Math 307 Review Notes

Reese Critchlow

Midterm 1

LU Decomposition

The gaussian elimination on a matrix A can be expressed using LU Decomposition. LU Decomposition follows the form:

$$A = LU$$

Where L is a unit lower triangular matrix and U is a upper triangular matrix.

Unit Lower Triangular Matricies

A unit lower triangular matrix has the following attributes:

- The matrix is square ($n \times n$).
- The diagonal entries of the matrix are ones and only zeroes are above the ones.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ x & 1 & 0 & 0 \\ x & x & 1 & 0 \\ x & x & x & 1 \end{bmatrix}$$

Upper Triangular Matricies

An upper triangular matrix has the following attribute:

- The matrix has only zeroes below the main diagonal.

$$\begin{bmatrix} 0 & x & x & x \\ 0 & 0 & x & x \end{bmatrix}$$

It is also important to note that $L = E_1^{-1}E_2^{-1}E_3^{-1}$.

To find the LU decomposition:

1. Perform Gaussian elimination to obtain the row echelon form of the matrix, by using matrix multiplication, noting each E along the way. The matrix in REF is the U part of the LU decomposition.
2. Compute the inverse of all of the E matricies.
3. Multiply the inverse matricies in reverse order together to obtain L.

Elementary Row Operations as Matrix Multiplications

The elementary row operations can be expressed as matrix multiplications. They are as follows:

Interchange rows i and j.

Modify the identity matrix such that:

$$a_{i,i} = 0 \qquad a_{j,j} = 0$$
$$a_{i,j} = 1 \qquad a_{j,i} = 1$$

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1_i & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1_j & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Multiply a row i by a scalar k.

Modify the identity matrix such that:

$$a_{i,i} = k$$

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & k_i & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Add c times row j to i.

Modify the identity matrix such that:

$$a_{i,j} = c$$

Note: it is required that $i > j$.

$$E = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ c_{i,j} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

It is also important to note that the E^{-1} can be acheived by

modifying the identity matrix such that:

$$a_{j,i} = -c.$$

Hence, Gaussian elimination can be expressed as $E_3E_2E_1A$.

It is important to note that the order of the inverse is paramount. The inverse of $E_3E_2E_1$ is $E_1^{-1}E_2^{-1}E_3^{-1}$, not $E_3^{-1}E_2^{-1}E_1^{-1}$. In short, inverses are to be applied in the opposite order that the original matricies were applied.

Generally, only the E transformation is used, because the rest of the transformations do not result in a lower triangular matrix.

Theorem: If a matrix A can be converted to row echelon form using only E row operations, then A has an LU decomposition.

Sometimes, the LU decomposition is not attainable, thus, other transformations are permitted, but the LU decomposition will take on the form of $A = PLU$ where P is a permuatation matrix.

Important Properties of the LU Decomposition:

- $\text{rank}(A) = \text{rank}(U)$
- If a is a square matrix, then: $\det(A) = \det(U)$ If A is a square matrix of full rank:
- $\det(A) \neq 0$.
- A is invertible.
- $A\vec{x} = \vec{b}$ has a unique solution.

Rank of a Matrix: As a review, the rank of a matrix is:

- The dimension of the span of the matrix.
- The number of non-zero leading/pivot entries in a matrix which is in row echelon form.

Row Echelon Form: As a review, a matrix is considered to be in row echelon form when

- All rows consisting of only zeroes are at the bottom.
- The pivot entry of any nonzero row is always strictly to the right of the leading coefficient of the row above it.

Notes on Rank:

- If $\text{rank}(A) = \text{rank}(A|B) =$ the number of rows in $\vec{x}$, then the system has a unique solution.
- If $\text{rank}(A) = \text{rank}(A|B) <$ the number of rows in $\vec{x}$, then the system has infinite solutions.
- If $\text{rank}(A) <$ the number of rows in A.
- Overarching theory: $\dim(N(A)) = \text{cols}_A - \text{rank}(A)$.

Inverting a Matrix: To invert a matrix A, create a amtrix $[A \mid I]$ and use row operations to transform the matrix to a form of $[I \mid A^{-1}]$.

Error Analysis and Matrix Norms

Matrix Norm: The matrix norm or operator norm is defined by:

$$||A|| = \max_{x \neq 0} \frac{||A\vec{x}||}{||\vec{x}||} \qquad \text{where } ||\vec{x}|| \text{is the } e^2 \text{ norm.}$$

The matrix norm describes the maximum stretch of a unit vector. Inverse Matrix Norm: If a matrix A is a square, non-singular matrix, then

$$||A^{-1}|| = \frac{1}{\min_{||\vec{x}||=1} ||A\vec{x}||}.$$

Significance of the Matrix Norm: If a matrix or a vector is obtained emperically, there may be errors in it. Thus, the matrix norm allows us to predict how large the effect of those errors may be.

Condition Number: The condition number of a matrix A is defined as:

$$\text{cond}(A) = ||A|| \cdot ||A^{-1}||$$

Error Bounding: Given a system $A\vec{x} = \vec{b}$, assuming that $A(\vec{x} + \Delta\vec{x}) = \vec{b} + \Delta\vec{b}$, then:

$$\frac{||\Delta\vec{x}||}{||\vec{x}||} \leq \text{cond}(A) \cdot \frac{||\Delta\vec{b}||}{||\vec{b}||}.$$

The equation above describes the error in $\vec{x}$ as a result of errors in $\vec{b}$.

One can also produce an expression for the effects of errors in A:

$$\frac{||\Delta\vec{x}||}{||\vec{x} + \Delta\vec{x}||} \leq \text{cond}(A) \cdot \frac{||\Delta A||}{||A||}$$

A common type of problem is to extract the condition number from the manipulation of an image.
Norm For Diagonal Matricies: The norm of a diagonal matrix D can be obtained with the following formula:

$$D = \max\{|d_k|\}$$

where d_k are the set of diagonal entries in the matrix.

Relations of the Condition Number: It is important to note that:

$$||AB|| \leq ||A|| \cdot ||B|| \qquad \text{thus} \qquad \text{cond}(AB) \leq \text{cond}(A) \cdot \text{cond}(B)$$

Polynomial Interpolation

A polynomial of the form:

$$P_\alpha = \left\{c_0 + c_1 t + c_2 t^2 + \cdots + c_\alpha t^\alpha : c_n \in \mathbb{R}\right\}$$

Can be used to interpolate α number of points, so long as $t_i \neq t_j$, and $y_i \neq t_i$.

Given a system with multiple points we can solve a system of equations to obtain the unknonw coefficients, c_n.

Such a system of equations takes the form of $A\vec{c} = \vec{y}$, where A is known as the Vandermonde Matrix.

$$\begin{bmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^\alpha \\ 1 & t_1 & t_1^2 & \cdots & t_1^\alpha \\ 1 & t_2 & t_2^2 & \cdots & t_2^\alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_\alpha & t_\alpha^2 & \cdots & t_\alpha^\alpha \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_\alpha \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_\alpha \end{bmatrix}$$

Determinant of a Vandermonde Matrix: The determinant of a Vandermonde matrix can be obtained by the following formula:

$$\det(A) = \prod_{0 \leq i < j \leq \alpha} (t_j - t_i)$$

It is important to note that in a Vandermonde matrix, the entries $t_0, t_1, \cdots, t_n$ represent time points and the y vector represents their corresponding values.

It is also important to note that the condition number of a Vandermonde matrix gets *very* large when the number of points increases.

Cubic Spline Interpolation

Another method for interpolation is the cubic spline method. Unlike the continuous nature of the Polynomial Interpolation, the cubic spline method is piecewise.

By definition, a cubic spline interpolation has the following features/properties.

- There are N cubic polynomials of the form:
  $p_k(t) = a_k(t - t_{k-1})^3 + b_k(t - t_{k-1})^2 + c_k(t - t_{k-1}) + d_k$
- $p(t)$, $p'(t)$, and $p''(t)$ are continuous
- $p(t_i) = y_i$ for all $i = 0, \cdots, N$

To find a cubic spline interpolation, one must use the coefficient matrix of a cubic spline, which has the following form:

$$C = \begin{bmatrix} a_1 & a_2 & \cdots & a_N \\ b_1 & b_2 & \cdots & b_N \\ c_1 & c_2 & \cdots & c_N \\ d_1 & d_2 & \cdots & d_N \end{bmatrix}$$

Hence, it can be concluded from this that every cubic spline interpolation coefficient matrix will have $4N$ terms, where N is the number of data points.

Equations Defining the Cubic Spline: There exist several methods to obtain the equations to solve for a cubic spline. They are as follows:

1. Interpolation
   (a) Left Endpoints
   Basis: $p_k(t_{k-1}) = y_{k-1}$
   Result: $d_k = y_{k-1}$
   Number of Equations: N
   (b) Right Endpoints
   Basis: $p_k(t_k) = y_k$
   Result: $a_k(t_k - t_{k-1})^3 + b_k(t_k - t_{k-1})^2 + c_k(t_k - t_{k-1}) + d_k = y_k$
   Number of Equations: N
2. Continuity of the Derivative
   Basis: $p'_k(t_k) = p'_{k+1}(t_k)$
   Result: $3a_k(t_k - t_{k-1})^2 + 2b_k(t_k - t_{k-1}) + c_k = c_{k+1}$
   Number of Equations: $N - 1$
3. Continuity of the Second Derivative
   Basis: $p''_k(t_k) = p''_{k+1}(t_k)$
   Result: $6a_k(t_k - t_{k-1}) + 2b_k = 2b_{k+1}$
   Number of Equations: $N - 1$
4. Secondary Methods
   (a) Neutral Cubic Spline Conditions Equations: $p''(t_0) = 0$ and $p''_N(t_N) = 0$

Obtaining the Coefficients for a Cubic Spline: First, it is known that $d_n = y_{n-1}$. Thus, the system that needs to be solved is as follows:

$$\begin{bmatrix} A(L_1) & B & 0\cdots 0 & & \\ 0 & A(L_2) & B\cdots 0 & & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T & 0 & 0 & V & \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ \vdots \\ a_N \\ b_N \\ c_N \end{bmatrix} = \begin{bmatrix} y_1 - y_0 \\ 0 \\ 0 \\ y_2 - y_1 \\ 0 \\ 0 \\ \vdots \\ y_N - y_{N-1} \\ 0 \\ 0 \end{bmatrix}$$

Where $L_k = t_k - t_{k-1}$ and:

$$A(L) = \begin{bmatrix} L^3 & L^2 & L \\ 3L^2 & 2L & 1 \\ 6L & 2 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -2 & 0 \end{bmatrix} \qquad T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad V = \begin{bmatrix} L_N^3 & L_N^2 & L_N \\ 0 & 0 & 0 \\ 6L_N & 2 & 0 \end{bmatrix}.$$

Subspaces

Definition: A subset $U \subseteq \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ under the following conditions:

1. U contains the zero vector $\vec{0} \in \mathbb{R}$.
2. Closed under addition: $\vec{u_1}, \vec{u_2} \in U \Rightarrow \vec{u_1} + \vec{u_2} \in U$.
3. Closed under scalar multiplication: $\vec{u} \in U, c \in \mathbb{R} \Rightarrow c\vec{u} \in U$.

For example, the smallest subspace of $\mathbb{R}^2$ is $\{\vec{0}\}$, and the largest subspace of $\mathbb{R}^2$ is simply $\mathbb{R}^2$. A common subspace however is any line passing through the origin.

Subspaces of $\mathbb{R}^3$ include lines through the origin and planes containing the origin.

Span

Definition: The span of a set of vectors is the set of all of the possible inear combinations of them.

Determining Span Membership: To determine whether a vector, $\vec{v}$ is within the span of a set of vectors, $\{\vec{u_n}\}$, one can write $\begin{bmatrix} \vec{u_1} & \vec{u_2} & \cdots & \vec{u_n} \mid \vec{v} \end{bmatrix}$ and solve the matrix.

Linear Independence

The vectors $\vec{u_1}, \cdots, \vec{u_m} \in \mathbb{R}$ are said to be linearly independent if $c_1\vec{u_1} + \cdots + c_m\vec{u_m} = \vec{0}$ if and only if the solution is trivial.

A more algorithmic approach to finding linear independence is to solve the matrix $\begin{bmatrix} \vec{u_1} \cdots \vec{u_n} \mid \vec{0} \end{bmatrix}$

Basis and Dimension

Definition: Let $U \subseteq \mathbb{R}^n$ be a subspace. A set of vectors $\{\vec{u_m}\}$ forms a basis of U if:

1. $\{\vec{u_m}\}$ is a linearly independent set.
2. $\text{span}\{\vec{u_m}\} = U$.

Remark: There are infinitley many different bases of a subspace U, but each basis of U has the same number of vectors.

Dimension: The dimension of U is the number of vectors in a basis of U. It is written as $\dim(U)$.

Finding Dimension and Span: Given a set of vectors $\{\vec{u_m}\}$, we can solve $\left[\vec{u_1} \cdots \vec{u_m} \mid \vec{0}\right]$ In row echelon form, redundant columns can be eliminated such that the associated matrix in row echelon form is full rank. Thus, $\dim(U)$ is the rank of the matrix, or the number of vectors in the span.

Nullspace $N(A)$

Definition: Let A be an $m \times n$ matrix. The nullspace of A is:

$$N(A) = \left\{\vec{x} \in \mathbb{R} : A\vec{x} = \vec{0}\right\}$$

In plain English, it is said that the nullspace is the set of vectors which by multiplication of A turn into the zero vector.

Finding the Nullspace: The nullspace can be obtained by solving the matrix $\left[A \mid \vec{0}\right]$, such that $A\vec{x} = \vec{0}$. The span of the solution to the matrix is the nullspace. The number of vectors in the span is the dimension of the matrix.

Range $R(A)$

Definition: Let A be an $m \times n$ matrix. The range of A is:

$$R(A) = \left\{ \vec{y} \in \mathbb{R} : A\vec{x} = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^M \right\}.$$

Finding the Range: Given a matrix A, use Gaussian elimination to bring it into row echelon form. The columns in the original matrix who have pivot entries in the REF matrix form the span of the range.

Rank Nullity Theorem

Let U be the row echelon form of A. We can generalize that:

$\dim(R(A))$: Number of columns in U with a leading nonzero pivot element (rank of the matrix).

$\dim(N(A))$: Number of columns in U without a leading nonzero pivot element.

Theorem: The rank-nullity theorem states that $\dim(R(A)) + \dim(N(A)) = n$ for an $n \times m$ matrix.

Implications of the Rank-Nullity Theorem

Theorem: Let $A = LU$ be the LU decomposition of A (if it exists), and let $r = \text{rank}(A)$.
Then, $R(A) = \text{span}\{\vec{l_1}, \cdots, \vec{l_r}\}$ where $\vec{l_n}$ are the first r columns of L.

Midterm 2

Inner Product

The inner product or dot product has some important properties:

- If two vectors are orthogonal, the inner product is zero.
- If two vectors are colinear, their inner product is the product of the magnitude of the two vectors.

Orthogonal Stuff

Orthogonal Sets: A set is orthogonal if all vectors in the set are orthogonal to each other.

Orthonormal Sets: A set is orthonormal if all vectors in the set are orthogonal to each other **and** the length of each vector is 1.

Remark: If vectors are orthogonal, they are linearly independent.

Orthogonal Complements

Orthogonal Subspaces: Two subspaces, S_1 and S_2 are are considered to be orthogonal subspaces if

$$\forall \vec{x} \in S_1, \vec{y} \in S_2, \langle \vec{x}, \vec{y} \rangle = 0.$$

Orthogonal Complements: An orthogonal complement of a matrix U, $U^\perp$ is the set of vectors:

$$U^\perp = \{\vec{x} \in \mathbb{R}^n : \langle \vec{x}, \vec{u} \rangle = 0, \forall \vec{u} \in U\}.$$

Remark: For a subspace $U \in \mathbb{R}^n$, $\dim(U) + \dim(U^\perp) = n$.

Finding an Orthogonal Complement: The orthogonal complement is simply the null space of the matrix. Hence, $U^\perp = N(A)$.

Properties of Orthogonal Complements:

$$N(A) = R(A^T)^\perp \qquad R(A) = N(A^T)^\perp$$
$$N(A)^\perp = R(A^T) \qquad R(A)^\perp = N(A^T)$$

Projections

The projection of a vector $\vec{x}$ onto another vector $\vec{u}$ is given by:

$$\text{proj}_{\vec{u}}(\vec{x}) = \frac{\langle \vec{x}, \vec{u} \rangle}{\langle \vec{u}, \vec{u} \rangle} \cdot \vec{u}.$$

Vector projections can also be given by matrix multiplication:

$$\text{proj}_{\vec{u}}(\vec{x}) = \frac{1}{\langle \vec{u}, \vec{u} \rangle} \vec{u}\vec{u}^T \vec{x}.$$

Hence, we can also define a projection matrix onto a vector, $\vec{u}$ as:

$$P = \frac{1}{||\vec{u}||^2} \vec{u}\vec{u}^T$$

The projection matrix has some interesting properties:

- $P^T = P$ (symmetric)

- $P^2 = P$
- $\text{rank}(P) = 1$

Projecting Vectors onto Subspaces: To project a vector onto a subspace U, we define $\text{proj}_U(\vec{x})$ as a linear combination of projections of $\vec{x}$ onto basis vectors of U.

We also would like to have an orthogonal basis of U, such that:

$$\vec{x} - \text{proj}_U(\vec{x}) \in U^\perp$$

Projecting onto the Orthogonal Complement: Let U be a subspace. The projection matrix $P^\perp$ onto the orthogonal complement of $U^\perp$ is given by:

$$P^\perp = I - P.$$

Projection Theorem: From the projection, we can state that

$$||\vec{x} - \text{proj}_U(\vec{x})|| = ||\text{proj}_{U^\perp}(\vec{x})|| \leq ||\vec{x} - \vec{u}||$$

for $\vec{u} \in U$. In other words, the vector $\text{proj}_U(\vec{x})$ is the vector in U that is closest to $\vec{x}$.

Constructing Orthogonal Bases

A set of vectors A is an orthogonal basis of a set of vectors, U, if

1. A is a basis of U.
2. A is an orthogonal set of vectors.

An important tool for constructing orthogonal bases is the Gram-Schmidt Algorithm. For a set of vectors $\{\vec{u_1}, \ldots, \vec{u_m}\}$, a basis of $U \subseteq \mathbb{R}^n$, the Gram-Schmidt Algorithm produces the following vectors:

$$\begin{aligned}
\vec{v}_1 &= \vec{u}_1 \\
\vec{v}_2 &= \vec{u}_2 - \text{proj}_{\vec{v}_1}(\vec{u}_2) \\
\vec{v}_3 &= \vec{u}_3 - \text{proj}_{\vec{v}_1}(\vec{u}_3) - \text{proj}_{\vec{v}_2}(\vec{u}_3) \\
&\vdots \\
\vec{v}_m &= \vec{u}_m - \text{proj}_{\vec{v}_1}(\vec{u}_m) - \cdots - \text{proj}_{\vec{v}_{m-1}}(\vec{u}_m)
\end{aligned}$$

Hence, it is implied that $\{\vec{v}_1, \ldots, \vec{v}_m\}$ is an orthogonal basis of U.

Furthermore, we can also define

$$\vec{w}_i = \frac{\vec{v}_i}{||\vec{v}_i||} i = 1, \ldots, m$$

Thus, we get that $\{\vec{w}_1, \ldots, \vec{w}_m\}$ is an orthnormal basis of U.

QR Decompositions

A QR decomposition on an $m \times n$ matrix with $\text{rank}(A) = n$ gives an alternate representation $A = QR$, which contains orthonormal bases of $R(A)$ and $R(A)^{\perp}$. This can be seen in $Q = [Q_1 Q_2]$, where Q_1 is the first n columns of Q, and is the orthonormal basis of $R(A)$, and the remaining columns form an orthonormal basis of $R(A)^{\perp}$.

It is also to be noted that Q is an orthogonal matrix with the same dimensions of the original matrix ($m \times n$) and R is an upper triangular matrix with the same dimensions of the original matrix.

Finding the QR Decomposition:

1. Obtain $R(A)$ and $R(A)^{\perp}$
2. Use the Gram-Schmidt algorithm to get an orthogonal basis of both $R(A)$ and $R(A)^{\perp}$, $\{\vec{w}\}$. These form Q_1 and Q_2, respectivley.
3. Construct R_1 given the following formula:

$$\begin{bmatrix} \langle \vec{w}_1, \vec{a}_1 \rangle & \langle \vec{w}_1, \vec{a}_2 \rangle & \cdots & \langle \vec{w}_1, \vec{a}_1 \rangle \\ 0 & \langle \vec{w}_2, \vec{a}_2 \rangle & \cdots & \vdots \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \vec{w}_n, \vec{a}_n \rangle \end{bmatrix}$$

4. For R_2, expand R_1 with zeroes until its dimensions are $m \times n$.

Just finding Q_1 and R_1 such that $A = Q_1 R_1$ is called the Thin QR Decomposition of the matrix A.

Least Squares Approximations

The least squares approximation is used for finding an exact solution for a system that has infinite solutions. In this, we seek to minimize $||A\vec{x} - \vec{b}||$.

By the projection theorem, we can visualize the least squares approximation as finding the vector in the range of our system $R(A)$ which is closest to $\vec{b}$. This is the orthogonal projection of $\vec{b}$ onto $R(A)$.

There exist two ways in which we can find a least squares solution to a problem:

1. Normal Equations: The least squares approximation to $A\vec{x} = \vec{b}$ is the solution to the normal equations system $A^T A\vec{x} = A^T\vec{b}$.
2. QR Decompositions: The least squares approximation to $A\vec{x} = \vec{b}$ is simply $R_1\vec{x} = Q_1^T\vec{b}$.

Remark: The QR decomposition method *apparently* has a lower condition number.

Eigenvalues and Eigenvectors

Most of this is review from earlier linear algebra courses but, there exist some important points:

- An eigenvalue is called defective if it is repeated in the characteristic equation.

Diagonalization

Another method for matrix decomposition is diagonalization. A matrix A is diagonalizable if you can bring it into the form:

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

Where for eigenvectors of the matrix A are $\vec{v}_k$ and eigenvalues for corresponding eigenvectors are λ_k, and

$$P = [\vec{v}_1 \cdots \vec{v}_n] \qquad D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Diagonalizable matricies have a special subcase, which is outlined in the spectral theorem. The spectral theorem states that: For a real, symmetric matrix A, there exists an orthogonal matrix P and a diagonal matrix D such that:

$$A = PDP^T.$$

Where P and D are defined in the same way as a normal diagonalizable matrix.

Singular Value Decomposition

Another decomposition that we can use is the singular value decomposition The singular value decomposition, or SVD, is similar to a diagonalization, but is different in some ways. Its form is given as:

$$A = P\Sigma Q^T$$

Where a singular value $\sigma_k = \sqrt{\lambda_k}$, for λ_k being an eigenvalue of the matrix AA^T, and where AA^T has r eigenvalues. Hence, we can describe each matrix as the following:

$$P = [\vec{p}_1 \cdots \vec{p}_r \ \vec{p}_{r+1} \cdots \vec{p}_m]_{m\times n} \qquad \vec{p}_k = \begin{cases} \frac{1}{\sigma_k} A\vec{q}_k & 0 \le k \le r \\ \text{orthonormal basis of } N(AA^T) & r < k \le m \end{cases}$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}_{m\times n} \qquad Q = [\vec{q}_1 \cdots \vec{q}_r \ \vec{0} \cdots \vec{0}]_{m\times n}$$

$$A^T A\vec{q}_k = \sigma_k^2 \vec{q}_k$$

Where σ_1 is the largest singular value, and σ_r is the smallest singular value.

Q can also be characterized as a matrix containing the eigenvectors of AA^T, in order of smallest to largest corresponding eigenvalue. However, these eigenvectors must be given as an orthonormal basis.

The "extension vectors" of P ($\vec{p}_k, k > r$) can also be characterized as the vectors in the nullspace of ($\vec{p}_k, k \le r$).

It is important to note that the elements of a SVD can be characterized as the following:

- P: Rotation/Reflection
- Σ: Stretch
- Q: Rotation/Reflection

Hence, the singular values provide us with important information on the norm and condition number of the matrix.

Thus, it can be said that $||A|| = \sigma_1$ and $||A^{-1}|| = \frac{1}{\sigma_n}$, so $\text{cond}(A) = \frac{\sigma_1}{\sigma_n}$.

Final Exam

Principle Component Analysis

Some matricies do not have inverses. Thus, we introduce the concept of a pseudoinverse, which builds off of the SVD to make something that's "good enough". Given some data matrix X which has dimensions $n \times p$, comprised of a set of vectors $\vec{x_1}, \ldots, \vec{x_n} \in \mathbb{R}^p$, assuming that the data is normalized such that

$$\sum_{k=1}^{n} \vec{x_n} = \vec{0}.$$

The Principle Component Analysis gives the vectors $\vec{w_q}, \ldots, \vec{w_n}$ Where $\vec{w}_1$ is the vector that maximizes $\sum_{k=1}^{n} ||\text{proj}_{\vec{w_1}}(\vec{x_k})||^2 = ||X\vec{w_1}||^2$

More generally, we can say that with the set of weight vectors $\{\vec{w_k}\}$, the k-th weight vector $\vec{w}_k$ is the unit vector which maximizes $||X_k\vec{w}_k||$, Where X_k is the projection of the data matrix onto span $\{\vec{w}_1, \ldots, \vec{w}_{k-1}\}$. This can also be written as:

$$X_k = X - \sum_{i=1}^{k=1} X\vec{w}_i\vec{w}_i^T$$

Hence, we can define the weight vectors as the right singular vectors of X, which are the columns of the Q matrix in the singular value decomposition.

$$\vec{w}_k = \vec{q}_k.$$

It's also important to note that each vector $\vec{w}_k$ is part of an orthonormal basis, thus for each $\vec{w}_k$, $||\vec{w}_k|| = 1$.

Pseudoinverses

Given an $m \times n$ matrix A, with SVD $A = P\Sigma Q^T$, the pseudoinverse of A, denoted A^+ is:

$$A^+ = Q\Sigma^+ P^T$$

Where Σ^+ is given by:

$$\Sigma^+ = \begin{bmatrix} \sigma_1^{-1} & & & \vdots \\ & \ddots & & 0 \\ & & \sigma_r^{-1} & \vdots \\ \cdots & 0 & \cdots & 0 \end{bmatrix}_{n \times m}$$

We can paraphrase this as Σ^+ being Σ, but with every singular value its reciprocal.

Thus, if A is invertible, then $A^+ = A^{-1}$.

The pseudoinverse has some interesting properties:

$$AA^+A = A \qquad\qquad A^+AA^+ = A^+.$$

The pseudoinverse also provides a new way to compute a least squares approximation:

$$\vec{x} = A^+\vec{b}.$$

An interesting implication of all things that we've done with the SVD is that we can take only the largest singular values in a SVD and get a result that has minimal data loss compared to our original matrix.

Complex Numbers

A complex number is of the form $z = a + ib$, where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$.

We can define some properties of a complex number z:

- Real Part The real part of z, is a, denoted $\text{Re}(z) = a$.
- Imaginary Part The imaginary part of z, is b, denoted $\text{Im}(z) = b$.
- Modulus The modulus, or length of z is given by $|z| = \sqrt{a^2 + b^2} = r$.
- Argument The argument, or angle of z is given by $\arg(z) = \arctan\left(\frac{b}{a}\right) = \theta$.

Given these definitions, we can express complex numbers in different ways:

- Euler's Formula: $z = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta)$.
- Polar Form: $z = re^{i\theta}$

We can also define the conjugate of z, denoted by $\overline{z}$ or z^*.

$$\overline{z} = a - ib = re^{-i\theta}.$$

Complex Vectors

Building off of the work done with complex numbers, we can now define complex vectors. A complex vector $\vec{v} \in \mathbb{C}^n$ is a vector with n complex entries.

Conjugates of Complex Vectors: We can define the conjugate of a complex vector, which for some complex vector $\vec{v}$ is given by:

$$\vec{v^*} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}^* = \begin{bmatrix} v_1^* \\ \vdots \\ v_n^* \end{bmatrix}$$

To paraphrase, a conjugate of a complex vector is just the original matrix with all of its entries conjugated.

Complex Inner Product: We can also define the inner product for two complex vectors $\vec{v}, \vec{w} \in \mathbb{C}^n$:

$$\langle \vec{v}, \vec{w} \rangle = \vec{v_1^T}\vec{w^*} = \langle \vec{v}, \vec{w^*} \rangle.$$

The complex inner product has some nice properties too:

- $\langle c\vec{v}, \vec{w} \rangle = c\langle \vec{v}, \vec{w} \rangle$
- $\langle \vec{v}, c\vec{w} \rangle = c\langle \vec{v}, \vec{w} \rangle$
- $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle^*$
- $\langle \vec{v}, \vec{v} \rangle = |v_1|^2 + \cdots + |v_n|^2$.

Hence, with the last property, we can say that the norm of a complex vector can be expressed by:

$$||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$$

We can also define another property for an $m \times n$ complex matrix, A, and two complex vectors $\vec{v} \in \mathbb{C}^m, \vec{w} \in \mathbb{C}^n$:

$$\langle A\vec{v}, \vec{w} \rangle = \langle \vec{v}, A^{T*}\vec{w} \rangle.$$

Hermetian Matricies: We call a complex matrix Hermetian if $A = A^{T*} = \overline{A^T}$

If A is Hermetian, then:

- $\langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A\vec{y}\rangle$.
- A has only real eignevalues and A is diagonalizable.
- The diagonal entries of A are real numbers.

Roots of Unity

We define a complex number $\omega in \mathbb{C}$ as an Nth root of unity if $\omega^N = 1$.
Generally, these are given in the form $\omega_n = e^{i\frac{2\pi}{N}}$.

Hence, we can define some important properties on roots of unity ω.

1. $\omega_N^N = 1$
2. $\omega_N^{N-1} = \overline{\omega_N} = \omega_N^*$
3. $\overline{\omega_N} = \omega_N^{-1}$
4. $\sum_{k=0}^{N-1} \omega_N^{kl} = 0, 0 < l < N$

Fourier Bases

As seen prior, a standard basis of $\mathbb{C}^n$ is $\vec{e}_0, \ldots, \vec{e}_{N-1}$ where $\vec{c}_k$ is given by:

$$\vec{e}_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Where the 1 is at entry k.

However, now we introduce the concept of a Fourier basis of $\mathbb{C}^n$, given by $\vec{f}_0, \ldots, \vec{f}_{N-1}$, where

$$\vec{f}_k = \begin{bmatrix} 1 \\ \omega_N^k \\ \omega_N^2 k \\ \vdots \\ \omega_N^{(N-1)k} \end{bmatrix}$$

Where $\omega_N = e^{i\frac{2\pi}{N}}$, and $n = N$.

Remark: A Fourier basis is an orthogonal basis of $\mathbb{C}^n$, where

$$\langle \vec{f}_k, \vec{f}_n \rangle = \begin{cases} N & k = n \\ 0 & k = n \end{cases}$$

Remark: $\vec{f}_k^* = \vec{f}_{N-k}$ for $0 < k < N$.

Discrete Fourier Transform

The Discrete Fourier Transform, or DFT, is the vector coefficients of $\vec{x}$ with respect to a Fourier basis. Hence, we can define the DFT as:

$$\text{DFT}(\vec{x}) = F_N\vec{x} = \begin{bmatrix} \vec{f_0^*}^T \\ \vdots \\ \vec{f_{N-1}^*}^T \end{bmatrix} = \begin{bmatrix} \vec{f_0}^T \\ \vec{f_{N-1}}^T \\ \vdots \\ \vec{f_1}^T \end{bmatrix}$$

Due to the properties of complex conjugates, a nice symmetry emerges from a Fourier transform. Given some $\vec{x} \in \mathbb{R}^n$ and $\vec{y} = \text{DFT}(\vec{x})$, then $\vec{y}[k] = \vec{y}[N-k]$ for $1 \leq k \leq N-1$ or $0 < k < N$.

Finally, we can also reconstruct $\vec{x}$ from $\text{DFT}(\vec{x})$, since F_N is invertible. Thus, we can obtain the inverse of F_N as

$$F_N^{-1} = \frac{1}{N}{F_N^*}^T.$$

Hence, the inverse discrete Fourier transform of $\vec{y} \in \mathbb{C}^N$ is:

$$\text{IDFT}(\vec{y}) = \frac{1}{N}{F_N^*}^T\vec{y}.$$

Sinusoids

If we divide a time range N into discrete time points such that:

$$\vec{t} = \begin{bmatrix} 0 \\ \frac{1}{N} \\ \frac{2}{N} \\ \vdots \\ \frac{N-1}{N} \end{bmatrix}.$$

Hence, we can also use this to turn functions into discretized vectors. For example, given a function $f(x)$, then $f(\vec{t})$ is given by:

$$f(\vec{t}) = \begin{bmatrix} f(0) \\ f\left(\frac{1}{N}\right) \\ f\left(\frac{2}{N}\right) \\ \vdots \\ f\left(\frac{N-1}{N}\right) \end{bmatrix}.$$

Hence for a sinusoid of the form:

$$f(t) = A\cos(2\pi k\vec{t} + \phi)$$

Where:

- A is the amplitude.
- k is the frequency.
- ϕ is the phase shift.

We can express sinusoids using the Fourier Basis:

- $\vec{f_k} = \cos(2\pi k\vec{t}) + i\sin(2\pi k\vec{t})$
- $\frac{1}{2}\left(\vec{f_k} + \vec{f_k}^*\right) = \cos(2\pi k\vec{t})$
- $\frac{1}{2i}\left(\vec{f_k} - \vec{f_k}^*\right) = \sin(2\pi k\vec{t})$

Finally, we can define the Discrete Fourier transform for a cosine function:

$$\text{DFT}(A\cos(2\pi k\vec{t} + \phi)) = \frac{AN}{2}e^{i\phi}\vec{e}_k + \frac{AN}{2}e^{-i\phi}\vec{e}_{N-k}$$

Stemplots and Reconstructions

Given a Discrete Fourier Tranform output, $\vec{y} \in \mathbb{C}^n$, we can obtain a plot of it's frequencies in the form of a stemplot. To obtain a stemplot, we divide the modulus of each entry by N if the entry is at index 0 or $\frac{N}{2}$, and if it is not at either of these entries, then we divide the modulus by $\frac{N}{2}$. This completes a magnitude stemplot. For a angle plot, we plot the argument of each entry. If the entry is positive and real, the argument or angle is 0. If the entry is negative and real, then the argument is π.
In order to reconstruct the signal as a sum of sinusoids, we can take every entry of the Fourier transform and take the stemplot amplitude as the amplitude of the signal A, the index of the entry k as the frequency, and the argument ϕ as the phase shift. Hence, we can reconstruct the signal as a sum of cosines. An important remark is that we shall only do this up until the $N/2$th entry.