EIGENVALE PROBLEMS AND THE SVD. [5.1 TO 5.3 & 7.4]

Eigenvalue Problems. Introduction

Let A an $n \times n$ real nonsymmetric matrix. The eigenvalue problem:

$$Au = \lambda u \qquad \begin{array}{l} \lambda \in \mathbb{C} : \text{eigenvalue} \\ u \in \mathbb{C}^n : \text{eigenvector} \end{array}$$

Example:

$$A = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix}$$

- $\lambda_1 = 1$ with eigenvector $u_1 = \binom{0}{1}$
- $\lambda_2 = 2$ with eigenvector $u_2 = \binom{1}{2}$
- The set of eigenvalues of A is called the spectrum of A

Eigenvalue Problems. Their origins

- Structural Engineering $[Ku = \lambda Mu]$
- Stability analysis [e.g., electrical networks, mechanical system,..]
- Quantum chemistry and Electronic structure calculations [Schrödinger equation..]
- Application of new era: page ranking on the world-wide web.

Basic definitions and properties

A scalar λ is called an eigenvalue of a square matrix A if there exists a nonzero vector u such that $Au = \lambda u$. The vector u is called an eigenvector of A associated with λ.

- The set of all eigenvalues of A is the 'spectrum' of A. Notation: $\Lambda(A)$.
- λ is an eigenvalue iff the columns of $A - \lambda I$ are linearly dependent.
- λ is an eigenvalue iff $\boxed{\det(A - \lambda I) = 0}$

Compute the eigenvalues of the matrix:

$$A = \begin{pmatrix} 2 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

Eigenvectors?

Basic definitions and properties (cont.)

➤ An eigenvalue is a root of the Characteristic polynomial:

$$p_A(\lambda) = \det(A - \lambda I)$$

➤ So there are n eigenvalues (counted with their multiplicities).

➤ The multiplicity of these eigenvalues as roots of p_A are called algebraic multiplicities.

✍ Find all the eigenvalues of the matrix:

$$A = \begin{bmatrix} 1 & 2 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Find the associated eigenvectors.

✍ How many eigenvectors can you find if a_{33} is replaced by one?

✍ Same questions if a_{12} is replaced by zero.

✍ What are all the eigenvalues of a diagonal matrix?

➤ Two matrices A and B are similar if there exists an invertible matrix V such that

$$A = VBV^{-1}$$

➤ A and B represent the same linear mapping in 2 different bases.

✍ Explain why [Hint: Assume a column of V represents one basis vector of the new basis expressed in the old basis...]

Solution: Let A be linear mapping represented in standard basis $e_1, \cdots, e_n$ (the 'old' basis). Consider a 'new' basis $v_1, v_2, \cdots, v_n$. Assume each v_i is expressed in the old basis and let $V = [v_1, v_2, ..., v_n]$. A vector s in the new basis is expressed as Vs in the old basis (explain). Linear mapping applied to this vector is $t = A(Vs)$. This is expressed in old basis. Then $t = V(V^{-1}AVs)$ expresses the result in new basis: $B = V^{-1}AVs$ represents mapping A in basis V. ■

✍ Show: A and B have the same eigenvalues. What about eigenvectors?

Definition: A is diagonalizable if it is similar to a diagonal matrix

➤ Note : not all matrices are diagonalizable

➤ Theorem 1: A matrix is diagonalizable iff it has n linearly independent eigenvectors

Example: Which of these matrices is/are diagonalizable

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

➤ Theorem 2: The eigenvectors associated with distinct eigenvalues are linearly independent

✍ Prove the result for 2 distinct eigenvalues

Solution: Let $Au_1 = \lambda_1 u_1$ and $Au_2 = \lambda_2 u_2$ with $\lambda_1 \neq \lambda_2$. We prove that if $\alpha_1 u_1 + \alpha_2 u_2 = 0$ then we must have $\alpha_1 = \alpha_2 = 0$. Multiply $\alpha_1 u_1 + \alpha_2 u_2 = 0$ by $A - \lambda_1 I$: then

$$\begin{aligned} (A - \lambda_1 I)\,[\alpha_1 u_1 + \alpha_2 u_2] &= 0 \rightarrow \\ \alpha_1 (A - \lambda_1 I) u_1 + \alpha_2 (A - \lambda_1 I) u_2 &= 0 \rightarrow \\ 0 + \alpha_2 (\lambda_2 - \lambda_1 I) u_2 &= 0 \end{aligned}$$

Since $\lambda_2 \neq \lambda_1$ we must have $\alpha_2 = 0$. Similar argument will show that $\alpha_1 = 0$. ■

➤ Consequence: if all eigenvalues of a matrix A are simple then A is diagonalizable.

➤ Theorem 3: A symmetric matrix has real eigenvalues and is diagonalizable. In addition A admits a set of orthonormal eigenvectors.

Transformations that Preserve Eigenstructure

Shift	$B = A - \sigma I$: $Av = \lambda v \iff Bv = (\lambda - \sigma)v$ eigenvalues move, eigenvectors remain the same.
Poly-nomial	$B = p(A) = \alpha_0 I + \cdots + \alpha_n A^n$: $Av = \lambda v \iff Bv = p(\lambda)v$ eigenvalues transformed, eigenvectors remain the same.
Invert	$B = A^{-1}$: $Av = \lambda v \iff Bv = \lambda^{-1} v$ eigenvalues inverted, eigenvectors remain the same.

✍ Let A be diagonalizable. How would you compute $p(A)$ if p is a high degree polynomial? [Hint: start with A^k]

The Singular Value Decomposition (SVD)

Theorem For any matrix $A \in \mathbb{R}^{m\times n}$ there exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that

$$A = U \Sigma V^T$$

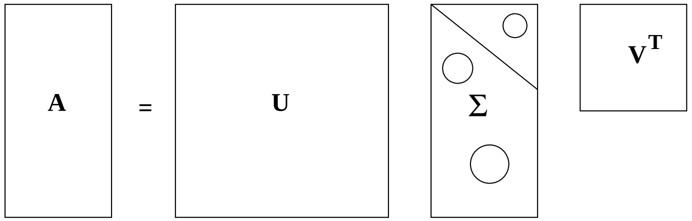
where Σ is a diagonal matrix with entries $\sigma_{ii} \geq 0$.

$$\sigma_{11} \geq \sigma_{22} \geq \cdots \sigma_{pp} \geq 0 \text{ with } p = \min(n, m)$$

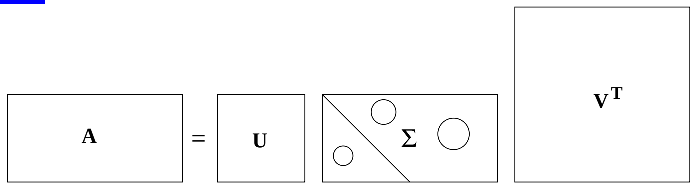
➤ The σ_{ii} are the singular values of A.

➤ σ_{ii} is denoted simply by σ_i

Case 1:



Case 2:



The "thin" SVD

➤ Consider the Case-1. It can be rewritten as

$$A = [U_1 U_2] \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix} V^T$$

Which gives:

$$A = U_1 \Sigma_1 V^T$$

where U_1 is $m \times n$ (same shape as A), and Σ_1 and V are $n \times n$

➤ referred to as the "thin" SVD. Important in practice.

✍ How can you obtain the thin SVD from the QR factorization of A and the SVD of an $n \times n$ matrix?

A few properties. Assume that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \text{ and } \sigma_{r+1} = \cdots = \sigma_p = 0$$

Then:

- $\text{rank}(A) = r =$ number of nonzero singular values.
- $\text{Ran}(A) = \text{span}\{u_1, u_2, \ldots, u_r\}$
- $\text{Null}(A) = \text{span}\{v_{r+1}, v_{r+2}, \ldots, v_n\}$
- The matrix A admits the SVD expansion:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Rank and approximate rank of a matrix

➤ The number of nonzero singular values r equals the rank of A

➤ In practice: zero singular values replaced by small values due to noise.

➤ Can define approximate rank: rank obtained by 'neglecting smallest singular values'

Example: Let A a matrix with singular values

$$\sigma_1 = 10.0; \quad \sigma_2 = 6.0; \quad \sigma_3 = 3.0;$$
$$\sigma_4 = 0.030; \quad \sigma_5 = 0.0130; \quad \sigma_6 = 0.0010;$$

➤ $\sigma_4, \sigma_5, \sigma_6$, are likely due to noise - so approximate rank is 3.

➤ Rigorous way of stating this exists – but beyond scope of this class [see csci 5304]

Right and Left Singular vectors:

$$A v_i = \sigma_i u_i$$
$$A^T u_j = \sigma_j v_j$$

➤ Consequence $A^T A v_i = \sigma_i^2 v_i$ and $A A^T u_i = \sigma_i^2 u_i$

➤ Right singular vectors (v_i's) are eigenvectors of $A^T A$

➤ Left singular vectors (u_i's) are eigenvectors of $A A^T$

➤ Possible to get the SVD from eigenvectors of AA^T and $A^T A$ – but: difficulties due to non-uniqueness of the SVD

A few applications of the SVD

Many methods require to approximate the original data (matrix) by a low rank matrix before attempting to solve the original problem

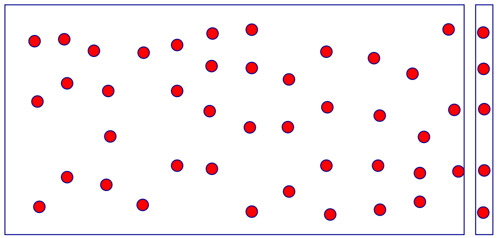
➤ Regularization methods require the solution of a least-squares linear system $Ax = b$ approximately in the 'dominant singular' space of A

➤ The Latent Semantic Indexing (LSI) method in information retrieval, performs the "query" in the dominant singular space of A

➤ Methods utilizing Principal Component Analysis, e.g. Face Recognition.

Information Retrieval: Vector Space Model

➤ Given: a collection of documents (columns of a matrix A) and a query vector q.



➤ Collection represented by an $m \times n$ term by document matrix with $\boxed{a_{ij} = L_{ij} G_i N_j}$

➤ Queries ('pseudo-documents') q are represented similarly to a column

Vector Space Model - continued

➤ Problem: find a column of A that best matches q

➤ Similarity metric: angle between column c and query q

$$\cos \theta(c, q) = \frac{|c^T q|}{\|c\| \|q\|}$$

➤ To rank all documents we need to compute

$$s = A^T q$$

➤ s = similarity vector.

➤ Literal matching – not very effective.

➤ Problems with literal matching: *polysemy*, *synonymy*,...

Use of the SVD

- Solution: Extract intrinsic information – or underlying "semantic" information –
- LSI: replace matrix A by a low rank approximation using the Singular Value Decomposition (SVD)

$$A = U\Sigma V^T \quad \rightarrow \quad A_k = U_k \Sigma_k V_k^T$$

- U_k : term space, V_k: document space.
- Refer to this as Truncated SVD (TSVD) approach
- Amounts to replacing small sing. values of A by zeros

New similarity vector:

$$s_k = A_k^T q = V_k \Sigma_k U_k^T q$$

LSI : an example

```
%% D1   :   INFANT & TODDLER first aid
%% D2   :   BABIES & CHILDREN's room for your HOME
%% D3   :   CHILD SAFETY at HOME
%% D4   :   Your BABY's HEALTH and SAFETY
%%      :   From INFANT to TODDLER
%% D5   :   BABY PROOFING basics
%% D6   :   Your GUIDE to easy rust PROOFING
%% D7   :   Beanie BABIES collector's GUIDE
%% D8   :   SAFETY GUIDE for CHILD PROOFING your HOME
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% TERMS:  1:BABY   2:CHILD  3:GUIDE  4:HEALTH  5:HOME
%%          6:INFANT  7:PROOFING   8:SAFETY  9:TODDLER
%% Source: Berry and Browne, SIAM., '99
```

- Number of documents: 8
- Number of terms: 9

- Raw matrix (before scaling).

$$A = \begin{array}{|cccccccc|l}
d1 & d2 & d3 & d4 & d5 & d6 & d7 & d8 & \\
\hline
 & 1 & & 1 & 1 & & 1 & & bab \\
 & 1 & 1 & & & & & 1 & chi \\
 & & & & & 1 & 1 & 1 & gui \\
 & & & 1 & & & & & hea \\
 & 1 & 1 & & & & & 1 & hom \\
1 & & & 1 & & & & & inf \\
 & & & & 1 & 1 & & 1 & pro \\
 & & 1 & 1 & & & & 1 & saf \\
1 & & & 1 & & & & & tod \\
\hline
\end{array}$$

Get the anwser to the query `Child Safety`, so

$$q = [0\ 1\ 0\ 0\ 0\ 0\ 0\ 1\ 0]$$

using cosines and then using LSI with $k = 3$.