# Chapter 1

# Introduction

One of our goals in this book is to equip the reader with a unifying view of linear algebra, or at least of what is traditionally studied under this name in university courses. With this mission in mind, we start with a *preview* of the subject, and describe its main achievements in lay terms.

To begin with a few words of praise: linear algebra is a very simple and useful subject, underlying most of other areas of mathematics, as well as its applications to physics, computer science, engineering, and economics. What makes linear algebra useful and efficient is that it provides ultimate solutions to several important mathematical problems. Furthermore, as should be expected of a truly fruitful mathematical theory, the problems it solves can be formulated in a rather elementary language, and make sense even before any advanced machinery is developed. Even better, the *answers* to these problems can also be described in elementary terms (in contrast with the *justification* of those answers, which better be postponed until adequate tools are developed). Finally, those several problems we are talking about are similar in their nature; namely, they all have the form of problems of *classification* of very basic mathematical objects.

Yet unready to discuss the general idea of classification in mathematics, we begin with presenting explicitly the main problems of linear algebra, and the answers to these problems, in elementary, down-to-earth terms. After that, we will work out a non-trivial example: classification of quadratic curves on the plane. With that example at hands, we will say a few words about the idea of classification in general, and about the layout of the book.

# 1 Problems of Linear Algebra

We formulate here the four model problems of linear algebra, solve them by bare hands in the simplest case of dimension 1, and state the answers in general.

## The Rank Theorem

`Question.` *Given $m$ linear functions in $n$ variables,*

$$\begin{array}{rcl} y_1 & = & a_{11}x_1 + ... + a_{1n}x_n \\ & ... & \\ y_m & = & a_{m1}x_1 + ... + a_{mn}x_n \end{array},$$

*what is the simplest form to which they can be transformed by linear changes of the variables,*

$$\begin{array}{rcl} y_1 & = & b_{11}Y_1 + ... + b_{1m}Y_m \\ ... & & \\ y_m & = & b_{m1}Y_1 + ... + b_{mm}Y_m \end{array}, \quad \begin{array}{rcl} x_1 & = & c_{11}X_1 + ... + c_{1n}X_n \\ & ... & \\ x_n & = & c_{n1}X_1 + ... + c_{nn}X_n \end{array} ?$$

`Example.` Consider a linear function in one variable: $y = ax$. We are allowed to make substitutions $y = bY$ and $x = cX$, where however $b \neq 0$ and $c \neq 0$ (so that we could reverse the substitutions). The substitutions will result in a new, transformed function: $Y = b^{-1}acX$. Clearly, if $a = 0$, then no matter what substitution we make, the linear function will remain identically zero. On the other hand, if $a \neq 0$, we can choose such values of $b$ and $c$ that the coefficient $b^{-1}ac$ becomes equal to 1 (e.g. take $b = 1$ and $c = a^{-1}$). Thus, every linear function $y = ax$ is either identically zero: $Y = 0$, or can be transformed to $Y = X$.

`Theorem.` ***Every system of*** $m$ ***linear functions in*** $n$ ***variables can be transformed by suitable linear changes of dependent and independent variables to exactly one of the normal forms:***

$$Y_1 = X_1, \quad \dots, \quad Y_r = X_r, \quad Y_{r+1} = 0, \quad \dots, \quad Y_m = 0,$$

***where*** $0 \leq r \leq m, n$***.***

The number $r$ featuring in the answer is called the **rank** of the given system of $m$ linear functions.

## The Inertia Theorem

`Question.` *Given a* **quadratic form** *(i.e. a homogeneous quadratic function) in $n$ variables,*

$$Q = q_{11}x_1^2 + 2q_{12}x_1x_2 + 2q_{13}x_1x_3 + ... + q_{nn}x_n^2,$$

*what is the simplest form to which it can be transformed by a linear change of the variables*

$$\begin{array}{rcl} x_1 & = & c_{11}X_1 + ... + c_{1n}X_n \\ & ... & \\ x_n & = & c_{n1}X_1 + ... + c_{nn}X_n \end{array} \quad ?$$

`Example.` A quadratic form in one variable, $x$, has the form $qx^2$. A substitution $x = cX$ (with $c \neq 0$), transforms it into $qc^2X^2$. Of course, if $q = 0$, no substitution will change the fact that the function is identically zero. When $q \neq 0$, we can make the absolute value of coefficient $qc^2$ equal to 1 (by choosing $c = \pm\sqrt{|q^{-1}|}$). However, no substitution will change the sign of the coefficient (that is, a positive quadratic form will remain positive, and negative will remain negative). Thus, every quadratic form in one variable can be transformed to exactly one of these: $X^2$, $-X^2$, or 0.

`Theorem.` ***Every quadratic form in*** $n$ ***variables can be transformed by a suitable linear change of the variables to exactly one of the normal forms:***

$$X_1^2 + ... + X_p^2 - X_{p+1}^2 - ... - X_{p+q}^2 \quad \textbf{\textit{where}}\ 0 \leq p + q \leq n.$$

Note that, in a way, the theorem claims that the $n$-dimensional case can be reduced to the sum (we will later call it "**direct sum**") of $n$ one-dimensional answers found in the example: $X^2$, $-X^2$, or 0. The possibility of such reduction of a higher-dimensional problem to the direct sum of one-dimensional problems is a standard theme of linear algebra.

The numbers $p$ and $q$ of positive and negative squares in the normal form are called **inertia indices** of the quadratic form in question. If the quadratic form $Q$ is known to be positive everywhere outside the origin, the Inertia Theorem tells us that in a suitable coordinate system $Q$ assumes the form $X_1^2 + ... + X_n^2$, i.e. its inertia indices are $p = n,\ q = 0$.

## The Orthogonal Diagonalization Theorem

`Question.` *Given two homogeneous quadratic forms in $n$ variables, $Q(x_1, ..., x_n)$ and $S(x_1, ..., x_n)$, of which the first one is known to be positive everywhere outside the origin, what is the simplest form to which they can be simultaneously transformed by a linear change of the variables?*

`Example.` In the case $n = 1$, we have $Q(x) = qx^2$, where $q > 0$, and $S(x) = sx^2$, where $s$ is arbitrary. As we know, the first quadratic form is transformed by the substitution $x = q^{-1/2}X$ into $X^2$. The same transformation will change $S$ into $\lambda X^2$ with $\lambda = sq^{-1}$. Of course, one can make $S$ to be $\pm X^2$ (if $s \neq 0$) by rescaling the variable once again, but this may destroy the form $X^2$ of the function $Q$. In fact the only substitutions $X = C\tilde{X}$ which *preserve* $Q$ (i.e. don't change the coefficient) are those with $C = \pm 1$. Unfortunately such substitutions do not affect at all the coefficient $\lambda$ in the function $S$: $\lambda X^2 = \lambda(\pm\tilde{X})^2 = \lambda\tilde{X}^2$. We conclude that each pair $Q$, $S$ can be transformed into one of the pairs $X^2, \lambda X^2$, where $\lambda$ is a real number, but two such pairs with different values of $\lambda$ cannot be transformed into each other.

`Theorem.` ***Every pair $Q$, $S$ of quadratic forms in $n$ variables, of which $Q$ is positive everywhere outside the origin, can be transformed by a linear change of the variables into exactly one of the normal forms***

$$Q = X_1^2 + ... + X_n^2, \;\; S = \lambda_1 X_1^2 + ... + \lambda_n X_n^2, \;\; \textbf{\textit{where}} \; \lambda_1 \geq ... \geq \lambda_n.$$

The real numbers $\lambda_1, \dots, \lambda_n$ are called **eigenvalues** of the given pair of quadratic forms (and are often said to form their **spectrum**).

Note that this theorem, too, reduces the $n$-dimensional problem to the "direct sum" of $n$ one-dimensional problems solved in our Example.

## The Jordan Canonical Form Theorem

The fourth question deals with a system of $n$ linear functions in $n$ variables. Such an object is the special case of systems of $m$ functions in $n$ variables when $m = n$. According to the Rank Theorem, such a system of rank $r \leq n$ can be transformed to the form $Y_1 = X_1, \dots, Y_r = X_r, Y_{r+1} = \cdots = Y_n = 0$ by linear changes of dependent and independent variables. There are many cases however where relevant information about the system is lost when dependent and independent variables are changed *separately*. This happens whenever both groups of variables describe objects in the same space (rather than in two different ones).

An important class of examples comes from the theory of Ordinary Differential Equations (ODE for short).

`Example.` Consider a linear first order ODE $\dot{x} = \lambda x$. It relates the values $x(t)$ of an unknown function, $x$, with its rate of change in time, $\dot{x}$ (which is the short notation for $dx/dt$). A rescaling of the function by $x = cX$ would make little sense if not accompanied with the simultaneous rescaling of the rate, $\dot{x} = c\dot{X}$ (we assume that the rescaling coefficient $c$ is time-independent). Unfortunately, such a rescaling does not affect the form of the equation: $\dot{X} = c^{-1}\lambda c X = aX$. We conclude that no two linear first order ODEs $\dot{x} = \lambda x$ with different values of the coefficient $\lambda$ can be transformed into each other by a linear change of the variable.

We will describe the fourth classification problem in the context of the ODE theory, although it can be stated more abstractly as a problem about $n$ linear functions in $n$ variables, to be transformed by a single linear change acting on both dependent and independent variables *the same way.*

`Question.` *Given a system of $n$ linear homogeneous 1st order constant coefficient ODEs in $n$ unknowns:*

$$\begin{array}{rcl} \dot{x}_1 & = & a_{11}x_1 + ... + a_{1n}x_n \\ & ... & \\ \dot{x}_n & = & a_{n1}x_1 + ... + a_{nn}x_n \end{array},$$

*what is the simplest form to which it can be transformed by a linear change of the unknowns:*

$$\begin{array}{rcl} x_1 & = & c_{11}X_1 + ... + c_{1n}X_n \\ & ... & \\ x_n & = & c_{n1}X_1 + ... + c_{nn}X_n \end{array} \quad ?$$

There is an advantage in answering this question *over* $\mathbb{C}$, i.e. assuming that the coefficients $c_{ij}$ in the change of variables, as well as the coefficients $a_{ij}$ of the given ODE system are allowed to be complex numbers. The advantage is due to the unifying power of the Fundamental Theorem of Algebra, discussed in Supplement "Complex Numbers."

`Example.` Consider a single $m$th order linear ODE of the form:

$$(\frac{d}{dt} - \lambda)^m y = 0, \quad \text{where } \lambda \in \mathbb{C}.$$

By setting

$$y = x_1,\ \frac{d}{dt}y - \lambda y = x_2,\ (\frac{d}{dt} - \lambda)^2 y = x_3,\ \ldots,\ (\frac{d}{dt} - \lambda)^{m-1} y = x_m,$$

the equation can be written as the following system of $m$ ODEs of the 1st order:

$$\begin{array}{lllllll} \dot{x}_1 & = & \lambda x_1 & + & x_2 & & \\ \dot{x}_2 & = & & & \lambda x_2 & + & x_3 \\ & & & \ldots & & & \\ \dot{x}_{m-1} & = & & & & \lambda x_{m-1} & + \; x_m \\ \dot{x}_m & = & & & & & \lambda x_m \end{array}.$$

Let us call this system the **Jordan block** of size $m$ with the eigenvalue $\lambda$. Introduce a **Jordan system** of several Jordan blocks of sizes $m_1, ..., m_r$ with the eigenvalues $\lambda_1, ..., \lambda_r$. It can be similarly compressed into the system

$$(\frac{d}{dt} - \lambda_1)^{m_1} y_1 = 0,\ ...,\ (\frac{d}{dt} - \lambda_r)^{m_r} y_r = 0$$

of $r$ *unlinked* ODEs of the orders $m_1, ..., m_r$.

The numbers $\lambda_1, \ldots, \lambda_r$ here are not assumed to be necessarily distinct. In fact they are the roots of a certain degree $n$ polynomial, $p(\lambda) = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_r)^{m_r}$, (called the **characteristic polynomial**), which can be associated with every linear ODE system, and does not change under the linear changes of the unknowns. When the polynomial has all its $n$ roots distinct (that is, all $m_i = 1$, and $r = n$), the Jordan system assumes the form $\dot{x}_1 = \lambda_1 x_1,\ \ldots,\ \dot{x}_n = \lambda_n x_n$ of $n$ unlinked first order ODEs discussed in our one-dimensional example. However, the theorem below implies that *not every* linear ODE system can be reduced to such a superposition (or direct sum) of one-dimensional ODEs. In particular, a single Jordan block of size $m > 1$ cannot be transformed into the superposition of one-dimensional ODEs.

`Theorem.` ***Every constant coefficient system of*** $n$ ***linear*** $1$***st order ODEs in*** $n$ ***unknowns can be transformed by a complex linear change of the unknowns into exactly one (up to reordering of the blocks) of the Jordan systems with*** $m_1 + ... + m_r = n$***.***

***EXERCISES***

***1.*** Transform explicitly one linear function $y = -3x$ to the normal form prescribed by the Rank Theorem.

***2.*** The same for the linear function $y = 3x_1 - 2x_2$.

***3.*** The same for the system: $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$.

***4.*** Prove that if two systems of $m$ linear functions in $n$ variables have the same rank, then they can be transformed into each other by linear changes of dependent and independent variables. ϟ

***5.*** Transform explicitly the quadratic forms $4x^2$ and $-9y^2$ to their normal forms prescribed by the Inertia Theorem.

***6.*** Transform the quadratic forms from the previous exercise into each other by a substitution $x = cy$ with possibly complex value of $c$.

***7.*** Classify quadratic forms $Q = ax^2$ in one variable with *complex* coefficients (i.e. $a \in \mathbb{C}$) up to complex linear changes: $x = cX$, $c \in \mathbb{C}, c \neq 0$. ✓

***8.***$^\star$ In the Inertia Theorem with $n = 2$, show that that there are six normal forms, and prove that they are pairwise non-equivalent. ϟ

***9.*** Find the indices of inertia of the quadratic form $Q(x, y) = xy$. ϟ

***10.*** Show that $X_1^2 + ... + X_n^2$ is the only one of the normal forms of the Inertia Theorem which is positive everywhere outside the origin.

***11.*** Sketch the surfaces $Q(X_1, X_2, X_3) = 0$ for all normal forms in the Inertia Theorem with $n = 3$.

***12.*** How many normal forms are there in the Inertia Theorem for quadratic forms in $n$ variables? ✓

***13.*** Transform explicitly the quadratic form $Q = 3x^2 + 16y^2 + 9z^2$ to its normal form prescribed by the Inertia theorem, and apply the same transformation to the quadratic form $S = x^2 - 4y^2 + 12yz$.

***14.*** Find the spectrum of the pair of quadratic forms: $Q = 3x^2 + 16y^2 + 9z^2$, $S = x^2 - 4y^2 + 12z^2$.

***15.*** Find the general solution to the differential equation $\dot{x} = \lambda x$. ✓

***16.*** Find the general solution to the system of ODE: $\dot{x} = 3x$, $\dot{y} = -y$, $\dot{z} = 0$. ✓

***17.*** Verify that $y(t) = e^{\lambda t}\left(c_0 + tc_1 + ... + c_{m-1}t^{m-1}\right)$, where $c_i \in \mathbb{C}$ are arbitrary constants, is the general solution to the ODE $(\frac{d}{dt} - \lambda)^m y = 0$.

***18.*** Rewrite the *pendulum* equation $\ddot{x} = -x$ as a system. ϟ

***19.***$^\star$ Identify the Jordan form of the system $\dot{x}_1 = x_2$, $\dot{x}_2 = -x_1$. ✓

***20.***$^\star$ Find the general solution to the system $\dot{x}_1 = x_2$, $\dot{x}_2 = 0$, and sketch the trajectories $(x_t(t), x_2(t))$ on the plane. Prove that the system cannot be transformed into any system $y_1 = \lambda_1 y_1$, $\dot{y}_2 = \lambda_2 y_2$ of two unlinked ODEs.

# 2 A Model Example: Quadratic Curves

We assume here some familiarity with vector geometry on the plane. The supplement "Vectors in Geometry" can be used as a reference.

## Conic Sections

On the coordinate plane, consider points $(x, y)$, satisfying an equation of the form

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

Generally speaking, such points form a curve. The set of all solutions to the equation is called a **quadratic curve**, provided that not all of the coefficients $a, b, c$ vanish.

Being a quadratic curve is a geometric property. Indeed, if the coordinate system is changed (say, rotated, stretched, or translated), the same curve will be described by a different equation, but the left-hand-side of the equation will remain a polynomial of degree 2.

Our goal in this section is to describe all possible quadratic curves geometrically (i.e. disregarding their positions with respect to coordinate systems); or, in other words, to *classify* quadratic equations in two variables up to suitable changes of the variables.

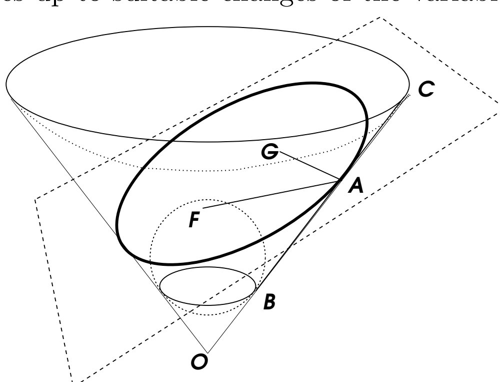



Figure 1. Dandelin's spheres

`Example:` **Dandelin's spheres.** The equation $x^2 + y^2 = z^2$ describes in a Cartesian coordinate system a cone (a half of which is shown on Figure 1). Intersecting the cone by planes, we obtain examples of quadratic curves. Indeed, substituting the equation $z = \alpha x + \beta y$ of a secting plane into the equation of the cone, we get a quadratic equation $x^2 + y^2 = (\alpha x + \beta y)^2$ (which actually describes the projection of the conic section to the horizontal plane).

The conic section on the picture is an **ellipse**. According to one of many equivalent definitions,[1] an ellipse consists of all points of the plane with a fixed sum of the distances to two given points (called **foci** of the ellipse). Our picture illustrates an elegant way[2] to locate the foci of a conic section.

Place into the conic cup two balls (a small and a large one), and inflate the former and deflate the latter until they touch the plane (one from inside, the other from outside). Then the points $F$ and $G$ of the tangency are the foci.

Indeed, let $A$ be an arbitrary point on the conic section. The segments $AF$ and $AG$ lie in the cutting plane and are therefore tangent to the balls at the points $F$ and $G$ respectively. On the generatrix $OA$, mark the points $B$ and $C$ where it crosses the circles of tangency of the cone with the balls. Then $AB$ and $AC$ are tangent at these points to the respective balls. All tangent segments from a given point to a given ball have the same length. Hence we find that $|AF| = |AB|$, and $|AG| = |AC|$. Therefore $|AF| + |AG| = |BC|$. But $|BC|$ is the distance along the generatrix between two parallel horizontal circles on the cone, and is the same for all generatrices. We conclude that the sum $|AF| + |AG|$ stays fixed when the point $A$ moves along our conic section.

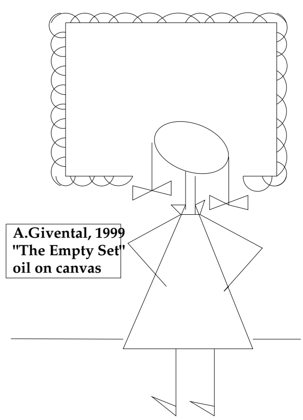



Figure 2. The Empty Set

Beside ellipses, we find among conic sections: **hyperbolas** (when a plane cuts through both halves of the cone), parabolas (cut by planes parallel to generatrices), and their degenerations (obtained

[1] According to a mock definition, "an ellipse is the circle inscribed into a square with unequal sides."

[2] Due to Germinal Pierre **Dandelin** (1794–1847).

when the cutting plane is replaced with the parallel one passing through the vertex $O$ of the cone): just one point $O$, pairs of intersecting lines, and "double-lines." We will see that this list exhausts all possible quadratic curves, except two degenerate cases: pairs of parallel lines and (yes!) empty curves.

The first step in obtaining this classification is a 2-dimensional version of the Orthogonal Diagonalization Theorem, which we prove here by bare hands.

## Orthogonal Diagonalization (Toy Version)

Let $(x, y)$ be Cartesian coordinates on a Euclidean plane, and let $Q$ be a **quadratic form** on the plane, i.e. a *homogeneous* degree-2 polynomial:

$$Q(x, y) = ax^2 + 2bxy + cy^2.$$

Theorem. ***Every quadratic form in a suitably rotated coordinate system assumes the form:***

$$Q = AX^2 + CY^2.$$

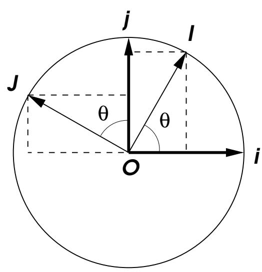



Figure 3. Rotation

Proof. Rotating the unit coordinate vectors $\mathbf{i}$ and $\mathbf{j}$ counterclockwise through the angle $\theta$ (Figure 3), we obtain the following expressions for the unit coordinate vectors $\mathbf{I}$ and $\mathbf{J}$ of the rotated coordinate system:

$$\mathbf{I} = (\mathbf{cos}\,\theta)\mathbf{i} + (\mathbf{sin}\,\theta)\mathbf{j} \quad \text{and} \quad \mathbf{J} = -(\mathbf{sin}\,\theta)\mathbf{i} + (\mathbf{cos}\,\theta)\mathbf{j}.$$

Next, we express the radius-vector of any point in both coordinate systems:

$$x\mathbf{i} + y\mathbf{j} = X\mathbf{I} + Y\mathbf{J} = (X\,\mathbf{cos}\,\theta - Y\,\mathbf{sin}\,\theta)\mathbf{i} + (X\,\mathbf{sin}\,\theta + Y\,\mathbf{cos}\,\theta).$$

This shows that the old coordinates $(x, y)$ are expressed in terms of the new coordinates $(X, Y)$ by the formulas

$$x = X\,\mathbf{cos}\,\theta - Y\,\mathbf{sin}\,\theta, \quad y = X\,\mathbf{sin}\,\theta + Y\,\mathbf{cos}\,\theta. \tag{$*$}$$

Substituting into $ax^2 + 2bxy + cy^2$, we rewrite the quadratic form in the new coordinates as $AX^2 + 2BXY + CY^2$, where $A, B, C$ are certain expressions of $a, b, c$ and $\theta$. We want to show that choosing the rotation angle $\theta$ appropriately, we can make $2B = 0$. Indeed, making the substitution explicitly and ignoring $X^2$- and $Y^2$-terms, we find $Q$ in the form

$$\cdots + XY\left(-2a\,\mathbf{sin}\,\theta\,\mathbf{cos}\,\theta + 2b(\mathbf{cos}^2\,\theta - \mathbf{sin}^2\,\theta) + 2c\,\mathbf{sin}\,\theta\,\mathbf{cos}\,\theta\right) + \dots$$

Thus $2B = (c-a)\,\mathbf{sin}\,2\theta + 2b\,\mathbf{cos}\,2\theta$. When $b = 0$, our task is trivial, as we can take $\theta = 0$. When $b \neq 0$, we can divide by $2b$ to obtain

$$\mathbf{cot}\,2\theta = \frac{a-c}{2b}.$$

Since **cot** assumes arbitrary real values, the theorem follows.

`Example.` For $Q = x^2 + xy + y^2$, we have $\mathbf{cot}\,2\theta = 0$, and find $2\theta = \pi/2 + \pi k$ ($k = 0, \pm 1, \pm 2, \dots$), i.e. up to multiples of $2\pi$, $\theta = \pm\pi/4$ or $\pm 3\pi/4$. (This is a general rule: together with a solution $\theta$, the angle $\theta + \pi$ as well as $\theta \pm \pi/2$, also work. Could you give an *a priori* explanation?) Taking $\theta = \pi/4$, we compute $x = (X - Y)/\sqrt{2}$, $y = (X + Y)/\sqrt{2}$, and finally find:

$$x^2 + y^2 + xy = X^2 + Y^2 + \frac{1}{2}(X^2 - Y^2) = \frac{3}{2}X^2 + \frac{1}{2}Y^2.$$

## Completing the Squares

In our study of quadratic curves, the plan is to simplify the equation of the curve as much as possible by changing the coordinate system. In doing so we may assume that the coordinate system has already been rotated to make the coefficient at $xy$-term vanish. Therefore the equation at hands assumes the form

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

where $a$ and $c$ cannot both be zero. Our next step is based on **completing squares**: whenever one of these coefficients (say, $a$) is non-zero, we can remove the corresponding linear term ($dx$) this way:

$$ax^2 + dx = a(x^2 + \frac{d}{a}x) = a\left((x + \frac{d}{2a})^2 - \frac{d^2}{4a^2}\right) = aX^2 - \frac{d^2}{4a}.$$

Here $X = x + d/2a$, and this change represents translation of the origin of the coordinate system from the point $(x, y) = (0, 0)$ to $(x, y) = (-d/2a, 0)$.

`Example.` The equation $x^2 + y^2 = 2ry$ can be rewritten by completing the square in $y$ as $x^2 + (y - r)^2 = r^2$. Therefore, it describes the circle of radius $r$ centered at the point $(0, r)$ on the $y$-axis.

With the operations of completing the squares in one or both variables, renaming the variables if necessary, and dividing the whole equation by a non-zero number (which does not change the quadratic curve), we are well-armed to obtain the classification.

## Classification of Quadratic Curves

***Case I:*** $a \neq 0 \neq c$. The equation is reduced to $aX^2 + cY^2 = F$ by completing squares in each of the variables.

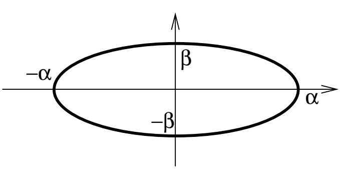
***Sub-case (i):*** $F \neq 0$. Dividing the whole equation by $F$, we obtain the equation $(a/F)X^2 + (c/F)Y^2 = 1$. When both $a/F$ and $c/F$ are positive, the equation can be re-written as

$$\frac{X^2}{\alpha^2} + \frac{Y^2}{\beta^2} = 1.$$

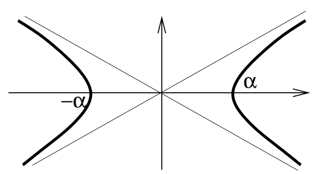
This is the equation of an ellipse with **semiaxes** $\alpha$ and $\beta$ (Figure 4). When one $a/F$ and $c/F$ have opposite signs, we get (possibly renaming the variables) the equation of a hyperbola (Figure 5)

$$\frac{X^2}{\alpha^2} - \frac{Y^2}{\beta^2} = 1.$$

When $a/F$ and $c/F$ are both negative, the equation has no real solutions, so that the quadratic curve is *empty* (Figure 2).




`Figure 4. Ellipse`




`Figure 5. Hyperbola`

***Sub-case (ii):*** $F = 0$. Then, when $a$ and $c$ have opposite signs (say, $a = \alpha^2 > 0$, and $c = -\gamma^2 < 0$), the equation $\alpha^2 X^2 = \gamma^2 Y^2$

describes a pair of intersecting lines $Y = \pm kX$, where $k = \alpha/\gamma$ (Figure 6). When $a$ and $c$ are of the same sign, the equation $aX^2 + cY^2 = 0$ has only one real solution: $(X, Y) = (0, 0)$. The quadratic curve is a "thick" point.[3]

***Case II: One of*** $a, c$ ***is*** $0$***.*** We may assume without loss of generality that $c = 0$. Since $a \neq 0$, we can still complete the square in $x$ to obtain an equation of the form $aX^2 + ey + F = 0$.

***Sub-case (i):*** $e \neq 0$***.*** Divide the whole equation by $e$ and put $Y = y - F/e$ to arrive at the equation $Y = -aX^2/e$. This curve is a **parabola** $Y = kX^2$, where $k = -a/e \neq 0$ (Figure 7).

***Sub-case (ii):*** $e = 0$***.*** The equation $X^2 = -F/a$ describes: a pair of parallel lines $X = \pm k$ (where $k = \sqrt{-F/a}$), or the empty set (when $F/a > 0$), or a "double-line" $X = 0$ (when $F = 0$).

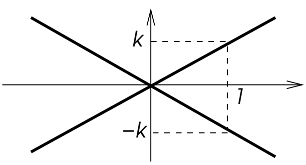


Figure 6. Crossing lines

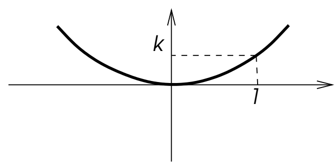


Figure 7. Parabola

We have proved the following:

Theorem. ***Every quadratic curve on a Euclidean plane is one of the following: an ellipse, hyperbola, parabola, a pair of intersecting, parallel, or coinciding lines, a "thick" point or the empty set. In a suitable Cartesian coordinate system, the curve is described by one of the standard equations:***

$$\frac{X^2}{\alpha^2} \pm \frac{Y^2}{\beta^2} = 1, -1, \ \textbf{\textit{or}}\ 0; \quad Y = kX^2; \quad X^2 = k.$$

***EXERCISES***

***21.*** Find the places of the following quadratic curves in our classification: $y = x^2 + x,\ xy = 1,\ xy = 0,\ xy = y,\ x^2 + x = y^2 - y,\ x^2 + x + y^2 - y = 0.$

***22.*** Following the steps of our classification, reduce the quadratic equation $x^2 + xy + y^2 + \sqrt{2}(x - y) = 0$ to one of the standard forms. Show that the curve is an ellipse, and find its semiaxes. ϟ ✓

[3] In fact this is the point of intersection of a pair of "imaginary" lines consisting of non-real solutions.

***23.*** Use our classification theorem to prove that, with the exception of parabolas, each conic section has a center of symmetry. $\lightning$

***24.*** Prove that a hyperbolic conic section consists of all points on the secting plane with a fixed *difference* of the distances to two points (called **foci**). Locate the foci by adjusting the construction of Dandelin's spheres.

***25.*** Locate foci of (a) ellipses and (b) hyperbolas given by the standard equations $x^2/\alpha^2 \pm y^2/\beta^2 = 1$, where $\alpha > \beta > 0$. $\checkmark$

***26.*** Show that "renaming coordinates" can be accomplished by a linear geometric transformation on the plane. $\lightning$

***27.*** A line is called an **axis of symmetry** of a given function $Q(x, y)$ if the function takes on the same values at every pair of points symmetric about this line. Prove that every quadratic form has two perpendicular axes of symmetry. (They are called **principal axes**.) $\lightning$

***28.*** Prove that if a line passing through the origin is an axis of symmetry of a quadratic form $Q = ax^2 + 2bxy + cy^2$, then the perpendicular line is also its axis of symmetry.$\lightning$

***29.*** Can a quadratic form on the plane have $> 2$ axes of symmetry? $\checkmark$

***30.*** Find axes of symmetry of the following quadratic forms $Q$:

$$(a)\ x^2 + xy + y^2, \quad (b)\ x^2 + 2xy + y^2, \quad (c)\ x^2 + 4xy + y^2.$$

Which of them have level curves $Q = const$ ellipses? hyperbolas? $\checkmark$

***31.*** Transform the equation $23x^2 + 72xy + 2y^2 = 25$ to one of the standard forms by rotating the coordinate system explicitly. $\lightning$ $\checkmark$

***32.*** Prove that ellipses are obtained by stretching (or shrinking) unit circles in two perpendicular directions with two different coefficients.

***33.*** Derive the Inertia Theorem for quadratic forms in two variables from the Orthogonal Diagonalization Theorem on the plane. That is, prove that every quadratic form on the plane in a suitable (but not necessarily Cartesian) coordinate system assumes one of the forms:

$$X^2 + Y^2,\ X^2 - Y^2,\ -X^2 - Y^2,\ X^2,\ -Y^2,\ 0.$$

Sketch graphs of these functions.

***34.*** Complete squares to find out which of the following curves are ellipses and which are hyperbolas: $\lightning$ $\checkmark$

$x^2 + 4xy = 1,\ x^2 + 2xy + 4y^2 = 1,\ x^2 + 4xy + 4y^2 = 1,\ x^2 + 6xy + 4y^2 = 1.$

***35.*** Show that a quadratic form $ax^2 + 2bxy + cy^2$ is, up to a sign $\pm$, the square $(\alpha x + \beta y)^2$ of a linear function if and only if $ac = b^2$. $\lightning$

***36.*** Show that if, in addition to rotation, reflection, and translation of coordinate systems, and multiplication of a quadratic equation by a non-zero constant, the change of scales of the coordinates is also allowed, then each quadratic equation can be transformed into one of the following 9 normal forms:

$$x^2 + y^2 = 1,\ x^2 + y^2 = 0,\ x^2 + y^2 = -1,\ x^2 - y^2 = 1,\ x^2 - y^2 = 0,$$
$$x^2 = y,\quad x^2 = 1,\quad x^2 = 0,\quad x^2 = -1.$$

***37.*** Examine the curves defined by the above equations to conclude that they fall into 8 different types.

***38.*** Find the place of the quadratic curve $x^2 - 4y^2 = 2x - 4y$ in the classification of quadratic curves. ✓

***39.***$^\star$ Prove that $ax^2 + 2bxy + cy^2 = 1$ is a hyperbola if and only if $ac < b^2$.

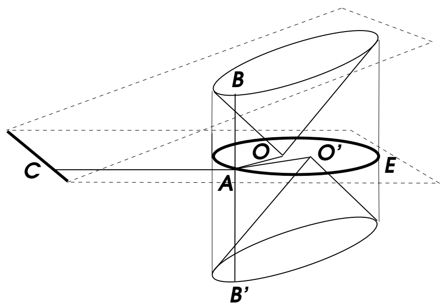



Figure 8. Conic sections

***40.*** Examine Figure 8 (showing two cones centrally symmetric to each other about the center of the ellipse $E$), and prove that $|AO| + |AO'| = |BB'|$. Derive that the vertex $O$ of the cone is a focus of the *projection* of a conic section along the axis of the cone to the perpendicular plane passing through its vertex.[4]

***41.***$^\star$ Prove that ellipses, parabolas and hyperbolas can be characterized as plane curves formed by all points with a fixed ratio $e$ (called **eccentricity**) between the distances to a fixed point (a **focus**) and a fixed line (called the **directrix**), and that $e > 1$ for ellipses, $e = 1$ for parabolas, and $1 > e > 0$ for ellipses (e.g. $e = |AO|/|AC|$ in Figure 8). ϟ

***42.***$^\star$ Prove that light rays emitted from one focus of an ellipse and reflected in it as in a mirror will focus at the other focus. Formulate and prove similar optical properties of hyperbolas and parabolas. ϟ

[4]The proof based on Figure 8 was invented by Anil **Hirani**, at that time a Caltech student.

# 3 Classifications in Mathematics

## What is classification?

Classifications are intended to bring order into seemingly complex or chaotic matters. Yet, there is a major difference between, say, our classification of quadratic curves and Carl **Linnaeus**' *Systema Naturae.*

For two quadratic curves to be in the same *class*, it is not enough that they share a number of features. What is required is a *transformation* of a prescribed type that would transform one of the curves into the other, and thus make them **equivalent** in this sense, i.e. *up to* such transformations.

What types of transformations are allowed (e.g., changes to *arbitrary* new coordinate systems, or only to *Cartesian* ones) may be a matter of choice. With every choice, the classification of objects of a certain kind (i.e. quadratic curves in our example) *up to* transformations of the selected type becomes a well-posed mathematical problem.

A complete answer to a classification problem should consist of
– a list of **normal** (or **canonical**) **forms**, i.e. representatives of the classes of equivalence, and
– a **classification theorem** establishing that each object of the kind (quadratic curve in our example) is equivalent to exactly one of the normal forms, i.e. in other words, that
(i) each object can be transformed into a normal form, and
(ii) no two normal forms can be transformed into each other.

Simply put, Linear Algebra deals with classifications of linear and/or quadratic equations, or systems of such equations. One might think that all that equations do is ask: *Solve us!* Unfortunately this attitude toward equations does not lead too far. It turns out that very few equations (and kinds of equations) can be explicitly *solved*, but all can be *studied* and many *classified.*

The idea is to replace a given "hard" (possibly unsolvable) equation with another one, the normal form, which should be chosen to be as "easy" as it is possible to find in the same equivalence class. Then the normal form should be studied (and hopefully "solved") thus providing information about the original "hard" equation.

What sort of information? Well, *any* sort that remains *invariant* under the equivalence transformations in question.

For example, in classification of quadratic curves up to changes of Cartesian coordinate systems, all equivalent ellipses are indistinguishable from each other*geometrically* (in particular, they have the same semiaxes) and differ only by the choice of a Cartesian coordinate system. However, if arbitrary rescaling of coordinates is also allowed, then all ellipses become indistinguishable from circles (but still different from hyperbolas, parabolas, etc.)

Whether a classification theorem really simplifies the matters, depends on the kind of objects in question, the chosen type of equivalence transformations, and the applications in mind. In practice, the problem often reduces to finding sufficiently simple normal forms and studying them in great detail.

### Fools and Wizards

The subject of linear algebra fits well into the general philosophy just outlined. In the opening section, we formulated four classification problems and respective answers. Together with a number of variations and applications, which will be presented later in due course, they form what is usually considered the main course of linear algebra.

Thus, in the rest of the book we will undertake a more systematic study of the four basic problems and prove the classification theorems stated here. However, the reader (not unlike a fairy-tale hero) should be prepared to meet the following three challenges of the next Chapter.

Firstly, linear algebra has developed an adequate language, based on the abstract notion of **vector space**. It allows one to represent relevant mathematical objects and results in ways much less cumbersome and thus more efficient than those found in the previous discussion. This language is introduced at the beginning of Chapter 2. The challenge here is to get accustomed to the abstract way of thinking.

Secondly, one will find there much more diverse material than what has been described in the Introduction. This is because many mathematical objects and classification problems about them can be *reduced* (speaking roughly or literally) to the four problems discussed above. The challenge is to learn how to recognize situations where results of linear algebra can be helpful. Many of those objects will be introduced in the middle section of Chapter 2.

Finally, we will encounter one more fundamental result of linear algebra, which is not a classification, but an important (and beautiful) formula. It answers the question: *Which substitutions of the form*

$$\begin{aligned} x_1 &= c_{11}X_1 + ... + c_{1n}X_n \\ &... \\ x_n &= c_{n1}X_1 + ... + c_{nn}X_n \end{aligned}$$

*are indeed changes of the variables and can therefore be inverted by expressing* $X_1, ..., X_n$ *linearly in terms of* $x_1, ..., x_n$*, and how to describe such inversion explicitly?* The answer is given in terms of the **determinant**, a remarkable function of $n^2$ variables $c_{11}, ..., c_{nn}$, which will also be studied in Chapter 2.

Let us describe now the principle by which our four main themes are grouped in Chapters 3 and 4.

Note that Jordan canonical forms and the normal forms in the Orthogonal Diagonalization Theorem do not form discrete lists, but instead depend on continuous parameters — the eigenvalues. Based on experience with many mathematical classifications, it is considered that the number of parameters on which equivalence classes in a given problem depend, is the right measure of complexity of the classification problem. Thus, Chapter 3 deals with **simple problems** of Linear Algebra, i.e. those classification problems where equivalence classes do not depend on continuous parameters. Respectively, the non-simple problems are studied in Chapter 4.

Finally, let us mention that the proverb: *Fools ask questions that wizards cannot answer,* fully applies in Linear Algebra. In addition to the four basic problems, there are many similarly looking questions that one can ask: for instance, to classify *triples* of quadratic forms in $n$ variables up to linear changes of the variables. In fact, in this problem, the number of parameters, on which equivalence classes depend, grows with $n$ at about the same rate as the number of parameters on which the three given quadratic forms depend. We will have a chance to touch upon such problems of Linear Algebra in the last section, in connection with *quivers.* The modern attitude toward such problems is that they are *unsolvable.*

***EXERCISES***

*43.* Using results of Section 2, derive the Inertia Theorem for $n = 2$.

*44.* Show that classification of real quadratic curves up to arbitrary linear inhomogeneous changes of coordinates consists of 8 equivalence classes. Show that if the coordinate systems are required to remain Cartesian, then there are infinitely many equivalence classes, which depend on 2 continuous parameters.

***45.*** Is there any difference between classification of quadratic equations in two variables $F(x,y)=0$ up to linear inhomogeneous changes of the variables and multiplications of the equations by non-zero constants, and the classification of quadratic curves, i.e. sets $\{(x,y)|F(x,y)=0\}$ of solutions to such equations, up to the same type of coordinate transformations? ✓

***46.***$^\star$ Derive the Orthogonal Diagonalization Theorem (as it is stated in Section 1) in the case $n=2$, using the “toy version” stated in Section 2, and *vice versa.* ϟ

***47.*** Let us represent a quadratic form $ax^2+2bxy+cy^2$ by the point $(a,b,c)$ in the 3-space. Show that the surface $ac=b^2$ is a cone. ϟ

***48.*** Locate the 6 normal forms $(x^2+y^2, x^2-y^2, -x^2-y^2, x^2, -y^2, 0)$ of the Inertia Theorem with respect to the cone $ac=b^2$ on Figure 9.

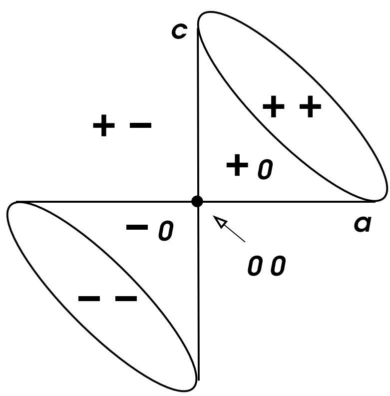



Figure 9. The Inertia Theorem

***49.*** The cone $ac=b^2$ divides the 3-space into three regions (Figure 9). Show that these three regions, together with the two branches of the cone itself, and the origin form the partition of the space into 6 parts which exactly correspond to the 6 equivalence classes of the Inertia Theorem in dimension 2.

***50.*** How many arbitrary coefficients are there in a quadratic form in $n$ variables? ✓

***51.***$^\star$ Show that equivalence classes of *triples* of quadratic forms in $n$ variables must depend on at least $n^2/2$ parameters. ϟ