Note 1: Notations below that appear as, e.g., "[1], p. 583," refer to page 583 in Reference [1], listed at the end of this document. All of the References are posted on the course webpage under the sub-rubric "Resources" of the rubric "Materials for the midterm project" (towards the end of the webpage). Note that materials presented in different Resources may duplicate one another.

Note 2: Also, within each cited Resource, there will unavoidably be material *not* related to the given presentation Topic. It is your responsibility to "filter out" such extraneous information and present only what is relevant to your Topic.

Note 3: If you are working on a Topic with the sequential number greater than 1, it may occur that you will need to understand some of the material presented in an earlier Topic (or Topics). If so, it will be *your responsibility to contact the person doing a presentation on that Topic* and ask them to help you understand that material, so that you could continue with preparation of your presentation.
Also, it will be a good idea for you to review descriptions of all Topics, and especially those adjacent to yours, to make sure that you do not unintentionally duplicate another speaker. If in doubt — communicate with the other speaker.

Note 4: You will also need to consult the entire posted document "Background information from Linear Algebra" to understand some of the concepts in your Topic.

Breakdown of the Midterm Project into Topics

1. Expansion (coordinates) of a vector in an orthonormal basis (just the result; no derivation, since it had to be in your Linear Algebra class). Projection of one vector on another (the formula with an explanation). Projection of a vector on a space spanned by several orthogonal vectors; the derivation must lead to the outer-product form of the projection matrix. The latter formula as a matrix–matrix product. If P is an $m \times m$ projection matrix, how is its rank related to m (and what does this mean)? A demonstration (in general terms: see below) that for an $m \times n$ matrix U with all orthonormal columns and $m > n$, $U^T U = I_n$.[1] A demonstration that for the same U, $U U^T \neq I_m$. (A demonstration for $m = 3,\ n = 2$ for a U with general, i.e., *not* numeric, entries will suffice. For the $U U^T$ case, you must show what the result of this matrix's acting on a vector of an appropriate dimension will be.)
   For am $m \times m$ matrix U with all orthonormal columns, rewrite the identity $I_m = U U^T$ as the sum of m outer products and present a geometric interpretation of this result for $m = 2$.

   Resources: [1], pp. 383–385, 391–395; [2], pp. 218–219, 366–376, 382–384; you can also find answers to some of the questions stated in compact form in Problems 25 and 26 in [2], p. 408; [3], pp. 570–571; [5], p. 359 (one line from Eq. (2)); [7], pp. 21-{18,19}; [6], pp. 227, 228–229, 234–236.

2. Geometry of the Least-squares problem $A\mathbf{x} = \mathbf{b}$. Interpretation of the best approximation to a vector as its orthogonal projection on a space spanned by a (not necessarily orthogonal) set of vectors. Derivation of the solution to the Least-squares problem leading to a matrix

[1] I_n is the $n \times n$ identity matrix.

$(A^T A)^{-1} A^T$. First encounter with the pseudoinverse matrix; its properties (prove either of the properties involving three matrices on the left-hand side; state the other). Matrix of orthogonal projection on a space spanned by a non-orthogonal set of vectors. Demonstration that it reduces to the formula in Topic 1 when the spanning vectors are orthogonal.

Resources: [1], pp. 406–410; [3], pp. 574–575, 583–586; [6], pp. 227–228.

3. Properties of the eigenvalues and eigenvectors of a symmetric[2] matrix A (real-valuedness, orthogonality). Discuss how the diagonalization $A = QDQ^T$ includes one of these properties but is more general than it. Spectral decomposition of a symmetric matrix. Geometric interpretation of the action of a symmetric matrix on a vector; must compare it with the geometric interpretation of the action of the identity matrix on a vector (see the end of Topic 1). (For this, you will need to briefly remind the audience the expressions of coordinates of a vector in an orthonormal basis, also found in Topic 1.)

   Resources: [5], pp. 348, 349, 352;[3] [1], pp. 383, 385, 443–446; [3], pp. 405–406; [7], pp. 21-{15–20}.

4. Singular Value Decomposition (SVD): must demonstrate why the columns of the U and V matrices are orthogonal and why singular values are real and positive. Reduced form of SVD.[4] In particular, you need to demonstrate, using properties of matrix–matrix multiplication, why the zero rows or columns of Σ do not contribute to the reduced form of SVD.

   Resources: [4], pp. 590, 592–594, 596; [8], pp. SVD-{1,3,4–7}; [9], pp. 91–94.

5. Outer-product form of SVD. Theorem 7.15(a,b,d) in [4].[5] Geometric interpretation of SVD (i.e., how a matrix acts on a vector) for the 2×2 and 2×3 cases. Compare the interpretation for the 2×2 case with the interpretation of the Spectral Decomposition of a symmetric matrix (Topic 3) of the same size.

   Resources: [4], pp. 596, 597, 598–599; [10]; [1], pp. 470–471; [8], pp. SVD-{7–9}; [9], pp. 95.

6. SVD and the pseudoinverse of a matrix. When does the pseudoinverse formula from Topic 2 reduce to the SVD-based pseudoinverse? When will the latter formula work but the former one will not? The SVD-based pseudoinverse gives the minimum-length solution to the Least-squares problem $A\mathbf{x} = \mathbf{b}$: present a derivation of this result. Connection of the solution to this problems to the column space of U and/or V. Brief presentation of Example 7.40 in [4]. Demonstration of either of the properties of the pseudoinverse mentioned in Topic 2 for the SVD-based pseudoinverse.

   Resources: [4], pp. 602–605; quad [1], p. 471; [6], pp. 347–349; quad [9], pp. 96–99.

---

[2] Recall that all matrices in this project are assumed to be real.

[3] This book talks about Hermitian matrices. These are complex counterparts of real symmetric matrices; so you need to replace "Hermitian" with "real symmetric" and the superscript 'H' with 'T'.

[4] Do not confuse the reduced SVD with the outer-product form of SVD, which is covered in the next Topic, or with a truncated SVD, considered in Topic 7.

[5] Disregard the proofs for parts (a,d) presented in the textbook. Instead, base your proofs on the following: (a) — materials found in "Background information from Linear Algebra"; (d) — Apply one of the earlier statements of this Theorem to A^T and connect the result to what is stated in (d).

7. Truncated SVD:
(a) Why is it needed? What is the trade-off? How is the approximation error related to the number k of singular values retained (formula only)?[6] (b) Application to image denoising. Details are as follows.

    (a) Save the provided image matrix X and Matlab code in a folder on your computer. Run the code to see the image. Plot singular values of X, obtained using command `svd`. Let the number of a singular value roughly corresponding to the "elbow" in the plot be k_0. Replot the image matrix truncated at $k = k_0$ terms. Repeat for the values k corresponding to the approximation error of (roughly) 10%, 20%, and 30%. What are the values of $k/\max(k)$ in these cases?

    (b) Add noise to the matrix X using the command
    `X + 1e-3 * max(X, [], 'all') * randn(size(X)).`[7]
    Plot the singular values of the original and contaminated matrices on the *logarithmic* scale (base 10). Explore where you need to truncate the SVD of the contaminated matrix to obtain a roughly optimal image quality. Present image figures and discuss your results.

    Resources: Matrix X (saved as a Matlab data file) and the code to load it are posted; [11]; [12]; [13].

8. The concept of the term–document matrix for text analysis. Its SVD and truncated SVD. Explain what meaning the number k of the retained singular values has in this context. Redo the analysis in [16] using the cosine similarity matrix instead of the "Spearman rank correlation" (which is used in [16]). Specifically, compare the original cosine similarity matrix and that based on the truncated-SVD matrix with $k = 2$ to verify if the "concepts/a.k.a. topics/a.k.a. features" have been indeed made clearer by the truncation. Repeat the analysis for the Example in [15]. Is the term 'die' closer to the Romeo-and-Juliet or the New-Hampshire-motto concept?

    Resources: [14]; [15], Secs. 3.1 and 3.2; [16], pp. 8–14; [5], pp. 227–229; [17]; [5], pp. 227–229 (for cosine similarity).

References

[1] Selected pages from D. Lay et al, *Linear Algebra and its applications*, 6th Ed.

[2] Selected pages on Projections from D. Poole, *Linear Algebra, a modern introduction*, 4th Ed.

[3] Selected pages on Least Squares from D. Poole, *Linear Algebra, a modern introduction*, 4th Ed.

[4] Selected pages on SVD from D. Poole, *Linear Algebra, a modern introduction*, 4th Ed.

[5] Selected pages from S. Leon, *Linear Algebra with applications*, 9th Ed.

[6] When answering this question, you will encounter the notion of the Frobenius norm of a matrix, which you will need to look up online and present only its definition.

[7] You need to understand what this command does.

[6] Selected pages from B. Noble & J. Daniel, *Applied Linear Algebra*, 3rd Ed.

[7] My handwritten notes on Spectral Decomposition of a symmetric matrix.

[8] My handwritten notes on SVD.

[9] Selected pages from M. Embree's Working notes on Linear Algebra. (The author considers complex-valued matrices. You should still consider real-valued ones and thus replace his notations as follows: $A^* \longrightarrow A^T$ and "unitary" $\longrightarrow$ "orthogonal.")

[10] D. Kalman, "A Singularly Valuable Decomposition: The SVD of a matrix," (2002); pp. 5–8.

[11] A. Falini, "A review on the selection criteria for the truncated SVD in Data Science applications," *J. Comput. Math. Data Sci.*, **5**, 100064 (2022). You will need only Sec. 1 up to and including Theorem 1.2 and Sec. 2.1.

[12] A GitHub page on Truncated SVD by Ilias Bilionis. Note that the plot of singular values is in the logarithmic scale, while you need to use the linear (i.e., usual) scale. Also, I will strongly prefer (although not require) that you use Matlab rather than Python.

[13] J. Simatupang, "Noise Reduction in Satellite Imagery Using Singular Value Decomposition," (2024).

[14] Blog entry of Mat Kelcey on Latent Semantic Analysis and SVD (2009): Introduction, Examples 1 and 2, and perhaps Example 3.

[15] Notes by Alex Thomo on Latent Semantic Analysis.

[16] T. Landauer, P. Foltz, D. Laham, "Introduction to Latent Semantic Analysis," *Discourse Processes*, **25**, 259–284 (1998).

[17] Wikipedia's entry on 'Latent Semantic Analysis', Sec. 'Derivation'.