CHAPTER 4 Vector Spaces

4.1 Vector Spaces and the Invertible Matrix Theorem

Let A be an $m \times n$ matrix $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$

The **column space** of A is the span of its columns: $\text{Col}(A) = \text{Span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$. It's dimension is called the **rank** of A.

The **null space** of A is all vectors solving the equation $A\boldsymbol{x} = \boldsymbol{0}$.

The sum of their dimensions must equal the size of A: $\text{Dim}(\text{Col}(A)) + \text{Dim}(\text{Nul}(A)) = n$.

A will map any vector from $\mathbb{R}^n$ onto some space $\mathbb{R}^k$, where $k = \text{Rank}(A)$. The dimensions of the column space therefor decides the dimension of the space we map onto. The null space then represents all the dimensions that "disappear" during the transform.

The Invertible Matrix Theorem

- A is invertible.
- The columns of A are linearly independent, and A spans $\mathbb{R}^n$.
- A is row-equivalent to the identity matrix I_n.
- A has n pivot columns.
- The dimensions of A's null space is zero. ($A\boldsymbol{x} = \boldsymbol{x}$ has only the solution $\boldsymbol{x} = \boldsymbol{0}$).
- 0 is not an eigenvalue of A.
- The determinant of A is different from zero.
- A^T is also invertible, and all of this applies to it.

4.2 Coordinate systems and mapping

Consider a vector $\boldsymbol{x}$ living in a vector space V. The vector $\boldsymbol{x}$ is an abstract concept, living in some abstract space V. It may have some physical or gemoetric meaning or whatnot.

We now enforce a *basis* onto V, called $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$. This makes V *behave like* $\mathbb{R}^n$, in the sense that each vector $\boldsymbol{x}$ in V is *mapped onto* a vector $[\boldsymbol{x}]_\mathcal{B}$ in $\mathbb{R}^n$. This is called a *coordinate mapping* $\boldsymbol{x} \mapsto [\boldsymbol{x}]_\mathcal{B}$ "onto" the basis $\mathcal{B}$. The vector space V might be foreign to us, and it can be important to create a mapping onto a more familiar vector space $\mathbb{R}^n$, which we know how behaves. This transformation is "one-to-one", mapping each point in V onto a point in $\mathbb{R}^n$, and vice versa. This relation is called an **isomorphism**, and makes any vector space V with a basis of n vectors indistinguishable from $\mathbb{R}^n$.

Usually, when a vector is written plainly as $\boldsymbol{x}$, we consider it to be written in a *standard basis* $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, meaning that $\boldsymbol{x} = [\boldsymbol{x}]_\mathcal{E}$.

The relation between $\boldsymbol{x}$ and $[\boldsymbol{x}]_\mathcal{B}$ is given by a **change-of-basis matrix** $P_\mathcal{B}$, which consists of the basis-vectors of $\mathcal{B}$, written in the basis of $\mathcal{E}$:

$$\boldsymbol{x} = P_\mathcal{B}[\boldsymbol{x}]_\mathcal{B} \qquad P_\mathcal{B} = [\boldsymbol{b}_1 \ \boldsymbol{b}_2 \ \ldots \ \boldsymbol{b}_n]$$

4.3 Change of basis

This change of basis is just a special case of a more general change of basis between two basises $\mathcal{B} = \{\boldsymbol{b}_1, \ldots \boldsymbol{b}_n\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \ldots \boldsymbol{c}_n\}$, both spanning the same vector space V. The general change of basis is then

$$[\boldsymbol{x}]_\mathcal{C} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\boldsymbol{x}]_\mathcal{B} \qquad \underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \Big[[\boldsymbol{b}_1]_\mathcal{C} \ [\boldsymbol{b}_2]_\mathcal{C} \ \ldots [\boldsymbol{b}_1]_\mathcal{C}\Big]$$

The change-of-basis matrix from $\mathcal{C}$ to $\mathcal{B}$ is simply the inverse: $\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = \left(\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}\right)^{-1}$

The change of basis of a linear transformation is

$$[T]_{\mathcal{C}} = P\,[T]_{\mathcal{B}}\,P^{-1}$$

where P is the change-of-basis matrix $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$

4.4 Linear transformations (mappings) between vector spaces

Any linear transformation is solely defined by how it acts on the unit vectors.

Consider two vector spaces V and W, with basises $\mathcal{B} = \{\boldsymbol{b}_1, \ldots \boldsymbol{b}_n\}$ and $\mathcal{C} = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m\}$ in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. We introduce a *linear transformation* $T : V \mapsto W$ such that $T(\boldsymbol{x}) = A\boldsymbol{x}$. This is all well and good, but we might only have the vector $\boldsymbol{x}$ represented in the basis $\mathcal{B}$, and usually want it written in the basis $\mathcal{C}$ after the transformation, as $\big[T(\boldsymbol{x})\big]_{\mathcal{C}}$.

What we want is some matrix M that carries us straight from $[\boldsymbol{x}]_{\mathcal{B}}$ to $\big[T(\boldsymbol{x})\big]_{\mathcal{C}}$. If we combine the change of basis with T, we get

$$\big[T(\boldsymbol{x})\big]_{\mathcal{C}} = M[\boldsymbol{x}]_{\mathcal{B}}$$

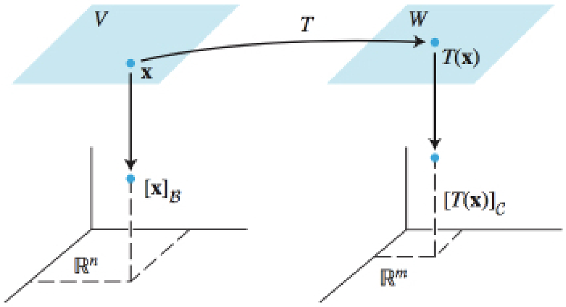

FIGURE 1 A linear transformation from V to W.

where

$$M = \Big[[T(\boldsymbol{b}_1)]_{\mathcal{C}} \;\ldots\; [T(\boldsymbol{b}_n)]_{\mathcal{C}}\Big] = \Big[[A\boldsymbol{b}_1]_{\mathcal{C}} \;\ldots\; [A\boldsymbol{b}_n]_{\mathcal{C}}\Big]$$

This matrix is called the **matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$.**

If M is invertible, there exists an inverse linear transform T^{-1} with matrix representation M^{-1}, mapping every point in $\mathcal{B}$ to a point in $\mathcal{C}$. This makes T an **isomorphism** (one-to-one mapping).

CHAPTER 5 Eigenvalues and Eigenvectors

5.1 Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ in decreasing order, and corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$.

- The eigenvalues of A is given by **the characteristic equation** $\mathrm{Det}(A - \lambda I) = 0$.
- If A is triangular, the eigenvalues are the entries of it's diagonal.
- If A has n distinct eigenvalues, the eigenvectors are linearly independent. If A has degenerate eigenvalues, we have to check whether the eigenvector are independent or not.

5.2 Diagonalization

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if it has n linearly independent eigenvectors.

We can then write

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

where D is a diagonal matrix and P is an invertable matrix such that

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \qquad P = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix}$$

5.3 Discrete dynamical systems

Cosider a difference equation on the form

$$\boldsymbol{x}_{i+1} = A\boldsymbol{x}_i \qquad \Rightarrow \qquad \boldsymbol{x}_k = A^k \boldsymbol{x}_0$$

where $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots\}$ is a vector sequence describing the properties of some system, and A is an $n \times x$ invertable matrix, giving it n descrete eigenvalues, with n orthogonal eigenvectors spanning $\mathbb{R}^n$. This means any vector x_i can be written as a linear combination of these:

$$x_i = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$$

such that the matrix-vector multiplication $A^k\boldsymbol{x}_0$ can be rewritten as

$$x_k = A^k x_0 = A^k(c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n) = (\lambda_1)^k c_1\boldsymbol{v}_1 + \cdots + (\lambda_n)^k c_n\boldsymbol{v}_n$$

In short, any such dynamical system can be decomposed into the eigenvectors of A, where the eigenvalues decides how the system behaves over time.

The coefficients are usually found as

$$\boldsymbol{c} = P^{-1}\boldsymbol{x}$$

5.3.1 Attractors

We see that if all eigenvalues of A are less than 1, the system will tend towards $\mathbf{0}$, and we say that $\mathbf{0}$ is an **attractor** for the system.

5.4 Complex Eigenvalues

Let A be a real 2×2 matrix with a complex eigenvalue $\lambda = a - bi$, and an associated eigenvector $\boldsymbol{v}$. Then

$$A = PCP^{-1}, \quad \text{where} \quad P = \begin{bmatrix} \operatorname{Re}(\boldsymbol{v}) & \operatorname{Im}(\boldsymbol{v}) \end{bmatrix} \quad \text{and} \quad C = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

5.5 Applications to Differential Equations

Consider the system of equations

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n \\ &\vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n \end{aligned}$$

where x_i are differentiable functions of t.

This problem can be written as the matrix problem

$$\boldsymbol{x}'(t) = A\boldsymbol{x}(t)$$

where we need to solve it for some known t.

CHAPTER 6 Orthogonality and least squares

6.1 Orthogonal Matrixes

Consider an $m \times n$ matrix U with orthonormal columns. We call this an orthonormal matrix. This matrix has the properties:

- U is orthonormal only if $U^T U = I$.
- U preserves length: $||U\boldsymbol{x}|| = ||\boldsymbol{x}||$.
- If U is square, we have that $U^{-1} = U^T$.

6.2 Projections

6.2.1 Projection of vector onto vector

The projection of a vector $\boldsymbol{y}$ onto another vector $\boldsymbol{x}$ is

$$\widehat{\boldsymbol{y}} = \text{proj}_{\boldsymbol{x}}(\boldsymbol{y}) = \frac{\boldsymbol{y} \cdot \boldsymbol{x}}{\boldsymbol{x} \cdot \boldsymbol{x}} \boldsymbol{x}$$

6.2.2 Projection of vector onto subspace

Let W be subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ and $\boldsymbol{y}$ be any vector in $\mathbb{R}^n$. Then the projection of $\boldsymbol{y}$ onto W is simply the projection onto each basis-vector:

$$\widehat{\boldsymbol{y}} = \text{proj}_W(\boldsymbol{y}) = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p} \boldsymbol{u}_p$$

If you have an <u>orthonormal</u> basis, the dot products reduces to 1, and this can be written.

$$\widehat{\boldsymbol{y}} = \text{proj}_W(\boldsymbol{y}) = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{y} \cdot \boldsymbol{u}_n)\boldsymbol{u}_n = UU^T \boldsymbol{y}$$

where $U = [\boldsymbol{u}_1, \ldots \boldsymbol{u}_n]$ contains the basis vectors.

6.3 The Gram-Schmidt process of orthogonal factorization

The Gram-Schmidt process takes any basis of vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p\}$ spanning a subspace in $\mathbb{R}^n$, and creates a new *orthogonal* basis of the same space, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$.

The idea is to create one and one new vector $\boldsymbol{v}_i$ from the corresponding $\boldsymbol{x}_i$, *but* subtract the projection of $\boldsymbol{x}_i$ onto each of the former vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{i-1}$, such that the new vector is orthogonal to all formerly created vectors.

- $\boldsymbol{v}_1 = \boldsymbol{x}_1$
- $\boldsymbol{v}_2 = \boldsymbol{x}_2 - \dfrac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1$
- $\boldsymbol{v}_3 = \boldsymbol{x}_3 - \dfrac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \dfrac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2} \boldsymbol{v}_2$
- $\ldots$

Remember to normalize afterwards if you need an *orthonormal* basis.

6.4 QR-factorization

If the columns of any $m \times n$ matrix A are linearly independent, it can be QR-factorized:

$A = QR$

- $Q = [\boldsymbol{u}_1\ \boldsymbol{u}_2 ... \boldsymbol{u}_n]$ where $\boldsymbol{u}_i$ is an orthonormal basis for the columnspace of A. Simply use Gram-Schmidt on the columns of A, and normalize them.
- $R = Q^{-1}A = Q^T A$, because $Q^T = Q^{-1}$ due to orthonormal columns. R will always be upper diagonal.

6.5 Least-sqaure problems

Let W be the column space of some matrix A: $W = \text{span}\{\text{col}(A)\}$.

Sometimes, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has no solutions, because $\boldsymbol{b}$ is not in W. We then often wish to find the solutions $\widehat{\boldsymbol{x}}$ which places the solutions $A\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}}$ as close to the actual solution $\boldsymbol{b}$ as possible.

This "closest solution" is the projection of $\boldsymbol{b}$ onto W, $\widehat{\boldsymbol{b}} = \text{proj}_W \boldsymbol{b}$.

The least-square solution to $A\boldsymbol{x} = \boldsymbol{b}$ is then:

$$A\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}} \quad \Rightarrow \quad \widehat{\boldsymbol{x}} = A^{-1}\widehat{\boldsymbol{b}} \qquad (= A^{-1}AA^T\boldsymbol{b})$$

If we have a QR factorization of A, we can solve the least square problem as

$$\widehat{x} = R^{-1}Q^T\boldsymbol{b}$$

6.5.1 Applications to linear models

Let's say we wish to fit some data $(x_1, y_1), (x_2, y_2), \ldots$ to a linear model $y(x) = \beta_0 + \beta_1 x$.

Predicted y**-value**		**Observed** y**-value**
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

We can write this system as

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

which produces a least-square solution $X\widehat{\beta} = \widehat{y}$.

CHAPTER 7 Symmetric Matrices and Quadratic Forms

7.1 Diagonalization of symmetric Matrices

A symetric matrix A is such that $A^T = A$.

If A is symetric, any two eigenvectors from different eigenspaces are orthogonal.

A is then orthogonally diagonalizable, such that $A = PDP^{-1} = PDP^T$, where P consists of A's orthogonal eigenvectors.

7.2 Quadratic Forms

Consider a function $Q : \mathbb{R}^n \mapsto \mathbb{R}$ given by $Q(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}$ called a **quadratic form**, defined by a symmetric $n \times n$ matrix A.
The result will then be a second order polynomial

$$c_{00}\, x_1^2 + c_{11}\, x_2^2 + c_{12}\, x_1 x_2 + ...$$

The matrix will then be a form

$$A = \begin{pmatrix} c_{00} & \frac{1}{2}c_{12} & \frac{1}{2}c_{13} & \cdots \\ \frac{1}{2}c_{12} & c_{11} & & \\ \vdots & & \ddots & \\ & & & c_{nn} \end{pmatrix}$$

The non-diagonals are halved, because they "share" the coefficients due to the symetry.

> If the eigenvalues of A are
> - All positive, $Q(\boldsymbol{x})$ takes only positive values in $\mathbb{R}$.
> - All negative, $Q(\boldsymbol{x})$ takes only negative values in $\mathbb{R}$.

7.2.1 Change of variable

Let $A = PDP^{-1}$. We can then substitute $\boldsymbol{x} = P\boldsymbol{y}$, giving

$$\boldsymbol{x}x^T A \boldsymbol{x} = \boldsymbol{y}^t D y$$

where D is diagonal.

7.3 Constraint optimization

We often wish to solve for the $\boldsymbol{x}$ that gives the maximum or minimum $Q(\boldsymbol{x})$ under some constraint $||\boldsymbol{x}|| = ||\boldsymbol{x}^2|| = \boldsymbol{x}^T \boldsymbol{x} = 1$.

- The maximum value of $Q(\boldsymbol{x})$ is the largest eigenvalue of A. The $\boldsymbol{x}$ giving this value is the corresponding eigenvector.
- The minimum value of $Q(\boldsymbol{x})$ it the smallest eigenvalue of A. The $\boldsymbol{x}$ giving this value is the corresponding eigenvector.

7.4 Singular Value Decomposition

We know the $A = PDP^{-1}$ diagonalization of A can only be applied to $n \times n$, diagonalizable matrices A.

However, the **singular value decomposition** $A = U\Sigma V$ can be applied to *any* matrix A. It reduces to PDP^{-1} factorization if A is diagonalizable.

- The singular values of A are the squareroots of the eigenvalues of $A^T A$, in decreasing order:

$$\sigma_i = \sqrt{\lambda_i}$$

- The singular vectors of A are the normalized eigenvectors of $A^T A$, ordered by decending eigenvalues.

Let A be some $m \times n$ matrix with rank r. Then A can be factorized into $A = U\Sigma V$ where

- Σ is a $m \times n$ matrix (same shape as A), where the diagonal values are all the singular values of A, with trailing zeros.

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- V is an orthogonal $n \times n$ matrix, consisting of the singular vectors of A:

$$V = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \end{pmatrix}$$

- U is an orthogonal $m \times m$ matrix where the columns are the normalized Av_i vectors:

$$U = \begin{pmatrix} \frac{A\boldsymbol{v}_1}{\sigma_1} & \frac{A\boldsymbol{v}_2}{\sigma_2} & \frac{A\boldsymbol{v}_3}{\sigma_3} \end{pmatrix}$$

  If there aren't m non-zero vectors to use, construct the rest of U as orthonormal vectors from Gram-Schmidt.

CHAPTER 8 Matlab

Poly(A)

Returns the negative of the characteristic polynomial of A.
Ex:

$$\begin{aligned} \text{INPUT:} \quad & \text{poly}(A) \\ \text{OUTPUT:} \quad & 1\ -4\ 3\ 0 \\ \Rightarrow \quad & -(\lambda^3 - 4\lambda^2 + 3\lambda + 0) = 0 \end{aligned}$$