Mathematics IB

Algebra Notes

School of Mathematical Sciences
The University of Adelaide
2016

Contents

References are given by section number to the 4th Edition of *Linear Algebra: A Modern Introduction* by Poole and the 7th edition of *Calculus* by Stewart.

Chapter 1

FURTHER RESULTS ON $\mathbb{R}^n$

1.1 Revision

Subspaces [Poole: §6.1]

Recall that a *subspace* W of $\mathbb{R}^n$ is a set of vectors which satisfy:

1. $\mathbf{0} \in W$
2. if $\mathbf{u}, \mathbf{v} \in W$ then $\mathbf{u} + \mathbf{v} \in W$
3. if $\mathbf{u} \in W$ and $c \in R$ then $c\mathbf{u} \in W$.

◊ **1.0: Examples of subspaces** ...

Linear Combinations [Poole: §433]

A vector $\mathbf{v}$ of $\mathbb{R}^n$ is a *linear combination* of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ (of $\mathbf{R}^n$) if

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k$$

for some real numbers $c_1, \ldots, c_k$.

span$\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is the set of *all* linear combinations of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$:

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{x_1\mathbf{v}_1 + \ldots + x_k\mathbf{v}_k \mid x_1, x_2, \ldots, x_k \in \mathbb{R}\}.$$
$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \text{ is a subspace of } \mathbb{R}^n.$$

Also recall that $\mathbf{v}$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2, \ldots,$ $\mathbf{v}_k$ if and only if the system $A\mathbf{x} = \mathbf{v}$ has a solution; where $A = [\mathbf{v}_1\mathbf{v}_2 \ldots \mathbf{v}_k]$.

◊ **1.1: Examples** ...

Note that if $\mathbf{v}_2 = k\mathbf{v}_1$ then span$\{\mathbf{v}_1, \mathbf{v}_2\}$ = span$\{\mathbf{v}_1\}$ = the line of vectors (through 0) containing $\mathbf{v}_1$.

Linear Independence [Poole: §6.2 pp443–446]

A list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is *linearly independent* if the only solution of the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \ldots + x_k\mathbf{v}_k = \mathbf{0}$$

is the trivial solution

$$x_1 = x_2 = \ldots = x_k = 0.$$

That is, the only linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ which can be $\mathbf{0}$ is the linear combination $0\mathbf{v}_1 + 0\mathbf{v}_2 + \ldots + 0\mathbf{v}_k = \mathbf{0}$. Vectors which are not linearly independent are *linearly dependent*. Recall that writing $A = [\mathbf{v}_1|\mathbf{v}_2|\ldots|\mathbf{v}_k]$ then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly dependent it is always possible to write at least one (maybe all) of the vectors as a linear combination of the others. When $\mathbf{v}_1 \neq \mathbf{0}$, for some $\mathbf{v}_r$, $r > 1$, we have

$$\mathbf{v}_r = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_{r-1}\mathbf{v}_{r-1}\,.$$

♢ **1.2: Examples** ...

Basis [Poole: §6.2,pp446-448]

A list of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ (of W) forms a basis for a subspace W if (i) $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are linearly independent (ii) $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ span W; i.e. $\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = W$.

One of the most important properties of a basis is that if $\mathbf{w}$ is any vector in W, $\mathbf{w}$ can be expressed *uniquely* as a linear combination of the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$:

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_r\mathbf{v}_r$$

and there is only one choice possible for each c_j.

♢ **1.3: Examples of bases** ...

1.2 Dimension [Poole: §6.2,pp452–456]

Our aim is to show that any two bases for a subspace have the same number of vectors.

Theorem 1.1. *: Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for the subspace W. Then every list of more than r vectors in W is linearly dependent.*

$\Diamond$ **1.4: Proof** ...

Theorem 1.2. *Any two bases of a subspace W contain the same number of vectors.*

$\Diamond$ **1.5: Proof** ...

Definition: The *dimension* of a subspace W is the number of vectors in a basis for W. We write dim W for the dimension of W.

◊ 1.6: Examples ...

Further Results on Bases

We know that if dim $W = r$ then

1. any list of $s > r$ vectors in W is linearly dependent
2. any basis for W has r vectors.

Suppose we have r linearly independent vectors in W. Must they form a basis, i.e., must they span W? For example, given 3 linearly independent vectors in $\mathbb{R}^3$, must they form a basis for $\mathbb{R}^3$?

Theorem 1.3. *Let dim $W = r$.*

1. *Any list of r linearly independent vectors in W is a basis for W.*
2. *Any list of r vectors that spans W is a basis for W.*
3. *If $s < r$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_s\}$ is any linearly independent list of vectors in W, then we can find $\mathbf{w}_{s+1}, \ldots, \mathbf{w}_r$ in W so that $\{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$ is a basis for W.*

◊ 1.7: Proof ...

1.3 The Row Space, Column Space and Null Space of a Matrix [Poole: §3.5]

Let

$$A = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 1 & 2 & 4 & 3 \\ 2 & 4 & 6 & 4 \end{bmatrix}.$$

The rows of A may be regarded as vectors in $\mathbb{R}^4$, called row vectors (of A). Let

$$\mathbf{r}_1 = (1, 2, 2, 1), \quad \mathbf{r}_2 = (1, 2, 4, 3), \quad \mathbf{r}_3 = (2, 4, 6, 4).$$

The *row space* of A is $U = \text{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}$. U is a subspace of $\mathbb{R}^4$; we denote it by Row A.

In general, if A is an $m \times n$ matrix, its *row space* is a subspace of $\mathbb{R}^n$. If $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m$ are the row vectors of A then Row $A = \text{span}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$. What is a basis for U and what is the dimension of U? The vectors $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ span U but they may not be linearly independent.

Theorem 1.4. *Elementary row operations do not change the row space of a matrix.*

♢ **1.8: Outline of proof:** ...

♢ **1.9: Example** ...

Theorem 1.5. *The non-zero row vectors in a row echelon form (or reduced row echelon form) of a matrix A form a basis for the row space of A.*

Proof: Omitted – the basic step is proving that such vectors are linearly independent.

♢ **1.10: Example** ...

The Null Space Poole,p197

Definition The *null space* of an $m \times n$ matrix A, written Nul$A$, is the set of solutions of the homogeneous equations

$$A\mathbf{x} = \mathbf{0}.$$
$$\text{Nul } A = \{\mathbf{x} \in R^n | A\mathbf{x} = \mathbf{0}\}.$$

Nul A is a subspace of R^n.

◊ **1.11: Examples** ...

The Column Space

[Poole,p201] **Definition.** The *column space* of an $m \times n$ matrix A, written as Col A, is the set of all linear combinations of the columns of A. If $A = [\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_n]$,

$$\text{Col } A = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$$

Thus Col A is a subspace of R^m.

We know $\mathbf{b}$ is a linear combination of the columns of $A = [\mathbf{a}_1 \ldots \mathbf{a}_n]$ if and only if the system of equations $A\mathbf{x} = \mathbf{b}$ has a solution.

Thus Col $A = \{\mathbf{b} \in R^m | \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in R^n\}$.

◊ **1.12: Example** ...

Algorithm to find Col A

Use elementary row operations to (reduced) row echelon form. List the columns in which the pivots occur; these columns from the original matrix A, are the *pivot columns of* A; they form a basis for Col A.

◊ **1.13: Examples** ...

Theorem 1.6. *The pivot columns of A form a basis for Col A.*

◊ 1.14: Proof ...

Since the number of vectors in a basis for Col A and Row A is the number of pivots, both dimensions are the same:

$$\dim \text{ Col } A = \dim \text{ Row } A.$$

These dimensions are called the *rank* of A. There is also a relationship between dim Nul A and the rank of A.

Theorem 1.7. *(The Rank Theorem) Poole §3.5, p205*

For an $m \times n$ matrix A, the dimension of Col A equals the dimension of Row A. This common dimension is called the rank of A which also satisfies

$$\textit{rank } (A) + \dim \textit{ Nul } A = n.$$

◊ 1.15: Example ...

1.4 Applications to systems of equations

Theorem 1.8. *The system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if*

$$\textit{rank}\ (A) = \textit{rank}\ ([A\mathbf{b}]).$$

♢ **1.16: Proof** ...

♢ **1.17: Examples** ...

Theorem 1.9. *(Extension of Theorem 1.4 from Maths IA):Poole §3.5 p206 The following statements are equivalent for an $n \times n$ matrix A.*

(1) A is invertible.

(2) A is row equivalent to I_n.

(3) $A\mathbf{x} = \mathbf{0}$ has only the zero solution $\mathbf{x} = \mathbf{0}$.

(4) $A\mathbf{x} = \mathbf{b}$ has a unique solution for any column matrix $\mathbf{b}$.

(5) The columns of A span $\mathbf{R}^n$.

(6) rank $A = n$.

(7) The row vectors of A are linearly independent.

(8) The column vectors of A are linearly independent.

Note that the equivalence of (1) to (5) was shown in Maths IA.

♢ **1.18: Proof** ...

1.5 Euclidean Inner Product [Poole §7.1]

If $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$, then their *inner product* or *dot product* or *scalar product* is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \ldots + u_n v_n \ .$$

◊ 1.19: Example ...

In $\mathbb{R}^2$ and $\mathbb{R}^3$ this is just the familiar dot product of vectors.

Properties of the inner product

For all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and any $k \in \mathbb{R}$,

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$;

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$;

(c) $(k\mathbf{u}) \cdot \mathbf{v} = k(\mathbf{u} \cdot \mathbf{v})$;

(d) $\mathbf{v} \cdot \mathbf{v} \geq 0$, $\mathbf{v} \cdot \mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

If $\mathbf{u} = (u_1, \ldots, u_n)$, its *length* or *norm* is

$$\begin{aligned} ||\mathbf{u}|| &= \sqrt{\mathbf{u} \cdot \mathbf{u}} \\ &= \sqrt{u_1^2 + \ldots + u_n^2} \ . \end{aligned}$$

The *distance* between $\mathbf{u}$ and $\mathbf{v}$ is $||\mathbf{u} - \mathbf{v}||$.

◊ 1.20: Example ...

1.6 Orthonormal Bases [Poole: §5.1]

Recall: If $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ then

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \ldots + u_n v_n \ , \quad ||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \ , \quad \mathbf{u} \neq \mathbf{0} \Leftrightarrow ||\mathbf{u}|| \neq 0.$$

Cauchy Schwarz Inequality: $(\mathbf{u} \cdot \mathbf{v})^2 \leq ||\mathbf{u}||^2 \ ||\mathbf{v}||^2$; with equality iff $\mathbf{u} = k\mathbf{v}, k \in R$.

Proof: Omitted.

Hence

$$-||\mathbf{u}|| \ ||\mathbf{v}|| \leq \mathbf{u} \cdot \mathbf{v} \leq ||\mathbf{u}|| \ ||\mathbf{v}|| \ ;$$

$$-1 \leq \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \ ||\mathbf{v}||} \leq 1, \quad \mathbf{u}, \mathbf{v} \neq \mathbf{0}.$$

Definition: The *angle* between non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ is θ, where

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \ ||\mathbf{v}||} \ , \quad 0 \leq \theta \leq \pi \ .$$

◊ **1.21: Examples** ...

Definition: Vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\mathbf{u} \cdot \mathbf{v} = 0$. (So, if $\mathbf{u}, \mathbf{v} \neq 0$, this means $\theta = \pi/2$.) A set $s = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthogonal set of vectors if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for all $i \neq j$.

◊ **1.22: Example** ...

Definition: Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be a basis for a subspace V of $\mathbb{R}^n$. If each vector $\mathbf{v}_i$ has length 1 and $\mathbf{v}_i, \mathbf{v}_j$ are orthogonal whenever $i \neq j$, then S is called an *orthonormal basis* of V. Thus

$$||\mathbf{v}_i|| = 1, \quad i = 1 \ldots r$$

and

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Theorem 1.10. *If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthogonal set of non-zero vectors in $\mathbb{R}^n$ then S is a linearly independent list of vectors.*

◊ **1.23: Proof** ...

◊ **1.24: Example** ...

Theorem 1.11. *If* $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ *is an orthonormal basis of* V *and* $\mathbf{u} \in V$*, then*

$$\mathbf{u} = (\mathbf{u} \cdot \mathbf{v}_1)\mathbf{v}_1 + \ldots + (\mathbf{u} \cdot \mathbf{v}_r)\mathbf{v}_r \; .$$

$\Diamond$ **1.25: Proof** ...

$\Diamond$ **1.26: Example** ...

Our Aim: given a subspace V, to find an orthonormal basis for V.

Orthogonal Projections[Poole§5.2]

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be an orthonormal basis for a subspace V of $\mathbb{R}^n$. Then for each $\mathbf{u} \in R^n$ there are vectors $\mathbf{w}_1, \mathbf{w}_2$ such that

(i) $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$;

(ii) $\mathbf{w}_1 \in V$;

(iii) $\mathbf{w}_2$ is orthogonal to V (i.e., to every vector in V).

$\mathbf{w}_1$ is the *orthogonal projection* of $\mathbf{u}$ on V.

$\mathbf{w}_2$ is the *component of* $\mathbf{u}$ *orthogonal to* V.

We can give formulae for $\mathbf{w}_1$ and $\mathbf{w}_2$:

$$\begin{aligned} \mathbf{w}_1 &= (\mathbf{u} \cdot \mathbf{v}_1)\mathbf{v}_1 + \ldots + (\mathbf{u} \cdot \mathbf{v}_r)\mathbf{v}_r \; ; \\ \mathbf{w}_2 &= \mathbf{u} - \mathbf{w}_1 \; . \end{aligned}$$

Why do these formulae work?

$\Diamond$ **1.27: Proof** ...

$\Diamond$ **1.28: Example** ...

1.7 Gram-Schmidt Process [Poole: §5.3]

Suppose we are given a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ of a subspace V. We wish to construct an orthonormal basis of V.

Step 1.	$\mathbf{v}_1 = \mathbf{u}_1/\|\|\mathbf{u}_1\|\|$.	(normalise)
Step 2.	$\mathbf{v}_2' = \mathbf{u}_2 - (\mathbf{u}_2 \cdot \mathbf{v}_1)\mathbf{v}_1$;	(component of $\mathbf{u}_2$ orthogonal to $\mathbf{v}_1$)
	$\mathbf{v}_2 = \mathbf{v}_2'/\|\|\mathbf{v}_2'\|\|$.	(normalise)
Step 3.	$\mathbf{v}_3' = \mathbf{u}_3 - (\mathbf{u}_3 \cdot \mathbf{v}_1)\mathbf{v}_1 - (\mathbf{u}_3 \cdot \mathbf{v}_2)\mathbf{v}_2$;	(component of $\mathbf{u}_3$ orthogonal to $\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2\}$)
	$\mathbf{v}_3 = \mathbf{v}_3'/\|\|\mathbf{v}_3'\|\|$.	(normalise)

Continue in this way to step n.

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ will be an orthonormal basis of V. This process gives the following:

Theorem 1.12. *Every subspace V of $\mathbb{R}^n$ has an orthonormal basis.*

◊ 1.29: Examples ...

Theorem 1.13. *The orthogonal projection $\mathbf{w_1}$ of $\mathbf{u}$ onto V is the unique vector minimising $||\mathbf{u} - \mathbf{v}||$ for all $v \in V$.*

This means that the orthogonal projection gives the closest point on V to $\mathbf{u}$.

Chapter 2

LINEAR TRANSFORMATIONS

2.1 Introduction [Poole: §6.4]

So far in the course, the functions discussed have been real valued functions of one or more variables. We now wish to extend our knowledge to vector-valued functions, i.e. given a vector $\mathbf{v} \in \mathbb{R}^n$, we operate on it by some function F to obtain another vector $\mathbf{w} \in \mathbb{R}^m$, i.e.

$$\mathbf{w} = F(\mathbf{v})$$

$\mathbf{w}$ is the *image* of $\mathbf{v}$ under the transformation F.

◊ 2.0: Example ...

In particular, we will restrict our attention to the study of functions that are known as Linear Transformations (L.T.).

Definition: A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* if the following two properties hold.

(i) $F(\mathbf{u}+\mathbf{v}) = F(\mathbf{u}) + F(\mathbf{v}) \qquad \forall\, \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ (Addition)

(ii) $F(k\mathbf{u}) = kF(\mathbf{u}) \qquad \forall\, \mathbf{u} \in \mathbb{R}^n,\ k \in \mathbb{R}$ (Scalar Multiplication)

Thus F maps subspaces of $\mathbb{R}^n$ to subspaces of $\mathbb{R}^m$.

◊ 2.1: Example ...

Theorem 2.1. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a L.T. Then*

(i) $F(\mathbf{0}) = \mathbf{0}$

(ii) $F(-\mathbf{u}) = -F(\mathbf{u})$

(iii) $F(k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \ldots + k_\ell\mathbf{u}_\ell) = k_1F(\mathbf{u}_1) + k_2F(\mathbf{u}_2) + \ldots + k_\ell F(\mathbf{u}_\ell)$

◊ 2.2: Proof ...

$\lozenge$ 2.3: Example ...

Example: Let $F : \mathbb{R}^n \to \mathbb{R}^n$, $F(\mathbf{u}) = t\mathbf{u}$ for some fixed number t; then F is a linear transformation.

If $t > 1$, F is a *dilation*.

If $0 < t < 1$, F is a *contraction*.

Note: $t = 1$ $\quad F \equiv I$, identity transformation
$t = 0$ $\quad F \equiv 0$, zero transformation.

Example: Orthogonal Projection.

Let W be an r-dimensional subspace of $\mathbb{R}^n$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be an orthonormal basis for W.

The (orthogonal) projection $F : \mathbb{R}^n \to W$,

$$F(\mathbf{v}) = (\mathbf{v} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{v} \cdot \mathbf{v}_2)\mathbf{v}_2 + \ldots + (\mathbf{v} \cdot \mathbf{v}_r)\mathbf{v}_r$$

is a linear transformation.

$\lozenge$ 2.4: Proof ...

Example: Matrix Multiplication.

Let A be an $m \times n$ matrix. $F : \mathbb{R}^n \to \mathbb{R}^m$ where

$$F(\mathbf{u}) = A\mathbf{u}$$

is a linear transformation.

Special case: Let $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

In this case $F : \mathbb{R}^2 \to \mathbb{R}^2$ represents a rotation through the angle θ.

$\lozenge$ 2.5: Showing that F represents a rotation ...

Let V be a subspace with basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

Suppose $F(\mathbf{v}_1) = \mathbf{w}_1, F(\mathbf{v}_2) = \mathbf{w}_2, \ldots, F(\mathbf{v}_r) = \mathbf{w}_r$ where

$$F : V \to W.$$

Thus for any vector $\mathbf{v} \in V$ we can determine $F(\mathbf{v}) \in W$. First express $\mathbf{v}$ as a linear combination of the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$, i.e.

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_r\mathbf{v}_r \ .$$

Then

$$\begin{aligned} F(\mathbf{v}) &= F(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_r\mathbf{v}_r) \\ &= F(a_1\mathbf{v}_1) + F(a_2\mathbf{v}_2) + \ldots + F(a_r\mathbf{v}_r) \\ &= a_1F(\mathbf{v}_1) + a_2F(\mathbf{v}_2) + \ldots + a_rF(\mathbf{v}_r) \\ &= a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \ldots + a_r\mathbf{w}_r \ . \end{aligned}$$

◊ **2.6: Example** ...

2.2 Kernel and Range [Poole: §6.5]

Definition: Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then the *kernel* or *nullspace* of F, denoted by ker F, is the set of all vectors $\mathbf{u} \in \mathbb{R}^n$ such that $F(\mathbf{u}) = \mathbf{0}$, i.e.

$$\ker(F) = \{\mathbf{u} \in \mathbb{R}^n \mid F(\mathbf{u}) = \mathbf{0}\}.$$

Definition: Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then the *range* of F, denoted by $\mathcal{R}(F)$, is the set of all vectors $\mathbf{w} \in \mathbb{R}^m$ such that $\mathbf{w} = F(\mathbf{v})$ for some $\mathbf{v} \in \mathbb{R}^n$, i.e.

$$\mathcal{R}(F) = \{F(\mathbf{v}) \mid \mathbf{v} \in \mathbb{R}^n\} \ .$$

◊ **2.7: Examples** ...

Theorem 2.2. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then*

(i) $\ker(F)$ *is a subspace of* $\mathbb{R}^n$.

(ii) $\mathcal{R}(F)$ *is a subspace of* $\mathbb{R}^m$.

◊ **2.8: Proof** ...

♢ 2.9: Example ...

Consider $F(\mathbf{v}) = A\mathbf{v}$ where A is an $m \times n$ matrix and $\mathbf{v} \in \mathbb{R}^n$.

Thus $F : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(\mathbf{v}) = A\mathbf{v}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$$

We wish to describe $\ker(F)$ and $\mathcal{R}(F)$.

♢ 2.10: Calculations ...

Thus we find that $\ker(F) \equiv$ Nul$A$ and $\mathcal{R}(F) =$Col(A).

♢ 2.11: Example ...

2.3 Matrix of a Linear Transformation [Poole: §6.6]

Let A be an $m \times n$ matrix and define a function $F : \mathbb{R}^n \to \mathbb{R}^m$ by $F(\mathbf{u}) = A\mathbf{u}$, where $\mathbf{u}$ is an $(n \times 1)$ column vector. Recall that F is a linear transformation.

♢ 2.12: proof ...

This example is extremely important because of the fact that *any* linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ can be represented in this way by some $m \times n$ matrix A. That is, the above example provides "the only example" of a linear transformation.

Given a linear transformation F, how do we find a matrix A such that $F(\mathbf{v}) = A\mathbf{v}$? We start with two examples.

Example: (a) $F : \mathbb{R}^2 \to \mathbb{R}^3$.

$$F\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ x+y \\ x-y \end{bmatrix}$$

◊ 2.13: Calculations . . .

(b) $F : \mathbb{R}^2 \to \mathbb{R}^3$ is linear and

$$F\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$F\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

◊ 2.14: Calculations . . .

From the above example we obtain the following definition.

Definition: Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a L.T. and let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the standard basis for $\mathbb{R}^n$

$$\left(\text{i.e. } \mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \to i\text{th position}\right)$$

The *standard matrix* for F is the $m \times n$ matrix with $F(\mathbf{e}_1), F(\mathbf{e}_2), \ldots, F(\mathbf{e}_n)$ as its columns (in this order).

Theorem 2.3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a L.T. Then the standard matrix A for F is the unique $m \times n$ matrix such that $\forall\, \mathbf{x} \in \mathbb{R}^n$,*

$$F(\mathbf{x}) = A\mathbf{x}.$$

◊ 2.15: Proof ...

Definition: Let $F : R^n \to R^m$ be a linear transformation. The *nullity* of F is the dimension of $\ker F$ and the *rank* of F is the dimension of $\mathcal{R}(F)$.

Theorem 2.4. *(Dimension Theorem) Let $F : R^n \to R^m$ be a linear transformation. Then*

$$\textit{rank of } F + \textit{ nullity of } F = n$$
$$(\textit{i.e.}\ \ \dim \mathcal{R}(F) + \dim \ker F = n).$$

◊ 2.16: Proof ...

◊ 2.17: Example ...

Definition: Let $G : \mathbb{R}^k \to \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be functions. Then the *composition*, $F \circ G$, of F and G is the function $(F \circ G) : \mathbb{R}^k \to \mathbb{R}^m$ defined by $(F \circ G)(\mathbf{u}) = F(G(\mathbf{u}))$.

Note:
(i) $(F \circ G)$ maps $\mathbb{R}^k$ to $\mathbb{R}^m$.
(ii) $(G \circ F)$ can only be defined if $m = k$ i.e. the range of F must lie in the domain of G.

Theorem 2.5. *Let $G : \mathbb{R}^k \to \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be two linear transformations with standard matrices $B(n \times k)$ and $A(m \times n)$ respectively. Then $F \circ G$ is a linear transformation with standard matrix $AB(m \times k)$.*

◊ 2.18: Proof ...

$\Diamond$ 2.19: Examples ...

Given two linear transformations $F, G : \mathbb{R}^n \to \mathbb{R}^n$, if the standard matrices of F and G are A and A^{-1} respectively, then the standard matrix of $F \circ G$ is simply

$$AA^{-1} = I_n \ ,$$

and so

$$(F \circ G)(\mathbf{u}) = I_n\mathbf{u} = \mathbf{u} \quad \forall\, \mathbf{u} \in \mathbb{R}^n \ .$$

This leads to the concept of invertible linear transformations.

- A linear transformation $F : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if there is a function $G : \mathbb{R}^n \to \mathbb{R}^n$ such that
$$(F \circ G)(\mathbf{u}) = \mathbf{u} = (G \circ F)(\mathbf{u}) \quad \forall\, \mathbf{u} \in \mathbb{R}^n \ .$$
- G, the *inverse* of F, is unique and is denoted by F^{-1}.
- If F has standard matrix A, then F is invertible iff A is invertible and the standard matrix of F^{-1} is A^{-1}.
- The composition of two invertible linear transformations is also an invertible linear transformation. The set of invertible linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ is known as the general linear group $GL(n, \mathbb{R})$ and has important applications in physics and geometry.
- A linear transformation $F : \mathbb{R}^n \to \mathbb{R}^n$ is invertible iff $\mathrm{Ker}F = \{\mathbf{0}\}$ and $\mathcal{R}(F) = \mathbb{R}^n$,

Chapter 3

SYMMETRIC MATRICES

3.1 The general quadratic equation in 2 variables [Poole: 5.5]

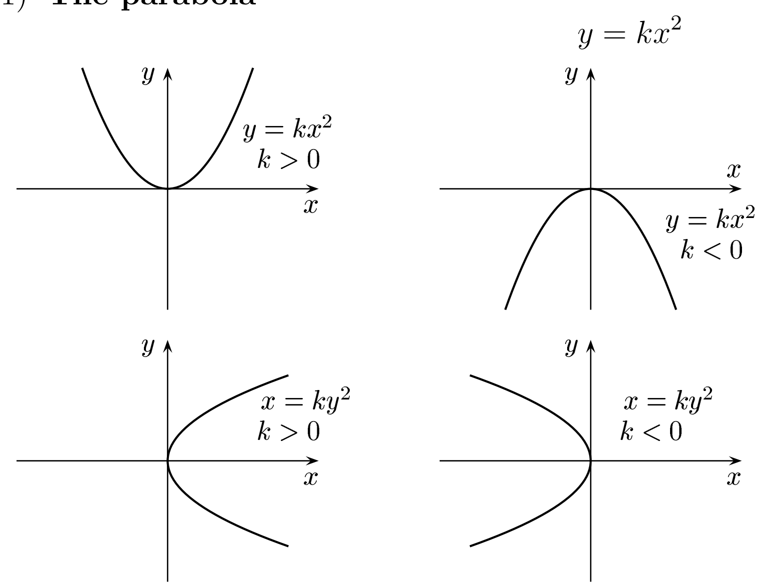
We will consider the general *quadratic* equation in two variables x, y :

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

where $a, b, c, \ldots$ are (constant) real numbers and at least one of $a, b, c \neq 0$. (if $a = b = c = 0$ we have a *linear* equation with the graph of a straight line.)

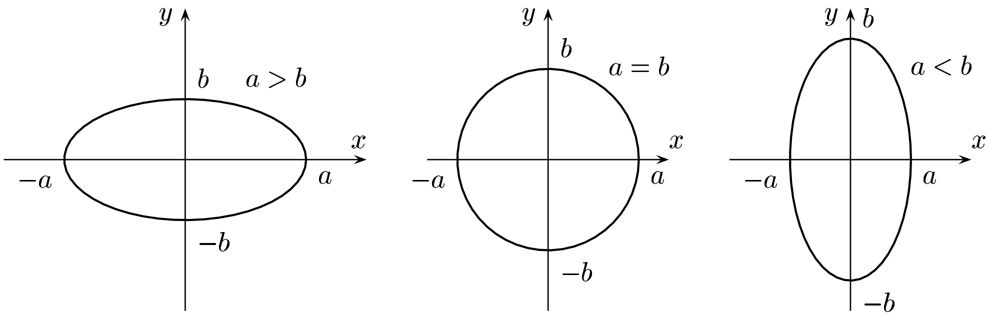
First we will look at the case when $b = 0$ and at least one of $a, c \neq 0$. We first list the possible curves in "standard form" — these are all called "conic sections". They are called "conic sections" because they all arise as plane sections of a cone; i.e. the curve formed when a plane cuts a cone at different angles. In the following lectures you will see how to eliminate the xy-term by "rotating axes", using eigenvalues.

(1) **The parabola**



(2) **The ellipse**

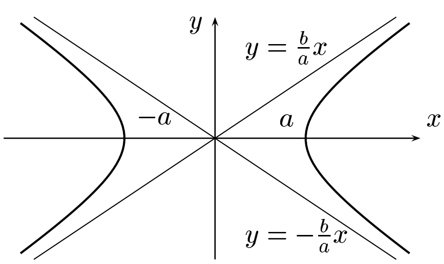
$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$



◊ 3.0: Key features of an ellipse ...

(3) **The hyperbola**

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$



◊ 3.1: Key features of a hyperbola ...

In addition there are some "degenerate" cases, one being

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$$

i.e.,

$$\begin{aligned} y^2 &= \frac{b^2x^2}{a^2} \\ y &= \pm \frac{b}{a}\, x. \end{aligned}$$

In this case we have a pair of straight lines.

The parabola, ellipse and hyperbola are the three different types of conic sections which can arise from the general quadratic. We now consider how to reduce the equation

$$ax^2 + cy^2 + dx + ey + f = 0$$

to one of these forms. The method is to "complete the square" on x if $a \neq 0$ and also on y when $c \neq 0$; and then translate the axes.

$\Diamond$ **3.2: Examples** ...

3.2 Eigen-properties [Poole: §4.3]

Recall that a non-zero vector $\mathbf{v}$ is an *eigenvector* of the $n \times n$ matrix A if there is a number λ, such that

$$A\mathbf{v} = \lambda\mathbf{v}$$

λ is the *eigenvalue* corresponding to $\mathbf{v}$.

To find the eigenvalues and eigenvectors of a matrix A:

(i) solve the *characteristic equation* $|\lambda I - A| = 0$ for λ;

(ii) for each eigenvalue λ, solve the vector equation $(\lambda I - A)\mathbf{x} = \mathbf{0}$ to determine the eigenvectors $\mathbf{x}$.

◊ 3.3: Example . . .

Diagonalisation [Poole: §4.4]

The $n \times n$ matrix A is diagonalized by the invertible matrix P if $P^{-1}AP = D$ is a diagonal matrix.

Not all matrices can be diagonalised. In fact A is diagonalisable if and only if A has n linearly independent eigenvectors and the diagonal matrix D consists of the n eigenvalues of A. The order in which the eigenvectors are placed in P determines the order of the corresponding eigenvalues in D.

◊ 3.4: Example . . .

Definition: A matrix A is *symmetric* if $A^t = A$.

◊ 3.5: Example . . .

Theorem 3.1. *The eigenvalues of a real symmetric matrix are real. Consequently, a basis for each eigenspace of A may be composed entirely of vectors in $\mathbb{R}^n$.*

◊ 3.6: Proof . . .

◊ **3.7: Example** ...

Theorem 3.2. *Let λ be an eigenvalue of a real symmetric matrix A with multiplicity m; then the corresponding eigenspace E_λ is m dimensional, i.e.*

$$\dim E_\lambda = \textit{multiplicity of } \lambda = m.$$

◊ **3.8: Example** ...

Theorem 3.3. *If A is a real symmetric matrix then the eigenvectors corresponding to distinct eigenvalues are orthogonal.*

◊ **3.9: Proof** ...

◊ **3.10: Example** ...

3.3 Orthogonal Diagonalisation [Poole: §5.4]

Definition: An $n \times n$ matrix P is said to be *orthogonal* if it is invertible and $P^{-1} = P^t$.

◊ 3.11: Example ...

Theorem 3.4. *The following statements are equivalent for an $n \times n$ matrix P.*

(i) P *is an orthogonal matrix.*

(ii) $P^tP = I$.

(iii) $PP^t = I$.

(iv) *The rows of P form an orthonormal basis for $\mathbb{R}^n$.*

(v) *The columns of P form an orthonormal basis for $\mathbb{R}^n$.*

◊ 3.12: Proof ...

Definition: An $n \times n$ matrix A is called *orthogonally diagonalisable* if there exists an orthogonal matrix P such that $D = P^{-1}AP (= P^tAP)$ is diagonal. The matrix P is said to *orthogonally diagonalise* the matrix A.

Evidently, an $n \times n$ matrix A is orthogonally diagonalisable if and only if it has an orthonormal set of n real eigenvectors (since A is diagonalisable and the matrix P which diagonalises is orthogonal).

Theorem 3.5. *A real $n \times n$ matrix A is orthogonally diagonalisable if and only if it is symmetric.*

◊ 3.13: Proof ...

◊ 3.14: Example of Orthogonal diagonalisation ...

3.4 Quadratic Forms: Conic Sections [Poole: §5.5]

The general quadratic equation in two variables

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

can always be written in matrix form:

$$\mathbf{x}^t A\mathbf{x} + K\mathbf{x} + f = 0$$

where $A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$ is a *symmetric* matrix, $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $K = \begin{bmatrix} d & e \end{bmatrix}$.

The first part of the equation, $\mathbf{x}^t A\mathbf{x}$ is called a *quadratic form* in two variables.

More generally, if A is a symmetric $n \times n$ (real) matrix and $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{x}^t A\mathbf{x}$ is a *quadratic form* in n variables.

◊ **3.15: Example** . . .

We use orthogonal diagonalisation to transform a quadratic equation into Standard Form.

We first consider the case when $K = \begin{bmatrix} 0 & 0 \end{bmatrix}$.

General Procedure

(a) Rewrite $ax^2 + bxy + cy^2 + f = 0$ as

$$\mathbf{x}^t A\mathbf{x} + f = 0$$

where $A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$ is real and symmetric.

(b) Find an orthogonal matrix P which (orthogonally) diagonalises A and such that the order of the columns of P ensure $\det P = 1$.

If P is an orthogonal matrix with $\det P = 1$ then P can be written in the form

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

(This follows because the columns of P form an orthonormal basis for $\mathbf{R}^2$.)

(c) Make the substitution $\mathbf{x} = P\mathbf{x}'$ to transform the quadratic equation to

$$(\mathbf{x}')^t \underbrace{P^t A P}_{D} \mathbf{x}' + f = 0$$

i.e.

$$\lambda_1 (x')^2 + \lambda_2 (y')^2 + f = 0$$

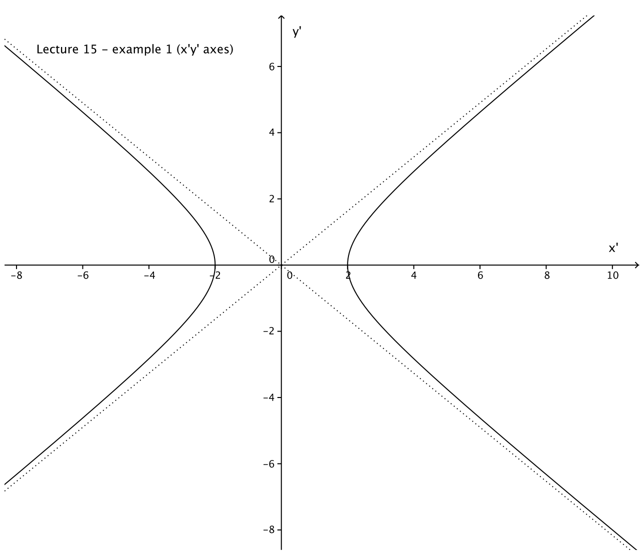
where $D = \operatorname{diag}(\lambda_1, \lambda_2)$;

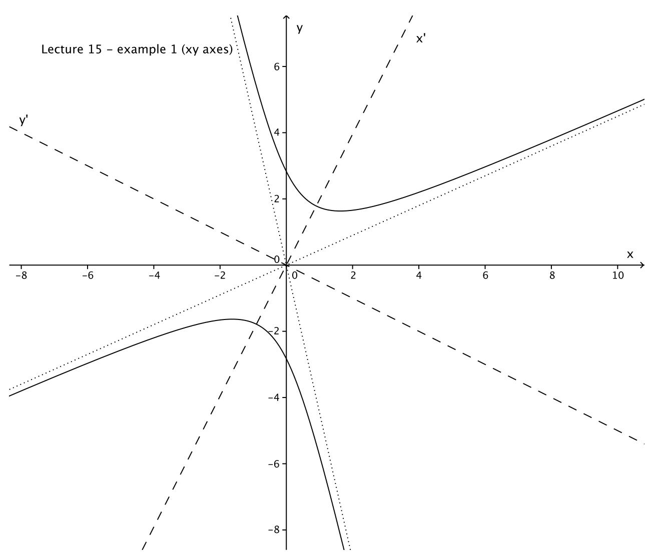
This substitution corresponds to a rotation of axes as $P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is a rotation.

(Refer to Chapter 7 on Linear Transformations.)

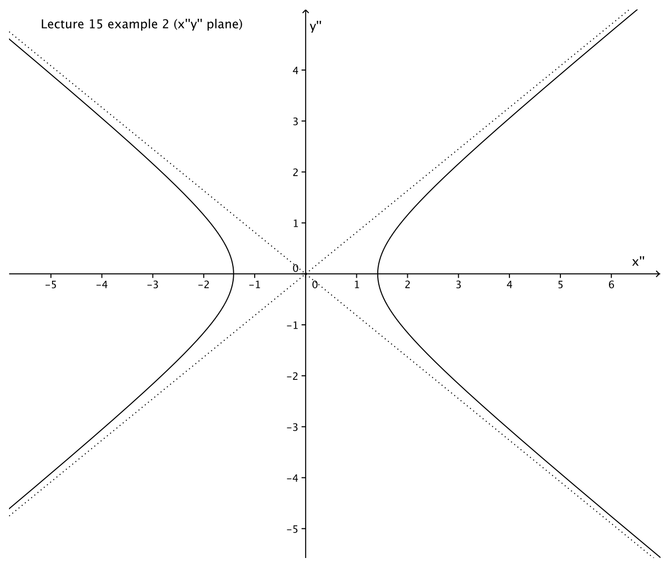
(d) Rearrange this simplified form into Standard Form.

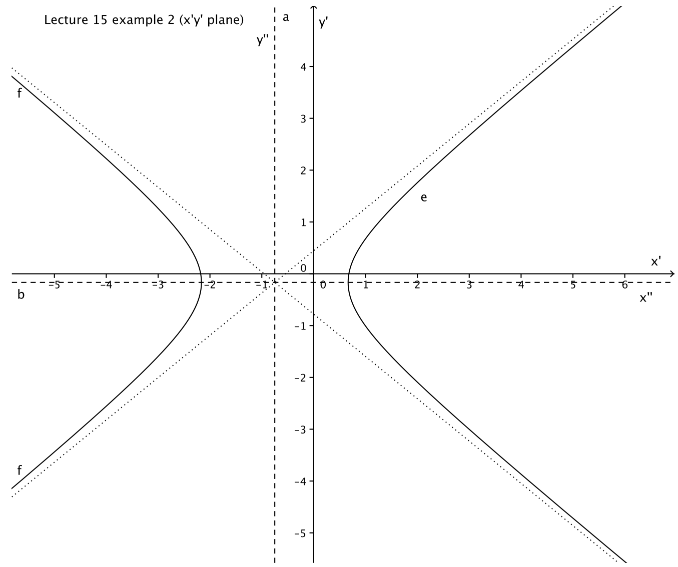
♢ **3.16: Example 1: Consider** $-2x^2 + 4xy + y^2 = 8$ ...

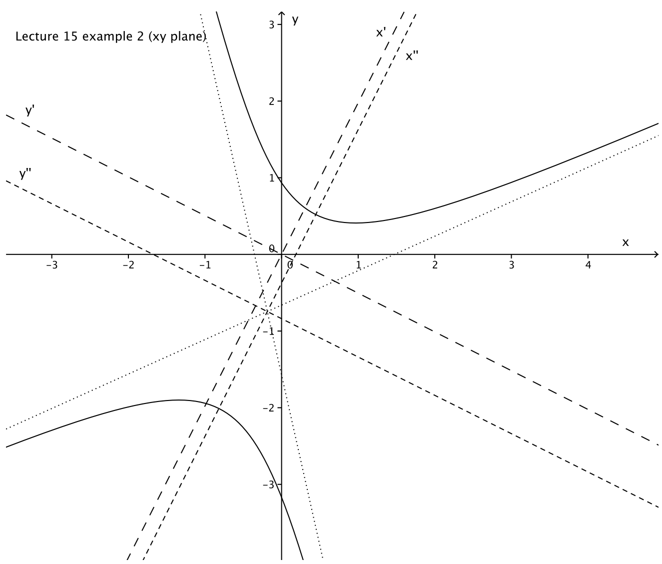




$\Diamond$ 3.17: Example 2: $(K \neq [0,0])$ Consider $-2x^2 + 4xy + y^2 + \sqrt{5}\,x + \sqrt{5}\,y = 71/24\ \ldots$







General Procedure $(K \neq [0,0])$

(a),(b),(c) as before.

(d) Complete the squares to obtain

$$\lambda_1(x' - x_0)^2 + \lambda_2(y' - y_0)^2 = f' \ .$$

(e) Substitute $\left.\begin{array}{l} x'' = x' - x_0 \\ y'' = y' - y_0 \end{array}\right\}$ translates $x'y'$ axes, x_0 units right, y_0 units up.

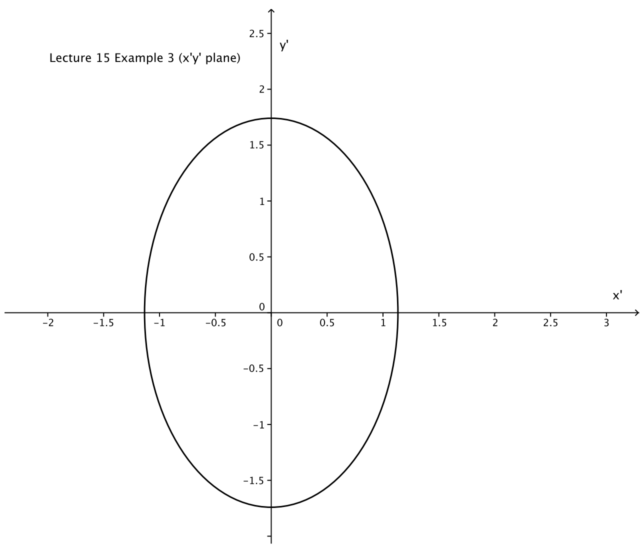
(f) Rearrange this simplified form into Standard Form.

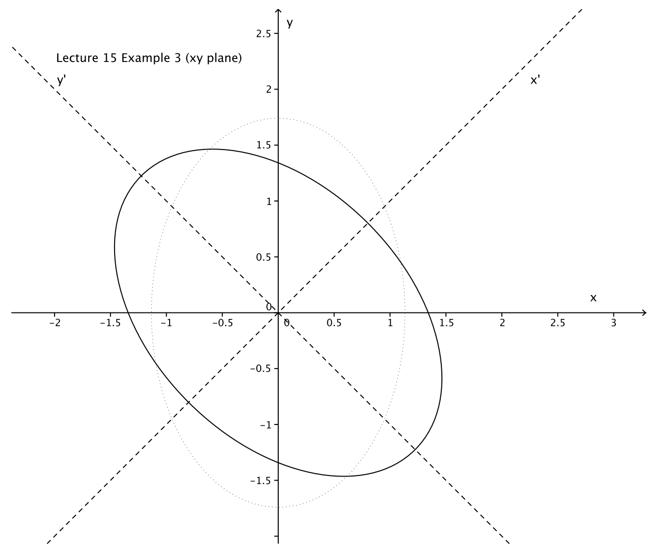
◊ 3.18: Example 3:

Consider the quadratic equation

$$5x^2 + 4xy + 5y^2 = 9$$

...





Chapter 4

CALCULUS OF MORE THAN ONE VARIABLE

4.1 Surfaces in 3 dimensions [St: §12.5-12.6]

In the plane, $f(x,y) = 0$ is usually a curve; in three dimensions $F(x,y,z) = 0$ usually represents a surface. Often we can solve for z and write $z = f(x,y)$, where f is a function of two variables.

Planes

You are already familiar with the simplest case – the linear equation

$$ax + by + cz + d = 0$$

which represents a plane.

Cylinders

Consider the equation $\frac{x^2}{3^2} + y^2 = 1$ in $\mathbb{R}^3$.

In $\mathbb{R}^2$ this represents an ellipse; as z does not appear, z may take any value whatsoever with x, y being related by $\frac{x^2}{3^2} + y^2 = 1$. (Compare with the equation of a line $x = 2$ in $\mathbb{R}^2$.) This is the surface of an elliptic cylinder in $\mathbb{R}^3$.

◊ 4.0: Examples ...

Quadric surfaces

An important class of surfaces, called the "quadric surfaces", arise from the general quadratic equation in three variables:

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0.$$

As for the two dimensional quadratic equation, this equation can be written in the form

$$\mathbf{x}^t A\mathbf{x} + K\mathbf{x} + j = 0$$

where A is a symmetric 3×3 matrix, $K = \begin{bmatrix} g & h & i \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

Using exactly the same method as for two dimensions, that is orthogonally diagonalizing A by an orthogonal matrix P (with $\det P = 1$), we can eliminate the xy, xz, yz terms. Completing the square then puts the equation in standard form.

$\Diamond$ **4.1: Example ...**

Some examples of quadric surfaces in standard form are now presented.

Examples of quadric surfaces:

(1) **The ellipsoid**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

$\Diamond$ **4.2: Calculations ...**

(2) **The hyperboloid of one sheet**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

$\Diamond$ **4.3: Calculations ...**

(3) **The hyperboloid of two sheets**

$$-\frac{x^2}{a^2} - \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

$\Diamond$ **4.4: Calculations ...**

(4) **The elliptic cone**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}\,.$$

(5) **The elliptic paraboloid**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c}\,.$$

(6) **The hyperbolic paraboloid**

$$-\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z}{c}\,, \qquad c > 0.$$

Some further possibilities: (7) no solution $-x^2/l^2 - y^2/m^2 - z^2/n^2 = 1$
(8) elliptic cylinder $x^2/l^2 + y^2/m^2 = 1$
(9) hyperbolic cylinder $x^2/l^2 - y^2/m^2 = 1$
(10) parabolic cylinder $x^2/l^2 + z = 0$.

$\Diamond$ **4.5: Examples ...**

4.2 Functions of Several Variables [St: §14.1]

So far in the Calculus section of Mathematics 1 we have only considered functions of one variable. However many real-world functions depend upon two or more variables. The main features of single-variable calculus — limits, derivatives, chain rule, maximum-minimum techniques — all generalise to functions of several variables.

Definition: Let $\mathcal{D}$ be a subset of R^2. Suppose there is a relation which assigns to each (x, y) in $\mathcal{D}$ a real number $f(x, y)$. Then f is said to be a function of two variables with domain $\mathcal{D}$.

◊ 4.6: Notes ...

Definition: Let f be a function of two variables with domain $\mathcal{D}$. The surface consisting of all points (x, y, z) of R^3 such that

$$z = f(x, y)$$

is called the *graph* of f.

◊ 4.7: Examples ...

Definition: The intersection of the horizontal plane $z = k$ with the surface $z = f(x, y)$ is called the *contour curve* of height k on the surface.

The vertical projection of this contour curve onto the xy-plane is called the *level curve* $f(x, y) = k$ of function f. Thus the level curves of f are curves in the xy-plane on which the value of f is constant.

◊ 4.8: Examples ...

Polar Coordinates [St: §10.3]

Polar coordinates give another way of specifying points in the cartesian plane. The point P has polar coordinates (r, θ), $r > 0$, if $|OP| = r$ and OP makes an angle θ with the positive x-axis.

Unlike cartesian coordinates, each point P has *many* polar coordinates. For example, (r, θ) and $(r, \theta + 2\pi)$ both represent P as do all points of the form $(r, \theta + 2k\pi)$, $k = 0, \pm 1, \pm 2, \ldots$.

◊ 4.9: Example ...

Relationship between polar and cartesian coordinates

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ r^2 &= x^2+y^2 \end{aligned}$$

$$\tan\theta = y/x \qquad \text{or} \qquad \theta = \arctan y/x.$$

◊ **4.10: Example** ...

4.3 Limits and Continuity [St: §14.2]

For functions of a single variable we effectively verify there is a limit L at a point $x = a$ by checking that the two one-sided limits exist and are equal to L. That is, we approach $x = a$ from two directions.

For functions of two variables we need to verify the limit is L at $(x, y) = (a, b)$ by checking from every direction in the plane of the domain.

Suppose f is a function of two variables with domain $\mathcal{D}$ and let $(a, b) \in R^2$. An intuitive description of the limit at (a, b) is:

A number L is the *limit* of $f(x, y)$ as (x, y) approaches (a, b) if $f(x, y)$ is close to L whenever (x, y) is close to (a, b).

In many cases no limit exists. When it exists, the limit L is unique and is denoted by

$$\lim_{(x,y)\to(a,b)} f(x,y).$$

We write $L = \lim_{(x,y)\to(a,b)} f(x,y)$.

◊ **4.11: Examples** ...

Theorem 4.1. *If* $\lim_{(x,y)\to(a,b)} f(x,y)$ exists *and equals* L*, then*

$$\lim_{x\to a} f(x,b) = \lim_{y\to b} f(a,y) = L. \tag{1}$$

Note: The converse is not true, as the limits (1) can exist and be equal without the double limit existing, as with $(x, y) \to (0, 0)$ for $f(x, y) = \dfrac{xy}{x^2+y^2}$ above.

Continuity

Definition: A function f of two variables is said to be continuous at the point (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x,y) \quad \text{exists and equals } f(a,b).$$

For the definition to be applicable, the domain $\mathcal{D}$ of f must contain a region/ball properly containing (a, b).

◊ 4.12: Notes ...

To check the continuity of functions defined by more than one rule, it is often necessary to evaluate limits at particular points.

◊ 4.13: Example ...

4.4 Partial Derivatives [St: §14.3]

If f is a function of several variables, we can investigate the rate of change of f with respect to just one of the variables.

Definition: Let f be a function of two variables x, y. The *partial derivative of f with respect to x* is the function f_x (of the two variables x, y) obtained by holding y constant and differentiating with respect to x. Thus if $(a, b) \in \mathcal{D}(f)$,

$$f_x(a, b) = \lim_{x \to a} \frac{f(x, b) - f(a, b)}{x - a}$$

if the limit exists.

Alternatively,

$$f_x(a, b) = \lim_{h \to 0} \frac{f(a + h, b) - f(a, b)}{h} .$$

Similarly, the *partial derivative of $f(x, y)$ w.r.t. y, f_y*, is obtained by holding x constant and differentiating w.r.t. y.

◊ 4.14: Example ...

Notation: If f is a function of variables x and y, and $z = f(x, y)$, common notations for the first derivative w.r.t. x are:

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x} = f_x(x, y) = D_x f(x, y) = D_1 f(x, y).$$

Note that the symbol "∂" is used instead of "d" to denote the partial derivative.

The subscript "1" in $D_1 f(x, y)$ denotes partial differentiation w.r.t. the first listed variable "x".

Similarly,

$$\frac{\partial z}{\partial y} = \frac{\partial f}{\partial y} = f_y(x, y) = D_y f(x, y) = D_2 f(x, y)$$

◊ 4.15: Example ...

The partial derivative F_x of a function of three variables x, y, z is the function of three variables obtained by holding y and z constant and differentiating w.r.t. x. Thus

$$F_x(x, y, z) = \lim_{h \to 0} \frac{F(x + h, y, z) - F(x, y, z)}{h}$$

and similarly for $F_y(x, y, z)$ and $F_z(x, y, z)$.

◊ 4.16: Example ...

Geometric interpretation of partial derivatives

Recall $z = f(x, y)$ represents a surface and if $c = f(a, b)$ then the point $P(a, b, c)$ lies on the surface. Taking $y = b$, we restrict our attention to the intersection of the plane $y = b$ and the surface, which is a curve in the plane.

The partial derivative $f_x(a, b)$ is the slope of the line at P tangent to the curve formed as the intersection of the plane $y = b$ and the surface $z = f(x, y)$.

Similarly, $f_y(a, b)$ is the slope of the tangent at P which is on the curve of intersection of the plane $x = a$ and the surface $z = f(x, y)$.

We illustrate this with an example.

◊ 4.17: Example ...

4.5 Tangent Planes [St: §14.4]

Consider the surface $z = f(x, y)$ with $c = f(a, b)$, so $P(a, b, c)$ is a point on the surface. Let C_1 be the curve which is the intersection of $z = f(x, y)$ and the plane $y = b$, with T_1 the tangent to C_1 at P. Similarly let T_2 be the tangent to C_2 at P, C_2 the curve which is the intersection of $z = f(x, y)$ and $x = a$.

These two tangent lines determine a unique *tangent plane* to the surface $z = f(x, y)$ at $P(a, b, c)$.

Definition: Suppose f is a function of two variables and suppose f_x and f_y are continuous. The *tangent plane* to the surface $z = f(x, y)$ at point $P(a, b, c)$ is the plane through P containing the tangent lines at P to

(i) $z = f(x, b)$ (with slope $f_x(a, b)$),

(ii) $z = f(a, y)$ (with slope $f_y(a, b)$).

♢ **4.18: Derivation of formula** . . .

The equation of the tangent plane at $P(a, b, c)$ is:

$$f_x(a, b)(x - a) + f_y(a, b)(y - b) - (z - c) = 0$$

♢ **4.19: Example** . . .

4.6 Higher Partial Derivatives [St: §14.3]

The first partial derivatives f_x and f_y are themselves functions of x and y. Hence we can take the partial derivatives of f_x and f_y (when they exist), obtaining the *second-order partial derivatives* of f.

Definition:

$$\begin{aligned}
(f_x)_x &= f_{xx} = \frac{\partial f_x}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x^2} = D_{11}f \\
(f_x)_y &= f_{xy} = \frac{\partial f_x}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x} = D_{12}f \\
(f_y)_x &= f_{yx} = \frac{\partial f_y}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial x \partial y} = D_{21}f \\
(f_y)_y &= f_{yy} = \frac{\partial f_y}{\partial y} = \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = \frac{\partial^2 f}{\partial y^2} = D_{22}f
\end{aligned}$$

If $z = f(x, y)$, we could replace f by z in the above expressions.

Note: f_{xy} is the second-order partial derivative of f w.r.t. x first, then y; f_{yx} is the second-order partial of f w.r.t. y first, then x.

Theorem 4.2. *If f_{xy} and f_{yx} exist in a neighbourhood of (a, b) and are continuous at (a, b) then*

$$f_{xy}(a, b) = f_{yx}(a, b)$$

◊ **4.20: Example** ...

4.7 The Chain Rule [St: §14.5]

We use the chain rule for differentiating functions of functions.

Recall that for functions of functions of a single variable x, e.g. if $g(x) = f(u(x))$, then

$$\frac{dg}{dx} = \frac{df}{du}\,\frac{du}{dx}\,.$$

For functions of several variables we consider two situations:

(i) $f(x, y)$ where x and y are functions of t, $x = g(t)$ and $y = h(t)$. Then $w = f(g(t), h(t))$ is a function of t and we can consider $w'(t) = dw/dt$.

(ii) $w = f(x, y)$ where x and y are each functions of u and v. Then w is a function of u and v and we can consider $\partial w/\partial u$ and $\partial w/\partial v$.

Theorem 4.3. *(Chain Rule (1): Suppose $w = f(x, y)$ has continuous first-order partial derivatives and that $x = g(t)$ and $y = h(t)$ are differentiable functions. Then w is a differentiable function of t and*

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\,\frac{dx}{dt} + \frac{\partial w}{\partial y}\,\frac{dy}{dt} \tag{2}$$

[Note: The partial derivatives here are to be evaluated at the point $(g(t), h(t))$, so (2) is more precisely stated as

$$\frac{d}{dt}\,f(g(t), h(t)) = \frac{\partial}{\partial x}\,f(g(t), h(t))\,\frac{dg}{dt} + \frac{\partial}{\partial y}\,f(g(t), h(t))\,\frac{dh}{dt}\,.]$$

We refer to w as the *dependent* variable, x and y as *intermediate* variables and t as the *independent* variable.

◊ **4.21: Example** ...

This theorem generalises to functions of 3 or more intermediate variables, all a function of the same independent variable.

Theorem 4.4. *(Chain Rule (2): Suppose $w = f(x, y)$ has continuous first-order partial derivatives. If $x = g(u, v)$ and $y = h(u, v)$ also have continuous first-order partial derivatives then*

$$\begin{aligned} \frac{\partial w}{\partial u} &= \frac{\partial w}{\partial x}\,\frac{\partial x}{\partial u} + \frac{\partial w}{\partial y}\,\frac{\partial y}{\partial u} \\ \textit{and} \qquad \frac{\partial w}{\partial v} &= \frac{\partial w}{\partial x}\,\frac{\partial x}{\partial v} + \frac{\partial w}{\partial y}\,\frac{\partial y}{\partial v} \end{aligned} \tag{3}$$

◊ **4.22: Example** ...

4.8 Directional Derivatives and the Gradient Vector [St: §14.6]

Recall that if $z = f(x, y)$, the partial derivatives f_x and f_y represent the rates of change (w.r.t. distance) of z in the x and y directions, respectively.

Suppose we wish to find the rate of change of z at (x_0, y_0) in the direction of the arbitrary unit vector $\mathbf{u} = (a, b)$.

◊ 4.23: Derivation of formula ...

Definition: The *directional derivative* of f at (x_0, y_0) in the direction of the unit vector $\mathbf{u} = (a, b)$ is

$$D_{\mathbf{u}}f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if this limit exists. $D_{\mathbf{u}}f(x_0, y_0)$ is often written as $f_{\mathbf{u}}(x_0, y_0)$.

Theorem 4.5. *If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $\mathbf{u} = (a, b)$ and*

$$D_{\mathbf{u}}f(x_0, y_0) = f_x(x_0, y_0).a + f_y(x_0, y_0).b$$

If the unit vector $\mathbf{u}$ makes an angle θ with the positive x-axis, in the xy-plane then $\mathbf{u} = (\cos\theta, \sin\theta)$ and

$$D_{\mathbf{u}}f(x, y) = f_x(x, y)\cos\theta + f_y(x, y)\sin\theta$$

◊ 4.24: Proof ...

◊ 4.25: Example ...

Definition: If f is a function of two variables x and y, the *gradient of f* at (x, y) is the vector function ∇f defined by

$$\nabla f(x, y) = f_x(x, y)\mathbf{i} + f_y(x, y)\mathbf{j} = (f_x(x, y), f_y(x, y))$$

∇f is also written "grad f" and ∇f is pronounced "del f".

Observe that $f_x(x, y)a + f_y(x, y)b = \nabla f(x, y) \cdot (a, b)$, i.e.

$$D_{\mathbf{u}}f(x, y) = \nabla f(x, y) \cdot \mathbf{u}$$

This expresses the directional derivative in the direction of $\mathbf{u}$ as the scalar projection of ∇f onto $\mathbf{u}$.

The directonal derivative may be described geometrically as the rate of increase of the function in the specified direction.

Question: In which direction (in the xy-plane) does $z = f(x, y)$ change fastest and what is the maximum rate of change?

Theorem 4.6. *Suppose f is a differentiable function of two variables x and y. The maximum value of $D_{\mathbf{u}}f(x,y)$ is $|\nabla f(x,y)|$ and it occurs when $\mathbf{u}$ has the same direction as $\nabla f(x,y)$.*

◊ 4.26: Proof . . .

◊ 4.27: Example . . .

The gradient vector ∇f points in the direction in which f increases most rapidly, and its length is the rate of increase of $z = f(x,y)$ (w.r.t. distance) in that direction.

When is $D_{\mathbf{u}}f(x,y)$ zero? Along a *level curve* in the domain, by definition, the value of $z = f(x,y)$ does not change ($z = f(x,y) =$ constant along a level/contour curve). That is, if $\mathbf{u}$ is in the direction of the level curve through (x_0, y_0) then

$$D_{\mathbf{u}}f(x_0,y_0) = 0, \quad \text{i.e.,} \quad \nabla f(x_0,y_0) \cdot \mathbf{u} = 0$$

So, provided $\nabla f \neq \mathbf{0}$, ∇f must be *perpendicular (normal)* to the level curve through (x_0, y_0).

◊ 4.28: Example . . .

The Second Derivative Test for Functions of Two Variables [St: §14.7]

How do we find the extrema (i.e. maxima and minima) of functions of two variables?

Theorem 4.7. *Necessary conditions for the existence of a local extremum at a point* (a, b) :

Let $f(x, y)$ *be continuous in some neighbourhood containing the point* (a, b)*. If* $f_x(a, b)$ *and* $f_y(a, b)$ *exist, then the* necessary *condition for* f *to have a relative extremum at* (a, b) *is that both* $f_x(a, b) = 0$ *and* $f_y(a, b) = 0$*.*

That is, the tangent plane to $z = f(x, y)$ must be horizontal at any local maximum or minimum point $(a, b, f(a, b))$.

A point (a, b) such that $f_x(a, b) = 0 = f_y(a, b)$ or where either partial derivative does not exist, is called a critical point.

◊ 4.29: Examples . . .

Note that a point where

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$$

may correspond to either a local maximum, minimum or neither. Thus this *theorem is not a sufficient condition* for a local extremum since it admits the case of a *saddle point* which is *not* an extremum.

◊ 4.30: Examples . . .

Theorem 4.8. *Second Derivative Test*
Let $f(x, y)$ *be a function of two variables whose first and second order partial derivatives are continuous in some neighbourhood of a point* (a, b)*. Note that this means* $f_{xy} = f_{yx}$*. Suppose* $f_x(a, b) = 0$ *and* $f_y(a, b) = 0$*, and let*

$$\begin{aligned} \Delta(a, b) &= f_{xx}(a, b) \cdot f_{yy}(a, b) - (f_{xy}(a, b))^2 \\ &= \left| \begin{matrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{matrix} \right|_{(a,b)} \end{aligned}$$

Then (i) f *has a relative minimum at* (a, b) *if* $\Delta(a, b) > 0$ and $f_{xx}(a, b) > 0$ *(or* $f_{yy}(a, b) > 0$*),*

(ii) f *has a relative maximum at* (a, b) *if* $\Delta(a, b) > 0$ and $f_{xx}(a, b) < 0$ *(or* $f_{yy}(a, b) < 0$*),*

(iii) f *has a saddle point at* (a, b) *if* $\Delta(a, b) < 0$*, and*

(iv) *if* $\Delta(a, b) = 0$*, no conclusion can be made.*

The expression

$$\Delta(a, b) = f_{xx}(a, b) \cdot f_{yy}(a, b) - [f_{xy}(a, b)]^2 = \left| \begin{matrix} f_{xx}(a, b) & f_{xy}(a, b) \\ f_{xy}(a, b) & f_{yy}(a, b) \end{matrix} \right|$$

is called the *discriminant* of $f(x,y)$ at (a,b).

Summary:

$$\left.\begin{array}{ccc} & \Delta > 0 & \\ \swarrow & & \searrow \\ f_{xx} > 0 & & f_{xx} < 0 \\ \text{rel. min.} & & \text{rel. max} \\ & \Delta < 0 & \\ & \text{saddle point} & \\ & \Delta = 0 & \\ & \text{no conclusion} & \end{array}\right\} \begin{array}{l} \text{only applies when} \\ f_x(a,b) = 0 = f_y(a,b). \end{array}$$

◊ 4.31: Examples ...