Analytic Geometry and Linear Algebra.

The basic idea of analytic geometry is to represent the plane (or 3-dimensional space) by coordinates.

The basic idea of analytic geometry is to represent the plane (or 3-dimensional space) by coordinates. If we denote by

$$\mathbb{R}^n = \{x = (x_1, ...x_n); x_j \in \mathbb{R}\},$$

this means that we think of the space of geometry as $\mathbb{R}^2$ or $\mathbb{R}^3$.

The basic idea of analytic geometry is to represent the plane (or 3-dimensional space) by coordinates. If we denote by

$$\mathbb{R}^n = \{x = (x_1, ...x_n); x_j \in \mathbb{R}\},$$

this means that we think of the space of geometry as $\mathbb{R}^2$ or $\mathbb{R}^3$. Notice that the idea that a line is represented by $\mathbb{R}$ is actually the main motivation or inspiration for the introduction of the real numbers.

The basic idea of analytic geometry is to represent the plane (or 3-dimensional space) by coordinates. If we denote by

$$\mathbb{R}^n = \{x = (x_1, ...x_n); x_j \in \mathbb{R}\},$$

this means that we think of the space of geometry as $\mathbb{R}^2$ or $\mathbb{R}^3$. Notice that the idea that a line is represented by $\mathbb{R}$ is actually the main motivation or inspiration for the introduction of the real numbers. The idea to represent higher dimensional spaces by coordinates is usually attributed to Descartes and Fermat (both lived in the first half of the seventeenth century), although similar ideas can be traced back to antiquity.

The basic idea of analytic geometry is to represent the plane (or 3-dimensional space) by coordinates. If we denote by

$$\mathbb{R}^n = \{x = (x_1, ...x_n); x_j \in \mathbb{R}\},$$

this means that we think of the space of geometry as $\mathbb{R}^2$ or $\mathbb{R}^3$. Notice that the idea that a line is represented by $\mathbb{R}$ is actually the main motivation or inspiration for the introduction of the real numbers. The idea to represent higher dimensional spaces by coordinates is usually attributed to Descartes and Fermat (both lived in the first half of the seventeenth century), although similar ideas can be traced back to antiquity.

Now the basic objects of geometry can be described in terms of coordinates.

1. A *point* is an element of $\mathbb{R}^n$.

1. A *point* is an element of $\mathbb{R}^n$.

2. A *line* through the point $a \in \mathbb{R}^n$ with direction $v \in \mathbb{R}^n$ is a set $L_{a,v} := \{a + tv; t \in \mathbb{R}\}$.

1. A *point* is an element of $\mathbb{R}^n$.

2. A *line* through the point $a \in \mathbb{R}^n$ with direction $v \in \mathbb{R}^n$ is a set $L_{a,v} := \{a + tv; t \in \mathbb{R}\}$.

3. The *distance* between two points x and y is $|x - y| = \sqrt{(x_1 - y_1)^2 + ...(x_n - y_n)^2}$

1. A *point* is an element of $\mathbb{R}^n$.

2. A *line* through the point $a \in \mathbb{R}^n$ with direction $v \in \mathbb{R}^n$ is a set $L_{a,v} := \{a + tv; t \in \mathbb{R}\}$.

3. The *distance* between two points x and y is $|x - y| = \sqrt{(x_1 - y_1)^2 + ...(x_n - y_n)^2}$

4. A *circle* or *sphere* is the set of points that satisfy $|x - c| = R$ for a fixed center c and radius R.

1. A *point* is an element of $\mathbb{R}^n$.

2. A *line* through the point $a \in \mathbb{R}^n$ with direction $v \in \mathbb{R}^n$ is a set $L_{a,v} := \{a + tv; t \in \mathbb{R}\}$.

3. The *distance* between two points x and y is $|x - y| = \sqrt{(x_1 - y_1)^2 + ...(x_n - y_n)^2}$

4. A *circle* or *sphere* is the set of points that satisfy $|x - c| = R$ for a fixed center c and radius R.

5. The *angle* between two directions v and w is given by

$$\arccos(\frac{v \cdot w}{|v||w|}),$$

where $v \cdot w = \sum v_j w_j$.

This way every statement in classical geometry corresponds to a statement about real numbers, and geometry becomes a part of the theory of real numbers.

This way every statement in classical geometry corresponds to a statement about real numbers, and geometry becomes a part of the theory of real numbers.

It is actually a rather small part: Geometry is *decidable* in the sense that for any statement it can be checked automatically if it holds or not (Tarski), whereas the theory of the real numbers is not (Gödel).

This way every statement in classical geometry corresponds to a statement about real numbers, and geometry becomes a part of the theory of real numbers.
It is actually a rather small part: Geometry is *decidable* in the sense that for any statement it can be checked automatically if it holds or not (Tarski), whereas the theory of the real numbers is not (Gödel). This does not mean that it is easy to check – the game of Chess is also decidable (whatever that means).

This way every statement in classical geometry corresponds to a statement about real numbers, and geometry becomes a part of the theory of real numbers.
It is actually a rather small part: Geometry is *decidable* in the sense that for any statement it can be checked automatically if it holds or not (Tarski), whereas the theory of the real numbers is not (Gödel). This does not mean that it is easy to check – the game of Chess is also decidable (whatever that means).
There is an extra (unexpected ?) bonus with the translation to coordinates: We can do geometry in any dimension, and it is in principle as easy as in two dimensions. Here is an example of this:

The method of least squares

Let $(x_1, y_1), ...(x_n, y_n)$ be a number of points in the plane. If $n > 2$ we cannot draw a line through all of the points in general.

The method of least squares

Let $(x_1, y_1), ...(x_n, y_n)$ be a number of points in the plane. If $n > 2$ we cannot draw a line through all of the points in general. But, we can try to find a line that ‘comes as close as possible’ to doing that.

The method of least squares

Let $(x_1, y_1), ...(x_n, y_n)$ be a number of points in the plane. If $n > 2$ we cannot draw a line through all of the points in general. But, we can try to find a line that 'comes as close as possible' to doing that.

Two numbers, a and b determine the line $y = ax + b$. Instead of trying to solve the *overdetermined* system of equations

$$y_j = ax_j + b$$

we try to minimize the error

$$\epsilon^2 = \sum_j (y_j - (ax_j + b))^2$$

over all choices of a and b.

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Any line in $\mathbb{R}^2$ corresponds to a choice of a, b and therefore to a point in P. The minimal error ϵ that we want to find is the distance from the point $\mathbf{y}$ in $\mathbb{R}^n$ to the plane P.

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Any line in $\mathbb{R}^2$ corresponds to a choice of a, b and therefore to a point in P. The minimal error ϵ that we want to find is the distance from the point $\mathbf{y}$ in $\mathbb{R}^n$ to the plane P. Why?

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Any line in $\mathbb{R}^2$ corresponds to a choice of a, b and therefore to a point in P. The minimal error ϵ that we want to find is the distance from the point $\mathbf{y}$ in $\mathbb{R}^n$ to the plane P. Why?

The distance from $\mathbf{y}$ to the plane is

$$d = \min |\mathbf{y} - \mathbf{z}|,$$

where $\mathbf{z}$ ranges over all points in the plane P.

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Any line in $\mathbb{R}^2$ corresponds to a choice of a, b and therefore to a point in P. The minimal error ϵ that we want to find is the distance from the point $\mathbf{y}$ in $\mathbb{R}^n$ to the plane P. Why?

The distance from $\mathbf{y}$ to the plane is

$$d = \min |\mathbf{y} - \mathbf{z}|,$$

where $\mathbf{z}$ ranges over all points in the plane P. But, any point $\mathbf{z}$ in the plane is of the form $\mathbf{z} = a\mathbf{x} + b\mathbf{1}$, so $d = \epsilon$.

Let $\mathbf{x} = (x_1, ...x_n)$, $\mathbf{1} = (1, 1...1)$ (**two points in $\mathbb{R}^n$!!**) and let P be the twodimensional plane

$$P = \{a\mathbf{x} + b\mathbf{1}; a, b \in \mathbb{R}\}.$$

Any line in $\mathbb{R}^2$ corresponds to a choice of a, b and therefore to a point in P. The minimal error ϵ that we want to find is the distance from the point $\mathbf{y}$ in $\mathbb{R}^n$ to the plane P. Why?

The distance from $\mathbf{y}$ to the plane is

$$d = \min |\mathbf{y} - \mathbf{z}|,$$

where $\mathbf{z}$ ranges over all points in the plane P. But, any point $\mathbf{z}$ in the plane is of the form $\mathbf{z} = a\mathbf{x} + b\mathbf{1}$, so $d = \epsilon$.

How do we find it?

It is clear from a figure that the minimum will occur in a point (a_0, b_0) such that $\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1})$ is perpendicular to any vector in the plane. (Excercise: prove this!).

It is clear from a figure that the minimum will occur in a point (a_0, b_0) such that $\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1})$ is perpendicular to any vector in the plane. (Excercise: prove this!). Since the plane is spanned by the vectors $\mathbf{x}$ and $\mathbf{1}$, his means that

$$[\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1}] \cdot \mathbf{x} = 0, \quad [\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1})] \cdot \mathbf{1} = 0.$$

This is an inhomogenous system of two equations and two unknowns which always has a solution (why?).

It is clear from a figure that the minimum will occur in a point (a_0, b_0) such that $\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1})$ is perpendicular to any vector in the plane. (Excercise: prove this!). Since the plane is spanned by the vectors $\mathbf{x}$ and $\mathbf{1}$, his means that

$$[\mathbf{y} - (a_0\mathbf{x}+b_0\mathbf{1}] \cdot \mathbf{x} = 0, \quad [\mathbf{y} - (a_0\mathbf{x} + b_0\mathbf{1})] \cdot \mathbf{1} = 0.$$

This is an inhomogenous system of two equations and two unknowns which always has a solution (why?). Observe that a_0 and b_0 are the unknowns, and $\mathbf{x}$, $\mathbf{y}$ are given!

In the same way we can find the best polynomial of degree p.

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.)

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.) Put

$$\vec{f_k} = (f_k(x_1), ...f_k(x_n)), \quad L := \{\sum_k a_k \vec{f_k}\} = [\vec{f_1}, ...\vec{f_q}].$$

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.) Put

$$\vec{f_k} = (f_k(x_1), ...f_k(x_n)), \quad L := \{\sum_k a_k \vec{f_k}\} = [\vec{f_1}, ...\vec{f_q}].$$

We want to minimize

$$\epsilon^2 = d(\vec{y}, L)^2.$$

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.) Put

$$\vec{f_k} = (f_k(x_1), ...f_k(x_n)), \quad L := \{\sum_k a_k \vec{f_k}\} = [\vec{f_1}, ...\vec{f_q}].$$

We want to minimize

$$\epsilon^2 = d(\vec{y}, L)^2.$$

The solution is when

$$\vec{y} - \sum a_k \vec{f_k} \perp \vec{f_l}, \quad \text{for all } l,$$

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.) Put

$$\vec{f_k} = (f_k(x_1), ...f_k(x_n)), \quad L := \{\sum_k a_k \vec{f_k}\} = [\vec{f_1}, ...\vec{f_q}].$$

We want to minimize

$$\epsilon^2 = d(\vec{y}, L)^2.$$

The solution is when

$$\vec{y} - \sum a_k \vec{f_k} \perp \vec{f_l}, \quad \text{for all } l,$$

i. e.

$$\sum a_k (\vec{f_k}, \vec{f_l}) = (\vec{y}, \vec{f_l}).$$

In the same way we can find the best polynomial of degree p. The general problem is: Let $f_1, ...f_q$ be a number of functions of x and try to minimize

$$\epsilon^2 := \sum_i (y_i - (a_1 f_1(x_i) + ...a_q f_q(x_i)))^2.$$

(Before we had $f_1(x) = x$ and $f_2(x) = 1$.) Put

$$\vec{f_k} = (f_k(x_1), ...f_k(x_n)), \quad L := \{\sum_k a_k \vec{f_k}\} = [\vec{f_1}, ...\vec{f_q}].$$

We want to minimize

$$\epsilon^2 = d(\vec{y}, L)^2.$$

The solution is when

$$\vec{y} - \sum a_k \vec{f_k} \perp \vec{f_l}, \quad \text{for all } l,$$

i. e.

$$\sum a_k(\vec{f_k}, \vec{f_l}) = (\vec{y}, \vec{f_l}).$$

The problem has a unique solution when the vectors $\vec{f_k}$ are linearly independent.

Notice that the method of least squares solves a problem in the plane by using geometry in n dimensions, where n is the number of points and can be arbitrary big.

Notice that the method of least squares solves a problem in the plane by using geometry in n dimensions, where n is the number of points and can be arbitrary big.
The method of least squares was probably first used by Gauss, who applied it to find a 'lost planet'.

The astronomers had found a dwarf planet Ceres and recorded its positions for some time, but suddenly they could not see it anymore, because it was close to aligned with the sun. Gauss extrapolated the position from the known trajectory and could tell the astronomers where to look.

The astronomers had found a dwarf planet Ceres and recorded its positions for some time, but suddenly they could not see it anymore, because it was close to aligned with the sun. Gauss extrapolated the position from the known trajectory and could tell the astronomers where to look.

The method of least squares is now indispensable in statistics, and is a good illustration of the use of abstractions.

The astronomers had found a dwarf planet Ceres and recorded its positions for some time, but suddenly they could not see it anymore, because it was close to aligned with the sun. Gauss extrapolated the position from the known trajectory and could tell the astronomers where to look.

The method of least squares is now indispensable in statistics, and is a good illustration of the use of abstractions.

Like Gauss, we can try to use it to predict many other things from known observations, like stockmarkets.

The astronomers had found a dwarf planet Ceres and recorded its positions for some time, but suddenly they could not see it anymore, because it was close to aligned with the sun. Gauss extrapolated the position from the known trajectory and could tell the astronomers where to look.

The method of least squares is now indispensable in statistics, and is a good illustration of the use of abstractions.

Like Gauss, we can try to use it to predict many other things from known observations, like stockmarkets.

Why least *squares*? Why not 'least sums'

$$\sum |y_i - (ax_i + b)|?$$

Why least *squares*? Why not ‘least sums’

$$\sum |y_i - (ax_i + b)|?$$

Least squares fit better with Euclidean geometry.

Why least *squares*? Why not 'least sums'

$$\sum |y_i - (ax_i + b)|?$$

Least squares fit better with Euclidean geometry. But, recent research in compressed sensing has indicated that least sums might be better in some cases!

Why least *squares*? Why not ‘least sums’

$$\sum |y_i - (ax_i + b)|?$$

Least squares fit better with Euclidean geometry. But, recent research in compressed sensing has indicated that least sums might be better in some cases!

The basic problem that the least squares method addresses is to describe data with many degrees of freedom (the points (x_i, y_i)) approximately with few parameters (a and b).

Why least *squares*? Why not 'least sums'

$$\sum |y_i - (ax_i + b)|?$$

Least squares fit better with Euclidean geometry. But, recent research in compressed sensing has indicated that least sums might be better in some cases!

The basic problem that the least squares method addresses is to describe data with many degrees of freedom (the points (x_i, y_i)) approximately with few parameters (a and b).

A similar problem arises when we try to compress a picture with many pixels to few kilobytes.

Why least *squares*? Why not 'least sums'

$$\sum |y_i - (ax_i + b)|?$$

Least squares fit better with Euclidean geometry. But, recent research in compressed sensing has indicated that least sums might be better in some cases!

The basic problem that the least squares method addresses is to describe data with many degrees of freedom (the points (x_i, y_i)) approximately with few parameters (a and b).

A similar problem arises when we try to compress a picture with many pixels to few kilobytes. This is where 'least sums' have proved to be surprisingly useful.

One central topic in linear algebra is the solution of linear systems of equations

$$\begin{aligned} a_{11}x_1 + \ldots.a_{1n}x_n = y_1 \\ a_{21}x_1 + \ldots a_{2n}x_n = y_2\ldots \\ a_{m1}x_1 + \ldots a_{mn}x_n = y_m \end{aligned},$$

or

$$Ax = y,$$

where A is the coefficient matrix of the system.

One central topic in linear algebra is the solution of linear systems of equations

$$a_{11}x_1 + ...a_{1n}x_n = y_1$$
$$a_{21}x_1 + ...a_{2n}x_n = y_2 ... \ ,$$
$$a_{m1}x_1 + ...a_{mn}x_n = y_m$$

or

$$Ax = y,$$

where A is the coefficient matrix of the system.
Here is the most important theorem in that context. We think of A as a linear map $x \to Ax$ from $\mathbb{R}^n$ to $\mathbb{R}^m$. Recall that $Ker(A) = \{x; Ax = 0\}$ and $Im(A) = \{Ax; x \in \mathbb{R}^n\}$; they are both linear subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively.

Theorem

Let A be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then

$$\dim(Ker(A)) + \dim(Im(A)) = n.$$

Theorem

Let A be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then

$$\dim(Ker(A)) + \dim(Im(A)) = n.$$

The statement and the proof hinges on the notion of dimension. A linear space, like $\mathbb{R}^n$ has many different bases, but they have all the same number of elements. (Exercise: Prove this!) This is the dimension of the space. Say the dimension of $Ker(A)$ is k, and let $e_1, ...e_k$ be a basis. We can find vectors in $\mathbb{R}^n$, $f_1, ...f_{n-k}$ that complete $e_1, ...e_k$ to a basis of $\mathbb{R}^n$. Let F be the linear span of $f_1, ...f_{n-k}$. Then the restriction of A to F is injective (why?).

Theorem

Let A be a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then

$$\dim(Ker(A)) + \dim(Im(A)) = n.$$

The statement and the proof hinges on the notion of dimension. A linear space, like $\mathbb{R}^n$ has many different bases, but they have all the same number of elements. (Exercise: Prove this!) This is the dimension of the space. Say the dimension of $Ker(A)$ is k, and let $e_1, ...e_k$ be a basis. We can find vectors in $\mathbb{R}^n$, $f_1, ...f_{n-k}$ that complete $e_1, ...e_k$ to a basis of $\mathbb{R}^n$. Let F be the linear span of $f_1, ...f_{n-k}$. Then the restriction of A to F is injective (why?). Hence, $Af_1, ...Af_{n-k}$ is a basis for $Im(A)$. Thus the dimension of $Im(A)$ is $n - k$ which is the statement.

The theorem can be reformulated in the following way. Let $G = \mathbb{R}^m / Im(A) := coker(A)$. Then

Theorem

$$ind(A) := \dim(Ker(A)) - \dim(cokernel(A)) = n - m.$$

The theorem can be reformulated in the following way. Let $G = \mathbb{R}^m / Im(A) := coker(A)$. Then

Theorem

$$ind(A) := \dim(Ker(A)) - \dim(cokernel(A)) = n - m.$$

The advantage with this formulation is that the kernel and the cokernel may have finite dimensions even if A acts on an infinite dimensional space.

The theorem can be reformulated in the following way. Let $G = \mathbb{R}^m / Im(A) := coker(A)$. Then

Theorem

$$ind(A) := \dim(Ker(A)) - \dim(cokernel(A)) = n - m.$$

The advantage with this formulation is that the kernel and the cokernel may have finite dimensions even if A acts on an infinite dimensional space. If $A : V \to V$ where V is a vector space of finite dimension, then the index is always zero.

The theorem can be reformulated in the following way. Let $G = \mathbb{R}^m / Im(A) := coker(A)$. Then

Theorem

$$ind(A) := \dim(Ker(A)) - \dim(cokernel(A)) = n - m.$$

The advantage with this formulation is that the kernel and the cokernel may have finite dimensions even if A acts on an infinite dimensional space. If $A : V \to V$ where V is a vector space of finite dimension, then the index is always zero. This is not always the case in infinite dimensions as we shall see later.

The theorem can be reformulated in the following way. Let $G = \mathbb{R}^m / Im(A) := coker(A)$. Then

Theorem

$$ind(A) := \dim(Ker(A)) - \dim(cokernel(A)) = n - m.$$

The advantage with this formulation is that the kernel and the cokernel may have finite dimensions even if A acts on an infinite dimensional space. If $A : V \to V$ where V is a vector space of finite dimension, then the index is always zero. This is not always the case in infinite dimensions as we shall see later. The index is an important object to study in the theory of partial differential equations, when A is a differential operator.

Matrices as we have seen arise in the study of linear maps between finite dimensional vector spaces, but they also appear in a somewhat different context.

Matrices as we have seen arise in the study of linear maps between finite dimensional vector spaces, but they also appear in a somewhat different context. Let

$$Q(x) = \sum a_{ij} x_i x_j$$

be a *quadratic form*.

Matrices as we have seen arise in the study of linear maps between finite dimensional vector spaces, but they also appear in a somewhat different context. Let

$$Q(x) = \sum a_{ij} x_i x_j$$

be a *quadratic form*. If $A = (a_{ij})$ we may write

$$Q(x) = x^t A x,$$

and we may assume that A is symmetric.

Matrices as we have seen arise in the study of linear maps between finite dimensional vector spaces, but they also appear in a somewhat different context. Let

$$Q(x) = \sum a_{ij} x_i x_j$$

be a *quadratic form*. If $A = (a_{ij})$ we may write

$$Q(x) = x^t A x,$$

and we may assume that A is symmetric. If we change basis in $\mathbb{R}^n$, $x = My$, where M is an invertible matrix, we have

$$Q(x) = y^t M^t A M y = Q'(y).$$

We now have the second important theorem of linear algebra:

Theorem

We may find an (orthonormal) M such that

$$Q'(y) = \sum \lambda_j y_j^2.$$

This is the *Spectral Theorem*. If we interpret A as a linear operator, $A' = M^{-1}AM$ is the matrix for the same operator in the new basis, where y are coordinates. But, since M is orthonormal, $M^t = M^{-1}$. hence the theorem says that we change coordinates so that A' is the diagonal with eigenvalues λ_j.

This is the *Spectral Theorem*. If we interpret A as a linear operator, $A' = M^{-1}AM$ is the matrix for the same operator in the new basis, where y are coordinates. But, since M is orthonormal, $M^t = M^{-1}$. hence the theorem says that we change coordinates so that A' is the diagonal with eigenvalues λ_j.

We are now ready to discuss the corresponding facts in infinite dimension.

Infinite dimension and Hilbert space.

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space.

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules.

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules. Nothing in this general picture presupposes that the space is finite dimensional, i e has a finite basis.

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules. Nothing in this general picture presupposes that the space is finite dimensional, i e has a finite basis.

However, just being a linear space is not enough structure to give interesting or useful mathematics. The interest starts when we introduce geometry, i e have a way to measure distances.

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules. Nothing in this general picture presupposes that the space is finite dimensional, i e has a finite basis.

However, just being a linear space is not enough structure to give interesting or useful mathematics. The interest starts when we introduce geometry, i e have a way to measure distances.

In the case of *Hilbert spaces*, this way of measuring distances comes from a *scalar product*:

$$(u, v).$$

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules. Nothing in this general picture presupposes that the space is finite dimensional, i e has a finite basis.

However, just being a linear space is not enough structure to give interesting or useful mathematics. The interest starts when we introduce geometry, i e have a way to measure distances.

In the case of *Hilbert spaces*, this way of measuring distances comes from a *scalar product*:

$$(u, v).$$

The length, or norm, of a vector is then given by

$$\|u\|^2 = (u, u).$$

The study of $\mathbb{R}^2$ and $\mathbb{R}^n$ from a geometric viewpoint, led to the notion of an abstract linear space. Its elements can be added and multiplied by scalars according to certain rules. Nothing in this general picture presupposes that the space is finite dimensional, i e has a finite basis.

However, just being a linear space is not enough structure to give interesting or useful mathematics. The interest starts when we introduce geometry, i e have a way to measure distances.

In the case of *Hilbert spaces*, this way of measuring distances comes from a *scalar product*:

$$(u, v).$$

The length, or norm, of a vector is then given by

$$\|u\|^2 = (u, u).$$

In the infinite dimensional case one needs an extra assumption (that is automatic in finite dimensions): The norm is *complete*.

In the infinite dimensional case one needs an extra assumption (that is automatic in finite dimensions): The norm is *complete*.
This means that Cauchy sequences are convergent, or equivalently that if

$$\sum \|u_j\| < \infty,$$

then

$$\lim \sum^{n} u_j$$

exists.

In the infinite dimensional case one needs an extra assumption (that is automatic in finite dimensions): The norm is *complete*.
This means that Cauchy sequences are convergent, or equivalently that if

$$\sum \|u_j\| < \infty,$$

then

$$\lim \sum^n u_j$$

exists.
In other words, there is an element u in the space such that

$$\|u - \sum^n u_j\| \to 0.$$

Example 1: Let
$V = \{u = (u_0, ...u_n, ...), u_k = 0 \quad \text{for k sufficiently large}\}$, with norm

$$\|u\|^2 = \sum |u_j|^2.$$

Example 1: Let
$V = \{u = (u_0, ...u_n, ...), u_k = 0 \quad \text{for k sufficiently large}\}$, with norm

$$\|u\|^2 = \sum |u_j|^2.$$

Example 2: Let $V = \{u = (u_0, ...u_n, ...), \sum |u_k|^2 < \infty\}$, with norm

$$\|u\|^2 = \sum |u_j|^2.$$

Example 1: Let
$V = \{u = (u_0, ...u_n, ...), u_k = 0 \quad \text{for k sufficiently large}\}$, with norm

$$\|u\|^2 = \sum |u_j|^2.$$

Example 2: Let $V = \{u = (u_0, ...u_n, ...), \sum |u_k|^2 < \infty\}$, with norm

$$\|u\|^2 = \sum |u_j|^2.$$

Example 2 is complete, Example 1 is not.

Theorem

Every Hilbert space V has an orthonormal basis, i e there is an orthonormal set of vectors $\{e_\alpha\}_{\alpha \in A}$ such that any vector in V can be written

$$x = \sum_A c_\alpha e_\alpha,$$

and

$$\|x\|^2 = \sum_A |c_\alpha|^2.$$

Theorem

Every Hilbert space V has an orthonormal basis, i e there is an orthonormal set of vectors $\{e_\alpha\}_{\alpha \in A}$ such that any vector in V can be written

$$x = \sum_A c_\alpha e_\alpha,$$

and

$$\|x\|^2 = \sum_A |c_\alpha|^2.$$

In practice, the most interesting case is when A is countable. The Hilbert space is then said to be *separable*.

Theorem

Every Hilbert space V has an orthonormal basis, i e there is an orthonormal set of vectors $\{e_\alpha\}_{\alpha \in A}$ such that any vector in V can be written

$$x = \sum_A c_\alpha e_\alpha,$$

and

$$\|x\|^2 = \sum_A |c_\alpha|^2.$$

In practice, the most interesting case is when A is countable. The Hilbert space is then said to be *separable*.

The theorem says that any separable (i e interesting) Hilbert space is isomorphic to

$$l^2 = \{(c_n)_{n \in N}; \sum |c_n|^2 < \infty\}.$$

Theorem

Every Hilbert space V has an orthonormal basis, i e there is an orthonormal set of vectors $\{e_\alpha\}_{\alpha \in A}$ such that any vector in V can be written

$$x = \sum_A c_\alpha e_\alpha,$$

and

$$\|x\|^2 = \sum_A |c_\alpha|^2.$$

In practice, the most interesting case is when A is countable. The Hilbert space is then said to be *separable*.

The theorem says that any separable (i e interesting) Hilbert space is isomorphic to

$$l^2 = \{(c_n)_{n \in N}; \sum |c_n|^2 < \infty\}.$$

Briefly, there is only one Hilbert space.

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$.

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$. This is a linear space, with a scalar product

$$(f, g) = \int_0^1 f\bar{g}dx,$$

and it is complete.

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$. This is a linear space, with a scalar product

$$(f, g) = \int_0^1 f\bar{g} dx,$$

and it is complete. It is a Hilbert space.

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$. This is a linear space, with a scalar product

$$(f, g) = \int_0^1 f\bar{g}dx,$$

and it is complete. It is a Hilbert space. Let

$$e_k(x) = e^{2\pi ikx}.$$

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$. This is a linear space, with a scalar product

$$(f, g) = \int_0^1 f\bar{g}dx,$$

and it is complete. It is a Hilbert space. Let

$$e_k(x) = e^{2\pi ikx}.$$

It is an orthonormal basis for $L^2(T) \sim l^2$; i. e. any element in L^2 can be written

$$f = \sum_{-\infty}^{\infty} c_k e^{2\pi ikx}.$$

Example: Let $L^2(T) = \{f; \int_0^1 |f(x)|^2 dx < \infty\}$. This is a linear space, with a scalar product

$$(f, g) = \int_0^1 f\bar{g}dx,$$

and it is complete. It is a Hilbert space. Let

$$e_k(x) = e^{2\pi ikx}.$$

It is an orthonormal basis for $L^2(T) \sim l^2$; i. e. any element in L^2 can be written

$$f = \sum_{-\infty}^{\infty} c_k e^{2\pi ikx}.$$

This is the Fourier series of f.

What is the origin of Hilbert spaces?

What is the origin of Hilbert spaces?
The Swedish mathematician Ivar Fredholm (1866 -1927) studied *integral equations* of the form

$$\int_0^1 K(x,y)f(x)dx = g(y)$$

(a *Fredholm equation of the first kind*).

What is the origin of Hilbert spaces?
The Swedish mathematician Ivar Fredholm (1866 -1927) studied *integral equations* of the form

$$\int_0^1 K(x,y)f(x)dx = g(y)$$

(a *Fredholm equation of the first kind*). Or

$$f(y) - \lambda \int_0^1 K(y,x)f(x)dx = g(y)$$

(a *Fredholm equation of the second kind*).

What is the origin of Hilbert spaces?
The Swedish mathematician Ivar Fredholm (1866 -1927) studied *integral equations* of the form

$$\int_0^1 K(x,y)f(x)dx = g(y)$$

(a *Fredholm equation of the first kind*). Or

$$f(y) - \lambda \int_0^1 K(y,x)f(x)dx = g(y)$$

(a *Fredholm equation of the second kind*).
Notice that they are continuous analogs of the matrix equations

$$Tf = g, \quad (I - \lambda T)f = g.$$

Equations of this type are important partly because many *differential* equations can be rewritten in this form.

Equations of this type are important partly because many *differential* equations can be rewritten in this form. Fredholm had the daring idea to define the determinants of 'infinite rank matrices',

$$\det(I - \lambda T).$$

Let

$$T_k = (1/k!) \int_{[0,1]^k} \det K(x_i, x_j) dx_1 dx_2 ... dx_k.$$

Then Fredholm's formula is

$$\det(I - \lambda T) = \sum_0^\infty (-\lambda)^k T_k.$$

Equations of this type are important partly because many *differential* equations can be rewritten in this form. Fredholm had the daring idea to define the determinants of 'infinite rank matrices',

$$\det(I - \lambda T).$$

Let

$$T_k = (1/k!) \int_{[0,1]^k} \det K(x_i, x_j) dx_1 dx_2 ... dx_k.$$

Then Fredholm's formula is

$$\det(I - \lambda T) = \sum_0^\infty (-\lambda)^k T_k.$$

Having defined determinants he could 'solve' the equations by Cramer's rule!

Theorem

*(**The Fredholm alternative**) Let*

$$Tf(x) = \int K(x,y)f(y)dy,$$

where K is continuous. Then, for any complex number λ, either the equation

$$(I - \lambda T)f = g$$

has a solution f for any choice of g, or the equation

$$(I - \lambda T)f = 0$$

has a non trivial solution.

The second alternative means that λ is an eigenvalue of T.

The second alternative means that λ is an eigenvalue of T. The full story is that the *index* of an operator $I - T$, where T is compact is always zero.

The second alternative means that λ is an eigenvalue of T. The full story is that the *index* of an operator $I - T$, where T is compact is always zero. Many differential equations can be rewritten in this form and the Fredholm-Hilbert theory is extremely important and useful in mathematics and applications.

The second alternative means that λ is an eigenvalue of T. The full story is that the *index* of an operator $I - T$, where T is compact is always zero. Many differential equations can be rewritten in this form and the Fredholm-Hilbert theory is extremely important and useful in mathematics and applications.

Fredholms article was published in 1903 and inspired Hilbert's general theory on integral equations and the solvability of 'equations in infinitely many variables'. (1912).

The second alternative means that λ is an eigenvalue of T. The full story is that the *index* of an operator $I - T$, where T is compact is always zero. Many differential equations can be rewritten in this form and the Fredholm-Hilbert theory is extremely important and useful in mathematics and applications.

Fredholms article was published in 1903 and inspired Hilbert's general theory on integral equations and the solvability of 'equations in infinitely many variables'. (1912).

The next big step was John von Neumann's general theory of Hilbert spaces (he introduced that name) as a foundation of quantum mechanics in 1932 (when von Neumann was 29 years old).

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space.

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space. In classical mechanics a state is given by the position and momentum of all particles in the system at a given time, i e a vector in $\mathbb{R}^{2n}$.

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space. In classical mechanics a state is given by the position and momentum of all particles in the system at a given time, i e a vector in $\mathbb{R}^{2n}$. An *observable quantity* in quantum theory is given by an (self adjoint) operator, and the eigenvalues of that operator correspond to the numbers that we can get when measuring the quantity.

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space. In classical mechanics a state is given by the position and momentum of all particles in the system at a given time, i e a vector in $\mathbb{R}^{2n}$. An *observable quantity* in quantum theory is given by an (self adjoint) operator, and the eigenvalues of that operator correspond to the numbers that we can get when measuring the quantity.

The quantum mechanical system can be an isolated system like one hydrogen atom, or the entire world. In both cases, a state is a vector in Hilbert space, or a 'wave function'.

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space. In classical mechanics a state is given by the position and momentum of all particles in the system at a given time, i e a vector in $\mathbb{R}^{2n}$. An *observable quantity* in quantum theory is given by an (self adjoint) operator, and the eigenvalues of that operator correspond to the numbers that we can get when measuring the quantity.

The quantum mechanical system can be an isolated system like one hydrogen atom, or the entire world. In both cases, a state is a vector in Hilbert space, or a 'wave function'. There is room for everybody and anything in Hilbert space!

In this theory the *state* of a quantum mechanical system is a vector in a Hilbert space. In classical mechanics a state is given by the position and momentum of all particles in the system at a given time, i e a vector in $\mathbb{R}^{2n}$. An *observable quantity* in quantum theory is given by an (self adjoint) operator, and the eigenvalues of that operator correspond to the numbers that we can get when measuring the quantity.

The quantum mechanical system can be an isolated system like one hydrogen atom, or the entire world. In both cases, a state is a vector in Hilbert space, or a 'wave function'. There is room for everybody and anything in Hilbert space!

In this way we can see Hilbert space as the mathematical theory of quantum mechanics, similarily to how Riemannian geometry is the mathematics of the theory general relativity.