Recitation 13

Thursday Oct 22, 2024

1 Recap

1.1 Diagonalization, Similar Matrices, and Matrix Powers

If a square matrix $m \times m$ has a *basis* of eigenvectors, i.e. it has m linearly independent eigenvectors $x_1, \ldots, x_m$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_m$, then we obtain the following factorization of A:

$$A = \underbrace{\begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix}}_{X} \underbrace{\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix}}_{\Lambda} X^{-1}$$

called the **diagonalization** of A. We say that such a matrix is *diagonalizable* — this is true of *almost all* square matrices, except for unusual cases involving repeated roots λ that never happen by chance. For now, we will mostly not worry about weird non-diagonalizable cases (called *defective*).

It follows that

$$A^n = X\Lambda^n X^{-1}$$

which can be interpreted as writing the input vector in the basis of eigenvectors (via X^{-1}), multiplying each coefficient by the corresponding λ_k^n (via Λ^n), and then adding up the linear combinations of eigenvectors (via X).

If we have a **dominant** eigenvalue $|\lambda_1| > |\lambda_k|$ for all $k \neq 1$, then for a sufficiently large n the matrix A^n becomes approximately rank 1 (all terms except λ_1^n become negligible). Equivalently, for a random input vector y we expect $A^n y \approx \lambda_1^n x_1 c_1$ for some coefficient c_1 (from solving $Xc = y$).

1.2 Determinant and trace

Let A is an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The following properties hold:

1. The determinant of A is the product of its eigenvalues:

$$\det(A) = \lambda_1 \lambda_2 \ldots \lambda_n$$

2. The trace of A (the sum of its diagonal elements) is equal to the sum of the eigenvalues:

$$\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

We can get both of these properties for diagonalizable A by simply plugging in the diagonalization. For trace, we need the key property that $\boxed{\text{tr}(AB) = \text{tr}(BA)}$ for any matrices A and B.

1.3 Similar matrices

We say that two $m \times m$ matrices A and B are **similar** if $A = SBS^{-1}$ for some invertible matrix S. Conceptually, they correspond to the "same" linear operation with a change of basis S (i.e. a change of "coordinate system").

We showed in class that similar matrices have the **same determinant**, the **same trace**, and the **same eigenvalues**. (However, they have *different eigenvectors* by a factor of S!).

A diagonalizable matrix A is similar to a diagonal matrix Λ (with $S = X$, i.e. in the basis of eigenvectors).

1.4 Application: Linear Dynamical Systems

Consider the *linear dynamical system* defined by the recurrence $y_{k+1} = Ay_k$. Suppose that A is diagonalizable, i.e. $A = X\Lambda X^{-1}$. What can we say about y_k as $k \to \infty$?

$$y_k \;= A^k y_0 \;= X\Lambda^k X^{-1} y_0 \;= X \begin{bmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix} X^{-1} y_0$$

Note the following:

- If $|\lambda_i| > 1$, then λ_i^k diverges ($|\lambda_i|^k \to \infty$)
- If $|\lambda_i| < 1$, then $\lambda_i^k \to 0$.
- If $|\lambda_i| < 1$ for every $i = 1, \ldots, n$, then $\Lambda^k \to 0$ as $k \to \infty$, and hence $y_k \to 0$. Sometimes we say this is a **stable** system.
- If $\lambda_j = 1$ for some j's and other $|\lambda_i| < 1$, then y_k goes to a nonzero **steady-state** vector (such as for Markov matrices, as in the PageRank example from class). (Assumes A is diagonalizable.)
- If $|\lambda_j| = 1$ but $\lambda_j \neq 1$ (e.g. it $= -1$ or $= i$ or $= e^{i\phi}$ for a real phase angle $0 < \phi < 2\pi$) for some j's and other $|\lambda_i| \leq 1$, then y_k **oscillates forever** (asymptotically, neither decaying nor growing). (Assumes A is diagonalizable.)
- If y_0 is an eigenvector of A, then $A^k y_0 = \lambda_0^k y_0$.
- If y_0 is a linear combination of eigenvectors, e.g. $y_0 = c_1 x_1 + c_2 x_2$, then $A^k y_0 = \lambda_1^k c_1 x_1 + \lambda_2^k c_2 x_2$. If one eigenvalue λ_1 has the biggest magnitude, then this term *dominates* and $A^k y_0 \approx \lambda_1^k c_1 x_1$

2 Exercises

1. For arbitrary $m \times m$ matrices A, B, C, which of the following *must* be equal to $\mathrm{tr}(ABC)$: $\mathrm{tr}(BCA)$, $\mathrm{tr}(BAC)$, $\mathrm{tr}(CAB)$, and/or $\mathrm{tr}(CBA)$?

2. Recall the Frobenius norm $\|A\|_F$ of any $m \times n$ matrix A is simply the square root of the sum of the squares of the matrix entries, i.e. $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2}$. It turns out that this has an important relationship to the trace operation:

   (a) Show that $\|A\|_F = \sqrt{\mathrm{tr}(A^T A)}$

   (b) From the compact SVD $A = \hat{U}\hat{\Sigma}\hat{V}^T$ and the trace formula, show that $\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$.

3. Consider the matrix
$$\begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix}$$
from lecture, which has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = 3$, and corresponding eigenvectors $x_1 = [1, 1]$ and $x_2 = [1, 2]$ (commas = column vectors).

   a) What do we get if we take the vector $x = [3, 4] = 2x_1 + x_2$ and multiply 100 times by A^{-1}?

   b) What happens if we take x and multiply many times by $(2A - 5I)^{-1}$? Does it converge to a particular vector?

   c) More generally, if we have an arbitrary matrix A with all eigenvalues distinct, and we multiply a vector x repeatedly by A^{-1}, it typically approaches what eigenvector? When might this fail to happen?

4. In this problem, we will explore one of many applications of diagonalization – solving a recurrence relation, similar to the Fibonacci example from class. In particular, suppose that we have a sequence of real numbers $\gamma_0 = \gamma_1 = -1$, and $\gamma_{n+1} = 5\gamma_n - 6\gamma_{n-1}$ for any $n \geq 1$. We want to determine a closed form of γ_k for any $k \geq 0$.

   (a) Compute γ_2, γ_3.

   (b) Let $g_i = \begin{bmatrix} \gamma_{i+1} \\ \gamma_i \end{bmatrix}$. What is g_0?

   (c) The relationship $\gamma_{n+2} = 5\gamma_{n+1} - 6\gamma_n$ implies $g_{n+1} = Ag_n$ for a proper choice of 2×2 matrix A. Find A. In other words, find the 2×2 matrix A such that
$$\begin{bmatrix} \gamma_{n+2} \\ \gamma_{n+1} \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} \gamma_{n+1} \\ \gamma_n \end{bmatrix} \quad \text{for any } n \geq 0.$$

   (d) Given the relationship $g_{i+1} = Ag_i$ for any $i \geq 0$, show that $g_k = A^k g_0$.

   (e) Find the eigenvalues λ_1, λ_2 and and their associated eigenvectors v_1, v_2 of A. Find its diagonalization. **Does γ_k go to zero, approach a steady state, oscillate forever, or diverge as $k \to \infty$?**

   (f) Compute $g_k = A^k g_0$. What is γ_k?

   (g) Compute $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ by the formula in the previous part. Do they match the answers from above?

   (h) Alternatively, write g_0 as a linear combination of v_1 and v_2. How does it aid us in computing $g_k = A^k g_0$?

3 Solutions

1. $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$. You can't arbitrarily re-order the entries. The identity $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ means that $\mathrm{tr}(ABC) = \mathrm{tr}(A(BC)) = \mathrm{tr}((BC)A)$, or $\mathrm{tr}(ABC) = \mathrm{tr}((AB)C) = \mathrm{tr}(C(AB))$. (This is a common confusion with this identity. Equivalently, you can only perform "cyclic" permutations of the matrices in the trace.)

2. (a) From the definition of trace and matrix product, we have $\mathrm{tr}(A^TA) = \sum_{j=1}^n (A^TA)_{jj} = \sum_{j=1}^n \left(\sum_{i=1}^m (A^T)_{ji}A_{ij}\right) = \sum_{j=1}^n \sum_{i=1}^m A_{ij}A_{ij}$, which is exactly the sum of the squares of the entries of A.

   (b) From, $A = \hat{U}\hat{\Sigma}\hat{V}^T$, we have:

$$\begin{aligned}
\mathrm{tr}(A^TA) &= \mathrm{tr}(\hat{V}\hat{\Sigma}^T\cancel{\hat{U}^T\hat{U}}\hat{\Sigma}\hat{V}^T) \\
&= \mathrm{tr}(\hat{V}\hat{\Sigma}^2\hat{V}^T) = \mathrm{tr}(\hat{\Sigma}^2\cancel{\hat{V}^T\hat{V}}) \\
&= \mathrm{tr}(\hat{\Sigma}^2) = \sigma_1^2 + \cdots + \sigma_r^2 \\
&= \|A\|_F^2
\end{aligned}$$

3. a) We know that the eigenvectors of A, A^{-1} are the same, and the eigenvalues are reciprocal. So the eigenvalues of A^{-1} are $\lambda_1 = 1/2$ and $\lambda_2 = 1/3$, and the corresponding eigenvectors are x_1, x_2 respectively. We also know that $A^{-n}x_i = (\lambda_i)^{-n}x_i$. So

$$A^{-100}x = 2(1/2)^{100}x_1 + (1/3)^{100}x_2 = (1/2^{99})x_1 + (1/3^{100})x_2.$$

   b) We can figure this out without ever computing $(2A - 5I)^{-1}$. This is because the eigenvectors of $A, 2A, 2A - 5I, (2A - 5I)^{-1}$ are all the same. So we just need to figure out how the eigenvalues of A and $(2A - 5I)$ are related.

   Scaling A scales the eigenvalues by the same amount, so the eigenvalues of $2A$ are $4, 6$. Translating $2A$ by $-5I$ adds -5 to each eigenvalue, so the eigenvalues of $2A - 5I$ are $-1, 1$. So the eigenvalues of $(2A - 5I)^{-1}$ are $-1, 1$ (since $-1 = 1/-1$ and $1 = 1/1$). This means

$$(2A - 5I)^{-n}x = 2(-1)^n x_1 + (1)^n x_2 = 2(-1)^n x_1 + x_2.$$

   So $(2A - 5I)^{-n}x$ does not converge to a particular vector! It vacillates between $-2x_1 + x_2$ and $2x_1 + x_2$ depending on if n is even or odd.

   c) Say the eigenvalues of A are $\lambda_1, \ldots, \lambda_r$ and the basis of eigenvectors is $v_1, \ldots, v_r$. Say we're interested in some vector x, which can be expressed in the basis of eigenvectors as

$$x = c_1v_1 + c_2v_2 + \cdots + c_rv_r.$$

   As we've seen above,

$$A^{-n}x = c_1(1/\lambda_1)^n v_1 + \cdots + c_r(1/\lambda_r)^n v_r.$$

   Usually, there will be one eigenvalue, say λ_1 with smallest modulus, so $1/\lambda_1$ has larger modulus than all the other eigenvalues of A^{-1}. The vector $A^{-n}x$ tends to approach an eigenvector with eigenvalue $1/\lambda_1$, since the coefficent

$c_1(1/\lambda_1)^n$ of v_1 is much much larger in modulus than the coefficients on the other eigenvectors.

There is an exception to this, though—when another eigenvalue, say $\lambda_2 \neq \lambda_1$, has the same modulus as λ_1. Then two coefficients are comparable in modulus, and $A^{-n}x$ does not approach an eigenvector — it *oscillates forever* (while decaying or growing, depending on the modulus).

4. (a) $\gamma_2 = 5\gamma_1 - 6\gamma_0 = 5(-1) - 6(-1) = 1, \gamma_3 = 5\gamma_2 - 6\gamma_1 = 5(1) - 6(-1) = 11.$

   (b) $g_0 = \begin{bmatrix}\gamma_1\\ \gamma_0\end{bmatrix} = \begin{bmatrix}-1\\ -1\end{bmatrix}.$

   (c) $\begin{bmatrix}\gamma_{n+2}\\ \gamma_{n+1}\end{bmatrix} = \begin{bmatrix}5\gamma_{n+1} - 6\gamma_n\\ \gamma_{n+1}\end{bmatrix} = \gamma_{n+1}\cdot\begin{bmatrix}5\\ 1\end{bmatrix} + \gamma_n\cdot\begin{bmatrix}-6\\ 0\end{bmatrix} = \begin{bmatrix}5 & -6\\ 1 & 0\end{bmatrix}\begin{bmatrix}\gamma_{n+1}\\ \gamma_n\end{bmatrix}$. This gives $A = \begin{bmatrix}5 & -6\\ 1 & 0\end{bmatrix}$.

   (d) It follows that $g_k = Ag_{k-1} = A(Ag_{k-2}) = A^2 g_{k-2} = ... = A^k g_0$.

   (e) An eigenvalue λ follows an equation $\det(\lambda I - A) = 0 \iff \det\begin{bmatrix}\lambda - 5 & 6\\ -1 & \lambda\end{bmatrix} = 0 \iff \lambda^2 - 5\lambda + 6 = 0 \iff (\lambda - 3)(\lambda - 2) = 0$. Thus, the two eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 2$. This immediately tells us that the sequence γ_k is **exponentially growing**, asymptotically growing proportional to 3^k.

   We can determine $v_1 \in N(3I - A)$ to be $\begin{bmatrix}3\\ 1\end{bmatrix}$, and $v_2 \in N(2I - A)$ to be $\begin{bmatrix}2\\ 1\end{bmatrix}$. The diagonalization $A = V\Lambda V^{-1}$ is therefore given by $V = [v_1\ v_2] = \begin{bmatrix}3 & 2\\ 1 & 1\end{bmatrix}$ and $\Lambda = \text{diagonal}(3, 2) = \begin{bmatrix}3 & 0\\ 0 & 2\end{bmatrix}$. In addition, we can compute $V^{-1} = \begin{bmatrix}1 & -2\\ -1 & 3\end{bmatrix}$.

   (f) $\begin{bmatrix}\gamma_{k+1}\\ \gamma_k\end{bmatrix} = g_k = A^k g_0 = V\Lambda^k V^{-1} g_0 = \begin{bmatrix}3 & 2\\ 1 & 1\end{bmatrix}\begin{bmatrix}3^k & 0\\ 0 & 2^k\end{bmatrix}\begin{bmatrix}1 & -2\\ -1 & 3\end{bmatrix}\begin{bmatrix}-1\\ -1\end{bmatrix} = \begin{bmatrix}3^{k+1} - 2^{k+2}\\ 3^k - 2^{k+1}\end{bmatrix}$. This gives $\gamma_k = 3^k - 2^{k+1}$.

   (g) Plug in $k = 0, 1, 2, 3$ into our formula from the previous part, we have $\gamma_0 = -1, \gamma_2 = -1, \gamma_2 = 1, \gamma_3 = 11$ which match the answers from above.

   (h) $g_0 = \begin{bmatrix}-1\\ -1\end{bmatrix} = \begin{bmatrix}3\\ 1\end{bmatrix} - 2\begin{bmatrix}2\\ 1\end{bmatrix} = v_1 - 2v_2$. Then we can compute $A^k g_0 = A^k(v_1 - 2v_2) = A^k v_1 - 2A^k v_2 = \lambda_1^k v_1 - 2\lambda_2^k v_2 = 3^k\begin{bmatrix}3\\ 1\end{bmatrix} - 2\cdot 2^k\begin{bmatrix}2\\ 1\end{bmatrix} = \begin{bmatrix}3^{k+1} - 2^{k+2}\\ 3^k - 2^{k+1}\end{bmatrix}$. This gives $\gamma_k = 3^k - 2^{k+1}$.