HOMOGENEOUS APPROXIMATION IN THE FIELD OF FORMAL POWER SERIES*

by SATISH K. AGGARWAL,† *Department of Mathematics, Ohio State University, Columbus, Ohio*

(Communicated by R. P. Bambah, F.N.I.)

(*Received January 10, 1967*)

Analogues of two classical theorems on Diophantine Approximations are proved in the field of formal power series over a field with the 'degree' valuation. In this set-up polynomials play the role of integers, rational functions of rational numbers and general power series that of real numbers.

§ 1. Let K be any field and t be an indeterminate. Let $K[t]$ be the ring of polynomials, $K(t)$ be the field of rational functions and $K\{t\}$ be the field of formal power series in t over K. The elements of $K\{t\}$ are of the form

$$x = \alpha_m t^m + \alpha_{m-1} t^{m-1} + \dots \text{ (up to } -\infty). \qquad \dots \qquad \dots \qquad (1.1)$$

In the field $K\{t\}$, we define the usual valuation $||$ as follows:

(i) $|0| = 0$.

(ii) $x = \alpha_m t^m + \alpha_{m-1} t^{m-1} + \dots, \alpha_m \neq 0$, then $|x| = e^m$.

If x is as in (1.1), we also define

$$\|x\| = |\alpha_{-1} t^{-1} + + \alpha_{-2} t^{-2} + \dots|.$$

Write $P_n = (K[t])^n = \{X = (x_1, \dots, x_n);\ x_i \in K[t]\}$ and $R_n = (K\{t\})^n = \{X = (x_1, \dots, x_n);\ x_i \in K\{t\}\}$. For $X = (x_1, \dots, x_n) \in R_n$, we define $|X| = \max(|x_1|, \dots, |x_n|)$.

It is well known that the fields $K\{t\}$, $K(t)$ and the integral domain $K[t]$ are similar in their relative structure to the real field, the rational field and the rational integers respectively. $\|x\|$ corresponds to the fractional part of the real number x, P_n corresponds to n-dimensional lattice and R_n corresponds to n-dimensional Euclidean space.

Dirichlet has proved:

(A) *If there are m real linear forms* $L_i(U) = L_i(u_1, \dots, u_n)$ $(1 \leq i \leq m)$,

---

* This is a part of the author's dissertation approved for the Ph.D. degree of the Ohio State University in December, 1965, written under the guidance of Professor R. P. Bambah.

† *Present address:* Mathematics Department, University of Illinois, Urbana, Illinois. 61801, U.S.A.

then to every real $t < 1$, there is a $U = (u_1, \ldots, u_n)$; u_i integers not all zero, such that

$$\| L_i(U) \| < t^{-\frac{n}{m}}, |u_j| \leq t, \quad (1 \leq i \leq m, 1 \leq j \geq n),$$

where $\| x \| = \min_{n \in Z} | x - n |$ *is the distance of x from the nearest rational integer.*

In the oppositie direction, the following result has been proved:

(B) *For any positive integers m, n, there exists a constant $\gamma(m, n) = \gamma > 0$ and real linear forms $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$, such that*

$$\left(\max_{1 \leq j \leq n} |u_j|\right)^n \left(\max_{1 \leq i \leq m} \| L_i(U) \|\right)^m \geq \gamma$$

for all $U = (u_1, \ldots, u_n)$; u_i integers not all zero (*see* e.g. Cassels 1957, pp. 15-16).

In § 3, we shall prove an analogue (theorem 1) of (B) in our set-up, provided the field K has at least $(m+n-1)$ distinct elements.

The analogue of (A) in this set-up can be proved by direct application of Mahler's theorem (Armitage 1957) on linear forms.

In the field of real numbers, a theorem of Dirichlet and Minkowski can be stated in the following form:

(C) *Let $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m < n)$ be m real linear forms, which are linearly independent over the field of real numbers, such that $L_i(U) \neq 0$ $(1 \leq i \leq m)$ for any $U = (u_1, \ldots, u_n)$; u_i integers not all zero. Then there exists a suitable positive constant C, such that*

$$\left(\max_{1 \leq i \leq m} |L_i(U)|\right)^m \left(\max_j |u_j|\right)^{n-m} \leq C$$

admits an infinite number of solutions $U = (u_1, \ldots, u_n)$, u_i integers.

An analogue to this result in the p-adic case was proved by E. Lutz (1955). In § 4 we prove an analogue (theorem 2) in the set-up discussed above.

§ 2. In this section, we give a few definitions and state precisely the results to be proved. The analogue of (B) is

THEOREM 1: *Let m, n be two positive integers and K be a field having at least $(m+n-1)$ distinct elements. Then there exists a constant $\mu > 0$ depending only on K, m, n and m linear forms $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$ in n variables $u_1, \ldots, u_n$ with coefficients from $K\{t\}$, such that*

$$|U|^n \left(\max_{1 \leq i \leq m} \| L_i(U) \|\right)^m \geq \mu$$

for all non-zero $U \in P_n$.

Now we shall make a few definitions.

In all that follows, all the linear forms will have coefficients in $K\{t\}$. When we say these are l.i., we shall mean $L_1, \ldots, L_m$ are linearly independent over $K\{t\}$.

Definition 1. Let $\wedge$ be a system of l.i. forms $L_1, \ldots, L_m$ in variables $u_1, \ldots, u_n$. Let A be the matrix of the coefficients. Define rational integers

$\delta = \delta(\wedge)$ and $\zeta = \zeta(\wedge)$ as follows:

$$e^{-\delta} = \max_{M} (|\det M|),$$

where M runs over the minors of rank m from the matrix A and

$$e^{\zeta} = \max_{M} \left(\min_{M^*} \left| \frac{\det M}{\det M^*} \right| \right),$$

where M runs over all the non-singular minors of the matrix A of rank $\leq m-1$ and M being given, M^* runs over non-singular minors of the matrix A containing M by adding one row and one column. (Since the linear forms are l.i., such an M^* always exists.)

Definition 2. A system $\wedge$ of linear forms $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$ is called *annulable* if there exists a non-zero vector $U \in P_n$ such that $L_i(U) = 0$ $(1 \leq i \leq m)$.

The analogue of (C) is

THEOREM 2: *Suppose* $\wedge : (L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m))$ *is a system of* m *l.i., linear forms with* $1 \leq m < n$ *and suppose* $\wedge$ *is not annulable. Then the inequality*

$$\left(\max_{1 \leq i \leq m} |L_i(U)| \right)^m |U|^{n-m} \leq \exp(n-m-\delta) \qquad \ldots \qquad \ldots \quad (2.1)$$

admits an infinite number of solutions $U \in P_n$ *where* δ *is the integer appearing in definition 1.*

§ 3. *Proof of Theorem 1*: Since K has at least $(m+n-1)$ distinct elements, there exist in $K\{t\}$, $(m+n)$ conjugate algebraic integers $\phi_1, \ldots, \phi_{m+n}$ over $K[t]$. Let $r = m+n$. Set

$$Q_K(U, V) = \sum_{i=1}^{m} \phi_k^{i-1} v_i + \sum_{j=1}^{n} \phi_k^{m+j-1} u_j \quad (1 \leq k \leq r).$$

Now we assert that for elements $0 \neq (u_1, \ldots, u_n, v_1, \ldots, v_m) \in P_{m+n}$, $Q_k(U, V) \neq 0$ for any k $(1 \leq k \leq r)$. If not, suppose $Q_k(U, V) = 0$ for some k. Then $Q_k(U, V) = 0$ for all k $(1 \leq k \leq r)$, because $Q_k(U, V)$ are all conjugate to each other. But the coefficients of $Q_k(U, V)$ $(1 \leq k \leq r)$ form a Vandermonde matrix. This has therefore a non-zero determinant, and hence the only solution of

$$Q_k(U, V) = 0 \quad (1 \leq k \leq r)$$

is $(U, V) = (0, 0)$.

Thus $\prod_k Q_k(U, V) \neq 0$ for $(U, V) \neq (0, 0)$. Also $\prod_k Q_k(U, V)$ is an element of $K[t]$ and hence

$$\left|\prod_k Q_k(U, V)\right| \geq 1 \text{ or } \prod_k |Q_k(U, V)| \geq 1. \qquad \ldots \qquad \ldots \quad (3.1)$$

Now consider the equations

$$\sum_{i=1}^{m} \phi_k^{i-1} w_i = -\sum_{j=1}^{n} \phi_k^{m+j-1} u_j \quad (1 \le k \le m) \qquad \ldots \qquad (3.2)$$

in variables $w_1, \ldots, w_m$. The martrix of these equations is a Vandermonde matrix and hence non-singular. So solving these equations by Cramer's rule we get

$$w_i = L_i(U) \quad (1 \le i \le m),$$

where $L_i(U)$ $(1 \le i \le m)$ are linear in $u_1, \ldots, u_n$. The coefficients of $L_i(U)$ $(1 \le i \le m)$ depend only on ϕ's.

Now we write $Q_k(U, V)$ $(m+1 \le k \le r)$ in the form

$$Q_k(U, V) = \sum_{i=1}^{m} \phi_k^{i-1}(v_i - L_i(U)) + \sum_{j=1}^{n} \gamma_{kj} u_j \quad (m+1 \le k \le r), \quad \ldots \quad (3.3)$$

where γ_{kj} $(m+1 \le k \le r,\ 1 \le j \le n)$ belong to $K\{t\}$ depending only on ϕ's.

Also $w_i = L_i(U)$ $(1 \le i \le m)$ are solutions of (3.2), so we get

$$\sum_{i=1}^{m} \phi_k^{i-1} L_i(U) = -\sum_{j=1}^{n} \phi_k^{m+j-1} u_j \quad (1 \le k \le m). \qquad \ldots \quad (3.4)$$

Let $U = (u_1, \ldots, u_n)$ be any non-zero vector of P_n.

Let

$$e^{\alpha} = \max_{1 \le j \le n} |u_j| = |U| \qquad \ldots \quad \ldots \quad \ldots \quad (3.5)$$

and

$$e^{\beta} = \max_{1 \le i \le m} \| L_i(U) \|. \qquad \ldots \quad \ldots \quad \ldots \quad (3.6)$$

Then

$$e^{\beta} \le e^{-1} < e^{\alpha}, \text{ i.e. } \beta < \alpha. \qquad \ldots \quad \ldots \quad \ldots \quad (3.7)$$

Let $v_1, \ldots, v_m$ be elements of $K[t]$, such that

$$\| L_i(U) \| = |L_i(U) - v_i| \quad (1 \le i \le m).$$

Then from (3.4) we have

$$|Q_k(U, V)| \le e^{\gamma_1 + \beta} \quad \text{for } 1 \le k \le m, \qquad \ldots \quad \ldots \quad (3.8)$$

where γ_1 depends only on ϕ's. Also from (3.3), (3.5), (3.6) and (3.7) we get

$$|Q_k(U, V)| \le \max\left(e^{\beta+\gamma_2}, e^{\alpha+\gamma_3}\right) \le e^{\alpha+\gamma_4} \quad \text{for } (m+1 \le k \le r). \qquad (3.9)$$

The γ's occurring in (3.9) depend only on ϕ's and hence on K.

From (3.1), (3.8) and (3.9) we get

$$1 \le \prod_k |Q_k(U, V)| \le \exp\left[m\gamma_1 + m\beta + n\alpha + n\gamma_4\right]$$

or

$$\exp\left[-m\gamma_1 - n\gamma_4\right] \le \exp(m\beta + n\alpha) = |U|^n \left(\max_i \| L_i(U) \|\right)^m$$

and this proves the theorem.

Proof of theorem 2.

§ 4. We shall first make a few definitions.

Definition 3. Let $F(X)$ be a function carrying R_n into $e^Z \cup \{0\}$, where Z is the set of rational integers and satisfying the following properties:

(i) $F(X) \neq 0$ for $0 \neq X \in R_n$.

(ii) $F(aX) = |a| F(X)$; $a \in K\{t\}$, $X \in R_n$.

(iii) $F(X \pm Y) \leq \max(F(X), F(Y))$; $X \in R_n$, $Y \in R_n$.

Then $F(X)$ is called *distance function.* Also a set $\{X \in R_n;\ F(X) \leq e^m$, where m is a rational integer$\}$ is called a *convex body.*

Definition 4. Let $F(X)$ be a distance function and $D(m)$ be the dimension of the vector space $\{X \in R_n;\ F(X) \leq e^m,\ m$ a rational integer$\}$ over K and $D_0(m)$ be the dimension of the vector space $\{X \in R_n;\ |X| \leq e^m\}$. We define the volume V of the convex body $\{X \in R_n;\ F(X) \leq 1\}$ by

$$V = \lim_{m \to \infty} \exp [D(m) - D_0(m].$$

(This limit is known to exist (Mahler 1941)).

Notation: We shall denote the volume of a set S by Vol. S.

Lemma 1: Let $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq n)$ be n linear forms with determinant $D \neq 0$. Then the volume of the set

$$\{U \in R_n;\ |L_i(U)| \leq 1 \quad (1 \leq i \leq n)\} \text{ equals } |D|^{-1}.$$

Proof (Lemma 1): By definition $\{X \in R_n;\ |X| \leq 1\}$ has volume 1 and lemma follows from a theorem of Mahler (1941, pp. 505–508).

Lemma 2: Let $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$ be a system $\wedge$ of linear forms. Suppose the matrix A of the coefficients has rank n. Then the volume V of the set

$$\mathfrak{J} = \left\{ U \in R_n;\ \max_{1 \leq i \leq m} |L_i(U)| \leq 1 \right\}$$

equals $\min_M (|\det M|^{-1})$, where M denotes an arbitrary minor of rank n of matrix A.

Proof (Lemma 2): Take all possible combinations of n linear forms among $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$. Let M denote the matrix of an arbitrary set of n forms. Consider $\max_M |\det M|$. Since the matrix A has rank n, $\max_M |\det M| \neq 0$. Without loss of generality we assume that the forms associated to $\max_M |\det M|$ are $L_1, \ldots, L_n$. Then the determinant D of the forms $L_1, \ldots, L_n$ is not zero.

Write

$$L_{n+k}(U) \equiv \sum_{i=1}^{n} \frac{L_i(U)}{D} y_{ki} \quad (1 \leq k \leq m-n).$$

Then by Cramer's rule we get $y_{ki} = D_{ki}$ $(1 \leq k \leq m-n;\ 1 \leq i \leq n)$, where D_{ki} is the determinant of the forms $L_1(U), \ldots, L_{i-1}(U), L_{i+1}(U), \ldots, L_n(U),$

$L_{n+k}(U)$. Further by our choice of D, $\left|\frac{y_{k+1}}{D}\right| \leq 1$, so that U which satisfies $L_i(U) \leq 1$ $(i = 1, \ldots, n)$ also satisfies $\left|\sum_{i=1}^{n} \frac{y_{ki}}{D} L_i(U)\right| \leq 1$ $(k = 1, \ldots, m-n)$. Thus the two sets $\mathfrak{J}$ and $\left\{U \in R_n;\ \max_{1 \leq i \leq m} |L_i(U)| \leq 1\right\}$ are identical. However, by lemma 1 the volume of the latter is $|D|^{-1}$. Hence the lemma follows.

LEMMA 3: Let $\wedge$ be a system of l.i. linear forms $L_i(U) = L_i(u_1, \ldots, u_n)$ $(1 \leq i \leq m)$. Let δ and ζ be as in definition 1. Let $\lambda_1, \ldots, \lambda_m$ be rational integers $\geqslant \zeta$.
Then

$$\mathrm{Vol}\,\{U \in R_n;\ |U| \leq 1,\ |L_i(U)| \leq e^{-\lambda_i},\ i = 1, \ldots, m\}$$
$$= \exp\left(\delta - \sum_{i=1}^{m} \lambda_i\right).$$

PROOF (LEMMA 3): Consider the following system of $m+n$ linear forms

$$\begin{array}{ll} t^{\lambda_i} L_i(U) & (i = 1, \ldots, m) \\ u_j & (j = 1, \ldots, n). \end{array}$$

Out of these $m+n$ linear forms, take an arbitrary set of n linearly independent forms. Let s $(0 \leq s \leq m)$ be the number of forms $t^{\lambda_i} L_i(U)$ among these and $(n-s)$ be the number of forms u_j. Then the valuation of the determinant of such a system is

$$\exp\left(\sum_{p \in S} \lambda_p\right) \cdot |\det M|,$$

where S denotes the set of indices $p \in S$ of the form $t^{\lambda_p} L_p(U)$ considered above and M is a certain minor of the coefficient matrix of rank s. If we denote by $\mathfrak{J}$ the set

$$\mathfrak{J} = \{U \in R_n;\ |U| \leq 1;\ |t^{\lambda_i} L_i(U)| \leq 1,\ i = 1, \ldots, m\},$$

we have by theorem 2,

$$\mathrm{Vol}\,\mathfrak{J} = \min\left(\exp\left(-\sum_{p \in S} \lambda_p\right) |\det M|^{-1}\right),$$

where minimum has been taken over all the possible combinations of n linearly independent forms.

Consider a value $s < m$, and adjoin to the forms $t^{\lambda_p} L_p(U)$ $(p \in S)$, already considered, an $(s+1)$th form of the above type. By hypothesis, these $(s+1)$ forms are linearly independent. This $(s+1)$th form is to be suitably chosen subsequently. For the time being, call it $t^{\lambda_*} L_*(U)$. With these $(s+1)$ forms, we take $(n-s-1)$ forms among the u_j already considered. The valuation of the new determinant is then

$$\left[\exp\left(\sum_{p \in S} \lambda_p + \lambda_*\right)\right] |\det M^*|,$$

where M^* is a certain minor of the coefficient matrix with rank $(s+1)$; and can be obtained from M by adding to it one row and one column. But by definition of ζ, $e^{\zeta} \geqslant \left|\frac{\det M}{\det M^*}\right|$ for all M of rank $\leq m-1$ and M being given, for some M^* containing M. Now we choose that form $t^{\lambda_*}L_*(U)$ for which $e^{\zeta} \geq \left|\frac{\det M}{\det M^*}\right|$. Then by hypothesis, $e^{\lambda_*} \geq e^{\zeta} \geq \left|\frac{\det M}{\det M^*}\right|$.

Hence the valuation of the new determinant is greater than or equal to the valuation of the old determinant. Hence the set S can be made to contain the m indices. Thus by definition of δ, we have

$$\operatorname{Vol} \mathfrak{J} = \exp\left(\delta - \sum_{i=1}^{m} \lambda_i\right),$$

and this proves the lemma.

LEMMA 4:

$$\operatorname{Vol}\left\{U \epsilon R_n;\ |U| \leq e^k,\ |L_i(U)| \leq e^{-\lambda_i},\ i = 1, \ldots, m\right\}$$
$$= \exp\left(\delta + (n-m)k - \sum_{i=1}^{m} \lambda_i\right),$$

where $k \geq 0$ is an integer.

PROOF: Let

$$\mathfrak{J} = \left\{U \epsilon R_n;\ |U| \leq e^k,\ |L_i(U)| \leq e^{-\lambda_i},\ i = 1, \ldots, m\right\}$$
$$= \left\{U \epsilon R_n;\ |t^{-k}U| \leq 1,\ |t^{\lambda_i}L_i(U)| \leq 1,\ i = 1, \ldots, m\right\}.$$

Let

$$F(U) = \max\left(|t^{-k}U|,\ |t^{\lambda_i}L_i(U)|,\ i = 1, \ldots, m\right).$$

Then $F(U)$ is a distance function and $\mathfrak{J}$ is the convex body $F(U) \leq 1$. Consider the transformation T defined by

$$V = t^{-k}U = TU.$$

This transformation maps $F(U) \leq 1$ into

$\mathfrak{J}^*$: $F^*(V) \leq 1$, where

$$F^*(V) = \max\left(|V|,\ |t^{k+\lambda_i}L_i(V)|,\ i = 1, \ldots, m\right),$$

i.e.

$$\mathfrak{J}^* = \left\{V \epsilon R_n;\ |V| \leq 1,\ |t^{k+\lambda_i}L_i(V)| \leq 1,\ i = 1, \ldots, m\right\}.$$

Using theorem 2, we have $\operatorname{Vol} \mathfrak{J}^* = \exp\left(\delta - mk - \sum_{i=1}^{m} \lambda_i\right)$. By using a result of Mahler (1941, pp. 505–508) we get

$$\operatorname{Vol} \mathfrak{J} = |\det T|^{-1} \operatorname{Vol} \mathfrak{J}^* = \exp\left(\delta + (n-m)k - \sum_{i=1}^{m} \lambda_i\right)$$

and this proves the lemma.

LEMMA 5: Let $k \geq 0$ be a rational integer. Then there exists $U \epsilon P_n$ satisfying

$$|L_i(U)| \leq e^{-\lambda_i} \quad (i = 1, \ldots, m), \quad 0 < |U| \leq e^k$$

if

$$(n-m)k \geq \sum_{i=1}^{m} \lambda_i - \delta(\wedge).$$

PROOF: Let $\mathfrak{J}$ be defined as in last corollary, then Vol $\mathfrak{J} \geq 1$, if

$$\delta + (n-m)k - \sum_{i=1}^{m} \lambda_i \geq 0,$$

i.e. if

$$(n-m)k \geq \sum_{i=1}^{m} \lambda_i - \delta(\wedge).$$

In this case, by a result of Mahler (1941, p. 511) we get $U \epsilon P_n$ satisfying the required properties.

PROOF (THEOREM 2): By taking $\lambda_1 = \lambda_m = \lambda$ in lemma 5, we get $U \epsilon P_n$ satisfying

$$|L_i(U)| \leq e^{-\lambda} \quad (i = 1, \ldots, m), \quad 0 < |U| \leq e^{\gamma},$$

where $\gamma \geq \dfrac{m\lambda - \delta}{n-m}$ is a non-negative integer. If λ is large, we define γ by

$$0 \leq \gamma - 1 < \frac{m\lambda - \delta}{n-m} \leq \gamma.$$

Thus for large values of λ, there exists $U \epsilon P_n$ satisfying

$$|L_i(U)| \leq e^{-\lambda} \quad (i = 1, \ldots, m), \quad 0 < |U| < \exp\left[\frac{m\lambda - \delta}{n-m} + 1\right],$$

and we get a solution $U \epsilon P_n$ of (2.1), unless $\max_{i=1,\ldots,m} |L_i(U)| = 0$ for some non-zero $U \epsilon P_n$, which is denied by the hypothesis. Hence the theorem follows.

ACKNOWLEDGEMENT

Finally the author thanks Professor R. P. Bambah, F.N.I., for his kind guidance and valuable help in writing this paper.

REFERENCES

Armitage, J. V. (1957). The product of n-linear forms in field of series. *Mathematika*, **4**, 132–137.

Cassels, J. W. S. (1957). An Introduction to Diophantine Approximation. Cambridge University Press.

Lutz, E. (1955). Sur les approximationes diophantiennes lineaires p-adiques. Actualities scientifique tracts, 1224. Hermann & Co., Paris.

Mahler, K. (1941). An analogue to Minkowski's geometry of numbers in a field of series. *Ann. Math.*, **42**, 488–522.