Analytic Geometry 1/2

Affine Spaces

Affine Subspace

Let V be a vector space, $\mathbf{x}_0 \in V$ and $U \subseteq V$ a subspace.

- Then the subset

$$
\begin{aligned}
L = \mathbf{x}_0 + U &:= \{\mathbf{x}_0 + \mathbf{u} : \mathbf{u} \in U\} && (1)\\
&= \{\mathbf{v} \in V \mid \exists \mathbf{u} \in U : \mathbf{v} = \mathbf{x}_0 + \mathbf{u}\} && (2)
\end{aligned}
$$

  is called *affine subspace* or *linear manifold* of V.

- U is called *direction* or *direction space*, and $\mathbf{x}_0$ is called support point.

Affine Spaces

- An affine space is not necessarily a vector subspace.
  - Unless $\mathbf{x}_0 \in U$
  - If $\mathbf{x}_0 \notin U$ then L does not include $\mathbf{0}$
  - Why?

$v = x_0 + u$. For $v = 0$ $\Rightarrow u = -x_0$. Since vector spaces are closed $\Rightarrow x_0 \in U$

- An affine space is often discussed in terms of a *frame* in computer graphics
- Affine subspaces are often represented as a *parametric equation*:
  - Let $L = \mathbf{x}_0 + U$ be a k-dimensional affine space, where $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$ is an ordered basis of U
  - The every element $\mathbf{x} \in L$ can be uniquely described as

$$\mathbf{x} = \mathbf{x}_0 + \lambda_1 \mathbf{b}_1 + \cdots + \lambda_k \mathbf{b}_k \tag{3}$$

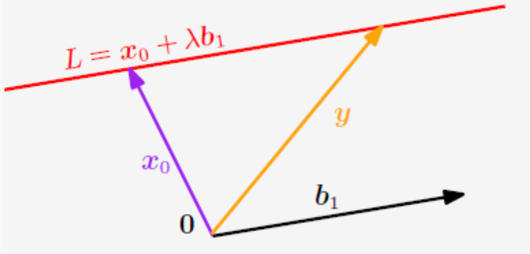
with each $\lambda_j \in \mathbb{R}$.

Affine Spaces: Example

- One-dimensional affine subspaces are called *lines* and can be written line as

$$\mathbf{y} = \mathbf{x}_0 + \lambda \mathbf{b}_1$$

  - where $\lambda \in \mathbb{R}$ and $U = span[\mathbf{b}_1] \subseteq \mathbb{R}^n$ where U is a one dimensional subspace of $\mathbb{R}^n$.



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

This idea can be directly extend in $\mathbb{R}^n$ to $(n-1)$-dimensional affine subspaces are called *hyperplanes*

Affine Spaces: Inhomogeneous systems of linear equations

$Ax = 0$
↳ Null Space
↳ Vector Space

Consider the system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^m$.

- The solution of the system is either the **empty set** or an **affine subspace** of $\mathbf{R}^n$ of dimension $n - rk(\mathbf{A})$.

Affine Mapping

For two vector spaces V, W, a linear mapping $\Phi : V \to W$ and $\mathbf{a} \in W$, the mapping

$$\begin{aligned} \phi : V &\to W && (5) \\ \mathbf{x} &\mapsto \mathbf{a} + \Phi(\mathbf{x}) && (6) \end{aligned}$$

is an affine mapping from V to W. The vector $\mathbf{a}$ is called the translation affine mapping vector of ϕ

Affine Mapping: Properties

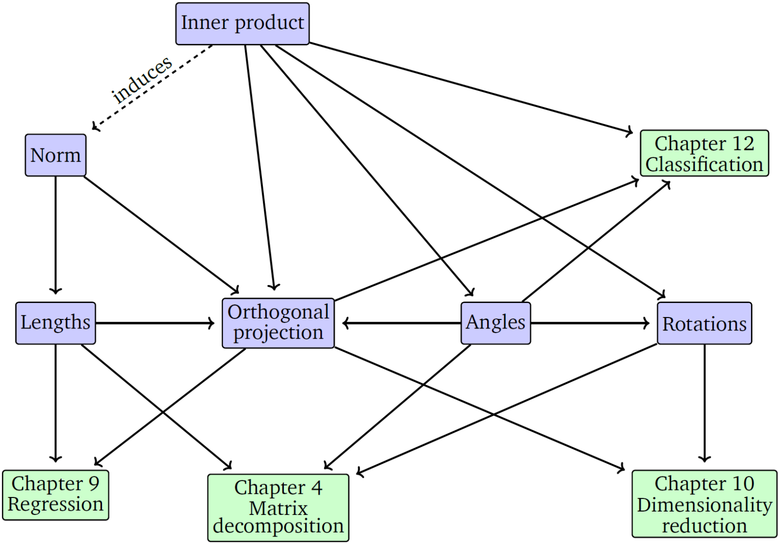
Many of the useful properties of linear mappings are still present with affine mapping

- The composition $\phi' \circ \phi$ of an affine mappings $\phi : V \rightarrow W$, $\phi' : W \rightarrow X$ is also affine
- Affine mappings keep the geometric structure invariant.
- Affine mappings also preserve the dimension and parallelism.

An affine mapping can also constructed from composing

- A linear mapping $\Phi : V \rightarrow W$ with
- A translation $\tau : W \rightarrow W$
- Namely: $\tau \circ \Phi$ where both Φ and τ are uniquely determined.

Analytic Geometry Mindmap



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

Norms

Norm

A norm on a vector space V is a function

$$\begin{aligned} \|\cdot\| : & V \to \mathbb{R}, && (7) \\ & \mathbf{x} \mapsto \|\mathbf{x}\|_2 && (8) \end{aligned}$$

which assigns each vector $\mathbf{x}$ its *length* $\|\mathbf{x}\| \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$ the following hold:

- Absolutely homogeneous: $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
- Triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
- Positive definite: $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$

$z = x + y$

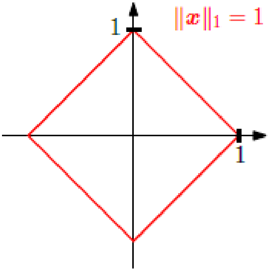
Norms: Example 1

Manhattan norm

The Manhattan norm on $\mathbb{R}^n$ as

$$\|\mathbf{x}\| := \sum_{i=1}^{n} |x_i| \tag{9}$$

where $|\cdot|$ is the absolute value. It is also called the ℓ_1 norm.



Source: M.P. Deisenroth *et al*, Mathematics for Machine Learning (First Edition)

Norms: Example 2

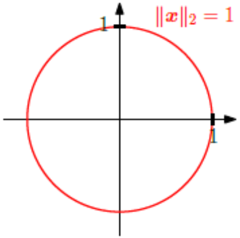
Euclidean norm

The Euclidean norm on $\mathbb{R}^n$ as

$$\|\mathbf{x}\| := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^T \mathbf{x}} \tag{10}$$

and computes the Euclidean distance of $\mathbf{x}$ from the origin. It is also called the ℓ_2 norm.

$\|x\|_p = (\sum |x_i|^p)^{1/p}$



Inner Products

The example of an inner product you have likely encountered is the *dot product*:

$$\mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i \tag{11}$$

In order to provide a compact definition of a general inner product a couple of preliminary concepts are needed.

Bilinear mapping

$\Omega : V \times V \to U$

Let V be a vector space the Ω is a *Bilinear mapping* if for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $\lambda, \psi \in \mathbb{R}$ that

$$\Omega(\lambda\mathbf{x} + \psi\mathbf{y}, \mathbf{z}) = \lambda\Omega(\mathbf{x}, \mathbf{z}) + \psi\Omega(\mathbf{y}, \mathbf{z}) \tag{12}$$

$$\Omega(\mathbf{x}, \lambda\mathbf{y} + \psi\mathbf{z}) = \lambda\Omega(\mathbf{x}, \mathbf{y}) + \psi\Omega(\mathbf{x}, \mathbf{z}) \tag{13}$$

Equation (12) asserts that Ω is linear in the first argument and Equation (13) asserts that Ω is linear in the second argument.

$$\Omega(\lambda x + \psi y, z) = \Omega(\lambda x, z) + \Omega(\psi y, z)$$
$$= \lambda\Omega(x, z) + \psi\Omega(y, z)$$

General Inner Products

Symmetric and positive definite

Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ a bilinear mapping. Then

- Ω is called *symmetric* if $\Omega(\mathbf{x}, \mathbf{y}) = \Omega(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in V$
- Ω is called *positive definite* if

$$\forall \mathbf{x} \in V \setminus \{\mathbf{0}\} \; : \Omega(\mathbf{x}, \mathbf{x}) > 0, \; \Omega(\mathbf{0}, \mathbf{0}) = 0 \tag{14}$$

General Inner Products

Inner product and inner product space

Let V be a vector space the $\Omega : V \times V \to \mathbb{R}$ be a Bilinear mapping. Then

- A positive definite symmetric bilinear mapping $\Omega : V \times V \to \mathbb{R}$ is called a inner product.
  - Typically we denote $\Omega(\mathbf{x}, \mathbf{y})$ as $\langle \mathbf{x}, \mathbf{y} \rangle$.
- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.
  - If we use the dot product defined in equation (11), $(V, \langle \cdot, \cdot \rangle)$ it is called the Euclidean vector space.

General Inner Products

There are a number of inner products on $\mathbb{R}^n$ that are not the dot product, for example, consider the following for $\mathbb{R}^2$

$$\langle \mathbf{x}, \mathbf{y} \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2 x_2 y_2 \tag{15}$$

- Show that equation (15) is in fact an inner product

Symmetric, Positive Definite Matrices

Symmetric, positive definite matrices and the related theory

- provides vital formalism for convex optimization and by extension
- many aspects of machine learning.

Relationship Between Inner Products and Matrices

There exists a direct relationship between an inner product on a finite dimensional vector space and a matrix operation. Specifically

- Let V be a n-dimensional vector space, with the inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, and the ordered basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$.
- Then for all $\mathbf{x}, \mathbf{y} \in V$ it follows that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \overbrace{\sum_{i=1}^{n} \psi_i \mathbf{b}_i}^{= x}, \overbrace{\sum_{j=1}^{n} \lambda_j \mathbf{b}_j}^{= y} \rangle \tag{16}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i \langle \mathbf{b}_i, \mathbf{b}_j \rangle \lambda_j \tag{17}$$

$$= \hat{\mathbf{x}}^T \mathbf{A} \hat{\mathbf{y}} \tag{18}$$

where $A_{ij} := \langle \mathbf{b}_i, \mathbf{b}_j \rangle$ and $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ are the coordinates of $\mathbf{x}$ and $\mathbf{y}$ with respect to the basis B.

Relationship Between Inner Products and Matrices

What can we now directly say about $\mathbf{A}$?

- $\mathbf{A}$ is symmetric, because $\langle \mathbf{b}_i, \mathbf{b}_j \rangle = \langle \mathbf{b}_j, \mathbf{b}_i \rangle$
- We have that

$$\forall \mathbf{x} \in V \setminus \{\mathbf{0}\} \; : \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \tag{19}$$

  because the inner product is positive definite

Symmetric, Positive Definite Matrix

Symmetric, Positive Definite Matrix

A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ that satisfied equation (19) *symmetric, positive definite.*

- If we do not have the strict inequality and only $\geq$ on equation (19) we say that the matrix is *symmetric, positive semidefinite.*

Symmetric, Positive Definite Matrix

Theorem 3.5.

For a real-valued, finite-dimensional vector space V and an ordered basis B of V, it holds that $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is an inner product **if and only if** there exists a symmetric, positive definite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with

$$\langle \mathbf{x}, \mathbf{y} \rangle = \hat{\mathbf{x}}^T \mathbf{A} \hat{\mathbf{y}} \tag{20}$$

Symmetric, Positive Definite Matrix

Two further properties of a symmetric positive definite matrix

- The null space (kernel) of A consists only of $\mathbf{0}$.
  - This can seen by noting that since $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$
    - $\implies$ that $\mathbf{A}\mathbf{x} \neq \mathbf{0}$ if $\mathbf{x} \neq 0$
  - The diagonal elements a_{ii} of $\mathbf{A}$ are positive.
    - Observe that if we use $\mathbf{e}_i$, the ith vector of the standard basis in $\mathbb{R}^n$, then we see that $a_{ii} = \mathbf{e}_i^T\mathbf{A}\mathbf{e}_i > 0$

Lengths and Distances

We have already discussed norms, but there is a special class of norms that are induced by inner products. Specifically,

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{21}$$

- All inner products induce a norm
- Not all norms are induced by a inner product.

A norm is induced from an inner product if following *parallelogram equality* holds

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2) \tag{22}$$

Induced Norms

Cauchy-Schwarz inequality

For an inner product vector space $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ the induced norm $\| \cdot \|$ satisfies the *Cauchy-Schwarz inequality*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\| \tag{23}$$

Distance and Metric

Metric

If X is set then the mapping

$$d : X \times X \to \mathbb{R} \tag{24}$$
$$(\mathbf{x}, \mathbf{y}) \mapsto d(\mathbf{x}, \mathbf{y}) \tag{25}$$

is a *metric* (and called a distance) if it satisfies

1. d is *positive definite*: $d(\mathbf{x}, \mathbf{y}) \geq 0 \ \forall \mathbf{x}, \mathbf{y} \in V$ and $d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$.
2. d is *symmetric*: $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in V$.
3. *Triangle inequality*: $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$.

We do not need to be this general in this course and will restrict ourselves to X as a vector space as apposed to an arbitrary set.

Distance and Metric

We can use a norm as a metric on a vector space V

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| \tag{26}$$

If V is a inner product space $(V, \langle \cdot, \cdot \rangle)$ we can go a step further and use the induced norm

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle} \tag{27}$$

Angles

From CS Eqn

$|\langle x, y\rangle| \leq \|x\|\|y\| \Rightarrow \frac{|\langle x, y\rangle|}{\|x\|\|y\|} \leq 1 \Rightarrow -1 \leq \frac{\langle x, y\rangle}{\|x\|\|y\|} \leq 1$

Inner products capture the geometry of a vector space by defining the angle ω between two vectors.

- Observer that we can obtain the following, directly from the Cauchy-Schwarz inequality

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|} \leq 1 \text{ for } \mathbf{x} \neq \mathbf{0}, \mathbf{y} \neq \mathbf{0} \tag{28}$$

- Furthermore there exists a unique $\omega \in [0, \pi]$ such that

$$cos(\omega) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{29}$$

The number ω is the *angle* (in radians) between the vectors $\mathbf{x}$ and $\mathbf{y}$

Orth: $\cos\left(\frac{\pi}{2}\right) = 0 = \frac{\langle x, y\rangle}{\|x\|\|y\|} \Rightarrow \langle x, y\rangle = 0$

Orthogonality

Orthogonality

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if and only if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

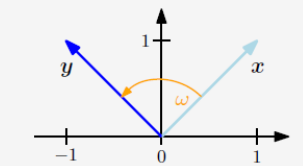
- Orthogonal vectors are denoted by $\mathbf{x} \perp \mathbf{y}$
- If in addition $\mathbf{x} \perp \mathbf{y}$ we have that $\|\mathbf{x}\| = 1 = \|\mathbf{y}\|$ we say $\mathbf{x}$ and $\mathbf{y}$ are *orthonormal*.

A direct implication of this definition is that the $\mathbf{0}$-vector is orthogonal to every vector in the vector space.

Orthogonality Over different Inner products

It is worth noting the originality of two vectors depends on the inner product used. For example

- Consider $\mathbf{x} = [1, 1]^T$ and $\mathbf{y} = [-1, 1]^T$ in $\mathbb{R}^2$



- If we use the dot product as the inner product we have that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = -1 + 1 = 0$$
$$\implies \text{angle of } \pi/2 \text{ radians} \implies \mathbf{x} \perp \mathbf{y}$$

- But if we define the inner product as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{y}$$
$$\implies \text{angle of } 1.91 \text{ radians}$$

Orthogonal Matrix

Orthogonal Matrix

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an *orthogonal matrix* if and only if its columns are orthonormal so that

$$\mathbf{A}\mathbf{A}^T = \mathbf{I} = \mathbf{A}^T\mathbf{A} \tag{30}$$

which implies that

$$\mathbf{A}^{-1} = \mathbf{A}^T \tag{31}$$

Orthogonal Matrix: Useful Properties

Transformations by orthogonal matrices, $\mathbf{A} \in \mathbb{R}^{n\times n}$, are special because the length of a vector $\mathbf{x}$ is preserved when $\mathbf{A}$ is applied. For example, in the case of the dot product

$$\begin{aligned}
\|\mathbf{Ax}\|^2 =& (\mathbf{Ax})^T(\mathbf{Ax}) \\
=& \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} \\
=& \mathbf{x}^T\mathbf{I}\mathbf{x} \\
=& \mathbf{x}^T\mathbf{x} \\
=& \|\mathbf{x}\|^2
\end{aligned}$$

Orthogonal Matrix: Useful Properties

Additionally the angle between two vector is also preserved when **A** is applied. For example, in the case of the dot product we have that

$$
\begin{aligned}
cos(\omega) &= \frac{\langle \mathbf{Ax}, \mathbf{Ay} \rangle}{\|\mathbf{Ax}\|\|\mathbf{Ay}\|} \\
&= \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{Ay}}{\|\mathbf{x}\|\|\mathbf{y}\|} \\
&= \frac{\mathbf{x}^T \mathbf{Iy}}{\|\mathbf{x}\|\|\mathbf{y}\|} \\
&= \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}
\end{aligned}
$$

Orthonormal Basis

We have previously used bases to perform a number of alebragic tasks. A further refinement of the concept is however useful. Namely:

Orthonormal Basis

Consider an n-dimensional vector space V and a basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of V. If

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0 \text{ for } i \neq j \tag{32}$$

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 1 \text{ for i=j} \tag{33}$$

holds for all $i, j = 1, \ldots, n$ then the basis is called an *orthonormal basis* (ONB).

- If only equation (32) is satisfied then the basis is called a *orthogonal basis*.

Orthogonal Complement

We can extend the idea of orthogonality to **whole** vector spaces.

- Consider a D-dimensional vector space V and an M-dimensional subspace $U \subseteq V$.
- Then U's *orthogonal complement* $U^{\perp}$ is a $(D - M)$-dimensional subspace of V
- and $U^{\perp}$ contains all vectors in V that are orthogonal to every vector in U.

Orthogonal Complement

We can also represent any $\mathbf{x} \in V$ as weighted sum of basis vectors from U and $U^{\perp}$.

- First note that $U \cap U^{\perp} = \{\mathbf{0}\}$, as such for any $\mathbf{x} \in V$ can be **uniquely** decomposed into

$$\mathbf{x} = \sum_{m=1}^{M} \lambda_m \mathbf{b}_m + \sum_{j=1}^{D-M} \psi_j \mathbf{b}_j^{\perp}, \ \lambda_m, \psi_j \in \mathbb{R} \tag{34}$$

where $(\mathbf{b}_1, \cdots, \mathbf{b}_M)$ is a basis of U and $(\mathbf{b}_1^{\perp}, \cdots, \mathbf{b}_{D-M}^{\perp})$ is a basis of $U^{\perp}$

Inner Product of Functions

Up until now we only really considered the inner products of finite-dimensional vectors. The concepts are far more powerful than this. Specifically, we can generalize to entities

- with countably infinite[1] entries $[x_1, x_2, \ldots]$
- with unaccountably infinitly[2] entries

$\hookrightarrow \mathbb{R}$

[1] For example $\mathbb{Z}$

[2] For example $(0, 1)$

Inner Product of Functions

Consider the two functions $u : \mathbb{R} \to \mathbb{R}$ and $v : \mathbb{R} \to \mathbb{R}$ then the following defines and inner product

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx \tag{35}$$

for finite a and b

- This inner product, as all inner products, induces a norm.
- We now have the concept of orthogonality of functions!

There is a bit more nuisance with regards to precisely defining the integral (measures, integral construction) and the concepts of a Hilbert space. We however can use equation (35) from a practical perspective in this course.