SINGULAR VALUE DECOMPOSITION

Dr. Matthieu R Bloch

Monday, November 15, 2021

LOGISTICS

General announcements

- Assignment 6 to be posted… (grading traffic jam)
- 7 lectures left!

Assignment 5 and Midterm 2:

- Grading starting, we'll keep you posted

WHAT'S ON THE AGENDA FOR TODAY?

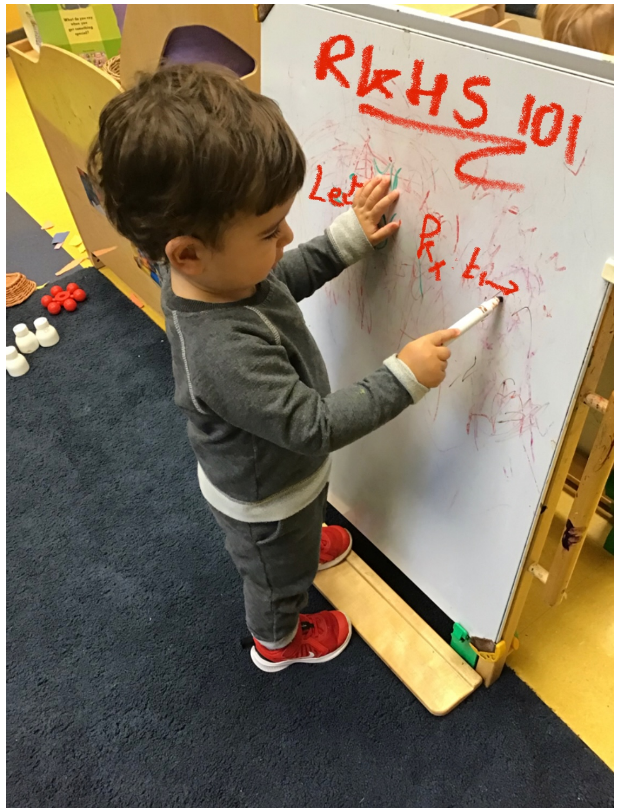
Last time:

- Singular value decomposition

Today

- Application to solving least squares

Reading: lecture notes 13/14



Toddlers can do it!

SINGULAR VALUE DECOMPOSITION

What happens for non-square matrice?

Theorem (Singular value decomposition)

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\operatorname{rank}(\mathbf{A}) = r$. Then $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ where

- $\mathbf{U} \in \mathbb{R}^{m \times r}$ such that $\mathbf{U}^\intercal \mathbf{U} = \mathbf{I}_r$ (orthonormal columns) $(UU^T \neq I)$ $U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix}$ $u_i^T u_j = 0$ $u_i^T u_i = 1$
- $\mathbf{V} \in \mathbb{R}^{n \times r}$ such that $\mathbf{V}^\intercal \mathbf{V} = \mathbf{I}_r$ (orthonormal columns) $(VV^T \neq I)$
- $\boldsymbol{\Sigma} \in \mathbb{R}^{r \times r}$ is diagonal with *positive entries*

$$\boldsymbol{\Sigma} \triangleq \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots \\ 0 & \sigma_2 & 0 & \cdots \\ \vdots & & \ddots & \\ 0 & \cdots & \cdots & \sigma_r \end{bmatrix}$$

and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. The σ_i are called the *singular values*

- We can write $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

We say that $\mathbf{A}$ is full rank is $r = \min(m, n)$

-

We can write $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal$

Hence $\quad U_1(U_1^T A V_1) V_1^T = (U_1 U_1^T) A (V_1 V_1^T) = A = U_1 \Sigma_1 V_1^T$

$$= \begin{array}{cc} \left[\begin{array}{c|c} U & U_0 \end{array}\right] & m \\ \begin{array}{cc} r & m-r \end{array} & \end{array} \quad \begin{array}{ccc} & n & \\ m & \left[\begin{array}{ccc|c} \sqrt{\lambda_1} & & & \\ & \ddots & & 0 \\ & & \sqrt{\lambda_r} & \\ \hline & 0 & & 0 \end{array}\right] & \begin{array}{c} r \\ \\ \\ m-r \end{array} \\ & \begin{array}{cc} r & n-r \end{array} & \end{array} \quad \left[\begin{array}{c|c} V & V_0 \end{array}\right]^T$$

$$= U \Sigma V^T \quad \text{w/} \quad \Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_r} \end{bmatrix}$$

The vectors $\{u_i\}_{i=1}^r$ form a basis for the column space of A

Check: Let's consider the span of $\{u_i\}_{i=1}^r$; note that $\forall i \in [\![1,r]\!]$ $u_i \triangleq \frac{1}{\sqrt{\lambda_i}} A v_i \in \mathrm{col}(A)$ so that $\mathrm{span}\{u_i\}_{i=1}^r \subset \mathrm{col}(A)$

By dimension considerations $\dim(\mathrm{span}\{u_i\}_{i=1}^r) = r = \mathrm{rank}(A) = \dim(\mathrm{col}(A))$ so that $\mathrm{span}\{u_i\}_{i=1}^r = \mathrm{Col}(A)$

Let $\{u_i\}_{i=r+1}^m$ be an orthobasis of $\mathrm{Ker}(A^T)$, recall that $\mathrm{Ker}(A^T) = \mathrm{Col}(A)^\perp$ (check using definition + dimension)

Hence $\{u_i\}_{i=1}^m$ form an orthobasis for $\mathbb{R}^m$

Recall: $V \triangleq \begin{bmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{bmatrix}$ $U \triangleq \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix}$

Define $V_0 \triangleq \begin{bmatrix} | & & | \\ v_{r+1} & \cdots & v_n \\ | & & | \end{bmatrix}$ $U_0 = \begin{bmatrix} | & & | \\ u_{r+1} & \cdots & u_m \\ | & & | \end{bmatrix}$

Note $V^TV = I$ $V_0^TV_0 = I$ $V^TV_0 = 0$

$U^TU = I$ $U_0^TU_0 = I$ $U^TU_0 = 0$

Construct $V_1 = [\underbrace{V}_{r} \mid \underbrace{V_0}_{n-r}] \updownarrow n$ $U_1 = [\underbrace{U}_{r} \mid \underbrace{U_0}_{m-r}] \updownarrow m$ such that $U_1^TU_1 = U_1U_1^T = I$ and $V_1^TV_1 = V_1V_1^T = I$

Compute $(U_1^TAV_1)_{ij} \triangleq u_i^TAv_j = \frac{1}{\sqrt{\lambda_i}} v_i^TA^TAv_j = \sqrt{\lambda_j}\, v_i^Tv_j = \begin{cases} \sqrt{\lambda_i} & \text{if } i=j \\ 0 & \text{else} \end{cases}$ $(U_1^TAV_1 \in \mathbb{R}^{m\times n})$

Hence $U_1^TAV_1 = \underbrace{\left[\begin{array}{ccc|c} \sqrt{\lambda_1} & & 0 & \\ & \ddots & & 0 \\ 0 & & \sqrt{\lambda_r} & \\ \hline & 0 & & 0 \end{array}\right]}_{n} \updownarrow m \triangleq \Sigma_1$

Proof. Start with $A^TA \in \mathbb{R}^{n\times n}$, which is symmetric

By the spectral theorem, we have $A^TA = \sum_{i=1}^{n} \lambda_i v_i v_i^T$ $\{v_i\}_{i=1}^n$ orthobasis of eigenvectors

Note that $\operatorname{rank}(A) = \operatorname{rank}(A^TA) = r \leq n$ so that $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_r > 0 = \lambda_{r+1} \cdots \lambda_n$ (recall $\lambda_i \geq 0\ \forall i$)

Consider $\{v_i\}_{i=1}^r$; they form an orthobasis for the column space of (A^TA), hence of $\operatorname{row}(A)$

Consider $\{v_i\}_{i=r+1}^n$; they form an orthobasis for the Kernel (null space) of A

Check. ① Since $\lambda_i = 0$ $A^TAv_i = 0$ for $i \in [\![r+1; n]\!]$
Hence $v_i \in \operatorname{Ker}(A^TA) = \operatorname{Ker}(A)$; $\operatorname{Span}\{v_i\}_{i=r+1}^n \subset \operatorname{Ker}(A)$

② We have a span of dimension $n-r$ and $\dim(\operatorname{Ker}(A)) = n - \operatorname{rank}(A) = n-r$ hence $\operatorname{span}\{v_i\}_{i=r+1}^n =$

Set $u_i \triangleq \frac{1}{\sqrt{\lambda_i}} A v_i$ for $i \in [\![1; r]\!]$; form $U = \overbrace{\begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix}}^{r} \updownarrow m$, form $V = \overbrace{\begin{bmatrix} | & & | \\ v_1 & \cdots & v_r \\ | & & | \end{bmatrix}}^{r} \updownarrow n$

Note that $u_j^T u_i = \frac{1}{\sqrt{\lambda_j \lambda_i}} v_j^T A^T A v_i = \frac{1}{\sqrt{\lambda_j \lambda_i}} \lambda_i v_j^T v_i = \sqrt{\frac{\lambda_i}{\lambda_j}} v_j^T v_i = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{else} \end{cases}$

Hence $U^TU = I$

IMPORTANT PROPERTIES OF THE SVD

The columns of $\mathbf{V}$ $\{\mathbf{v}_i\}_{i=1}^r$ are eigenvectors of the psd matrix $\mathbf{A}^\top\mathbf{A}$. $\{\sigma_i : 1 \leq i \leq n \text{ and } \sigma_i \neq 0\}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}^\top\mathbf{A}$.

Note: Since $A = U\Sigma V^T$ $\quad A^TA = V\Sigma U^TU\Sigma V^T = V\Sigma^2V^T$

IMPORTANT PROPERTIES OF THE SVD

The columns of $\mathbf{V}$ $\{\mathbf{v}_i\}_{i=1}^r$ are eigenvectors of the psd matrix $\mathbf{A}^\intercal\mathbf{A}$. $\{\sigma_i : 1 \leq i \leq n \text{ and } \sigma_i \neq 0\}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}^\intercal\mathbf{A}$.

The columns of $\mathbf{U}$ $\{\mathbf{u}_i\}_{i=1}^r$ are eigenvectors of the psd matrix $\mathbf{A}\mathbf{A}^\intercal$. $\{\sigma_i : 1 \leq i \leq n \text{ and } \sigma_i \neq 0\}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}\mathbf{A}^\intercal$.

Note: $AA^T = U\Sigma V^T V \Sigma U^T = U\Sigma^2 U^T = \sum_{i=1}^{r} \sigma_i^2 u_i u_i^T$

IMPORTANT PROPERTIES OF THE SVD

The columns of $\mathbf{V}$ $\{\mathbf{v}_i\}_{i=1}^r$ are eigenvectors of the psd matrix $\mathbf{A}^\intercal\mathbf{A}$. $\{\sigma_i : 1 \leq i \leq n \text{ and } \sigma_i \neq 0\}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}^\intercal\mathbf{A}$.

The columns of $\mathbf{U}$ $\{\mathbf{u}_i\}_{i=1}^r$ are eigenvectors of the psd matrix $\mathbf{A}\mathbf{A}^\intercal$. $\{\sigma_i : 1 \leq i \leq n \text{ and } \sigma_i \neq 0\}$ are the square roots of the non-zero eigenvalues of $\mathbf{A}\mathbf{A}^\intercal$.

The columns of $\mathbf{V}$ form an orthobasis for $\operatorname{row}(\mathbf{A})$

The columns of $\mathbf{U}$ form an orthobasis for $\operatorname{col}(\mathbf{A})$

Equivalent form of the SVD: $\mathbf{A} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}^T$ where

- $\widetilde{\mathbf{U}} \in \mathbb{R}^{m\times m}$ is orthonormal
- $\widetilde{\mathbf{V}} \in \mathbb{R}^{n\times n}$ is orthonormal
- $\widetilde{\mathbf{\Sigma}} \in \mathbb{R}^{m\times n}$ is

$$\widetilde{\mathbf{\Sigma}} \triangleq \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

SVD AND LEAST-SQUARES

When we cannot solve $\mathbf{y} = \mathbf{A}\mathbf{x}$, we solve instead

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_2^2 \text{ such that } \mathbf{A}^\intercal\mathbf{A}\mathbf{x} = \mathbf{A}^\intercal\mathbf{y}$$

- This allows us to pick the minimum norm solution among potentially infinitely many solutions of the normal equations.

Recall: when $\mathbf{A} \in \mathbb{R}^{m\times n}$ is of rank n, then $\mathbf{x} = \mathbf{A}^\intercal(\mathbf{A}\mathbf{A}^\intercal)^{-1}\mathbf{y}$

Proposition (General solution) The solution of

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{x}\|_2^2 \text{ such that } \mathbf{A}^\intercal\mathbf{A}\mathbf{x} = \mathbf{A}^\intercal\mathbf{y}$$

is

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\intercal\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}\langle\mathbf{y}, \mathbf{u}_i\rangle\,\mathbf{v}_i$$

where $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the SVD of $\mathbf{A}$.

Proof. Let $x \in \mathbb{R}^n$; then $x = \underbrace{V\underline{\alpha}}_{\in \mathrm{row}(A)} + \underbrace{V_0 \underline{\alpha_0}}_{\in \mathrm{Ker}(A)}$ w/ V & V_0 defined earlier in SVD analysis

Hence $V^T x = V^T V \underline{\alpha} = \underline{\alpha}$ and $V_0^T x = V_0^T V_0 \underline{\alpha_0} = \underline{\alpha_0}$; $\|x\|_2^2 = \|V\underline{\alpha}\|_2^2 + \|V_0 \alpha_0\|_2^2 = \alpha^T V^T V \alpha + \alpha_0^T V_0^T V_0 \alpha_0$

$= \|\alpha\|_2^2 + \|\alpha_0\|_2^2$

Let $y \in \mathbb{R}^m$; then $y = \underbrace{U\beta}_{\in \mathrm{col}(A)} + \underbrace{U_0 \beta_0}_{\in \mathrm{Ker}(A^T)}$ w/ U and U_0 as before

Hence $U^T y = \underline{\beta}$ and $U_0^T y = \beta_0$, $\|y\|_2^2 = \|\underline{\beta}\|_2^2 + \|\underline{\beta_0}\|_2^2$

Then $y - Ax = U\underline{\beta} + U_0 \underline{\beta_0} - U\Sigma V^T(V\underline{\alpha} + V_0 \underline{\alpha_0}) = U\underline{\beta} + U_0\beta_0 - U\Sigma\alpha = \underline{U_0\beta_0} + U(\beta - \Sigma\alpha)$

$\|y - Ax\|_2^2 = \|\beta_0\|_2^2 + \|\beta - \Sigma\alpha\|_2^2$

PSEUDO INVERSE

$\mathbf{A}^{+} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\intercal}$ is called the *pseudo-inverse*, Lanczos inverse, or Moore-Penrose inverse of $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{T}$.

If $\mathbf{A}$ is square invertible then $\mathbf{A}^{+} = \mathbf{A}$

If $m \geq n$ (tall and skinny matrix) of rank n then $\mathbf{A}^{+} = (\mathbf{A}^{\intercal}\mathbf{A})^{-1}\mathbf{A}^{\intercal}$

If $m \geq m$ (short and fat matrix) of rank m then $\mathbf{A}^{+} = \mathbf{A}^{\intercal}(\mathbf{A}\mathbf{A}^{\intercal})^{-1}$

Note $\mathbf{A}^{+}$ is as “close” to an inverse of $\mathbf{A}$ as possible

STABILITY OF LEAST SQUARES

What if we observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$ and we apply the pseudo inverse? $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{y}$

We can separate the error analysis into two components

$$\hat{\mathbf{x}} - \mathbf{x}_0 = \underbrace{\mathbf{A}^+\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0}_{\text{null space error}} + \underbrace{\mathbf{A}^+\mathbf{e}}_{\text{noise error}}$$

$$= A^+y - x_0 = A^+(Ax_0 + e) - x_0$$

Note $A^+Ax_0 = V\Sigma^{-1}U^TU\Sigma V^Tx_0 = VV^Tx_0 \neq I x_0$ in general, which means that we don't reconstruct x_0 exactly

(when $\operatorname{rank}(A) < n$, i.e. $\operatorname{Ker}(A)$ non trivial)

STABILITY OF LEAST SQUARES

What if we observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$ and we apply the pseudo inverse? $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{y}$

We can separate the error analysis into two components

$$\hat{\mathbf{x}} - \mathbf{x}_0 = \underbrace{\mathbf{A}^+\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0}_{\text{null space error}} + \underbrace{\mathbf{A}^+\mathbf{e}}_{\text{noise error}}$$

We will express the error in terms of the SVD $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal$ With

- $\{\mathbf{v}_i\}_{i=1}^{r}$ orthobasis of $\mathrm{row}(\mathbf{A})$, augmented by $\{\mathbf{v}_i\}_{i=1}^{r+1} \in \ker \mathbf{A}$ to form an orthobasis of $\mathbb{R}^n$
- $\{\mathbf{u}_i\}_{i=1}^{r}$ orthobasis of $\mathrm{col}(\mathbf{A})$, augmented by $\{\mathbf{u}\}_{i=1}^{r+1} \in \ker \mathbf{A}^\intercal$ to form an orthobasis of $\mathbb{R}^m$

The null space error is given by

$$\left\|\mathbf{A}^+\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0\right\|_2^2 = \sum_{i=r+1}^{n} \left|\langle \mathbf{v}_i, \cancel{x}_0 \rangle\right|^2$$

Note:

$$x_0 = \sum_{i=1}^{n} \langle x_0, v_i \rangle v_i = \underbrace{\sum_{i=1}^{r} \langle x_0, v_i \rangle v_i}_{} + \boxed{\underbrace{\sum_{i=r+1}^{n} \langle x_0, v_i \rangle v_i}_{}}$$

$$\sum_{i=1}^{r} \langle x_0, v_i \rangle v_i = V V^T x_0 = \sum_{i=1}^{r} v_i v_i^T x_0 = A^+ A x_0$$

$$\sum_{i=r+1}^{n} \langle x_0, v_i \rangle v_i = V_0 V_0^T x_0 = \sum_{i=r+1}^{n} v_i v_i^T x_0$$

$$\| A^+ A x_0 - x_0 \|_2^2 = \sum_{i=r+1}^{n} | \langle x_0, v_i \rangle |^2$$

STABILITY OF LEAST SQUARES

What if we observe $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{e}$ and we apply the pseudo inverse? $\hat{\mathbf{x}} = \mathbf{A}^+\mathbf{y}$

We can separate the error analysis into two components

$$\hat{\mathbf{x}} - \mathbf{x}_0 = \underbrace{\mathbf{A}^+\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0}_{\text{null space error}} + \underbrace{\mathbf{A}^+\mathbf{e}}_{\text{noise error}}$$

We will express the error in terms of the SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$ With

- $\{\mathbf{v}_i\}_{i=1}^r$ orthobasis of $\operatorname{row}(\mathbf{A})$, augmented by $\{\mathbf{v}_i\}_{i=1}^{r+1} \in \ker \mathbf{A}$ to form an orthobasis of $\mathbb{R}^n$
- $\{\mathbf{u}_i\}_{i=1}^r$ orthobasis of $\operatorname{col}(\mathbf{A})$, augmented by $\{\mathbf{u}\}_{i=1}^{r+1} \in \ker \mathbf{A}^\intercal$ to form an orthobasis of $\mathbb{R}^m$

The null space error is given by

$$\left\|\mathbf{A}^+\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0\right\|_2^2 = \sum_{i=r+1}^{n} \left|\langle \mathbf{v}_i, x_0\rangle\right|^2$$

The noise error is given by

$$\left\|\mathbf{A}^+\mathbf{e}\right\|_2^2 = \sum_{i=1}^{r} \frac{1}{\sigma_i^2}\left|\langle \mathbf{e}, \mathbf{u}_i\rangle\right|^2$$

$$A^+ e = V\Sigma^{-1}U^T e = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T e = \sum_{i=1}^{r} \frac{1}{\sigma_i} \langle u_i, e\rangle v_i$$

$$\|A^+ e\|_2^2 = \sum_{i=1}^{r} \frac{1}{\sigma_i^2} |\langle u_i, e\rangle|^2$$

STABLE RECONSTRUCTION BY TRUNCATION

How do we mitigate the effect of small singular values in reconstruction?

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\intercal\mathbf{y} = \sum_{i=1}^{r} \frac{1}{\sigma_i}\langle \mathbf{y}, \mathbf{u}_i\rangle \mathbf{v}_i$$

Truncate the SVD to $r' < r$

$$\mathbf{A}_t \triangleq \sum_{i=1}^{r'} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal \qquad \mathbf{A}_t^+ = \sum_{i=1}^{r'} \frac{1}{\sigma_i} \mathbf{u}_i \mathbf{v}_i^\intercal$$

Reconstruct $\hat{\mathbf{x}}_t = \sum_{i=1}^{r'} \frac{1}{\sigma_i}\langle \mathbf{y}, \mathbf{u}_i\rangle \mathbf{v}_i = \mathbf{A}_t$

Error analysis:

$$\left\|\hat{\mathbf{x}}_t\right\|_2^2 = \sum_{i=r+1}^{n} \left|\langle \mathbf{x}_0, \mathbf{v}_i\rangle\right|^2 + \sum_{i=r'+1}^{r} \left|\langle \mathbf{x}_0, \mathbf{v}_i\rangle\right|^2 + \sum_{i=1}^{r\,'} \frac{1}{\sigma_i^2}\left|\langle \mathbf{e}, \mathbf{u}_i\rangle\right|^2$$

STABLE RECONSTRUCTION BY REGULARIZATION

Regularization means changing the problem to solve

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \qquad \lambda > 0$$

The solution is

$$\hat{\mathbf{x}} = (\mathbf{A}^\intercal\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^\intercal\mathbf{y} = \mathbf{V}(\mathbf{\Sigma}^2 + \lambda\mathbf{I})^{-1}\mathbf{\Sigma}\mathbf{U}^\intercal\mathbf{y}$$

NUMERICAL METHODS

We have seen several solutions to systems of linear equations $\mathbf{A}\mathbf{x} = \mathbf{y}$ so far

- $\mathbf{A}$ full column rank: $\hat{\mathbf{x}} = (\mathbf{A}^\intercal\mathbf{A})^{-1}\mathbf{A}^\intercal\mathbf{y}$
- $\mathbf{A}$ full row rank: $\hat{\mathbf{x}} = \mathbf{A}^\intercal(\mathbf{A}\mathbf{A}^\intercal)^{-1}\mathbf{y}$
- Ridge regression: $\hat{\mathbf{x}} = (\mathbf{A}^\intercal\mathbf{A} + \delta\mathbf{I})^{-1}\mathbf{A}^\intercal\mathbf{y}$
- Kernel regression: $\hat{\mathbf{x}} = (\mathbf{K} + \delta\mathbf{I})^{-1}\mathbf{y}$
- Ridge regression in Hilbert space: $\hat{\mathbf{x}} = (\mathbf{A}^\intercal\mathbf{A} + \delta\mathbf{G})^{-1}\mathbf{A}^\intercal\mathbf{y}$

Extension: constrained least-squares

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \text{ s.t. } \mathbf{x} = \mathbf{B}\boldsymbol{\alpha} \text{ for some } \boldsymbol{\alpha}$$

- The solution is $\hat{\mathbf{x}} = \mathbf{B}(\mathbf{B}^\intercal\mathbf{A}^\intercal\mathbf{A}\mathbf{B})^{-1}\mathbf{B}^\intercal\mathbf{A}^\intercal\mathbf{y}$

All these problems involve a symmetric positive definite system of equations.

- Many methods to achieve this based on matrix factorization

EASY SYSTEMS

Diagonal system

- $\mathbf{A} \in \mathbb{R}^{n \times n}$ invertible and diagonal
- $O(n)$ complexity

Orthogonal system

- $\mathbf{A} \in \mathbb{R}^{n \times n}$ invertible and orthogonal
- $O(n^2)$ complexity

Lower triangular system

- $\mathbf{A} \in \mathbb{R}^{n \times n}$ invertible and lower diagonal
- $O(n^2)$ complexity

General strategy: factorize $\mathbf{A}$ to recover some of the structures above