MATH 661.FA23 Midterm Examination 1 - Solution

Solve the problems for your appropriate course track. Problems probe understanding of the definitions and results from the module on floating point arithmetic and linear algebra. Formulate your answers clearly, cogently, and include a concise description of your approach. Each question is meant to be completely answered within ten minutes. Allowed test time is 75 minutes.

1 Common problems

1. Matrix $\boldsymbol{A}\in\mathbb{R}^{m\times m}$, $\operatorname{rank}(\boldsymbol{A})=m$, has the singular value decomposition (SVD) $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ ($\boldsymbol{U}\boldsymbol{U}^T=\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}$, $\boldsymbol{V}\boldsymbol{V}^T=\boldsymbol{V}^T\boldsymbol{V}=\boldsymbol{I}$, $\boldsymbol{\Sigma}=\operatorname{diag}(\sigma_1,..,\sigma_m)$, $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_m>0$) and the pseudoinverse $\boldsymbol{A}^+=\boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$, $\boldsymbol{\Sigma}^+=\operatorname{diag}(1/\sigma_1,..,1/\sigma_m)$. Find the SVDs of:

    a) $\boldsymbol{A}^T\boldsymbol{A}$;

    b) $(\boldsymbol{A}^T\boldsymbol{A})^+$;

    c) $(\boldsymbol{A}^T\boldsymbol{A})^+\boldsymbol{A}^T$;

    d) $\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^+$;

    e) $\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^+\boldsymbol{A}^T$.

    Solution.

    a) $\boldsymbol{A}^T\boldsymbol{A}=(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T$, since $\boldsymbol{\Sigma}^T=\boldsymbol{\Sigma}$. The SVD is

    $$\boldsymbol{A}^T\boldsymbol{A}=\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T$$

    and the singular values of $\boldsymbol{A}^T\boldsymbol{A}$ are the squares of those of $\boldsymbol{A}$.

    b) $(\boldsymbol{A}^T\boldsymbol{A})^+=(\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T)^+=\boldsymbol{V}(\boldsymbol{\Sigma}^2)^+\boldsymbol{V}^T$. This is not an SVD since

    $$(\boldsymbol{\Sigma}^2)^+=\operatorname{diag}(1/\sigma_1^2,..,1/\sigma_m^2), 1/\sigma_1^2\leqslant 1/\sigma_2^2\leqslant\cdots\leqslant 1/\sigma_m^2.$$

    Introduce permutation matrix $\boldsymbol{P}=[\ \boldsymbol{e}_m\ \ \boldsymbol{e}_{m-1}\ \ ...\ \ \boldsymbol{e}_1\ ]$ which is orthogonal, $\boldsymbol{P}\boldsymbol{P}^T=\boldsymbol{I}$ and symmetric $\boldsymbol{P}=\boldsymbol{P}^T$ to obtain

    $$(\boldsymbol{\Sigma}^2)^+\boldsymbol{P}=\operatorname{diag}(1/\sigma_m^2,..,1/\sigma_1^2)=\boldsymbol{\Lambda},$$

    the correct ordering and the SVD

    $$(\boldsymbol{A}^T\boldsymbol{A})^+=\boldsymbol{V}(\boldsymbol{\Sigma}^2)^+\boldsymbol{V}^T=\boldsymbol{V}(\boldsymbol{\Sigma}^2)^+\boldsymbol{P}\boldsymbol{P}^T\boldsymbol{V}^T=\boldsymbol{V}\boldsymbol{\Lambda}(\boldsymbol{V}\boldsymbol{P})^T,$$

    since $\boldsymbol{V}\boldsymbol{P}$, the product of two orthogonal matrices, is itself orthogonal.

    c) Calculate

    $$(\boldsymbol{A}^T\boldsymbol{A})^+\boldsymbol{A}^T=\boldsymbol{V}\boldsymbol{\Lambda}(\boldsymbol{V}\boldsymbol{P})^T\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T=\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{U}^T.$$

Since

$$\boldsymbol{P}^T\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_m, .., \sigma_1),$$

obtain

$$\boldsymbol{\Lambda}\boldsymbol{P}^T\boldsymbol{\Sigma} = \operatorname{diag}(1/\sigma_m, .., 1/\sigma_1) = \boldsymbol{\Gamma},$$

and the SVD

$$(\boldsymbol{A}^T\boldsymbol{A})^+\boldsymbol{A}^T = \boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{U}^T.$$

d) Calculate

$$\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^+ = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Lambda}(\boldsymbol{V}\boldsymbol{P})^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Lambda}(\boldsymbol{V}\boldsymbol{P})^T = \boldsymbol{U}\boldsymbol{\Gamma}(\boldsymbol{V}\boldsymbol{P})^T,$$

which is an SVD.

e) Use above and $\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{I}$ to write

$$\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^+\boldsymbol{A}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{U}^T = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{U}^T = \boldsymbol{I}.$$

2 Track 1

1. Write pseudo-code to accurately evaluate the sum

$$S_{2n} = \sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} x^k$$

in floating point arithmetic when $x = 1 + \varepsilon$, $1 \gg \varepsilon > 0$. ($\lim_{n\to\infty} S_{2n} = \ln(1+x)$).

Solution. There is possible loss of precision from successive terms of alternating signs, the effect of which can be attenuated by adding two terms at a time to the sum accumulator

$$S_{2n} = \sum_{k=1}^{n} x^{2k-1}\left(\frac{1}{2k-1} - \frac{x}{2k}\right)$$

$S = 0;\ y = x;\ x2 = x^2$
for k=1 to n
  $l = 2k; d = 1/(l-1) - x/l$
  $t = y \cdot l;\ S = S + t;\ y = y \cdot x2$

2. Use the SVD of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ to express the Moore-Penrose pseudoinverse as a sum of rank-one matrices.

Solution. The SVD of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ and the pseudoinverse is written as

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T = [\ \boldsymbol{v}_1\ \ ..\ \ \boldsymbol{v}_n\ ]\begin{bmatrix} 1/\sigma_1 & & & \\ & \ddots & & \\ & & 1/\sigma_r & \\ & & & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} = \sum_{j=1}^{r} \frac{1}{\sigma_j}\boldsymbol{v}_j\boldsymbol{u}_j^T,$$

a sum of the rank-1 updates $\boldsymbol{v}_j\boldsymbol{u}_j^T/\sigma_j$.

3 Track 2

1. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Show that the Moore-Penrose pseudoinverse $\boldsymbol{X} = \boldsymbol{A}^{+}$ minimizes $\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{I}\|_F$ over all n by m matrices.

   Solution. The squared Frobenius norm of $\boldsymbol{A}\boldsymbol{X} - \boldsymbol{I} = [\ \boldsymbol{A}\boldsymbol{x}_1 - \boldsymbol{e}_1 \ \ .... \ \ \boldsymbol{A}\boldsymbol{x}_n - \boldsymbol{e}_n\ ]$ is the sum of its squared column vector 2-norms

$$\|\boldsymbol{A}\boldsymbol{X} - \boldsymbol{I}\|_F^2 = \sum_{j=1}^{n} \|\boldsymbol{A}\boldsymbol{x}_j - \boldsymbol{e}_j\|_2^2,$$

   and the minimum is attained by the solution of the n least squares problems

$$\min_{\boldsymbol{x}_j} \|\boldsymbol{A}\boldsymbol{x}_j - \boldsymbol{e}_j\|_2 \Rightarrow \boldsymbol{x}_j = \boldsymbol{V}\boldsymbol{\Sigma}^{+}\boldsymbol{U}^T\boldsymbol{e}_j \Rightarrow \boldsymbol{X} = \boldsymbol{A}^{+}\boldsymbol{I} = \boldsymbol{A}^{+}.$$

2. Let $\boldsymbol{A} \in \mathbb{C}^{m \times m}$ be skew-Hermitian, i.e., $\boldsymbol{A}^* = -\boldsymbol{A}$. Prove that:

   a) $\boldsymbol{I} - \boldsymbol{A}$ is nonsingular;

   b) $\boldsymbol{C} = (\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} + \boldsymbol{A})$ is unitary.

   Solution. a) For $m = 1$, $a + a^* = 0$, and $b = 1 - a$ nonsingular implies $\|b\|_2^2 = b^* b > 0$, readily verified

$$b^* b = (1 - a^*)(1 - a) = 1 + a^* a > 0.$$

   This also holds for $m > 1$, $\boldsymbol{A} + \boldsymbol{A}^* = 0$, and $\boldsymbol{B} = \boldsymbol{I} - \boldsymbol{A}$ nonsingular implies $\|\boldsymbol{B}\boldsymbol{x}\|_2^2 = \|\boldsymbol{y}\|_2^2 > 0$ for any $\boldsymbol{x}$ of unit norm. Compute

$$\boldsymbol{y}^*\boldsymbol{y} = \boldsymbol{x}^*\boldsymbol{B}^*\boldsymbol{B}\boldsymbol{x} = \boldsymbol{x}^*(\boldsymbol{I} + \boldsymbol{A}^*\boldsymbol{A})\,\boldsymbol{x} = 1 + \|\boldsymbol{y}\|_2^2 > 0.$$

   b) Again, use $m = 1$ to gain insight in which case $c = (1 - a)^{-1}(1 + a)$ and compute

$$c^* c = (1 + a^*)(1 - a^*)^{-1}(1 - a)^{-1}(1 + a) = (1 - a)[(1 - a)(1 + a)]^{-1}(1 + a) = (1 - a)(1 - a^2)^{-1}(1 + a)$$

$$c^* c = (1 - a)[(1 + a)(1 - a)]^{-1}(1 + a) = (1 - a)(1 - a)^{-1}(1 + a)^{-1}(1 + a) = 1$$

   Similarly, for $m > 1$

$$\boldsymbol{C}^*\boldsymbol{C} = (\boldsymbol{I} + \boldsymbol{A}^*)(\boldsymbol{I} - \boldsymbol{A}^*)^{-1}(\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} + \boldsymbol{A}) = (\boldsymbol{I} - \boldsymbol{A})[(\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A}^*)]^{-1}(\boldsymbol{I} + \boldsymbol{A}) \Rightarrow$$

$$\boldsymbol{C}^*\boldsymbol{C} = (\boldsymbol{I} - \boldsymbol{A})[(\boldsymbol{I} + \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A})]^{-1}(\boldsymbol{I} + \boldsymbol{A}) = (\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} + \boldsymbol{A})^{-1}(\boldsymbol{I} + \boldsymbol{A}) = \boldsymbol{I}.$$