CERTAIN ALGEBRAIC AND GEOMETRIC ASPECTS OF THE COMPLEX FIELD

By

MICHAEL CLYDE CLOCK

Bachelor of Science
Lewis and Clark College
Portland, Oregon
1957

Master of Science
Oregon State University
Corvallis, Oregon
1961

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF EDUCATION
July, 1967

CERTAIN ALGEBRAIC AND GEOMETRIC

ASPECTS OF THE COMPLEX FIELD

Thesis Approved:

Thesis Adviser

Dean of the Graduate College

ACKNOWLEDGEMENTS

In reflecting on the past two years, during which time this dissertation was conceived and developed and much of the allied course of study completed, I find myself deeply indebted to many people. It is not possible to list all those who have contributed directly or indirectly. However, any such reference would have to include Dr. Gerald Goff who has provided continuing counsel and assistance in the preparation of this paper. In addition, the constructive criticism offered by Dr. L. Wayne Johnson and Dr. Vernon Troxel, as members of my advisory committee, has been much appriciated. Furthermore, in retrospect, I view the disciplined approach to the study of mathematics suggested by Dr. Jeanne Agnew as a significant step in the direction of personal mathematical maturity.

Finally, no list of acknowledgements would be complete without expressing gratitude to my wife, Carol, and our four children, Tom, Kathy, Mary, and Paul, who sacrificed much, and without whose encouragement and cooperation this task could not have been completed.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Background

The past 15 years has been a period of ferment in mathematics. In particular, there has been a vast amount of time and energy expended in an attempt to determine the appropriate scope and sequence of mathematics instruction from the grammar school level to the graduate level.

Questions have arisen as to the reasons for instigating some rather radical changes in a curriculum, which has remained relatively static for a long period of time. It appears that there are at least two major factors that have influenced those concerned. One has been the extraordinary growth of pure mathematics in recent times. The other is the increasing dependence of scientific thought upon mathematical methods, coupled with an urgent demand for the services of scientists in almost every phase of endeavor. Thus, regardless of profession, the contention is that mathematics will profoundly influence the life of modern man.

Unfortunately, scrutiny of the mathematics curriculum of a decade ago indicates that most of the mathematics presented to students up to 19 or 20 years of age was at least 200 years old. Paradoxically, it has been conservatively estimated that more mathematics has been discovered in the last 100 years than in all of the previous history of mankind. A major contention then is that the educated man, whom we envision as the end product of our educational process, should not be left 200 years behind the times in mathematics.

One outgrowth of the almost universal concern for the direction of the mathematics curriculum has been the establishment of several national and international committees. Many of these groups have convened for the express purpose of determining the content and tenor of mathematics instruction in the immediate future.

A few of these groups, most notably perhaps the School Mathematics Study Group, have established writing committees in an effort to produce instructional materials commensurate with their recommendations. In the main, however, these groups have been content to make suggestions regarding the appropriate general content and sequence of the mathematics curriculum. The result has been that many of the topics recommended for inclusion, particularly at the secondary level, are not readily available to those who teach on this level. The complex number system is included among these. In order to implement the curriculum suggested, it appears both desirable and expedient to produce self contained papers that might be used by the instructor and students to gain the required insight into those areas where there is a deficiency of available materials. The production of such materials seems best-fitted to those with a backlog of teaching experience on the secondary level, considerable mathematical maturity, and time. These three ingredients appear necessary in order to insure that the most significant aspects of the material will be presented in a consistent, rigorous, and teachable manner. It was with these thoughts in mind that this work was undertaken.

Objectives

The paper focuses on certain algebraic and geometric aspects of complex numbers that might be presented to an audience having a foundation in elementary algebra, coordinate geometry, the real number system, trigonometry, and elementary functions, with a degree of rigor and completeness. Specifically, the presentation is accomplished without recourse to the limit concept, the sole exception being the fundamental theorem of algebra. The work is self-contained to the extent that results used, which are not generally encountered in the aforementioned five areas, are either stated without proof or demonstrated. In general the results stated without proof are readily available in standard texts on modern algebra or complex variables.

Although several classical results are demonstrated, or illustrated in some detail, the intent was to direct attention to those aspects of complex numbers that are not currently treated on either the high school or undergraduate level. Little of what is included can be termed truly original, although a review of the literature seems to suggest that the setting in which many of the results appear is somewhat unique.

Scope and Sequence

The initial portion of the paper is concerned with the development of complex numbers as an algebraic system. In addition to a detailed presentation of the complex number system as a two dimensional extension of the real numbers, attention is given to the allied question of the existence of a 3, 4, 5, ..., n dimensional extension of the real numbers. The discussion points up the unique algebraic position of the complex number system as a field extension of the reals, while providing a natural setting for an acquaintance with some significant algebraic structures that fail to possess all the characteristics of a field. The progressive nature of the theorems in Chapter IV was deemed desirable from the standpoint of the audience prescribed and the relative sophistication of the terminal results. This is in keeping with the overall tone of the presentation.

The second major aspect of the work deals with a mathematical model of the complex number system, namely the isometries of the Euclidean Plane. The focus here is on the algebraic development of these transformations, although the impetus is clearly geometric. Throughout this portion of the paper the associated geometry is used to motivate, illustrate and clarify the basic propositions.

Review of the Literature

A broad survey of the literature was made initially in an attempt to determine those aspects of complex variables that might profitably be discussed within the limitations of the paper. After delimiting the scope of the paper, an intensive review of the literature pertaining to the selected areas was undertaken. The Mathematical Review, indices to books in print, the card catalog, indices of The American Mathematical Monthly, and bibliographies of texts served as primary tools. In general, there was a dearth of reference material relating directly to this work, although some portion of the literature was suggestive of almost everything undertaken.

CHAPTER II

A HISTORICAL OVERVIEW

The purpose of this chapter is to give the reader some insight into the etiology of the complex number concept. There are three principal reasons for including such a discussion. First, it was felt that such an initial chapter would provide a framework to which the reader could relate all subsequent aspects of the work. Secondly, in view of the rather formal nature of the work in Chapters III through VI it seemed desirable to give the prescribed audience some insight into the rather erratic and informal historical evolution of the number concept. Finally, for the sake of completeness, material has been included which alludes to the physical applications of complex numbers. In the authors eyes such an inclusion has the additional advantage of giving credence to complex numbers, where the reader is reluctant to accept them on a purely mathematical basis.

It is the author's contention that most beginning students fail to see the human element in the development of mathematics. Too often they envision mathematics as having evolved in the same continuous deductive fashion in which it appears in their texts. It is hoped that a brief exposure to the history of the complex number concept will, among other things, reveal the fallacy of such a notion.

The early history of complex numbers is strikingly similar to that of the negative reals, a record of blind manipulations unrelieved by any serious attempt at interpretation. The first recorded evidence of recognition of imaginaries is that of Mahavira, the Indian mathematician of the ninth century. He was content to observe that "in the nature of things a negative number has no square root." [3;175] The next intrusion of imaginaries came in the sixteenth century with the work of the Italian mathematicians; specifically Cardan and Bombelli. Cardan in his quest for a solution to the reduced cubic was the first to symbolize the imaginaries, although he apparently rejected them as numbers. The crux of the matter was that Cardan's formula for the reduced cubic gave a quite satisfactory result for the real root of a cubic having, as we know it today, two complex roots. However, in the case where all three roots were real the formula gave illusory results for one of the real values. Cardan, and later Bombelli, were bold (or foolish, according to the readers whim) enough to attempt to manipulate these conjured, and admittedly fictitious symbols, in an effort to achieve a complete solution.

Consider, for example, the equation $x^3 = 15x + 4$ treated by Bombelli in his algebra published in 1572. [10] The equation has three real solutions $-2 + \sqrt{3}$, $-2 - \sqrt{2}$, and 4, yet application of the Cardan formula lead to the mystic expression $\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$, in place of the rational value 4. It occurred to Bombelli that the two radicals might represent expressions of the form $a + \sqrt{-b}$, $a - \sqrt{-b}$ where a and b are positive, in which case the sum would be independent of the imaginary symbol $\sqrt{-b}$. With some effort, and no small amount of misgiving over the undertaking, he was able to show that the two radicals did indeed resolve into $2 + \sqrt{-1}$ and $2 - \sqrt{-1}$, the sum of which is 4. Encouraged by his initial success, Bombelli proceeded to develop rules for operations on these mystic beings. Apart from notation the gifted Italian had all the rules in essentially their current form. It is not surprising that Bombelli's operational rules would parallel current definitions when one realizes that it was a widely held belief in his time that algebraic consistency was dependent on obediance to the manipulative principles for positive numbers.

Thus, we see a mathematician, emminent in his own time, devising rules for manipulating meaningless, though not altogether useless, symbols. The work of Bombelli marks the beginning of an era of blind formalism in connection the complex number symbol, a period that lasted approximately two hundred years. Listed among those who followed Bombelli in this mysterious play on symbols are some of the great mathematicians of the seventeenth and eighteenth centuries.

There are a couple of observations worth bringing to the fore in connection with the development of complex numbers to this point. First, it is interesting to note that at the time Bombelli was taking the initial steps in the area of complex numbers the real number system was entirely without foundation as we know it today. [10] As a matter of fact, negatives were not fully understood nor widely accepted in his day! Secondly, we note that in contrast to the logical current practice of introducing complex numbers following a discussion of quadratic equations, they initially came to the fore during an attempt to solve the cubic. The foregoing provide graphic illustration of the fact that mathematics in the making often bears little resemblance to the systematic exposition of the textbook.

The imaginary beings of Bombelli found little acceptance, had no real foundation, nor were they given any interpretation for over two hundred years, yet it is interesting to note that formalism alone produced some results of considerable significance. [3] About 1710 an Englishman, Cotes, discovered what later was recognized as the equivalent of Eulers famous relationship between e, i, π, and 1, namely that $i\theta = \log_e(\cos\theta + i\sin\theta)$. The second result of this period was DeMoivres' discovery of the trigonometric identity which bears his name, namely that $\cos n\theta + i\sin n\theta = (\cos\theta + i\sin\theta)^n$, n a natural number. This relationship gave the mystic numbers a new air of permanency by linking them to trigonometry. The prolific Euler introduced the transcendental e and extended the result of DeMoivre to arbitrary integral values for n. The famous special case of the foregoing result which bears Eulers name being $e^{i\pi} + 1 = 0$. Even today one can but marvel at this simple identity that involves some of the most important symbols of mathematics, each of historically disparate origin.

In addition to the preceeding developments it was reasonable to inquire as to whether the system created by the adjunction of complex numbers was adequate for the solution of the fundamental problem of algebra: determining the root of the most general polynomial equation.

In view of the Cardan formula and its predecessor the quadratic formula, it was evident in Bombelli's time that the complex numbers provided a complete solution for polynomial equations of degree three or less having real coefficients. The Ferrari method for solving the quartic, developed contemporaneously with Cardans' result, allowed extension of the above conclusion to degree four. The quest for a sharper result in this connection was centered around a necessarily futile attempt to derive expressions for the roots of higher ordered equations in terms of the coefficients and the basic arithmetic operations. Of course, the impossibility of producing such formulas was not established until the ingenious, but ill-fated, Galios provided the answer in 1830. By this time Gauss had already published (1799) his proof of the now classical Fundamental Theorem of Algebra. The combined results of Gauss and Galois answered with finality the age old questions of existence and radical solvability. That the foregoing questions were raised and completely answered prior to the acceptance of the complex number system is additional testimony to the logical irregularities in the development of mathematics. The following comment due to Euler (1770), though somewhat predating the works mentioned, was apparently characteristic of this period, and surely serves to dramatically illustrate the status of complex numbers at that time.

> All such expressions as $\sqrt{-1}$, $\sqrt{-2}$, etc., are consequently impossible or imaginary numbers, since they represent roots of negative numbers; and of such numbers we may truly assert that they are neither nothing, nor greater than nothing, nor less than nothing, which necessarily constitutes them imaginary or impossible. [10;191]

Certainly, the etymology of the word imaginary as applied to roots of negative reals needs no further clarification!

Part of the difficulty in accepting complex numbers stemmed from the fact that no one had been successful in giving a consistent, useful interpretation of them prior to Gauss' time. It is true that both the Norwegian, Wessel (1797), and the Swiss, Argand (1806), preceded Gauss in giving the now familiar vector interpretation. [12] Unfortunately, their results were not widely recognized, and it remained for Gauss to rediscover and present the essence of their works. Interestingly enough, just at the time when the long sought interpretation was achieved, the mathematical world arrived at a level of sophistication which deemed inacceptable a geometric foundation for a number system. In response to this, Gauss gave the first recorded formal treatment of complex numbers as ordered pairs in 1831. At last, man had given the mystical numbers of Bombelli both a postulational foundation and an intuitively appealing interpretation.

To one unfamiliar with the pattern of mathematical history it might appear that 1831 marks the terminus of one phase of endeavor. Quite the contrary, our vantage point reveals that this merely signaled the end of the beginning. The immediate stimulus for extensions of the number concept was the geometric description of the rotations of the plane afforded by Gauss' interpretation of complex numbers. The response was almost immediate.

The Irishman, Hamilton, reasoned that it should be possible to generate a number system that could be used to describe rotations in the space of three dimensions. The hurdle that blocked Hamilton's path in his initial attempts to achieve the desired algebraic description was that any such system would lack the commutative property. It must be recognized that in Hamilton's time the opinion was still widely held that one could avoid contradictory results only by adhering to the properties inherent in the rational numbers. Hamilton ultimately had the conviction to proceed in the endeavor, and in 1843 he presented his quaternion algebra. [16] Hamilton subsequently devoted the greater portion of life in a vain attempt to convince physicists and geometers that his quaternions held the key to major advances in their disciplines. Although they never received the attention that their inventor imagined, they do find some current application in both areas.

In the long run, the permanent residue of Hamilton's labor seems to be that he demonstrated a self consistent algebra in which the commutative law fails to hold. In so doing he paved the way for a host of algebras, in which one after another of the principles of rational arithmetic were discarded or modified. It is interesting to note that the appearance of an abstract approach to algebra coincides historically with the freeing of geometry from the bondage of Euclid's fifth postulate. Thus, almost simultaneously, geometers and algebraists perceived that mathematical systems are not supernaturally imposed on human beings, rather they are creations of the mind. In retrospect, it seems surprising that such a notion was so long in coming to the fore.

In addition to Hamilton, history reveals another very fertile mind at work in the area of hypercomplex numbers during this period. [3] The German mathematician, Grassman, essentially considered the much more general problem of defining a product on ordered n-tuples in such a way that it satisfied certain predetermined properties. The unfortunate aspect of Grassmans work was that his notation and style of writing were so unusual that his work found little acceptance in his own time. The scope of his theory was not fully appreciated until the current century when it was revealed that his work not only included complex numbers and quaternions as a potential detail, but aspects of matrices and tensor calculus as well. Thus, history shows that an unfortunate method of presentation obscured a work which might have advanced this area of algebra some fifty years.

After Hamilton's epochal revelation, the development of hypercomplex numbers, or linear algebras as they are often called, follows in three principal phases. [3] The first phase was represented by such work as that of the American, B. Pierce, who was very active in the

1870's. His efforts were directed toward the problem of exhibiting all linear associative algebras of a given finite dimension, having real or complex coordinates. The second phase is exemplified in the works of the German mathematician, Frobenius, who established a general result that described the nature of the totality of linear associative algebras of finite dimension over the real field. In addition, his work suggested the extension of the discussion of hypercomplex number systems to n-tuples whose coordinates were from fields other than the real or complex. The third phase is characterized by the work of the Scotchman, J.H.M. Wedderburn. In the early 1900's he established a series of theorems that in essence exposed the fundamental structure of a linear associative algebra of arbitrary dimension over any field. Subsequent efforts in the area of hypercomplex numbers have been directed toward the discovery of the analog of Wedderburns results for non-associative systems. One can but speculate that if this quest is successful mathematical desire for generality will culminate in a search for a theory linking the associative and non-associative algebras.

At this point a principal historical sequence, which originated with Cardan's mystical symbols, has been traced to current research in the area of linear algebra. In addition to being instrumental in the achievement of previously stated objectives, the development provides an example that serves to illustrate the usual path to abstractness and generality in mathematics.

The final portion of this chapter is devoted to a brief consideration of the applications of complex numbers. It must be remarked in passing that the study of complex numbers, or more generally complex function theory, requires no further justification for the pure mathematician then the inherent beauty of the structure. Nevertheless, it seems quite satisfying to see that a branch of mathematics, as unmotivated by physical observation and experience as complex numbers, does find application in physics. Hopefully, the reader recognizes that any relationship that finds expression in terms of complex variables can be formulated solely in terms of reals. The fact is that the complex representation often provides a much more elegant and penetrating formulation in physics.

The areas of physics that have proven most amenable to complexification are quantum mechanics, electricity, and optics. [15][31] In the main, the applications of complex representation occur at fairly high level sophistication in these disciplines, thus making discussion of them difficult here. A single example from the field of optics was selected because of its availability to the reader and its striking illustration of the extent to which complex numbers find an interpretive reality.

In elementary physics, Snell's Law asserts that if light passes from one transparent media to another then the ratio $\frac{\sin i}{\sin r}$ is constant. [30] In this expression i and r represent the magnitudes of the angles between the direction of propagation of the incident and refracted rays and the respective directed normals. In keeping with both experimental results and Huygen's wave model of light the constant value of this ratio turns out to be the ratio of the velocity of light in the incident media (v_i) to that of light in the refracting media (v_r). Thus, if the velocities are known and the angle of incidence is given, the angle of refraction can be determined.

In 1823 the French physicist, Fresnel, took an additional step toward the completion of a comprehensive theory of light. Beginning with a limited number of propositions about the behavior of light, he was able to show that for light polarized in the plane of reflection the ratio of amplitudes of reflected and incident light is $-\frac{\sin(i - r)}{\sin(i + r)}$. [33] This relationship had been previously suggested by empirical evidence.

In connection with the foregoing result, Fresnel recognized that if $\frac{v_i}{v_r} = k < 1$, then there exists a value x, $0 < x < \pi/2$, such that for $i > x$, $\sin i > k$. If the sine function is restricted to the real field, then in this case there is no corresponding solution for r. If, however, one considers the extended sine function, then the equation $\sin r = \frac{\sin i}{k}$ has a solution for all real values of i. [27] The multi-valued complex solution for r when $\sin i > k$ has no recognizable physical interpretation as an angle, however, Fresnel considered these solutions in connection with his result relating the amplitudes of reflected and incident light. In particular, he observed that for these values of r the ratio $\frac{\sin(i - r)}{\sin(i + r)} = e^{i\theta}$, θ real. Fresnel conjectured that this indicated that total reflection occurs, and that the incident and reflected waves have the same amplitude, but differ in phase by an amount θ. These statements were subsequently completely confirmed by experimentation! [5] The fascinating aspect of Fresnel's work here is that one sees complex numbers playing a role in physics that supercedes that of merely providing an elegant symbolic formulation of an already conceived theory. In particular, one sees laws of nature being abstracted from a branch of mathematics that is not at all an obvious abstraction from the physical world. Contemporary physics reveals that Fresnel's work merely set the stage for more extensive exploitation of complex variables in physics.

CHAPTER III

THE COMPLEX FIELD

The focus of the current chapter is on the development of the complex field and the basic properties of the system that are either necessary in the sequel or desirable for completeness. The initial stimulus for the development is the desire for an algebraic solution to polynomial equations over the real field. Specifically, attention is directed to the problem of enlarging the real number system in such a way that the equation $x^2 = -1$ will have a solution.

Some of the results of this chapter are readily available elsewhere and the proofs of these were omitted where it was felt that such a demonstration would contribute little toward the achievement of the objectives of the paper. These theorems are recognized by the fact that an appropriate letter of the alphabet follows the identification number. Throughout the paper references to all definitions, theorems, and corollaries are indicated by the corresponding number preceded by D,T and C respectively.

In order to implement the development certain preliminary notions are introduced.

Definition 3.1. A non-empty set F on which two binary operations + and $\circ$ are defined is a field if and only if the following conditions hold:

(i) F is closed with respect to + and $\circ$;

(ii) $+$ and $\circ$ are commutative;

(iii) $+$ and $\circ$ are associative;

(iv) There exist distinct elements $0, 1$ in F such that $x + 0 = x$, $x \circ 1 = x$, for every $x \in F$;

(v) For each $x \in F$ there exists $-x \in F$ such that $x + (-x) = 0$;

(vi) For each $x \in F$, $x \neq 0$, there exists $x^{-1} \in F$ such that $x \circ x^{-1} = 1$;

(vii) $\circ$ distributes over $+$.

In connection with the above definition the operations $+$ and $\circ$ will be referred to as addition and multiplication respectively. Further, for $y \neq 0$, the symbol $\frac{x}{y}$ is defined to mean $x \circ y^{-1}$. In like manner $x - y$ means $x + (-y)$. Finally, when the operations of two different fields are used in a single setting, it is assumed that the context will suffice to clarify the meaning.

In general, it is presumed that the reader is familiar with the basic properties of a field, having encountered them in the development of the rational number system. The following examples of fields will be referred to on occasion and the associated symbols will be used to expedite this.

Example 3.1. The set of rational numbers with ordinary addition and multiplication forms a field denoted Q.

Example 3.2. The set of integers mod p, p prime, with the usual modular sum and product will be designated I_p.

Example 3.3. The set of expressions of the form $a + b\sqrt{2}$, where $a, b \in Q$, with $(a + b\sqrt{2}) + (c + d\sqrt{2}) = ([a + c] + [b + d]\sqrt{2})$ and $(a + b\sqrt{2})(c + d\sqrt{2})$

$= ([ac + 2bd] + [ad + bc]\sqrt{2})$, forms a field denoted $Q(\sqrt{2})$.

The following concept plays a significant role in subsequent work with fields.

Definition 3.2. If K is a subset of a field F, then K is a subfield of F if and only if K is a field with respect to the operations on F.

Example 3.4. The set of all elements of $Q(\sqrt{2})$ of the form $a + 0\sqrt{2}$ forms a subfield of $Q(\sqrt{2})$.

The following characterization of a subfield will prove useful in practice.

Theorem 3.1. In order that K, a non-empty subset of a field F, be a subfield it is necessary and sufficient that:

(i) $-1 \in K$;

(ii) Whenever $x, y \in K$, then $x + y$, $x \cdot y \in K$ and, provided $x \neq 0$, $x^{-1} \in K$.

Proof. The necessity is almost immediate from the fact that K is a field.

The sufficiency requires showing that K possesses the properties (i) - (vii) of D.3.1. That K is closed with respect to + and $\cdot$ follows from T 3.1.(ii). The commutativity and associativity of + and $\cdot$ is apparent since $K \subseteq F$. Now $-1 \in K$, hence $(-1)(-1) \in K$, or $1 \in K$. Also $1 + (-1) \in K$, or $0 \in K$; thus, D 3.1.(iv) is satisfied. If $x \in K$, then $(-1)x \in K$, or $-x \in K$ and D 3.1.(v) follows. Similarly, if $x \in K$, $x \neq 0$, then $x^{-1} \in K$. The distributivity of $\cdot$ over + in K is again a result of the fact that $K \subseteq F$.

Definition 3.3. A field F is said to be totally ordered if and only if there exists a non-empty subset of F, denoted P, such that:

(i) For every $x \in F$ exactly one of the following holds: $x = 0$ or $x \in P$ or $-x \in P$;

(ii) If $x, y \in P$, then $x + y$, $x \cdot y \in P$.

The set P is called a set of positive elements of F.

The next result sheds some light on the structure of a totally ordered field.

Theorem 3.2. If F is a totally ordered field, then there exists a relation $\leq$ on F such that for every $x, y, z \in F$ the following hold:

(i) $x \leq x$;

(ii) $x \leq y$ and $y \leq x$ implies $x = y$;

(iii) $x \leq y$ and $y \leq z$ implies $x \leq z$;

(iv) $x \leq y$ or $y \leq x$;

(v) $x \leq y$ implies $x + z \leq y + z$;

(vi) $x \leq y$ and $z \geq 0$ implies $x \cdot z \leq y \cdot z$.

Proof. Define $x \leq y$ if and only if $x = y$ or $y - x \in P$. In view of the similarity of technique used in showing (i) - (vi) only the demonstration of (iii) is presented here. If $x \leq y$ and $y \leq z$, then $x = y$ or $y - x \in P$ and $y = z$ or $z - y \in P$. If $x = y$, then $x = y = z$ or $z - x = z - y \in P$, hence in either case $x \leq z$. If $y - x \in P$, then $z - x = y - x \in P$ or $z - x = (y - x) + (z - y) \in P$ and $x \leq z$. Therefore, in any case, $x \leq z$.

In the sequel the relation $\leq$ is assumed to be defined as in the foregoing argument unless otherwise indicated.

The converse of T 3.2 is also valid thus providing a characterization of total ordering.

Theorem 3.3. If F is a field and $\leq$ is a relation on F such that for $x,y \in F$, $x \leq y$ if and only if $x < y$ or $x = y$ and $\leq$ satisfies (i) - (vi) of T 3.2, then $P = \{x \in F \mid x > 0\}$ is a set of positive elements for a total ordering of F.

Proof. Clearly $P \subseteq F$. Since $0,1,-1 \in F$ and $1 \neq 0$, then (iv) implies that $1 < 0$ or $0 < 1$. In case $1 > 0$, then $1 \in P$ and $P \neq \emptyset$. If $1 < 0$ then $-1 + 1 < -1 + 0$, or $0 < -1$, hence $-1 \in P$. Thus, in any event, $P \neq \emptyset$.

Let $x \in F$, $x \neq 0$, then as in the case of 1 either $x \in P$ or $-x \in P$. Suppose both $x \in P$ and $-x \in P$, then $x > 0$ and $-x > 0$, whence $x + (-x) > x + 0$, or $0 > x$. Thus, $x > 0$ and $0 > x$ and (ii) implies $x = 0$, which is a contradiction. Therefore, for every $x \in F$, exactly one of the following holds: $x = 0$, $x \in P$, $-x \in P$.

Now if $x,y \in P$, then $x > 0$ and $y > 0$, hence $x + y \geq x + 0 = x$. Utilizing (iii) $x + y \geq 0$. Suppose $x + y = 0$, then $y = -x$ and $-x \in P$. But $x \in P$ and $-x \in P$ is in contradiction to the result of the preceding paragraph. Therefore, $x + y > 0$, or $x + y \in P$. Finally, for $x > 0$, $y > 0$ (vi) implies that $x \cdot y \geq x \cdot 0 = 0$. However, $x \cdot y \neq 0$, since x and y are nonzero elements of a field. Thus, $x \cdot y > 0$, or $x \cdot y \in P$, which completes the proof.

Example 3.5. The field of rational numbers with the usual ordering is a totally ordered field.

Definition 3.4. If G is a nonempty subset of an ordered field and there exists an $x \in F$ such that $x \geq y$ for every $y \in G$, then x is said to be an upper bound for G. If G is bounded above and there exists an upper bound $z \in F$ such that $z \leq y$ for every upper bound y, then z is said to be the least upper bound for G.

Definition 3.5. If F is an ordered field and every nonempty subset of F that is bounded above has a least upper bound, then F is said to be a complete ordered field.

Definition 3.6. If F and F' are fields and $f:F \to F'$ is a one to one mapping of F into F' such that for every $x,y \in F$:

(i) $f(x + y) = f(x) + f(y)$;

(ii) $f(x \cdot y) = f(x) \cdot f(y)$;

Then f is called an isomorphism of F into F' and F and f(F) are said to be isomorphic. In case $F = f(F)$, then f is called an automorphism.

The following result establishes the nature of the range of a field isomorphism. The proof furnishes an application of T 3.1.

Theorem 3.4. If F,F' are fields and $f:F \to F'$ is a field isomorphism of F into F' then f(F) is a subfield of F'.

Proof. Clearly $f(F) \subseteq F'$ and $f(F) \neq \emptyset$. If $f(x),f(y) \in f(F)$, then $x,y \in F$, hence $f(x + y) \in f(F)$. However, $f(x + y) = f(x) + f(y)$, thus $f(x) + f(y) \in f(F)$. Similarly, $f(x) \cdot f(y) \in f(F)$. Now, there exists $0 \in F$ and for every $x \in F$, $f(x) = f(x + 0) = f(x) + f(0) = f(0) + f(x)$. Specifically then, $f(x) = f(x) + f(0)$. Moreover, since F' is a field $-f(x) \in F'$ and from $-f(x) + f(x) = -f(x) + [f(x) + f(0)]$, it follows that $f(0) = 0$. Also note that $f(1) = f(1 \cdot 1) = f(1) \cdot f(1)$. However, since F' is a field and $f(1) \neq 0$, then $[f(1)]^{-1} \in F'$; hence the preceding equality can be used to show that $f(1) = 1$. Now,

$0 = f(0) = f(1+-1) = f(1) + f(-1)$. Consequently, $f(-1) = -f(1) = -1$, or $-1 \in f(F)$. Finally, if $f(x) \in f(F)$, $f(x) \neq 0$, then $x \in F$ and $x \neq 0$. Hence $x^{-1} \in F$. Therefore, $1 = f(1) = f(x \cdot x^{-1}) = f(x) \cdot f(x^{-1})$, or since F' is a field and $f(x) \neq 0$, then the result that $f(x^{-1}) = [f(x)]^{-1}$ implies $[f(x)]^{-1} \in f(F)$. The hypothesis of T 3.1. being satisfied, it follows that f(F) is a subfield of F'.

In view of the definition and T 3.4. it is reasonable to interpret a field isomorphism as a one to one correspondence between two fields that preserves the operations. Even more loosely speaking, two fields are isomorphic if they differ only in notation. The following will serve to clarify this important concept.

Example 3.6. The mapping $f: Q(\sqrt{2}) \rightarrow Q(\sqrt{2})$ such that $f(a + b\sqrt{2}) = a - b\sqrt{2}$ is an automorphism of $Q(\sqrt{2})$.

Example 3.7. If F is any field, the identity map $g: F \rightarrow F$ is a field automorphism of F.

Example 3.8. The function $h: Q \rightarrow Q(\sqrt{2})$ such that $h(a) = a + 0 \cdot \sqrt{2}$ is an isomorphism of Q into $Q(\sqrt{2})$.

Having introduced the necessary preliminary concepts, the following definition is formulated.

Definition 3.7. A field F is called a field of real numbers if and only if F is a complete ordered field.

It is of note that the foregoing definition admits the possibility of more than one example of a real number system. Indeed this is the case. The reader may find this a little perplexing since it is common practice to use the term in the singular. As one might anticipate, it can be shown that, within an isomorphism, all complete ordered fields are identical. [4] In view of this, the convention of referring to a particular model of a complete ordered field as the real number system seems appropriate. Thus, in the sequel the symbol R will be used to denote the familiar set of real numbers. In addition, R with its usual structure will be referred to as the system of real numbers.

Although the properties of the real number system are assumed in this development, it is of interest to note that without recourse to sophisticated techniques one can give credence to the foregoing definition. Observe that the familiar fields Q and I_p, which are obviously not isomorphic to R due to differences in cardinality, also fail to satisfy the conditions for a complete ordered field. Specifically, the rational number system is an ordered field yet fails to be complete in that such sets as $\{x \in Q \mid x^2 < 2\}$ are bounded above but have no least upper bound in Q. On the other hand I_p does not possess a total ordering. To see this suppose H_p is a nonempty set of positive elements of I_p and let $x \in H_p$, then the sum of p x's, denoted px, must again be an element of H_p. It suffices to note that $px = 0$, mod p, and that $0 \notin H_p$.

A currently popular pedagogical device for motivating the various extensions of the number concept, when one develops the real number properties from those of the natural numbers, is to allude to the insolvability of certain simple polynomial equations in a given system. In keeping with this approach, consider the problem of extending the concept of number so as to obtain a field in which the quadratic $x^2 = -1$ has a solution. It is clear that in developing such a system it would be desirable to do so in such a way that the additional algebraic strength afforded by the real number system not be sacrificed. In particular, consider the necessary properties of a field containing a subfield isomorphic to R and an element e such that $e^2 = -1$.

Theorem 3.5. If F is a field that contains a subfield R' that is isomorphic to R, and there exists $e \in F$ such that $e^2 = -1$, then:

(i) $D = \{x + ey \mid x,y \in R'\}$ is a subfield of F;

(ii) R' is a subfield of D.

Proof. (i) T 3.1. is applied. Clearly D is a nonempty subset of F. Let $x_1 + e\cdot y_1$, $x_2 + e\cdot y_2 \in D$, then

(1) $(x_1 + ey_1) + (x_2 + ey_2) = ([x_1 + x_2] + e[y_1 + y_2])$

and (2) $(x_1 + ey_1) \cdot (x_2 + ey_2) = ([x_1x_2 + e^2y_1y_2] + e[x_1y_2 + x_2y_1])$

$$= ([x_1x_2 - y_1y_2] + e[x_1y_2 + x_2y_1]),$$

using the distributive, commutative, and associative properties of the field F. Furthermore, since R' is a subfield of F and $x_1,y_1,x_2,y_2 \in R'$, then $x_1 + y_1$, $x_2 + y_2 \in R'$ and $x_1x_2 - y_1y_2$, $x_1y_2 + x_2y_1 \in R'$. Therefore, D is closed with respect to addition and multiplication. Since $e = 0 + e\cdot 1$, then $e \in D$ and the closure of D relative to multiplication yields the immediate result that $-1 \in D$. Now note that $0 + 0\cdot e \in D$ and $(x + ey) + (0 + e\cdot 0) = (0 + e\cdot 0) + (x + ey) = (x + ey)$ for every $x + ey \in D$. Hence $0 + e\cdot 0$ is the additive identity. If $x,y \in R'$ and $x + e\cdot y \neq 0 + e\cdot 0$, then $x \neq 0$ or $y \neq 0$. Since R' is isomorphic to R, it possesses a total order $\leq$, thus utilizing the properties in T 3.2. it can be shown that for x and y as above $x^2 + y^2 > 0$. But $x^2 + y^2 \in R'$ and $x^2 + y^2 \neq 0$, thus

$\frac{x}{x^2 + y^2}$, $\frac{-y}{x^2 + y^2} \in R'$. From the foregoing it follows that

$\frac{x}{x^2 + y^2} + e \circ \frac{-y}{x^2 + y^2} \in D$. Now, using (2) and the field properties of R' it is clear that

$$(x + e \circ y) \circ \left(\frac{x}{x^2 + y^2} + e \cdot \frac{-y}{x^2 + y^2}\right) = \frac{x^2 - y(-y)}{x^2 + y^2} + e \circ \left[\frac{x(-y) + y \circ x}{x^2 + y^2}\right]$$

$$= \frac{x^2 + y^2}{x^2 + y^2} + e \circ \left[\frac{0}{x^2 + y^2}\right]$$

$$= 1 + e \circ 0$$

$$= 1$$

The foregoing shows that $\frac{x}{x^2 + y^2} + e \circ \frac{-y}{x^2 + y^2}$ is the multiplicative inverse of $x + e \circ y$. Thus, $(x + e \circ y)^{-1} \in D$. Therefore, the conditions of T 3.1. are satisfied and D is a subfield of F and (i) is established.

(ii) That R' is a subfield of D is apparent from the fact that for every $x \in R'$, $x = x + e \circ 0$.

The preceding result sheds light on the nature of any field containing a system of real numbers and a solution of $x^2 = -1$, but it does not insure the existence of such a field. Nevertheless, the conclusion of the theorem, coupled with (1) and (2), give direction to the formation of the desired system. The following construction produces a field satisfying these necessary conditions.

Theorem 3.6. There exists a field G containing an element e such that:

(i) $e^2 = -1$;

(ii) A subfield of G, denoted R', is isomorphic to R;

(iii) For each $z \in G$ there exist unique elements $x, y \in R'$ such that $z = x + e \cdot y$.

Proof. Let $G = R \times R$ and for every $(x_1,y_1),(x_2,y_2) \in G$ define

(1) $(x_1,y_1) = (x_2,y_2)$ if and only if $x_1 = x_2$ and $y_1 = y_2$,

(2) $(x_1,y_1) + (x_2,y_2) = (x_1 + x_2,\ y_1 + y_2)$,

and (3) $(x_1,y_1) \cdot (x_2,y_2) = (x_1x_2 - y_1y_2\ ,\ x_2y_1 + x_1y_2)$

Clearly, addition and multiplication are well defined binary operations on the nonempty set G. Furthermore, the closure, commutativity, and associativity of addition and multiplication in G follow from the corresponding properties of R. In like manner it can be shown that $\cdot$ distributes over + in G. It is not difficult to show that (0,0) and (1,0) are the respective additive and multiplicative identities and that $-(x,y) = (-x,-y)$. Finally, for $(x,y) \in G, (x,y) \neq (0,0)$, it follows by direct application of the definition of multiplication that $(x,y)^{-1} = \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2}\right)$. Therefore, G forms a field with respect to the prescribed operations.

(i) is established by letting $e = (0,1)$ and noting that $e^2 = (0 \cdot 1 - 1 \cdot 1,\ 0 \cdot 1 + 0 \cdot 1) = (-1,0) = -(1,0)$.

Using T 3.1. it is almost immediate that $R' = \{(x,0) \in G\}$ is a subfield of G, however, consider $f: R \to R'$, where for every $x \in R$, $f(x) = (x,0)$. That f is a function is clear. Furthermore, f is one to one, since if $x_1, x_2 \in R$ and $f(x_1) = f(x_2)$, then $(x_1,0) = (x_2,0)$ which implies that $x_1 = x_2$. It is apparent that f maps R onto R'. Finally, if $x,y \in R$, then $f(x + y) = (x + y,0) = (x,0) + (y,0) = f(x) + f(y)$. Similarly, $f(x \cdot y) = f(x) \cdot f(y)$, hence f preserves the operations. Therefore, f is an isomorphism of R onto R', or R and R' are isomorphic.

The conclusion (ii) is now verified by applying T 3.4.

(iii) follows by noting that for $(x,y) \in G$, $(x,y) = (x,0) + (0,1) \cdot (y,0)$. Recall that $(a,b) = (c,d)$ if and only if $a = c$ and $b = d$.

The foregoing theorem provides an affirmative reply to the initial question regarding the existence of a field containing an isomorphic copy of R and a solution to $x^2 = -1$. Although the approach to the above problem may be new to the reader, it is anticipated that the constructed field is a familiar one. Having experienced success in producing a concrete example of a field satisfying the necessary conditions of T 3.5., the question naturally arises as to what extent the solution is unique. The answer is provided in T 3.8. The following theorem expedites the proof T 3.8. and other subsequent results. Actually the symmetry property of field isomorphisms, which is stated formally below, was tacitly assumed in D 3.6. The proof of this result is well within the means of the reader, but is not presented here.

Theorem 3.7.A. If F, G, and H are fields and $f:F \to G$ and $g:G \to H$ are field isomorphisms onto G and H respectively, then f^{-1} and $f \circ g$ are isomorphisms of G and F onto F and H respectively. [20]

Theorem 3.8. If G_1 and G_2 are two fields satisfying the hypothesis of T 3.6., then G_1 is isomorphic to G_2.

Proof. Let e_1 and e_2 represent the elements of G_1 and G_2 respectively whose square is the additive inverse of unity. Also, denote the subsets of G_1 and G_2 that are isomorphic to R by S_1 and S_2 respectively. By T 3.7. S_1 is isomorphic to S_2, thus let $g:S_1 \to S_2$ be an isomorphism from

S_1 onto S_2. Now, every element in G_1 has a unique representation in the form $x + e_1y$, where $x,y \in S_1$, hence $f:G_1 \to G_2$ such that $f(x + e_1y) = g(x) + e_2 \cdot g(y)$ is a well defined mapping of G_1 into G_2. That f is onto G_2 follows from the fact that g maps S_1 onto S_2 and each element in G_2 has a unique representation in the form $a + e_2b$, $a,b \in S_2$. Furthermore, if $x_1 + e_1y_1$, $x_2 + e_1y_2 \in G_1$ and $f(x_1 + e_1y_1) = f(x_2 + e_1y_2)$, then $g(x_1) + e_2 \cdot g(y_1) = g(x_2) + e_2 \cdot g(y_2)$. Again utilizing the uniqueness of the representation in G_2, the foregoing implies that $g(x_1) = g(x_2)$ and $g(y_1) = g(y_2)$. The fact that g is one to one yields $x_1 = x_2$ and $y_1 = y_2$, or $x_1 + e_1 \cdot y_1 = x_2 + e_1 \cdot y_2$, hence f is one to one. Finally, if $x_1 + e_1 \cdot y_1$, $x_2 + e_1y_2 \in G$, then the statements (1) and (2) of T 3.5. insure that $f([x_1 + e_1y_1] + [x_2 + e_1y_2]) = f(x_1 + e_1y_1) + f(x_2 + e_1y_2)$ and $f([x_1 + e_1y_1] \cdot [x_2 + e_1y_2]) = f(x_1 + e_1y_1) \cdot f(x_2 + e_1y_2)$. Therefore, f is an isomorphism of G_1 onto G_2 and G_1 is isomorphic to G_2.

Theorems 3.6 and 3.8 establish the existence and uniqueness, within an isomorphism, of a field satisfying the conditions of T 3.5. In view of this the following definition is in order.

Definition 3.8. A field G is called a field of complex numbers if and only if:

(i) There exists an element $e \in G$ such that $e^2 = -1$;

(ii) There exists a subfield R' of G isomorphic to R;

(iii) For every $z \in G$ there exist unique elements $x,y \in R'$ such that $z = x + e \cdot y$.

As in the case of a real field, a particular model of a field of complex numbers is singled out and given special status. In this paper the field developed in the following theorem is the designated one and will subsequently be referred to as the field of complex numbers. The symbol C will be used to represent this field.

Theorem 3.9. If C is the set of expressions of the form a + bi, where $a,b \in R$ and equality, addition, and multiplication are defined by the following:

(i) $(x_1 + y_1 i) = (x_2 + y_2 i)$ if and only if $x_1 = x_2$ and $y_1 = y_2$,

(ii) $(x_1 + y_1 i) + (x_2 + y_2 i) = [x_1 + x_2] + [y_1 + y_2]i$,

(iii) $(x_1 + y_1 i)\cdot(x_2 + y_2 i) = [x_1 \cdot x_2 - y_1 \cdot y_2] + [x_1 y_2 + x_2 \cdot y_1]i$,

then C is a field of complex numbers.

Proof. The argument completely parallels T 3.6.

The elements of C of the form a + 0i will be referred to as real complex numbers, or simply real numbers where there is no ambiguity. The expressions in C of the form $0 + b\cdot i$, or briefly denoted bi, will be called imaginary numbers.

The following theorem, which is almost immediate, provides an alternative definition of a field of complex numbers.

Theorem 3.10. A field G is a field of complex numbers if and only if G and C are isomorphic.

Proof. If G is a field of complex numbers, then the isomorphism of G and C is immediate from T 3.8.

If C is isomorphic to G and $f: C \to G$ is an isomorphism of C onto G, where R' is the subset of C isomorphic to R, then $f(R') \subseteq G$ is also isomorphic to R. The foregoing is justified in view of T 3.7. Since $i \in C$ and $i^2 = -1$, then, utilizing the argument in T 3.4, where it was shown that $g(-1) = -g(1) = -1$ for any field isomorphism g, it follows that $f(i^2) = f(-1) = -1$. Thus, f(i) is an element of G whose square is the additive inverse of the multiplicative identity. Finally, if $z \in G$, then since f is a one-to-one mapping of C onto G there exists a unique element $w \in C$ such that $f(w) = z$. But, corresponding to every $w \in C$ there are unique elements $x, y \in R'$ such that $w = x + yi$. Thus, $f(w) = f(x + y \cdot i) = f(x) + f(y) \cdot f(i) = z$, since f preserves operations. In view of the foregoing the elements f(x) and f(y) are clearly the only elements of f(R') satisfying the condition that $z = f(x) + f(i) \cdot f(y)$. Thus, the conditions of D 3.8. are satisfied and G is a field of complex numbers.

In order that the reader be aware of the fact that there exist examples of complex fields where the isomorphism with C is not transparent, another model is considered. A prerequisite of the development of this model is a brief acquaintance with matrices. In particular, it is assumed that the reader is familiar with the matrix operations of sum and product. The following theorem is the focal point of this discussion.

Theorem 3.11. The set G of all real 2×2 matrices of the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, with the usual matrix operations, is a field of complex numbers.

Proof. Although D 3.8. affords a simple proof of this result the characterization of T 3.10. is used since it seems to find wider application. Consider the relation $f: C \to G$, where $f(a + bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ for every $a + bi \in C$. f is a function, since if $a + bi$, $c + di \in C$ and $a + bi = c + di$, then $a = c$ and $b = d$. Hence, $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} c & d \\ -d & c \end{pmatrix}$.

Thus, $f(a + bi) = f(c + di)$. By essentially reversing the foregoing argument it follows that f is one to one. Now, if $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \in G$, then $a, b \in R$; thus, $a + bi \in C$ and $f(a + bi) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Therefore, f is onto G. Finally, if $a + bi, c + di \in C$, then $f[(a + bi)\circ(c + di)]$ $= f[(ac - bd) + (ad + bc)i] = \begin{pmatrix} ac - bd & ad + bc \\ -[ad + bc] & ac - bd \end{pmatrix}$. By applying the definition of matrix product and the field properties of R, $f(a + bi)\circ f(c + di) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}\circ\begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} ac - bd & ad + bc \\ -[ad + bc] & ac - bd \end{pmatrix}$. Thus, $f[(a + bi)\circ(c + di)] = f(a + bi)\circ f(c + di)$. It follows similarly that $f[(a + bi) + (c + di)] = f(a + bi) + f(c + di)$. Therefore, f is an isomorphism of C onto G and by T 3.10. it follows that G is a field of complex numbers.

It is of interest to note that the set G discussed in the preceding theorem is a subset of the set of all real 2×2 matrices, which does not itself form a field with respect to the given operations.

For the reader familiar with the role that matrices play in the theory of linear transformations on a vector space the foregoing matrix model of a field of complex numbers gives some insight into their geometry. An acquaintance with this application of matrices will not be assumed in the sequel.

At this point, having developed in detail a number system containing a subsystem isomorphic to R and a solution to $x^2 + 1 = 0$, the question arises as to what extent the field C provides solutions for other real polynomial equations. The answer is truly amazing. The fact is that the field C not only contains a root to every real polynomial equation, but provides a solution to every polynomial equation having complex coeficients as well. The reader is undoubtedly familiar with this result which is generally termed the Fundamental Theorem of

Algebra. As noted earlier it was first demonstrated by Gauss in 1799. In a sense the Fundamental Theorem of Algebra is not an algebraic theorem at all, since every known proof relies on notions which are foreign to algebra. [4] Close examination of this result leads one to suspect that any proof will lean heavily on topological notions and continuity. Thus, the proof of this result is beyond the scope of this paper. Nevertheless, the theorem is used on occasion and is stated here precisely for reference.

Theorem 3.12.B. If $f(z)$ is a polynomial of degree n, $n \geq 1$, having real or complex coefficients, then the equation $f(z) = 0$ has at least one root in C. [27]

In view of the foregoing result it is clear that no further generalization of the number concept can logically be based on the desire for algebraic completeness. As noted in Chapter II there are several interesting extensions based on other considerations. One of these is examined in Chapter IV. The remainder of the current chapter is devoted primarily to an exposition of the fundamental algebraic properties of C necessary in the sequel. The following result is of this nature.

Theorem 3.13.C. If $P(x)$ is a polynomial with real coefficients, then $P(x)$ can be expressed as a product of factors each of which is of the form $ax + b$ or $cx^2 + dx + e$, where $a,b,c,d,e \in R$. [4]

The proof of the foregoing, though not presented here, is readily accessible to the reader. In connection with the above theorem the necessity of $P(x)$ having real coefficients should be carefully noted.

In view of the inherent strength afforded the real number system by the total ordering that it possesses, it is reasonable to consider the possibility of imposing such an ordering on C. The futility of such a quest is pointed up in the following theorem.

Theorem 3.14. The field C is not totally ordered.

Proof. The proof is by contradiction. Suppose that P is a nonempty subset of C satisfying D 3.3. Now $i \in C$ and $i \neq 0$, hence either $i \in P$ or $-i \in P$ and not both. In case $i \in P$, then $i \cdot i \in P$, or $-1 \in P$. Reapplying the second condition of the definition yields $-1 \cdot i \in P$, or $-i \in P$, which is a contradiction. A similar argument shows that $-i \in P$ leads to a contradiction. Therefore, $i \neq 0$, $i \notin P$ and $-i \notin P$; hence the initial assumption regarding the existence of P must be invalid. Therefore C is not totally ordered.

The widespread utility of order relations in algebraic structures in general leads to an inquiry into the possibility of defining an ordering on C that possesses some of the desirable features of a total order. This is indeed possible and an examination of the properties of a total order exposed in T 3.2. leads to the following result.

Theorem 3.15. The field C possesses an ordering $\leq$ such that conditions (i) - (v) of T 3.2. are satisfied.

Proof. If $x, y \in C$ and $x = a + bi$, $y = c + di$, then define $x \leq y$ if and only if $a < c$, or $a = c$ and $b \leq d$. $\leq$ as used in connection with a, b, c, d is the standard order on R. Let $x, y, z \in C$ with $x = a + bi$, $y = c + di$, $z = e + fi$, then;

(i) clearly $x \leq x$.

(ii) If $x \leq y$ and $y \leq x$, then elimination of the impossible cases produces $a = c$, $b \leq d$, and $d \leq b$, or $a = c$ and $b = d$. Thus, $a + bi = c + di$, or $x = y$.

(iii) If $x \leq y$, then $a < c$, or $a = c$ and $b \leq d$. If also $y \leq z$, then $c < e$, or $c = e$ and $d \leq f$. In considering each of the cases the conclusion $a \leq e$ and $b \leq f$ is valid, thus $x \leq z$.

(iv) Since $a,b,c,d \in R$ and $\leq$ is a total ordering on R, then $a = c$, $a < c$ or $c < a$ and $b \leq d$ or $d \leq b$. A casewise discussion is again in order. If $a = c$ and $b \leq d$, then $x \leq y$. If $a = c$ and $d \leq b$, then $y \leq x$. In case $a < c$, then $x \leq y$, and if $c < a$, then $y \leq x$. Thus, in any event either $x \leq y$ or $y \leq x$.

(v) If $x \leq y$, then $a < c$, or $a = c$ and $b \leq d$. Now, either $a + e < c + e$, or $a + e = c + e$ and $b + f \leq d + f$; hence $x + z \leq y + z$.

In light of the results of T 3.3. it is apparent that the ordering outlined in the foregoing proof fails to be a total ordering of C only on one count. It is of interest to note that the order relation described in T 3.15. is compatible with the standard ordering on R. By compatability with the conventional ordering of R, it is meant that if x and y represent complex numbers of the form $a + 0i$ and $c + 0i$ respectively, then $x \leq y$ if and only if $a \leq c$. The ordering of C outlined in T 3.15. might appropriately be termed a lexicographic ordering. Since the concept of a linear ordering is so well established in mathematics, it is worth mentioning that the lexicographic ordering of C is also a linear ordering. A linear order on a set is one satisfying conditions (i) - (iv) of T 3.15.

In view of the fact that C fails to possess a total ordering, it is clear that the notion of completeness, as it was defined in D 3.5., cannot be extended to the complex field. It turns out that there is a characterization of completeness for ordered fields that can be extended to certain fields that fail to possess a total ordering. This property is enjoyed by the field of complex numbers. [24] The development of this characterization is beyond the scope of this paper.

The following theorem exposes a unique relationship that exists between certain pairs of complex numbers.

Theorem 3.16. If $z_1, z_2 \in C$ with $z_1 = a + bi$ and $z_2 = c + di$, $d \neq 0$, then $z_1 \cdot z_2$ and $z_1 + z_2$ are both real complex numbers if and only if $a = c$ and $b = -d$.

Proof. Suppose $z_1 \cdot z_2$ and $z_1 + z_2$ are both real. Now $z_1 \cdot z_2 = (ac - bd) + (ad + bc)i$ and $z_1 + z_2 = (a + c) + (b + d)i$. The fact that the sum and product are both real imply that $b + d = 0$ and $ad + bc = 0$. From the first equation it follows that $b = -d$. Substituting $b = -d$ in the second equation produces $ad - dc = 0$, or $(a - c)d = 0$. Since $d \neq 0$, then $a = c$. In summary $a = c$ and $b = -d$, or $a + bi = c - di$.

Conversely, if $a = c$ and $b = -d$, then the conclusion that $z_1 \cdot z_2$ and $z_1 + z_2$ are real complex numbers follows readily.

The foregoing theorem motivates the following definition.

Definition 3.9. If $z \in C$ and $z = a + bi$, then $a - bi$ is called the conjugate of z and is denoted $\bar{z}$.

Theorem 3.16. rephrased in terms of the preceding terminology asserts that the sum and product of two nonreal complex numbers are both real if and only if the two numbers are conjugates of each other.

The following algebraic results involving conjugation will be necessary in the sequel. If $z = a + bi$, the notation $Re(z)$ and $Im(z)$ will be used to denote a and b respectively.

Theorem 3.17.D. If $z_1, z_2 \in C$ and $z_1 = a + bi$ then

(i) $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$,

(ii) $\overline{z_1 \cdot z_2} = \overline{z}_1 \cdot \overline{z}_2$,

(iii) $\overline{z_1 - z_2} = \overline{z}_1 - \overline{z}_2$,

(iv) $\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\overline{z}_1}{\overline{z}_2}$,

(v) $\overline{\overline{z}}_1 = z_1$,

(vi) $z_1 \cdot \overline{z}_1 = a^2 + b^2$,

(vii) $z_1 + \overline{z}_1 = 2\,Re(z_1)$ and $z_1 - \overline{z}_1 = 2i\,Im(z_1)$,

(viii) z is real if and only if $z - \overline{z} = 0$.

The proofs of the foregoing results are readily accessible to the reader. [27] Parts (i) and (ii) of T 3.17.D. suggest that conjugation, considered as a mapping of C into itself, might well be an automorphism. This is indeed the case and is stated formally in the following theorem.

Theorem 3.18. The function $f: C \to C$ such that $f(z) = \overline{z}$ is an automorphism of C.

Proof. The brief argument required uses the uniqueness of representation in C and the results of T 3.17.D.

It can be shown that the conjugation map outlined above is the only nonidentical automorphism of C that sends real complex numbers into themselves. Furthermore, although it is not developed here, it can be demonstrated that there are no nonidentical automorphisms of the rational field or the real field. [34] The foregoing facts tend to suggest that the conjugation automorphism is a very fundamental feature of the complex field. The following characterization of a field of complex numbers supports this point of view. [28]

Theorem 3.19. In order that a field G be a field of complex numbers it is necessary and sufficient that G satisfy the following:

(i) $G = \{x^2 \mid x \in G\}$;

(ii) There exists a function $' : G \to G$ such that for every $x, y \in G$ the following hold:

(1) $(x + y)' = x' + y'$;

(2) $(x \circ y)' = x' \circ y'$;

(3) $(x')' = x$;

(4) If x and y are nonzero, there exists $z \in G$, $z \neq 0$, such that $xx' + yy' = zz'$;

(5) If $x = x'$, then there exists $z \in G$ such that $x^2 = (zz')^2$;

(iii) If $R' = \{x \in G \mid x = x'\}$ and $P = \{x \in G \mid$ there exists $y \in G$, $y \neq 0$, and $x = yy'\}$ and A,B are nonempty subsets of G such that $A - B = \{a - b \mid a \in A \text{ and } b \in B\} \subseteq P$, then there exists $c \in R'$ such that $A - \{c\} \subseteq P \cup \{0\}$ and $\{c\} - B \subseteq P \cup \{0\}$.

Proof. The conditions (i) - (iii) are sufficient. The procedure is to verify that D 3.8. is satisfied.

R' is a subfield of G. Clearly $R' \subseteq G$. Using (2) and (3) $x \circ 1' = (x' \circ 1)' = (x')' = x$ for every $x \in G$. In like manner $1' \circ x = x$ for every $x \in G$. Thus, $1' = 1$, or $1 \in R'$ and $R' \neq \emptyset$. Similarly, $0' = (0 \circ 0')' = 0' \circ (0')' = 0' \circ 0 = 0$ and $0 \in R'$. Furthermore, $(1 + -1) = 0$; hence $(1 + -1)' = 0' = 0$, or $1' + (-1)' = 1 + (-1)' = 0$. Thus, $(-1)' = -1$ and $-1 \in R'$. Now, if $x, y \in R'$, then applying (1) and (2) yeilds $(x + y)' = x' + y' = x + y$ and $(x \circ y)' = x' \circ y' = x \circ y$. Hence, R' is closed with respect to addition and multiplication. If $x \in R'$, $x \neq 0$, then $x(x^{-1}) = 1$ and $[x(x^{-1})]' = 1' = 1$. Also, $[x \circ (x^{-1})]' = x' \circ (x^{-1})' = x \circ (x^{-1})'$. Thus, $x \circ (x^{-1})' = 1$, or $(x^{-1})' = x^{-1}$ and $x^{-1} \in R'$. Therefore, the conditions of T 3.1 are satisfied.

P totally orders R'. $P \neq \emptyset$, since $1 = 1 \circ 1'$. $P \subseteq R'$, for if $x \in P$, then $x = yy'$ for some $y \in G$ and $x' = (y \circ y')' = y' \circ y = y \circ y' = x$. Now, if $x, y \in P$, then there exist $a, b \in G$, $a \neq 0$, $b \neq 0$, such that $x = a \circ a'$ and $y = b \circ b'$. Thus, $x \circ y = (a \circ a')(b \circ b') = (a \circ b)(a' \circ b') = (a \circ b)(a \circ b)'$, where $a \circ b \neq 0$. Also, $x + y = a \circ a' + b \circ b'$ and (4) assures the existence of $c \in G$, $c \neq 0$, such that $a \circ a' + b \circ b' = c \circ c'$; hence $x + y = cc'$. Combined, the preceding imply that $x \circ y$, $x + y \in P$, or that P is closed with respect to addition and multiplication. If $x \in R'$ and $x = 0$, then $x \notin P$; for suppose $0 \in P$, then there exists $y \in G$, $y \neq 0$, such that $0 = y \circ y'$. However, $y \circ y' = 0$ implies $y' = 0$; thus using (3), $(y')' = 0'$, or $y = 0$, which is a contradiction. If $x \in R'$ and $x \neq 0$, then $x = x'$ and (5) guarantees the existence of $z \in G$ such that $x^2 = (zz')^2$. Thus, $x = zz'$ or $-x = zz'$. Furthermore, $z \neq 0$ since $x \neq 0$. From the foregoing either $x \in P$ or $-x \in P$.

Finally, not both $x \in P$ and $-x \in P$, for if so, then using the previously established closure $x + (-x) \in P$. But $x + (-x) = 0$ and $0 \notin P$, hence a contradiction.

R' is complete. Let $\leq$ be the order induced by P as described in T 3.2. Consider H, a nonempty subset of R' that is bounded above and define $A = \{x \in R' \mid x > c \text{ for every } c \in H\}$ and $B = \{x \in R' \mid x \leq c$ for some $c \in H\}$. From the fact that H is nonempty and bounded above, it follows that $A \neq \emptyset$ and $B \neq \emptyset$. Now, if $a \in A$ and $b \in B$, then $a > c$ for every $c \in H$ and $b \leq c$ for some $c \in H$; thus, there exists $c_0 \in H$ such that $a > c_0 \geq b$. Therefore, $a > b$, or $a - b > 0$. Thus, $a - b \in P$ and $A - B \subseteq P$. (iii) insures the existence of $c_1 \in R'$ such that $A - c_1 \subseteq P \cup \{0\}$ and $c_1 - B \subseteq P \cup \{0\}$. From the preceding and the definition of $\leq$ it follows $a - c_1 \geq 0$ and $c_1 - b \geq 0$, for every $a \in A$, $b \in B$. Clearly, c_1 is an upper bound for H, since $H \subseteq B$. Furthermore, if $d \in R'$ and d is an upper bound for H distinct from c_1, then $d \in A$; hence $d > c_1$. Therefore, c_1 is the least upper bound for H.

In summary R' is a system of real numbers.

There exists an element e in G such that $e^2 = -1$. This follows from (i) and the fact that $-1 \in G$.

Every element of G has a unique representation in the form $x + e \cdot y$, where $x, y \in R'$. By an argument paralleling that of T 3.14. it follows that $e \notin R'$, thus $e \neq e'$. Since $e^2 = -1$ and $-1 \in R'$, then $e^2 = (e^2)' = (e')^2$, hence $e' = -e$. Also, $1 \in R'$ and R' is closed under addition and nonzero division; thus, $2 \in R'$ and $1/2 \in R'$. Now, if $z \in G$, then $z = 1/2(z + z') + 1/2(z' - z) = 1/2(z + z')$ $+ e[1/2(z - z')e]$. But, $[1/2(z + z')]' = (1/2)'(z + z')' = 1/2(z' + z)$ $= 1/2(z + z')$ and $[1/2(z - z')e]' = (1/2)'(z - z')'e' =$

$1/2(z' - z)(-e) = 1/2(z - z')e$. Therefore, $1/2(z + z')$, $1/2(z - z')e \in R'$, or $z = x + ey$, where $x,y \in R'$. Finally, the representation is unique, for suppose $a,b,c,d \in R'$ and $a + eb = c + ed$, then $a - c = e(d - b)$. However, $a - c \in R'$, thus $e(d - b) \in R'$ and $[e(d - b)]' = e'(d' - b') = (-e)(d - b) = -[e(d - b)]$. From this it follows that $e(d - b) = 0$, or $d = b$. Also, since $a - c = e(d - b) = 0$, then $a = c$.

The proof of the necessity is sketched. If G is a field of complex numbers, then the isomorphism $f:G \to C$ guaranteed by T 3.10., coupled with the fact that $C = \{x^2 \mid x \in C\}$, leads to verification of (i). Furthermore, since the conjugate automorphism of C possesses all the properties outlined in (ii), then it follows that there exists an automorphism of G satisfying (ii). Finally, the isomorphism of G and C, together with the fact that $S = \{z \in C \mid z = \overline{z}\}$ and $T = \{z \in C \mid$ there exists $y \in C$, $y \neq 0$, and $z = y\overline{y}\}$ represent a field of real numbers and its positive elements, can be used to show that the corresponding elements of G satisfy (iii).

It is of note that the foregoing theorem not only presents a characterization of a complex field that focuses on the role of the conjugate relationship, but also provides a definition that does not explicitly assume the existence of a real subfield. The proof of T 3.19., though lengthy, does afford a rare opportunity for the reader to become better acquainted with the definitive properties of the real number system.

A very important notion in the real field is that of absolute value. This concept can be extended to the complex field in the following way.

Definition 3.10. If $z \in C$, $z = a + bi$, then the absolute value of z, denoted $|z|$, is defined to be $\sqrt{a^2 + b^2}$. I.e., $|z| = \sqrt{a^2 + b^2}$.

It is of note that if z is a real complex number, then the absolute value of z agrees with absolute value as defined in R. Furthermore, the following theorem shows that absolute value as defined in C shares the main properties of the absolute value function in R. The proof of this result is readily accessible to the reader. [27]

Theorem 3.20.E. If $z_1, z_2 \in C$, then

(i) $|z_1| \geq 0$ and $|z| = 0$ if and only if $z = 0$,

(ii) $|z_1 z_2| = |z_1||z_2|$,

(iii) $\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}$, if $z_2 \neq 0$,

(iv) $|z_1 + z_2| \leq |z_1| + |z_2|$,

(v) $|z_1 - z_2| \leq ||z_1| - |z_2||$,

(vi) $|z_1 + z_2|^2 + |z_1 - z_2|^2 = 2|z_1|^2 + 2|z_2|^2$,

(vii) $|z_1|^2 = z_1\bar{z}_1$,

(viii) $|z_1| = |\bar{z}_1|$.

With the foregoing theorem the principal properties of the complex field that do not hinge on geometric or trigonometric notions have been presented. In particular, those algebraic aspects of C that are necessary for the development in Chapter IV have been exposed. Chapter V provides an appropriate setting for the presentation of those results that have a geometric flavor.

CHAPTER IV

HYPERCOMPLEX NUMBER SYSTEMS

The intent of this chapter is to view the complex field from another vantage point. The development here is concerned with the area of algebra outlined historically in Chapter II. Specifically, attention is directed toward an elementary exposition of certain results that might appropriately be included in a study of finite linear algebras over the real field. In view of the approach taken in this paper it will be unnecessary to define formally the concept of a linear algebra. The informed reader will recognize that most of the theory of this section could have been presented much more elegantly (and esoterically) in a vector space setting. Furthermore, it will be noted that the approach taken in this chapter is again a progressive one, where brevity is often sacrificed in an effort to motivate subsequent aspects of the work. In the main, the propositions in this section are unique to this paper. However, there are references that contain results that relate directly to certain aspects of the development. [21] [22]

Historically the initial stimulus for the development of hypercomplex number systems was geometric in nature. As noted in Chapter II, Hamilton's discovery of the real quaternion algebra was the initial step in this direction following Gauss' treatment of complex numbers as ordered pairs of real numbers. The reader will be in a better position to appreciate the geometric considerations which stimulated Hamilton after viewing the simplicity afforded the study of the rigid motions of the plane using complex numbers. Nevertheless, the appeal at this point does not rely on a knowledge of such.

In Chapter III the definition of a complex number system was precipitated by the desire for a field containing a solution to a certain polynomial equation and having a subfield isomorphic to R. It was observed that the field C fulfilled these conditions and furthermore that it contained a root to every polynomial equation with complex coefficients. Having achieved complete success in one direction it is in the nature of a mathematician to seek alternative avenues of generalization. Specifically, after viewing the construction of a field whose elements are ordered pairs of real numbers (see T 3.6.) it is reasonable to raise the question as to the possibility of defining sum and product for ordered triples of real numbers in such a way that the resultant system is a field. Finally, it is recognized that there is nothing magical about the number three, thus the foregoing question might well be posed for systems where the elements are ordered n tuples of real numbers. This question forms the framework for the investigations in this chapter.

The perceptive reader will recognize that the scope of the presentation here could readily be enlarged by considering the possibility of defining binary field operations on n tuples where the coordinates are from fields other than R. In addition, one could further broaden the development by allowing infinite tuples or by surpressing certain of the field axioms. Of course, the intent here is not to encompass the field of linear algebra, but to focus on a single aspect of it that is compatible with the central theme of the paper. In keeping with this, all subsequent references will be to real n tuples. A system whose elements are ordered n tuples will be referred to as n dimensional.

In an effort to reduce the problem to manageable proportions it is necessary to impose certain restrictions on the definitions of sum and product. The appropriate conditions are suggested by both algebraic and geometric considerations. From an algebraic standpoint it is reasonable to seek a definition of sum and product in such a way that the base field R is a subfield of the proposed system. Geometrically it is the vector interpretation of ordered pairs, and more generally of ordered n tuples, which gives direction to the quest for an appropriate set of restrictions. It is assumed that the reader has been exposed to vector methods in plane geometry, thus putting him in a position to recognize that if an algebra of n tuples is to find application in geometry the operations defined on them should reflect basic vector operations.

Specifically, vector considerations point directly to defining the sum of two n tuples component-wise. The appropriate restriction on the product is not as apparent. However, a structure isomorphic to R can be attained by demanding that multiplication of an arbitrary n tuple by one having at most a nonzero entry in the first position also be defined component-wise. Geometrically such a restriction on the definition insures that the product of an arbitrary n tuple (vector) and a real n tuple (one having zeroes in the 2nd through nth position) will produce a result analogous to that of scalar multiplication. Finally, both algebraic and geometric considerations suggest the sort of uniqueness of representation inherent in a component-wise definition of equality.

It is under the restrictions outlined in the preceding paragraph that the investigations of the current chapter are carried out. The following theorem is presented to expose the necessary conditions imposed on the definition of multiplication in the two dimensional case. To expedite the demonstration of this result and subsequent similar theorems the n tuples of the form $(x,0,0,\ldots,0)$ will simply be denoted x on occasion. Furthermore, these n tuples possess all the properties of R, the isomorphism being transparent in light of the restrictions. In view of this, these elements will be termed real elements of the system or briefly, real numbers. The context will clarify the meaning.

Theorem 4.1. If $G = R \times R$ is a field, where for all $x,y,u,v \in R$

(i) $(x,y) = (u,v)$ if and only if $x = u$ and $y = v$,

(ii) $(x,y) + (u,v) = (x + u,\ y + v)$,

(iii) $(x,0)(u,v) = (xu,xv)$,

then (iv) $(x,y)(u,v) = (xu + yvm,\ xv + yu + yvn)$, where $(m,n) = (0,1)^2$ and $m^2 + 4n < 0$.

Proof. Note that for $(x,y) \in G$, $(x,y) = (x,0) + (0,y) = (x,0)(1,0) + (y,0)(0,1)$, or using the aforementioned convention $(x,y) = x(1,0) + y(0,1)$. Thus, if $(x,y),(u,v) \in G$, then $(x,y)(u,v) = [x(1,0) + y(0,1)]\cdot[u(1,0) + v(0,1)]$; using the field properties $(x,y)(u,v) = xu(1,0)^2 + (xv + yu)(1,0)(0,1) + yv(0,1)^2$. Further simplification produces $(x,y)(u,v) = (xu,\ xv + yu) + yv(0,1)^2$. Thus, letting $(0,1)^2 = (m,n)$, $m,n \in R$, $(x,y)(u,v) = (xu + yvm,\ xv + yu + yvn)$. However, G being a field requires that for each $(x,y) \in G$, $(x,y) \neq (0,0)$, there exists a unique $(u,v) \in G$, such that $(x,y)(u,v) = (xu + yvm,\ xv + yu + yvn) = (1,0)$; $(1,0)$ clearly being the multiplicative identity for G. But (xu + yvm, xv + yu + yvn) = (1,0) implies that (ym)v + xu = 1 and (x + yn)v + yu = 0. However, this system of equations possesses a unique solution for u and v if and only if

$$\begin{vmatrix} ym & x \\ x + yn & y \end{vmatrix} \neq 0,$$

or equivalently if and only if $y^2m - x(x + yn) \neq 0$. The foregoing can be expressed in the form $x^2 + (yn)x - y^2m \neq 0$, and treating this as a quadratic in x produces

$$x \neq \frac{-yn \pm \sqrt{y^2n^2 + 4y^2m}}{2} = \left(\frac{-n \pm \sqrt{n^2 + 4m}}{2}\right) y.$$

Thus, the condition that every nonzero (x,y) have a unique inverse fails to be satisfied if and only if there exists an $(x,y) \neq (0,0)$ such that

$$x = \left(\frac{-n \pm \sqrt{n^2 + 4m}}{2}\right) y.$$

Now, if this requirement fails it does so for nonzero y. For suppose y = 0 and

$$x = \left(\frac{-n \pm \sqrt{n^2 + 4m}}{2}\right) y,$$

then x = 0; hence (x,y) = (0,0), which is a contradiction. However, given $y \in R$, $y \neq 0$, there exists an x in R such that

$$x = \left(\frac{-n \pm \sqrt{n^2 + 4m}}{2}\right) y$$

if and only if $n^2 + 4m \geq 0$; since if $n^2 + 4m < 0$, then

$$\left(\frac{-n \pm \sqrt{n^2 + 4m}}{2}\right) y$$

is not real. Therefore, since G is a field it is necessary that (m,n) be such that $n^2 + 4m < 0$. The conclusion follows from the foregoing and the fact that $(x,y)(u,v) = (xu + yvm, xv + yu + yvn)$.

The reader has undoubtedly recognized that the model of a complex field developed in T 3.6. satisfies all the conditions of the hypothesis in the preceding theorem. In that particular two dimensional field the square of $(0,1)$ was defined to be $(-1,0)$ which is, of course, in keeping with the results of T 4.1. e.g., $0^2 + 4(-1) < 0$ and $(x,y)(u,v) = (xu - yv, xv + yu) = (xu + yv(-1), xv + yu + yv0)$.

The series of equivalent statements occuring in the proof of the foregoing theorem suggest that the conditions of the conclusion may be sufficient to insure that the set G satisfying (i) - (iii) be a field. Indeed this is the case and as a matter of fact the resultant fields have a familiar structure.

Theorem 4.2. If $G = R \times R$ and G satisfies conditions (i) - (iv) of T 4.1., then G is a field of complex numbers.

Proof. The result is established by:

(1) verifying that with the given hypothesis G is a field;

(2) exhibiting an isomorphism between C and G and invoking T 3.10.

The demonstration of the fact that G is a field with respect to the prescribed operations is tedious but straightforward and is left to the reader. It should be observed that the if and only if statements of T 4.1. are sufficient to insure the existence of a multiplicative inverse for each nonzero element of G, thus establishing the most difficult portion of this proof.

To show that C is isomorphic to G consider the relation $f: C \rightarrow G$ such that for every $a + bi \in C$,

$$f(a + bi) = \left(a + bn\sqrt{\frac{-1}{n^2 + 4m}},\ -2b\sqrt{\frac{-1}{n^2 + 4m}}\right).$$

The righthand member of the foregoing equality is an element of G, since

$$a, b, n, \sqrt{\frac{-1}{n^2 + 4m}} \in R.$$

Note that

$$\sqrt{\frac{-1}{n^2 + 4m}} \in R$$

is insured since, by hypothesis, $n^2 + 4m < 0$, which implies that

$$\frac{-1}{n^2 + 4m} > 0.$$

To facilitate the presentation

$$\sqrt{\frac{-1}{n^2 + 4m}}$$

will be denoted e throughout the remainder of the proof. In this notation $f(a + bi) = (a + bne, -2be)$.

The relation f defined above is a function, since if $a + bi$, $c + di \in C$ and $a + bi = c + di$, then $a = c$ and $b = d$. Hence $a + bne = c + dne$ and $-2be = -2de$. Thus, from the definition of equality in G it follows that $(a + bne, -2be) = (c + dne, -2de)$, or $f(a + bi) = f(c + di)$. Furthermore, the mapping is one to one, for if $a + bi$, $c + di \in G$ and $f(a + bi) = f(c + di)$, then $(a + bne, -2be) = (c + dne, -2de)$, or $a + bne = c + dne$ and $-2be = -2de$. Since $e \neq 0$ the last equality implies that $b = d$ and substituting into its predecessor yields $a + bne = c + bne$, from which it follows that $a = c$. Therefore, $a + bi = c + di$.

To show that the function maps C onto G consider $(x,y) \in G$. Then f is onto G if and only if there exists $a + bi \in C$ such that

$f(a + bi) = (x, y)$, or equivalently if and only if there exist real numbers a and b such that $a + bne = x$ and $-2be = y$. However, the existence of a real solution to these equations is assured since the coefficients are real and $\begin{vmatrix} 1 & ne \\ 0 & -2e \end{vmatrix} = -2e \neq 0$. Therefore, the range of f is G.

To verify that the operations are preserved under f let $a + bi$, $c + di \in C$, then $f[(a + bi) + (c + di)] = f[(a + c) + (b + d)i] = ([a + c] + [b + d]ne, -2[b + d]e)$. Utilizing the definition of addition in G and the field properties of R, it follows that $([a + c] + [b + d]ne, -2[b + d]e) = (a + bne, -2be) + (c + dne, -2de)$. Thus, $f[(a + bi) + (c + di)] = f(a + bi) + f(c + di)$ and sums are preserved under f. Now, $f[(a + bi)(c + di)] = f[(ac - bd) + (ad + bc)i] = ([ac - bd] + [ad + bc]ne, -2[ad + bc]e)$. Employing the definition of multiplication in G leads to $f(a + bi) \cdot f(c + di) = (a + bne, -2be) \cdot (c + dne, -2de) = ([a + bne][c + dne] + [-2be][-2de]m, [a + bne][-2de] + [-2be][c + dne] + [-2be][-2de]n)$. To see that the preceding ponderous expression does indeed reduce to the expression for $f[(a + bi)(c + di)]$ consider the first component $[a + bne][c + dne] + [-2be][-2de]m$. Using the field properties of R this can be written $ac + bde^2[n^2 + 4m] + [ad + bc]ne$. However

$$e^2 = \frac{-1}{n^2 + 4m}$$

and substituting this produces,

$$ac + bd\left[\frac{-1}{n^2 + 4m}\right][n^2 + 4m] + [ad + bc]ne = [ac - bd] + [ad + bc]ne,$$

which is the first component in the expansion of $f[(a + bi)(c + di)]$. A similar approach can be used to verify the equality of the second components, hence $f[(a + bi)(c + di)] = f(a + bi) \cdot f(c + di)$.

Therefore, f is an isomorphism of C onto G, or C and G are isomorphic.

In addition to showing the necessity and sufficiency of condition (iv) of T 4.1. the preceding two theorems establish the uniqueness (within an isomorphism) of the complex field as a two dimensional extension of the real field. It is of interest to note that conditions (i)-(iv) expose a means of constructing a variety of ordered pair models of the complex field where multiplication is strikingly (though not abstractly) different from the model in T 3.6. At this point the reader might reasonably raise the question as to whether or not there exists a simple distinguishing feature which separates the familiar two dimensional complex field of T 3.6. from the infinitude of distinct complex fields assured by T 4.2. That such a condition does exist is established by the following theorem. [17]

Theorem 4.3. If $G = R \times R$ satisfies (i)-(iv) of T 4.1. and for every $(x,y),(u,v) \in G$ $|(x,y)(u,v)| = |(x,y)||(u,v)|$, where $|(x,y)| = \sqrt{x^2 + y^2}$, then $(x,y)(u,v) = (xu - yv, xv + yu)$.

Proof. First note that $|(x,y)(u,v)| = |(x,y)||(u,v)|$ implies that $|(x,y)(u,v)|^2 = |(x,y)|^2|(u,v)|^2$. Now, using the definition of product in G in condition (iv) and the definition of absolute value, the foregoing can be written $(xu + yvm)^2 + (xv + yu + yvn)^2 = (x^2 + y^2)(u^2 + v^2)$. Expanding and rearranging terms into a convenient form yields

(1) $(1 - m^2 - n^2)y^2v^2 = 2(m + 1)xyuv + 2n(uvy^2 + xyv^2)$.

However, $|(m,n)| = |(0,1)^2| = |(0,1)||(0,1)|$, or equivalently $m^2 + n^2 = (0^2 + 1^2)(0^2 + 1^2) = 1$. Hence, $m^2 + n^2 = 1$, or alternately $1 - m^2 - n^2 = 0$. Substituting 0 for $1 - m^2 - n^2$ in (1) produces

(2) $0 = 2(m + 1)xyuv + 2n(uvy^2 + xyv^2)$.

Since x,y,u,v are arbitrary elements of R, then in particular (2) is valid for $u = v = y = 1$ and $x = 0$, in which case the equality becomes $0 = 2n$. Therefore, $n = 0$. Replacing n by 0 in (2) and letting $x = y = u = v = 1$ yields $0 = 2(m + 1)$, whence $m = -1$. Upon substituting $m = -1$ and $n = 0$ in (iv) the conclusion follows.

In view of the algebraic and geometric significance of the absolute value function as defined in D 3.10. The foregoing theorem suggests that no other ordered pair model of the complex field could play the functional role of the model of T 3.6. The importance of the product relationship for absolute value in R and C points to the desirability of seeking hypercomplex fields which satisfy the prescribed conditions and also have this feature. Unfortunately no such field exists for dimension n, $n > 2$. In fact, no higher dimensional fields exist that satisfy only the three initial restrictions. Because the proof of the last result is somewhat more sophisticated the weaker theorem is also demonstrated here. It is perhaps instructive to note that the author developed the weaker implication after initial efforts to prove the stronger result failed.

The symbol R^n will be used henceforth to denote the set of all ordered n tuples of real numbers.

Theorem 4.4. If $G = R^n$, $n > 2$, and + and $\cdot$ are binary operations on G such that for every $(x_1,x_2,\ldots,x_n)$, $(y_1,y_2,\ldots,y_n) \in G$

(i) $(x_1,x_2,\ldots,x_n) = (y_1,y_2,\ldots,y_n)$ if and only if $x_i = y_i$, $1 \leq i \leq n$,

(ii) $(x_1,x_2,\ldots,x_n) + (y_1,y_2,\ldots,y_n) = (x_1 + y_1,x_2 + y_2,\ldots,x_n + y_n)$,

(iii) $(x_1,0,0,\ldots,0)(y_1,y_2,\ldots,y_n) = (x_1y_1,x_1y_2,\ldots,x_1y_n)$,

(iv) $|(x_1,x_2,\ldots,x_n)(y_1,y_2,\ldots,y_n)| =$

$|(x_1,x_2,\ldots,x_n)||(y_1,y_2,\ldots,y_n)|$,

where $|(x_1,x_2,\ldots,x_n)| = \sqrt{x_1^2 + x_2^2 \ldots + x_n^2}$, then G is not a field.

Proof. The proof is by contradiction. To expedite the argument let e_i denote the element of G having 1 in the ith position and zeroes elsewhere. E.g., $e_1 = (1,0,0,\ldots,0)$, $e_2 = (0,1,0,0,\ldots,0)$, etc.

Suppose $G = R^n$, $n > 2$, is a field satisfying the given conditions, then each element of G can be written in the form $a_1e_1 + a_2e_2 +\ldots+ a_ne_n$, where a_i is a real element of G. Also note that $|a_1e_1 + a_2e_2 +\ldots+ a_ne_n| = \sqrt{a_1^2 + a_2^2 + \ldots + a_n^2}$. In particular, let $e_2^2 = b_1e_1 + b_2e_2 + \ldots + b_ne_n$, where b_i is real, $1 \leq i \leq n$. Utilizing the assumed field properties of G and the fact that $e_1^2 = e_1$ produces $(e_1 + e_2)(e_1 - e_2) = e_1^2 - e_2^2 = e_1 - (b_1e_1 + b_2e_2 + \ldots b_ne_n)$, or simply $(e_1 + e_2)(e_1 - e_2) = (1 - b_1)e_1 - b_2e_2 - \ldots - b_ne_n$. Thus, condition (iv) yields $|e_1 + e_2||e_1 - e_2| = |(1 - b_1)e_1 - b_2e_2 - \ldots - b_ne_n|$, or upon squaring and substituting $(1^2 + 1^2)(1^2 + 1^2) = (1 - b_1)^2 + b_2^2 + \ldots + b_n^2$. Equivalently (1) $4 = 1 - 2b_1 + (b_1^2 + b_2^2 + \ldots + b_n^2)$. Now, $|e_2^2|^2 = |e_2|^4 = |b_1e_1 + b_2e_2 + \ldots + b_ne_n|^2 = b_1^2 + b_2^2 + \ldots + b_n^2$. Also, $|e_2|^4 = |(0,1,0,0,\ldots,0)|^4 = 1$. Therefore, $b_1^2 + b_2^2 + \ldots + b_n^2 = 1$ and substituting into (1) produces $4 = 1 - 2b_1 + 1$. Hence $b_1 = -1$ and $1 = b_1^2 + b_2^2 + \ldots + b_n^2 = 1^2 + b_2^2 + \ldots + b_n^2$. Thus, $b_2^2 + b_3^2 +\ldots+b_n^2 = 0$, from which it follows that $b_2 = b_3 = \dots = b_n = 0$. Therefore, $e_2^2 = -e_1$. Since $n > 2$, $e_3 \in G$ and precisely the same argument can be used to show that $e_3^2 = -e_1$. Thus, $(e_2 + e_3)(e_2 - e_3) = e_2^2 - e_3^2 = 0$. Since neither $e_2 + e_3$ nor $e_2 - e_3$ is zero the foregoing contradicts the fact that in a field there are no divisors of zero. Hence the assumption that G is a field is invalid and the conlusion follows.

The foregoing theorem assures the futility of any further quest for a field over R of dimension greater than two, which possesses the desirable features outlined in the hypothesis of T 4.4. It is of note however, that Hamiltons quaternion algebra does satisfy conditions (i)-(iv) and, in fact, fails to be a field only in that multiplication is not commutative. A system satisfying all the field properties except the commutative law of multiplication is called a skew field. In light of the historical significance of Hamilton's system and its relevance to the material of this chapter the essential structure of quaternion algebra is outlined. Many of the details are omitted.

If for each $x,y \in Q = R^4$, $x = (x_1,x_2,x_3,x_4)$, $y = (y_1,y_2,y_3,y_4)$, equality and addition are defined component-wise and multiplication is performed according to the equation below, then the resultant system is the real quaternion algebra.

$$(x_1,x_2,x_3,x_4)(y_1,y_2,y_3,y_4) = ([x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4],$$
$$[x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3], [x_1y_3 + x_3y_1 + x_4y_2 - x_2y_4],$$
$$[x_1y_4 + x_4y_1 + x_2y_3 - x_3y_2]).$$

With the exception of the associative and inverse properties of multiplication it is relatively easy to show that the conditions for a skew field are satisfied by the foregoing system. To prove that every nonzero quaternion has a multiplicative inverse define the conjugate of $x = (x_1, x_2, x_3, x_4)$ to be $x^* = (x_1, -x_2, -x_3, -x_4)$. Note that $xx^* = x_1^2 + x_2^2 + x_3^2 + x_4^2$, so that xx^* is a non-negative real number. Now it is readily seen that the absolute value of x, defined by $|x| = \sqrt{xx^*}$, satisfies $|x| = \sqrt{xx^*} = \sqrt{x^*x} = \sqrt{x_1^2 + x_2^2 + x_3^2 + x_4^2}$. From this it is clear that $|x| > 0$, if $x \neq 0$. Thus, for $x \neq 0$, $(\frac{1}{|x|^2}, 0,0,0) \in Q$, or briefly $\frac{1}{|x|^2} \in Q$. By direct application of the definition of multiplication it follows that $\frac{1}{|x|^2} x^* = x^{-1}$, for $x \neq 0$. The brute force approach could be used to validate the associative property, but there is a slightly simpler attack outlined in Birkhoff. [4]

At this point it is not difficult to show that Q satisfies conditions (i) - (iv) or T 4.4. The first three are almost immediate. To establish (iv) consider $x,y,z \in Q$, where x and y are arbitrary and z is real, then one can verify that $(xy)^* = y^*x^*$ and $xz = zx$. Using these two facts and the associative property of multiplication it follows that $|xy|^2 = (xy)(xy)^* = (xy)(y^*x^*) = x(yy^*)x^* = (xx^*)(yy^*) = |x|^2|y|^2$. Hence, $|xy| = |x||y|$. Thus, Q is a four dimensional skew field over R having all the features of the hypothesis of T 4.4.

The reader has undoubtedly suspected by now (see T 4.1.) that the essence of the problem of defining a product on R^n so that the resultant system will be a field satisfying the given conditions, is that of defining multiplication for the units of the system. In the notation of T 4.4. the units are the elements e_i, $1 \leq i \leq n$, having 1 in the ith position and zero elsewhere and such that each $x \in R^n$ can be written in the form $a_1e_1 + a_2e_2 + \ldots + a_ne_n$, where a_i is real. Similarly, for the four dimensional quaternion skew field multiplication is determined once the unital products are given. Thus, although the prospect of finding the product of two quaternions is frightening at first glance, the essence of multiplication is embodied in the following table. e_1, e_2, e_3, and e_4 represent the units of Q.

$\circ$	e_1	e_2	e_3	e_4
e_1	e_1	e_2	e_3	e_4
e_2	e_2	$-e_1$	e_4	$-e_3$
e_3	e_3	$-e_4$	$-e_1$	e_2
e_4	e_4	e_3	$-e_2$	$-e_1$

In the event the reader has not already confirmed the noncommutativity of multiplication an examination of the unit products above will expose this. E.g., $e_2e_3 = -e_3e_2$. Furthermore, associativity could be established by considering the various combinations of unit products and noting that real quaternions associate.

In addition to the basic properties of quaternion algebra outlined above this system has several other interesting features that are not pursued here. For example, it has been shown that every polynomial equation over Q contains a root in Q, a result analogous to the Fundamental Theorem of Algebra. [25] In view of the rather nice behavior of quaternions it is not surprising that mathematicians have addressed themselves to the question of the existence of other similar structures. An eight dimensional system, called the Cayley algebra, was discovered about 1850 by the man whose name it bears. This system failed to be a field in that multiplication was neither commutative nor associative. [21] However, in spite of these deficiencies, Cayley numbers can be shown to possess all the desirable features of T 4.4. Although considerable effort was expended in this direction between 1850 and 1950, it has only recently been shown that, aside from isomorphic copies of the quaternion algebra and Cayley algebra, all other hypercomplex systems are degenerate to the point of having divisors of zero. [6][7] This result is incompatible with the development here, nevertheless the following theorem is a step in this direction.

Theorem 4.5. There exists no field of dimension n, $n > 2$, over R satisfying conditions (i) - (iii) of T 4.4.

Proof. The proof is by contradiction. Suppose $G = R^n$, $n > 2$, is such a field and that e_i, $i \leq i \leq n$, are the units. Consider the elements $e_2^n, e_2^{n-1}, \ldots, e_2$. The assertion is that there exist real elements $x_n, x_{n-1}, \ldots, x_1, x_0 \in G$, such that $x_n e_2^n + x_{n-1} e_2^{n-1} + \ldots + x_1 e_2^1 + x_0 = 0$, where at least one $x_i \neq 0$. This is clearly equivalent to claiming that e_2 is a root of a real polynomial of degree n, where $n \geq 2$. Now let $e_2^i = a_{i1}e_1 + a_{i2}e_2 + \ldots + a_{in}e_n$ and note that in view of conditions (i) - (iii) the existence of x_i's satisfying the foregoing is contingent of the existence of a nontrivial solution to the following system of n real homogeneous equations in n + 1 variables.

$$\begin{aligned}
a_{n1}x_n + a_{(n-1)1}x_{n-1} + \cdots + a_{11}x_1 + x_0 &= 0 \\
a_{n2}x_n + a_{(n-1)2}x_{n-1} + \cdots + a_{12}x_1 \quad &= 0 \\
\vdots \qquad \vdots \qquad \qquad \vdots \quad & \\
a_{nn}x_n + a_{(n-1)n}x_{n-1} + \cdots + a_{1n}x_1 \quad &= 0
\end{aligned}$$

However, such a system always has a nontrivial solution in R. Thus, let $x_n, x_{n-1}, \ldots, x_1, x_0$ be real elements of G satisfying the above, then e_2 is a root of (1) $x_n y^n + x_{n-1}y^{n-1} + \ldots \; x_1 y + x_0 = 0$. Moreover, T 3.13.C. asserts that every polynomial with real coefficients can be expressed as a product of quadratic and linear factors having real coefficients. Since e_2 satisfies (1) and is not real, then it must be a root of an irreducible quadratic polynomial equation with coefficients in R. Suppose $ae_2^2 + be_2 + c = 0$, where a,b and c are real, is such an equation. Then,

$$e_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \text{ or alternately } \left(\frac{2ae_2 + b}{\sqrt{4ac - b^2}}\right)^2 = -1.$$

Letting

$$k_1 = \frac{2a}{\sqrt{4ac - b^2}},\ k_2 = \frac{b}{\sqrt{4ac - b^2}},$$

the preceding can be written $(k_1 e_2 + k_2)^2 = -1$, where k_1 and k_2 are real since $b^2 - 4ac < 0$ implies $4ac - b^2 \geq 0$. Using an argument parallel to the above it follows that there exist ℓ_1, ℓ_2 real such that $(\ell_1 e_3 + \ell_2)^2 = -1$. Now note that in view of conditions (i) - (iii) $\ell_1 e_3 + \ell_2 \neq \pm (k_1 e_2 + k_2)$; for if so, then $\ell_1 e_3 - k_1 e_2 = k_2 - \ell_2$, or

$\ell_1 e_3 + k_1 e_2 = -k_2 - \ell_2$. Considering the first case in the original notation yields the equivalent condition $(0, -k_1, \ell_1, 0, 0, \ldots, 0)$ $= (k_2 - \ell_2, 0, 0, \ldots, 0)$, which implies that $k_2 - \ell_2 = -k_1 = \ell_1 = 0$. This leads to $(k_1 e_2 + k_2)^2 = k_2^2 = -1$, which is impossible since k_2 is real. Similarly, $\ell_1 e_3 + \ell_2 \neq -(k_1 e_2 + k_2)$. Letting $k_1 e_2 + k_2 = i$ and $\ell_1 e_3 + \ell_2 = j$, it follows that $i^2 - j^2 = 0$, where $i \neq \pm j$. Using the assumed commutativity of multiplication and distributivity of multiplication over addition, the preceding can be written $(i + j)(i - j) = 0$. Since neither $i + j$ nor $i - j$ are zero this contradicts the fact that for elements of a field $ab = 0$ if and only if $a = 0$ or $b = 0$. Therefore, the assumption made is false and the result is established.

In summary the theorems of this chapter point up the unique position of the complex field as a finite dimensional extension of R. In addition to the central theme, the discussion affords the reader a glimpse of a branch of algebra that is a direct descendent of investigations of the complex field. Furthermore, the development provides a natural setting for an exposure to some recent fruits of mathematical research.

CHAPTER V

GEOMETRY OF COMPLEX NUMBERS

The current chapter is devoted to an elementary exposition of certain results which might appropriately belong in a study of the complex analytic geometry of the plane. Most of the propositions presented here are available elsewhere. [13][35] There are two principal reasons for including such a discussion in this paper. First of all, any introductory treatment of the complex field would be incomplete without some reference to the geometric interpretation of complex numbers provided by Wessel, Argand and Gauss. One might well justify attention to their interpretation solely on a historical basis. However, the writer draws support for the inclusion from the fact that the geometry of the complex plane can be a significant intuitive aid in studying functions of a complex variable. Second, although the treatment of the isometries of Euclidean two space as presented in Chapter VI is basically algebraic, it is clear that the motivation for such a discussion is geometric. In view of this subsequent chapter the reader might anticipate that the current section would show a bias in favor of those results which are pertinent to the development in Chapter VI. Indeed this is the case.

Before directing attention to the central notions of the chapter it is appropriate to point out that the ensuing presentation is not as axiomatic as in the two preceding sections. The somewhat informal approach taken in this chapter is not only expedient, but is in keeping with the fact that in this paper geometry is utilized primarily as a vehicle for motivating the work in Chapter VI. In addition to the foregoing, the reader will note that the development in this section relies heavily on results from trigonometry as well as elementary geometry. This is consistent with the background assumptions made at the outset.

The point of departure for a geometric interpretation of complex numbers is T 3.6. This theorem suggests that it is natural to represent elements of C as points of the plane. The obvious correspondence is that of associating the complex number $x + iy$ with the point having Cartesian coordinates (x,y). When used in this fashion for the purpose of displaying complex numbers the rectangular coordinate system is generally referred to as the Argand plane, or simply the complex plane. The horizontal and vertical axes are referred to as the real and imaginary axis respectively.

With the foregoing representation in mind it is not difficult to see that the conjugation mapping corresponds to a reflection in the real axis. Similarly, it is almost immediate that the additive inverse of z corresponds to the image of the point associated with z under a reflection inthe origin. Finally, it is clear that $|z|$ is representative of the distance from the origin to the point corresponding to z. Figure 5.1 illustrates the foregoing.

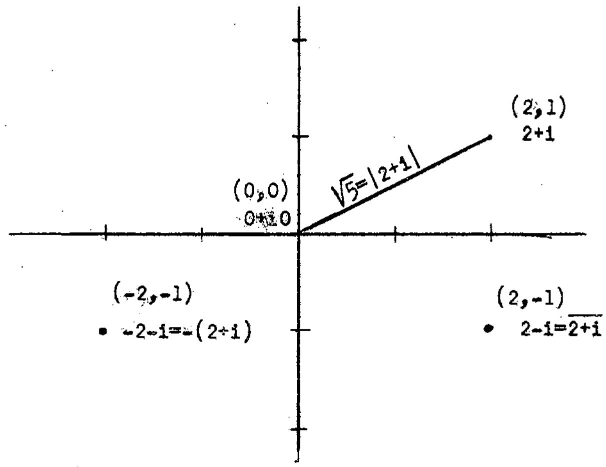

Figure 5.1.

In addition to the point interpretation of complex numbers it is apparent that each complex number z can be identified with the directed line segment, or vector, from the origin to the point associated with z. Those familiar with a vector approach to geometry will recognize that it is more appropriate to identify z with all directed segments in the plane having the same length and sense as the vector from the origin to the point corresponding to z. This association between complex numbers and classes of directed segments will, on occasion, provide the most revealing interpretation of C. In other instances the point interpretation of z will be more appropriate. The symbol z will be used interchangeably to represent the number, the associated point, and the corresponding class of vectors. The context will clarify the meaning. The phrase 'the vector z' will be used to refer to any element of the class of vectors identified with z.

The geometric representation of the elements of C as points of the plane, or vectors, is not revealing in itself. The aspect of these interpretations that provides insight into the structure of C is a result of the fact that to each of the fundamental operations on complex numbers there corresponds a geometric construction. These constructions form the basis for the analytic geometry of the Argand plane.

If $z = x + iy$ and $w = u + iv$, then $z + w = (x + u) + i(y + v)$ and it is easy to verify that the point representing $z + w$ can be obtained from the points $0, z, w$ by completing the parallelogram having $\overline{0z}$ and $\overline{0w}$ as a pair of adjacent sides. The fourth vertex of this quadrilateral is the point corresponding to $z + w$. Of course, this is essentially the parallelogram rule for finding the vector sum in the plane. See Figure 5.2.

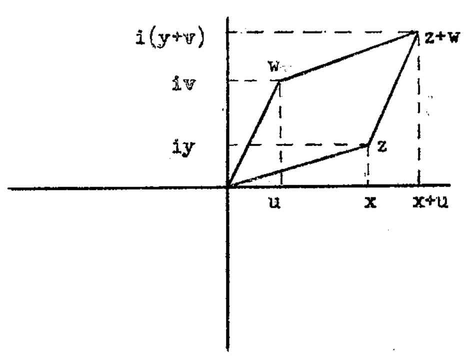

Figure 5.2.

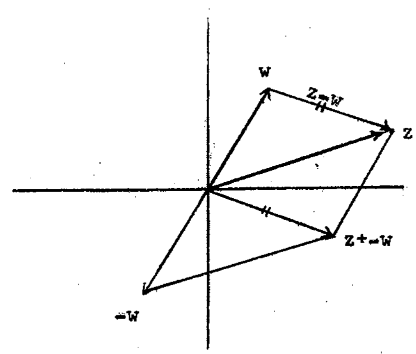

Figure 5.3.

After observing that $z - w = z + -w$ the construction of $z - w$ can be accomplished using the method outlined for addition. Vectorially, $z - w$ can be represented by a vector from the point w to the point z. See Figure 5.3.

The construction for the product and quotient of the complex numbers is somewhat more complicated than that for addition or subtraction. To expedite the discussion of these it is desirable to consider an alternate representation of complex numbers. The reader is hopefully acquainted with plane polar coordinates, thus recognizing that if $z = x + iy$, then $x = r \cos \theta$ and $y = r \sin \theta$, where $r = \sqrt{x^2 + y^2} = |z|$ and $\tan \theta = y/x$. With this in mind $z = r(\cos \theta + i \sin \theta)$. The expression $r(\cos \theta + i \sin \theta)$ is called the polar form of the complex number z. The angle θ (determined only up to multiples of 2π) is referred to as the argument of z, or briefly $\theta = \arg z$. The relationship between x,y,r,θ is depicted in Figure 5.4.

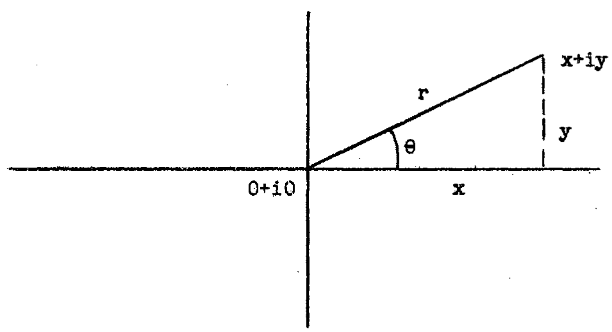

Figure 5.4

If $z = r(\cos \theta + i \sin \theta)$, $-\pi \leq \theta < \pi$, then θ is often denoted Arg z and is called the principal argument of z. Observe that for $r > 0$ and $-\pi \leq \theta < \pi$, every complex number determines a unique $r = \sqrt{x^2 + y^2}$ and a unique $\theta = \cos^{-1} \frac{x}{r} = \sin^{-1} \frac{y}{r}$. In case $z = 0$, then $r = 0$ and θ is arbitrary. Note also, that if $z_1 = r_1(\cos \theta_1 + i \sin \theta_1)$ and $z_2 = r_2(\cos \theta_2 + i \sin \theta_2)$, then $z_1 = z_2$ if and only if $r_1 = r_2$ and $\theta_1 = \theta_2$, where equality of angles is up to multiples of 2π. It will be apparent that equality is used in this sense in the sequel.

A third form is frequently used for expressing complex numbers. If $z = r(\cos \theta + i \sin \theta)$, then $z = re^{i\theta}$, where $e^{i\theta} = \cos \theta + i \sin \theta$.

The notation $e^{i\theta}$ for $\cos\theta + i\sin\theta$ can be justified, but the discussion is beyond the scope of this paper. The expression $re^{i\theta}$ is referred to as the exponential form of the complex number z. In case $r = 1$, z is called a turn. The letters s and t will be used to represent turns in the remainder of this work.

To expose the relationship between the algebraic operation $z \cdot w$ and its geometric interpretation consider the product in polar form. Suppose $z = r_1(\cos\theta + i\sin\theta)$ and $w = r_2(\cos\phi + i\sin\phi)$, where $|z| = r_1$, $|w| = r_2$, $\theta = \arg z$ and $\phi = \arg w$, then

$$z \cdot w = r_1 r_2\{(\cos\theta\cos\phi - \sin\theta\sin\phi) + i(\sin\theta\cos\phi + \cos\theta\sin\phi)\}.$$

From trigonometry it follows that

$$\cos\theta\cos\phi - \sin\theta\sin\phi = \cos(\theta + \phi)$$

and

$$\sin\theta\cos\phi + \cos\theta\sin\phi = \sin(\theta + \phi).$$

Therefore, $z \cdot w = r_1 r_2\{\cos(\theta + \phi) + i\sin(\theta + \phi)\}$, or alternately $z \cdot w = r_1 r_2 e^{i(\theta + \phi)}$. From this it is clear that $\arg z \cdot w = \arg z + \arg w$, where it is understood that the equality is valid within a multiple of 2π. The following theorem is a formal summary of the foregoing discussion.

Theorem 5.1. If $z, w \in C$, $z = r_1(\cos\theta + i\sin\theta)$, $w = r_2(\cos\phi + i\sin\phi)$, then $z \cdot w = r_1 r_2\{\cos(\theta + \phi) + i\sin(\theta + \phi)\}$. In exponential form $z \cdot w = r_1 r_2 e^{i(\theta + \phi)}$.

Geometrically the length of the vector $z \cdot w$ is equal to the products of the lengths of z and w. The angle between the directed segment $z \cdot w$ and the positive real axis is the sum of the angles arg z and arg w. Figure 5.5 illustrates the situation.

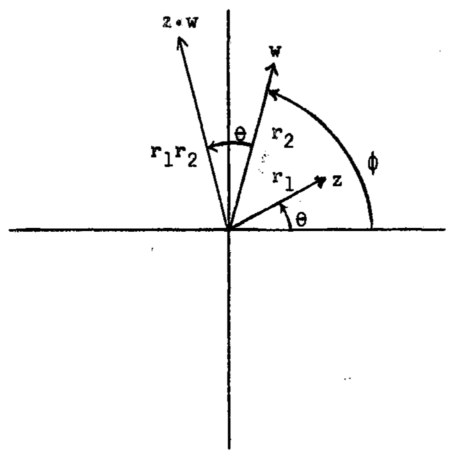

Figure 5.5.

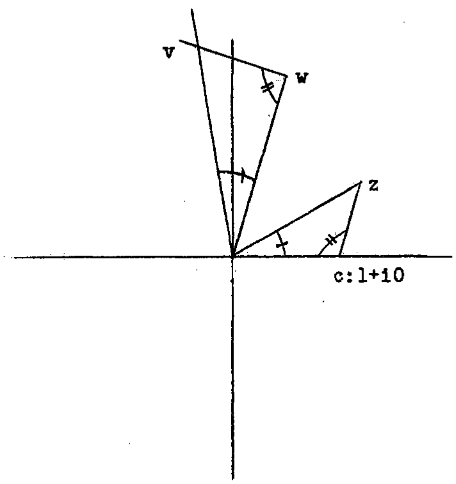

Figure 5.6.

In addition to the preceding graphic relationship between factors and their product it is of interest to note that multiplication can be performed by purely geometric means. In particular, if z and w are arbitrary points of the plane and c is the point corresponding to $1 + i0$, then the point corresponding to $z \cdot w$ is the third vertex of the triangle Owv which is directly similar to triangle Ocz. The essence of the construction is suggested by Figure 5.6.

The following special products merit some attention. If z is an arbitrary complex number and t is a turn, then the point corresponding to zt can be obtained by rotating z about O through arg t. In case z is arbitrary and r is a real complex number, then the point associated with rz lies on the ray Oz at a distance $r|z|$ from O. In a vector setting the relationship would be one of rz being a scalar multiple of z.

Since division is the inverse of multiplication the problem of determining v so that $v = \frac{z}{w}$, $w \neq 0$, is equivalent to finding v so that $z = vw$. Thus, if $z = r_1e^{i\theta_1}$, $w = r_2e^{i\theta_2}$, $v = re^{i\theta}$, then r and θ must be such that $r_1 e^{i\theta_1} = re^{i\theta} \cdot r_2 e^{i\theta_2} = rr_2 e^{i(\theta + \theta_2)}$. However, this implies that $r_1 = rr_2$ and $\theta_1 = \theta + \theta_2$. Since $w \neq 0$, then $r_2 > 0$ and it follows that $r = r_1/r_2$ and $\theta = \theta_1 - \theta_2$. The following theorem provides a concise statement of this result.

Theorem 5.2. If $z,w \in C$, $z = r_1(\cos\theta_1 + i\sin\theta_1)$, $w = r_2(\cos\theta_2 + i\sin\theta_2) \neq 0$, then $\frac{z}{w} = \frac{r_1}{r_2}\{\cos(\theta_1 - \theta_2) + i\sin(\theta_1 - \theta_2)\}$. In exponential form $\frac{z}{w} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$.

Corollary 5.2. If t is a turn, then $\bar{t} = \frac{1}{t} = t^{-1}$.

Geometrically the length of the vector z/w is the quotient of the lengths of the vectors z and w respectively. The inclination of vector z/w to the positive real axis is arg z - arg w. Of course the quotient can be constructed by essentially inverting the process of multiplication. These purely geometric means of determining the product and quotient will play no role in the sequel. The graphical relationship between the divisor, dividend, and quotient will prove a valuable intuitive guide and is illustrated in Figure 5.7.

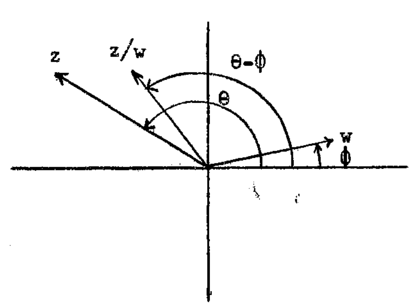

Figure 5.7

Consider T 5.1. in the case where $z = w$. If $z = r(\cos\theta + i\sin\theta) = w$, then $z \cdot w = z^2 = r^2(\cos 2\theta + i\sin 2\theta)$. This suggests the following theorem, which is generally referred to as DeMoivres theorem. The proof is within the grasp of the reader acquainted with mathematical induction arguments. [24]

Theorem 5.3.A. If n is any integer and $z = \cos\theta + i\sin\theta = e^{i\theta}$, then $z^n = (\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta$. In exponential form, $z^n = e^{in\theta}$.

The significance of DeMoivres Theorem becomes apparent in the following result. This is essentially a corollary of T 5.3.A.

Theorem 5.4.B. If $a = r(\cos\theta + i\sin\theta)$ and n is a positive integer, then the numbers $r^{1/n}(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n})$, $k = 0,1,2,3,\ldots,n-1$ are the roots of the equation $z^n = a$. I.e., the nth roots of a. These numbers are distinct if $a \neq 0$.

Geometrically it is clear from our identification of complex numbers with points of the plane that T 5.4.B indicates that the nth roots of $a \neq 0$ are represented by n points spaced equally around a circle of radius $|a|^{1/n}$. See Figure 5.8. for the case where $n = 5$.

Observe that if $a \in C$, $a \neq 0$, then there exist exactly two elements $z,w \in C$ such that $z^2 = w^2 = a$. Furthermore, it follows from T 5.4.B. that precisely one of these numbers will be such that $0 \leq \text{Arg } z < \pi$. The symbol $\sqrt{a}$ will henceforth be used to refer to this root of a. In keeping with this convention $-\sqrt{a}$ will be used to denote the square root of a having an argument equal to $\text{Arg}\sqrt{a} - \pi$.

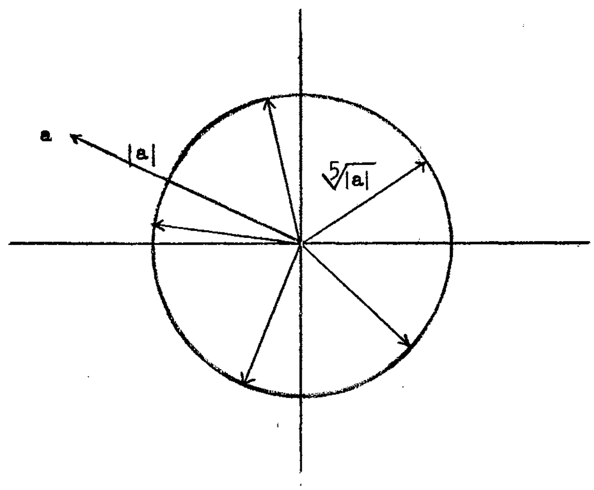

Figure 5.8.

At this point, having considered the geometric interpretation of the fundamental operations on complex numbers, attention is focused on certain linear aspects of the analytic geometry of the Argand plane. The point of departure for such a development is a recognition of the relationship between the cartesian coordinates of a point and the complex number identified with the point. Specifically, note that if (x,y) and $z = x + iy$ are the respective labels for a point of the plane, then

$$x = \frac{z + \bar{z}}{2} \text{ and } y = \frac{z - \bar{z}}{2i} .$$

With this in mind the initial step in the direction of the complex analytic geometry of the line is the following theorem.

Theorem 5.5. The general equation of a line in the Argand plane is of the form $\bar{\alpha}z + \alpha\bar{z} + \beta = 0$, where $\alpha \neq 0$ and β is real. This line contains the point $\dfrac{-\beta\alpha}{2|\alpha|^2}$.

Proof. The general equation of a line in Cartesian coordinates is $ax + by + c = 0$, where a,b and c are real and $a^2 + b^2 \neq 0$. Using the aforementioned relationship between ordinary rectangular coordinates and the associated complex number, one gets upon substitution $a(\frac{z + \overline{z}}{2}) + b(\frac{z - \overline{z}}{2i}) + c = 0$, or $a(z + \overline{z}) + (-i)\frac{b(z - \overline{z})}{(-i)(i)} + 2c = 0$. The preceding can be written $(a - ib)z + (a + ib)\overline{z} + 2c = 0$. Letting $\alpha = a + ib$, $\beta = 2c$, this becomes the indicated equation. Thus, if ℓ is a line in the plane ℓ has an equation of the form $\overline{\alpha}z + \alpha\overline{z} + \beta = 0$, where $\alpha \neq 0$ and β is real. Conversely, it is not difficult to see that an equation of this form can be written in the form $ax + by + c = 0$, where a,b,c are real and $a^2 + b^2 \neq 0$. Thus, the corresponding locus is a line. That $\frac{-\beta\alpha}{2|\alpha|^2}$ is on the line follows by substitution.

Corollary 5.5. Every line in the complex plane has an equation of the form $z - a\overline{z} - b = 0$, where $a,b \in C$, $|a| = 1$.

Often in this work no distinction will be made between a certain locus of points and the corresponding equation. For example, an equation of the form $\overline{\alpha}z + \alpha\overline{z} + \beta = 0$, $\alpha \neq 0$, β real, will be referred to, on occasion, as a line.

The following result establishes the analytic condition for perpendicularity of lines in the Argand plane.

Theorem 5.6. If $\ell_1: \overline{\alpha}z + \alpha\overline{z} + r = 0$ and $\ell_2: \overline{\beta}z + \beta\overline{z} + p = 0$ are lines in the complex plane, then ℓ_1 is perpendicular to ℓ_2 if and only if $\alpha\overline{\beta} + \overline{\alpha}\beta = 0$.

Proof. It is clear from the demonstration of T 5.5. that if

$a_1x + b_1y + c_1 = 0$ and $a_2x + b_2y + c_2 = 0$ are the corresponding Cartesian coordinate equations of ℓ_1 and ℓ_2 respectively, then $\alpha = a_1 + ib_1$ and $\beta = a_2 + ib_2$. For $b_1 \neq 0 \neq a_2$, $\ell_1 \perp \ell_2$ if and only if $-\frac{a_1}{b_1} = \frac{b_2}{a_2}$, or equivalently if and only if $a_1a_2 + b_1b_2 = 0$. Now observe that $\alpha\bar{\beta} + \bar{\alpha}\beta = (a_1 + ib_1)(a_2 - ib_2) + (a_1 - ib_1)(a_2 + ib_2)$ $= 2(a_1a_2 + b_1b_2)$. Thus, $\alpha\bar{\beta} + \bar{\alpha}\beta = 0$ if and only if $a_1a_2 + b_1b_2 = 0$, or alternately if and only if $\ell_1 \perp \ell_2$. A similar discussion disposes of the case where $b_1 = 0$ or $a_2 = 0$.

Corollary 5.6. If $\ell_1 : z - m\bar{z} - p = 0$ and $\ell_2 : z - n\bar{z} - q = 0$ are lines, then $\ell_1 \perp \ell_2$ if and only if $m = -n$.

The following definition extends the notion of perpendicularity to vectors in the natural way.

Definition 5.1. If $a, b \in C$, $a \neq 0 \neq b$, then vector a is perpendicular to vector b if and only if ℓ_1 is perpendicular to ℓ_2, where ℓ_1, ℓ_2 are the lines passing through the origin containing a and b respectively. A nonzero vector a is perpendicular to a line ℓ if and only if the line determined by a and 0 is perpendicular to ℓ.

The ensuing analytic characterizations of the above notions prove useful.

Theorem 5.7. If $a, b \in C$, $a \neq 0 \neq b$, then vector a is perpendicular to vector b if and only if $a\bar{b} + \bar{a}b = 0$.

Proof. Observe that 0 and a are on $\ell_1 : \overline{ia}z + ia\bar{z} = 0$ and 0 and b are on

$\ell_2: \overline{ib}z + ib\overline{z} = 0$. Thus, $a \perp b$ if and only if $\ell_1 \perp \ell_2$, or alternately if and only if $(ia)(\overline{ib}) + (\overline{ia})(ib) = 0$. However, $(ia)(\overline{ib}) + (\overline{ia})(ib) = a\overline{b} + \overline{a}b$ and the conclusion follows.

Corollary 5.7. If $a, b \in C$ and $a \perp b$, then $r_1 a \perp r_2 b$, for every nonzero real choice of r_1 and r_2.

Theorem 5.8. If $c \in C$, $c \neq 0$, and $\ell: \overline{a}z + a\overline{z} + b = 0$ is a line, then ℓ is perpendicular to c if and only if $a\overline{c} - \overline{a}c = 0$.

Proof. Since 0 and c are on $\ell: \overline{ic}z + ic\overline{z} = 0$, then $c \perp \ell$ if and only if $a(\overline{ic}) + \overline{a}(ic) = 0$. But $a(\overline{ic}) + \overline{a}(ic) = -i(a\overline{c} - \overline{a}c)$, hence $c \perp \ell$ is equivalent to $a\overline{c} - \overline{a}c = 0$.

Corollary 5.8. If $\ell: \overline{a}z + a\overline{z} + b = 0$ is a line, then $a \perp \ell$.

The reader will note that C 5.8. together with T 5.5. indicates that $\ell: \overline{a}z + a\overline{z} + b = 0$ is the line perpendicular to a and at a vector distance $\left(\frac{-b}{2|a|^2}\right)\cdot a$ from 0. In view of this it is not difficult to see that if $\ell_1: \overline{a}z + a\overline{z} + b = 0$ and $\ell_2: \overline{a}z + a\overline{z} + c = 0$, then the directed distance from ℓ_1 to ℓ_2 is $-\left(\frac{c - b}{2|a|^2}\right)a$. In particular, if $|a| = 1$, then the vector distance from ℓ_1 to ℓ_2 is $-\left(\frac{c - b}{2}\right)a$. Figure 5.9 illustrates the situation where a is a turn.

Having developed a set of necessary and sufficient analytic conditions for perpendicularity in the Argand plane, attention is now centered on the analogous results for parallelism. The proofs of these theorems are omitted, since they generally parallel the demonstrations of the preceding propositions. The reader will find it instructive to fill in the details.

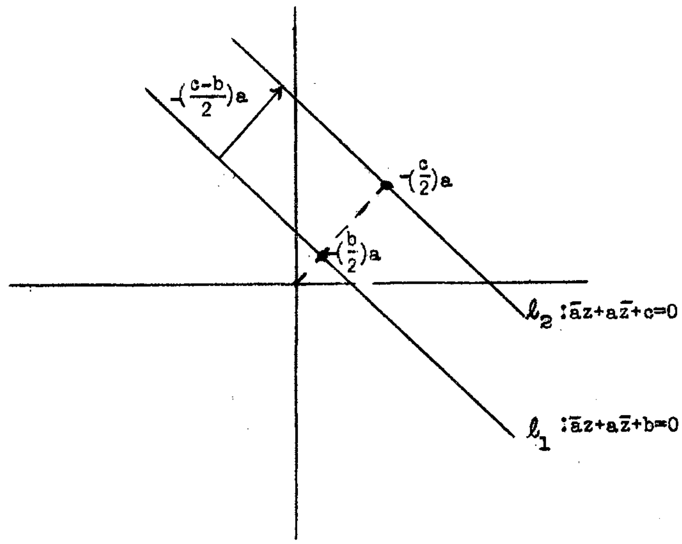

Figure 5.9.

Theorem 5.9.C. If $\ell_1:\bar{a}z + a\bar{z} + b = 0$ and $\ell_2:\bar{c}z + c\bar{z} + d = 0$ are lines in the complex plane, then ℓ_1 is parallel to ℓ_2 if and only if $a\bar{c} - \bar{a}c = 0$.

Corollary 5.9. If the equations of lines ℓ_1 and ℓ_2 respectively are written in the form $z - m\bar{z} - p = 0$ and $z - n\bar{z} - q = 0$, then $\ell_1 \,||\, \ell_2$ if and only if $m = n$.

Parallelism is extended to vectors in the following definition. The reader should note that this characterization of parallel vectors is not the usual one. This anomalous definition causes no difficulty in this work and is expedient because it leads to a particularly simple analytic characterization of parallelism.

Definition 5.2. If $a,b \in C$, $a \neq 0 \neq b$, then vector a is parallel to vector b if and only if 0, a, and b are collinear. A nonzero vector a is parallel to a line ℓ if and only if the line containing 0 and a is parallel to ℓ.

Theorem 5.10.D. If $a,b \in C$, $a \neq 0 \neq b$, then in order that vector a be parallel to vector b it is necessary and sufficient that $a\bar{b} - \bar{a}b = 0$.

Corollary 5.10. If $a,b \in C$ and $a \,||\, b$, then $r_1 a \,||\, r_2 b$ for every nonzero real r_1, r_2.

Theorem 5.11.E. If $a \in C$, $a \neq 0$ and $\ell: \bar{c}z + c\bar{z} + d = 0$ is a line, then a is parallel to ℓ if and only if $a\bar{c} + \bar{a}c = 0$.

Corollary 5.11. If $\ell: \bar{c}z + c\bar{z} + d = 0$ is a line, then $ic \,||\, \ell$.

Observe that in view of T 5.2. and T 5.11.E. it follows that if $\ell_1: \bar{a}z + a\bar{z} + b = 0$ and $\ell_2: \bar{c}z + c\bar{z} + d = 0$ are lines, then the directed angles between ℓ_2 and ℓ_1 (in that order) are given by Arg $\frac{ia}{ic}$ and Arg $\frac{-ia}{ic}$. Since Arg $\frac{ia}{ic}$ = Arg $\frac{a}{c}$ = Arg $a\bar{c}$, these angles are alternately denoted Arg $a\bar{c}$ and Arg$(-a\bar{c})$. Figure 5.10. illustrates the situation for lines through the origin.

There is one remaining result of a linear nature that provides some insight into the work in Chapter VI. The proposition involved is contingent on the concept of projection. To introduce the notion in a complex setting consider a nonzero vector b and a line ℓ parallel to the vector c. Without loss of generality one can suppose $|c| = 1$. See Figure 5.11. In the traditional sense of the word the projection

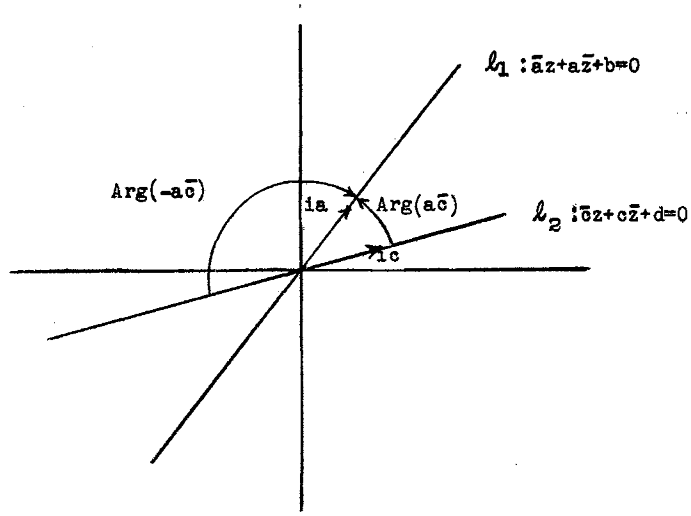

Figure 5.10.

of b on ℓ would be $|b \cos \theta|$, where θ = Arg c - Arg b. In a vector treatment of geometry $(|b| \cos \theta)c$ would be the projection of b on ℓ. Interestingly enough the foregoing can be expressed somewhat more elegantly in our complex setting.

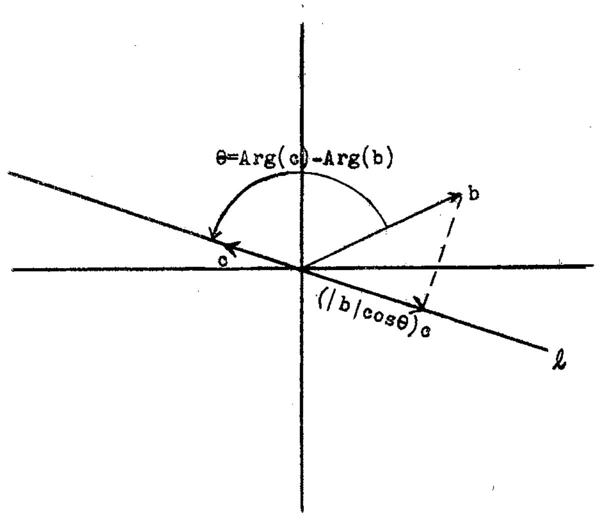

Figure 5.11.

Theorem 5.12. If $b, c \in C$, $b \neq 0$, $|c| = 1$, and ℓ is a line parallel to c, then the vector projection of b on ℓ is

$$\frac{c^2\bar{b} + b}{2}.$$

Proof. Let $\theta = \text{Arg } c$ and $\phi = \text{Arg } b$, then $c = \cos\theta + i\sin\theta = e^{i\theta}$. Thus, the vector projection of b on ℓ is $[|b| \cos(\theta - \phi)]c =$ $|b|\{\cos(\theta - \phi) \cos\theta + i \cos(\theta - \phi) \sin\theta\}$. Using the appropriate trigonometric product identities, the preceding can be written

$$\frac{|b|}{2}\{[\cos\phi + \cos(2\theta - \phi)] + i[\sin(2\theta - \phi) - \sin(-\phi)]\}$$

$$= \frac{|b|}{2}\{[\cos(2\theta - \phi) + i\sin(2\theta - \phi)] + [\cos\phi + i\sin\phi]\}$$

$$= \frac{|b|}{2}\{e^{i(2\theta - \phi)} + e^{i\phi}\}$$

$$= \frac{|b|}{2}\{e^{i2\theta} \cdot e^{-i\phi} + e^{i\phi}\}$$

$$= \frac{|b|}{2}\{c^2 \cdot \frac{\bar{b}}{|b|} + \frac{b}{|b|}\}$$

$$= \frac{c^2\bar{b} + b}{2}.$$

Although the foregoing result is not essential to the proof of any of the theorems in Chapter VI the reader will find it invaluable when seeking a geometric interpretation of certain propositions in that section.

CHAPTER VI

THE ISOMETRIES OF THE ARGAND PLANE

The current chapter is directed toward a systematic analytic development of the Euclidean transformations. The significance of these mappings in plane geometry is widely recognized. The intent here is not to dwell on the geometric aspects of these transformations, rather to focus on the problem of algebraically developing the relationships between them in a complex setting. The treatment is not exhaustive in this regard. In particular, attention is given to those results which lend themselves to a logical exposition of the fundamental nature of reflections.

The informed reader will recognize that few of the propositions in this chapter are truly original. However, the literature suggests that these results have been given only cursory attention in the setting in which they appear here. [3][13] It is the writers contention that the elegance afforded by a complex analytic treatment justifies their inclusion. In addition to the foregoing, the well versed reader will observe that the notion of a group could have been utilized to unify certain aspects of the discussion. This notion was not introduced in an effort to keep to a minimum the number of concepts marginally related to the central theme. Finally, although the allied geometry is given little attention in this paper, the reader will find it instructive to interpret the various propositions in this chapter geometrically.

The formal development is initiated by giving careful attention to the technical terms used rather glibly in the preceding paragraphs.

Definition 6.1. A function $f:A \to B$ is a transformation if and only if f is one to one and onto.

In this paper the only transformations of interest will be those having the complex field as domain and range. These will be referred to as complex transformations. The reader will note that if f and g are complex transformations, then the composition, $f \circ g$, is a complex transformation. If $h = f \circ g$, then h will be referred to as the product of f and g, or variously f and g will be called factors of h. In addition to the foregoing, it is clear that if f,g and h are complex transformations, then $(f \circ g) \circ h$ is well defined and $(f \circ g) \circ h = f \circ (g \circ h)$. Furthermore, since f is both one to one and onto C, then f^{-1} exists and is itself a complex transformation. Finally, it is not difficult to show that $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.

In general the transformations of interest are those which have an invariant feature. In particular, attention here is focused on those complex transformations that preserve distance in the absolute value sense. The following definition presents the concept formally.

Definition 6.2. If $f:C \to C$, then f is an isometry if and only if $|f(z_1) - f(z_2)| = |z_1 - z_2|$, for every $z_1, z_2 \in C$.

The following two results are basic to any discussion of isometries. The proof of each is almost immediate.

Theorem 6.1. If f and g are isometries, then $f \circ g$ is also an isometry.

Proof. $f \cdot g$ is a well defined complex transformation since f and g are complex transformations. Furthermore, as a result of the fact that f and g are isometries

$|f \cdot g(z_1) - f \cdot g(z_2)| = |f[g(z_1)] - f[g(z_2)]| = |g(z_1) - g(x_2)| = |z_1 - z_2|$

for every $z_1, z_2 \in C$.

Theorem 6.2. If f is an isometry, then f^{-1} is an isometry.

Proof. The details are omitted.

In light of the descriptive nature of definition 6.2. it is reasonable to seek a constructive characterization of the isometry concept. The following theorem provides this.

Theorem 6.3. $f: C \to C$ is an isometry if and only if $f(z) = az + b$ or $f(z) = a\overline{z} + b$, where $a,b \in C$ and $|a| = 1$.

Proof. The sufficiency is not difficult. If $f(z) = az + b$, where $a,b \in C$, $|a| = 1$, then it follows readily that f is a one to one mapping of C onto C. Furthermore, $|f(z_1) - f(z_2)| = |(az_1 + b) - (az_2 + b)|$ $= |a(z_1 - z_2)| = |a||z_1 - z_2|$. Since a is a turn the distance preserving quality of f is apparent. A similar discussion disposes of the case where $f(z) = a\overline{z} + b$, $a,b \in C$, $|a| = 1$.

Now, suppose f is an isometry, then $|f(z) - f(1)| = |z - 1|$, for every $z \in C$. This implies the identity $|f(z) - f(1)|^2 = |z - 1|^2$. Using T 3.20.E. it follows that $[f(z) - f(1)][\overline{f(z)} - \overline{f(1)}] =$ $[z - 1][\overline{z} - 1]$, or $f(z)\overline{f(z)} - f(z)\overline{f(1)} - \overline{f(z)}f(1) + f(1)\overline{f(1)}$ $= z\overline{z} - z - \overline{z} + 1$, or

$$(1) \quad |f(z)|^2 - f(z)\overline{f(1)} - \overline{f(z)}f(1) + |f(1)|^2 = |z|^2 - z - \overline{z} + 1.$$

In the event that $f(0) = 0$, then $|f(z)| = |z|$, or a alternately $|f(z)|^2 = |z|^2$. Thus, when the origin is preserved under f equation (1) can be written $-f(z)\overline{f(1)} - \overline{f(z)}f(1) = -z - \overline{z}$, or

(2) $f(z)\overline{f(1)} + \overline{f(z)}f(1) = z + \overline{z}$.

Since $f(1)\overline{f(z)} = \overline{\overline{f(1)}f(z)}$, then (2) coupled with T 3.17.D. implies that

(3) $\text{Re}[\overline{f(1)}f(z)] = \text{Re}(z)$, for every $z \in C$.

Also note, however, that under the assumption that $f(0) = 0$ it follows that $|\overline{f(1)}f(z)|^2 = |\overline{f(1)}|^2|f(z)|^2 = |f(z)|^2 = |z|^2$. The fact that $|\overline{f(1)}f(z)|^2 = |z|^2$ is equivalent to the statement that

(4) $[\text{Re}(\overline{f(1)}f(z))]^2 + [\text{Im}(\overline{f(1)}f(z))]^2 = [\text{Re}(z)]^2 + [\text{Im}(z)]^2$.

Utilizing the previous observation that $\text{Re}(\overline{f(1)}f(z)) = \text{Re}(z)$, then (4) can be used to assert that $[\text{Im}(\overline{f(1)}f(z))]^2 = [\text{Im}(z)]^2$, or

(5) $\text{Im}(\overline{f(1)}f(z)) = \pm \text{Im}(z)$.

Identities (3) and (5) combined imply that

(6) $\overline{f(1)}f(z) = z$ or $\overline{f(1)}f(z) = \overline{z}$.

As previously noted, under the assumption that the origin is mapped onto itself it follows that $\overline{f(1)}$ is a turn. Thus, $[\overline{f(1)}]^{-1} = \overline{\overline{f(1)}}$ $= f(1)$, by C.5.2. Using this fact the equalities (6) can be written

(7) $f(z) = f(1)z$ or $f(z) = f(1)\overline{z}$.

Hence if $f(0) = 0$, the result is apparent.

Now, suppose that $f:C \to C$ is a distance preserving transformation and $f(0) \neq 0$. Consider the function $g:C \to C$ such that $g(z) = z - f(0)$. g is an isometry from the sufficiency argument, hence $g \circ f(z) =$ $f(z) - f(0)$ is an isometry by T 6.1. Clearly $g \circ f(0) = 0$. Thus, it follows that $g \cdot f(z) = [g \cdot f(1)]z$, or $g \cdot f(z) = [g \cdot f(1)]\overline{z}$. But, $g \cdot f(z) = f(z) - f(0)$. Therefore,

$$f(z) - f(0) = [f(1) - f(0)]z \text{ or } f(z) - f(0) = [f(1) - f(0)]\overline{z}.$$

After adding $f(0)$ to each member of the foregoing equations and observing that $|f(1) - f(0)| = |1 - 0| = 1$, the conclusion follows.

In view of the geometric interpretations of addition and multiplication as outlined in Chapter V it is not too surprising that the isometries take the two general forms exhibited in T 6.3. Of course, the reader has observed that it requires some algebraic finesse to establish a result that, at least in retrospect, is geometrically apparent.

Theorem 6.3. suggests an initial classification of isometries according to the form of the functional relationship. This turns out to be appropriate and the following definition is reasonably well established.

Definition 6.3. A function $f: C \to C$ of the form $f(z) = tz + b$, where $t, b \in C$ is called a direct isometry. Any isometry that is not direct is opposite.

The above terminology has its roots in the elusive concept of orientation. A mathematically exact description of this notion is not appropriate here.

The reader is perhaps familiar with the following expression.

Definition 6.4. If f is an isometry and $z \in C$ such that $f(z) = z$, then z is called an invariant point under f. There is an isometry under which every point is invariant, namely $f(z) = z$. This function will be referred to as the identity, or variously the trivial isometry.

It turns out that an examination of invariant points leads to an appropriate subclassification of direct isometries. The following simple result exposes the conditions under which a direct isometry has an invariant point.

Theorem 6.4. If $f(z) = tz + b$ is a non-trivial isometry, then f has an invariant point if and only if $t \neq 1$. For a nonidentical direct isometry there is at most one such point, namely $\frac{b}{1 - t}$.

Proof. A point z is preserved under f if and only if $z = tz + b$. However, $z = tz + b$ is equivalent to $(1 - t)z = b$, and where $b \neq 0$ this has a solution if and only if $1 - t \neq 0$. It is clear that the number $\frac{b}{1 - t}$ is the only root when a solution exists.

Geometric considerations suggest the appropriateness of the following terminology.

Definition 6.5. A distance preserving transformation of the form $f(z) = z + b$ is called a translation through b, or simply a translation.

It should be noted that, aside from the identity transformation, the translations are precisely those direct isometries that have no invariant points. In addition, it is not difficult to see that if f is a translation through b, then $f^{-1}(z) = z - b$. Furthermore, it is almost immediate that the product of two translations is a translation. On occasion, the suggestive notation $T(b)$ will be used to refer to the function $f(z) = z + b$. Using this notation it is apparent that $[T(b)]^{-1} = T(-b)$ and $T(a) \cdot T(b) = T(b) \cdot T(a) = T(a + b)$. The latter result is stated as a theorem for reference pusposes.

Theorem 6.5. If f and g are translations through a and b respectively, then $f \circ g$ is a translation through $a + b$ and $f \circ g = g \circ f$.

In the event that a function of the form $f(z) = tz + b$ has a single fixed point $c = \frac{b}{1 - t}$, a little algebra can be used to verify that $f(z) = t(z - c) + c$. It is not difficult to see that such a representation of f is unique. In this paper the foregoing form of such a transformation will be called canonical. The canonical form of a direct isometry having exactly one invariant point indicates that such a function is the product of $T(-c)$, $g(z) = tz$, and $T(c)$. Since $t \neq 1$ is a turn, the geometric interpretation of multiplication given in Chapter V suggests that g might be termed a rotation about the origin. The foregoing observation and the fact that $f = T(c) \circ g \circ T(-c)$ indicates that f might appropriately be called a rotation about the point c. This is in keeping with the following definition.

Definition 6.6. A direct isometry f is a rotation if and only if f has an invariant point. In case $f(z) = tz + b$ has exactly one fixed point c, f will be called a rotation of arg t about c.

The notation $R(t,c)$ will be used to denote the rotation of arg t about c. With this convention the canonical form of a rotation f, of arg t about c, becomes $f = T(c) \circ R(t,0) \circ T(-c)$. Conversely, it is clear that every product of the form $T(c) \circ R(t,0) \circ T(-c)$ is a rotation of arg t about c. These observations are summarized in the next theorem.

Theorem 6.6. A nontrivial direct isometry f is a rotation of arg t about c if and only if $f = T(c) \cdot R(t,0) \cdot T(-c)$.

Observe that the identity transformation is the only direct isometry that is both a rotation and translation. Furthermore, a little computation reveals that if $f(z) = tz + b$ is a rotation, then $f^{-1}(z) = \bar{t}z - \bar{t}b$. Additional computation shows that the canonical form of f^{-1}, for a nonidentical rotation f, is $T(c) \cdot R(\bar{t},0) \cdot T(-c)$, where $c = \frac{b}{1 - t}$. Thus, if f is a rotation of arg t about c, f^{-1} is a rotation of -arg t about c.

The following special type of rotation merits individual attention.

Definition 6.7. A rotation of the form $f(z) = -z + 2b$ is called a half turn about b.

In keeping with the earlier convention a half turn about b can be written $R(-1,b)$. On occasion it will prove more suggestive to write $H(b)$ in place of $R(-1,b)$.

The following result establishes an interesting relationship between translations and half turns.

Theorem 6.7. If f and g are half turns about a and b respectively, then $f \cdot g$ is a translation through $2(a - b)$. Symbolically $H(a) \cdot H(b) = T(2[a - b])$.

Proof. $f(z) = -z + 2a$, $g(z) = -z + 2b$, hence $f \cdot g(z) = -(-z + 2b) + 2a = z + 2(a - b)$. Therefore, $f \cdot g = H(a) \cdot H(b) = T(2[a - b])$.

The foregoing theorem suggests the possibility of factoring any translation into the product of two half turns. This can be done and it is easy to see that such a decomposition is not unique. The next proposition, which is essentially a corollary of T 6.7., establishes the nature of such a factorization.

Theorem 6.8. If $f(z) = z + a$, then f can be factored into the product of two half turns. Specifically, $T(a) = H(b) \cdot H(c)$, where $b = c + \frac{a}{2}$.

Proof. Apply T 6.7.

The reader has perhaps observed that the set of direct isometries is closed under composition. With this in mind it is clear that product of two rotations is either a rotation or translation. Theorem 6.7. indicates that the set of rotations is not closed under composition, although the set of translations is (T 6.5.). The conditions under which the product of two rotations is again a rotation are exhibited in the following proposition.

Theorem 6.9. If $f(z) = tz + a$ and $q(z) = sz + b$ are rotations, then $f \cdot g$ is a translation or nontrivial rotation according to whether $t = \overline{s}$ or not.

Proof. $f \cdot g(z) = t(sz + b) + a = (ts)z + (tb + a)$. Thus, $f \cdot g$ is a translation if and only if $ts = 1$. However, $ts = 1$ is equivalent to $t = \overline{s}$ as a result of the fact that s is a turn. Since the product is either a translation or nontrivial rotation this completes the demonstration.

The preceding result suggests the possibility of generalizing T 6.8. This is possible, but there will be no reason to state this formally. The reader will do well to consider the geometry associated with the two alternatives in T 6.9.

Attention is now centered on the opposite, or indirect, distance preserving complex transformations. An initial observation in this direction is that the composition of two opposite isometries is a direct isometry. In particular, if $f(z) = t\overline{z} + a$, $g(z) = s\overline{z} + b$, then $f \cdot g(z) = t(\overline{s\overline{z} + b}) + a = (t\overline{s})z + t\overline{b} + a$. Again, it is an examination of invariant points that leads to the appropriate classification of opposite isometries.

Theorem 6.10. If $f(z) = t\overline{z} + a$, then a necessary and sufficient condition for the existence of an invariant point under f is that $t\overline{a} + a = 0$.

Proof. In case z is an invariant point under f, then $z = t\overline{z} + a$. However, $z = t\overline{z} + a$ implies that $\overline{z} = \overline{t}z + \overline{a}$. Substituting $\overline{t}z + \overline{a}$ for $\overline{z}$ in the former equation yields $z = t(\overline{t}z + \overline{a}) + a$, or $z = (t\overline{t})z + t\overline{a} + a = z + t\overline{a} + a$. From this it follows that $t\overline{a} + a = 0$ and the necessity is established.

To see that the condition is sufficient observe that if $t\overline{a} + a = 0$, then $\frac{a}{2} = \frac{a}{2} + 0 = \frac{a}{2} + \frac{t\overline{a} + a}{2} = \frac{t\overline{a}}{2} + a = f(\frac{a}{2})$. In other words $\frac{a}{2}$ is an invariant point under f.

Although the foregoing theorem exposes a simple condition which characterizes point invariance under an opposite isometry, it is not too geometrically revealing. The following result sheds some light in this direction.

Theorem 6.11. If $f(z) = t\overline{z} + a$, $a \neq 0$, then $t\overline{a} + a = 0$ if and only if a is perpendicular to $\ell: \overline{\sqrt{-t}}\, z + \sqrt{-t}\, \bar{z} = 0$. When f has an invariant point, then the set of all such points is the line $m: \overline{\sqrt{-t}}\, z + \sqrt{-t}\, \bar{z} - \overline{\sqrt{-t}}\, a = 0$.

Proof. $t\bar{a} + a = 0$ implies that $-t\bar{a} - a = 0$, or alternately,

(1) $\sqrt{-t}\sqrt{-t}\, \bar{a} - a = 0$.

Since t is a turn it follows that -t and $\sqrt{-t}$ are also turns. Thus, $(\sqrt{-t})^{-1} = \overline{\sqrt{-t}}$. Multiplying (1) by $\overline{\sqrt{-t}}$ and applying C 5.2. yeilds

(2) $\sqrt{-t}\, \bar{a} - \overline{\sqrt{-t}}\, a = 0$.

But, the preceding equality is precisely the condition required for the perpendicularity of a and ℓ. The converse follows by essentially reversing the foregoing steps.

When $t\bar{a} + a = 0$, then $\{z \mid z = t\bar{z} + a\}$ is the set of invariant points. However, $z = t\bar{z} + a$ can be written

(3) $z - t\bar{z} - a = 0$.

Again using the fact that $\sqrt{-t}$ is a turn and multiplying both sides of (3) by $\overline{\sqrt{-t}}$ produces

(4) $\overline{\sqrt{-t}}\, z + \sqrt{-t}\, \bar{z} - \overline{\sqrt{-t}}\, a = 0$.

Now, the preceding is the equation of a line in the Argand plane if $\overline{\sqrt{-t}}\, a$ is real. But, T 3.17.D. insures that $\overline{\sqrt{-t}}\, a$ is real if $\sqrt{-t}\, \bar{a} - \overline{\sqrt{-t}}\, a = 0$, and this was established in (2). Therefore, $m: \overline{\sqrt{-t}}\, z + \sqrt{-t}\, \bar{z} - \overline{\sqrt{-t}}\, a = 0$ is the set of points preserved under f.

The foregoing theorem suggests the appropriateness of the following terminology.

Definition 6.8. An opposite isometry f is a line reflection if and only if f has an invariant point.

In light of the proof of T 6.11. and the observations following Corollaries 5.8 and 5.11, it can be seen that the invariant line m, under $f(z) = t\overline{z} + a$, is parallel to $i\sqrt{-t}$ and at a vector distance $\frac{a}{2}$ from 0. Since $i\sqrt{-t} \,||\, \sqrt{t}$, one can also write $m \,||\, \sqrt{t}$. Moreover, a line ℓ is determined if a vector parallel to the line is given and a point through which ℓ passes is known. As a result of this it is fitting to denote the reflection in m by $E(\sqrt{t}, \frac{a}{2})$. More generally, the reflection in ℓ, where $\ell \,||\, c$ and passes through d, will be symbolized by E(c,d). Using this notation it is not too difficult to see that $E(\sqrt{t}, \frac{a}{2}) = T(\frac{a}{2})\circ E(\sqrt{t},0)\circ T(-\frac{a}{2})$. To establish the corresponding factorization in the more general case it helps to first verify that if $\ell{:}\overline{a}z + a\overline{z} + b = 0$ is a line, then the reflection in ℓ is the transformation $f(z) = -\frac{a}{\overline{a}}\,\overline{z} - \frac{b}{\overline{a}}$. It now follows that the reflection in the line $m{:}\overline{ic}(z - d) + ic(\overline{z - d}) = 0$, parallel to c and passing through d, is $f(z) = \frac{c}{\overline{c}}(\overline{z - d}) + d$.

From the foregoing it is relatively easy to see that $E(c,d) = f = T(d)\circ E(c,0)\circ T(-d)$. Conversely, it is almost immediate that every product of the form $T(d)\circ E(c,0)\circ T(-d)$ is a line reflection. These observations result in the following characterization of an opposite isometry having an invariant point.

Theorem 6.12. An opposite isometry f has an invariant point if and only if $f = T(d)\circ E(c,0)\circ T(-d)$.

Proof. The argument is sketched in the preceding paragraph.

In general the product of T(b) and E(c,d), where $b \perp c$, is a line reflection. This result concerning the composition of reflections and translations is manifest in the following proposition.

Theorem 6.13. The product of a translation through b, $b \neq 0$, and a reflection in ℓ, in either order, is a reflection if and only if $b \perp \ell$. Specifically, $T(b)\cdot E(c,d) = E(c,d + b/2)$, if $b \perp c$.

Proof. Let E(c,d) be a reflection in $\ell \parallel c$. Now, by the discussion following D 6.8. $E(c,d) = f$, where $f(z) = \frac{c}{\overline{c}}(\overline{z - d}) + d$. Also $g = T(b)$, where $g(z) = z + b$. Consequently, $g \cdot f(z) = \frac{c}{\overline{c}}(\overline{z - d}) + d + b = \frac{c}{\overline{c}}\,\overline{z} + \frac{-c}{\overline{c}}\,\overline{d} + d + b$. Since $g \cdot f$ is an opposite isometry, then in accordance with T 6.10. it will be a reflection if and only if $\frac{c}{\overline{c}}\overline{\left[\frac{-c}{\overline{c}}\,\overline{d} + d + b\right]} + \frac{-c}{\overline{c}}\,\overline{d} + d + b = 0$. However, $\frac{c}{\overline{c}}\overline{\left[\frac{-c}{\overline{c}}\,\overline{d} + d + b\right]} + \frac{-c}{\overline{c}}\,\overline{d} + d + b = -d + \frac{c}{\overline{c}}\,\overline{d} + \frac{c}{\overline{c}}\,\overline{b} + \frac{-c}{\overline{c}}\,\overline{d} + d + b = \frac{c}{\overline{c}}\,\overline{b} + b$. Therefore $g \cdot f$ is a reflection if and only if $\frac{c}{\overline{c}}\,\overline{b} + b = 0$, or alternately if and only if $c\overline{b} + \overline{c}b = 0$. However, the latter is precisely the analytic condition for the perpendicularity of b and c. Since $c \parallel \ell$ it follows that $g \cdot f$ is an opposite isometry with a fixed point if and only if $b \perp \ell$.

To see that $T(b)\cdot E(c,d) = E(c,d + b/2)$ when $b \perp \ell$, note that $g \cdot f(d + b/2) = d + b/2$, and that the invariant line under $g \cdot f$ is parallel to ℓ.

The demonstration is similar when $f \cdot g$ is considered.

Before outlining the principal composition theorems for reflections a couple of other observations merit some attention. First, as one might anticipate, every reflection is its own inverse. This can be readily verified by direct composition. Second, although the general significance of the condition $t\bar{a} + a = 0$ is apparent in T 6.10.-T.6.13., one can give a direct geometric interpretation of this equality. In particular, if $f(z) = t\bar{z} + a$ is a reflection, then $t\bar{a} + a = 0$ and $\ell: \sqrt{-t}\, z + \overline{\sqrt{-t}}\, \bar{z} - \overline{\sqrt{-t}}\, a = 0$ is invariant under f. In keeping with C 5.11. and the projection theorem at the end of Chapter V, it follows that $\frac{t\bar{a} + a}{2}$ is the projection of a on ℓ. Appropriately this is zero when a is perpendicular to ℓ.

The reader will recall that a primary objective of the current chapter was to give a motivated exposition of the fundamental nature of reflections. The following two results are pivotal in this regard.

Theorem 6.14. The product of two reflections is a translation or nontrivial rotation according to whether the invariant lines are parallel or intersect in a single point.

Proof. Let $f(z) = t\bar{z} + a$ and $g(z) = s\bar{z} + b$ be reflections in ℓ_1 and ℓ_2 respectively. Then $f \cdot g(z) = t(\overline{s\bar{z} + b}) + a = (t\bar{s})z + t\bar{b} + a$, which is a direct isometry. Thus, $f \cdot g$ is a translation or nontrivial rotation depending on whether $t\bar{s} = 1$ or $t\bar{s} \neq 1$ respectively.

In case $t\bar{s} = 1$, then $t = s$, since s is a turn. However, $\ell_1: z - t\bar{z} - a = 0$ and $\ell_2: z - s\bar{z} - b = 0$, and in keeping with C 5.9. $\ell_1 \parallel \ell_2$ if and only if $t = s$. Therefore, $f \cdot g$ is a translation if and only if $\ell_1 \parallel \ell_2$. This essentially completes the proof, since if ℓ_1 and ℓ_2 are not parallel, then they intersect in a single point and it follows that $t \neq s$, or alternately $t\bar{s} \neq 1$.

Theorem 6.14 suggests the possibility of factoring any direct isometry into the product of a pair of line reflections. The validity of this conjecture becomes apparent in the following theorem.

Theorem 6.15. If h is a direct isometry, then h can be factored into the product of reflections in a pair of parallel or intersecting lines according to whether h is a translation or nontrivial rotation.

Proof. In case $h(z) = z + c$ consider $f \cdot g$, where f and g respectively are reflections in $\ell_1: \bar{c}z + c\bar{z} + a = 0$ and $\ell_2: \bar{c}z + c\bar{z} + b = 0$. Here b is an arbitrary real number and $a = \bar{b} - |c|^2$. By an earlier discussion $f(z) = -\frac{c}{\bar{c}}\bar{z} - \frac{a}{\bar{c}}$ and $g(z) = -\frac{c}{\bar{c}}\bar{z} - \frac{b}{\bar{c}}$. Thus, $f \cdot g(z) = -\frac{c}{\bar{c}}\overline{\left(-\frac{c}{\bar{c}}\bar{z} - \frac{b}{\bar{c}}\right)} - \frac{a}{\bar{c}} = z + \frac{(\bar{b} - a)}{\bar{c}}$. However, $\bar{b} - a = |c|^2$; hence $f \cdot g(z) = z + \frac{|c|^2}{\bar{c}} = z + c$. Therefore, $h = f \cdot g$.

Now, suppose $h(z) = az + c$, $|a| = 1$, $a \neq 1$. Let $u = \frac{c}{1 - a}$. To see that h can be factored into reflections consider the isometries $g(z) = t(\overline{z - u}) + u$ and $f(z) = s(\overline{z - u}) + u$, where $s\bar{t} = a$. In light of T 6.12. f and g are reflections. By direct composition $f \cdot g(z) = s[\overline{t(\overline{z - u}) + u - u}] + u = (s\bar{t})z - s\bar{t}u + u$. But, $s\bar{t} = a$, hence $f \cdot g(z) = az - au + u = a(z - u) + u$. However, the last expression is precisely the rotation h in canonical form. Consequently, $h = f \cdot g$. Finally, since $f \cdot g$ is not a translation the lines of reflection must intersect by T 6.14.

Although T 6.15 is algebraically complete, it requires some inspection to gain insight into the geometry of the indicated products.

Of course, there are certain results in the preceding chapter that the reader will find pertinent to a geometric interpretation of T 6.15. However, perhaps the best approach is the synthetic one, with subsequent reference to the analytic results of Chapter V. In any event, one should observe that the factorization of a direct isometry into a pair of reflections is not unique. This is suggested algebraically by the arbitrariness of one of the constants in each of the factorizations outlined in the proof of T 6.15. Specifically, in the case of a translation any line m perpendicular to the translation vector c can be selected as the initial line of reflection. The second line must be the image of m under a translation through c/2. In factoring a rotation about u, of arg a, $|a| = 1$, into reflections, any line ℓ containing u can be picked for the initial invariant line. The second line of reflection must be the image of ℓ under a rotation about u through 1/2 arg a. Symbolically these factorizations can be written $T(c) = E(ic,a)\cdot E(ic,b)$, where $a = b + c/2$, and $R(a,u) = E(s,u)\cdot E(t,u)$, where $s = t\sqrt{a}$.

The preceding observations terminate the investigation into the matter of decomposing direct isometries into a product of reflections. Attention is now focused on the opposite isometries having no fixed points.

Definition 6.9. If $f(z) = t\bar{z} + a$, then f is a glide reflection if and only if $t\bar{a} + a \neq 0$.

The appropriateness of the foregoing terminology becomes apparent in the following proposition.

Theorem 6.16. If $f(z) = t\overline{z} + a$ is a glide reflection, then $f = g \cdot h$ where h is a line reflection and g is a nontrivial translation parallel to h.

Proof. Write $f(z) = (t\overline{z} - [\frac{t\overline{a} - a}{2}]) + \frac{t\overline{a} + a}{2} = (t(\overline{z - a/2}) + a/2) + \frac{t\overline{a}+a}{2}$

Let $h(z) = t(\overline{z - a/2}) + a/2$ and $g(z) = z + \frac{t\overline{a} + a}{2}$. Then h is a reflection by T 6.12. Furthermore, the invariant line under h is $\ell : \overline{\sqrt{-t}}\, z + \sqrt{-t}\, \overline{z} - \overline{\sqrt{-t}}\, a = 0$. Thus, it remains to show that the nonzero vector $\frac{t\overline{a} + a}{2}$ is parallel to ℓ, or alternately that $-t\overline{a} - a$ is parallel to ℓ. T 5.11.E. and a little algebra can be used to verify this.

Since the product of a translation and an opposite isometry is again an opposite isometry, T 6.13. can be used to establish the converse of the preceding result.

In connection with glide reflections it is not difficult to see how to construct the isometry that corresponds to a reflection in a given line followed by a nontrivial translation parallel to that line. Specifically, the synthesis of such a function could be accomplished by constructing the appropriate line reflection by the procedure outlined earlier, then composing this with the given translation. It is of note that such a product can be shown to be commutative.

In reflecting on the proof of Theorem 6.16 the reader might well seek some motivation for the factorization given. The interpretation of $\frac{t\overline{a} + a}{2}$ outlined in the discussion preceding T 6.14. proves enlightening in this regard.

Theorem 6.16 not only serves to disclose the geometric nature of a glide reflection, but it is the final result needed to exhibit the fundamental character of line reflections. In particular, the following proposition is a consequence of T 6.15. and T 6.16.

Theorem 6.17. Every isometry f can be expressed as the product of no more than three line reflections. If f has an invariant point, then no more than two factors are required.

Proof. Theorem 6.15 insures that every direct isometry can be so factored. Theorem 6.16 implies that a glide reflection can be written as the product of a reflection and translation. But the translation can be factored into two reflections, hence the glide reflection can be written as the product of the three reflections. A reflection trivially satisfies the conclusion, thus the first part of the theorem is established. The second part is immediate, since only the glide reflection requires three such factors and it has no invariant points.

The foregoing is the focal point of the chapter in view of the stated objectives. However, there is one additional result of a similar type that brings to light the fundamental nature of reflections in a broader sense. The term reflection used in the more encompassing sense refers to any isometry of the form $f(z) = t\bar{z} + a$, $t\bar{a} + a = 0$, or $f(z) = -z + 2b$. The reader undoubtedly recognizes the appropriateness of calling a transformation of the latter form a reflection, or more specifically a point reflection. Such terminology was not adopted in this paper to eliminate any ambiguity in the use of the term reflection. Of course this convention will be continued, nevertheless the following results appear to be more in keeping with the spirit of the chapter if both half turns and line reflections are viewed as reflections.

In connection with the thought posed in the preceding paragraph, it is clear in light of T 6.17. that any interesting result regarding the decomposition of isometries into reflections, in the broader sense, must involve no more than two factors. Furthermore, previous considerations indicate that, with the exception of glide reflections, every isometry can be written as a product of no more than two line reflections. Thus, to establish the possibility of factoring every isometry into one or two reflections, in the more encompassing sense of the word, it remains to represent a glide reflection as the product of a half turn and line reflection. A reasonable approach to the problem of determining whether such a decomposition exists, would be to multiply a line reflection and half turn and see if the product could take on the form of a glide reflection. To this end consider the following theorem.

Theorem 6.18. The product of a half turn about b and a reflection in ℓ is a glide reflection if and only if b is not on ℓ.

Proof. Let $f(z) = t\overline{z} + a$ and $g(z) = -z + 2b$ be the given reflection and half turn. Note that the invariant line under f is $\ell: z - t\overline{z} - a = 0$. Now, $f \circ g(z) = t(\overline{-z + 2b}) + a = (-t)\overline{z} + 2t\overline{b} + a$. Clearly $f \circ g$ is an opposite isometry, thus it is a glide reflection if and only if $-t(\overline{2t\overline{b} + a}) + (2t\overline{b} + a) \neq 0$. Moreover, $-t(\overline{2t\overline{b} + a}) + 2t\overline{b} + a$ $= -2b - t\overline{a} + 2t\overline{b} + a = -2b - t\overline{a} - a + 2t\overline{b} + 2a$. But, $-ta - a = 0$, hence $-t(\overline{2t\overline{b} + a}) + 2t\overline{b} + a = -2b + 2t\overline{b} + 2a$. From this it follows that f·g is a glide reflection if and only if $-2b + 2t\overline{b} + 2a \neq 0$, or equivalently if and only if $b - t\overline{b} - a \neq 0$. However, $b - t\overline{b} - a \neq 0$ is precisely the condition that b not be incident with ℓ.

In addition to being a logical antecedent of any proposition regarding the desired factoring of a glide reflection, the foregoing is of note in another respect. In particular, observe that multiplying a half turn and reflection always results in an opposite isometry, thus in light of T 6.18. such a product will be a reflection if and only if b lies on ℓ. Hence a corollary of the foregoing theorem provides a necessary and sufficient condition for incidence of a point and line in terms of a product of reflections (in the broad sense). Actually the condition can be refined somewhat, by verifying that the product of a half turn and a reflection is a reflection if and only if it is commutative. At any rate this result is one of several such propositions which afford a characterization of a geometric notion in terms of a condition on the product of half turns and reflections. [32] These are not developed here.

The following theorem provides the answer to the question which precipitated T 6.18.

Theorem 6.19. If $f(z) = t\overline{z} + a$, $t\overline{a} + a \neq 0$, then $f = g \cdot h$, where h is a half turn and g is a reflection.

Proof. Consider $g(z) = -t\overline{z} + b$ and $h(z) = -z + 2c$, where b is such that $-t\overline{b} + b = 0$ and $c = (\overline{\frac{b - a}{2t}})$. Then by direct composition

$$g \cdot h(z) = -t\left(\overline{-z + [\overline{\frac{b - a}{t}}]}\right) + b = t\overline{z} - b + a + b = t\overline{z} + a.$$

The appropriate factors in the foregoing were suggested by the demonstration of T 6.18. Again the decomposition is not unique. Geometrically the preceding is not very revealing, but the reader should recognize that the choice of -t as the coefficient of $\bar{z}$, is equivalent to selecting ℓ, the invariant line under g, perpendicular to m, the reflection line under f. The arbitrary nature of b corresponds to the fact that, aside from the aforementioned perpendicularity condition, the choice of ℓ is arbitrary. The restriction on c is somewhat more obscure, but it can be shown to imply that c is the point on m at a vector distance $-\frac{d}{2}$ from the point of intersection of ℓ and m. Here d is the translation vector under f. In spite of the interesting geometric implications of T 6.19. it was developed here primarily because it leads to the following result.

Theorem 6.20. Every isometry can be written as the product of no more than two reflections (point or line).

Proof. The result is apparent in view of T 6.17. and T 6.19.

The foregoing proposition, coupled with T 6.17., firmly establishes the fundamental character of point and line reflections.

CHAPTER VII

A FINAL ANALYSIS

Summary

The salient features of this paper were sketched in Chapter I. However, there are certain aspects of the presentation which are more appropriately examined in retrospect. In Chapter II the following three things are apparent. First, the development was such that the disparity between mathematics in the making and the formal presentation of the subject was brought to the fore. Second, the discussion was encompassing enough to provide a historical framework for all subsequent aspects of the work. Finally, some attention was given to external applications of complex numbers. Chapter III provided a rigorous development of the complex field motivated by the classical desire for algebraic completeness. Following certain preliminary results the basic properties of the real number system were exposed in D 3.7. Comparative reference was made to these properties after the development of the complex field.

In Chapter IV the possibility of constructing a field extension of R, satisfying conditions markedly different from those of the preceding chapter, was explored. The endeavor seems appropriate from two standpoints. First, it provides a glimpse of a currently fertile branch of mathematics, distinct from complex analysis, which evolved out of mans investigation of complex numbers. Second, after viewing the complex algebraic treatment of the isometries of the plane outlined in Chapter VI the geometric significance of constructing a higher dimensional field extension of R becomes apparent. The results of Chapter IV suggest that any algebraic treatment of the isometries of higher dimensional Euclidean spaces must be based on number systems which fail to satisfy certain of the field properties. Simultaneously these theorems serve to establish the unique position of the complex number system as a finite field extension of R. Thus, Chapters III and IV point to the peculiar position of the field C from two different vantage points.

Chapter V provided a desirable link between the arithmetic operations on C and the geometry of the plane. These results are instrumental in interpreting the propositions in Chapter VI. Perhaps the most significant aspect of Chapter VI is that it utilizes complex numbers to produce an algebraic model of a geometric notion.

Educational Implications

It often happens that the sincere student of mathematics is formally introduced to the complex field in a graduate level course in complex analysis. In part, this appears to be due to the fact that little has been written on the subject with the undergraduate in mind. It is anticipated that this paper will contribute to the literature by making available a compendium of results about the complex field, which are accessible to one having the mathematical maturity of a good high school senior or undergraduate. It is foreseeable that the audience might well include secondary teachers of mathematics. It is hoped that an acquaintance with the development in this thesis will not only leave the reader better informed regarding the complex field, but promote a continuing interest in mathematics.

The writer can personally attest to the fact that this paper has already proven to be a valuable educational device.

A SELECTED BIBLIOGRAPHY

1. Albert, A. A., "Absolute-Valued Algebraic Algebras," Bulletin of the American Mathematical Society, Vol. 55, (1949), pp. 763-768.

2. Artzy, Rafael, Linear Geometry, Reading, Mass.: Addison Wesley Pub. Co., Inc., 1965.

3. Bell, E. T., The Development of Mathematics, New York: McGraw-Hill Co., Inc., 1945.

4. Birkhoff, Garrett and MacLane, Saunders, A Survey of Modern Algebra, New York: MacMillian Co., 1965.

5. Bochner, Salomon, The Role of Mathematics in the Rise of Science, Princeton, New Jersey: Princeton University Press, 1966.

6. Bott, R. and Milnor, J., "On the Parallelizability of the Spheres," Bulletin of the American Mathematical Society, Vol. 64, (1958), pp. 87-89.

7. Bruck, R. H. and Kleinfield, E., "The Structure of Alternative Division Rings," Procedings of the American Mathematical Society, Vol. 2, (1951), pp. 878-890.

8. Coxeter, H.S.M., "Quaternions and Reflections," The American Mathematical Monthly, Vol. 53, (1946), pp. 136-146.

9. _____, Introduction to Geometry, New York: John Wiley and Sons, Inc., 1961.

10. Dantzig, Tobias, Number the Language of Science, New York: MacMillan Co., 1954.

11. Deaux, Roland, Introduction to the Geometry of Complex Numbers, New York: Frederick Ungar Pub. Co., 1956.

12. Dubish, Roy, The Nature of Number, New York: The Ronald Press Co., 1952.

13. Eves, Howard, A Survey of Geometry, Vol. 2, Boston: Allyn and Bacon, Inc., 1965.

14. Feferman, Solomon, The Number Systems, Reading, Mass.: Addison-Wesley Pub. Co., 1964.

15. Guilleman, E. A., Communication Networks, Vol. 1, New York: John Wiley and Sons, Inc., 1931.

16. Hamilton, W. R., Lectures on Quaternions, Dublin: Hodges and Smith Pub. Co., 1853.

17. Hansen, D. J., "A Note on the Operation of Multiplication in the Complex Plane," The American Mathematical Monthly, Vol. 71, pp. 185-186.

18. Herstein, I. N., Topics in Algebra, New York: Blaisdell Pub. Co., 1964.

19. Iaglom, I. M., Geometric Transformations, New York: Random House, Inc., 1962.

20. Johnson, R. E., First Course in Abstract Algebra, New York: Prentice-Hall, Inc., 1953.

21. Kurosh, A. G., Lectures on General Algebra, New York: Chelsea Pub. Co., 1963.

22. McKirnan, Sharon K., "A Generalization of Multiplication of Complex Numbers," Pi Mu Epsilon Journal, Vol. 4, (1965), pp. 49-56.

23. Meserve, B. E., Fundamental Concepts of Geometry, Reading, Mass.: Addison-Wesley Pub. Co., Inc., 1955.

24. Mostow, G. D., Sampson, J. H. and Meyer, J., Fundamental Structures of Algebra, New York: McGraw-Hill Co., Inc., 1963.

25. Niven, Ivan, "The Roots of a Quaternion," The American Mathematical Monthly, Vol. 49, (1942), pp. 386-388.

26. Paige, L. J. and Swift, J. D., Elements of Linear Algebra, Boston: Ginn and Co., 1961.

27. Pennisi, L. L., Elements of Complex Variables, New York: Holt, Rinehart and Winston, Inc., 1963.

28. Potts, D. H., "Axioms for Complex Numbers," The American Mathematical Monthly, Vol. 70, (1963), pp. 974-975.

29. Sawyer, W. W., "Trigonometry Abstractly Treated," The American Mathematical Monthly, Vol. 64, (1957), pp. 734-737.

30. Sears, F. W. and Zemansky, M. W., College Physics, Reading, Mass.: Addison-Wesley Pub. Co., 1960.

31. Strong, John, Concepts of Classical Optics, San Francisco: W. H. Freeman and Co., Inc., 1958.

32. Thomsen, G., "The Treatment of Elementary Geometry by Group-Calculus," The Mathematical Gazette, Vol. 17, (1933), pp. 230-242.

33. Whittaker, E. T., History of the Theories of Aether and Electricity, New York: Philosophical Library Inc., 1951.

34. Yale, P. B., "Automorphisms of the Complex Numbers," Mathematics Magazine, Vol. 39, (1966), pp. 135-141.

35. Zwikker, C., Advanced Plane Geometry, Amsterdam, Netherlands, North-Holland Pub. Co., 1950.

VITA

Michael Clyde Clock

Candidate for the Degree of

Doctor of Education

Thesis: CERTAIN ALGEBRAIC AND GEOMETRIC ASPECTS OF THE COMPLEX FIELD

Major Field: Higher Education

Minor Field: Mathematics

Biographical:

Personal Data: Born in Portland, Oregon, November 17, 1935, the son of Charles and Ellen Clock.

Education: Attended the elementary school at Ladd Hill, Oregon; graduated from Newberg High School, Newberg, Oregon in 1953; received the Bachelor of Science degree from Lewis and Clark College, Portland, Oregon in 1957 with majors in mathematics and physics; received the Master of Science degree from Oregon State University in 1961 with a major in general science; attended Portland State College in the summers of 1958, 1959, 1963, Reed College Portland, Oregon in the summer of 1960, Oregon State University from September 1960 to August 1961, University of Santa Clara, Santa Clara, California during the summer of 1964 and Oklahoma State University from June 1965 through July 1967.

Professional Experience: Served as an instructor in physics at Tigard High School, Tigard, Oregon in 1957-58; taught mathematics at Sunset High School, Beaverton, Oregon during the academic years beginning in 1958, 1959 and 1961 - 1964; served as a consultant for the Mathematics Curriculum Revision Committee of School District 48, Beaverton, Oregon, in the summer of 1962; was a graduate assistant in mathematics at Oklahoma State University, 1965 - 1967.

Professional Organizations: Member of the Mathematical Association of America and Pi Mu Epsilon.