbol{b}$$

- Gather all above steps into a single matrix $\boldsymbol{A}^+=\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$ called the *pseudo-inverse* of $\boldsymbol{A}$.
- The solution then to a linear system (either exact solution of best approximation) is

$$\boldsymbol{x}=\boldsymbol{A}^+\boldsymbol{b}$$

- In Julia, Matlab, Octave, above procedure is implemented through `x=A\b`.