Some topics in the theory of Diophantine approximation

R.F.C. Walters

January, 1967

SOME TOPICS IN THE THEORY OF DIOPHANTINE APPROXIMATION

by

ROBERT FRANK CARSLAW WALTERS

A thesis submitted for the degree of Master of Science

at the University of Queensland, January 1967.

Acknowledgements

I wish to thank Professor C.S. Davis for his helpful supervision of this thesis. Financial support during the work on the thesis - in 1965 and 1966 - was provided by a Commonwealth Post-Graduate Award. I further wish to thank Mr. K.R. Matthews for his interest in the topic of this thesis and his collaboration on one or two points.

Contents

Introduction

Let Λ be a singular matrix with real elements such that there are no integral vector solutions $\underset{\sim}{x}$, other than $\underset{\sim}{0}$, of the equation

$$\Lambda \underset{\sim}{x} = \underset{\sim}{0} . \qquad \ldots(1)$$

Then by a theorem of Minkowski's (see Corollary 3 of Appendix B in Cassells [3]) we can choose integral vectors $\underset{\sim}{x} \neq \underset{\sim}{0}$ such that the components of $\Lambda \underset{\sim}{x}$ are as small as we please. A central problem in the theory of Diophantine Approximation is to find out how 'small' $\Lambda \underset{\sim}{x}$ can be, with respect to the 'size' of the integral vector $\underset{\sim}{x}$. More precisely, if $m(\underset{\sim}{x})$ is a measure of the size of $\Lambda \underset{\sim}{x}$ and $c(\underset{\sim}{x})$ is a measure of the size of $\underset{\sim}{x}$, we wish to find two functions $\varphi(t)$, $\psi(t)$ both tending to zero as t tends to infinity, with the properties:

I. There are an infinite number of integral vectors $\underset{\sim}{x}$ such that

$$m(\underset{\sim}{x}) < \varphi\{c(\underset{\sim}{x})\};$$

II. For all but a finite number of integral vectors

$$m(\underset{\sim}{x}) > \psi\{c(\underset{\sim}{x})\}.$$

A closely related problem, also important, is to determine, given a particular real matrix Λ, whether there are integral vector solutions $\underset{\sim}{x} \neq \underset{\sim}{0}$ of equation (1). This includes the problem of the transcendence of the real number α when we consider the sequence of matrices $\Lambda_n = (\alpha^n, \alpha^{n-1}, \ldots, \alpha, 1)$.

The concern of this thesis is to deal with certain cases of these problems from the point of view of a certain infinite process involving matrices which we call the κ-process:- Let $A_1, A_2, A_3, \ldots$ be a sequence of $m \times m$ matrices over the field of real numbers, and let $\underset{\sim}{e}$ be a unit m-vector. Then frequently there exists a sequence of constants $\{k_n\}$ such that

$$k_n A_1 A_2 \cdots A_n \underset{\sim}{e} \rightarrow \underset{\sim}{\alpha} \neq \underset{\sim}{0}$$

as $n \to \infty$ (where $\underset{\sim}{x}_n \to \underset{\sim}{\alpha}$ means that the components of $\underset{\sim}{x}_n$ tend to the components of $\underset{\sim}{\alpha}$).

The relevance of this process to the outlined problems is that a special case of the process, the continued fraction algorithm, has proved to be of great importance when Λ is a 1×2 matrix. Many authors have written on generalizations of the continued fraction process mainly with a view to extending the very special number theoretic properties of the regular continued fraction algorithms - an attempt which has not met with any notable success. Almost all the generalizations are included in the κ-process. A list of these authors (before 1936) is contained in Koksma [8, pp.50-51]. In this thesis we do not attempt to find an exact generalization of the regular continued fraction process, but rather we develop the theory of the κ-process, and obtain, from this point of view, what applications to Diophantine Approximation we can. More specifically, chapters 1, 2 and 3 contain the definition, some convergence theorems, and some transformation theorems for the κ-process. In chapter 4 there is a brief treatment of continued fractions, and a mildly interesting generalization. In chapter 5 a variety of particular κ-expansions involving well known functions are given. This chapter includes a rather different proof of the regular continued fraction expansions of $e^{2/k}$ and $\tan(1/k)$ (where k is a positive integer). In chapter 6 there is a proof of a very special arithmetic property (the best approximation property) of regular continued fractions and a partial generalization. The final chapter contains two further applications to Diophantine Approximation of the κ-process.

1. An Infinite Process Involving Matrices

Notation

1. In this work matrices are understood to have complex numbers as elements. We denote $m \times m$ matrices by capital letters or by (a_{ij}) where the a_{ij} are the elements of the matrix, and m-vectors ($m \times 1$ matrices) we denote by $\underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{\alpha}, \underset{\sim}{\beta}$ etc.. The unit vectors we write as usual $\underset{\sim}{e}_1, \underset{\sim}{e}_2, \cdots, \underset{\sim}{e}_m$ and the zero vector we write as $\underset{\sim}{o}$.

2. If $\{\underset{\sim}{x}_n\}$ is a sequence of vectors, by the phrase '$\underset{\sim}{x}_n \to \underset{\sim}{\alpha}$' we mean that the components of $\underset{\sim}{x}_n$ tend to the components of $\underset{\sim}{\alpha}$.

3. If $A = (a_{ij})$ and $B = (b_{ij})$, by $A > B$ we mean

$$a_{ij} \geqslant b_{ij} \quad (i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, m) \text{ and } A \neq B.$$

We use with similar meaning the phrase '$\underset{\sim}{x} > \underset{\sim}{y}$'.

We now introduce our fundamental definitions.

Definition 1.1

Two vectors $\underset{\sim}{\alpha}, \underset{\sim}{\beta}$ are said to be equivalent if there exists a non-zero number k such that

$$\underset{\sim}{\alpha} = k\, \underset{\sim}{\beta}.$$

We write $\underset{\sim}{\alpha} \sim \underset{\sim}{\beta}$.

It is easy to see that the relation is indeed an equivalence relation over the set of all vectors.

Definition 1.2

If $\{\underset{\sim}{x}_n\}$ is a sequence of m-vectors, then we say that $K\,\underset{\sim}{x}_n$ converges to $\underset{\sim}{\alpha}$ if $\underset{\sim}{\alpha} \neq \underset{\sim}{0}$ and there exists a sequence of (non-zero) numbers $\{k_n\}$ such that

$$k_n \underset{\sim}{x}_n \rightarrow \underset{\sim}{\alpha} \qquad \text{as} \qquad n \rightarrow \infty .$$

We write this as

$$K\,\underset{\sim}{x}_n \rightarrow \underset{\sim}{\alpha} \qquad \text{or} \qquad \underset{\sim}{\alpha} = \lim_{n\rightarrow\infty} K\,\underset{\sim}{x}_n .$$

Evidently $K\,\underset{\sim}{x}_n \rightarrow \underset{\sim}{\alpha}$ and $K\,\underset{\sim}{x}_n \rightarrow \underset{\sim}{\beta}$ if and only if $\underset{\sim}{\alpha} \sim \underset{\sim}{\beta}$. Thus although there is no unique limit of a sequence of vectors, the various limits are precisely the members of a certain equivalence class. It is clear that if $K\,\underset{\sim}{x}_n \rightarrow (\alpha_i)$ and α_j is a non-zero component, then there is a (β_i) such that $K\,\underset{\sim}{x}_n \rightarrow (\beta_i)$ and $\beta_j = 1$.

Definition 1.3

If $A_1, A_2, A_3, \cdots$ is a sequence of $m \times m$ matrices and

$$K\,A_1 A_2 \cdots A_n \underset{\sim}{e}_i \rightarrow \underset{\sim}{\alpha}_i \qquad \text{as} \qquad n \rightarrow \infty$$

then we write

$$\underset{\sim}{\alpha}_i \sim K_i\, A_1 A_2 A_3 \cdots$$

or

$$\underset{\sim}{\alpha}_i \sim K_i \prod_{n=1}^{\infty} A_n .$$

If $\underset{\sim}{\alpha}_i$ exists for $i = 1, 2, \cdots, m$ and $\underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2 = \cdots = \underset{\sim}{\alpha}_m = \underset{\sim}{\alpha}$ then we write simply

$$\underset{\sim}{\alpha} \sim K\, A_1 A_2 A_3 \cdots .$$

We call this infinite process the K-process, and particular examples of the K-process we call K-products, or K-expansions. When we wish to refer to the size of matrix involved, we speak of $m \times m$ K-products etc..

We call the vectors $A_1 A_2 \cdots A_n \underset{\sim}{e}_i$ the ith column convergents or the i-convergents of the K-expansion. The set of all i-convergents for $(i=1,2,\cdots,m)$ we call simply the convergents of the K-expansion.

Remark 1.1

If $K \underset{\sim}{x}_n \rightarrow \underset{\sim}{\alpha}$ where $\underset{\sim}{x}_n = (x_{in})$ and $\underset{\sim}{\alpha} = (\alpha_i)$ and $x_{jn} \neq 0$, $\alpha_j \neq 0$ for some particular j and all n, then

$$\frac{x_{in}}{x_{jn}} = \frac{k_n x_{in}}{k_n x_{jn}} \longrightarrow \frac{\alpha_i}{\alpha_j} \quad (i=1,2,\cdots,m) \text{ as } n \rightarrow \infty .$$

Conversely if $\underset{\sim}{x}_n = (x_{in})$ $(n=1,2,3,\cdots)$ is a sequence of vectors and for some j

$$\frac{x_{in}}{x_{jn}} \longrightarrow \frac{\alpha_i}{\alpha_j} \quad (i=1,2,\cdots,m) \quad \text{as } n \rightarrow \infty$$

then by choosing k_n to be α_j / x_{jn} we easily see that

$$K \underset{\sim}{x}_n \longrightarrow (\alpha_i) .$$

Some facts immediately consequent on our definitions are given (without proof) in the following lemmas.

Lemma 1.1

If $\underset{\sim}{\alpha} \sim K A_1 A_2 A_3 \cdots$ and $B\underset{\sim}{\alpha} \neq \underset{\sim}{0}$, then

$$B\underset{\sim}{\alpha} \sim K\, B A_1 A_2 \cdots .$$

If $\det B \neq 0$, then the condition $B\underset{\sim}{\alpha} \neq \underset{\sim}{0}$ is obviously satisfied.

Lemma 1.2

If $\{k_n\}$ is a sequence of non-zero numbers, then

$$K A_1 A_2 A_3 \cdots \sim K (k_1 A_1)(k_2 A_2)(\cdots ,$$

if either side exists.

Lemma 1.3

$$K A_1 A_2 \cdots A_p A_{p+1} \cdots A_q A_{q+1} \cdots \sim K (A_1 A_2 \cdots A_p)(A_{p+1} \cdots A_q)(A_{q+1} \cdots$$

if the left hand side exists.

We now prove a theorem which establishes a connection between continued fractions (not necessarily regular) and the K-process. Continued fractions have been treated from the point of view of the K-process by Chatelet [4] and Kolden [9] . See also Perron [11, pp. 12-13] .

Theorem 1.1

If the continued fraction

$$\frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots \qquad \text{(in an obvious notation)}$$

converges to α and its convergents are $\frac{p_n}{q_n}$, then

$$\begin{pmatrix} q_n \\ p_n \end{pmatrix} \sim \begin{pmatrix} b_1 & 1 \\ a_1 & 0 \end{pmatrix}\begin{pmatrix} b_2 & 1 \\ a_2 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} b_1 & 1 \\ a_1 & 0 \end{pmatrix} \cdots \begin{pmatrix} b_{n+1} & 1 \\ a_{n+1} & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and hence the K-product $K\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} b_1 & 1 \\ a_1 & 0 \end{pmatrix}\begin{pmatrix} b_2 & 1 \\ a_2 & 0 \end{pmatrix} \cdots$ converges to $\begin{pmatrix} \alpha \\ 1 \end{pmatrix}$.

Proof

We prove by induction on k that for $k < n$

$$\begin{pmatrix} 1 \\ \frac{a_{n-k}}{b_{n-k}+} \frac{a_{n-k+1}}{b_{n-k+1}+} \cdots \frac{a_n}{b_n} \end{pmatrix} \sim \begin{pmatrix} b_{n-k} & 1 \\ a_{n-k} & 0 \end{pmatrix}\begin{pmatrix} b_{n-k+1} & 1 \\ a_{n-k+1} & 0 \end{pmatrix} \cdots \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The statement is true for $k=0$. The deduction from k to $k+1$ follows from the simple assertion that

$$\begin{pmatrix} 1 \\ \dfrac{a_{n-k-1}}{b_{n-k-1}+x} \end{pmatrix} \sim \begin{pmatrix} b_{n-k-1} & 1 \\ a_{n-k-1} & 0 \end{pmatrix}\begin{pmatrix} 1 \\ x \end{pmatrix}.$$

The convergence of the K-product to $\begin{pmatrix} \alpha \\ 1 \end{pmatrix}$ follows from Remark 1.1 .

2. Two Convergence Theorems

For the K-process to be useful, K-products must converge in reasonably general circumstances. We give two simple theorems which will be adequate for our purposes. Much more detailed convergence theorems for the case of 2×2 matrices are contained in Perron [12, Chapter II]. They are stated in terms of continued fractions. Convergence theorems for $m\times m$ K-expansions, where $m>2$ have been given by various authors in relation to special generalizations of the continued fraction process. For example, see Perron [10, pp. 12-17]. As we have mentioned in the introduction, Koksma [8, pp. 50-51] has a list of such authors.

A convenient theorem for 2×2 K-expansions is the following:

Theorem 2.1

If $\{A_n\}$ is a sequence of 2×2 matrices of the form $A_n = \begin{pmatrix} a_n & 1 \\ b_n & 0 \end{pmatrix}$ where the a_n, b_n are real and $a_n \geq b_n \geq 1$, then $K\, A_1 A_2 A_3 \cdots$ converges. Further there is a limit (α_i) with $\alpha_1 \geq \alpha_2 = 1$.

Proof

Put $A_1 A_2 \cdots A_n = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = P_n$.

Now it is clear that all later convergents are of the form $P_n \begin{pmatrix} u \\ v \end{pmatrix}$ where u and v are non-negative. It is also clear that q_n increases monotonically with n and is greater than 1 for $n>0$.

Hence for $n>1$

$$\left| \frac{u p_n + v p_{n-1}}{u q_n + v q_{n-1}} - \frac{p_n}{q_n} \right| = \left| \frac{v \det P_n}{q_n(u q_n + v q_{n-1})} \right| \leqslant \left| \frac{\det P_n}{q_n q_{n-1}} \right| = \left| \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} \right| .$$

Thus if we can prove that $\frac{\det P_n}{q_n q_{n-1}} \to 0$ as $n \to \infty$, we will have proved the convergence of the sequence $\frac{p_n}{q_n}$ and by our Remark 1.1, this implies the convergence of $K A_1 A_2 A_3 \cdots$, and the fact that we can choose α_2 to be 1.

Now for $n>2$

$$\left| \frac{\det P_n}{q_n q_{n-1}} \right| = \left| \frac{b_n q_{n-2}}{(a_n q_{n-1} + b_n q_{n-2})} \right| \left| \frac{\det P_{n-1}}{q_{n-1} q_{n-2}} \right|$$
$$\leq \frac{1}{2} \left| \frac{\det P_{n-1}}{q_{n-1} q_{n-2}} \right| .$$

Hence $\left| \frac{\det P_n}{q_n q_{n-1}} \right| \leqslant \frac{1}{2^{n-2}} \left| \frac{\det P_2}{q_2 q_1} \right|$

and we have our result.

The remark that $\alpha_1 \geqslant \alpha_2$ when α_2 is chosen to be 1 follows from the obvious fact that $p_n \geqslant q_n$ for all n.

A convergence criterion for 3×3 K-expansions is

<u>Theorem 2.2</u>

If $\{A_n\}$ is a sequence of matrices where the A_n are either of the form $\begin{pmatrix} a_n & 1 & 0 \\ b_n & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ $a_n \geqslant b_n \geqslant 1$, or $\begin{pmatrix} c_n & 0 & 1 \\ d_n & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ $c_n > d_n \geqslant 1$ with c_n, d_n integers then $K_1 A_1 A_2 A_3 \cdots$ converges. If only a finite number of the matrices are of the *second* type then $K A_1 A_2 A_3 \cdots$, converges. In any case there is a limit (α_i) of the K-product with $\alpha_1 \geqslant \alpha_2 \geqslant \alpha_3 = 1$.

We need a preliminary result.

Lemma 2.1

If $\underset{\sim}{x}^{(r)} = (x_i^{(r)})$ and the j th components of $\underset{\sim}{x}^{(r)}$ $(r = 1, 2, \cdots, n)$ and $\sum_{r=1}^{n} \alpha_r \underset{\sim}{x}^{(r)}$ are not zero, then if we define $d_{ij}(\underset{\sim}{x}, \underset{\sim}{y})$ to be $\left| \frac{x_i}{x_j} - \frac{y_i}{y_j} \right|$ we have

$$d_{ij}\left(\underset{\sim}{x}, \sum_{r=1}^{n} \alpha_r \underset{\sim}{x}^{(r)}\right) \leq \sum_{r=1}^{n} \frac{|\alpha_r x_j^{(r)}| \; d_{ij}(\underset{\sim}{x}, \underset{\sim}{x}^{(r)})}{|\alpha_1 x_j^{(1)} + \alpha_2 x_j^{(2)} + \cdots + \alpha_n x_j^{(n)}|}$$

The proof follows by a simple manipulation and the triangle inequality.

Proof of Theorem 2.2

We need to consider two cases.

Case 1 - in which an infinite number of the A_n are of the first form.
Put $A_1 A_2 \cdots A_n = (\underset{\sim}{x}_n, \underset{\sim}{y}_n, \underset{\sim}{z}_n)$ and $\underset{\sim}{x}_n = (x_{i,n})$, $\underset{\sim}{y}_n = (y_{i,n})$, $\underset{\sim}{z}_n = (z_{i,n})$ $(i = 1, 2, 3)$.

Further let $A_{n(1)}, A_{n(2)}, A_{n(3)}, \cdots$ be the matrices of the first type in the expansion.

Now if $A_{m-k} A_{m-k+1} \cdots A_m \underset{\sim}{e}_1 = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ (where $k < m$), then

$$\alpha \geqslant \beta \geqslant \gamma \geqslant 1. \qquad \ldots(1)$$

It is simple to prove this by induction on k. Hence all 1-convergents subsequent to $\underset{\sim}{x}_n$ are of the form $\alpha \underset{\sim}{x}_n + \beta \underset{\sim}{y}_n + \gamma \underset{\sim}{z}_n$ where $\alpha \geqslant \beta \geqslant \gamma \geqslant 1$.
Now in view of our Remark 1.1, and the obvious fact that $x_{3,n} > 0$ for all n, we need only show that the sequences

$$\frac{x_{1,n}}{x_{3,n}} \quad \text{and} \quad \frac{x_{2,n}}{x_{3,n}}$$

are convergent.

We shall prove that $\frac{x_{1,n}}{x_{3,n}}$ converges; the proof that the other sequence converges is similar.

It is clear that for $n > n(1)$ we have $y_{3,n} > 0$ and $z_{3,n} > 0$.

From Lemma 2.1, if $m > n > n(1)$ and $\underset{\sim}{x}_m = \alpha \underset{\sim}{x}_n + \beta \underset{\sim}{y}_n + \gamma \underset{\sim}{z}_n$,

then

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) \leqslant \frac{\beta y_{3,n}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{y}_n)}{(\alpha x_{3,n} + \beta y_{3,n} + \gamma z_{3,n})} + \frac{\gamma z_{3,n}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{z}_n)}{(\alpha x_{3,n} + \beta y_{3,n} + \gamma z_{3,n})}.$$

Now it is clear that

$$\underset{\sim}{x}_n > \underset{\sim}{y}_n \quad \text{and} \quad \underset{\sim}{x}_n > \underset{\sim}{z}_n,$$

and since $\alpha \geqslant \beta \geqslant \gamma$ we have either

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \tfrac{1}{2}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{y}_n) + \tfrac{1}{3}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{z}_n)$$

or

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \tfrac{1}{3}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{y}_n) + \tfrac{1}{2}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{z}_n)$$

according as $\beta y_{3,n}$ is greater than, or less than, $\gamma z_{3,n}$.

In either case we have

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \tfrac{5}{6} \max\{ d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{y}_n), d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{z}_n)\}. \qquad \ldots(2)$$

Now if $n(2k) < n \leq n(2k+2)$, from the types of matrices involved

$$\underset{\sim}{y}_n = \underset{\sim}{x}_p \quad \text{where} \quad n(2k-2) < p < n \qquad \ldots(3)$$

and

$$\underset{\sim}{z}_n = \underset{\sim}{x}_q \quad \text{where} \quad n(2k-2) < q < n. \qquad \ldots(4)$$

Thus by successive use of inequality (2) we may show that

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \left(\tfrac{5}{6}\right)^k \max\{ d_{1,3}(\underset{\sim}{x}_t, \underset{\sim}{y}_t), d_{1,3}(\underset{\sim}{x}_t, \underset{\sim}{z}_t)\}$$

where $n(2) < t \leqslant n(4)$.

Hence if $L = \max\limits_{n(2) < t \leq n(4)} \{ d_{1,3}(\underset{\sim}{x}_t, \underset{\sim}{y}_t), d_{1,3}(\underset{\sim}{x}_t, \underset{\sim}{z}_t)\}$

we have $d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \left(\tfrac{5}{6}\right)^k L$.

Evidently $k \to \infty$ as $n \to \infty$ and hence the sequence $\frac{x_{1,n}}{x_{3,n}}$ converges.

It is clear from (3) and (4) that $K\underset{\sim}{y}_n$ and $K\underset{\sim}{z}_n$ also converge, and and to the same limit as $K\underset{\sim}{x}_n$.

If $\underset{\sim}{\alpha}=(\alpha_i)$ is a limit of the K-product then the nature of our proof assures us that $\alpha_3 \neq 0$. Hence we can choose $\alpha_3 = 1$. The fact that $\alpha_1 \geqslant \alpha_2 \geqslant \alpha_3$ when $\alpha_3 = 1$ follows from (1) when $k = m-1$.

<u>Case 2</u> - in which all but a finite number of the A_n are of the second form.

By Lemma 1.1, for convergence considerations we can discard a finite number of the A_n (since $|\det A_n| = 1$). Hence we may assume that all of the A_n are of the second form. Further by Lemma 1.1, we can consider the K-product

$$K_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} A_1 A_2 A_3 \cdots \quad \text{rather than} \quad K_1 A_1 A_2 A_3 \cdots .$$

Let $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} A_1 A_2 \cdots A_n = (\underset{\sim}{x}_n, \underset{\sim}{e}_3, \underset{\sim}{z}_n)$, $\underset{\sim}{x}_n = (x_{i,n})$, $\underset{\sim}{z}_n = (z_{i,n})$.

It is clear that for $n > 1$, $x_{3,n}$ and $z_{3,n}$ are positive, and that $\underset{\sim}{x}_n > \underset{\sim}{z}_n = \underset{\sim}{x}_{n-1}$.

As in Case 1, all 1-convergents subsequent to $\underset{\sim}{x}_n$ are of the form $\alpha \underset{\sim}{x}_n + \beta \underset{\sim}{e}_3 + \gamma \underset{\sim}{z}_n$ where $\alpha \geqslant \beta \geqslant \gamma \geqslant 1$.

From Lemma 2.1, with $m > n > 1$, $\underset{\sim}{x}_m = \alpha \underset{\sim}{x}_n + \beta \underset{\sim}{e}_3 + \gamma \underset{\sim}{z}_n$,

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) \leqslant \frac{\beta\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{e}_3)}{(\alpha x_{3,n} + \beta + \gamma z_{3,n})} + \frac{\gamma z_{3,n} + d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{z}_n)}{(\alpha x_{3,n} + \beta + \gamma z_{3,n})} \quad \ldots(5)$$

If we put $n = 2$ in this inequality we see that $d(\underset{\sim}{x}_m, \underset{\sim}{x}_2)$ is bounded as $m \to \infty$ and hence that $d(\underset{\sim}{x}_m, \underset{\sim}{e}_3)$ is bounded by L, say, as $m \to \infty$.

Further $x_{3,n} \geqslant x_{3,n-1} + x_{3,n-2} \geqslant 2x_{3,n-2}$

and hence $x_{3,n} > 2^{n/2} M$ (where M is a positive constant).

Thus from (5)

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \frac{L}{2^{n/2}M} + \frac{1}{2}\, d_{1,3}(\underset{\sim}{x}_n, \underset{\sim}{x}_{n-1}). \qquad \ldots(6)$$

Applying inequality (6) successively we obtain

$$d_{1,3}(\underset{\sim}{x}_m, \underset{\sim}{x}_n) < \frac{L(n-1)}{2^{n/2}M} + \frac{d_{1,3}(\underset{\sim}{x}_2, \underset{\sim}{x}_1)}{2^{n-1}}.$$

Hence the sequence $\frac{x_{1,n}}{x_{3,n}}$ converges. The proof that the sequence $\frac{x_{2,n}}{x_{3,n}}$ converges is similar.

Now from exactly the same considerations as in case 1 we can show that $K_1A_2A_3A_4\cdots$ has a limit (β_i) with $\beta_1 \geqslant \beta_2 \geqslant \beta_3 \geqslant 0$ and $\beta_2 \neq 0$.

Then

$$\begin{pmatrix} a_1 & 0 & 1 \\ b_1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \sim \begin{pmatrix} a_1 + \beta_3/\beta_1 \\ b_1 + \beta_2/\beta_1 \\ 1 \end{pmatrix}$$

Evidently $a_1 + \beta_3/\beta_1 \geqslant b_1 + \beta_2/\beta_1 \geqslant 1$ and so there is a limit (α_i) of $K_1A_1A_2A_3\cdots$ with $\alpha_1 \geqslant \alpha_2 \geqslant \alpha_3 = 1$.

Remark 2.1

It is clear from the method of proof of Theorem 2.2 that if $A_1, A_2, A_3, \ldots$ is a sequence of matrices satisfying the conditions of the theorem, and $\underset{\sim}{\alpha} \sim K_1A_1A_2A_3\cdots$, then $\underset{\sim}{\alpha} \sim \lim_{n\to\infty} KA_1A_2\cdots A_n \begin{pmatrix} \alpha_n \\ \beta_n \\ \gamma_n \end{pmatrix}$ if $\begin{pmatrix} \alpha_n \\ \beta_n \\ \gamma_n \end{pmatrix}$ is any sequence of vectors with $\alpha_n \geqslant \beta_n \geqslant \gamma_n > 0$.

3. Transformations of K-products

There are various ways of transforming a given K-expansion into a different K-expansion with the same limit. Lemmas 1.2 and 1.3 are simple examples of such transformations. This chapter is concerned mainly with a treatment of some classical transformations for continued fractions, as given by Perron [12, Chapter I] from the viewpoint of the K-process. We also generalize some of these transformations to apply to 3×3 K-expansions.

A large number of transformations arise from the following theorem:

Theorem 3.1

(I) If $K A_1 A_2 A_3 \cdots$ converges to $\underset{\sim}{\alpha}$, and the convergents of the K-product $K B_1 B_2 B_3 \cdots$ are (apart from a finite number) non-zero multiples of convergents of the K-product $K A_1 A_2 A_3 \cdots$, then $K B_1 B_2 B_3 \cdots$ converges, and has limit $\underset{\sim}{\alpha}$.

(II) If $K A_1 A_2 A_3 \cdots$ and $K B_1 B_2 B_3 \cdots$ both converge and an infinite number of convergents of the first expansion are non-zero multiples of convergents of the second expansion, then they both have the same limit.

Both results are obvious.

A further fruitful source of transformations is the observation that if $B_{r-1} A_r = C_r B_r \ (r = 1,2,3,\cdots)$, then $B_0 A_1 A_2 \cdots A_n = C_1 C_2 \cdots C_n B_n$. It seems plausible then that in many cases

$\lim_{n\to\infty} K C_1C_2\cdots C_n B_n \underset{\sim}{e}_i \sim K C_1C_2C_3\cdots \quad (i=1,2,3,\cdots,m)$ and hence that

$$K\, B_0\, A_1 A_2 \cdots \sim K\, C_1 C_2 C_3 \cdots .$$

Theorems 3.2 and 3.3 give conditions under which this occurs .

Theorem 3.2

If $A_1, A_2, A_3 \cdots$, and $B_1, B_2, B_3, \cdots$ are positive $m \times m$ matrices (that is, $A_n > 0$, $B_n > 0$) and none of the column vectors of B_n are $\underset{\sim}{0}$, then if $K A_1 A_2 A_3 \cdots$ converges, so does $K A_1 A_2 \cdots A_n B_n \underset{\sim}{e}_i$ $(i=1,\cdots,m)$ and to the same limit.

Proof

If $(\alpha_i) \sim K A_1 A_2 A_3 \cdots$ then one of the components of (α_i) is non-zero. We may suppose without loss of generality that $\alpha_m \neq 0$. Then by Remark 1.1, if

$$A_1 A_2 \cdots A_n = (\underset{\sim}{x}_{1,n}, \underset{\sim}{x}_{2,n}, \cdots, \underset{\sim}{x}_{m,n})$$

$$d_{i,m}(\underset{\sim}{x}_{1,n}, \underset{\sim}{x}_{j,n}) \to 0 \qquad (i=1,2,\cdots,m-1;\ j=2,3,\cdots,m)$$

as $n \to \infty$. Now let

$$A_1 A_2 \cdots A_n B \underset{\sim}{e}_1 = a_1 \underset{\sim}{x}_{1,n} + a_2 \underset{\sim}{x}_{2,n} + \cdots + a_m \underset{\sim}{x}_{m,n} = \underset{\sim}{y}_n .$$

By Lemma 2.1 $\quad d_{i,m}(\underset{\sim}{x}_{1,n}, \underset{\sim}{y}_n) \leq \sum_{j=2}^{m} d_{i,m}(\underset{\sim}{x}_{1,n}, \underset{\sim}{x}_{j,n}) \quad (i=1,2,\cdots,m-1),$

and hence, again by Remark 1.1, $K\underset{\sim}{y}_n \to \underset{\sim}{\alpha}$.

Theorem 3.3

If $B, A_1, A_2, A_3, \cdots$ and $C_1, C_2, C_3, \cdots$ are 2×2 matrices over the ring of Gaussian integers, with $|\det A_r| = 1 \quad (r=1,2,3,\cdots)$ and no column vector of B is $\underset{\sim}{0}$, and if $B C_r = A_r B \quad (r=1,2,3,\cdots)$ then if $K A_1 A_2 A_3 \cdots$ converges to $\underset{\sim}{\alpha}$, $K B C_1 C_2 C_3 \cdots$ also converges and has limit $\underset{\sim}{\alpha}$.

Proof

It suffices to verify that $K\ A_1A_2\cdots A_nB\underset{\sim}{e}_1 \to \underset{\sim}{\alpha}$ as $n\to\infty$ since the proof that $K\ A_1A_2\cdots A_nB\underset{\sim}{e}_2 \to \underset{\sim}{\alpha}$ is similar. If $\underset{\sim}{\alpha} = (\alpha_i)$ is the limit of $KA_1A_2A_3\cdots$, again we suppose without loss of generality that $\alpha_2 \neq 0$.

Put $A_1A_2\cdots A_n = \begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix} = P_n$ and $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Now since $\frac{p_n}{q_n}, \frac{r_n}{s_n}$ both tend to the same limit $\frac{\alpha_1}{\alpha_2} = \alpha$, and $\left|\frac{p_n}{q_n} - \frac{r_n}{s_n}\right| = \frac{|\det P_n|}{q_n s_n} = \frac{1}{q_n s_n}$

we must have $q_n s_n \to \infty$ as $n\to\infty$.

Now if we write $\frac{ap_n + cr_n}{aq_n + cs_n} = \alpha_n$, then

$$\left|\alpha_n - \frac{p_n}{q_n}\right| \left|\alpha_n - \frac{r_n}{s_n}\right| = \frac{|ac|}{|q_n s_n|\,|aq_n + cs_n|^2} .$$

It is easy to see from this that if $|a q_n + c s_n| \geq 1$ for all large n, then $\alpha_n \to \alpha$; for then $\frac{ac}{|q_n s_n|\,|aq_n + cs_n|^2} \to 0$ and both $\frac{p_n}{q_n}, \frac{r_n}{s_n} \to \alpha$ as $n\to\infty$.

Now since $aq_n + cs_n$ are Gaussian integers $|aq_n + cs_n| \geq 1$ if $aq_n + cs_n \neq 0$; but $(q_n, s_n) = 1$, since $|p_n s_n - q_n r_n| = 1$, and so $aq_n + cs_n = 0$ implies that $q_n s_n$ divides ac, which is impossible for sufficiently large n.

We will now give some transformation lemmas which are specializations of these theorems.

Lemma 3.1

$$K\prod_{n=1}^{\infty} \begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix} \sim K\begin{pmatrix} p_1 & 1 \\ q_1 & 0 \end{pmatrix} \prod_{n=1}^{\infty} \begin{pmatrix} p_{n+1}q_n + q_{n+1}s_n & q_n \\ -q_{n+1}\Delta_n & 0 \end{pmatrix}$$

where $\Delta_n = p_n s_n - q_n r_n$, if the left hand side exists, and $q_n \neq 0$.

This transformation is due to K.R. Matthews.

The proof follows from the observation that

$$\begin{pmatrix} p_n & 1 \\ q_n & 0 \end{pmatrix}\begin{pmatrix} p_{n+1}q_n + q_{n+1}s_n & q_n \\ -q_{n+1}\Delta_n & 0 \end{pmatrix} = q_n \begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix}\begin{pmatrix} p_{n+1} & 1 \\ q_{n+1} & 0 \end{pmatrix}.$$

This identity implies that

$$\left\{ \prod_{n=1}^{N} q_n \begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix}\right\} \begin{pmatrix} p_{N+1} & 1 \\ q_{N+1} & 0 \end{pmatrix} = \begin{pmatrix} p_1 & 1 \\ q_1 & 0 \end{pmatrix} \prod_{n=1}^{N} \begin{pmatrix} p_{n+1}q_n + q_{n+1}s_n & q_n \\ -q_{n+1}\Delta_n & 0 \end{pmatrix},$$

and hence that the convergents of the second expansion (with the exception of $\binom{1}{0}$) are non-zero multiples of the 1-convergents of the first expansion. Theorem 3.1 applies.

If $q_n = 0$ then the convergents (other than $\binom{1}{0}$) of the expansion $K \begin{pmatrix} p_1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} p_2 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} p_3 & 1 \\ 0 & 0 \end{pmatrix}\cdots$

are easily seen to be the 1-convergents of the expansion

$$K \prod_{n=1}^{\infty} \begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix}.$$

<u>Lemma 3.2</u>

$$K \prod_{n=1}^{\infty} \begin{pmatrix} a_n & d_n & g_n \\ b_n & e_n & h_n \\ c_n & f_n & j_n \end{pmatrix} \sim K \begin{pmatrix} a_1 & 1 & 0 \\ b_1 & 0 & 1 \\ c_1 & 0 & 0 \end{pmatrix} \prod_{n=1}^{\infty} A_n$$

Where
$$A_n = \begin{pmatrix} a_{n+1}c_n + b_{n+1}f_n + c_{n+1}j_n & c_n & f_n \\ b_{n+1}(c_nd_n - a_nf_n) + c_{n+1}(c_ng_n - a_nj_n) & 0 & c_nd_n - a_nf_n \\ b_{n+1}(c_ne_n - b_nf_n) + c_{n+1}(c_nh_n - b_nj_n) & 0 & c_ne_n - b_nf_n \end{pmatrix}$$

if the first expansion converges and $c_n \neq 0$.

This lemma is a generalization of Lemma 3.1.

It follows from the identity

$$\begin{pmatrix} a_n & 1 & 0 \\ b_n & 0 & 1 \\ c_n & 0 & 0 \end{pmatrix} A_n = c_n \begin{pmatrix} a_n & d_n & g_n \\ b_n & e_n & h_n \\ c_n & f_n & j_n \end{pmatrix}\begin{pmatrix} a_{n+1} & 1 & 0 \\ b_{n+1} & 0 & 1 \\ c_{n+1} & 0 & 0 \end{pmatrix}.$$

It is clear that with the exception of $\begin{pmatrix}1\\0\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\\0\end{pmatrix}$ the 1-convergents and the 2-convergents of the second expansion are non-zero multiples of the 1-convergents of the first expansion, while the 3rd column convergents of the second expansion are the 2nd column convergents of the first expansion .

If we apply this transformation again to the expansion $K\prod_{n=1}^{\infty} A_n$ the matrices we obtain are all of the form $\begin{pmatrix} p_n & s_n & 0 \\ q_n & 0 & t_n \\ r_n & 0 & 0 \end{pmatrix}$ and the convergents (except for a finite number) are the 1-convergents of $K\prod_{n=1}^{\infty} \begin{pmatrix} a_n & d_n & g_n \\ b_n & e_n & h_n \\ c_n & f_n & j_n \end{pmatrix}$. (or multiples of the 1-convergents).

Lemma 3.3

If one of the following expansions converges, and $\{c_n\}$ is a sequence of non-zero numbers, then

$$K\begin{pmatrix} c_0 & 0 \\ 0 & 1 \end{pmatrix}\prod_{n=0}^{\infty}\begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix} \sim K\prod_{n=0}^{\infty}\begin{pmatrix} p_n c_n & r_n c_n c_{n+1} \\ q_n & s_n c_{n+1} \end{pmatrix}.$$

This transformation is essentially Satz 1.2 of Perron [12, p.5].

Proof

$$c_{n+1}\begin{pmatrix} c_n & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} p_n & r_n \\ q_n & s_n \end{pmatrix} = \begin{pmatrix} p_n c_n & r_n c_n c_{n+1} \\ q_n & s_n c_{n+1} \end{pmatrix}\begin{pmatrix} c_{n+1} & 0 \\ 0 & 1 \end{pmatrix}.$$

The convergents of one expansion are non-zero multiples of the convergents of the other (except for the isolated convergents $\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}c_0\\0\end{pmatrix}$ of the first expansion).

An example of a generalization of Lemma 3.3 to 3×3 K-products is

Lemma 3.4

If $c_0, c_1, c_2, \cdots$ is a sequence of non-zero constants, and either expansion converges, then

$$K\begin{pmatrix} c_0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \prod_{n=0}^{\infty} \begin{pmatrix} p_n & s_n & x_n \\ q_n & t_n & y_n \\ r_n & u_n & z_n \end{pmatrix} \sim K \prod_{n=0}^{\infty} \begin{pmatrix} c_n p_n & c_n c_{n+1} s_n & c_n c_{n+1} x_n \\ q_n & c_{n+1} t_n & c_{n+1} y_n \\ r_n & c_{n+1} u_n & c_{n+1} z_n \end{pmatrix}.$$

The proof is clear in view of the proof of Lemma 3.3.

A transformation which produces a new expansion whose 1-convergents are a certain sub-sequence of the 1-convergents of the original expansion, is called in the case of continued fractions, a contraction (see Perron [12, pp.10–14]). Contractions for 2×2 K-expansions may be accomplished by using Lemmas 1.3 and 3.1. We will not prove this but will give as an illustration Lemma 3.5 (given in Perron [12 , pp. 13–14]).

Lemma 3.5

If the left hand side exists

$$K\prod_{n=1}^{\infty} \begin{pmatrix} x_n & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \sim K\begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix}\begin{pmatrix} ax_1+b & 1 \\ 1 & 0 \end{pmatrix}\prod_{n=2}^{\infty}\begin{pmatrix} ax_n+b+c & 1 \\ \Delta & 0 \end{pmatrix}$$

where $\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$, $\det\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \Delta = (-1)^k$,

and the a_k are positive integers.

Proof

By Lemma 1.3 the left hand side is equivalent to

$$K \prod_{n=1}^{\infty} \begin{pmatrix} ax_n+b & cx_n+d \\ a & c \end{pmatrix}$$

It is clear that $a \neq 0$ and so, using Lemma 3.1 the left hand side is equivalent to

$$K \begin{pmatrix} ax_1+b & 1 \\ a & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} ax_n+b+c & 1 \\ \Delta & 0 \end{pmatrix}$$

$$\sim K \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} ax_1+b & 1 \\ 1 & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} ax_n+b+c & 1 \\ \Delta & 0 \end{pmatrix}.$$

It is clear from the proof that, except initially, the convergents of the right hand expansion of Lemma 3.5 are the $n(k+1)$ th 1-convergents (for $n=1,2,3,\ldots$) of the left hand expansion of Lemma 3.5.

We now give as an application of Theorem 3.2 and Lemma 3.1 a proof of the transformation of Bauer and Muir (see §7 of Perron [12, pp.25-30]).

Lemma 3.6

If $a_0, a_1, a_2, \ldots ; b_0, b_1, b_2, \ldots ; r_0, r_1, r_2, \ldots$ are positive sequences and $\varphi_n = a_n - r_{n-1}(b_n + r_n) \neq 0$, then

$$K\begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \prod_{n=1}^{\infty} \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix} \sim K \begin{pmatrix} b_0+r_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_1+r_1 & 1 \\ \varphi_1 & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} (b_n+r_n)\varphi_{n-1} - r_{n-2}\varphi_n & \varphi_{n-1} \\ a_{n-1}\varphi_n & 0 \end{pmatrix}$$

provided the left hand side exists.

Proof

From the identity

$$\begin{pmatrix} 1 & 0 \\ r_{n-1} & 1 \end{pmatrix} \begin{pmatrix} b_n+r_n & 1 \\ \varphi_n & -r_{n-1} \end{pmatrix} = \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r_n & 1 \end{pmatrix}$$

and Theorem 3.2 we have, since the matrices $\begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ r_n & 1 \end{pmatrix}$ are positive, that $K \prod_{n=1}^{\infty} \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix} \sim K \begin{pmatrix} 1 & 0 \\ r_0 & 1 \end{pmatrix} \prod_{n=1}^{\infty} \begin{pmatrix} b_n+r_n & 1 \\ \varphi_n & -r_{n-1} \end{pmatrix}$.

Now by Lemma 3.1, since $\varphi_n \neq 0$

$$K \prod_{n=1}^{\infty} \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix} \sim K \begin{pmatrix} 1 & 0 \\ r_0 & 1 \end{pmatrix} \begin{pmatrix} b_1+r_1 & 1 \\ \varphi_1 & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} (b_n+r_n)\varphi_{n-1} - r_{n-2}\varphi_n & \varphi_{n-1} \\ a_{n-1}\varphi_n & 0 \end{pmatrix}$$

and hence

$$K \begin{pmatrix} b_0 & 1 \\ 1 & 0 \end{pmatrix} \prod_{n=1}^{\infty} \begin{pmatrix} b_n & 1 \\ a_n & 0 \end{pmatrix} \sim K \begin{pmatrix} b_0+r_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_1+r_1 & 1 \\ \varphi_1 & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} (b_n+r_n)\varphi_{n-1} - r_{n-2}\varphi_n & \varphi_{n-1} \\ a_{n-1}\varphi_n & 0 \end{pmatrix}.$$

Perron gives some alternative conditions under which this transformation is valid.

4. Two General K-expansions

4.1 The continued fraction process

If $a_0, a_1, a_2, \ldots$ is a sequence of positive integers, from Theorem 2.1 the K-product,

$$K \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \cdots \cdot ,$$

converges to a limit (α_i) for which $\alpha_1 \geqslant \alpha_2 = 1$.

An expansion of this type is called a regular (or simple) continued fraction expansion of α_1. In Theorem 1.1 we have shown the equivalence of this definition and the usual one, given say in Hardy and Wright [7, Chapter X]. (It is usual to allow a_0 to be negative or zero, but for simplicity we have excluded this case.)

Some important (and quite standard) facts concerning the regular continued fraction process are contained in

Theorem 4.1

(1) To any real irrational number $\alpha > 1$, there corresponds a sequence $a_0, a_1, a_2, \ldots$ of positive integers such that

$$\begin{pmatrix} \alpha \\ 1 \end{pmatrix} \sim K \prod_{n=0}^{\infty} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

(2) Furthermore, to any sequence $a_0, a_1, a_2, \ldots$ of positive integers there corresponds a real irrational number $\alpha > 1$ such that

$$\begin{pmatrix} \alpha \\ 1 \end{pmatrix} \sim K \prod_{n=0}^{\infty} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

(3) The correspondence between the real numbers α and the sequences $a_0, a_1, a_2, \ldots$ is one-to-one.

Proof of (1)

If $\alpha > 1$ is an irrational number, it is easy to see that there exists an a_0 such that

$$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim \begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}\alpha_1\\1\end{pmatrix}$$

where α_1 is greater than 1 and is irrational.

For this means that

$$\alpha = a_0 + \frac{1}{\alpha_1}$$

so $[\alpha]$ (the integral part of α) is the appropriate integer a_0.

In this way we can choose a sequence $a_0, a_1, a_2, \ldots$ of positive integers such that

$$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim \begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix}\cdots\cdots\begin{pmatrix}a_{n-1} & 1\\1 & 0\end{pmatrix}\begin{pmatrix}\alpha_n\\1\end{pmatrix}$$

for all n , where α_n is irrational and $\alpha_n > 1$.

Now as we have mentioned

$$K\prod_{n=0}^{\infty}\begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix}$$

converges, and hence by Theorem 3.2

$$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim \lim_{n\to\infty} K\begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\cdots\begin{pmatrix}a_{n-1} & 1\\1 & 0\end{pmatrix}\begin{pmatrix}\alpha_n\\1\end{pmatrix} \sim K\prod_{n=0}^{\infty}\begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix}.$$

Proof of (2)

Theorem 2.1 assures us of the existence of a real number α such that

$$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim K\prod_{n=0}^{\infty}\begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix}.$$

We must prove that this α is greater than 1 and irrational.

Now if α_1 is defined by

$$\begin{pmatrix}\alpha_1\\1\end{pmatrix} \sim K \prod_{n=1}^{\infty} \begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix}$$

a consequence of Theorem 2.1 is that $\alpha_1 \geqslant 1$.

Then $\begin{pmatrix}\alpha\\1\end{pmatrix} \sim \begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}\alpha_1\\1\end{pmatrix}$ and hence $\alpha = a_0 + \frac{1}{\alpha_1}$.

Thus $1 \leqslant a_0 < \alpha \leqslant a_0+1$.

To prove that α is irrational we note that if

$$\begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix} \cdots \begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix} = \begin{pmatrix}p_n & p_{n-1}\\q_n & q_{n-1}\end{pmatrix} = P_n$$

then

$$\alpha = \frac{\alpha_n p_n + p_{n-1}}{\alpha_n q_n + q_{n-1}} \quad \text{where} \quad \begin{pmatrix}\alpha_n\\1\end{pmatrix} \sim \prod_{k=n+1}^{\infty} \begin{pmatrix}a_k & 1\\1 & 0\end{pmatrix}.$$

Hence $|p_n - \alpha q_n| = \dfrac{1}{q_n\alpha_n + q_{n-1}} = \varepsilon_n \to 0$ as $n \to \infty$.

Suppose α is rational. Then there exist integers a, b not both zero such that $a - \alpha b = 0$. Since the rank of P_n is 2, either $\det\begin{pmatrix}a & p_n\\b & q_n\end{pmatrix}$ or $\det\begin{pmatrix}a & p_{n-1}\\b & q_{n-1}\end{pmatrix}$ is not zero. Without loss of generality we suppose the first. Then from the equations

$$a - \alpha b = 0$$

$$p_n - \alpha q_n = \varepsilon_n$$

by Cramer's rule we have

$$-\det\begin{pmatrix}a & b\\p_n & q_n\end{pmatrix}\alpha = \det\begin{pmatrix}a & 0\\p_n & \varepsilon_n\end{pmatrix} = a\,\varepsilon_n \to 0 \text{ as } n \to \infty.$$

This is impossible however, since $\det\begin{pmatrix}a & b\\p_n & q_n\end{pmatrix}$ is a non-zero integer.

Proof of (3)

Suppose two sequences $a_0, a_1, \cdots, a_n, \cdots$ and $a_0, a_1, \cdots, b_n, \cdots$ agree for the first n terms, and that

$$K\begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix}\cdots\begin{pmatrix}a_n & 1\\1 & 0\end{pmatrix}\cdots \sim K\begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\cdots\begin{pmatrix}b_n & 1\\1 & 0\end{pmatrix}\cdots.$$

Then by Lemma 1.1

$$K\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_{n+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \sim K\begin{pmatrix} b_n & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} b_{n+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots .$$

But as we proved in the proof of (2), if the common limit of the left and right hand sides is $\binom{\alpha}{1}$ then $a_n < \alpha \leqslant a_n+1$, $b_n < \alpha \leqslant b_n+1$ and so $a_n = b_n$.

Remark 4.1

This theorem is already of considerable interest in respect to the second problem mentioned in the introduction; for the question of whether integral vector solutions $\underset{\sim}{x} \neq \underset{\sim}{o}$ exist for the equation, $\Lambda \underset{\sim}{x} = \underset{\sim}{o}$, when $\Lambda = (\alpha, 1)$ is the question of whether α is rational. Theorem 4.1 provides a criterion for the irrationality of α. In chapter 5 we shall obtain regular continued fraction expansions of several interesting numbers, and their irrationality will thus be demonstrated.

It is convenient to introduce here a definition pertaining to the second problem of the introduction.

Definition 4.1

If p is the largest number of linearly independent integral vector solutions of

$$\underset{\sim}{x} \cdot \underset{\sim}{\alpha} = \underset{\sim}{o} \qquad (\underset{\sim}{x} \cdot \underset{\sim}{\alpha} = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_m\alpha_m)$$

we call p the index of $\underset{\sim}{\alpha}$.

An immediate observation on this definition is that the index of $\underset{\sim}{\alpha}$ and $A\underset{\sim}{\alpha}$ are the same if A is a non-singular integral matrix.

4.2 A Generalization of the Regular Continued Fraction Process

In this section we consider 3×3 K-products of the following types:

(I) $K_1 A_1 A_2 A_3 \cdots,$

and (II) $K_1 A_1 A_2 \cdots A_n B C_1 C_2 C_3 \cdots$ (or $K_1 B C_1 C_2 C_3 \cdots$),

where the matrices A, B and C belong to the special types given below. (a and b are understood to be integers subject to the stated restrictions.)

Either (1) $$A = \begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad a \geq b \geq 1,$$

or (2) $$A = \begin{pmatrix} a & 0 & 1 \\ b & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad a > b \geq 1.$$

Either (3) $$B = \begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad a \geq b > 1,$$

or (4) $$B = \begin{pmatrix} a & 0 & 1 \\ b & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad a > b \geq 1,$$

or (5) $$B = \begin{pmatrix} a & 1 & 1 \\ b & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad a > b \geq 1.$$

(6) $$C = \begin{pmatrix} a & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad a \geq 1.$$

Such expansions we call G-expansions. The interest in these expansions lies in the following analogue of Theorem 4.1.

Theorem 4.2

(1) If α, β are real numbers, $\alpha > \beta > 1$, and the index of $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ is less than 2, then $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ is the limit of a G-expansion.

(2) Every G-expansion converges to a vector $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ with index less than 2 and $\alpha > \beta > 1$.

(3) The correspondence between G-expansions and real vectors $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ with index less than 2 and $\alpha > \beta > 1$, is one to one.

The methods of proof are similar to those of Theorem 4.1 but the situation is much more complicated.

Proof of (1)

As in the proof of Theorem 4.1 we inductively derive the expansion corresponding to $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$, but here there are six different situations that can arise.

Let $\alpha > \beta > 1$, and the index of $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ be less than 2.

(i) If $\alpha - [\alpha] > \beta - [\beta] > 0$ then

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 1 & 0\\ [\beta] & 0 & 1\\ 1 & 0 & 0\end{pmatrix} \begin{pmatrix}\alpha_1\\ \beta_1\\ 1\end{pmatrix}$$

where $\alpha_1 > \beta_1 > 1$ and evidently $[\alpha] \geqslant [\beta] \geqslant 1$.

(ii) If $\beta - [\beta] > \alpha - [\alpha] > 0$ then

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 0 & 1\\ [\beta] & 1 & 0\\ 1 & 0 & 0\end{pmatrix} \begin{pmatrix}\alpha_1\\ \beta_1\\ 1\end{pmatrix}$$

where $\alpha_1 > \beta_1 > 1$, and evidently $[\alpha] > [\beta] \geqslant 1$.

(iii) If $\alpha-[\alpha]=0$, then since the index of $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ is less than 2, $\beta-[\beta]>0$ and

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 0 & 1\\ [\beta] & 1 & 0\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}\alpha_1\\ 1\\ 0\end{pmatrix}$$

where $\alpha_1>1$, and evidently $[\alpha]>[\beta]\geqslant 1$.

(iv) If $\beta-[\beta]=0$, then as in (iii) $\alpha-[\alpha]>0$, and

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 1 & 0\\ [\beta] & 0 & 1\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}\alpha_1\\ 1\\ 0\end{pmatrix}$$

where $\alpha_1>1$ and evidently $[\alpha]\geqslant[\beta]>1$.

(v) If $\alpha-[\alpha]=\beta-[\beta]$, then both quantities must be greater than zero, and

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 1 & 1\\ [\beta] & 1 & 0\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}\alpha_1\\ 1\\ 0\end{pmatrix}$$

where $\alpha_1>1$ and evidently $[\alpha]>[\beta]\geqslant 1$.

If $\alpha>1$ and $\begin{pmatrix}\alpha\\ 1\\ 0\end{pmatrix}$ has index less than 2, then

(vi)

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim \begin{pmatrix}[\alpha] & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\alpha_1\\ 1\\ 0\end{pmatrix}$$

where $\alpha_1>1$ and evidently $[\alpha]\geqslant 1$.

In all cases the vectors $\begin{pmatrix}\alpha_1\\ \beta_1\\ 1\end{pmatrix}$ and $\begin{pmatrix}\alpha_1\\ 1\\ 0\end{pmatrix}$ also have index less than 2.

Using these six facts, we see that for any $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$ satisfying the conditions of the theorem we can choose a sequence of matrices $D_1, D_2, D_3, \ldots$ such that $K_1 D_1 D_2 D_3 \ldots$ is a G-expansion and

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim D_1 D_2 D_3 \cdots D_n \begin{pmatrix}\alpha_n\\ \beta_n\\ \gamma_n\end{pmatrix}$$

where either $\alpha_n > \beta_n > \gamma_n = 1$ or $\alpha_n > \beta_n = 1$, $\gamma_n = 0$.
Now if $K_1 D_1 D_2 D_3 \ldots$ is a G-expansion of type I, it follows by Theorem 2.2 and Remark 2.1 that $K_1 D_1 D_2 \cdots$ converges and
$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim K_1 D_1 D_2 D_3 \ldots\ldots$

If $K_1 D_1 D_2 D_3 \ldots$ is a G-expansion of type II, its convergence follows by Theorem 2.1 since a matrix of type 6 is a direct sum of a 2×2 matrix and a 1×1 matrix and for the convergence of 1-convergents we need only consider the 2×2 submatrix.
Then using Theorem 3.2, again we have

$$\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \sim K_1 D_1 D_2 D_3 \ldots .$$

<u>Proof of 2</u>

As we have noted in the proof of part (1), every G-expansion converges. From Theorems 2.2 and 4.1 we obtain the further information that

$$K_1 A_1 A_2 A_3 \ldots \sim \begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \text{ where } \alpha \geqslant \beta \geqslant 1,$$

and $K_1 C_1 C_2 C_3 \ldots \sim \begin{pmatrix}\alpha\\ 1\\ 0\end{pmatrix}$ where $\alpha > 1$,

(where the matrices A are of types 1 or 2 and the matrices C are of type 6).

We make some further observations.

(a) If A is a matrix of type 1 or 2 and

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \sim A\begin{pmatrix}\alpha'\\ \beta'\\ 1\end{pmatrix}$$

then we can choose $\gamma = 1$ and then

$\alpha' \geqslant \beta' \geqslant 1$ implies that $\alpha \geqslant \beta > \gamma = 1$,

and $\alpha' \geqslant \beta' > 1$ implies that $\alpha > \beta > \gamma = 1$.

We will prove this when A is of type 1. Then

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \sim \begin{pmatrix}a & 1 & 0\\ b & 0 & 1\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}\alpha'\\ \beta'\\ 1\end{pmatrix} \sim \begin{pmatrix}a+\beta'/\alpha'\\ b+1/\alpha'\\ 1\end{pmatrix}$$

and since $a \geqslant b \geqslant 1$, the result follows. The proof when A is of type 2 is similar.

(b) If B is a matrix of type 3, 4, or 5 and

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \sim B\begin{pmatrix}\alpha'\\ 1\\ 0\end{pmatrix}$$

and $\alpha > 1$ then we can choose $\gamma = 1$ and

$$\alpha > \beta > \gamma = 1.$$

We will prove this when B is of type 5. Then

$$\begin{pmatrix}\alpha\\ \beta\\ \gamma\end{pmatrix} \sim \begin{pmatrix}a & 1 & 1\\ b & 1 & 0\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}\alpha'\\ 1\\ 0\end{pmatrix} \sim \begin{pmatrix}a+1/\alpha'\\ b+1/\alpha'\\ 1\end{pmatrix}$$

and since $a > b \geqslant 1$, the result follows.

The proofs when B is of type 3 or 4 are similar.

Now using these facts we can easily prove that if

$K_1 D_1 D_2 D_3 \ldots$ is a G-expansion then

$$K_1 D_1 D_2 D_3 \ldots \sim \begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix} \quad \text{with} \quad \alpha > \beta > 1.$$

Suppose $K_1 A_1 A_2 \ldots$, is of type I, then by our first remarks

$$K_1 A_3 A_4 \ldots \sim \begin{pmatrix} \alpha_3 \\ \beta_3 \\ 1 \end{pmatrix} \quad \text{with } \alpha_3 \geq \beta_3 \geq 1 \text{ which by}$$

(a) implies that $K_1 A_2 A_3 A_4 \cdots \sim \begin{pmatrix} \alpha_2 \\ \beta_2 \\ 1 \end{pmatrix}$ with $\alpha_2 \geq \beta_2 > 1$,

and again by (a) that $K_1 A_1 A_2 A_3 \ldots \sim \begin{pmatrix} \alpha_1 \\ \beta_1 \\ 1 \end{pmatrix}$ with $\alpha_1 > \beta_1 > 1$.

Suppose $K_1 A_1 A_2 \ldots A_n B C_1 C_2 \ldots$, is of type II, then by our first remarks

$$K_1 C_1 C_2 \ldots \sim \begin{pmatrix} \alpha'' \\ 1 \\ 0 \end{pmatrix} \text{ with } \alpha'' > 1.$$

Then by (b) $K_1 B C_1 C_2 \ldots \sim \begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix}$ with $\alpha' > \beta' > 1$,

and by (a) $K_1 A_1 A_2 \ldots A_n B C_1 C_2 \cdots \sim \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ with $\alpha > \beta > 1$.

We now prove that the index of the limit $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ of a G-expansion is less than 2.

If the G-expansion is of type II the result follows from Theorem 4.1 for $\begin{pmatrix} \alpha \\ 1 \\ 0 \end{pmatrix}$ has index 1 if α is irrational.

Otherwise let

$$A_1 A_2 \cdots A_n = \begin{pmatrix} p_n & u_n & x_n \\ q_n & v_n & y_n \\ r_n & w_n & z_n \end{pmatrix} = P_n .$$

Then $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim P_n \begin{pmatrix} \alpha_n \\ \beta_n \\ 1 \end{pmatrix}$ where $\alpha_n > \beta_n > 1$.

Put $\det \begin{pmatrix} q_n & v_n \\ r_n & w_n \end{pmatrix} = X_n$, $\det \begin{pmatrix} u_n & p_n \\ w_n & r_n \end{pmatrix} = Y_n$, $\det \begin{pmatrix} p_n & u_n \\ q_n & v_n \end{pmatrix} = Z_n$.

Then

$$|X_n\alpha + Y_n\beta + Z_n| = \frac{|X_n(\alpha_n p_n + \beta_n u_n + x_n) + Y_n(\alpha_n q_n + \beta_n v_n + y_n) + Z_n(\alpha_n r_n + \beta_n w_n + z_n)|}{(\alpha_n r_n + \beta_n w_n + z_n)}$$

$$= \frac{1}{(\alpha_n r_n + \beta_n w_n + z_n)} = \varepsilon_n,$$

(since $X_n p_n + Y_n q_n + Z_n r_n = 0$, $X_n u_n + Y_n v_n + Z_n w_n = 0$, $X_n x_n + Y_n y_n + Z_n z_n = \det P_n$).

Now it is clear that $\varepsilon_n \to 0$ as $n \to \infty$ and $\varepsilon_n \neq 0$.

Suppose there are two linearly independent vectors (a,b,c), (d,e,f) such that

$$a\alpha + b\beta + c = 0$$
$$d\alpha + e\beta + f = 0.$$

Now (X_n, Y_n, Z_n) cannot depend linearly on $(a,b,c),(d,e,f)$ or we would have $X_n\alpha + Y_n\beta + Z_n = 0$.

Hence $\det\begin{pmatrix} X_n & Y_n & Z_n \\ a & b & c \\ d & e & f \end{pmatrix} \neq 0$ for any n.

Consider the equations in $\alpha, \beta, 1$:

$$a\alpha + b\beta + c = 0$$
$$d\alpha + e\beta + f = 0$$
$$X_n\alpha + Y_n\beta + Z_n = \varepsilon_n.$$

By a familiar property of determinants,

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ X_n & Y_n & Z_n \end{pmatrix} = \det\begin{pmatrix} a & b & 0 \\ d & e & 0 \\ X_n & Y_n & \varepsilon_n \end{pmatrix} = \varepsilon_n \det\begin{pmatrix} a & b \\ d & e \end{pmatrix} \to 0 \text{ as } n \to \infty.$$

This is impossible since the left hand side is a non-zero integer for all n.

Proof of (3)

Suppose two different G-expansions have the same limit. Using Lemma 1.1 we can first remove the common initial matrices. Now, by studying the nature of the limit of a truncated G-expansion, we will show two such expansions with a common limit have the same initial matrix.

The information we require is the following (where the limit of the truncated G-expansion is $\begin{pmatrix}\alpha\\ \beta\\ 1\end{pmatrix}$):

(1) If the initial matrix is $\begin{pmatrix} a & 1 & 0\\ b & 0 & 1\\ 1 & 0 & 0\end{pmatrix}$ of type 1 or 3

then $[\alpha]=a$, $[\beta]=b$ and $\alpha-[\alpha]>\beta-[\beta]\geqslant 0$.

(2) If the initial matrix is $\begin{pmatrix} a & 0 & 1\\ b & 1 & 0\\ 1 & 0 & 0\end{pmatrix}$ of type 2 or 4

then $[\alpha]=a$, $[\beta]=b$ and $\beta-[\beta]>\alpha-[\alpha]\geqslant 0$.

(3) If the initial matrix is $\begin{pmatrix} a & 1 & 1\\ b & 1 & 0\\ 1 & 0 & 0\end{pmatrix}$ of type 5

then $[\alpha]=a$, $[\beta]=b$ and $\beta-[\beta]=\alpha-[\alpha]>0$.

If the limit of the truncated G-expansion is $\begin{pmatrix}\alpha\\ 1\\ 0\end{pmatrix}$, then the initial matrix is $\begin{pmatrix} [\alpha] & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}$.

We will prove (1). The proofs of the others are similar.

If $\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ is of type 1 and $D_1, D_2, D_3, \ldots$ are other matrices in the expansion, then

$$K_1 D_1 D_2 D_3 \ldots \sim \begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix} \quad \text{with} \quad \alpha' > \beta' > 1 .$$

Then $\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix} \sim \begin{pmatrix} a + \frac{\beta'}{\alpha'} \\ b + \frac{1}{\alpha'} \\ 1 \end{pmatrix}$

and we have our result.

If $\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ is of type 3 and $D_1, D_2, D_3, \ldots$ are the remaining matrices, then

$$K_1 D_1 D_2 D_3 \ldots \sim \begin{pmatrix} \alpha \\ 1 \\ 0 \end{pmatrix}, \quad \alpha > 1,$$

and $\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ 1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} a + 1/\alpha \\ b \\ 1 \end{pmatrix}$

and our result clearly follows.

It is clear from a consideration of these facts that two different G-expansions cannot have a common limit.

This completes the proof of Theorem 4.2.

5. Some Particular K-expansions

In this chapter we derive a variety of K-expansions, some with and some without number theoretic interest. Most of the expansions are well known, though some of the proofs are new. For a multitude of other expansions see Perron [12].

Expansion 5.1

$$\begin{pmatrix} \sqrt[3]{4} \\ \sqrt[3]{2} \\ 1 \end{pmatrix} \sim K \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \prod_{n=1}^{\infty} \begin{pmatrix} 3 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

We will require this expansion in chapter 6. It provides an illustration of the evaluation of periodic K-products.

Let $A_1, A_2, A_3, \ldots$ be the matrices which occur in this expansion.

By Theorem 2.2 the right hand side converges to $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ where $\alpha \geqslant \beta \geqslant 1$.

Then $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim A_1 A_2 \begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix} \sim A_1 A_2 A_3 \begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix}$ since $A_3 = A_4 = \ldots$.

Hence $\begin{pmatrix} \alpha' \\ \beta' \\ 1 \end{pmatrix} \sim A_2^{-1} A_1^{-1} \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim A_3^{-1} A_2^{-1} A_1^{-1} \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$

and thus $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim A_1 A_2 A_3^{-1} A_2^{-1} A_1^{-1} \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim \begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$;

which amounts to the equations

$$\alpha\beta = 2 \;, \quad \beta^2 = \alpha \;; \qquad \text{or} \qquad \beta^3 = 2 \;, \quad \alpha = \beta^2 .$$

But β is real, and there is only one real cube root of 2.

The following four sections amount to a derivation of the regular continued fraction expansion of $e^{2/k}$ and $\tan(\frac{1}{k})$ for k a positive integer. For other derivations of the expansion of $e^{2/k}$ see C.S. Davis [6] and Perron [11, pp. 123-125]. The expansion for $\tan(1/k)$ may be derived from Lambert's semi-regular expansion for $\tan(1/k)$ (see Perron [11, pp. 148-149] and [12, p. 157]).

Expansion 5.2

$$\begin{pmatrix} e^x \\ 1 \end{pmatrix} \sim K \prod_{m=0}^{\infty} \begin{pmatrix} (2m+1)+x & 2m+1 \\ (2m+1) & (2m+1)-x \end{pmatrix}$$

for all complex x.

Proof

We first show that

$$\prod_{m=1}^{n} \begin{pmatrix} (2m-1)+x & (2m-1) \\ (2m-1) & (2m-1)-x \end{pmatrix} = \begin{pmatrix} f_n(x) & g_n(x) \\ h_n(x) & k_n(x) \end{pmatrix} \qquad \ldots(1)$$

where $h_n(x) = g_n(-x)\,,\ k_n(x) = f_n(-x)$ ...(2)

and $$f_n(x) = \sum_{k=0}^{n} n\, c_{n,k}\, x^k$$

$$g_n(x) = \sum_{k=0}^{n} (n-k)\, c_{n,k}\, x^k$$

with $$c_{n,k} = \frac{(2n-k-1)!}{(n-k)!\,k!}\,.$$

The relations (2) follow immediately from the observation that the left hand side of (1) is unchanged on interchanging rows and then columns of each of the matrices and replacing x by $-x$.

We now proceed by induction on n. The result is clearly true for $n=1$, and we assume it true for some $n \geqslant 1$. To prove it true for $n+1$ it suffices, in view of (2), to show that

$$(2n+1)\{f_n(x)+g_n(x)\}+xf_n(x) = f_{n+1}(x),$$
$$(2n+1)\{f_n(x)+g_n(x)\}-xg_n(x) = g_{n+1}(x),$$

and it is easily verified that these relations hold.

To establish the expansion it is sufficient to prove that

$$\frac{f_n(x)}{n(n+1)\cdots(2n-1)} \to e^{\frac{1}{2}x}, \quad \frac{g_n(x)}{n(n+1)\cdots(2n-1)} \to e^{\frac{1}{2}x}.$$

We prove the first of these. Using the expression for $f_n(x)$, we have for all complex x

$$\frac{f_n(x)}{n(n+1)\cdots(2n-1)} = 1 + \sum_{k=1}^{n} \frac{n(n-1)\cdots(n-k+1)}{(2n-1)(2n-2)\cdots(2n-k)} \frac{x^k}{k!}$$

$$= 1 + \sum_{k=1}^{n} \frac{(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})}{(1-\frac{1}{2n})(1-\frac{2}{2n})\cdots(1-\frac{k}{2n})} \frac{(\frac{1}{2}x)^k}{k!}$$

$$= 1 + \sum_{k=1}^{n} a_{n,k} \frac{(\frac{1}{2}x)^k}{k!}, \text{ say.}$$

Clearly $a_{n,k} \to 1$ as $n \to \infty$ for fixed k, and also

$$a_{n,k} < \frac{1}{(1-\frac{k}{2n})^k} \leqslant \frac{1}{(1-\frac{1}{2})^k} = 2^k,$$

so the result stated follows from Tannery's theorem [2, p. 136]. The second follows from the first and the relation

$$(2n+1)\{f_n(x)+g_n(x)\} + xf_n(x) = f_{n+1}(x)$$

(if we divide by $(n+1)(n+2)\cdots(2n+1)$ and let $n \to \infty$).

We now deduce some regular continued fraction expansions from Expansion 5.2.

We use a shorter notation for regular continued fractions.

For $$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim K\begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_2 & 1\\1 & 0\end{pmatrix}\cdots$$

we write $\alpha = [a_0, a_1, a_2, \dots]$;

and, for example, by $[2, \overline{1, 2n, 1}]_{n=1}^{\infty}$ we shall mean $[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, \dots]$

<u>Expansions 5.3</u>

$$e^{1/k} = [\overline{1, (2n+1)k-1, 1}]_{n=0}^{\infty} \qquad \text{where } k>1 \text{ is an integer,}$$

and $$e = [2, \overline{1\ 2n, 1}]_{n=1}^{\infty}.$$

<u>Proof</u>

If we put $x = 1/k$ in Expansion 5.2, with integral $k>0$, and use Lemma 1.2 we obtain

$$\begin{pmatrix}e^{1/k}\\1\end{pmatrix} \sim K\prod_{n=0}^{\infty}\begin{pmatrix}(2n+1)k+1 & (2n+1)k\\(2n+1)k & (2n+1)k-1\end{pmatrix}.$$

Using the result

$$\begin{pmatrix}a+1 & a\\a & a-1\end{pmatrix} = \begin{pmatrix}1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a-1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\1 & 0\end{pmatrix}$$

with $a = (2n+1)k$, this gives

$$\begin{pmatrix}e^{1/k}\\1\end{pmatrix} \sim K\prod_{n=1}^{\infty}\begin{pmatrix}1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}(2n+1)k-1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\1 & 0\end{pmatrix},$$

which is a regular continued fraction if $k>1$.

If $k=1$, we have

$$e = [1, 0, 1, \overline{1, 2n, 1}]_{n=1}^{\infty}$$

$$= [2, \overline{1, 2n, 1}]_{n=1}^{\infty} \qquad \left(\text{since } \begin{pmatrix}1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\1 & 0\end{pmatrix} = \begin{pmatrix}2 & 1\\1 & 0\end{pmatrix}\right).$$

Expansions 5.4

$$e^{2/k} = \left[1, \overline{\tfrac{1}{2}\{(6n+1)k-1\}, 6(2n+1)k, \tfrac{1}{2}\{(6n+5)k-1\}, 1}\right]_{n=0}^{\infty}$$

for an odd integral $k>0$, and

$$e^{2} = \left[7, \overline{3n+2, 1, 1, 3n+3, 6(2n+3)}\right]_{n=0}^{\infty}$$

Proof

If we put $x = 2/k$ in Expansion 5.2 we obtain

$$\begin{pmatrix} e^{2/k} \\ 1 \end{pmatrix} \sim K \prod_{\nu=0}^{\infty} \begin{pmatrix} (2\nu+1)k+2 & (2\nu+1)k \\ (2\nu+1)k & (2\nu+1)k-2 \end{pmatrix} .$$

We may transform the product of three successive factors in this expression, given by $\nu = 3n, 3n+1, 3n+2,$ into a form appropriate when k is an odd integer. We observe that

$$\begin{aligned} \begin{pmatrix} a+2 & a \\ a & a-2 \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \tfrac{1}{2}(a-1) & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix} \\ &= \tfrac{1}{2} \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 2a & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} \tfrac{1}{2}(a-1) & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \end{aligned}$$

and apply these factorizations in this order to the three matrices specified, noting that

$$\begin{pmatrix} 2 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} = 4 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This yields

$$\begin{pmatrix} (6n+1)k+2 & (6n+1)k \\ (6n+1)k & (6n+1)k-2 \end{pmatrix} \begin{pmatrix} (6n+3)k+2 & (6n+3)k \\ (6n+3)k & (6n+3)k-2 \end{pmatrix} \begin{pmatrix} (6n+5)k+2 & (6n+5)k \\ (6n+5)k & (6n+5)k-2 \end{pmatrix}$$

$$= 8 \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \tfrac{1}{2}\{(6n+1)k-1\} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 6(2n+1)k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \tfrac{1}{2}\{(6n+5)k-1\} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

and hence the results stated (which are regular continued fractions for the specified k).

Expansion 5.5

$$\tan \tfrac{1}{k} = [0, k-1, \overline{1, (2n+1)k-2}\,]_{n=1}^{\infty}$$
$$\tan 1 = [\overline{1, 2n-1}\,]_{n=1}^{\infty}$$

Proof

Since $$\cot(\tfrac{1}{k}) - 1 = \frac{(i-1)e^{2i/k} + (i+1)}{e^{2i/k} - 1} \qquad (i^2 = -1),$$

application of Lemma 1.1 to Expansion 5.2 with $x = \frac{2i}{k}$ gives

$$\cot(\tfrac{1}{k}) - 1 \sim k \begin{pmatrix} i-1 & i+1 \\ 1 & -1 \end{pmatrix} \prod_{n=0}^{\infty} \begin{pmatrix} (2n+1)k+2i & (2n+1)k \\ (2n+1)k & (2n+1)k-2i \end{pmatrix}.$$

Since $$\begin{pmatrix} i-1 & i+1 \\ 1 & -1 \end{pmatrix} \cdot \tfrac{1}{2} \begin{pmatrix} a+2i & a \\ a & a-2i \end{pmatrix} = \begin{pmatrix} a-1 & a-2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} i-1 & i+1 \\ 1 & -1 \end{pmatrix}$$

and $$\begin{pmatrix} a-1 & a-2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} a-2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

Theorem 3.3 leads to the result $\cot(\frac{1}{k}) - 1 = [\overline{(2n+1)k-2, 1}\,]_{n=0}^{\infty}$.

It follows that $$\cot \tfrac{1}{k} = [k-1, \overline{1, (2n+1)k-2}\,]_{n=1}^{\infty},$$

and so $$\tan \tfrac{1}{k} = [0, k-1, \overline{1, (2n+1)k-2}\,]_{n=1}^{\infty};$$

these are regular continued fractions for integers $k > 0$ and $k > 1$ respectively.

Finally,

$$\tan \tfrac{1}{k} = [\overline{1, 2n-1}\,]_{n=1}^{\infty}.$$

Expansion 5.6

$$\begin{pmatrix}\log 2\\ \pi\\ 1\end{pmatrix} \sim K\begin{pmatrix}25 & 15 & 30\\ 104 & 80 & 120\\ 30 & 30 & 30\end{pmatrix}\prod_{n=2}^{\infty}\begin{pmatrix}-2n^2+n+4 & (2n+3)(n+2) & 0\\ 2n^2+7n+2 & 0 & (2n+3)(n+2)\\ (2n-1)n & 0 & 0\end{pmatrix}$$

Proof

This is an example of the following equivalence between a pair of series and a 3×3 K-product:

$$\begin{pmatrix}a_1+a_2+a_3+\cdots\\ b_1+b_2+b_3+\cdots\\ 1\end{pmatrix}\sim K\begin{pmatrix}a_1+a_2+a_3 & a_1+a_2 & a_1\\ b_1+b_2+b_3 & b_1+b_2 & b_1\\ 1 & 1 & 1\end{pmatrix}\prod_{n=2}^{\infty}\begin{pmatrix}\Delta_{1,n}+\Delta_{3,n} & \Delta_{1,n} & 0\\ -\Delta_{3,n}+\Delta_{2,n} & 0 & \Delta_{1,n}\\ -\Delta_{2,n} & 0 & 0\end{pmatrix}$$

where $\Delta_{1,n}=a_{n+1}b_n-a_nb_{n+1}$, $\Delta_{2,n}=a_{n+1}b_{n+2}-a_{n+2}b_{n+1}$, and $\Delta_{3,n}=a_{n+2}b_n-a_nb_{n+2}$, if the left hand side exists, and if $\Delta_{1,n}\neq 0$ for any n.

Our expansion is obtained by using the well-known series

$$\frac{1}{2}\log 2 = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{2n},$$

$$\frac{\pi}{4} = \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{(2n-1)}.$$

To prove the equivalence, it is a simple exercise in induction to show that if $S_n=\sum_{k=1}^{n}a_k$, $T_n=\sum_{k=1}^{n}b_k$,

then

$$\Delta_{1,2}\Delta_{1,3}\cdots\Delta_{1,n-2}\begin{pmatrix}S_n & S_{n-1} & S_{n-2}\\ T_n & T_{n-1} & T_{n-2}\\ 1 & 1 & 1\end{pmatrix}=\begin{pmatrix}S_3 & S_2 & S_1\\ T_3 & T_2 & T_1\\ 1 & 1 & 1\end{pmatrix}\prod_{k=2}^{n-2}\begin{pmatrix}\Delta_{1,k}+\Delta_{3,k} & \Delta_{1,k} & 0\\ -\Delta_{3,k}+\Delta_{2,k} & 0 & \Delta_{1,k}\\ -\Delta_{2,k} & 0 & 0\end{pmatrix}.$$

Expansion 5.7

If $a>0, b>0$ then

$$\begin{pmatrix} iJ_{\frac{b}{a}}\left(\frac{2i}{a}\right) \\ J_{\frac{b}{a}+1}\left(\frac{2i}{a}\right) \end{pmatrix} \sim K \prod_{k=1}^{\infty} \begin{pmatrix} (ak+b) & 1 \\ 1 & 0 \end{pmatrix}$$

where $J_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{r=0}^{\infty} \frac{(-1)^r \left(\frac{1}{2}z\right)^{2r}}{r!\,\Gamma(\nu+r+1)}$ is the Bessel function of the first kind of order ν.

For other proofs see Perron [12, pp. 103-104; p. 281].

Proof

As for Expansion 5.2, we can determine the convergents of this K-expansion. We have

$$\prod_{k=1}^{n} \begin{pmatrix} ak+b & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} f_n & f_{n-1} \\ g_n & g_{n-1} \end{pmatrix}$$

where $$f_n = \sum_{k=0}^{[\frac{n}{2}]} \binom{n-k}{k} \prod_{r=k+1}^{n-k} (ar+b)$$

and $$g_n = \sum_{k=0}^{[\frac{n-1}{2}]} \binom{n-k-1}{k} \prod_{r=k+2}^{n-k} (ar+b) .$$

(Empty products are taken to be 1 and empty sums 0.)

The proof by induction rests on the two identities (for $n>0$)

$$\{a(n+1)+b\}.f_n + f_{n-1} = f_{n+1}$$

$$\{a(n+1)+b\}g_n + g_{n-1} = g_{n+1}$$

which are easily verified.

We now use Tannery's Theorem to show that if $a>0, b>0$

$$K \prod_{k=1}^{\infty} \begin{pmatrix} ak+b & 1 \\ 1 & 0 \end{pmatrix} \sim \begin{pmatrix} \displaystyle\sum_{k=0}^{\infty} \frac{1}{k!\, a^k (a+b)(2a+b)\cdots(ak+b)} \\ \displaystyle\sum_{k=0}^{\infty} \frac{1}{k!\, a^k (a+b)(2a+b)\cdots(a(k+1)+b)} \end{pmatrix} \quad \ldots(3)$$

This formula, by comparison with the defining series for the Bessel functions, then yields the required expansion.

From

$$\frac{f_n}{(a+b)(2a+b)\cdots(an+b)} = \sum_{k=0}^{[\frac{n}{2}]} \frac{(n-k)(n-k-1)\cdots(n-2k+1)(a(k+1)+b)\cdots(a(n-k)+b)}{(a+b)(2a+b)\cdots(an+b)\ k!}$$

$$= \sum_{k=0}^{[\frac{n}{2}]} \frac{(1-\frac{k}{n})(1-\frac{k+1}{n})\cdots(1-\frac{2k-1}{n})}{(1-\frac{k-1}{n}+\frac{b}{an})\cdots(1+\frac{b}{an})\,(a+b)(2a+b)\cdots(ak+b)\,a^k k!}$$

$$= \sum_{k=0}^{[\frac{n}{2}]} u_{n,k}\,\frac{1}{a^k(a+b)(2a+b)\cdots(ak+b)\,k!}\ ,$$

Evidently $u_{n,k} \to 1$ as $n \to \infty$ for fixed k , and also

$$u_{n,k} < \frac{1}{(1-\frac{k-1}{n})^k} < \frac{1}{(1-\frac{1}{2})^k} = 2^k.$$

So Tannery's Theorem applies, and as $n \to \infty$

$$\frac{f_n}{(a+b)(2a+b)\cdots(an+b)} \to \sum_{k=0}^{\infty} \frac{1}{k!\,a^k\,(a+b)(2a+b)\cdots(ak+b)}$$

We can prove in a similar way that

$$\frac{g_n}{(a+b)(2a+b)\cdots(an+b)} \to \sum_{k=0}^{\infty} \frac{1}{k!\,a^k\,(a+b)(2a+b)\cdots(a(k+1)+b)}\ ,$$

and since it is clear that both of the limits are not zero when $a>0$, $b>0$, we have proved (3) and hence Expansion 5.7.

Expansion 5.8

If $a>0$, $a^2n^2+bn+c=a_n>0$, $a_{n+1}-a_n>0$ $(n=1,2,3,\ldots)$ then

$$\begin{pmatrix} \sum_{r=0}^{\infty} \frac{1}{a_1a_2\cdots a_{r+1}a^r r!} \\ \sum_{r=0}^{\infty} \frac{1}{a_1a_2\cdots a_{r+2}a^r r!} \\ \sum_{r=0}^{\infty} \frac{1}{a_1a_2\cdots a_{r+3}a^r r!} \end{pmatrix} \sim K \prod_{r=1}^{\infty} \begin{pmatrix} a_{r+1} & \frac{a_{r+2}-a_{r+1}}{a} & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Remark

This expansion is essentially a special case of a set of $n\times n$ K-expansions due to Todd [14] (who gives them in terms of recurrence relations). Notice that in this proof we do not obtain the convergents as we did in Expansions 5.2, 5.7.

Proof

We first show that if

$$G_n = \sum_{r=0}^{\infty} \frac{1}{a_1a_2\cdots a_{r+n}a^r r!} \qquad (n>0)$$

then

$$\begin{pmatrix} G_{n-1} \\ G_n \\ G_{n+1} \end{pmatrix} = \begin{pmatrix} a_n & \frac{a_{n+1}-a_n}{a} & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} G_n \\ G_{n+1} \\ G_{n+2} \end{pmatrix} \qquad \ldots(4)$$

This amounts to

$$\sum_{r=1}^{\infty} \frac{1}{a_1a_2\cdots a_{n+r-1}a^r r!} = \sum_{r=2}^{\infty} \frac{a_n+r(a_{n+1}-a_n)+a^2r(r-1)}{a_1a_2\cdots a_{n+r}a^r r!} + \frac{1}{a_1a_2\cdots a_{n-1}} + \frac{1}{a_1\cdots a_n a},$$

and our result follows from the fact that

$$a_n + r(a_{n+1}-a_n) + a^2r(r-1) = a_{n+r}.$$

From ( ) it obviously follows that

$$\begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} = \left\{ \prod_{r=2}^{n} \begin{pmatrix} a_r & \frac{a_{r+1}-a_r}{a} & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\} \begin{pmatrix} G_n \\ G_{n+1} \\ G_{n+2} \end{pmatrix}$$

It is also quite clear that

$$\begin{pmatrix} G_n \\ G_{n+1} \\ G_{n+2} \end{pmatrix} > \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \text{for all } n > 0 .$$

Thus if we can prove the convergence of

$$K \prod_{r=2}^{\infty} \begin{pmatrix} a_r & \frac{a_{r+1}-a_r}{a} & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} ,$$

then by Theorem 3.2 we will have the stated result.

Now consider the product

$$P_n = \prod_{r=2}^{n} \begin{pmatrix} c_r & d_r & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

Using the identity

$$\begin{pmatrix} c_n & d_n & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_{n+1} & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} c_n & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_{n+1} & 1 & 0 \\ d_n & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

we obtain

$$P_n \begin{pmatrix} c_{n+1} & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} c_2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \prod_{r=2}^{n} \begin{pmatrix} c_{r+1} & 1 & 0 \\ d_r & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} ,$$

and hence

$$P_n = \begin{pmatrix} c_2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left\{ \prod_{r=2}^{n-1} \begin{pmatrix} c_{r+1} & 1 & 0 \\ d_r & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right\} \begin{pmatrix} 1 & 0 & 0 \\ 0 & d_n & 1 \\ 0 & 1 & 0 \end{pmatrix} .$$

Hence, again using Theorem 3.2, the convergence of $K \prod_{r=2}^{\infty} \begin{pmatrix} c_r & d_r & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$,

follows from the convergence of $K \prod_{r=2}^{\infty} \begin{pmatrix} c_{r+1} & 1 & 0 \\ d_r & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, if $d_r \geq 0$ $(r \geq 2)$;

and for $c_n = a_n$, $d_n = \dfrac{a_{n+1} - a_n}{a}$ the convergence of this second K-product follows by Theorem 2.2. This completes the proof of Expansion 5.8. Notice that at the same time we have proved the expansion

$$\begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} \sim K \begin{pmatrix} a_2 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} a_{n+1} & 1 & 0 \\ \frac{a_{n+1} - a_n}{a} & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} .$$

If we specialize a_n to be $(3n-1)(3n-2)$, the functions on the left hand side of Expansion 5.8 can be expressed in terms of $e, e^{\omega}, e^{\omega^2}$ where $\omega \neq 1$ is a cube root of 1.

6. Best Approximations

Throughout this chapter we use Latin letters for integral vectors and Greek letters for real vectors.

6.1 General discussion

We shall now informally introduce the idea of 'best approximations' in relation to the first problem mentioned in the introduction. We shall suppose that $\Lambda, m(\underset{\sim}{x}), c(\underset{\sim}{x}), \varphi(t), \psi(t)$ have the meanings assigned to them in the introduction. Now the problem of finding a function φ is theoretically more simple than that of finding a function ψ, for the first only requires the knowledge of a certain sequence of vectors, whereas the second requires knowledge of all but a finite number of vectors. The solutions of both problems, and especially the second, would be facilitated if we could find the vectors $\underset{\sim}{x}$ for which $m(\underset{\sim}{x})$ is 'smallest' (in some sense) with respect to $c(\underset{\sim}{x})$. We would then know that for all other vectors $m(\underset{\sim}{x})$ is 'larger' and thus we would only have to estimate $m(\underset{\sim}{x})$ in terms of $c(\underset{\sim}{x})$ for these 'best' vectors in order to obtain a function ψ; and of course, the 'better' the vectors $\underset{\sim}{x}$ are, the better the function φ we can obtain by estimating $m(\underset{\sim}{x})$ for these vectors.

In this direction we make the following definition:

The vector $\underset{\sim}{y}$ is called a best approximation with respect to Λ if for all other vectors $\underset{\sim}{x} \neq \underset{\sim}{0}$ with $c(\underset{\sim}{x}) < c(\underset{\sim}{y})$ we have $m(\underset{\sim}{x}) > m(\underset{\sim}{y})$.

Now if the function $c(\underset{\sim}{x})$ has the property that for any real number K there are only a finite number of vectors with $c(\underset{\sim}{x}) < K$, we can order the complete set of best approximations in a sequence $\{\underset{\sim}{x}_n\}$ so that

$$c(\underset{\sim}{x}_1) \leq c(\underset{\sim}{x}_2) \leq c(\underset{\sim}{x}_3) \leq c(\underset{\sim}{x}_4) \leq \ldots$$

Furthermore we can prove

<u>Lemma 6.1</u>

Suppose there exists a function $\psi(t)$ which decreases steadily as t increases, such that

$$m(\underset{\sim}{x}) > \psi\{c(\underset{\sim}{x})\}$$

for all best approximations, then there are no vectors $\underset{\sim}{x} \neq \underset{\sim}{0}$ such that

$$m(\underset{\sim}{x}) \leq \psi\{c(\underset{\sim}{x})\}.$$

<u>Proof</u>

Suppose that S, the set of all vectors $\underset{\sim}{x} \neq 0$, with $m(\underset{\sim}{x}) < \psi\{c(\underset{\sim}{x})\}$, is not empty. Then belonging to S, there is a vector $\underset{\sim}{y}$ with $c(\underset{\sim}{y}) \leq c(\underset{\sim}{x})$ for all $\underset{\sim}{x} \in S$.
The existence of $\underset{\sim}{y}$ follows from the nature of the function $c(\underset{\sim}{x})$.
Now obviously no member of S can be a best approximation. Hence there exists a vector with

$$c(\underset{\sim}{z}) < c(\underset{\sim}{y}) \quad \text{and} \quad m(\underset{\sim}{z}) \leq m(\underset{\sim}{y}).$$

Then by the property of $\underset{\sim}{y}$, $\underset{\sim}{z}$ does not belong to S, and hence

$$\psi\{c(\underset{\sim}{z})\} < m(\underset{\sim}{z}) \leq m(\underset{\sim}{y}) < \psi\{c(\underset{\sim}{y})\},$$

which is not compatible with the equation:

$c(\underset{\sim}{z}) < c(\underset{\sim}{y})$ since $\psi(t)$ is a decreasing function of t.

One obvious result of this lemma is that $m(\underset{\sim}{x}_n) \to 0$ as $n \to \infty$ if $\{x_n\}$ is the sequence of best approximations (since we can choose vectors for which $m(\underset{\sim}{x})$ is as small as we please).

In the following sections we prove a theorem (Theorem 6.1) which sometimes enables us to find complete sets of best approximations when Λ is of the form

$$\Lambda = \begin{pmatrix} \alpha_m & 0 & \dots & 0 & -\alpha_1 \\ 0 & \alpha_m & \dots & 0 & -\alpha_2 \\ 0 & 0 & \alpha_m \dots & 0 & -\alpha_3 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & \alpha_m & -\alpha_{m-1} \end{pmatrix} \qquad (\alpha_m \neq 0)$$

For this special case we use rather different language than we have been using. Instead of 'best approximation with respect to Λ'

we say 'best approximation to $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m-1} \\ \alpha_m \end{pmatrix}$' and we say that $m(\underset{\sim}{x})$

measures how closely $\underset{\sim}{x}$ approximates $\underset{\sim}{\alpha}$, rather than saying that $m(\underset{\sim}{x})$ measures the size of $\Lambda\underset{\sim}{x}$, because in this case, the best approximations $\underset{\sim}{x}_n$ have the property that $K\underset{\sim}{x}_n \to \underset{\sim}{\alpha}$ as $n \to \infty$.

A further important point about the following sections is that the method of proof of Theorem 6.1 requires that we give a rather different definition of best approximations; however we shall relate it to the one just given.

6.2 Measures of Approximation

Suppose $\underset{\sim}{\alpha} = (\alpha_i)$, α_m (the last component) $\neq 0$, $\underset{\sim}{\alpha} > 0$. We will call a function, $M(\underset{\sim}{x})$, a measure of approximation of $\underset{\sim}{x}$ to $\underset{\sim}{\alpha}$ if M has the following properties:

I. $M(\underset{\sim}{x}) > 0$ for all $\underset{\sim}{x} \neq \underset{\sim}{0}$.

II. $M(\lambda \underset{\sim}{x}) = |\lambda|^k M(\underset{\sim}{x})$ for some k (where λ is any real number).

III. If $\{\underset{\sim}{x}_n\}$ is a sequence such that $M(\underset{\sim}{x}_n) \to 0$ as $n \to \infty$, then $\underset{\sim}{x}_n \to \underset{\sim}{\alpha}$ as $n \to \infty$.

IV. If $\{\underset{\sim}{x}_n\}$ is a sequence with $\underset{\sim}{x}_n = (x_{i,n}) \neq \underset{\sim}{0}$ such that

$$\det \begin{pmatrix} x_{i,n} & \alpha_i \\ x_{m,n} & \alpha_m \end{pmatrix} \to 0 \text{ as } n \to \infty \text{ for } i = 1, 2, \ldots, m-1, \text{ then}$$

$M(\underset{\sim}{x}_n) \to 0$ as $n \to \infty$. (This is the assumption, tacit throughout Section 6.1, that $m(\underset{\sim}{x}_n) \to 0$ if the components of $\Lambda \underset{\sim}{x}_n$ tend to zero.)

V. If $M(\underset{\sim}{x}) < M(\underset{\sim}{e}_i)$ $(i = 1, 2, \ldots, m)$ then either $\underset{\sim}{x} < \underset{\sim}{0}$, $\underset{\sim}{x} = \underset{\sim}{0}$, or $\underset{\sim}{x} > \underset{\sim}{0}$.

Two further properties depend on a definition.

Definition 6.1

$\underset{\sim}{y}$ is a best approximation to $\underset{\sim}{\alpha}$ (with respect to function $M(\underset{\sim}{x})$) if $\underset{\sim}{y} > \underset{\sim}{0}$ and if, when $\underset{\sim}{x}$ is any vector with $\underset{\sim}{0} < \underset{\sim}{x} < \underset{\sim}{y}$, we have $M(\underset{\sim}{x}) > M(\underset{\sim}{y})$.

It is clear that unit vectors are best approximations to any $\underset{\sim}{\alpha}$ with respect to any function $M(\underset{\sim}{x})$. Other best approximations we call proper.

We then have the two further properties of a measure of approximation.

VI. There exists a least proper best approximation $\underset{\sim}{a}$.

VII. If $M(\underset{\sim}{x}) \leqslant M(\underset{\sim}{y}) \leqslant M(\underset{\sim}{a})$ then either $\underset{\sim}{x} < \underset{\sim}{y}$ or $\underset{\sim}{x} = \underset{\sim}{y}$ or $\underset{\sim}{x} > \underset{\sim}{y}$.

We now give another definition of best approximation in the spirit of Section 6.1, and we shall show that the two definitions are almost equivalent.

Definition 6.2

$\underset{\sim}{y}$ is a best approximation to $\underset{\sim}{\alpha}$ if when $\underset{\sim}{x}$ is any vector such that $0 < \sum_{i=1}^{m} |x_i| < \sum_{i=1}^{m} |y_i|$, we have $M(\underset{\sim}{y}) < M(\underset{\sim}{x})$.

We shall now discuss the correspondence between Definition 6.1 and Definition 6.2.

Suppose $\{\underset{\sim}{x}_n\}$ are the best approximations according to Definition 6.2. Then $M(\underset{\sim}{x}_n) \to 0$ as $n \to \infty$, and hence for all n larger than some n_0, $M(\underset{\sim}{x}_n) < M(\underset{\sim}{e}_i)\ (i=1,2,\dots,m)$ and so $\underset{\sim}{x}_n > \underset{\sim}{0}$ or $-\underset{\sim}{x}_n > \underset{\sim}{0}$. Further it is clear that if $\underset{\sim}{x}_n$ is a best approximation then so is $-\underset{\sim}{x}_n$. If we consider the positive half of the pairs $\underset{\sim}{x}_n, -\underset{\sim}{x}_n\ (n > n_0)$ and look at the vectors $\underset{\sim}{y}$ such that $\underset{\sim}{0} < \underset{\sim}{y} \leqslant \underset{\sim}{x}_n$, then obviously $C(\underset{\sim}{y}) < C(\underset{\sim}{x}_n)$. Hence since $\underset{\sim}{x}_n$ is a best approximation according to Definition 6.2, $M(\underset{\sim}{y}) > M(\underset{\sim}{x}_n)$. Hence the positive vectors among the $\underset{\sim}{x}_n\ (n > n_0)$ are best approximations according to Definition 6.1.

Now consider the set of proper best approximations according to Definition 6.1. By properties VI and VII we can write them as a sequence $\{\underset{\sim}{x}_n\}$ such that

$$\underset{\sim}{a} = \underset{\sim}{x}_1 < \underset{\sim}{x}_2 < \underset{\sim}{x}_3 < \cdots .$$

Now by a very similar method to the proof of Lemma 6.1 one can prove that $M(\underset{\sim}{x}_n) \to 0$ as $n \to \infty$. Hence for all n greater than some n_1, $M(\underset{\sim}{x}_n) < M(\underset{\sim}{e}_i)$ $(i = 1, 2, \cdots, m)$. Consider, for each $n > n_1$, the vectors $\underset{\sim}{y} \neq \underset{\sim}{0}$ with

$$c(\underset{\sim}{y}) < c(\underset{\sim}{x}_n) \text{ and } M(\underset{\sim}{y}) \leq M(\underset{\sim}{x}_n) \quad \ldots (*)$$

Evidently $\varepsilon \underset{\sim}{y} > \underset{\sim}{0}$ where $\varepsilon = \pm 1$. Further $\underset{\sim}{0} < \varepsilon \underset{\sim}{y} < \underset{\sim}{x}_n$ since $M(\varepsilon \underset{\sim}{y}) = M(\underset{\sim}{y}) \leq M(\underset{\sim}{x}_n) \leq M(\underset{\sim}{a})$, and $c(\underset{\sim}{y}) < c(\underset{\sim}{x}_n)$. Clearly, since $\underset{\sim}{x}_n$ is a best approximation according to Definition 6.1, there are no vectors $\underset{\sim}{y}$ satisfying $(*)$. Thus we have shown that the best approximations according to Definition 6.1 are, apart from a finite number, best approximations according to Definition 6.2.

We now give some belated remarks on the definition of measure of approximation. Throughout the following, 'best approximation' means 'best approximation according to Definition 6.2.'

<u>Remarks 6.1</u>

(1) If M is a measure of approximation to $\underset{\sim}{\alpha}$ and $\underset{\sim}{\alpha} \sim \underset{\sim}{\beta}$, then M is a best approximation to $\underset{\sim}{\beta}$; the best approximations to $\underset{\sim}{\alpha}$ with measure M are precisely the best approximations to $\underset{\sim}{\beta}$ with measure M.

(2) If M is a measure of approximation to $\underset{\sim}{\alpha}$, then so is $M_1 = \lambda M$ if λ is a positive constant. We call two such measures equivalent, and write $M_1 \approx M$. The best approximations to $\underset{\sim}{\alpha}$ with measure M are precisely the best approximations to $\underset{\sim}{\alpha}$ with measure M_1.

Theorem 6.1

If (1) $\underset{\sim}{\alpha}_1 \sim K \prod_{k=1}^{\infty} A_k$ ...(1)

where $A_k = (\underset{\sim}{a}_k, \underset{\sim}{e}_1, \underset{\sim}{e}_2, \cdots \underset{\sim}{e}_{n-1})$ and $\underset{\sim}{e}_m \cdot \underset{\sim}{a}_k = 1$, $\underset{\sim}{e}_r \cdot \underset{\sim}{a}_k \neq 0 \; (r = 1, 2, \cdots, m-1)$

(2) $M_n(\underset{\sim}{x}) \approx M_1\left[\left(\prod_{k=1}^{n-1} A_k\right)\underset{\sim}{x}\right] \quad (n = 1, 2, 3, \cdots)$

are respectively measures of approximation to

$\underset{\sim}{\alpha}_n$ where $\underset{\sim}{\alpha}_n \sim K \prod_{k=n}^{\infty} A_k$,

and (3) $\underset{\sim}{a}_n$ is the least best approximation to $\underset{\sim}{\alpha}_n$ with measure M_n,

then the 1-convergents of expansion (1) are all the proper best approximations to $\underset{\sim}{\alpha}_1$ with measure M_1.

The proof is conveniently divided into two parts.

Part 1

We shall prove first that if $\underset{\sim}{x}$ is a best approximation to $\underset{\sim}{\alpha}_2$ with measure M_2, then $A_1\underset{\sim}{x}$ is a best approximation to $A_1\underset{\sim}{\alpha}_2 \sim \underset{\sim}{\alpha}_1$ with measure M_1

[It follows from this that $\left(\prod_{k=1}^{n} A_k\right)\underset{\sim}{e}_1 \; (n = 1, 2, 3, \ldots)$ are best approximations to $\underset{\sim}{\alpha}_1$ with measure M_1. For $\underset{\sim}{e}_1$ is a best approximation to $\underset{\sim}{\alpha}_{n+1}$ with measure M_{n+1}; hence $A_n \underset{\sim}{e}_1$ is a best approximation to $\underset{\sim}{\alpha}_n$ with measure M_n; and $A_{n-1}A_n\underset{\sim}{e}_1$ is a best approximation to $\underset{\sim}{\alpha}_{n-1}$ with measure M_{n-1}; and so on. Finally $A_1 A_2 \ldots A_n \underset{\sim}{e}_1$ is a best approximation to $\underset{\sim}{\alpha}_1$ with measure M_1.]

Proof of Part 1

If $\underset{\sim}{x}$ is a unit vector, Part 1 is evidently true, so we assume that $\underset{\sim}{x}$ is a proper best approximation. It is clear that $A_1\underset{\sim}{x} > \underset{\sim}{o}$. Suppose in contradiction to Part 1 that there exists a $\underset{\sim}{y}$ with

$$\underset{\sim}{o} < \underset{\sim}{y} < A_1\underset{\sim}{x} \quad \text{and} \quad M_1(\underset{\sim}{y}) \leqslant M_1(A_1\underset{\sim}{x});$$

then $\quad M_2(A_1^{-1}\underset{\sim}{y}) \leqslant M_2(\underset{\sim}{x})\,. \qquad \ldots(2)$

From (2) and property V we have either

$$A_1^{-1}\underset{\sim}{y} < \underset{\sim}{o} \quad \text{or} \quad A_1^{-1}\underset{\sim}{y} > \underset{\sim}{o}\,.$$

Now
$$A_1^{-1} = \begin{pmatrix} \underset{\sim}{e}_m' \\ \underset{\sim}{e}_1' - (\underset{\sim}{e}_1\cdot\underset{\sim}{a}_1)\underset{\sim}{e}_m' \\ \underset{\sim}{e}_2' - (\underset{\sim}{e}_2\cdot\underset{\sim}{a}_1)\underset{\sim}{e}_m' \\ \vdots \\ \underset{\sim}{e}_{m-1}' - (\underset{\sim}{e}_{m-1}\cdot\underset{\sim}{a}_1)\underset{\sim}{e}_m' \end{pmatrix},$$

Thus either $\underset{\sim}{e}_m\cdot\underset{\sim}{y} > \underset{\sim}{o}$ whence $A_1^{-1}\underset{\sim}{y} > \underset{\sim}{o}$, or $\underset{\sim}{e}_m\cdot\underset{\sim}{y} = \underset{\sim}{o}$

whence
$$A_1^{-1}\underset{\sim}{y} = \begin{pmatrix} 0 \\ \underset{\sim}{e}_1\cdot\underset{\sim}{y} \\ \vdots \\ \underset{\sim}{e}_{m-1}\cdot\underset{\sim}{y} \end{pmatrix} > \underset{\sim}{o}\,.$$

In either case we have $A_1^{-1}\underset{\sim}{y} > \underset{\sim}{o}$.

Further from (2) and VII, and the fact that $\underset{\sim}{x}$ is a proper best approximation to $\underset{\sim}{\alpha}_2$ with measure M_2, we have $A_1^{-1}\underset{\sim}{y} \geqslant \underset{\sim}{x} > \underset{\sim}{o}$. Multiplying this inequality by the positive matrix A_1, we have

$$\underset{\sim}{y} \geqslant A_1\underset{\sim}{x}\,.$$

But this conflicts with our supposition that $\underset{\sim}{y} < A_1\underset{\sim}{x}$ and we have our result.

Part 2

We prove in Part 2 that if $\underset{\sim}{x}_1$ is a best approximation to $\underset{\sim}{\alpha}_1$ with measure $\underset{\sim}{M}_1$, $\underset{\sim}{x}_1 > \underset{\sim}{a}_1$ and $\underset{\sim}{x}_1 = A_1 \underset{\sim}{x}_2$, then $\underset{\sim}{x}_2$ is a proper best approximation to $\underset{\sim}{\alpha}_2$ with measure M_2 (where $\underset{\sim}{\alpha}_2 \sim A_1^{-1}\underset{\sim}{\alpha}_1$).

[This implies that any proper best approximation to $\underset{\sim}{\alpha}_1$ with measure M_1 is one of the 1-convergents of expansion (1). For if $\underset{\sim}{x}_1$ is a proper best approximation to $\underset{\sim}{\alpha}_1$ with measure M_1, then either $\underset{\sim}{x}_1 = \underset{\sim}{a}_1$ or $\underset{\sim}{x}_1 = A_1\underset{\sim}{x}_2$ and $\underset{\sim}{x}_2$ is a proper best approximation to $\underset{\sim}{\alpha}_2$ with measure M_2, and this implies that either $\underset{\sim}{x}_2 = \underset{\sim}{a}_2$ or $\underset{\sim}{x}_1 = A_1 A_2 \underset{\sim}{x}_3$ where $\underset{\sim}{x}_3$ is a proper best approximation to $\underset{\sim}{\alpha}_3$ with measure M_3; and so on. But this process must come to an end.

For

$$\underset{\sim}{x}_{n+1} = \begin{pmatrix} \underset{\sim}{e}_m \cdot \underset{\sim}{x}_n \\ \underset{\sim}{e}_1 \cdot \underset{\sim}{x}_n - (\underset{\sim}{e}_1 \cdot \underset{\sim}{a}_n)(\underset{\sim}{e}_m \cdot \underset{\sim}{x}_n) \\ \vdots \\ \underset{\sim}{e}_{m-1} \cdot \underset{\sim}{x}_n - (\underset{\sim}{e}_{m-1} \cdot \underset{\sim}{a}_n)(\underset{\sim}{e}_m \cdot \underset{\sim}{x}_n) \end{pmatrix}$$

and since $\underset{\sim}{e}_i \cdot \underset{\sim}{x}_n > 0$, $\underset{\sim}{e}_i \cdot \underset{\sim}{a}_n > 0$ ($\underset{\sim}{x}_n$ is a proper best approximation, and the least proper best approximation $\underset{\sim}{a}_n$ has $\underset{\sim}{e}_i \cdot \underset{\sim}{a}_n \neq 0$), we have

$$0 < \sum_{i=1}^{m} \underset{\sim}{e}_i \cdot \underset{\sim}{x}_{n+1} < \sum_{i=1}^{m} \underset{\sim}{e}_i \cdot \underset{\sim}{x}_n \, . \Big]$$

Proof of Part 2

We must first show that $\underset{\sim}{x}_2 = A_1^{-1}\underset{\sim}{x}_1 > \underset{\sim}{0}$. We have $M_1(\underset{\sim}{x}_1) < M_1(\underset{\sim}{a}_1)$ and hence $M_2(\underset{\sim}{x}_2) < M_2(\underset{\sim}{e}_1)$. Further $M_1(\underset{\sim}{a}_1) < M_1(\underset{\sim}{e}_i)$ and hence $M_2(\underset{\sim}{x}_2) < M_2(\underset{\sim}{e}_{i+1})$ $(i=1,2,..m-1)$. Hence by V $\underset{\sim}{x}_2 < \underset{\sim}{0}$ or $\underset{\sim}{x}_2 > \underset{\sim}{0}$.

By an exactly similar argument to the one in the proof of Part 1 we can show that $\underset{\sim}{x}_2 > \underset{\sim}{0}$. Now suppose that, contrary to Part 2, there is a $\underset{\sim}{y}$ with $\underset{\sim}{0} < \underset{\sim}{y} < \underset{\sim}{x}_2$ and $M_2(\underset{\sim}{y}) \leq M_2(\underset{\sim}{x}_2)$.

Then $M_1(A_1\underset{\sim}{y}) \leqslant M_1(\underset{\sim}{x}_1)$.

By VII and the fact that $\underset{\sim}{x}_1$ is a proper best approximation, we have

$$A_1\underset{\sim}{y} > \underset{\sim}{x}_1 \qquad (A_1\underset{\sim}{y} \neq \underset{\sim}{x}_1)$$

But from our supposition $A_1\underset{\sim}{y} \leqslant A_1\underset{\sim}{x}_2 = \underset{\sim}{x}_1$, and these are incompatible.

In the remaining sections we consider two special cases of Theorem 6.1.

6.3 The Regular Continued Fraction Process

If $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\underset{\sim}{\alpha} = \begin{pmatrix} \alpha \\ 1 \end{pmatrix}$, $\alpha > 1$, a suitable measure of approximation of $\underset{\sim}{x}$ to $\underset{\sim}{\alpha}$ is

$$M(\underset{\sim}{x}) = |x_1 - \alpha x_2|.$$

Conditions I, II, III, IV are obviously satisfied by M if α is irrational.

$\begin{pmatrix} a_0 \\ 1 \end{pmatrix}$ where a_0 is $[\alpha]$ is evidently the least proper best approximation to α.

Regarding V, if neither $\underset{\sim}{x} < \underset{\sim}{0}$ nor $\underset{\sim}{x} = \underset{\sim}{0}$, nor $\underset{\sim}{x} > \underset{\sim}{0}$ then we must have $x_1 < 0,\ x_2 > 0$ or $x_1 > 0, x_2 < 0$,

and in either case $M(\underset{\sim}{x}) = |x_1| + \alpha|x_2| > 1$.

Regarding VII, if neither $\underset{\sim}{x} \leqslant \underset{\sim}{y}$ nor $\underset{\sim}{x} > \underset{\sim}{y}$, for similar reasons

$$2 < 1+\alpha \leqslant |x_1 - y_1| + \alpha|x_2 - y_2| \leqslant |x_1 - \alpha x_2| + |y_1 - \alpha y_2|,$$

so we cannot have both $M(\underset{\sim}{x})$ and $M(\underset{\sim}{y})$ less than 1.

We can apply Theorem 6.1 if α is irrational to find all the best approximations to α . For if $\alpha > 1$ is irrational we can expand α

as a regular continued fraction

$$\begin{pmatrix}\alpha\\1\end{pmatrix} \sim K \begin{pmatrix}a_0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_2 & 1\\1 & 0\end{pmatrix}\cdots$$

(where the a_k are positive integers).

Further $M_2(\underset{\sim}{x}) \sim \left|(a_1x_1+x_2)\alpha_1 - x_1(a_1\alpha_1+1)\right| = \left|x_1-\alpha_1x_2\right|$

where $\quad \underset{\sim}{\alpha}_1 = \begin{pmatrix}\alpha_1\\1\end{pmatrix} \sim K\begin{pmatrix}a_1 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}a_2 & 1\\1 & 0\end{pmatrix}\cdots.$

Hence M_n $(n=1,2\ldots)$ are measures to their respective $\underset{\sim}{\alpha}_{n-1}$, and further $\begin{pmatrix}a_n\\1\end{pmatrix}(r=1,2,\ldots)$ are the least proper best approximations to

$$\begin{pmatrix}\alpha_n\\1\end{pmatrix} = \underset{\sim}{\alpha}_n \sim K\prod_{k=n}^{\infty}\begin{pmatrix}a_k & 1\\1 & 0\end{pmatrix} \quad \text{since} \quad a_n = [\alpha_n].$$

Hence all the proper best approximations to α with measure M are given by the 1-convergents of the regular continued fraction expansions of α.

6.4 Simultaneous Approximation to $(\sqrt[3]{4}, \sqrt[3]{2})$

If $\underset{\sim}{x} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$, $\underset{\sim}{\alpha}_1 = \begin{pmatrix}\alpha_1\\\beta_1\\1\end{pmatrix} = \begin{pmatrix}\sqrt[3]{4}\\\sqrt[3]{2}\\1\end{pmatrix}$ we define the function $M_1(\underset{\sim}{x})$ to be $M_1(\underset{\sim}{x}) = X_1^2 - \beta_1X_1Y_1 + \alpha_1Y_1^2$

where $X_1 = (x_1 - \alpha_1x_3)$ and $Y_1 = (x_2 - \beta_1x_3)$.

We will prove that M_1 is a measure of approximation to $\underset{\sim}{\alpha}_1$.

Then the main result of this section is

Theorem 6.2

The proper best approximations to $\underset{\sim}{\alpha}_1$ with measure M_1 are the 1-convergents of the expansion:

$$\underset{\sim}{\alpha}_1 \sim k \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \prod_{k=1}^{\infty} \begin{pmatrix} 3 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

(This is Expansion 5.1.)

Let $A_1, A_2, A_3, \ldots$ be the sequence of matrices in this expansion. Now if $M_n(\underset{\sim}{x})$ and $\underset{\sim}{\alpha}_n = \begin{pmatrix} \alpha_n \\ \beta_n \\ 1 \end{pmatrix}$ are defined as in the statement of Theorem 6.1 and $X_n = (x_1 - \alpha_n x_3)$, $Y_n = (x_2 - \beta_n x_3)$

then $M_2(\underset{\sim}{x}) \approx (\beta_2^2 - \beta_2 + 1) X_2^2 + (\alpha_2 - 2\alpha_2\beta_2 + 1) X_2 Y_2 + \alpha_2^2 Y_2^2$,

$$M_3(\underset{\sim}{x}) \approx (3\beta_3^2 - 8\beta_3 + 9) X_3^2 + (8\alpha_3 + 6 - 6\alpha_3\beta_3 - 3\beta_3) X_3 Y_3 + (3\alpha_3^2 + 3\alpha_3 + 1) Y_3^2,$$

and
$$M_4(\underset{\sim}{x}) \approx (12\beta_4^2 - 33\beta_4 + 37) X_4^2 + (-24\alpha_4\beta_4 + 33\alpha_4 - 10\beta_4 + 21) X_4 Y_4 + (12\alpha_4^2 + 10\alpha_4 + 3) Y_4^2.$$

Now it is quite clear that $\underset{\sim}{\alpha}_3 \sim \underset{\sim}{\alpha}_4 \sim \underset{\sim}{\alpha}_5 \sim \ldots$ and $A_3 = A_4 = A_5 = \cdots$; and we can now show that $M_3(\underset{\sim}{x}) \approx M_4(\underset{\sim}{x}) \approx M_5(\underset{\sim}{x}) \ldots$.
We need only prove that $M_3(\underset{\sim}{x}) \approx M_4(\underset{\sim}{x})$ and the rest follows easily.
Now α_3, β_3 satisfy

$$\alpha_3^3 = 3\alpha_3^2 + 3\alpha_3 + 1$$

and
$$\beta_3 = \alpha_3^2 - 3\alpha_3,$$

and from these we can deduce that

$$\alpha_3(3\beta_3^2 - 8\beta_3 + 9) = (12\beta_3^2 - 33\beta_3 + 37)$$

$$\alpha_3(8\alpha_3 + 6 - 6\alpha_3\beta_3 - 3\beta_3) = (-24\alpha_3\beta_3 + 33\alpha_3 - 10\beta_3 + 21)$$

and
$$\alpha_3(3\alpha_3^2 + 3\alpha_3 + 1) = (12\alpha_3^2 + 10\alpha_3 + 3),$$

which is our result.

Now it is any easy calculation to show that if there exists a least proper best approximation to $\underset{\sim}{\alpha}_1$ with measure M_1 then it must be $\begin{pmatrix}1\\1\\1\end{pmatrix}$. All we need to show is that $\begin{pmatrix}1\\1\\1\end{pmatrix}$ is a best approximation, and that if $\underset{\sim}{0} < \underset{\sim}{x} < \begin{pmatrix}1\\1\\1\end{pmatrix}$ then $\underset{\sim}{x}$ is not a best approximation. We can deal similarly with $\begin{pmatrix}3\\2\\1\end{pmatrix}$ and $\begin{pmatrix}3\\3\\1\end{pmatrix}$.

All that remains to be proved then, is that M_1, M_2, M_3 are measures to $\underset{\sim}{\alpha}_1$, $\underset{\sim}{\alpha}_2$ and $\underset{\sim}{\alpha}_3$ respectively. We will frequently use a

<u>Lemma 6.2</u>

If $Ax^2+2Bxy+Cy^2 \le D$ and $AC > B^2$, $A>0$,

then $|px+qy| \le \left(\dfrac{D(Aq^2-2Bpq+Cp^2)}{(AC-B^2)}\right)^{\frac{1}{2}}$ (for any real numbers p, q).

<u>Proof</u>

We may suppose that p and q are not both zero. Then

$$Ax^2+2Bxy+Cy^2 = \frac{(AC-B^2)}{(Aq^2-2Bpq+Cp^2)}\left(px+qy\right)^2 + \left((Aq-Bp)x+(Bq-Cp)y\right)^2.$$

Now both of these terms are non-negative for all x, y,

hence $$\left(\frac{AC-B^2}{Aq^2-2Bpq+Cp^2}\right)(px+qy)^2 \le D$$

and the conclusion follows.

6.5 Proof that M_1 is a Measure of Approximation to $\underset{\sim}{\alpha}_1$

M_1 evidently has properties I, II, III, IV. We need to discuss properties V, VI and VII in detail.

Property V

Let $\underset{\sim}{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. If $M_1(\underset{\sim}{x}) < M_1(\underset{\sim}{e_i})$ $(i = 1, 2, 3)$ then $M_1(\underset{\sim}{x}) < 1$.

Now applying Lemma 6.2 to $X_1^2 - \beta_1 X_1 Y_1 + \alpha_1 Y_1^2$ with $D = 1$ and with $(p, q) = (1, 0), (0, 1)$ and $(-\beta_1, \alpha_1)$ we obtain

$$|x - \alpha_1 z| < 1 \cdot 2$$

$$|y - \beta_1 z| < 1 \cdot 0$$

and $$|\alpha_1 y - \beta_1 x| < 1 \cdot 5 .$$

Suppose $\underset{\sim}{x}$ is not comparable with $\underset{\sim}{0}$. Then either

$$|x - \alpha_1 z| \geqslant 1 + \alpha_1 > 2 \cdot 5$$

or $$|y - \beta_1 z| \geqslant 1 + \beta_1 > 2 \cdot 2$$

or $$|\alpha_1 y - \beta_1 x| \geqslant \alpha_1 + \beta_1 > 2 \cdot 8$$

and we have a contradiction.

Property VI

In order to prove that $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is the least proper best approximation to $\underset{\sim}{\alpha_1}$ with measure M_1 , we need only show that if $\underset{\sim}{x}$ is not comparable with $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ then $\underset{\sim}{x}$ is not a best approximation, since we have disposed of the other possibilities.

Suppose $\underset{\sim}{x} > \underset{\sim}{0}$ is not comparable with $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Then $\left\{ \underset{\sim}{x} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ must have a least one positive, and at least one negative component.

Thus either $\quad 1+\alpha_1 \leqslant |(x-1)-\alpha_1(z-1)| \leqslant |x-\alpha_1 z| + |1-\alpha_1|$

and hence $\quad |x-\alpha_1 z| \geqslant 2\,;$

or $\quad 1+\beta_1 \leqslant |(y-1)-\beta_1(z-1)| \leqslant |y-\beta_1 z| + |1-\beta_1|$

and hence $\quad |y-\beta_1 z| \geqslant 2\,;$

or $\quad \alpha_1+\beta_1 \leqslant |\alpha_1(y-1)-\beta_1(x-1)| \leqslant |\alpha_1 y-\beta_1 x| + |\alpha_1-\beta_1|$

and hence $\quad |\alpha_1 y-\beta_1 x| > 2{\cdot}5\,.$

Now in order for $\underset{\sim}{x}$ to be a best approximation to $\underset{\sim}{\alpha}_1$ it must at least be a better approximation that $\underset{\sim}{e}_1, \underset{\sim}{e}_2$ or $\underset{\sim}{e}_3$.

Since $M_1(\underset{\sim}{e}_3)>M_1(\underset{\sim}{e}_2)>M_1(\underset{\sim}{e}_1)$, we must, in fact, have $M_1(\underset{\sim}{x}) < M_1(\underset{\sim}{e}_3)$.

Applying Lemma 6.2 with $D = M_1(\underset{\sim}{e}_3)$

we obtain $\quad |x-\alpha_1 z| < 1{\cdot}7$

$$|y-\beta_1 z| < 1{\cdot}4$$

and $\quad |\alpha_1 y-\beta_1 x| < 2{\cdot}2\,.$

Hence $\underset{\sim}{x}$ is not a best approximation to $\underset{\sim}{\alpha}_1$.

Property VII

Suppose $M_1(\underset{\sim}{x}) \leqslant M_1(\underset{\sim}{u}) \leqslant M_1\left[\begin{pmatrix}1\\1\\1\end{pmatrix}\right]$ where $\underset{\sim}{x}=\begin{pmatrix}x\\y\\z\end{pmatrix}, \underset{\sim}{u}=\begin{pmatrix}u\\v\\\omega\end{pmatrix}$.

Applying Lemma 6.2 with $\quad D = M_1\left[\begin{pmatrix}1\\1\\1\end{pmatrix}\right] < 0{\cdot}26$

we have $\quad |x-\alpha_1 z| < 0{\cdot}6$

$$|y-\beta_1 z| < 0{\cdot}5$$

and $\quad |\alpha_1 y-\beta x| < 0{\cdot}8\,.$

We obtain three more inequalities if we replace (x, y, z) by (u, v, ω) in the above.

Now suppose $\underset{\sim}{x}$ and $\underset{\sim}{u}$ are incomparable. Then $(\underset{\sim}{x}-\underset{\sim}{u})$ has a positive component and a negative component.

Hence either

$$2\cdot 5 < 1+\alpha_1 \leq |(x-u)-\alpha_1(z-\omega)| \leq |x-\alpha_1 z| + |u-\alpha_1\omega|$$

or $$2\cdot 2 < 1+\beta_1 \leq |(y-v)-\beta_1(z-\omega)| \leq |y-\beta_1 z| + |v-\beta_1\omega|$$

or $$2\cdot 8 < \alpha_1+\beta_1 \leq |\alpha_1(y-v)-\beta_1(x-u)| \leq |\alpha_1 y-\beta_1 x| + |\alpha_1 v-\beta_1 u|$$

and we have a contradiction.

Similar arguments enable one to deduce that M_2 and M_3 are measures to α_2 and α_3 . We have thus completed the proof of Theorem 6.2.

Remarks 6.2

(1) We do not make any estimates of $\psi(t)$ (see Lemma 6.1) for any sets of best approximations that we have obtained. However for best approximations to $\binom{\alpha}{1}$, one can always make an estimate (though perhaps a crude one) in terms of the partial quotients of the continued fraction expansion of α . For more delicate estimates for the best approximations to e see C.S. Davis [6, p.198].

(2) For a review of Simultaneous Diophantine Approximation (and a comment on analogues of the continued fraction process) see Davenport [5].

7. Two Further Applications in the Theory Diophantine Approximation

The first application is contained in

Theorem 7.1

If p, q are integers, $q > p > 0$, and ϵ is any positive real number, then there are only a finite number of integral vectors $\underset{\sim}{u} = \begin{pmatrix} u \\ v \end{pmatrix}$, with $v > 0$, such that

$$\left| e^{p/q} - \frac{u}{v} \right| < \frac{1}{v^{2+\epsilon}}$$

Remarks 7.1

(1) This result is an example of a variety of much more extensive results which can be found in Chapter IV of Schneider [13].

(2) Theorem 7.1 is a solution to the first problem of the introduction, part II, with

$$\Lambda = (1, -e^{p/q}), \quad m(\underset{\sim}{u}) = |u - e^{p/q} v|, \quad c(\underset{\sim}{u}) = |v|.$$

The function $\psi(t)$ we obtain is $\frac{1}{2} t^{-(1+\epsilon)}$. The method of proof is quite different from that discussed in chapter 6 and depends on

Lemma 7.1

If $\{p_n\}$, $\{q_n\}$ are sequences of integers

$$q_n^{-\omega} < | p_n - \alpha q_n | < \tfrac{1}{2} q_n^{-\nu} \qquad (\omega > \nu > 0),$$

and $0 < q_n < q_{n+1} < q_n^{\lambda}$ $(\lambda > 1)$ then there are only a finite number of vectors (r, s) with

$$|r - \alpha s| < \tfrac{1}{4} s^{-\sigma} \qquad \text{where} \qquad (1+\sigma)\nu = (1+\omega)\lambda .$$

(This Lemma is modelled on Lemma 5 of Baker's paper [1] but the idea behind it is well known.)

Proof

Suppose $|r-\alpha s| < \frac{1}{2}s^{-\sigma}$, $s^{\lambda/\nu} > q_1 \, (s>0)$.

Now $\sigma = \frac{\lambda}{\nu}(1+\omega)-1 = \frac{\lambda}{\nu} + \frac{\lambda\omega}{\nu} - 1 > \frac{\lambda}{\nu}$

(since $\lambda > 1$, $\omega > \nu > 0$). Put $\frac{\lambda}{\nu} = \delta$.

Then $|r-\alpha s| < \frac{1}{2}s^{-\delta}$.

Further, for some n, $q_n \leqslant s^{\delta} < q_{n+1}$.

For this n we have

$$s < q_{n+1}^{1/\delta} < q_n^{\lambda/\delta} = q_n^{\nu}.$$

Suppose $\frac{p_n}{q_n} \neq \frac{r}{s}$, then

$$1 \leq |p_n s - q_n r| \leqslant s|p_n - \alpha q_n| + q_n|r-\alpha s| < \frac{s}{2q_n^{\nu}} + \frac{q_n}{2s^{\delta}} < 1.$$

Thus we must have $\frac{p_n}{q_n} = \frac{r}{s}$; but then

$$|r-\alpha s| = \frac{s}{q_n}|p_n - \alpha q_n| > \frac{s}{q_n^{(1+\omega)}} \geqslant \frac{1}{s^{\delta(1+\omega)-1}} = \frac{1}{s^{\sigma}}.$$

We have completed the proof of Lemma 7.1.

Proof of Theorem 7.1

We use Expansion 5.2 with $x = \frac{p}{q}$. That is

$$\begin{pmatrix} e^{p/q} \\ 1 \end{pmatrix} \sim K \prod_{k=1}^{\infty} \begin{pmatrix} (2k-1)q+p & (2k-1)q \\ (2k-1)q & (2k-1)q-p \end{pmatrix}.$$

Let $A_1, A_2, A_3, \ldots$ be the sequence of matrices in this expansion, and

let $\prod_{k=1}^{n} A_k = P_n = \begin{pmatrix} x_n & z_n \\ y_n & w_n \end{pmatrix}$

Now since for all n, $A_n > 0$,

$$\begin{pmatrix} e^{p/q} \\ 1 \end{pmatrix} \sim P_n \begin{pmatrix} \alpha_n \\ \beta_n \end{pmatrix}$$

where α_n, β_n are non-negative.

Hence $\left|\frac{x_n}{y_n} - e^{p/q}\right| = \frac{\beta_n \omega_n}{(\alpha_n y_n + \beta_n \omega_n)} \; \frac{|\det P_n|}{y_n \omega_n}$

and $\left|\frac{z_n}{\omega_n} - e^{p/q}\right| = \frac{\alpha_n y_n}{(\alpha_n y_n + \beta_n \omega_n)} \; \frac{|\det P_n|}{y_n \omega_n}$.

Now either $\frac{\beta_n \omega_n}{(\alpha_n y_n + \beta_n \omega_n)} \geqslant \frac{1}{2}$ or $\frac{\alpha_n y_n}{(\alpha_n y_n + \beta_n \omega_n)} > \frac{1}{2}$,

and hence for all n , one of $\frac{x_n}{y_n}$, $\frac{z_n}{\omega_n}$ (we shall call it $\frac{u_n}{v_n}$) has the property

$$\frac{|\det P_n|}{2 y_n \omega_n} \leqslant \left|\frac{u_n}{v_n} - e^{p/q}\right| \leqslant \frac{|\det P_n|}{y_n \omega_n}. \qquad \ldots(1)$$

Now in order to apply Lemma 7.1 to the sequences u_n, v_n, we need to estimate $\frac{|\det P_n|}{y_n \omega_n}$ in terms of v_n , and also $\frac{v_n}{v_{n-1}}$ in terms of v_{n-1} .

Now evidently, for $k = 1, 2, 3, \ldots$

$$0 < \{(2k-1)q - p\}\begin{pmatrix}1&1\\1&1\end{pmatrix} < A_k < \{(2k-1)q + p\}\begin{pmatrix}1&1\\1&1\end{pmatrix},$$

and hence

$$(q-p)(3q-p)\cdots((2n-1)q-p)\begin{pmatrix}1&1\\1&1\end{pmatrix}^n \leqslant P_n \leqslant (q+p)(3q+p)\cdots((2n-1)q+p)\begin{pmatrix}1&1\\1&1\end{pmatrix}^n.$$

Now since $q \geqslant p+1$, and $\begin{pmatrix}1&1\\1&1\end{pmatrix}^n = 2^{n-1}\begin{pmatrix}1&1\\1&1\end{pmatrix}$,

$$2^{2n-2}(n-1)!\, q^{n-1}\begin{pmatrix}1&1\\1&1\end{pmatrix} < P_n < 2^{2n-1}\, n!\, q^n \begin{pmatrix}1&1\\1&1\end{pmatrix}.$$

Thus v_n satisfies

$$2^{2n-2}(n-1)!\, q^{n-1} < v_n < 2^{2n-1} n!\, q^n$$

and $y_n \omega_n$ satisfies

$$2^{4n-4}\{(n-1)!\}^2 q^{2n-2} < y_n \omega_n < 2^{4n-2}(n!)^2 q^{2n}.$$

Hence

$$\frac{v_n^2}{4n^2q^2} < y_n\omega_n < 4n^2q^2v_n^2 \qquad \ldots(2)$$

and $1 < \frac{v_n}{v_{n-1}} < 8n(n-1)q^2$. $\qquad \ldots(3)$

Further, since

$$(n-1)! \geqslant e^{-n+1}n^{n-1} \qquad (n \geqslant 1)$$

we have $\quad v_n > \left(\frac{4q}{e}\right)^{n-1} n^{n-1}$.

Now, for $n \geqslant 1$ since $q \geqslant 2$, $\frac{4q}{e} \geqslant \frac{8}{e} > 2$ and $2^{n-1} \geqslant n$,

so $\quad v_n > n^n$. $\qquad \ldots(4)$

Now in order to apply Lemma 7.1 we wish to prove that for any $\epsilon > 0$ and all sufficiently large n

(A) $\frac{1}{v_n^{2+\epsilon}} < \left| \frac{u_n}{v_n} - e^{p/q} \right| < \frac{1}{2v_n^{2-\epsilon}}$

and (B) $1 < \frac{v_n}{v_{n-1}} < (v_{n-1})^{\epsilon}$.

Now to prove (A) it is sufficient to prove, in view of (1),

that $\frac{|\det P_n|}{y_n\omega_n} < \frac{1}{2v_n^{2-\epsilon}}$ and $\frac{|\det P_n|}{2y_n\omega_n} > \frac{1}{v_n^{2+\epsilon}}$

or that $\quad v_n^{\epsilon} > \frac{2v_n^2 p^{2n}}{y_n\omega_n} \qquad (|\det P_n| = p^{2n})$

and $\quad v_n^{\epsilon} > \frac{2y_n\omega_n}{v_n^2 p^{2n}}$.

But by (2) $\quad \frac{2y_n\omega_n}{v_n^2p^{2n}} < \frac{8n^2q^2}{p^{2n}} < 16n^2q^2p^{2n}$,

and $\quad \frac{2v_n^2p^{2n}}{y_n\omega_n} < 8n^2q^2p^{2n} < 16n^2q^2p^{2n}$.

Further by (3) $\frac{v_{n+1}}{v_n} < 8n(n+1)q^2 < 16n^2q^2p^2 \quad (n \geqslant 1)$.

Hence to prove (A) and (B), we need only prove that for any $\epsilon > 0$, and all sufficiently large n,

$$v_n^{\epsilon} > 16 n^2 q^2 p^{2n}.$$

It is clear that our estimate (4) is sufficient.

It is also clear that Lemma 7.1 applied to (A) and (B) yields the result of Theorem 7.1.

The second application of the K-process theory to Diophantine Approximation is

Theorem 7.2

$$\text{If} \quad \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix} \sim K \prod_{k=1}^{\infty} \begin{pmatrix} a_k & 1 & 0 \\ b_k & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

where a_n and b_n are positive integers and $a_n > 2b_n + 2$, then the index (see Definition 4.1) of $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ is zero.

Proof

Suppose $A_1, A_2, A_3, \dots$ is the sequence of matrices of the above expansion. We put

$$A_k A_{k+1} \cdots A_n = \begin{pmatrix} p_{n,k} & p_{n-1,k} & p_{n-2,k} \\ q_{n,k} & q_{n-1,k} & q_{n-2,k} \\ r_{n,k} & r_{n-1,k} & r_{n-2,k} \end{pmatrix},$$

but where $k = 1$ we drop the subscript k.

By Theorem 4.2 we can put $\begin{pmatrix} \alpha_k \\ \beta_k \\ 1 \end{pmatrix} \sim K A_k A_{k+1} A_{k+2} \cdots$,

with $\alpha_k > \beta_k > 1$. Evidently $\begin{pmatrix} \alpha_1 \\ \beta_1 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$.

Lastly we put $P_{n,k} = |p_{n,k} - \alpha_k r_{n,k}|$ and $Q_{n,k} = |q_{n,k} - \beta_k r_{n,k}|$.

Now by definition

$$\left.\begin{aligned} p_{n,k-1} &= a_{k-1} p_{n,k} + q_{n,k} \\ q_{n,k-1} &= b_{k-1} p_{n,k} + r_{n,k} \\ r_{n,k-1} &= p_{n,k} \\ \text{and} \quad \alpha_{k-1} &= a_{k-1} + \frac{\beta_k}{\alpha_k} \\ \beta_{k-1} &= b_{k-1} + \frac{1}{\alpha_k} \end{aligned}\right\} \qquad \ldots(5)$$

Hence $P_{n,k-1} = \frac{1}{\alpha_k} \left|(a_{k-1} p_{n,k} + q_{n,k})\alpha_k - (a_{k-1}\alpha_k + \beta_k) p_{n,k}\right|$

$$= \frac{1}{\alpha_k} \left|\alpha_k q_{n,k} - \beta_{n,k} p_{n,k}\right| \leqslant \frac{\beta_k}{\alpha_k} P_{n,k} + Q_{n,k}.$$

Further $Q_{n,k-1} = \frac{1}{\alpha_k}\left|(b_{k-1} p_{n,k} + r_{n,k})\alpha_k - (b_{k-1}\alpha_k + 1) p_{n,k}\right|$

$$= \frac{1}{\alpha_k} P_{n,k}.$$

Now by equations (5), replacing k by $k+1$, and the fact that

$\alpha_{k+1} > \beta_{k+1} > 1$, $\alpha_k > [\alpha_k] = a_k$ and $\beta_k > [\beta_k] = b_k$.

Hence

$$\begin{pmatrix} P_{n,k-1} \\ Q_{n,k-1} \end{pmatrix} < \begin{pmatrix} \frac{b_k+1}{a_k} & 1 \\ \frac{1}{a_k} & 0 \end{pmatrix} \begin{pmatrix} P_{n,k} \\ Q_{n,k} \end{pmatrix}.$$

It is a simple consequence of this that

$$\begin{pmatrix} |p_n - \alpha r_n| \\ |q_n - \beta r_n| \end{pmatrix} = \begin{pmatrix} P_{n,1} \\ Q_{n,1} \end{pmatrix} < \begin{pmatrix} \frac{b_2+1}{a_2} & 1 \\ \frac{1}{a_2} & 0 \end{pmatrix} \begin{pmatrix} \frac{b_3+1}{a_3} & 1 \\ \frac{1}{a_3} & 0 \end{pmatrix} \cdots \begin{pmatrix} \frac{b_n+1}{a_n} & 1 \\ \frac{1}{a_n} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

since $\begin{pmatrix} P_{n,n} \\ Q_{n,n} \end{pmatrix} = \begin{pmatrix} a_n - \alpha_n \\ b_n - \beta_n \end{pmatrix} < \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$

Now $a_n > 2b_n + 2$, so

$$\begin{pmatrix} |p_n - \alpha r_n| \\ |q_n - \beta r_n| \end{pmatrix} < \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{4} & 0 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

It is an easy matter to prove by induction that

$$\begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}^n \leqslant \begin{pmatrix} (3.5)^n & 4(3.5)^{n-1} \\ (3.5)^{n-1} & 4(3.5)^{n-2} \end{pmatrix}.$$

It is obviously true for $n=1$. Suppose it is true for some $n \geq 1$.

Then

$$\begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}^{n+1} \leq \begin{pmatrix} (3\cdot5)^n & 4\,(3\cdot5)^{n-1} \\ (3\cdot5)^{(n-1)} & 4(3\cdot5)^{n-2} \end{pmatrix} \begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} (3\cdot5)^{n-1}(2(3\cdot5)+4) & 4(3\cdot5)^n \\ (3\cdot5)^{n-2}(2(3\cdot5)+4) & 4\,(3\cdot5)^{n-1} \end{pmatrix}$$

The truth for $n+1$ thus follows, since $11 = 2\,(3\cdot5)+4 < (3\cdot5)^2 = 12\cdot25$.

It is clear then that

$$\begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{4} & 0 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{4^n}\begin{pmatrix} 2 & 4 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{as } n \to \infty .$$

Thus we have proved that if $p_n - \alpha r_n = \varepsilon_n$, $q_n - \beta r_n = \delta_n$, then $\varepsilon_n \to 0$ and $\delta_n \to 0$ as $n \to \infty$.

We are now ready to prove that the index of $\begin{pmatrix} \alpha \\ \beta \\ 1 \end{pmatrix}$ is zero.

Suppose there is a non-zero vector (a, b, c) such that $a\alpha + b\beta + c = 0$.

Then $(a,b,c)A_1A_2 \cdots A_n \neq \underline{0}$ because the matrices are non-singular. Thus at least one out of every three successive convergents has

$$p_n a + q_n b + r_n c \neq 0 .$$

Consider, for these convergents, the equations:

$$\begin{aligned} a\alpha + b\beta + c &= 0 \\ -r_n\alpha \qquad + p_n &= \varepsilon_n \\ -r_n\beta + q_n &= \delta_n . \end{aligned}$$

By a familiar property of determinants

$$\det \begin{pmatrix} a & b & 0 \\ -r_n & 0 & \varepsilon_n \\ 0 & -r_n & \delta_n \end{pmatrix} = \det \begin{pmatrix} a & b & c \\ -r_n & 0 & p_n \\ 0 & -r_n & q_n \end{pmatrix},$$

or $\quad a r_n \varepsilon_n + b r_n \delta_n = a p_n r_n + b q_n r_n + c r_n^2$.

Evidently for $n \geq 1, r_n > 0$ so

$$a p_n + b q_n + c r_n = a\varepsilon_n + b\delta_n \to 0 \quad \text{as } n \to \infty.$$

But the left hand side is a non-zero integer for all n.

This completes the proof of Theorem 7.2.

We can apply Theorem 7.2 to some cases of Expansion 5.8.

If $a_n = (an+f)(an+g)$ where $a > 0,\ f > 0,\ g > 0$ are integers, then a variant of Expansion 5.8 given on page 43 becomes

$$\begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} \sim K \begin{pmatrix} (2a+f)(2a+g) & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \prod_{n=2}^{\infty} \begin{pmatrix} (a(n+1)+f)(a(n+1)+g) & 1 & 0 \\ a(2n+1)+f+g & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Clearly all but a finite number of the matrices of this expansion are of the type specified in Theorem 7.2 and thus the index of $\begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix}$ is zero.

References

1. BAKER, A.
Rational approximations to certain algebraic numbers
Proc. London Math. Soc. (3) 14, 385-98, 1964.

2. BROMWICH, T.J.I.
An introduction to the theory of infinite series. 2nd ed.rev.
London, Macmillan, 1926 reprinted 1955.

3. CASSELS, J.W.S.
An introduction to diophantine approximation
(Cambridge tracts in mathematics and mathematical physics, no.45)
Cambridge, University Press, 1957 reprinted 1965.

4. CHÂTELET, A.
Contribution à la théorie des fractions continues arithmétiques
Bull. Soc. Math. France. 40, 1-25, 1912.

5. DAVENPORT, Harold
Simultaneous diophantine approximation
Proceedings of the International Congress of Mathematicians
[12th] 1954, Amsterdam. Vol. 3, 9-12.
Groningen, Noordhoff, 1956.

6. DAVIS, C.S.
On some simple continued fractions connected with e
J. London Math. Soc. 20, 194-198, 1945.

7. HARDY, G.H. and WRIGHT, E.M.
An introduction to the theory of numbers. 4th ed.
Oxford, Clarendon Press, 1960 reprinted 1965.

8. KOKSMA, J.F.
Diophantische Approximationen
(Ergebnisse der Mathematik und ihrer Grenzgebiete, Bd.5, [Heft] 4)
New York, Chelsea, c1936.

9. KOLDEN, Kjell
Continued fractions and linear substitutions
Arch. Math. Naturvid. 50, no.6, 141-196, 1949.

10. PERRON, Oskar
Grundlagen für eine Theorie des Jacobischen Kettenbruch-algorithmus
Math. Ann. 64, 1-76, 1907.

11. PERRON, Oskar
Die Lehre von den Kettenbrüchen. 3rd ed. Bd.1. Elementare Kettenbrüche
Stuttgart, Teubner, 1954.

12. PERRON, Oskar
Die Lehre von den Kettenbrüchen. 3rd ed. Bd.2. Analytisch-funktionentheoretische Kettenbrüche
Stuttgart, Teubner, 1957.

13. SCHNEIDER, Theodor
Einführung in die transzendenten Zahlen
(Die Grundlehren der Mathematischen Wissenschaften, Bd.81)
Berlin, Springer, 1957.

14. TODD, H.
On diophantine approximation to certain exponential and Bessel functions
Proc. London Math. Soc. (2) 50, 550-559, 1949.