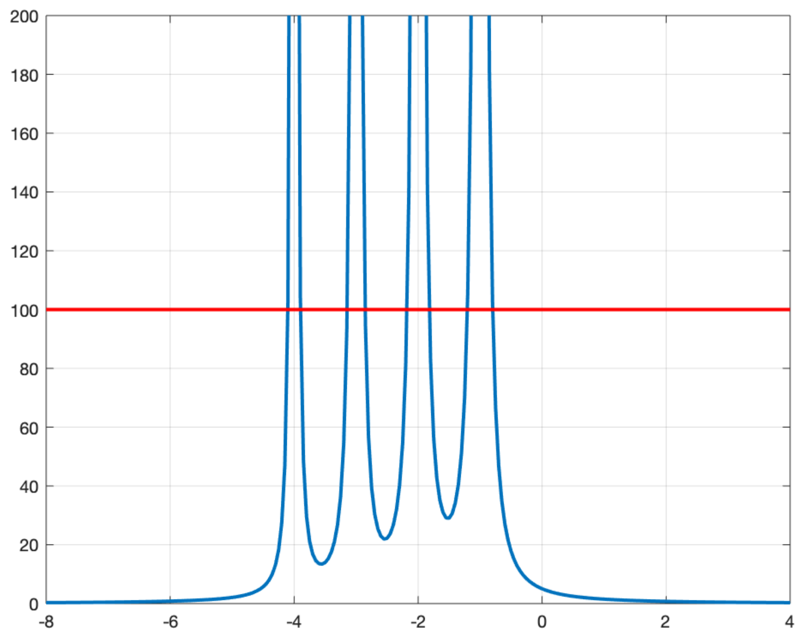
STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2023
LECTURE 6

1. LEAST SQUARES WITH NORM CONSTRAINTS

- let $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and α be some given positive number
- we wish to solve the problem
$$\begin{aligned} \text{minimize} \quad & \|A\mathbf{x} - \mathbf{b}\|_2 \\ \text{subject to} \quad & \|\mathbf{x}\|_2 \le \alpha \end{aligned} \tag{1}$$
- this problem is known as *least squares with quadratic constraints*
- arises in many situations:
  - ridge regression
  - Tychonov regularization
  - generalized cross-validation (GCV)
- note that if $\alpha \ge \|A^\dagger \mathbf{b}\|_2$, the unconstrained minimum norm solution $A^\dagger\mathbf{b}$ would already be a solution
- so for a non-trivial solution, we assume that $\alpha < \|A^\dagger \mathbf{b}\|_2$ and in which case the solution $\mathbf{x}$ to (1) must sit on the boundary of the ball of radius α, i.e., $\|\mathbf{x}\|_2 = \alpha$, and we have
$$\begin{aligned} \text{minimize} \quad & \|A\mathbf{x} - \mathbf{b}\|_2 \\ \text{subject to} \quad & \|\mathbf{x}\|_2 = \alpha \end{aligned} \tag{2}$$
- prove the statement above as an exercise, i.e., when $\alpha < \|A^\dagger \mathbf{b}\|_2$, then (1) and (2) have the same solution
- to solve this problem, we define the *Lagrangian*
$$L(\mathbf{x}, \mu) = \|A\mathbf{x} - \mathbf{b}\|_2^2 + \mu(\|\mathbf{x}\|^2 - \alpha^2)$$
where μ is called the *Lagrange multiplier*
- first-order condition for minimality: set derivative to zero
$$\mathbf{0} = \nabla_{\mathbf{x}} L(\mathbf{x}, \mu) = -2A^{\mathsf{T}}\mathbf{b} + 2A^{\mathsf{T}}A\mathbf{x} + 2\mu\mathbf{x}$$
$$0 = \frac{\partial}{\partial \mu} L(\mathbf{x}, \mu) = \|\mathbf{x}\|^2 - \alpha^2$$
- we obtain
$$\begin{cases} (A^{\mathsf{T}}A + \mu I)\mathbf{x} = A^{\mathsf{T}}\mathbf{b}, \\ \|\mathbf{x}\|^2 = \alpha^2 \end{cases} \tag{3}$$
- note that if $(\mathbf{x}_1, \mu_1)$ and $(\mathbf{x}_2, \mu_2) \in \mathbb{R}^{n+1}$ are both solutions to (3), then
$$\begin{aligned} (A^{\mathsf{T}}A + \mu_1 I)\mathbf{x}_1 &= A^{\mathsf{T}}\mathbf{b}, \\ (A^{\mathsf{T}}A + \mu_2 I)\mathbf{x}_2 &= A^{\mathsf{T}}\mathbf{b} \end{aligned} \tag{4}$$
and $\|\mathbf{x}_1\|^2 = \|\mathbf{x}_2\|^2 = \alpha^2$
- left multiply the first equation in (4) by $\mathbf{x}_1^{\mathsf{T}}$ and the second equation by $\mathbf{x}_2^{\mathsf{T}}$ and subtract, we get
$$\|A\mathbf{x}_2\|_2^2 - \|A\mathbf{x}_1\|_2^2 - \mathbf{b}^{\mathsf{T}}A(\mathbf{x}_2 - \mathbf{x}_1) = \mu_1\|\mathbf{x}_1\|^2 - \mu_2\|\mathbf{x}_2\|^2 \tag{5}$$

Date: October 16, 2023, version 1.0. Comments, bug reports: `lekheng@uchicago.edu`.

- left multiply the first equation in (4) by $\mathbf{x}_2^\mathsf{T}$ and the second equation by $\mathbf{x}_1^\mathsf{T}$ and subtract, we get
$$-\mathbf{b}^\mathsf{T} A(\mathbf{x}_2 - \mathbf{x}_1) = -(\mu_1 - \mu_2)\mathbf{x}_1^\mathsf{T}\mathbf{x}_2 \tag{6}$$
- adding (5) and (6), and noting that $\|\mathbf{x}_1\|^2 = \|\mathbf{x}_2\|^2$, we get
$$\|A\mathbf{x}_2 - \mathbf{b}\|_2^2 - \|A\mathbf{x}_1 - \mathbf{b}\|_2^2 = \frac{\mu_1 - \mu_2}{2}\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2$$
- hence
$$\mu_1 > \mu_2 \quad \Longrightarrow \quad \|A\mathbf{x}_2 - \mathbf{b}\|_2 > \|A\mathbf{x}_1 - \mathbf{b}\|_2$$
- the solution $\mathbf{x}$ to (2) is given by the solution $(\mathbf{x}, \mu) \in \mathbb{R}^{n+1}$ to (3) with the largest μ
- to solve (3), we see that we need to compute
$$\mathbf{x} = (A^\mathsf{T} A + \mu I)^{-1} A^\mathsf{T}\mathbf{b} \tag{7}$$
where
$$\mathbf{x}^\mathsf{T}\mathbf{x} = \mathbf{b}^\mathsf{T} A(A^\mathsf{T} A + \mu I)^{-2} A^\mathsf{T}\mathbf{b} = \alpha^2$$
- we want to avoid actually forming $A^\mathsf{T} A$ for reasons that will be clear when we discuss normal equations later and to do this we use SVD
- if $A = U\Sigma V^\mathsf{T}$ is the full SVD of A, we let $\mathbf{c} = U^\mathsf{T}\mathbf{b}$, then we have
$$\begin{aligned}
\alpha^2 &= \mathbf{b}^\mathsf{T} U\Sigma V^\mathsf{T}(V\Sigma^\mathsf{T}\Sigma V^\mathsf{T} + \mu I)^{-2} V\Sigma^\mathsf{T} U^\mathsf{T}\mathbf{b} \\
&= \mathbf{c}^\mathsf{T}\Sigma[(V\Sigma^\mathsf{T}\Sigma V^\mathsf{T} + \mu I)V]^{-1}[V^\mathsf{T}(V\Sigma^\mathsf{T}\Sigma V^\mathsf{T} + \mu I)]^{-1}\Sigma^\mathsf{T}\mathbf{c} \\
&= \mathbf{c}^\mathsf{T}\Sigma(V\Sigma^\mathsf{T}\Sigma + \mu V)^{-1}(\Sigma^\mathsf{T}\Sigma V^\mathsf{T} + \mu V^\mathsf{T})^{-1}\Sigma^\mathsf{T}\mathbf{c} \\
&= \mathbf{c}^\mathsf{T}\Sigma[(\Sigma^\mathsf{T}\Sigma V^\mathsf{T} + \mu V^\mathsf{T})(V\Sigma^\mathsf{T}\Sigma + \mu V)]^{-1}\Sigma^\mathsf{T}\mathbf{c} \\
&= \mathbf{c}^\mathsf{T}\Sigma(\Sigma^\mathsf{T}\Sigma + \mu I)^{-2}\Sigma^\mathsf{T}\mathbf{c} \\
&= \sum_{i=1}^{r} \frac{c_i^2\sigma_i^2}{(\sigma_i^2 + \mu)^2} \\
&=: f(\mu)
\end{aligned}$$
where $\mathbf{c} = (c_1, \dots, c_m)^\mathsf{T}$
- we find all solutions of $f(\mu) = \alpha^2$, a univariate nonlinear equation, using any standard solvers, e.g., Newton–Raphson method, to get all roots μ and pick the largest one μ_*
- note the function f has poles at $-\sigma_i^2$ for $i = 1, \dots, r$, i.e., f blows up to $+\infty$ at these points and they can never be solutions of $f(\mu) = \alpha^2$
- furthermore, $\lim_{\mu\to\pm\infty} f(\mu) = 0$ so there can only be a finite number of solutions — see Figure 1 for an example
- to summarize, the algorithm for solving this problem, given A, $\mathbf{b}$, and α^2, is as follows:
  - step 1: compute SVD of A to obtain $A = U\Sigma V^\mathsf{T}$
  - step 2: compute $\mathbf{c} = U^\mathsf{T}\mathbf{b}$
  - step 3: solve $f(\mu_*) = \alpha^2$ with Newton–Raphson method and select largest μ_*
  - step 4: use the SVD to compute
$$\mathbf{x} = (A^\mathsf{T} A + \mu I)^{-1} A^\mathsf{T}\mathbf{b} = V(\Sigma^\mathsf{T}\Sigma + \mu I)^{-1}\Sigma^\mathsf{T} U^\mathsf{T}\mathbf{b}$$
- this is an example of an 'near closed form' solution: we have an analytic expression for $\mathbf{x}$ that depends on just one unknown parameter μ_*, which is the root of a univariate nonlinear equation
- as a toy example, Figure 1 shows the intersections of the blue curve $y = f(x)$ where
$$f(x) = \frac{4}{(1+x)^2} + \frac{3}{(2+x)^2} + \frac{2}{(3+x)^2} + \frac{1}{(4+x)^2}$$



and the red horizontal line $y = 10^2$, the right most intersection point of the red line and blue curve gives us μ_*

- a better alternative is that instead of using Newton–Raphson method on $f(\mu) = \alpha^2$ directly, we could instead solve $1/f(\mu) = 1/\alpha^2$, which is a much better strategy
- as an aside that will make sense after we discuss condition number in the future, note that if we denote the eigenvalues of $A^{\mathsf{T}}A$ by

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

- then eigenvalues of $A^{\mathsf{T}}A + \mu I$ are

$$\lambda_1 + \mu, \dots, \lambda_n + \mu$$

- if $\mu \geq 0$, then $\kappa_2(A^{\mathsf{T}}A + \mu I) \leq \kappa_2(A^{\mathsf{T}}A)$, because

$$\frac{\lambda_1 + \mu}{\lambda_n + \mu} \leq \frac{\lambda_1}{\lambda_n}$$

- so $A^{\mathsf{T}}A + \mu I$ is better conditioned

2. MATRIX APPROXIMATION PROBLEMS

- we next examine problems like

$$\min_{X^{\mathsf{T}}=X} \|A - X\|_{\mathsf{F}}, \qquad \min_{X^{\mathsf{T}}X=I} \|A - X\|_{\mathsf{F}}, \qquad \min_{\operatorname{rank}(X)\leq r} \|A - X\|_{\mathsf{F}}$$

for a given $A \in \mathbb{R}^{n\times n}$

- the second and third problems are great applications of SVD

3. CLOSEST HERMITIAN/SYMMETRIC MATRIX

- this one doesn't require SVD but is interesting nonetheless

- given $A \in \mathbb{C}^{n\times n}$, find its closest Hermitian matrix
$$\min_{X^*=X} \|A - X\|_{\mathsf{F}} \tag{8}$$
or its closest skew-Hermitian matrix
$$\min_{X^*=-X} \|A - X\|_{\mathsf{F}} \tag{9}$$
- note that any square matrix can be written as a sum of a Hermitian matrix and a skew-Hermitian matrix
$$A = \frac{1}{2}(A + A^*) + \frac{1}{2}(A - A^*)$$
- the solutions to (8) and (9) are given by $X = \frac{1}{2}(A + A^*)$ and $X = \frac{1}{2}(A - A^*)$ respectively (why?)
- for $A \in \mathbb{R}^{n\times n}$ these yield the closest symmetric and skew-symmetric matrices to A
- a more general problem is that given $A, B \in \mathbb{C}^{n\times n}$, we want X such that
$$\min_{X^*=X} \|A - BX\|_{\mathsf{F}}$$
and you will be guided through its proof in Homework 2

4. closest unitary/orthogonal matrix

- let $\mathrm{U}(n)$ be the set of all $n \times n$ unitary matrices
- given $A \in \mathbb{C}^{n\times n}$, we wish to find the matrix $X \in \mathrm{U}(n)$ that satisfies
$$\min_{X\in \mathrm{U}(n)} \|A - X\|_{\mathsf{F}}$$
- first note that
$$\|A - X\|_{\mathsf{F}}^2 = \operatorname{tr}(A^*A) + \operatorname{tr}(X^*X) - 2\operatorname{Re}\operatorname{tr}(A^*X) = \|A\|_{\mathsf{F}}^2 + n - 2\operatorname{Re}\operatorname{tr}(A^*X)$$
since $X^*X = I$
- so minimizing $\|A - X\|_{\mathsf{F}}^2$ is equivalent to maximizing $\operatorname{Re}\operatorname{tr}(A^*X)$
- let $A = U\Sigma V^*$ be the SVD of A
- then writing $Z = U^*XV$, we get
$$\operatorname{Re}\operatorname{tr}(A^*X) = \operatorname{Re}\operatorname{tr}(V\Sigma^*U^*X) = \operatorname{Re}\operatorname{tr}(\Sigma^*U^*XV) = \operatorname{Re}\operatorname{tr}(\Sigma^*Z) = \sum_{i=1}^{p} \sigma_i \operatorname{Re}(z_{ii}) \le \sum_{i=1}^{p} \sigma_i$$
where the last inequality follows since Z is an orthogonal matrix and so $\operatorname{Re}(z_{ii}) \le |z_{ii}| \le 1$
- the upper bound is attained by making $Z = I$, i.e.,
$$X = UV^*$$
- exercise: show that $\|\cdot\|_2$ and $\|\cdot\|_{\mathsf{F}}$ are unitarily invariant, i.e., satisfies
$$\|UXV\| = \|X\|$$
for all $X \in \mathbb{C}^{m\times n}$ whenever U and V are unitary matrices
- if we set
$$X = UV^*,$$
then
$$\|A - X\|_{\mathsf{F}}^2 = \|U(\Sigma - I)V^*\|_{\mathsf{F}}^2 = \|\Sigma - I\|_{\mathsf{F}}^2 = (\sigma_1 - 1)^2 + \cdots + (\sigma_n - 1)^2$$
- for real matrices A, one could also ask for
$$\min_{X\in \mathrm{O}(n)} \|A - X\|_{\mathsf{F}}$$
which is just a special case

- a more general problem is that given $A, B \in \mathbb{C}^{n\times n}$, we want $X \in \mathrm{U}(n)$ such that
$$\min_{X\in\mathrm{U}(n)} \|A - BX\|_{\mathsf{F}}$$
and you asked to do this in Homework 2

5. Closest rank-r matrix

- given $A \in \mathbb{C}^{m\times n}$, we want to find $X \in \mathbb{C}^{m\times n}$ of rank not more than r so that $\|A - X\|$ is minimized
- in notations, we want
$$\min_{\operatorname{rank}(X)\le r} \|A - X\| \tag{10}$$
- such an X is called a best rank-r approximation to A or a rank-r projection of A
- if $r \ge \operatorname{rank}(A)$, then clearly $X = A$ and the problem is trivial
- so we shall always assume that $r < \operatorname{rank}(A)$
- we will start with the classical case where $\|\cdot\|$ is the matrix 2-norm or spectral norm

Theorem 1 (Eckart–Young). *Let the* SVD *of A be*
$$A = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_i \mathbf{u}_i \mathbf{v}_i^*, \quad \sigma_1 \ge \cdots \ge \sigma_{\operatorname{rank}(A)} > 0.$$
Then for any $r \in \{1, \dots, \operatorname{rank}(A) - 1\}$, a solution to (10) *when $\|\cdot\| = \|\cdot\|_2$ is given by*
$$X = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*. \tag{11}$$
Furthermore, we have
$$\min_{\operatorname{rank}(X)\le r} \|A - X\|_2 = \sigma_{r+1}. \tag{12}$$
In matrix form, we have
$$X = U \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix} V^*, \tag{13}$$
where $A = U\Sigma V^*$ is the SVD *of A.*

Proof. Suppose there is a $B \in \mathbb{C}^{m\times n}$ with $\operatorname{rank}(B) \le r$ and $\|A - B\|_2 < \sigma_{r+1}$. Then by the rank-nullity theorem
$$\operatorname{rank}(B) + \dim(\ker(B)) = n$$
and so
$$\dim(\ker(B)) \ge n - r.$$
Let $\mathbf{w} \in \ker(B)$. Then $B\mathbf{w} = \mathbf{0}$ and so
$$\|A\mathbf{w}\|_2 = \|(A - B)\mathbf{w}\|_2 \le \|A - B\|_2 \|\mathbf{w}\|_2 < \sigma_{r+1}\|\mathbf{w}\|_2. \tag{14}$$
Let $\mathbf{w} \in W := \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_{r+1}\}$. Then $\mathbf{w} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_{r+1}\mathbf{v}_{r+1}$. Rewriting this in matrix form
$$\mathbf{w} = V_{r+1} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{r+1} \end{bmatrix} = V_{r+1}\boldsymbol{\alpha}$$

where $V_{r+1} = [\mathbf{v}_1, \dots, \mathbf{v}_{r+1}] \in \mathbb{C}^{n \times r}$, i.e., the first $r+1$ columns of V.

$$\|A\mathbf{w}\|_2^2 = \|U\Sigma V^* V_{r+1}\boldsymbol{\alpha}\|_2^2 = \left\|\Sigma \begin{bmatrix} I_{r+1} \\ O \end{bmatrix} \boldsymbol{\alpha}\right\|_2^2 = \left\|\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_{r+1} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{r+1} \end{bmatrix}\right\|_2^2$$

$$= \sum_{i=1}^{r+1} \sigma_i^2 |\alpha_i|^2 \geq \sigma_{r+1}^2 \sum_{i=1}^{r+1} |\alpha_i|^2 = \sigma_{r+1}^2 \|\mathbf{w}\|_2^2.$$

Hence if $\mathbf{w} \in W$, then

$$\|A\mathbf{w}\|_2 \geq \sigma_{r+1}\|\mathbf{w}\|_2. \tag{15}$$

But since $\dim(\ker(B)) \geq n - r$ and $\dim(W) = r + 1$, the two subspaces must intersect nontrivially, i.e., $\dim(\ker(B) \cap W)) \geq 1$ and so there exists a non-zero vector $\mathbf{w} \in \ker(B) \cap W$. Such a vector would satisfy both (14) and (15), a contradiction. Hence our original assumption is false: There is no rank-r matrix B that could beat the bound in (12). On the other hand it is easy to verify that the choice of X in (13) satisfies (12):

$$\|A - X\|_2 = \left\|U \begin{bmatrix} 0 & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & & & & & \\ & & & \sigma_{r+1} & & & & \\ & & & & \ddots & & & \\ & & & & & \sigma_{\operatorname{rank}(A)} & & \\ & & & & & & 0 & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix} V^*\right\|_2 = \sigma_{r+1}.$$

$\square$

- the generalization of Eckart–Young theoem to any arbitrary unitarily invariant norm is due to Mirsky and this theorem is sometimes also called the Eckart–Young–Mirsky theorem
- note that the general theorem only says that (11) is the best rank-r approximation of A, the value in (12) would in general be different
- for example if we use the Frobenius norm
$$\min_{\operatorname{rank}(X) \leq r} \|A - X\|_{\mathsf{F}} = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_{\operatorname{rank}(A)}^2}.$$
- a more general problem is to find $X \in \mathbb{C}^{n \times p}$ such that
$$\min_{\operatorname{rank}(X) \leq r} \|A - BX\|_{\mathsf{F}}$$
for given $A \in \mathbb{C}^{m \times p}$ and $B \in \mathbb{C}^{m \times n}$

6. Solving total least squares problems

- assume $A \in \mathbb{C}^{m \times n}$ has full column rank, i.e., $\operatorname{rank}(A) = n \leq m$
- in ordinary least squares problem, we solve
$$A\mathbf{x} = \mathbf{b} + \mathbf{r}, \quad \|\mathbf{r}\|_2 = \min$$

- in *total least squares* problem, we wish to solve

$$(A+E)\mathbf{x} = \mathbf{b} + \mathbf{r}, \quad \|E\|_{\mathsf{F}}^2 + \lambda^2 \|\mathbf{r}\|_2^2 = \min$$

- note that if $\mathbf{b} \in \operatorname{im}(A)$, then the solution is given by setting $E = O$, $\mathbf{r} = \mathbf{0}$ and choosing $\mathbf{x}$ to be any solution of $A\mathbf{x} = \mathbf{b}$
- so assume $\mathbf{b} \notin \operatorname{im}(A)$ and therefore

$$\operatorname{rank}([A, \mathbf{b}]) = n + 1$$

- from $A\mathbf{x} - \mathbf{b} + E\mathbf{x} - \mathbf{r} = \mathbf{0}$ we obtain the system

$$\begin{bmatrix} A & \mathbf{b} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} + \begin{bmatrix} E & \mathbf{r} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} = \mathbf{0}$$

  or

$$(C + F)\mathbf{z} = \mathbf{0} \tag{16}$$

- since

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} \neq \mathbf{0} \tag{17}$$

  we must have $\operatorname{nullity}(C + F) \geq 1$
- and since $\operatorname{rank}(C) = n + 1$, we must have

$$\operatorname{rank}(C + F) \leq n$$

- we need the matrix $C + F$ to have $\operatorname{rank}(C + F) \leq n$, and we want to minimize $\|F\|_{\mathsf{F}}$
- to solve this problem, we compute the SVD of $C \in \mathbb{C}^{m \times (n+1)}$

$$C = \begin{bmatrix} A & \mathbf{b} \end{bmatrix} = U\Sigma V^* = U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ & & & \sigma_{n+1} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} V^*$$

- note that $\sigma_{n+1} > 0$ since $\operatorname{rank}(C) = n + 1$
- we want F so that $\operatorname{rank}(C + F) \leq n$ so need to zero out σ_{n+1}, i.e., we want

$$C + F = U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ & & & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} V^* \tag{18}$$

  or to be more precise, we want

$$\min_{\operatorname{rank}(C+F) \leq n} \|F\|_{\mathsf{F}} = \min_{\operatorname{rank}(C+F) \leq n} \|C - (C + F)\|_{\mathsf{F}} = \min_{\operatorname{rank}(X) \leq n} \|C - X\|_{\mathsf{F}}$$

  and Eckhart–Young theorem tells us that $X \coloneqq C + F$ must take the form in (18)

- so pick

$$F = U \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & -\sigma_{n+1} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} V^*$$

and note that this F would produce the effect needed for (18)
- let $V = [\mathbf{v}_1, \dots, \mathbf{v}_{n+1}] \in \mathbb{C}^{(n+1)\times(n+1)}$ where $\mathbf{v}_i \in \mathbb{C}^{n+1}$ is the ith column of V note that $\mathbf{v}_i^* \mathbf{v}_{n+1} = 0$ for all $i = 1, \dots, n$
- we have

$$(C+F)\mathbf{v}_{n+1} = U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_n & \\ & & & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} V^* \mathbf{v}_{n+1} = U \begin{bmatrix} \sigma_1 \mathbf{v}_1^* \\ \vdots \\ \sigma_n \mathbf{v}_n^* \\ \mathbf{0}^\mathsf{T} \\ \mathbf{0}^\mathsf{T} \\ \vdots \\ \mathbf{0}^\mathsf{T} \end{bmatrix} \mathbf{v}_{n+1} = U \begin{bmatrix} \sigma_1 \mathbf{v}_1^* \mathbf{v}_{n+1} \\ \vdots \\ \sigma_n \mathbf{v}_n^* \mathbf{v}_{n+1} \\ \mathbf{0}^\mathsf{T} \mathbf{v}_{n+1} \\ \mathbf{0}^\mathsf{T} \mathbf{v}_{n+1} \\ \vdots \\ \mathbf{0}^\mathsf{T} \mathbf{v}_{n+1} \end{bmatrix} = U \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

- so the vector $\mathbf{v}_{n+1}$ ought to be a candidate for the solution $\mathbf{z}$ in (16) but there is one caveat — the last coordinate of $\mathbf{z}$ must be -1 by (17)
- how do we achieve that? we divide $\mathbf{v}_{n+1}$ by the negative of its last coordinate, so

$$\begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} = \mathbf{z} = -\frac{1}{v_{n+1,n+1}} \mathbf{v}_{n+1}$$

provided that $v_{n+1,n+1} \neq 0$
- this gives the solution

$$\mathbf{x} = -\begin{bmatrix} v_{1,n+1}/v_{n+1,n+1} \\ \vdots \\ v_{n,n+1}/v_{n+1,n+1} \end{bmatrix}$$

where the v_{ij} refers to the entries of $V = [v_{ij}]_{i,j=1}^{n+1}$

7. Other applications

- next lecture we will also look at SVD as a notion of "numerical rank"
- in the homework you see yet other uses of SVD
- here are some other uses of SVD that we didn't have time to consider:
  - least squares with linear constraints (we will discuss this under QR though)
  - least squares with quadratic constraints
  - finding angles between subspaces
  - orthonormal basis for intersection of subspaces
  - subset selection
- all these should convince you that SVD truly is a swiss army knife of matrix computations