MATHEMATICS OF COMPUTATION, VOLUME 33, NUMBER 145
JANUARY 1979, PAGES 201–216

Minkowski Reduction of Integral Matrices

By John L. Donaldson

Abstract. In 1905 Hermann Minkowski introduced his theory of reduction of positive definite quadratic forms. Recently, Hans J. Zassenhaus has suggested that this theory can be applied to the problem of row reduction of matrices of integers. Computational investigations have shown that for matrices with more columns than rows, the number of steps required for reduction decreases drastically. In this paper it is proved that as the number of columns increases, the probability that a matrix is Minkowski reduced approaches one. This fact is the motivation behind the introduction of a modified version of Minkowski reduction, resulting in a reduction procedure more suitable for computation.

1. Introduction. In 1905 Hermann Minkowski introduced his theory of reduction of positive definite quadratic forms [1]. This theory is one of the essential foundations of the geometry of numbers. Recently, Hans J. Zassenhaus has suggested that Minkowski reduction can be applied to the problem of row reduction of matrices of integers [2]. It is the study and development of this idea that forms the basis of this paper; emphasis is placed particularly on the reduction algorithm, as adapted to machine computation. Section 2 is devoted to the theoretical foundations of the subject, including an outline of Minkowski's original work. Statistical methods are used in Section 3 to examine the algorithm, and the result is the main theorem of the dissertation. This theorem gives a relationship between the dimensions of a matrix and the probability that it is reduced. In Section 4, motivated by the theorem, a modified version of Minkowski reduction is defined and developed.

2. Theoretical Background. We are concerned with matrices of integers. Two matrices A and B are said to be unimodularly equivalent if and only if there exists a unimodular integral matrix U such that $A = UB$. Our aim here is to define a canonical representative, the "reduced" matrix, in each class, and to provide an algorithm for finding the reduced matrix equivalent to a given matrix.

2.1. *The Hermite Normal Form.* The usual method of reduction of integral matrices was introduced by Hermite in 1851 [3]. Among more recent accounts of his theory is the one by MacDuffee [4]. Hermite showed that every matrix is unimodularly equivalent to one in upper triangular form, such that each entry above the diagonal is bounded by one-half the magnitude of the diagonal entry directly below it. Such a matrix is said to be in Hermite normal form.

Received November 6, 1975; revised April 7, 1978.
AMS (MOS) subject classifications (1970). Primary 10E25, 15A36.

0025-5718/79/0000-0013/$05.00

Several algorithms to compute the Hermite normal form of a given matrix have have been published, all of them similar [5], [6], [7]. Implementation of these algorithms for machine computation suffers one major drawback, that of overflow. The computer investigations by Zassenhaus and David Ford verified the fact that the numbers generated by the reduction procedure can increase rapidly in magnitude, beyond the capacity of the typical machine, even when the matrices to be reduced have as few as three or four rows and entries bounded in magnitude by 100.

2.2. *Matrices and Modules.* Before we pursue our alternative to Hermitian reduction, we shall mention the connection between a matrix and its associated Z-module, first developed by Chatelet [8]. Suppose A is an $m \times n$ real matrix, with rows $a_1, a_2, \ldots, a_m$. With addition and scalar multiplication defined componentwise, the Z-module generated by $a_1, \ldots, a_m$ is called the *row module* of A. $a_1, \ldots, a_m$ form a basis for the row module if and only if A is nonsingular. If A is nonsingular, there is a one-to-one correspondence between matrices unimodularly equivalent to A (modulo the ordering of the rows) and bases of the row module. Thus, reduction is equivalent to selecting a canonical basis for the row module. It is this correspondence which leads us to develop the theory of Minkowski reduction in terms of modules.

Remark. The row module of A may be thought of as a lattice in real n-dimensional space. It is here that the connection with the geometry of numbers lies. Minkowski's work is done in terms of lattices.

Remark. Although the primary concern is with integral matrices, the reduction theory is valid for real matrices. Most of the work following will be in this more general setting.

2.3. *Gauge Functions.* The basic idea of Minkowski reduction is to choose a matrix from each equivalence class whose rows are as short as possible, according to some definition of length. Minkowski used the common Euclidean length; this was generalized to the gauge function by Weyl [9]. Since we can consider any Z-module of row vectors as a subset of a real n-dimensional vector space, we shall define gauge functions on real vector spaces.

Definition. Let V be a real vector space. A gauge function on V is a function $f : V \longrightarrow R$ such that for $x, y \in V$ we have:

1. $f(x) > 0$, except $f(0) = 0$.
2. $f(tx) = |t| f(x)$ for all $t \in R$.
3. $f(x + y) \leqslant f(x) + f(y)$.

With f as the norm, V becomes a finite-dimensional normed linear space. It follows that all gauges on V generate the same topology [10].

By means of the correspondence between matrix rows and elements of the row module we now have a means by which to measure the length of rows.

2.4. *Minkowski Reduction.* We shall define a Minkowski reduced matrix by first considering the row module M. We assume we are given a gauge function f on M.

Definition. A set of elements of M, $x_1, \ldots, x_k$, is called a *primitive system* if it can be extended to a basis of M.

Definition. A basis of M is called *Minkowski reduced* if the following properties are satisfied:

1. x_1 is a shortest element of M.

2. x_k $(k = 2, \ldots, m)$ is a shortest element among all elements of M which together with $x_1, \ldots, x_{k-1}$ form a primitive system.

This definition is equivalent to the following:

Definition. A basis of M is called *Minkowski reduced* if the following condition is satisfied:

$$f(x_k) \leqslant f(s_1x_1 + s_2x_2 + \cdots + s_mx_m) \text{ for all systems } s_1, \ldots, s_m \text{ of integers with greatest common divisor } (s_k, \ldots, s_m) \text{ equal to one.} \tag{1.5.1}$$

The second definition is most useful for matrices, for the inequalities (1.5.1) can be checked once we are given the matrix and the function f. The algorithm for reducing a matrix is based on the first so-called finiteness theorem:

THEOREM. *When the gauge function used is the Euclidean norm* $f(x_1, \ldots, x_n) = (x_1^2 + \cdots + x_n^2)^{1/2}$, *then we can select finitely many of the inequalities* (1.5.1) *from which all the rest follow.*

The theorem is true only for gauge functions in the form of positive definite quadratic forms [11].

Thus, the algorithm is as follows: We are given a matrix A, with rows $a_1, \ldots, a_m$. We test all the finitely many inequalities. If all are satisfied, we are done. If on the contrary $f(a_k) > f(s_1a_1 + \cdots + s_ma_m)$, we may apply a unimodular transformation to replace a_k by $s_1a_1 + \cdots + s_ma_m$ and leave $a_1, \ldots, a_{k-1}$ fixed (that this is possible follows from the fact that $(s_k, \ldots, s_m) = 1$). Eventually, all the inequalities must be satisfied, and the matrix will be Minkowski reduced.

THEOREM. *If all the inequalities* (1.5.1) *are strict, the Minkowski reduced basis is unique. If not, there are at most finitely many different such bases.*

Since our concern is the determination of a unique representative of each class, this is important. In the case where we must select from several different reduced matrices in a class, Zassenhaus uses the lexicographic ordering [2]. This consideration completes the algorithm.

When $m = 2$, 3, or 4, the necessary inequalities are those for which the coefficients s_i take on the values 0 or ± 1. In these cases the number of tests is small, and the unimodular transformations used in the reduction steps consist merely of replacing the row a_k by the row $s_1a_1 + \cdots + s_ma_m$.

However, when $m \geqslant 5$, the number of inequalities to be tested begins to increase rapidly. A table of these for $m \leqslant 6$ is found in Tammela [16]. When $m \geqslant 7$, we may no longer assume $s_k = 1$ in (1.5.1), and this is another significant complication [12].

2.5. *The Zassenhaus-Ford Conjecture.* This algorithm for Minkowski reduction of matrices was programmed by David Ford, a student of Zassenhaus, for matrices of 2, 3, or 4 rows and a varying number of columns. The reduction procedure was

applied to randomly generated matrices, and the average number of steps required was computed for each dimension pair m, n. Ford's results produced the following unexpected phenomenon: Among matrices with a fixed number of rows, those with a larger number of columns required significantly fewer steps, on the average, to reduce. Zassenhaus and Ford conjectured that there was some provable connection between the number of columns and the average number of steps. This connection is the subject of Section 2.

The conjecture is important because of its implication for the efficiency of the reduction algorithm. It would indicate another advantage of Minkowski reduction as opposed to Hermitian reduction, at least for matrices with more columns than rows. It is clear that no similar phenomenon should occur in the Hermitian case, since only the leftmost square of the matrix is actually used to make decisions in the algorithm, with the rest of the columns merely being carried along.

It is important to note that we must investigate the algorithm from a statistical point of view. Since any matrix can be lengthened by adding columns of zeros, without affecting its reduction, we cannot expect that any absolute measure of the speed of the algorithm would improve as the number of columns increased. We consider instead the average behavior of matrices.

3. A Theorem on Minkowski Reduction of Matrices. In this section we assume that the gauge function is the Euclidean norm, $f(x_1, \ldots, x_n) = (x_1^2 + \cdots + x_n^2)^{1/2}$, so the reduction inequalities (1.5.1) may be written as

$$(s_1a_1 + \cdots + s_ma_m)^2 \geqslant a_k^2 \text{ for all systems } s_1, \ldots, s_m \text{ with } (s_k, \ldots, s_m) = 1. \quad (x^2 \text{ is used to denote } x \cdot x \text{ throughout.}) \tag{3.0.1}$$

We note, in particular, that in a reduced matrix the rows must be in increasing order according to length, that is

$$a_1^2 \leqslant a_2^2 \leqslant \cdots \leqslant a_m^2. \tag{3.0.2}$$

We will assume here that all matrices have already been put in this form.

3.1. *The Probability Model.* We are going to investigate the probability that a real $m \times n$ matrix is reduced; and, therefore, we need to make the set of all such matrices into probability space. We shall consider each matrix as a point in mn-dimensional real space. Ideally, we would consider a random variable distributed uniformly in R^{mn}, so that probabilities of events A would be defined by

$$P(X \in A) = \frac{m(A)}{m(R^{mn})},$$

where m denotes Lebesgue measure and A is any measurable set. Obviously, we cannot do this, so we have a number of alternatives:

1. Give X some distribution other than uniform; say its density function is f.

Then

$$P(X \in A) = \int_A f(X). \tag{3.1.1}$$

2. X can have a uniform distribution if we think of R^{mn} as a conditional probability space [13]. Then

$$P(X \in A \mid X \in B) = \frac{m(A \cap B)}{m(B)} . \tag{3.1.2}$$

The disadvantage is that we can determine only conditional probabilities.

3. Since we are concerned primarily with integer matrices, we might define the density of a set of integer points in R^{mn} as follows:

$$D(A) = \lim_{k \to \infty} \frac{(\#\text{ of integer points in } A \cap S_k)}{(\#\text{ of integer points in } S_k)} , \tag{3.1.3}$$

where S_k denotes the sphere of radius k centered at the origin. This D is not, however, a probability, since it is only finitely additive.

In our computations in the following sections, we investigate directly the conditional probability $P_{m,n,k} = P(K \cap S_k \mid S_k)$ as defined by (3.1.2), where K is the set of nonreduced matrices in R^{mn}. We will show later the connection between P_k and the probabilities defined in (3.1.1) and (3.1.3).

3.2. *The Reduction Inequalities.* The set of reduced matrices, the measure of which is to be computed, is determined by the inequalities (3.0.1). To simplify matters, we make some observations on these inequalities.

In our list of necessary inequalities, we may assume we never have $s_k = 0$. For if $(s_k, \ldots, s_m) = 1$ and $s_k = 0$, then for some $t \geqslant k$, $s_t \neq 0$, $(s_t, \ldots, s_m) = 1$, and $(s_1 a_1 + \cdots + s_m a_m)^2 \geqslant a_t^2 \geqslant a_k^2$.

We may write the inequality (3.0.1) as

$$\sum_{i=1}^{m} \sum_{j=1}^{m} s_i s_j (a_i \cdot a_j) \geqslant a_k^2 \tag{3.2.1}$$

or

$$\sum_{i=1}^{m} s_i^2 a_i^2 - a_k^2 \geqslant \sum_{i=1}^{m} \sum_{j=1;j\neq i}^{m} s_i s_j (a_i \cdot a_j). \tag{3.2.2}$$

Inequality (2.2.2) will hold if

$$\sum_{i=1}^{m} s_i^2 a_i^2 - a_k^2 \geqslant \sum_{i=1}^{m} \sum_{j=1;j\neq i}^{m} |s_i s_j| \, |a_i \cdot a_j|. \tag{3.2.3}$$

To further simplify (3.2.3) we use the following lemma:

LEMMA. *If* $|a_i \cdot a_j| / a_i^2 \leqslant |s_i| / m |s_j|$ *for all* $1 \leqslant i, j \leqslant m$; s_i, $s_j \neq 0$, *then* (3.2.3) *is satisfied.*

Proof. Since we have observed that in (3.2.3) $s_k \neq 0$, we may assume without loss of generality that $s_i \neq 0$ for $i = 1, \ldots, m$. Now, if necessary, relabel so that

$a_m^2 \geqslant a_i^2$ for $i = 1, \ldots, m-1$, and further that $s_{m-1}^2 s_{m-1}^2 \geqslant \cdots \geqslant s_2^2 a_2^2 \geqslant s_1^2 a_1^2$. Then

$$\sum_{i=1}^{m} \sum_{j=1;j\neq i}^{m} |s_i s_j|\,|a_i \cdot a_j|$$

$$= 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} |s_i s_j|\,|a_i \cdot a_j| \leqslant 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{s_i^2 a_i^2}{m}$$

$$= 2 \sum_{i=1}^{m-1} \frac{m-i}{m} s_i^2 a_i^2 = \sum_{i=1}^{m-1} \frac{m+(m-2i)}{m} s_i^2 a_i^2$$

$$= \sum_{i=1}^{m-1} s_i^2 a_i^2 + \sum_{i=1}^{m-1} \frac{m-2i}{m} s_i^2 a_i^2$$

$$= \sum_{i=1}^{m-1} s_i^2 a_i^2 + \sum_{1\leqslant i<m/2} \frac{m-2i}{m} + \sum_{m/2<i\leqslant m-1} \frac{m-2i}{m} s_i^2 a_i^2.$$

Letting $j = m-1$ in the third summation, the right side becomes

$$\sum_{i=1}^{m-1} s_i^2 a_i^2 + \sum_{1\leqslant i<m/2} \frac{m-2i}{m} s_i^2 a_i^2 + \sum_{1\leqslant j<m/2} \frac{2j-m}{m} s_{m-j}^2 a_{m-j}^2$$

$$= \sum_{i=1}^{m-1} s_i^2 a_i^2 + \sum_{1\leqslant i<m/2} \frac{m-2i}{m} (s_i^2 a_i^2 - s_{m-i}^2 a_{m-i}^2).$$

The second summation here is less than or equal to 0, so we have

$$\sum_{i=1}^{m} \sum_{j=1;j\neq i}^{m} |s_i s_j|\,|a_i \cdot a_j|$$

$$\leqslant \sum_{i=1}^{m-1} s_i^2 a_i^2 \leqslant \sum_{i=1}^{m-1} s_i^2 a_i^2 + (s_m^2 a_m^2 - a_k^2) = \sum_{i=1}^{m} s_i^2 a_i^2 - a_k^2.$$

This is (3.2.3), so the lemma is proved.

COROLLARY. *If* $|a_i \cdot a_j|/a_i^2 \leqslant |s_i|/m\,|s_j|$ *for all* $1 \leqslant i, j \leqslant m$ *with* $s_i, s_j \neq 0$, *for each of the inequalities* (3.0.1), *then* A *is Minkowski reduced.*

Let M be a bound for the coefficients s_i in the inequalities (3.0.1). Then M depends on m but not n.

COROLLARY. *If*

$$\frac{|a_i \cdot a_j|}{a_i^2 + a_j^2} \leqslant \frac{1}{2(mM+1)} \quad \text{and} \quad \frac{a_i^2}{a_i^2 + a_j^2} \geqslant \frac{mM}{2(mM+1)}$$

for all $1 \leqslant i \leqslant j \leqslant m$, A *is Minkowski reduced.*

Proof. Multiplying the inequalities gives

$$\frac{|a_i \cdot a_j|}{a_i^2 + a_j^2} \frac{mM}{2(mM+1)} \leqslant \frac{a_i^2}{a_i^2 + a_j^2} \frac{1}{2(mM+1)}$$

or

$$\frac{|a_i \cdot a_j|}{a_i^2} \leqslant \frac{1}{mM} \leqslant \frac{|s_i|}{m|s_j|}.$$

3.3. *Probability Distributions of Vectors in R^{2n}.* The last corollary suggests that we examine the probability distributions of the functions $|x \cdot y|/(x^2 + y^2)$ and $x^2/(x^2 + y^2)$, where x and y are row vectors of R^n. According to the discussion in Section 3.1, we let

$$H_n(t) = P\left(\frac{x^2}{x^2 + y^2} \leqslant t \mid x^2 + y^2 \leqslant k\right)$$

(3.3.1)
$$= \frac{m\left(\left\{(x, y) : x^2 + y^2 \leqslant k, \frac{x^2}{x^2 + y^2} \leqslant t\right\}\right)}{m(\{(x, y) : x^2 + y^2 \leqslant k\})} \qquad (0 \leqslant t \leqslant 1)$$

and

$$G_n(t) = P\left(t \leqslant \frac{|x \cdot y|}{x^2 + y^2} \;\middle|\; x^2 + y^2 \leqslant k\right)$$

(3.3.2)
$$= \frac{m\left(\left\{(x, y) : x^2 + y^2 \leqslant k, t \leqslant \frac{|x \cdot y|}{x^2 + y^2}\right\}\right)}{m(\{(x, y) : x^2 + y^2 \leqslant k\})} \qquad (0 \leqslant t \leqslant \tfrac{1}{2}).$$

Using the substitution $kw = (x_1^2 + \cdots + x_m^2)(\operatorname{sgn} x_m)$ and the formula for the volume of an m-dimensional sphere, the resulting mutiple integral can be reduced to

$$(3.3.3)\quad H_n(t) = \frac{\Gamma(n + 1)}{\Gamma(\frac{1}{2}n + 1)\Gamma(\frac{1}{2}n)} \left[\int_0^t (1 - z)^{\frac{1}{2}n} z^{\frac{1}{2}n - 1}\, dz - \frac{1}{n}(1 - t)^{\frac{1}{2}n} t^{\frac{1}{2}n}\right].$$

For G_n the change of variables

$$x_i = (\mu_i + \upsilon_i)/\sqrt{2}, \qquad y_i = (\mu_i - \upsilon_i)/\sqrt{2},$$

results in an integral similar to the one for H_n, giving finally

$$(3.3.4)\quad G_n(t) = \frac{2\Gamma(n + 1)}{\Gamma(\frac{1}{2}n + 1)\Gamma(\frac{1}{2}n)} \left[\int_t^{\frac{1}{2}} (\tfrac{1}{2} + z)^{\frac{1}{2}n}(\tfrac{1}{2} - z)^{\frac{1}{2}n - 1}\, dz - \frac{1}{n}(\tfrac{1}{2} + t)^{\frac{1}{2}n}(\tfrac{1}{2} - t)^{\frac{1}{2}n}\right].$$

Simple estimates of these integrals yield the following upper bounds:

$$(3.3.5)\quad H_n(t) \leqslant \frac{\Gamma(n + 1)}{\Gamma(\frac{1}{2}n + 1)^2} \frac{(t - t^2)^{\frac{1}{2}n}}{2(1 - 2t)},$$

$$(3.3.6)\quad G_n(t) \leqslant \frac{\Gamma(n + 1)}{\Gamma(\frac{1}{2}n + 1)^2} \frac{(\frac{1}{2} - t^2)^{\frac{1}{2}n}}{2t}.$$

3.4. *The Conditional Probability of Reduction.* Let K be the set of nonreduced matrices in R^{mn}. Let $B_{i,j}$ be the set of matrices for which $|a_i \cdot a_j|/(a_i^2 + a_j^2) > 1/2(mM + 1)$ and $C_{i,j}$ $(i < j)$ be the set of matrices for which $a_i^2/(a_i^2 + a_j^2) < mM/2(mM + 1)$. Then by the corollary of Section 3.2,

$$K \cap S_k \subset \left(\bigcup_{i,j} B_{i,j} \bigcup_{i<j} C_{i,j} \right) \cap S_k,$$

so

$$\begin{aligned} P(K \cap S_k \mid S_k) &\leqslant \sum_{i,j} P(B_{i,j} \cap S_k \mid S_k) + \sum_{i<j} P(C_{i,j} \cap S_k \mid S_k) \\ &= \sum_{i,j} H_n\left(\frac{mM}{2(mM+1)}\right) + \sum_{i<j} G_n\left(\frac{1}{2(mM+1)}\right) \\ &= m(m-1)H_n\left(\frac{mM}{2(mM+1)}\right) + \frac{m(m-1)}{2} G_n\left(\frac{1}{2(mM+1)}\right) \end{aligned}$$

so

$$P_{m,n,k} \leqslant m(m-1) \frac{\Gamma(n+1)}{\Gamma(\tfrac{1}{2}n+1)^2} \left(\frac{m^2M^2 + 2mM}{4(mM+1)^2}\right)^{\frac{1}{2}n} (mM+1). \tag{3.4.1}$$

Applying Stirling's formula to estimate the Gamma function, we get

$$P_{m,n,k} \leqslant m(m-1) \frac{\sqrt{2n\pi}\left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n-1}\right)}{n\pi\left(\frac{n}{2e}\right)^n} \left(\frac{m^2M^2 + 2mM}{(mM+1)^2}\right)^{\frac{1}{2}n} \frac{1}{2^n} (mM+1),$$

(3.4.2)

$$P_{m,n,k} \leqslant m(m-1)(mM+1)\sqrt{\frac{2}{n\pi}} \left(1 + \frac{1}{12n-1}\right) \left(\frac{m^2M^2 + 2mM}{(mM+1)^2}\right)^{\frac{1}{2}n}.$$

Since $(m^2M^2 + 2mM)/(mM+1)^2 < 1$, we have

THEOREM. *For fixed m, as $n \to \infty$, $P_{m,n,k} \to 0$.*

Remark. Since A is reduced if and only if any scalar multiple of A is reduced, $P_{m,n,k}$ does not actually depend on k. Therefore, from now on we shall simply write $P_{m,n}$.

3.5. *The Other Probability Models.* We now consider the other situations discussed in Section 3.1. Suppose X, the random variable representing an $m \times n$ matrix, has density function f on R^{mn}. If f is of a certain type, then $P(K) = \int_K f(X) = P_{m,n}$.

Suppose f depends only on $|X|$, the Euclidean length of X. Then

$$\begin{aligned} P(K) &= \lim_{k\to\infty} P(X \in K \cap S_k) = \lim_{k\to\infty} P\left(\frac{X}{|X|} \in K \text{ and } |X| \leqslant k\right) \\ &= \lim_{k\to\infty} P\left(\frac{X}{|X|} \in K\right) P(|X| \leqslant k) \\ &= P(X \in K \mid |X| = 1) \lim_{k\to\infty} P(|X| \leqslant k) = P_{m,n}. \end{aligned}$$

Now we consider the density of the integer points of K.

$$D(K) = \lim_{k\to\infty} \frac{(\#\text{ of integer points in } K \cap S_k)}{(\#\text{ of integer points in } S_k)}$$

$$= \lim_{k\to\infty} \frac{(\#\text{ of integer points in } k(K \cap S_1))}{(\#\text{ of integer points in } kS_1)}$$

$$= \lim_{k\to\infty} \frac{k^{mn}m(K \cap S_1)}{k^{mn}m(S_1)} = \frac{m(K \cap S_1)}{m(S_1)} = P_{m,n}.$$

4. Extended Reduction.

4.1. *Reduction by Submatrices.* The purpose of this section is to investigate Minkowski reduction, with the aim of making it more suitable for computation. We recall that the main difficulty in Minkowski reduction is the large number of inequalities to be tested. The theorem of Section 2 suggests that some type of block-wise reduction might be advantageous; since long thin matrices are easy to reduce, perhaps we can develop some connection between reduction of a matrix and its submatrices. For example, given a 24×24 matrix A, we break it up into six 4×24 matrices $A_1, \ldots, A_6$. By reducing $A_1, \ldots, A_6$, we hope to be able to find the reduced form of A.

Unfortunately, even if all submatrices of a matrix are reduced, there is still a possibility that a linear combination of all its rows may give a further reduction. In fact, the inequalities to be checked involving all the rows will be the most numerous and difficult to use.

To circumvent this difficulty, we must be able to make final the choice of the submatrices $A_1, \ldots, A_k$ from the row module; that is, no more operations involving rows of different submatrices should be necessary, once $A_1, \ldots, A_k$ are chosen. Also, further reduction of the individual submatrices should be possible without altering this choice.

Minkowski's definition of reduction will not permit this type of scheme; we need a new definition. To do so, we need a way of choosing submatrices of a matrix which generalizes Minkowski's method of choosing rows, which is based on the gauge function. We will obtain a gauge function on matrices by first looking at the Grassmann algebra of a matrix.

4.2. *The Extension of the Gauge Function.* Our gauge functions were originally defined on a real vector space V, and we shall extend them to the Grassmann algebra G of that vector space. We recall that the Grassmann algebra is an anticommutative associative algebra with vector space basis $\{e_1^{r_1} \wedge e_2^{r_2} \wedge \cdots \wedge e_m^{r_m} : r_i = 0 \text{ or } 1\}$, where $e_1, \ldots, e_m$ is a basis of V. We can write $G = G_0 \oplus G_1 \oplus \cdots \oplus G_m$, where G_i is the set of homogeneous elements of degree i. In particular, $G_1 \cong V$.

Definition. Let f be a gauge function on V. Then $\bar{f} : G \to R$ is called an extension of f if:

1. $\bar{f}|_{G_1} = f$.

2. $\bar{f}$ is a gauge function on G, considered as a vector space.

3. $\Lambda : G \times G \longrightarrow G$ is continuous in the topology defined by $\bar{f}$.

In fact, condition 3 is not necessary, for every bilinear mapping of finite-dimensional normed linear spaces is continuous.

4.3. *The Gauge Function on Submatrices.* We construct a map q from the set of matrices into the Grassmann algebra as follows: Let A be a matrix, made up of the row vectors $a_1, \ldots, a_m$. Then $q(A) = a_1 \wedge a_2 \wedge \cdots \wedge a_m$. If f is a gauge function on a vector space containing $a_1, \ldots, a_m$, and if $\bar{f}$ is its extension to the Grassmann algebra, then $\bar{f} \circ q$ will be the gauge by which we measure the size of submatrices of a matrix. To complete the connection we note that

PROPOSITION. *If L is any linear transformation of V, and $a_1, \ldots, a_m$ is a basis of V, then*

$$L(a_1) \wedge \cdots \wedge L(a_m) = (\det L) a_1 \wedge \cdots \wedge a_m \quad [17].$$

COROLLARY. *If B is any $m \times m$ matrix, then $f(BA) = |\det B| f(A)$.*

COROLLARY. *If U is any $m \times m$ unimodular matrix, then $f(UA) = f(A)$.*

COROLLARY. *A is row-dependent if and only if $f(A) = 0$.*

These are the important properties we need for our applications to matrices. Corollary 2 allows us to consider f as a function on submodules of the row module, as follows: Let N be a submodule. Let $b_1, \ldots, b_k$ be a basis of N. Define $f(N) = f(b_1 \wedge \cdots \wedge b_k)$. Since any other basis of N is obtained from $b_1, \ldots, b_k$ by a unimodular transformation, f is well-defined.

Example. The function $f(A) = \sqrt{\det(AA^{\mathrm{T}})}$ is a gauge on matrices. It corresponds to the function f on the Grassmann algebra defined by $f(\Sigma x_i g_i) = \sqrt{\Sigma x_i^2}$, where $g_1, \ldots, g_k$ is the standard basis of G. In this case we have $f(A \wedge B) \leqslant f(A) f(B)$.

4.4. *Reduction by Submodules.* In Minkowski reduction, we selected a basis for the row module M of a matrix by choosing basis vectors as short as possible, according to the gauge function f. We will now define a reduction procedure which selects submodules of M as small as possible according to the extended gauge function, which we shall also call f.

Definition. Let $l_1, l_2, \ldots, l_k$ be positive integers such that $l_1 + l_2 + \cdots + l_k \leqslant m$. Let M be an m-dimensional free Z-module. A system of submodules $N_1, \ldots, N_k$ of M is called $(l_1, \ldots, l_k)$ primitive if

1. $\dim N_i = l_i$, $i = 1, \ldots, k$.

2. $N_1, \ldots, N_k$ are linearly independent.

3. $r_1 a_1 + \cdots + r_k a_k \in A_1 + \cdots + A_k$ (r_i real, $a_i \in A_i$), then $r_1, \ldots, r_k \in Z$.

In particular, an (l_1) primitive system is called a primitive submodule. A primitive $(1, 1, \ldots, 1)$ system is called a primitive system of vectors. If $l_1 + \cdots + l_k = m$, a primitive $(l_1, \ldots, l_k)$ system is called a basic $(l_1, \ldots, l_k)$ system.

THEOREM; *If $l_1 + \cdots + l_p = m = \dim M$, and $1 \leqslant k \leqslant p$, then any primitive $(l_1, \ldots, l_k)$ system of submodules of M can be extended to a basic $(l_1, \ldots, l_p)$ system of M.*

Proof. First, we show that any l_1-dimensional primitive submodule N_1 of M can be extended to a basic $(l_1, \ldots, l_p)$ system. Let $b_1, \ldots, b_{l_1}$ be a basis of the submodules N_1. We show that $b_1, \ldots, b_{l_1}$ is a primitive system of vectors.

Suppose $r_1 b_1 + \cdots + r_{l_1} b_{l_1} \in N_1$ (r_i real). Since $N_1 = Zb_1 + \cdots + Zb_{l_1}$, $r_1 b_1 + \cdots + r_{l_1} b_{l_1}$ can be expressed as $z_1 b_1 + \cdots + z_{l_1} b_{l_1}$, with integers $z_1, \ldots, z_{l_1}$. Since $b_1, \ldots, b_{l_1}$ are linearly independent, $r_i = z_i$, so $r_i \in Z$ for $i = 1, \ldots, l_1$.

Since $b_1, \ldots, b_{l_1}$ is a primitive system of vectors, it can be extended to a basis of M [14]. If we group these basis vectors in groups of $l_2, \ldots, l_p$ in any fashion, we obtain the desired basic $l_1, \ldots, l_p$ system for M.

Now suppose we start with a primitive $(l_1, \ldots, l_k)$ system $N_1, \ldots, N_k$. Then $N_1 + \cdots + N_k$ is a primitive submodule, for if $r(a_1 + \cdots + a_k) \in N_1 + \cdots + N_k$, where $a_i \in N_i$, than $ra_1 + \cdots + ra_k \in N_1 + \cdots + N_k$, and so $r \in Z$. We can extend $N_1 + \cdots + N_k$ to a basic $(l_1 + \cdots + l_k, l_{k+1}, \ldots, l_p)$ system $N_1 + \cdots + N_k$, $N_{k+1}, \ldots, N_p$, and $N_1, \ldots, N_p$ is a basic $(l_1, \ldots, l_p)$ system. This completes the proof.

We assume now that we have some gauge function f defined on submodules of M. If $f(N_1) \leqslant f(N_2)$, we say N_1 is smaller than N_2.

LEMMA. *Let M be a Z-module with basis $a_1, \ldots, a_m$, q the mapping of submodules into the Grassmann algebra G. Then $\{q(N) : N$ is a submodule of $M\}$ is a discrete set in G.*

Proof. Suppose $q(N_n)$ converges to $x \in G$. Let $\{b_{n1}, \ldots, b_{nk_n}\}$ be a basis of N_n. Let $b_{ni} = \Sigma_{j=1}^{m} c_{ni}^{(j)} a_j$. Then

$$q(N_n) = \left(\sum_{j=1}^{m} c_{n1}^{(j)} a_j\right) \wedge \cdots \wedge \left(\sum_{j=1}^{m} c_{nk_n}^{(j)} a_j\right) = \sum_{s=1}^{t} z_{ns} g_s,$$

where z_{ns} are integers and g_s are the elements of the form $a_1^{d_1} \wedge \cdots \wedge a_m^{d_m}$, $d_i = 0$ or 1. Since $\{g_s\}$ is linearly independent, we must have $\{z_{ns}\}$ converges for each s. Therefore, $z_{ns} = z_s$ for some integer z_s for every s for sufficiently large n. Therefore, $q(N_n) = \Sigma z_s g_s$ for sufficiently large n.

THEOREM. *In every nonempty set of submodules there is a smallest one.*

Proof. By the lemma, $\{f(N) : N$ is a submodule$\}$ is a discrete set of nonnegative real numbers; and, therefore, every subset has a least element r. Given any set of submodules S, and $N \in S$ with $f(N) = r$ has the required property.

Definition. A basic $(l_1, \ldots, l_p)$ system $N_1, \ldots, N_p$ of submodules of M is called reduced with respect to the extended gauge function f if the following conditions are fulfilled:

1. N_1 is the smallest l_1-dimensional submodule of M.

2. For $k = 2, \ldots, p$, N_k is the smallest l_k-dimensional submodule of M which, together with $N_1, \ldots, N_{k-1}$, forms a primitive $(l_1, \ldots, l_k)$ system of submodules.

In particular, a basic $(1, 1, \ldots, 1)$ system is a Minkowski reduced basis for M.

THEOREM. *Every $l_1 + \cdots + l_p$-dimensional Z-module M has a reduced basic $(l_1, \ldots, l_p)$ system.*

Proof. We construct the system inductively. The second theorem says we can choose N_1 to satisfy condition 1. The first theorem says that the set of submodules which together with $N_1, \ldots, N_{k-1}$ form a primitive system is nonempty; the second allows us to select a smallest one.

$(l_1, \ldots, l_p)$ reductions may be used to select a canonical basis for M, by further reducing each submodule N_i once it has been chosen. By combining $(l_1, \ldots, l_p)$ reductions we have many different reduction schemes, each of which yields a basis of M.

The consideration of the uniqueness of the basic system is similar to that of Minkowski reduction. Because of the discreteness, every choice made in the proof of Theorem 3 is from a finite number of submodules. Therefore, a module can have only finitely many reduced basic systems.

4.5. $(l_1, \ldots, l_p)$ *Reduction of Matrices.* A matrix A with rows $a_1, \ldots, a_m$ may be thought of as consisting of p submatrices $A_1, \ldots, A_p$, such that A_1 consists of the rows $a_1, \ldots, a_{l_1}$, A_2 consists of the rows $a_{l_1} + 1, \ldots, a_{l_2}$ and so on. A will be called $(l_1, \ldots, l_p)$ reduced if the submodules $N_1, \ldots, N_p$ associated with $A_1, \ldots, A_p$ form a basic $(l_1, \ldots, l_p)$ system in the row module M. Again, by combining reductions we can construct a reduction scheme which selects a particular matrix from each equivalence class. As an example, we shall investigate the computation of the reduced form of a matrix according to a particular scheme, namely successive $(1, m-1)$ reductions.

4.6. *The* $(1, m-1)$ *Reduction Scheme.* The particular reduction scheme will be defined as follows: Denote by $A^{(k)}$ the matrix formed by the last k rows of A. Then A is reduced if and only if for each $k = 2, \ldots, m$, we have that $A^{(k)}$ is a $(1, k-1)$ reduced matrix. The procedure used for finding the reduced matrix according to this scheme would be simply to first $(1, m-1)$ reduce A, then $(1, m-2)$ reduce $A^{(m-1)}$, then $(1, m-3)$ reduce $A^{(m-2)}$, and so on.

The method for the actual $(1, k-1)$ reduction of each matrix is given by the following:

PROPOSITION. *Let A be an $m \times n$ matrix with rows $a_1, \ldots, a_m$. Then A is $(1, m-1)$ reduced if and only if $f(a_1) \leqslant f(s_1a_1 + \cdots + s_ma_m)$ for any relatively prime integers $s_1, \ldots, s_m$, and*

$$f\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix} \leqslant f\begin{pmatrix} a_2 + t_2a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m + t_ma_1 \end{pmatrix}$$

for any system of integers $t_2, \ldots, t_m$. Here

$$\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a_2 + t_2a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m + t_ma_1 \end{pmatrix}$$

denote the matrices whose rows are $a_2, \ldots, a_m$ *and* $a_2 + t_2a_1, \ldots, a_m + t_ma_1$, *respectively.*

Proof. First, we prove the sufficiency. If $f(s_1a_1 + \cdots + s_ma_m) < f(a_1)$, then the first condition of the definition of reduction is not satisfied; if

$$f\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix} > f\begin{pmatrix} a_2 + t_2a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m + t_ma_1 \end{pmatrix},$$

then the second condition is not satisfied.

Next the necessity. Suppose A is not $(1, m-1)$ reduced. If a_1 is not the shortest vector in the row module of A, then $f(a_1) > f(s_1a_1 + \cdots + s_ma_m)$ for some integers $s_1, \ldots, s_m$. If $d = \text{g.c.d.}(s_1, \ldots, s_m)$, then

$$f\left(\frac{s_1}{d}a_1 + \cdots + \frac{s_m}{d}a_m\right) = \frac{1}{d}f(s_1a_1 + \cdots + s_ma_m) < f(a_1)$$

which contradicts the first hypothesis.

If

$$\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix}$$

is not the smallest $(m-1)$-dimensional submodule of the row module which along with a_1 forms a primitive system, then

$$f\begin{pmatrix} b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix} < f\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix},$$

where $a_1, b_2, \ldots, b_m$ is a primitive system, and therefore a basis of the row module. We then have that

$$\begin{pmatrix} a_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix} = U\begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix}$$

for some unimodular U in the form

$$U = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline t_2 & & & \\ t_3 & & & \\ \vdots & & U^* & \\ t_m & & & \end{array}\right).$$

Let

$$V = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ 0 & & & \\ \vdots & & (U^*)^{-1} & \\ 0 & & & \end{array}\right).$$

Then

$$V\begin{pmatrix} a_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = VU\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline t_2 & & & \\ \vdots & & (Id_{m-1}) & \\ t_2 & & & \end{array}\right)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}.$$

$$= \begin{pmatrix} a_1 & & \\ a_2 & + & t_2 a_1 \\ & \vdots & \\ a_m & + & t_m a_1 \end{pmatrix}.$$

So

$$(U^*)^{-1}\begin{pmatrix} b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix} = \begin{pmatrix} a_2 + t_2 a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m + t_m a_1 \end{pmatrix}.$$

Thus,

$$f\begin{pmatrix} a_2 + t_2 a_1 \\ \cdot \\ \cdot \\ a_m + t_m a_1 \end{pmatrix} = f\left((U^*)^{-1}\begin{pmatrix} b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix}\right) = f\begin{pmatrix} b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix} < f\begin{pmatrix} a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{pmatrix},$$

which contradicts the second hypothesis.

To obtain the reduced matrix, we must, therefore, be able to minimize the functions

$$F(s_1, \ldots, s_m) = f(s_1 a_1 + \cdots + s_m a_m)$$

and

$$D(t_2, \ldots, t_m) = f\begin{pmatrix} a_2 + t_2 a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_m + t_m a_1 \end{pmatrix}$$

over Z. We can simplify

$$D(t_2, \ldots, t_m) = f((a_2 + t_2 a_1) \wedge \cdots \wedge (a_m + t_m a_1))$$

$$= f(a_2 \wedge \cdots \wedge a_m + t_2 A_{(2)} + t_3 A_{(3)} + \cdots + t_m a_{(m)}),$$

where $A_{(i)} = a_2 \wedge \cdots \wedge a_{i-1} \wedge a_1 \wedge a_{i+1} \wedge \cdots \wedge a_m$.

In the case where the gauge function is the familiar Euclidean norm, F is just a positive definite quadratic form in $s_1, \ldots, s_m$ and D is a positive definite quadratic polynomial.

The improvement therefore over the standard Minkowski reduction is that instead of finding the successive minima of a module or form, we need only find the absolute minimum. In addition, we can expect on the basis of the theorem of Section 2 that the task of $(1, k-1)$ reducing each submatrix will become easier as we proceed, since the number of columns will remain fixed as the number of rows decreases.

APPENDIX

Computer Results. To verify the theorem of Section 2, a computer program was written by Merle Owdom of Ohio State's Instruction and Research Computer Center. The program generated matrices using the system's random number generator, then tested whether or not they were Minkowski reduced. The matrix entries were integers bounded in magnitude by 100, and were assumed to be distributed uniformly. For each dimension pair, 1000 matrices were tested, using the reduction inequalities. The results follow:

TABLE 1. *Percentage of Reduced Matrices*

m	n	% reduced	m	n	% reduced
2	4	47.4	3	4	13.7
2	8	73.4	3	8	44.9
2	12	84.3	3	12	65.2
2	16	93.1	3	16	77.6
2	20	95.1	3	20	87.9
2	24	97.1	3	24	94.0
2	28	98.5	3	28	96.0
2	32	99.2	3	32	98.2
2	36	99.1	3	36	98.1
2	40	99.8	3	40	99.3
2	44	99.9	3	44	99.6
2	48	99.9	3	48	99.7
2	52	100.0	3	52	99.9
2	56	100.0	3	56	100.0
2	60	100.0	3	60	100.0

Department of Mathematics
California State University, Los Angeles
Los Angeles, California 90032

1. HERMANN MINKOWSKI, *Gesammelte Abhandlungen.* II, pp. 53–100.
2. HANS ZASSENHAUS, "Bilinear spaces and reduction," Unpublished manuscript.
3. C. HERMITE, *J. Reine Angew. Math.*, v. 41, 1851, pp. 191–216.
4. C. C. MacDUFFEE, *The Theory of Matrices,* Chelsea, New York, 1956.
5. G. H. BRADLEY, *Math. Comp.*, v. 25, 1971, pp. 897–907.
6. J. B. ROSSER, *J. Res. Nat. Bur. Standards,* v. 49, 1952, pp. 349–358.
7. W. A. BLANKINSHIP, *Comm. ACM,* v. 9, 1966, p. 513.
8. A. CHATELET, *Ann. Ecole Norm.* III, v. 28, 1911, pp. 105–202.
9. H. WEYL, *Trans. Amer. Math. Soc.*, v. 48, 1940, pp. 126–164.
10. G. J. O. JAMESON, *Topology and Normed Spaces,* Chapman and Hall, London, 1974.
11. P. W. AITCHISON, *J. Austral. Math. Soc.*, v. 14, 1972, pp. 336–351.
12. S. S. RYSKOV, *Soviet Math. Dokl.*, v. 12, 1971, pp. 946–950.
13. A. RENYI, *Probability Theory,* North-Holland, Amsterdam and London, 1970.
14. B. van DER WAERDEN, *Acta Math.*, v. 96, 1956, pp. 265–309.
15. C. CHEVALLEY, *Fundamental Concepts of Algebra,* Academic Press, New York, 1956.
16. P. TAMMELA, *Soviet Math. Dokl.*, v. 14, 1973, p. 651.
17. MARVIN MARCUS, *Finite Dimensional Multilinear Algebra,* Part 1, Dekker, New York, 1973.