PRINCIPAL RINGS AND INVARIANT FACTORS

Jon A. Sjogren

Air Force Office of Scientific Research

Submitted 17 February 2003

to Detlef Laugwitz

Abstract. In this article we put forward a new look at the theory of principal and Invariant-Factor rings, with a view toward facilitating the formalization, automation, and archiving of results and their proofs. We take an elementary and constructive approach: standard techniques such as prime ideals and factorization of elements are avoided, and determinant constructions are minimized. Using such "computationally friendly" methods, the main existence and uniqueness results on invariant factors for a f.g. torsion module are derived, and several new algebraic constructions and results are found. The lattice of principal integral ideals for any commutative Bézoutian ring is explicitly constructed based on a first-order proof overlooked in the literature, together with a proof that this lattice is distributive. A "Lagrange quotient" theorem for finitely generated modules over any principal ring is stated for the first time. A very constructive new proof is given that a principal ring has the Hermite property, so is also an Invariant-Factor ring. A calculus that is needed in the ideal lattice, naturally yields a number of formulas valid for a function lattice.

Introduction

A movement has begun toward the systematic archiving of mathematical theories together with proofs that are more complete than is usual for printed publications. In the near future an announcement of a new result will include links that allow the reader (or "browser") to work through the proof to any desired degree of detail. Lemmas, definitions, theorems and proofs will easily be modified, and a fast proof checker will validate the entire working document.

Within a particular topic, there are many choices in putting together a foundation and further development. Certainly, the more basic theories from algebra, analysis and geometry should be developed first. More advanced results may then "import" or reference already established truths. The most ambitious project along these lines, underway before electronic storage and transmission of text was prevalent, was of course Bourbaki's *Éléments de Mathématiques*. Among the choices made by Bourbaki was to keep as a foundation for the rest of mathematics, the Zermelo-Frankel theory of sets. Bourbaki was able to assemble and integrate large

Key words and phrases. Bézoutian ring, module capacity, Smith canonical form, distributive semi-group lattice, invariant basis property.

areas of mathematics, using only the assumptions of the Z-F theory, new definitions, and the standard rules of inference. Proofs were rigorous in principle, though fully detailed proofs would be too lengthy and tedious to display.

In this article we examine one topic from commutative algebra relating to rings whose finitely presented modules may be given as a sum of cyclic modules. We call such rings (which are commutative with a unity 1) *Invariant-Factor* rings, but they are exactly the same as "Elementary Divisor Rings", and thus have a considerable literature. The purpose is to examine some of the fundamental results of this theory from the point of view of economy of logical development and algorithmic implementation. For example, if we are working with Euclidean rings, that possess a "fast" algorithm for the greatest common divisor, we should prefer to use such constructions and avoid factorization into prime factors. As is known, such primary factorization does not afford a mechanical algorithm of reasonable cost, even in specific cases (e.g., the ring of rational integers), much less in a uniform way over all rings.

We thus avoid the use of prime elements, prime or maximal ideals, factorization into irreducibles, Jacobson radical, fields of quotients, or localization. We not only present alternative constructions as algorithms, we also avoid use of such "computationally intensive" concepts in carrying out proofs. Thus we do not invoke the Jordan-Hölder theorem, Krull-Schmidt theorem, or use composition series methods. We take the point of view that a general determinant is a "computationally intensive" object. We avoid expansion of determinants, as well as co-factors, classical adjoints and exterior algebra constructions, in reproving results such as the uniqueness of the "Smith form" for any f.p. module over an IF ring. An important result that apparently is new is "Lagrange's theorem" for f.p. modules over a principal ring (P.I.R.). For this result we had to make use of some basic properties of linear systems, including Cramer's rule, but not classical adjoints.

In fact we have two approaches to the uniqueness of the invariant factors. One of them applies to all IF rings and does not use determinants, but does require the polynomial ring construction. The second, based on Lagrange's theorem, applies to all principal rings (P.I.R.) and requires the most elementary properties of the determinant. This approach to the uniqueness result, that the Smith form depends only on the module type and not the particular presentation, is then a motivation and an application of Lagrange's theorem.

Many important results in commutative and non-commutative algebra cannot efficiently be proved under our restrictions (without using maximal ideals and localization for example). Deep results about non-commutative principal rings contained in [10], and other results of "Goldie Theory" require powerful tools, that go beyond algorithmic constructibility. But in the commutative arena, it is surprising that classical results can be proved as well or better with primitive, and algorithmically robust, methods. In fact, new results in algebra are found, not in spite of, but because of the search for elementary proofs.

To obtain some results actually new to algebra, we examine the "module capacity" under mappings between f.g. torsion modules. In [15, pg. 191] "Lagrange's theorem for modules" is stated: the co-kernel of one module injected into another has for its *capacity* the quotient of the corresponding capacities. We show that this Lagrange theorem holds for P.I.R. (it is usually given less generally over a Euclidean domain) and conjecture that the same result holds over any IF domain. This result is then applied to the uniqueness of invariant factor decomposition.

The other example where we redo a classical proof is in showing that any Principal Ideal Ring (P.I.R.) is a Hermite ring, thus is also an IF ring. This latter follows since the process of putting a matrix successively into upper followed by lower triangular form terminates only with a diagonal matrix. The process must terminate or we would have an infinite chain of elements in strict division, which cannot hold in a principal ring (Noetherian property). In his celebrated paper [14], the author uses the standard tools of commutative algebra for this result, but it can be derived more nearly from first principles, without mentioning the Jacobson radical, Axiom of Choice, or using structural theorems involving valuation rings. We do need and provide a structural characterization of principal rings all of whose zero-divisors are nilpotent (NP rings or "special rings", [3]).

We give a first-order proof of the result implicit in [17], that the principal ideals of a Bézoutian ring form a lattice. The fact that this lattice is a distributive lattice is a key element of our new "Lagrange Theorem". Thus we have in some direction a generalization of theorems of [12]. In Appendix A, we list some join and meet formulas that are needed, and in Appendix B, show how these equations correspond to true formulas relating sup and inf for functions, whose domain can be almost anything, into a totally ordered abelian group.

1. Presentation and relation matrices

Good references for notation and basic results on invariant factor rings (IF rings or IFR), also known as "elementary divisor rings" are the papers of [14], [7] and the books of [3] and [15]. A ring $\mathcal{R}$ (meaning commutative with unity 1) may have the property that any matrix may be put into *lower triangular form* by *left* multiplication by an *invertible* square matrix. Thus if A is an $m \times n$ matrix over $\mathcal{R}$, there exists $n \times n$ matrices V and W such that $V \cdot W = I_{n\times n}$ and $A \cdot V = B$ where B has the same dimensions as A, and is such that its entries satisfy $B_{ij} = 0$ if $j > i$. By considering the transpose, it is seen that any matrix with entries from $\mathcal{R}$ may be put into *upper* triangular form through *left* multiplication with an invertible matrix of the appropriate size. We now give a name to such commutative rings (the non-commutative case is also studied), as *Hermite* rings.

It is known from the references that an Hermite ring has the Bézout property: any finitely generated ideal is principal. Furthermore, any *entire* Bézoutian ring (integral domain) is a Hermite domain. In Section 4 we prove again a result of Kaplansky, that a P.I.R. is an IF ring. In fact there is a persistent interest in IF rings that have zero-divisors, remarks of [15, pg. 169] notwithstanding.

A matrix A is *diagonalizable* if there are unimodular (invertible) matrices U and V over $\mathcal{R}$, of appropriate sizes, so that the entries of $B = UAV$ are such that $B_{ij} = 0$ unless $i = j$. A usual criterion is when $b_1 = B_{11}, \ldots$, for these diagonal entries successively to divide each other, that is

$$b_1 \,|\, b_2 \,|\, \cdots \,|\, b_k, \tag{1}$$

where k is the smaller of the dimensions of A. If so, the new diagonal matrix is said to be in "Smith form". A result in [17] that we revisit in Section 2 shows that all matrices over $\mathcal{R}$ are diagonalizable exactly when all matrices are diagonalizable with condition (1). Such a ring is now called an Invariant-Factor ring, or IFR.

Special forms for matrices over $\mathcal{R}$ are closely related to the decomposition of a finitely generated (f.g.) module. Such a module M can be *presented* by a presentation matrix P, where we take the convention that the columns represent *generators*

of M, and the rows represent their mutual *relations*. A module M is *finitely presented*, (f.p.) by P, if P is a finite matrix (has a finite number of rows). If $\mathcal{R}$ is Noetherian, say a P.I.R., then any f.g. module M is also f.p. Now suppose that $\mathcal{R}$ is a Hermite ring, not necessarily principal, that is, not necessarily satisfying an ascending chain condition on ideals. A f.p. module M is represented by an $m \times n$ matrix P. If $m \geq n$, P can be put into upper triangular form by means of a left multiplication. Since the left-multiplying matrix U is invertible, we obtain a valid (defining) presentation matrix for M, $Q = UP$, by reducing the relations. The rows of Q beyond the n-th row should be zero rows; they may be deleted to yield a *square* presentation matrix Q'. If $m < n$, we put P into lower triangular form by means of a right multiplication (assigning new generating elements in terms of the old ones). This process is reversible so we still have a presentation for M, $R = PV$. The columns of R beyond the n-th column are zero columns: they represent free factors of the module. Let R' be the leftmost square part of R (also obtained by removing these "free columns". Also if R' or in the other case Q' has any 0 diagonal entry, the module M has a free part, it is not torsion.

To sum up, if $\mathcal{R}$ is Hermite, there is an equivalence between f.p. torsion $\mathcal{R}$-modules, and matrices P with at least as many rows as columns, and such that the upper triangular form of P has no zeroes on its diagonal (i.e., no all-zero columns). Hence, in dealing with Hermite rings and f.p. torsion modules, one need only deal with a square matrix, the upper triangular, row-truncated, form of P.

Putting the presentation matrix in triangular form amounts to giving a "composition series" for the torsion module M. that is

$$M_1 \subset M_2 \subset \cdots \subset M_n = M,$$

where each quotient module M_{i+1}/M_i is a cyclic module, and thus isomorphic to $\mathcal{R}/I_i$ for some ideal I_i. If we succeed in finding a *diagonal* square presentation matrix P for M, the quotient factors split and we can express M as a direct sum

$$M \simeq \frac{\mathcal{R}}{I_1} \oplus \cdots \oplus \frac{\mathcal{R}}{I_n}.$$

In case $\mathcal{R}$ is a P.I.D., the ideals I_i have a single generator and can be written (r_i), giving the well-known decomposition of an arbitrary f.g. module over such a ring.

Over a skew-field (division ring), a square matrix P can be reduced to diagonal form using only row operations familiar from "Gaussian elimination". For a general commutative ring, we cannot "zero out" an entry b in row j by an element a in row i unless $a \,|\, b$. Multiplying row j by c so that $b \cdot c$ is divisible by a makes possible the zeroing out. If the matrix of interest is a relation matrix P however, this operation of multiplying row j by c may not preserve the property "defining matrix of relations", but may be useful nevertheless. In fact, all the rows of P' (after this operation) still represent valid relations in the module M represented. The module presented by P' is a submodule of M. We say that P' constitutes a *valid matrix of relations* (VMR) for the f.g. module M.

A Euclidean ring is always Hermite, and we can obtain from a square presentation matrix P for a f.g. torsion module M, another such presenting matrix P', now upper triangular, through a sequence of row operations from the following list, making use of the Euclidean algorithm to give the gcd of two elements.

(1) interchange two rows;

(2) add a multiple of one row to another row, giving a new row to replace the second-named row;
(3) multiply a row by a unit, giving a row to replace the old row.

Starting with one VMR P for M, we obtain another VMR (Q) by performing any sequence of the above row operations and in addition

(3') multiply a row by a non-zero element, giving a row to replace the old row.

Thus the rows of Q are *valid* in the sense that if the row is $[\beta_1, \ldots, \beta_n]$ and the columns (generators) are denoted by $\{\xi_i\}$, $i = 1, \ldots, n$, we have that the element

$$\sum_{i=1}^{n} \beta_i \xi_i$$

is zero in the module M. Now, using the three rules we can mimic the "row-echelon" construction of Gaussian elimination over a field and in fact obtain a *diagonal* VMR. Calling this matrix A, we may see that $\gamma = \prod_{i=1}^{n} a_{ii}$ is an *annihilator* of M ($\gamma \cdot M = 0$). Over a P.I.D., we could have used $\mathrm{lcm}(a_{11}, \ldots, a_{nn})$ as annihilator instead.

Proposition 0. *Let $\mathcal{R}$ be an arbitrary commutative ring, M an $\mathcal{R}$-module, and let A be a valid matrix of relations for M, which is square diagonal. Then the product of the diagonal entries of A annihilates all of M. If $\mathcal{R}$ is a ring where least common multiplier is defined (or its principal ideal), then the* lcm *of the diagonal entries is an annihilator.*

Now let $\mathcal{R}$ be any commutative ring. Of special interest is the case of an f.g. torsion module W over the ring $\mathcal{S} = \mathcal{R}[X]$ of polynomials of finite degree. As is well-known [15, chpt. 9], the module structure may be defined by a square matrix over $\mathcal{R}$ called A. Then we get a module isomorphism $W \simeq \mathcal{R}^n$. A presentation matrix can also readily be written out:

$$XI - A = \begin{bmatrix} X - \alpha_{11} & -\alpha_{12} & \cdots & -\alpha_{1n} \\ -\alpha_{21} & X - \alpha_{22} & & \\ \vdots & & \ddots & \vdots \\ -\alpha_{n1} & & & X - \alpha_{nn} \end{bmatrix}.$$

Here, $\{\alpha_{ij}\}$ are the entries of A. The observation we made that W has a diagonal VMR, called B, shows already that A satisfies a polynomial relation itself. In fact we found an annihilator for all of W, in the form of a polynomial in X. When A is substituted for X in this polynomial, we obtain a matrix which has value zero on *any* n-vector of $\mathcal{R}$ entries. Hence this matrix must be the zero matrix. In case the ground ring $\mathcal{R}$ is a field, we have that $\mathcal{S}$ is Euclidean, hence a P.I.D., and obtain a conceptual proof of the Cayley-Hamilton theorem.

Proposition 1 "Cayley-Hamilton". *If A is an $n \times n$ matrix over the field $\mathcal{R}$, put the polynomial matrix $XI - A$ into upper triangular form Q by means of row operations (1)-(3). These operations only change the "determinant"* $\det(XI - A)$ *by multiplication by a non-zero $\beta \in \mathcal{R}$ (unit of $\mathcal{S}$). Let $q(X) = \prod_{i=1}^{n} q_{ii}$, then A is a matrix root of $q(X)$, that is, $q(A)$ is the $n \times n$ matrix of zeros.*

Proof. Using the Hermite property of $\mathcal{S}$ and the Euclidean algorithm through row operation (2), we obtained a square upper-triangular presentation matrix Q of the

same module W. Now one can use element q_{22} to eliminate element $q_{1,2}$ by first multiplying the first row by q_{22} according to operation (3'), obtaining a new VMR. Proceeding to entry q_{33} and continuing in this manner ultimately gives a diagonal VMR for module W, called Q' whose corner entry q'_{ii} is divisible by $q'_{i-1,i-1}$. Hence q'_{11} is an annihilator for W, but by construction it is just the product of the q_{ii}, $i = 1, \ldots, n$. Hence $q(A) = \mathbf{0}$, since its action on every basis element of $\mathcal{R}^n$ gives $\vec{0}$.

We mention one more application of VMR's which arise in another way from an $\mathcal{S}$-module. The result exhibits a duality with the Cayley-Hamilton Theorem.

Proposition 2 "Nakayama's Lemma"[1]. *Let M be an f.g. torsion module over the polynomial ring $\mathcal{S} = \mathcal{R}[X]$ and Q be square of size n, a valid matrix of relations for M, satisfying*

$$Q_{ij} \equiv \begin{cases} 0 \pmod{X}, & i \neq j \\ 1 \pmod{X}, & i = j \end{cases}.$$

Thus diagonal entries have constant term unity, and off-diagonal entries are divisible by X. Then there is an annihilator of M,

$$\theta(X) = \sum_{j=0}^{K} \beta_j X^j$$

with constant term $\beta_0 = 1$

Proof. Starting with the $(1,1)$ entry of Q, we perform a "pivot", zeroing out the rest of the first column by means of VMR operations 1) – 3) above. Specifically, multiply each row except the first by Q_{11}. This operation preserves the form of the entries $\pmod{X}$ as in the hypothesis. Next, subtract the appropriate multiple of the first row from each of the other rows to zero out the entries $Q_{j,1}$, $j > 1$. In this latter operation, all entry changes involved were addition by a polynomial which is divisible by X, except for zeroing out of off-diagonal entries in the first column. Thus the congruences of the hypothesis are maintained in the modified matrix Q'. Now repeat this procedure using Q'_{22} as a pivot to zero out the second column, and proceed similarly down the diagonal. All these entries are non-zero and thus suitable to use as a pivot. Diagonal elements that have been used as a pivot never change again since their column has otherwise been zeroed out. Finally we have a diagonal VMR called R, each of whose diagonal entries is $\equiv 1 \pmod{X}$. By Proposition 0, their product $\theta(X)$, which is $\equiv 1 \pmod{X}$ will serve as an annihilator for M.

2. Diagonalization algorithms for a P.I.D.

In this Section we review diagonalization algorithms for an f.p. module over a principal ideal domain (P.I.D.). In Section 4 we establish that any principal ring is a Hermite ring, and thus the results of this Section will also apply to any P.I.R. Recall that a *Hermite ring* is a commutative ring $\mathcal{R}$ such that given a two-vector $[a \quad b]$ with entries in $\mathcal{R}$, there is $c \in \mathcal{R}$ and B, an invertible two-by-two matrix, such that

$$[c \quad 0] = [a \quad b] \cdot B.$$

[1] See the book of [18] for historical remarks on this result.

An Hermite ring is easily seen to be a Bézoutian ring (or simply *Bézout ring*), that is, in $\mathcal{R}$ the sum of two principal ideals is principal. A characterization of Hermite rings that we will use can be found as Proposition T [7]. A Bézoutian *domain* is Hermite, as proved in [14], so a P.I.D. is Hermite. From this information, it follows *by algorithmic construction* that a P.I.D. $\mathcal{R}$ is also an Invariant-Factor ring, and hence that we may diagonalize any square matrix over $\mathcal{R}$ by left and right invertible matrices (change of generators and relations). If the diagonalization satisfies the division condition, we have a Smith form for the matrix. By the result of [17], we can always obtain a Smith form. The algorithm is exposed independently here as the results do not seem to have reached the engineering community [13, pg. 391]. Some applications of the Smith form and related canonical forms to numerical analysis, in the case of a real or complex polynomial ring, are given in [21].

We present two diagonalization algorithms which will serve different purposes. The idea of the first algorithm appeared in [17] and is an improvement on algorithms from the engineering literature that may have been inspired by the treatment in [23, pg. 5], for example. Most of these proofs are actually given for the case of a Euclidean domain, and may begin by "take the smallest entry in absolute value". No absolute value or Euclidean algorithm is generally available in the P.I.D. case however. The second algorithm is useful for proving "Lagrange's Theorem".

Recall from sources such as [1, pg. 288], that over a P.I.D., left multiplication by an invertible matrix is effected through generalized row and column operations as follows:

(1) interchange two rows;
(2) add a multiple of one row to another row, giving a new row to replace the second-named row;
(3) given two rows, form two linear combinations of the rows. The first linear combination replaces the first of the two rows, the second linear combination replaces the second of the rows, and this is done in an invertible manner, so that the leading entry of the first row is the gcd of the two leading entries of the original rows, and the lead entry of the other row becomes 0.
(4) multiply a row by a unit of the ring, giving a row to replace the old row.

and similar column operations for right multiplication by an invertible matrix. Operations of type (2) are included for generality. In fact the matrix reductions we are performing can be done using only the other three operations.

Algorithm I. *Let there be given a square matrix B with entries in a P.I.D. called $\mathcal{R}$. Then one may form an equivalent diagonal matrix $D = E \cdot B \cdot F$, where E, F are invertible, through a sequence of row and column operations on B.*

Method. For two non-zero entries in the first column of B, perform row operations upon their respective rows as an operation of type (3), resulting in the gcd of the two entries and an additional zero in the column. Proceeding in this manner results finally in but one non-zero entry. (There should be at least one since we are tacitly supposing that our matrix A presents a *torsion* module.)

This solitary entry may be called x_1, and it may also happen to be the gcd of the vector of its own row (after those row operations were effected). If so, we may zero out that row in the well-known manner, by column operations, obtaining x_1 as sole non-zero entry in both its row and its column. The entry may now be placed by row and column transpositions to the $(1,1)$ position. We then proceed in the

same manner on the remaining $(n-1) \times (n-1)$ sub-matrix, until the whole matrix B has been transformed to diagonal D by allowable operations, or in other words by a "congruence" $D = E \cdot B \cdot F$, where E, F are invertible.

In case x_1 did not divide all entries of its row, we perform (generalized) column operations so that a solitary y_1 appears in this row. We must have that $y_1 | x_1$ but x_1 does not divide y_1. If y_1 divides all entries in its (new) column, we may move it so that it becomes a corner entry, solitary in both its row and column. If not, we proceed as before, creating a sequence of strict divisions:

$$\cdots \mid y_k \mid x_k \mid y_{k-1} \mid x_{k-1} \mid \cdots \mid y_1 \mid x_1.$$

But this is impossible since the union of the principal ideals $\cdots \subset (x_i) \subset (y_i) \subset (x_{i+1}) \subset \ldots$ has a generator z which must lie in some (x_j) (Noetherian property).

This method looks similar to the proofs of "decomposition over a P.I.D." in the standard texts, but we note that we have not yet obtained the "division condition" (1) for the diagonal entries. In [13] it is proposed, once a solitary corner entry is found, to use it to reduce (assuming the Euclidean algorithm is available) the rest of the matrix into remainders, which normally destroys the partly diagonal form just painstakingly built up! Thus the corner element is no longer solitary in its row and column and we have to start again.

The point is that it is better first to diagonalize the matrix, and subsequently to change the diagonal entries to effect the division property, obtaining the actual invariant factors of the Smith form. The evident technique of how to do this was apparently first written down in the paper of [17]. This is linked to the new observation that in a Bézoutian ring, the intersection of two principal ideals is principal, so a lcm of two elements is well-defined, as an ideal. This result is a sentence in a first-order theory (Bézout rings) and should really have a first-order proof. Thus there should be a proof involving only variables and constants on ring elements and no modules, matrices, decomposition or equivalences. Such a simple first-order proof is presented in Section 3.

Thus there remains to put a diagonal matrix into a Smith form where the entries satisfy the division property (1), through allowable row and column operations. Thus if we have the presentation matrix now in the form

$$D = \begin{bmatrix} \alpha_{11} & 0 & & 0 \\ 0 & \alpha_{22} & & \\ & & \ddots & \\ 0 & 0 & & \alpha_{kk} \end{bmatrix},$$

we say that D expresses a *diagonal presentation* of G. We formally state how two given non-zero elements can be made to satisfy "division".

Lemma A. *Given a diagonal presentation D for G as above, for fixed indices $1 \leq i < j \leq k$ let*

$$g(i,j) = \gcd(\alpha_{ii}, \alpha_{jj})$$
$$l(i,j) = \mathrm{lcm}(\alpha_{ii}, \alpha_{jj}) = \alpha_{ii} \cdot \alpha_{jj} / g(i,j).$$

Then D' which results from D when the i-th diagonal entry is replaced by $g(i,j)$ and the j-th diagonal entry is replaced by $l(i,j)$ is also a diagonal presentation for G.

Proof. To modify D we perform row and column operations only on rows labeled i, j and columns labeled i, j. These operations are induced from operations performed on the 2×2 matrix

$$D_{ab} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix},$$

where we simplified the notation. Using the integers f, g given in the formula

$$fa + gb = g(i,j) = \gcd(a,b) \quad ,$$

we may transform D_{ab} into

$$D'_{ab} = \begin{bmatrix} a & g \\ 0 & b \end{bmatrix}.$$

The operations actually performed on the original D are of course, to add f times the i-th column to the j-th column, followed by adding g times the j-th row to the i-th row. Call the resulting matrix D' Since $g \mid a$ and $g \mid b$, we may "zero out" the $(1,1)$ and $(2,2)$ entries to obtain

$$\begin{bmatrix} 0 & g \\ -ab' & 0 \end{bmatrix}$$

where $b' \cdot g = b$. Permuting the columns (generators) yields entries up to multiplicative unit, arrayed as

$$\begin{bmatrix} g & 0 \\ 0 & h \end{bmatrix},$$

where $h = \operatorname{lcm}(a,b)$. Again we mimic these operations and perform the corresponding operation on D' obtaining D'', where now $\alpha''_{ii} \mid \alpha''_{jj}$.

Repeatedly applying this lemma to the (updated) matrix D will finally put it into Smith form. In fact, apply Lemma A successively to all pairs of diagonal indices, ordered lexicographically. That is, we apply Lemma A to $(i,j) = (1,2), (1,3), \ldots, (1,n), \ldots, (n-1,n)$. At the end of the process, the division property (1) is easily seen to be satisfied.

Proposition 3. *Use of Algorithm I with repeated application of Lemma A will put a square (presentation) matrix over a principal ideal domain, into Smith normal form and thus exhibit its invariant factors, which as always are defined up to multiplication by a unit of the ring.*

Software with instructions on using this algorithm is provided at the site, [6]. All routines apply to the Euclidean ring of rational integers $\mathbb{Z}$. The programs written to put into effect the above algorithm, and the one preferred in [13] differ only by a few lines. A file with parameter values is provided to allow one easily to construct a selection of initial presentations that have interesting invariants, that is, certainly more than one invariant factor not equal to 1.

Using the matrix routines, we compare the computational cost of the two methods. To this end, square matrices of size 5 to 9 were selected "at random" with

Size	Flops (K)	SD %	Flops (LCM)	SD %	Ratio %
5	1948.4	22.99	583.99	29.72	29.97
6	3468.2	23.75	758.89	36.44	21.88
7	5725.9	25.65	1530.9	27.28	26.74
8	8863.7	25.69	1552.8	42.33	17.52
9	13132.0	26.54	3162.0	30.62	24.08

TABLE I. Floating Point Operations for Kailath and LCM Diagonalization Algorithms

a high probability of having multiple invariant factors. Five hundred trials with different matrices were performed for each size. The table above then lists the mean number of floating point operations ("flops") required for the conventional method of diagonalizing to Smith form, followed by the standard deviation of the data, in percent. The average number of flops and the SD then follow for the new method that forms the invariant factors from a diagonal matrix by means of repeated applications of gcd and lcm. The ratio in percent, of the average computation by the latter to computation by the former method is given. We observe a substantial speed-up and simplification, resulting from not having to perform any modulo reductions with a corner element. In the case of 9×9 matrices, the LCM method shows nearly a four-fold speed-up over the Kailath algorithm.

As a guide to the software, the routine "FormNew" collects statistics using the conventional method with parameters, "SetUp", and the new method, "smith", itself a small script for the main routine "DirSum". The random generation depends on some *ad hoc* parameters as well as a generalized routine "MatFact", which produces a random diagonal matrix G, followed by "RandT", that produces a row and column equivalent matrix $K = KL * G * KR$. Both methods ("Director" and "DirSum") rely on finding an entry, after row and column operations, that divides both its row and its column (routines "corner" and "corner2"). The two "corner" routines in turn rely on producing a solitary entry in a row or column through use of the Euclidean algorithm (routine "euclid").

Let now B present an f.g. torsion module M over a P.I.D. $\mathcal{R}$, so that we may take B to be square of size n.

Algorithm II. *There is a sequence of upper triangular matrices B_i, and lower triangular matrices C_i, so that $B_1 = A_1 \cdot B$, $C_i = B_i \cdot E_i$, $B_{i+1} = A_{i+1} \cdot C_i$, where A_i, E_i are all invertible of the same size as B, and C_K is diagonal for some index $K > 1$.*

Proof of Algorithm II. We can start out by making B_1 upper triangular as indicated, as $\mathcal{R}$ is an Hermite ring. Now in B_1 we ask where each diagonal entry β_{ii} divides its entire row. If so, B_1 is easily diagonalized by a right E_1, and we are finished. Otherwise, there is a smallest $i \geq 1$ with $j > i$ and $d_1 = \gcd(\beta_{ii}, \beta_{ij})$ strictly divides β_{ii}. Following from the proof of Proposition 3.25 of [14], we get B_1 into lower triangular form through right multiplication by an invertible matrix, such that the new C_1 is "diagonal" for row index $k < i$, and in fact where the diagonal entry $\gamma_i = \gcd(\beta_{ii}, \beta_{i\,i+1}, \ldots, \beta_{i\,n})$ strictly divides β_{ii}.

We arrived at lower triangular C_1 in accordance with the Algorithm. We can now apply the same argument just used for B_1 on C_1, after transposing everything. There is a column such that the gcd of the column strictly divides its diagonal entry, and after the re-triangularization is done, that gcd will sit in the diagonal position. Since the matrix is of finite size, continuing this process will lead to a sequence of strict divisions of some of these diagonal entries,

$$\cdots, e_{h+1} \,|\, e_h \,|\, \cdots \,|\, e_2 \,|\, e_1,$$

which since $\mathcal{R}$ is principal, hence Noetherian, is impossible. Hence after a finite number of steps we would have reached a diagonal matrix, which justifies the method of Algorithm II.

We reiterate that in fact Algorithms I and II, and Proposition 3, also apply to matrices over any principal ring, possibly with zero-divisors. The "domain" condition was used only to establish that the ring $\mathcal{R}$ satisfies the Hermite property. In Section 4, Theorem 1 establishes that this property is true in general for a P.I.R.

3. The Principal Lattice of a Bézoutian ring

The theory and techniques of Lattice Theory have proven to be well suited for the study of commutative rings, especially for Bézoutian integral domains. Facts of lattice theory needed for our discussion can be found in Appendix 1 of [5], or [2]. The lattice that is usually defined for an entire Bézoutian ring is in fact a "lattice-ordered group", a lattice endowed with a compatible structure of an abelian group. This implies immediately, as in [8] that the lattice is a distributive lattice.

For any Bézoutian ring we have the structure of a "semi-lattice" on its set of principal ideals. That is, we can form the "meet" of two ideals $(a), (b)$, through summing the ideals, and representing the new ideal $(\gcd(a, b))$ by a generator (c). There is, however, no reason why this construction should not be enhanced to a lattice structure: we would need to define the "join" operation, which will of course be an lcm of the generators a and b.

The resulting lattice will not be an "ordered group", but will be compatible with the multiplication of ideals, which is a commutative monoid. Thus $(a) \leq (c)$ if $(c) \subset (a)$. The ideal $\mathcal{R} = (1)$ is a universal lower bound, and (0) is a universal upper bound. In the case of a Bézoutian domain, the "lattice-ordered monoid" that we are defining is the same as the "positive cone" of the "lattice of principal fractional ideals in the field of fractions" prevalent in the literature.

For this to work, we need to show that the "join", $A \vee B$ is always defined. We generalize slightly a result appearing in [17].

Proposition 4. *Let $\mathcal{T}$ be any commutative ring. Suppose that $a, b \in \mathcal{T}$ are elements such that the ideal they generate is principal, $(a, b) = (\Delta)$, then the intersection of the ideals they each generate is also principal, that is*

$$(a) \cap (b) = (\Gamma)$$

for some $\Gamma \in \mathcal{T}$.

Before we give the proof of Proposition 4, let us examine its statement in terms of first-order logic. P.M. Cohn has observed that the theory of Bézoutian rings is

a first-order theory, and this applies to the theory of commutative rings in general. To define a principal ring, however, requires talking about ideals in general, subsets, predicates or some equivalent second-order concept. So a statement of the Proposition, in a first-order theory where the axioms of a commutative ring do hold, ought to be

$$\forall a, b, f, g, \Delta, k, \ell, \{\Delta = ka + \ell b \,\&\, a = f\Delta \,\&\, b = g\Delta \Rightarrow \\ \exists r, s \,\{ra = sb \,\&\, \forall \alpha, \beta(\alpha \cdot a = \beta \cdot b \supset \exists F(F \cdot ra = \alpha a))\}\}.$$

Remark 1. Computers are rather good at determining whether such a formal statement is true, or at least determining whether a given putative proof is valid. Automated tools do not generally try to determine, however, whether the formal sentence actually expresses the mathematical idea that is desired. Human mathematicians do not fare much better at this. Only through experience and testing out proofs of various related theorems can one approach a level of certainty that the formalization was actually done correctly. For example, several plausible definitions can be given for the concept "connected graph". One of these arises from the definition based on "union of disjoint open sets" in topology, another from the concept of "path-connectedness", and a third that is unique to graphs, recursively defined from the edge and vertex degrees. It has been verified through an automated proof-check, [4], that all three definitions are equivalent. This is evidence that one has arrived at the "correct" definition of graph connectedness.

Proof. We are given for $a, b \in \mathcal{T}$ a "greatest common divisor" $\Delta = ka + lb$, $a = f\Delta$, $b = g\Delta$, which follow from the three "containment" relations among the ideals (a), (b), (a, b), and (Δ). The element $\Gamma = fg\Delta$ is in the intersection of (a) and (b). Now given $c = \alpha a = \beta b$, let $F = \alpha \cdot \ell + \beta \cdot k$, so

$$F \cdot \Gamma = g(\beta k \Delta f + \alpha \ell \Delta f) = g\beta(k\Delta f + \ell \Delta g) = g\beta(ka + \ell b) = g\Delta\beta = \beta b = c.$$

Thus Γ divides any such c in the intersection, and is a generator of the intersection ideal.

One could suppose that this result would be quite suitable for an automatic proof by some dedicated first-order proving engine. The system "Otter", developed at the Argonne National Laboratories, treats first-order algebraic theories, [22]. A proof was attempted for the Proposition in case $\mathcal{T}$ is an integral domain, where the result is well-known and certainly easier. Even in this case it was indicated [20], that the substitution required (element F in the proof) would never be found by Otter. This is disappointing since that actual substitution requires only four symbols, and is symmetrical.

Essentially, one may ask how the elements Γ, F whose existence is required by Proposition 4 are to be found. One approach that could be tried by a more "intelligent" prover might be at first further to restrict the theory. This is exactly what the Argonne experts did when checking the Proposition on a domain instead of a general commutative ring. Such a system might argue as follows: The rings we know about that produce sums and intersections of ideals are Euclidean domains, such as $\mathbb{Z}$. Given $c = \alpha a = \beta b$, using the Euclidean algorithm we find $\Delta = ka + \ell b$, which leads to $\alpha fk + \alpha g\ell = \alpha$ which shows that g divides α (since g divides αf), so in fact $c = \alpha a = \alpha' g a = \alpha' f g \Delta = \alpha' \Gamma$, so we have

found the $\mathrm{lcm}(a,b)$ "constructively". Then α' can also be found (in a domain). If Proposition 2 indeed holds in a general commutative ring, surely the same formulas such as $F = \alpha' = \alpha\ell + \beta k$ must apply in the general (non-domain) case. Thus the "necessary substitutions" have been provided by another theory where the result is more obvious.

We have shown that gcd and lcm for a general Bézoutian ring are defined (non-uniquely), so we have a "semigroup-ordered lattice" of ring elements in a certain equivalence relation. Here the elements $\gcd(a,b)$ may not be uniquely defined, but they do uniquely define a principal ideal.

Definition 1. *A distributive lattice is one where either of the equivalent universal identities hold:*

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$$
$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z).$$

A good discussion of the distributive property, in addition to the lattice theory sources previously mentioned, can be found in the book [9].

We sometimes write $[x,(y,z)] = ([x,y],[x,z])$ for the first formula, where we can take (y,z) as the ideal generated by the two elements, or the gcd, and similarly $[y,z]$ as the intersection of principal ideals or lcm.

Similarly we define a Bézoutian ring to be *distributive* if its lattice of principal ideals is a distributive lattice. We use the distributivity property of the lattice $\mathcal{L}(\mathcal{T})$ to prove Lagrange's theorem on module capacity (similar to *order* for a finite abelian group), where $\mathcal{T}$ is a P.I.R.

A P.I.D. is distributive as it is an entire Bézoutian ring. We show that a Bézoutian domain is distributive. This follows "classically" in at least two ways. For one thing, we have the principal ideal lattice embedded in the group-ordered lattice of principal fractional ideals, and a group-ordered lattice is always distributive, see [8]. Alternatively, it is known that the universal formula $x \wedge y = x \wedge z \ \ \& \ \ x \vee y = x \vee z \Rightarrow y = z$ is an equivalent condition for distributivity. For a Bézoutian domain this follows, since essentially we have by hypothesis

$$[x,y] = \frac{xy}{(x,y)} = \frac{xz}{(x,z)} = [x,z],$$

which implies $xy = xz$, or $y = z$, all up to a multiplicative unit. This latter proof appears to have "first-order" quality but it is not so satisfactory. The characterization of distributive lattice this way (if meet and join with a third element are equal, so are the elements themselves) is not obvious, see [2]. We give a third proof directly from the definitions. Remarkably, the result does not exclude zero-divisors from the ring.

Proposition 5. *Any Bézoutian ring is distributive.*

Proof. Given principal ideals of $\mathcal{R}$, x, y, z, we may write for the gcd, $x \wedge y = x \cdot y = xy$ and $y \vee z = y + z$ as is common lattice algebra notation. The distributive law is then,

$$x(y+z) = xy + xz. \tag{2}$$

The statement $s \leq t$ is the same as $s+t = t$. From elementary properties of a partial ordering we have universally $xy \leq y \leq (y+z)$, where one can "multiply through by x" obtaining $xy = x \cdot xy \leq x \cdot (y+z)$, and by symmetry $xz \leq x \cdot (y+z)$, so the join on the two inequalities satisfies $xy + xz \leq x \cdot (y+z)$. In ring notation, if we have an element u in $\gcd(a, \mathrm{lcm}(b,c))$, seen as a principal ideal, then $u = \alpha a + \beta k$ where $k = \gamma b = \delta c$. Thus clearly u is both in $\gcd(a,b)$ and $\gcd(a,c)$ so is in their join (intersection). Thus set-theoretically if $x = (a)$ and so forth, $x(y+z) \subset xy + xz$, but this means the same as $xy + xz \leq x(y+z)$.

Hence it is only necessary to show that

$$x \cdot (y+z) \leq xy + xz. \tag{3}$$

This is equivalent to another characterization of distributivity, namely

$$\forall x, y, z \quad (x \vee y) \wedge z \leq x \vee (y \wedge z). \tag{D}$$

That this follows from (3) is easy. It is seen to imply (3), [9, pg. 36], by taking $x = a, y = b, z = a \vee c$ and one obtains (universally) $(a \vee b) \wedge (a \vee c) \leq a \vee (b \wedge c)$. Note also that (D) as a universal formula is self-dual: taking the dual results in the same formula merely with x and z interchanged. Now setting $a = x$, $b = y$, $c = z$ and taking the dual (interchanging $\wedge$ and $\vee$, $\leq$ and $\geq$) gives precisely (3). Thus it is enough to prove (D) for a Bézoutian ring. We revert to ring operations of addition and multiplication and denote by x, y and so forth, ring elements corresponding to the lattice elements (associativity classes) x, y. Note that to prove (3) directly is tantamount to being able, given $u = \alpha x + \beta y = \kappa x + \lambda z$, to re-write $u = \alpha' x + w$, where $w = \beta' y = \lambda' z$. How to do this directly is not so obvious, but will be carried out subsequent to the present proof.

To continue with the proof of universal inequality (D), we must show that the ideal given on the *right* hand side is *contained* in the ideal of the left hand side. Thus given $f = \gamma x \in \mathcal{R}$ which also is of the form $\alpha y + \beta z$, we have to show how to write $f = \gamma' q + \delta z$, where $q = \pi x = \sigma y$ for suitable coefficients from the ring $\mathcal{R}$. If in fact we could take $q = \pi' f$ and $\delta z = \delta' \beta z$ and $\sigma y = \sigma' \alpha y$, we also would be done *a fortiori*. Thus we may take $f = y + z$, and set $\gamma = (y, z)$ any generator of the sum of the principal ideals (y) and (z). Thus for suitable ring elements, $y = G\gamma$, $z = H\gamma$, so

$$f = G\gamma + H\gamma = r\gamma.$$

Letting $Ly + Mz = \gamma$, we have

$$r \cdot Ly + r \cdot Mz = r\gamma = f. \tag{4}$$

But $rLy = rL(G\gamma) = GL \cdot r = GLf$, and similarly $r \cdot Mz = rMH\gamma = HMf$, so by (4)

$$f = rLy + rMz$$

where each term is actually in the ideal (f). This is a stronger conclusion than needed to establish (D).

We promised to give a procedure to show that the equivalent formulation (3) of "distributivity" holds, for any Bézoutian ring. Again, this formula is perhaps not so obvious even over the ring $\mathbb{Z}$. First we restate what we just did.

Lemma X. *Over a Bézoutian ring, if we have $x = y + z$ we may write*

$$x = (G + H)LG\gamma + (G + H)MH\gamma, \tag{5}$$

where the first term is in $(x) \cap (y)$ and the second term is in $(x) \cap (z)$. Here, $\gamma = Ly + Mz$, $G\gamma = y$, $H\gamma = z$.

Now in formulation (3) we start with an element

$$w = \begin{cases} \alpha a + \beta b \\ \gamma a + \delta c \end{cases}.$$

In particular we have $w = u = \alpha a + \beta b$, so applying Lemma X yields $u = Bu + \theta\alpha a$ where $Bu = \eta\beta b$. Using the definition of w we also obtain $\eta\beta b = Bu = B\gamma a + B\delta c$. Applying Lemma X to the element $\eta\beta b$ in turn yields (for some new coefficients)

$$\eta\beta b = TB\delta c + VB\gamma a,$$

where also $TB\delta c = R\eta\beta b$. Putting two formulas together gives

$$w = u = VB\gamma a + \theta\alpha a + TB\delta c,$$

or $w = Ka + Hc = Ka + Jb$ which was to be proved. Thus we have an algorithm through Lemma X to validate either formulation (3) or (D) of the distributive property for the principal ideal lattice of any Bézoutian ring, proving Proposition 5.

4. The Structure of a Principal Ideal ring

If we wish to consider general commutative rings, with zero-divisors, an appropriate class of Invariant-Factor rings is the class of Principal Ideal rings. A strong characterization of such rings, based on a decomposition into well-understood rings is given by Theorem 12.3 in [14] and has a modern exposition in [3]. We give here a similar characterization that is sufficient for our purposes and consistent with our elementary and constructive approach.

A *sub-direct sum* decomposition, [19, pg. 52], is given by projections π_j of a commutative ring $\mathcal{R}$ onto rings $\mathcal{S}_j$, such that the induced homomorphism ϕ of $\mathcal{R}$ to the direct sum ring of the $\mathcal{S}_j$ is injective. Thus $\mathcal{R}$ is embedded into the direct sum, in such a way that the image in each summand is surjective. Recall that an element $x \in \mathcal{R}$ is *nilpotent* if for some integer power $r \geq 1$, we have $x^r = 0$.

Lemma Z. *If $\mathcal{R}$ is a (commutative) principal ideal ring having a zero-divisor z, $z \cdot a = 0$, which is* **not** *nilpotent, then $\mathcal{R}$ is a sub-direct sum of two summands. Each projection has a non-trivial kernel.*

Proof. Let a, z be as in the hypothesis. Let $Ann(z^n)$ be the set of $b \in \mathcal{R}$ with $bz^n = 0$. This is an ideal and furthermore $Ann(z^n) \subset Ann(z^{n+1})$; their union is an ideal which must be generated by an element y which is in some $Ann(z^k)$. Hence all subsequent $Ann(z^j)$ with $j > k$ are the same as this ideal (y). If $\kappa a = f \cdot z^k$, then f is seen to be an annihilator of z^{k+1}, hence and annihilator of z^k, which means that $\kappa a = 0$. Thus

$$(z^k) \cap (a) = (0).$$

The theorem of E. Lasker asserts that for a Noetherian ring, ideals that are irreducible according to intersection of ideals, and ideals that are "primary", coincide. The above argument is key to this theorem, as given for example in [1, pg.83]. Now Theorem 3.9 of [19] amounts to the easy fact that the projections

$$\pi_1 : \mathcal{R} \to \frac{\mathcal{R}}{(z^k)}, \qquad \pi_2 : \mathcal{R} \to \frac{\mathcal{R}}{(a)}$$

express $\mathcal{R}$ in the desired form.

Proposition 6. *A P.I.R. can be written as a finite sub-direct sum where each summand is a P.I.R., all of whose zero-divisors are nilpotent (Property NP).*

Proof. If there are no such zero-divisors, we may take a single summand (degenerate case of sub-direct sum). Otherwise Lemma Z tells us that there is a direct sum of two rings $\mathcal{R}_1^1 \oplus \mathcal{R}_0^1$ where we are on "level" 1. If both these summands have the NP property, we are finished. If not, we may continue a dyadic splitting into two summands for each summand. On a level d, a summand (*node* of the dyadic tree) is *live* if it possesses zero divisors not nilpotent, and is *dead* if it satisfies NP. We move to a higher level if and only if there is a live node at level d. Each new node represents a P.I.R. since the property "principal ring" is preserved under projection to a non-trivial ring. A live node is split into two summands as before, while a dead node $\mathcal{R}_k^d$, $0 \leq k \leq 2^d - 1$ is made into a sub-direct sum of two identical rings, with both projections π_1, π_2 the identity. This is a degenerate case of splitting, but useful for bookkeeping purposes.

If there is no bound on the level number where the construction continues, there exists an infinite chain of live nodes. Indeed, the node $\mathcal{R} = \mathcal{R}_0^0$ is live, and is the first link of our chain. If for an arbitrary d there is always a live node $\mathcal{R}_k^d$, we note that either (or both) $\mathcal{R}_1^1$ or $\mathcal{R}_0^1$ is live. In fact, the sub-tree generated by one of them must have live nodes of unbounded level. If that node is $\mathcal{R}_0^1$, the subtree consists of those nodes whose lower index has most significant bit $= 0$ in binary. Continuing to level d, by induction our last link is a live note $\mathcal{R}_k^d$ and one of its descendants generates a sub-tree with live nodes of unbounded level. Choose that descendant node and we have constructed a chain of live nodes, each node $\mathfrak{a}^d$ projecting onto its successor $\mathfrak{a}^{d+1}$ with a non-zero kernel.

But such a chain is impossible for a P.I.R. since it leads to a strictly ascending sequence (chain) of ideals under inclusion. The first ideal is the kernel of $\pi : \mathcal{R} \to \mathfrak{a}^1$ and the second ideal is the pullback along π of the kernel of $\pi^1 : \mathfrak{a}^1 \to \mathfrak{a}^2$. Since at every link of the chain of projections, the kernel is non-zero, we obtain a strictly increasing sequence of ideals of $\mathcal{R}$. Since the union of the sequence is principal and has a generator, the sequence must stabilize when this generator enters, after a finite number of steps (Noetherian property). This contradiction shows that there must be a bound D on the levels where live nodes can exist. If necessary for efficiency, one can do some "clean-up" to remove those final nodes that do not have a live parent. There is a projection from $\mathcal{R}$ to each of these final (dead) nodes at a level D or less, given by composition of the projections along a chain of live nodes reaching from $\mathcal{R}$ to the final node (say) $\mathcal{R}_k^D$. All conclusions of Proposition 6 are now fulfilled.

Next we characterize those P.I.R.'s with the property NP.

Proposition 7. *Suppose $\mathcal{T}$ is a principal ring where any zero-divisor is nilpotent. Then either there are no zero-divisors ($\mathcal{T}$ is then a P.I.D.) or else let the ideal S of nilpotent elements be of the form $S = (s)$, with r the least power $1 < r$ so that $s^r = 0$. Then*

i) *any element not in S is a unit of $\mathcal{T}$,*
ii) *any $t \neq 0$ satisfies $t = u \cdot s^k$ for some $0 \leq k < r$, where u is a unit.*

Proof. Consider first an element $t \mid s$. Taking $a \cdot t = s$, $a \neq 0$, we have $at \cdot s^{r-1} = 0$, so t is a zero-divisor, hence nilpotent, unless $as^{r-1} = 0$. In the latter case a must also be nilpotent so $a = c \cdot s$ and we get

$$(ct - 1) \cdot s = 0$$

which shows that $(ct - 1)$ is nilpotent, and hence by expansion of the power, ct is a unit, and t is a unit. Thus any divisor of s is either nilpotent or a unit.

Now suppose we have an element $a \notin S$. We form $d = \gcd(a, s)$ which cannot be nilpotent since it has a as a multiple. But $d \mid s$ so by the above d must be a unit, so we could actually find coefficients such that

$$\alpha a + \beta s = 1.$$

Since $1 - \beta s$ has an explicit inverse by telescoping expansion, αa and a are units. This verifies conclusion i). But now if $t \neq 0$, among the finite set (interval) $[1, \ldots, r-1]$ there is a maximal k so that $t = q \cdot s^k$ for some $c \in \mathcal{R}$, which may not be in S by the maximality, hence by i) is a unit. This settles conclusion ii) of the Proposition.

Note that the classical result is a decomposition of any P.I.R into a finite *direct* sum of domains and certain "valuation rings" which are the same as NP rings, [14, pg. 486]. This stronger characterization of a P.I.R. is used to show that such a ring is Hermite, hence an IF ring. But our semi-direct sum characterization is also sufficient to prove this result.

This is a convenient juncture to state a result, which when combined with Theorem 3 of [7] shows that a Bézoutian domain is Hermite, and also implies Theorem 3.2 of [14].

Lemma T. *Suppose that $\mathcal{R}$ is a commutative ring where every zero-divisor is nilpotent. Then if for a, $b \in \mathcal{R}$ $\gcd(a, b) = d$, elements a_1, b_1 that satisfy*

$$a = a_1 d, \quad b = b_1 d, \quad d = fa + gb$$

satisfy $\gcd(a_1, b_1) = 1$.

Proof. We obtain immediately $d = (fa_1 + gb_1) \cdot d$, whence by the hypothesis we must have that $c = 1 - fa_1 - gb_1$ is nilpotent. Expanding $c^n = (1 - fa_1 - gb_1)^n = 0$ yields $(fa_1 + gb_1) \cdot w = 1$, for some $w \in \mathcal{R}$, so we have $\gcd(a_1, b_1) = 1$ (defined up to a unit).

Finally we may state the main result of this Section.

Theorem 1. *A P.I.R. $\mathcal{T}$ is an Invariant-Factor ring (IFR).*

Proof. Assuming that the ring $\mathcal{T}$ is Hermite, and that any ascending chain of ideals must become stable, we can apply Algorithm I to obtain an invariant factor decomposition, or Smith form, of any f.g. torsion module. Alternatively, Theorem 5.4 of [14] implies the same thing. For the Hermite property, it is sufficient to show that property T of [7] holds. But by Proposition 6, $\mathcal{T}$ may be written as a subdirect sum, so we may work with the summand coordinates of any element. If the summands obey condition T, so does the sub-direct sum. But each summand is a ring that satisfies property NP, and hence Lemma T applies. In fact this says that the summand must have property T, and hence is Hermitian. As we noted, then $\mathcal{R}$ will have property T, so is Hermitian, so also is an IFR by Lemma T.

For the sequel we note two general facts about ideals in a commutative ring.

Fact 1. *Let M be a cyclic module over $\mathcal{R}$, and* $\mathfrak{i}$ *its annihilating ideal so that*

$$M \simeq \frac{\mathcal{R}}{\mathfrak{i}}.$$

Then for $c \in \mathcal{R}$, $cM \simeq \dfrac{\mathcal{R}}{\mathfrak{j}}$, where

$$\mathfrak{j}(c) = \{a \in \mathcal{R} \mid a \cdot c \in \mathfrak{i}\}.$$

Proof. If x is a generator for M, let $\mathfrak{j} = \mathrm{Ann}(cx)$ and suppose $\kappa \in \mathfrak{j}$, so then $\kappa cx = 0$. Thus $\kappa \cdot c \in \mathfrak{i}$. Conversely, if $\kappa c \in \mathfrak{i}$ then $\kappa c \cdot x = 0$ so $\kappa \cdot cx = 0$ and $\kappa \in \mathfrak{j}$.

Fact 2. *With the above notation if* $\mathfrak{i}_2 \subset \mathfrak{i}_1$ *and*

$$M_i = \frac{\mathcal{R}}{\mathfrak{i}_i}, \quad i = 1, 2$$

then for any $c \in \mathcal{R}$, $\mathfrak{j}_2(c) \subset \mathfrak{j}_1(c)$.

Proof. Let $y \in \mathfrak{j}_2(c)$, then $yc \in \mathfrak{i}_2 \subset \mathfrak{i}_1$, hence by Fact 1 $y \in \mathfrak{j}_1(c)$ as was to be proved.

5. Module capacity and Lagrange's theorem

If a square matrix Θ is put into triangular form, and there are no zero elements on the diagonal, we have a presentation of a torsion module. In constructing invariants from the diagonal entries, the natural sort of commutative ring to work over would be a Hermite ring, where such a triangular reduction is always possible.

Definition 2. *Given a Hermite ring $\mathcal{R}$ and a torsion module M over N presented by a square matrix Θ, we define the capacity ideal (or "capacity") of the presented module to be the product of the principal ideals given by the diagonal elements for some equivalent triangular form of Θ. The capacity (generator) is written χ_Θ.*

Remark 2. Although the definition does not depend on which equivalent triangular form is used, it may well depend on the original presentation, not only on the module itself. In fact at least for the Hermitian case, the capacity ideal is indeed an invariant of the module's isomorphism class ("module type"). We discuss various

proofs of this result further on, and it also follows (allowing free use of determinantal ideals) from the theory of Fitting ideals [15].

If $\mathcal{R}$ is the ring of rational integers $\mathbb{Z}$, then the capacity of a torsion module (finite abelian group) is the (ideal generated by) the order of the group. Here it is clear from the invariant meaning of "order" that this ideal is uniquely determined by the group. Also in the case of the Euclidean ring $\mathbb{F}[X]$ acting on $\mathbb{F}^n$, the capacity equals the characteristic polynomial (defined up to multiplicative unit), and elementary linear algebra shows that it is an invariant.

An embedded abelian group has order dividing the larger group, and the quotient group has order which is the quotient of these two orders. In the linear algebra case, a "submodule" is an invariant subspace of $\mathbb{F}^n$, which will yield a new characteristic polynomial dividing the original one.

Next we state a possible generalization of "Lagrange's theorem" to the general case of an Invariant-Factor ring. If the ring is a Euclidean domain, the result is quoted in [15]. We are able to formulate the Lagrange theorem precisely in the P.I.R. case, and even prove it. This is a significant generalization of the Euclidean domain case, and the proof does not involve primary decomposition.

Conjecture 1. *Over an IF ring $\mathcal{R}$, let $N \subset M$ be an embedding of f.p. torsion modules, and $R = M/N$ the quotient module. Given presentation matrices Θ_N and Θ_M, there is a presentation matrix Θ_R such that*

$$\chi(\Theta_M) = \chi(\Theta_R) \cdot \chi(\Theta_N)$$

considered as an equality of principal ideals.

We prove the conjecture in the case where $\mathcal{T}$ is a P.I.R. This Lagrange theorem in the P.I.R. case is used to show that the ideal generated by $\chi(\Theta_M)$, does not depend on the particular presentation of M, and so is an invariant of module type. Another proof of this invariance does not depend on the P.I.R. property, or even the Hermite property, but uses the extended ring $\mathcal{R}[X]$ and Nakayama's lemma (Proposition 2). This result at least leads to being able to define $\chi(M)$ as an invariant of module type for a ring where it is defined on f.p. modules. Once it is established that capacity is a module invariant for a torsion module over a P.I.R., it is straightforward to show that the whole Smith form is also such an invariant.

The following lemma appears to require the existence of the general determinant and its most elementary properties. It would be of interest to find a proof that relies only on the more limited concept of capacity (for triangular matrices) just defined.

Lemma B. *Let Q be an $n \times n$ upper triangular matrix over a P.I.R. $\mathcal{T}$, so that the capacity ideal and its generator $\chi(Q)$ are defined. Then Q is equivalent to a diagonal matrix R which gives the same capacity ideal.*

Proof. By means of right multiplication by a unimodular matrix F, we may form $Q' = Q \cdot F$ to be *lower* triangular, since in particular $\mathcal{T}$ is a Hermite ring. We wish to show that $\chi(Q') = \chi(Q)$. The matrix multiplication includes the first rows:

$$[q_{11} \cdots q_{nn}] \cdot F = [q'_{11}\ 0 \cdots\ 0].$$

By Cramer's rule we may write $q_{11} \cdot \det(F) = \det(G)$, where G is the matrix obtained from F by replacing the first row by $[q'_{11}\ 0 \cdots\ 0]$. But expanding $\det(G)$

gives $q_{11} \cdot \det(F) = q'_{11} \cdot \det(F_{11})$ where F_{jj} is *not* defined to be the minor at (j,j) but rather the matrix formed by deleting from F all rows and columns up to and including the j-th one. In this case where $j = 1$, F_{11} *is* the usual minor.

The purpose of this construction is seen as we observe similarly

$$[q_{jj} \cdots q_{nn}] \cdot F_{j-1,\,j-1} = [q'_{jj}\ 0 \cdots\ 0].$$

which leads by Cramer's rule to $q_{jj} \cdot \det(F_{j-1,j-1}) = q'_{jj} \cdot \det(F_{j,j})$. By telescoping we see that as generators of principal ideals

$$\det(F) \cdot q_{11} \cdots q_{nn} = q'_{11} \cdots q'_{nn}. \tag{6}$$

We are assuming that F is unimodular but we do not want to quote the fact that in this case $\det(F)$ is a unit, as this depends on the formula for the determinant of a product of matrices. The proof of the product formula is an order of magnitude more involved than that of Cramer's rule, and we will not require it. Letting H be the inverse matrix to F, we have $Q = Q' \cdot H$, so the product on the diagonals is evaluated in a manner similar to before, working instead from the final row of Q'. Specifically, we change $Q = Q' \cdot H$ into another valid matrix equation by reversing the rows and columns of all the matrices, replacing the $(i,\,j)$ entry q_{ij} by $q_{n-i+1,\,n-j+1}$. Writing $\tilde{Q} = \tilde{Q}' \cdot \tilde{H}$, we notice that $\det(\tilde{H}) = \det(H)$. The matrix equation has the same form as examined above, an upper triangular matrix multiplied on the right to yield a lower triangular matrix. We obtain as before,

$$\det(H) \cdot q'_{nn} \cdots q'_{11} = q_{nn} \cdots q_{11}$$

which when combined with (4) shows that the capacity ideals from matrix Q and matrix Q' are the same.

There remains to show that Q can be reduced to a diagonal matrix by transformations of this type, a succession of alternately upper and lower triangular matrices. But this is just the content of Algorithm II, applied to a P.I.R. according to Theorem 1.

Lemma B'. *Given a P.I.R. $\mathcal{T}$ and a triangularly presented torsion module M over $\mathcal{T}$, M has a diagonal presentation with the same capacity.*

Proof. This follows immediately from Lemma B and the equivalence between full-rank square matrices and presentations of torsion modules.

For the following results let $\mathcal{T}$ be a Principal Ideal Ring (with zero-divisors). First we make some remarks about uniqueness of the construction that we make. In a commutative ring, we may have $a \cdot c = b$, with other solutions $c' \neq c$ satisfying $ac' = b$. In a domain, if $a \neq 0$, then $c' = c$ is unique. If there are zero-divisors, it may even be so that $c' \mid c$ strictly. This comes up in particular in our NP rings from Proposition 6, that we use to decompose a P.I.R. These are the commutative rings where any zero-divisor is nilpotent. Let a *minimal quotient element* of b by a be an element c such that $ac = b$, and if $ac' = b$ with $c' \mid c$, then also $c \mid c'$.

Lemma I. *In a P.I.R. $\mathcal{T}$, if a divides b, then there is a minimal quotient (element) for b by a.*

Proof. If not, we would have a descending chain of elements with $c_0 = c$ and $c_{i+1} \mid c_i$ where none of the divisibilities can be reversed, hence a strictly ascending chain of ideals which is impossible in $\mathcal{T}$.

Lemma J. *Suppose that Δ divides A and A_1 is a minimal quotient element for A by Δ, so $A = A_1\Delta$. Further, suppose that $(\kappa, A_1) = (1)$ and $A \mid \mathfrak{g}\kappa\Delta$. Then $A_1 \mid \mathfrak{g}$.*

Proof. By the decomposition result in Lemma Z, we may examine for any ring element, its coordinates. Each coordinate is an element of a P.I.D. or of a ring with a principal nilpotent ideal and other properties indicated in Proposition 7. If we show that the conclusion $A_1 \mid \mathfrak{g}$ holds for each coordinate, it will hold also in the large. Suppose first that the chosen coordinate ring $\mathcal{R}^\alpha$, representing the α-th coordinate, is a P.I.D. If $\Delta = 0$, by the minimal property we must have A_1 a unit, so take $A_1 = 1$, and this will divide any $\mathfrak{g}$. If $\Delta \neq 0$, the hypothesis asserts that for some $c \in \mathcal{T}$, the equality $cA_1\Delta = \mathfrak{g}\kappa\Delta$ holds. But since there are no zero-divisors in the coordinate ring, we obtain $c \cdot A_1 = \mathfrak{g}\kappa$, equivalent to what we wanted to prove. The other case is where $\mathcal{R}^i$ is a principal ring where every non-nilpotent element is a unit, and every element is the product of a unit and a power of the nilpotent generator s. We take the nilpotency of s to be r. The case $\Delta = 0$ is handled as in the previous case: A_1 is taken to be 1. If $\Delta = s^k$, then κ must be a unit by the condition on (A_1, κ). If $A = 0$, we must choose by minimality $A_1 = s^{r-k}$, so let $f = r - k$. The coordinate $\mathfrak{g}^\alpha = \mathfrak{g}$ must be a unit times s^f to give $\mathfrak{g} \cdot \kappa\Delta = 0$ which is required. Then obviously $\mathfrak{g}$ is divisible by A_1. In case $A \neq 0$, we examine both sides of $cA_1\Delta = \mathfrak{g}\kappa\Delta$ which since κ is a unit, and all non-zero elements are uniquely represented in the form $u \cdot s^j$, $0 \leq j < r$, we must also have $A_1 \mid \mathfrak{g}$ which was the last case needing proof.

For the remainder of Section 5, $\mathcal{T}$ is a principal ring.

Proposition 8. *Let $\mathcal{T}$-modules N and M be explicitly presented by*

$$N = \frac{\mathcal{T}}{(x_1)} \oplus \cdots \oplus \frac{\mathcal{T}}{(x_s)}, \qquad M = \frac{\mathcal{T}}{(y_1)} \oplus \cdots \oplus \frac{\mathcal{T}}{(y_t)}, \tag{7}$$

so are diagonally presented by Θ_N, Θ_M, and suppose that as modules $N \simeq M$. Then (as ideals) $\chi(\Theta_N) = \prod_{i=1}^{s} x_i = \prod_{i=1}^{t} y_i = \chi(\Theta_M)$. Hence $\chi(\Theta_M)$ does not depend on the particular diagonal presentation and is an invariant of M.

In fact to prove Proposition 8 it is enough to show the following.

Lemma C. *Given a injective module homomorphism $\phi : N \to M$ between two diagonally presented modules as above, we have $\chi(N) \mid \chi(M)$.*

Recall that if $a, b \in \mathcal{T}$, $a \sim b$ means that the ideals (a), (b) are the same (a and b are *associates*.) If we establish Lemma C, Proposition 5 is proven since the inverse of an isomorphism ϕ, $\phi^{-1} : M \to N$ is injective and hence $\chi(M)$ divides $\chi(N)$ capacities (or generators) are associate, verifying Proposition 8.

The key technical result that allows us to prove Lemma C is the following.

Lemma D. *Let $V = \dfrac{\mathcal{T}}{\mathfrak{g}}$ be a cyclic module and consider an injective homomorphism*

$$\psi : V \to M \simeq \frac{\mathcal{T}}{\mathfrak{b}_1} \oplus \cdots \oplus \frac{\mathcal{T}}{\mathfrak{b}_k}.$$

Then the f.g. torsion module $W = \dfrac{M}{\psi(V)}$

(1) *has a diagonal presentation of size* $\leq k$,
(2) *satisfies* $\mathfrak{g} \cdot \chi(\Theta_W) = \chi(\Theta_M)$ *where* Θ_M *is the given diagonal presentation.*

Proof of Lemma C from Lemma D. Given $\phi : N \to M$ injective, consider the restriction ϕ_f to $\frac{\mathcal{T}}{(x_1)} = N_1$, the first summand of N. Since ϕ_f is injective, we can apply Lemma D. By part (1) of the conclusion, the quotient

$$W = \frac{M}{\phi_f(N)}$$

by the image of the first summand, itself has t or fewer summands in some diagonal presentation. Let $\hat{N}$ be the complement of $\mathcal{T}(x_1)$ in the direct sum and ϕ_ℓ the restriction of ϕ to $\hat{N}$. Since ϕ is injective, also the composition of ϕ_ℓ with the projection mapping $\pi : M \to W$ is injective. We employ induction on the number of summands in N, namely s. The induction works since the ground case for $s = 1$ is essentially Lemma D itself. Also by part (1) of Lemma D, a module W that arises in the construction has a diagonal presentation (direct sum) with equal or fewer summands than the previous target module considered. This assertion as part of Lemma D relies upon Lemma B' (where some determinants were employed). Thus the induction hypothesis allows us to assume that $\chi(\hat{N}_1) \mid \chi(\Theta_W)$, but by part (2) of Lemma D, $\chi(\Theta_M) = \chi(\Theta_W) \cdot (x_1)$. Also $\chi(N) = \chi(\hat{N}) \cdot (x_1)$. (We suppress some of the Θ presentation notation.) Putting these three formulas together yields $\chi(N) \mid \chi(M)$ as needed to verify the conclusion of Lemma C, and also Proposition 8.

Proof of Lemma D. Given the injective mapping $\psi : \dfrac{\mathcal{T}}{\mathfrak{g}} \to M \simeq \dfrac{\mathcal{T}}{\mathfrak{b}_1} \oplus \cdots \oplus \dfrac{\mathcal{T}}{\mathfrak{b}_k}$, we can consider each summand component of $\psi(1)$, the image of the unity 1 of $\dfrac{\mathcal{T}}{\mathfrak{g}}$ in M. That is, the projection onto the j-th summand of $\psi(1)$ is called v_j. The element $v_j \in \mathcal{T}$ gives a coset modulo the ideal $\mathfrak{b}_j$ in $\mathcal{T}$ according to the mapping $\pi_j \circ \psi$. That coset is mapped by multiplication into the ideal $\mathfrak{b}_j$ by some elements $y \in \mathcal{T}$, which in fact form an ideal containing $\mathfrak{b}_j$. This ideal $Ann(\mathfrak{v}_j)$, is of course the annihilator of $\mathfrak{v}_j = (v_j)$.

A set of relations for $W \simeq \dfrac{M}{\psi(V)}$ is obtained from any set of relations for $\mathcal{U}$ by adjoining a relation equivalent to the element $\psi(1)$ in M. From the standard diagonal presentation matrix for M, we may now form a presentation matrix of the module $W = \dfrac{M}{\psi(V)}$:

$$\mathcal{F} = \begin{matrix} y_1 & 0 & & 0 \\ 0 & y_2 & & \\ & & \ddots & \\ 0 & 0 & & y_k \\ v_1 & v_2 & & v_k \end{matrix} .$$

At this point we notice that we have proven the first conclusion of Lemma D. The presentation matrix $\mathcal{F}$ for the torsion module W has k columns, and hence

any diagonal presentation matrix derived from $\mathcal{F}$ by allowable operations will be square of size k or less. Working with $\mathcal{F}$, we use row operations to obtain an upper triangular form and reach conclusions about $\chi(W)$.

We compute $\Delta_j = \gcd(y_j, v_j)$ and choose a minimal quotient A_j of y_j by Δ_j. Thus we can take $y_j = A_j\Delta_j$ and $v_j = \kappa_j\Delta_j$, and A_j and κ_j are relatively prime, by the characterization of a Hermite ring discussed in Lemma T, [7]. In other words, A_j is the extra multiplier needed to adjoin to v_j in order to put it into $\mathfrak{b}_j$. This discussion now yields a new way of looking at the presentation, namely

$$\mathcal{F} = \begin{matrix} A_1\Delta_1 & 0 & & 0 \\ 0 & A_2\Delta_2 & & \\ & & \ddots & \\ 0 & 0 & & A_k\Delta_k \\ \kappa_1\Delta_1 & \kappa_2\Delta_2 & & \kappa_k\Delta_k \end{matrix},$$

where $\gcd(\kappa_j, \Delta_j) = 1$.

Henceforth we sometimes write for the gcd of a set of elements of $\mathcal{T}$, $\Gamma = (a_1, \ldots, a_s)$ and for the least common multiple (lcm), the expression $\Lambda = [b_1, \ldots, b_t]$. Concerning the presentation $\mathcal{F}$ we note two critical facts. Since ψ is a well-defined module homomorphism we have that $\mathfrak{g}$, the annihilator for $\mathcal{V}$, must be contained in the annihilator for each component $\mathfrak{v}_i$. Therefore $y_i \mid \mathfrak{g}\mathfrak{v}_i$ for each $i = 1, \ldots, k$, so by Lemma J we obtain $A_i \mid \mathfrak{g}$. Therefore $\Lambda = \operatorname{lcm}_{i=1}^{k} A_i = [A_1, \ldots, A_k]$ is such that $\Lambda \mid \mathfrak{g}$, or $\mathfrak{g} \subset (\Lambda)$.

On the other hand, the mapping ψ is injective, and we notice that since $A_j v_j = 0 \pmod{\mathfrak{b}_j}$ for each summand, we also have $\Lambda\psi(1) = 0$ in M. By injectivity $\Lambda = 0$ in $\mathcal{V} \simeq \frac{\mathcal{T}}{\mathfrak{g}}$, hence $\mathfrak{g} \mid \Lambda$. Putting the two results together yields, as ideals of $\mathcal{T}$,

$$(\Lambda) = [A_1, \ldots, A_k] = \mathfrak{g}.$$

Thus if we can show that $\Lambda \cdot \chi(W) = \chi(M) = \prod_{i=1}^{k} y_i$ we will have completed the proof of Lemma D. We modify the presentation $\mathcal{F}$ so that it takes an upper triangular form, with certain factors deleted from from the "main diagonal". The product of these deleted ring elements generates the "dividing ideal". Step by step we show that this dividing ideal is just Λ, and this completes the proof. We are now ready to begin a reduction of the presentation $\mathcal{F}$ by means of generalized row operations.

Focusing on the first two columns of $\mathcal{F}$, we transform the matrix through left unimodular multiplication to zero out the $(k+1, 1)$ entry and replace the $(1, 1)$ entry by the gcd of those original entries, which equals Δ. The unimodular matrix employed is essentially a 2×2 matrix. The resulting matrix portion is

$$\mathcal{F}' = \begin{matrix} \hat{A}_1\Delta_1 & * & \\ & A_2\Delta_2 & \\ \vdots & & \\ 0 & A_1\kappa_2\Delta_2 & \ldots \end{matrix},$$

We write $\hat{A}_1$ in the first row, even though the A_1 factor was eliminated by the operation, so that it appears clearly as a factor in the dividing ideal. The unknown

entry $*$ arises through the mixing of the $k+1$-st row with the first row under the generalized row operation. The factor A_1 now reoccurs in the $(k+1,2)$ entry for the since the determinant of the 2×2 matrix stays invariant under the unimodular transformation. (Use of small determinant expressions is allowed.)

The sequence of operations applied to $\mathcal{F}$ so far has modified every column of the matrix. New elements represented by $*$ have been adjoined but only to the first row (above the main diagonal). The first entry in the $k+1$-st row has been zeroed out. We now put the remaining submatrix of interest, of rows two and higher, and columns two and higher, into a form similar to the original $\mathcal{F}$. We illustrate this on column two. The entries now in position $(2,2)$ and $(k+1,2)$ are $A_2\Delta_2$ and $v_2' = A_1\kappa_2\Delta_2$ respectively. We need to find the annihilator of v_2'. This is done exactly as before by using $\rho = \gcd(A_1, A_2)$. All explicit quotient elements are taken to be minimal. Letting $\Delta_2' = \rho\Delta_2$, $\kappa_2' = A_1\cdot\kappa_2/\rho$, we may now write the transformed $\mathcal{F}$ as

$$\mathcal{F}'' = \begin{matrix} \hat{A}_1\Delta_1 & * & \\ & \frac{A_2}{(A_1,A_2)}\Delta_2' & \\ \vdots & & \\ 0 & \kappa_2'\Delta_2' & \dots \end{matrix},$$

We see that $(A_1/\rho)\cdot\kappa_2$ is coprime to A_2/ρ by multiplying together the two sides respectively of defining equations

$$\begin{aligned} fA_1 + gA_2 &= \rho \\ n\kappa_2 + mA_2 &= 1. \end{aligned} \tag{8}$$

The new annihilator in the second column of $\mathcal{F}''$ is equal to $A_2' = A_2/\rho$. We may then continue this process of eliminating sub-diagonal elements, on to the second column. In particular, we see that A_2' is a factor of the dividing ideal. Appealing to Lemma B, the "upper triangular" elements $*$ that arise will not affect the determinant of the diagonal presentation matrix which will arise from $\mathcal{F}$.

Examining once more the new presentation matrix $\mathcal{F}''$ for the module M, we see that if the first row and column are deleted, we obtain the truncated matrix

$$\mathcal{K} = \begin{matrix} A_2'\Delta_2' & * & * \\ & \ddots & \\ 0 & \dots & A_k'\Delta_k' \\ \kappa_2'\Delta_2' & \dots & \kappa_k'\Delta_k' \end{matrix},$$

Here $A_k' = \frac{A_k}{\gcd(A_1,A_k)}$. The truncated matrix $\mathcal{K}$ has a form similar to the original $\mathcal{F}$, and we may perform row operations as before, obtaining more zeroes in the final row, and adding in new elements only above the diagonal. Clearly the dividing ideal of $\mathcal{F}$ is A_1 times the dividing ideal of $\mathcal{K}$. In case $k=1$, $\mathcal{F}'$ is diagonal (plus a final "row" of zeroes), $\hat{A}_1 = \hat{A}$ must generate $\mathfrak{g}$, and also generates the dividing ideal. To carry through the induction step from size $k-1$ to size k (number of columns) of the presentation matrices $\mathcal{K}$ and $\mathcal{F}$ it is now only necessary to prove that

$$(\Lambda) \qquad \operatorname{lcm}[A_1,\dots,A_k] \sim A_1\cdot\operatorname{lcm}[\frac{A_2}{\gcd(A_1,A_2)},\dots,\frac{A_k}{\gcd(A_1,A_k)}].$$

This formula follows from the "distributive law" in the principal ideal lattice, and so holds for a distributive Bézoutian ring, and thus for a P.I.R. such as $\mathcal{T}$, as shown in Proposition 15 of Appendix A.

6. Uniqueness of the Smith Form

By the results of the last Section we now know that over a principal ring, two isomorphic f.g. torsion modules possess the same capacity ideals. In this Section we see how this implies that the entire Smith forms are also equal, first for a P.I.D., and then re-using parts of the argument, for a general P.I.R. These uniqueness results are then proved for a completely general commutative ring (where f.g. torsion modules are not always diagonalizable). The cost in getting this stronger result is the use of the ring $\mathcal{R}[X]$, which leads beyond an elementary ring-theoretic formal framework, and the "sophisticated" linear algebra needed for Nakayama's lemma.

We have defined a "Smith form" over any commutative ring $\mathcal{R}$ to apply to a f.p. module M and consist of a finite vector of elements

$$d_1 \mid d_2 \mid \cdots \mid d_s. \tag{9}$$

The relation to the diagonal presentation matrix is that $M \simeq \bigoplus_{i=1}^{s} \frac{\mathcal{R}}{(d_i)}$. At the beginning of the sequence there may be some $d_k = \text{unit}$, so $(d_k) = \mathcal{R}$, which gives a zero term in the direct sum representation of M. At the end of the sequence there may be entries $d_k = 0$ which result in summands $\simeq \mathcal{R}$, that cannot occur when M is a torsion module. However, it is useful not to rule out this case completely, for we will see that zeros may occur in some of the *coordinates* of such elements $\{d_i\}$.

We say that two Smith forms are equal if the lengths of the vectors above are equal and the corresponding ideals (d_i) and (d_i') are the same ideal. Generally we deal with a *reduced* Smith form, where any initial units in the vector (9) have been removed.

Proposition 9. *Over a P.I.D. $\mathcal{T}$, given a f.g. torsion module M, any two reduced Smith forms are the same. Thus for such modules, the (reduced) Smith form is an invariant of module type.*

Proof. We are given two reduced "Smith" vectors $\vec{a}$, $\vec{b}$ that represent M:

$$\begin{array}{cccc} a_1 & a_2 & \dots & a_s \\ b_1 & b_2 & \dots & b_t, \end{array}$$

with $(a_s) \subset (a_{s-1}) \subset \cdots \subset (a_1)$ and similarly for the ideals of vector $\vec{b}$. We will see that it is not hard to show the desired uniqueness once we know that for *all* f.g. torsion modules N, a reduced Smith vector has a determined length. Thus in particular we wish to prove that $s = t$. We assume that $s > t$ and that for $s' < s$ or $s' = s$, $t' < t$, we have that any module N represented in two ways with lengths s', t' actually has only one reduced Smith representation, in particular necessarily $s' = t'$.

Next we note, using Facts 1 and 2, that we may take $a = a_1 \mid b_1$, or $b_1 = f_1 a_1$, since if (b_1) has an element x not in (a_1), the module $x \cdot M$ has a reduced Smith vector of length s from $\vec{a}$, but a reduced Smith vector of length $t' < t$ from vector $\vec{b}$, and this contradicts our induction hypothesis.

Next we have by Proposition 5 that the capacities are equal:

$$(\prod \vec{a}) = (\prod \vec{b}). \tag{10}$$

But in forming the module $a \cdot M$, $a = a_1$, we obtain from the two given Smith forms, two new forms (which might not be reduced),

$$\begin{array}{cccc} \frac{a_1}{a} & \frac{a_2}{a} & \cdots & \frac{a_s}{a} \\ \frac{b_1}{a} & \frac{b_2}{a} & \cdots & \frac{b_t}{a}. \end{array}$$

making use of Facts 1 and 2. Applying Proposition 5 to the new module aM gives $a^s \prod \vec{a} = a^t \prod \vec{b}$ which when combined with (10), since we are in a domain, implies that $a = a_1$ is a unit, contradicting the reduced quality of the Smith vector $\vec{a}$.

Finally we come to "the Kaplansky argument", from [14, pg. 479] which cannot much be improved upon. Once we know that all Smith forms have the same length, we look for the first stage at which they differ: $(a_j) \neq (b_j)$. Taking an element x from one ideal, say (a_j), that is not in the other, we express the module xM as the direct sum of quotient modules in two ways, one with not more than $s-j$ summands (coming from $\vec{a}$) and the other with at least $s - j + 1$ non-zero summands. The final argument by induction in [14] is unnecessary: we already have a contradiction since all reduced Smith forms must have the same length.

Remark 3. The Kaplansky approach to the uniqueness of the Smith form was taken up in a similar treatment in [3, pg. 190] The proof in [15] applies only to a Euclidean ring and relies on primary factorization. One reason other approaches have been preferred may be Kaplansky's Lemma 9.2, relying on a result from [11] in an earlier style. This argument is replaced by [3, Lemma 15.12], relying on non-constructive techniques. The "most general" result on number of generators is discussed below, owing to results of Vasconcelos and Strook, [16, pg. 226]. Another reason that the Kaplansky method did not find favor may be that the author asserted, [14, pg. 479], that his arguments apply not only to commutative rings, but also to "duo" rings, where all one-sided ideals are actually two-sided. This result apparently has not been quoted since then[2].

With this preamble it is relevant to consider "invariant basis" results. Though these Propositions concern free modules, we gain insight into torsion modules as well. We emerge with three distinct proofs that a P.I.D. $\mathcal{T}$ has the property that a f.g. free module has a determined number of generators. The second proof applies to any P.I.R., and the third comes from the Vasconcelos/Strook result and applies to all commutative rings.

Proposition 10. *Any P.I.D. $\mathcal{T}$ has the invariant basis property.*

Proof. Suppose that there is an isomorphism $\phi : \mathcal{T}^n \to \mathcal{T}^m$ with an inverse ψ, and $n > m$. Let $\mathbf{e}_j$ be the standard basis element of $\mathcal{T}^n$, written as a row vector, whose only non-zero entry is 1 in the j-th position. Similarly we write a standard basis $\{\mathbf{f}_k\}$ for $\mathcal{T}^m$. We express for $1 \leq j \leq m$, the m-vector

$$\phi(\mathbf{e}_j) = (a_{j1}, a_{j2}, \ldots, a_{jm}).$$

By the Hermite property the $m \times m$ matrix $A = [a_{jk}]$ can be put into upper triangular form through left multiplication by a unimodular $m \times m$ matrix U. The

[2]The fact that for square matrices in a duo ring, $AB = I$ implies $BA = I$, was shown by K. Goodearl using methods of Artin rings, according to a personal communication of T.Y. Lam.

given generators $\{\mathbf{e}_j\}$ are well expressed by the $m \times n$ matrix T, which is the $m \times m$ identity matrix with a $m \times (n-m)$ block of zeros appended. The rows of $U \cdot T$ now represent elements from the pre-image of ϕ that map into the respective rows of $A^u = U \cdot A$. Therefore none of the rows of A^u are zero, else it comes from a non-zero row of UT which violates the injectivity of ϕ. The problem now is to construct $y \in \mathcal{T}^m$ such that $\psi(y) = \mathbf{e}_{m+1}$.

The final row of A^u has only a single non-zero entry, the (m, m) corner entry c. Now if $\psi(\mathbf{f}_m)$ had a non-zero component of $\mathbf{e}_{m+1}$, this would be killed by c, since the pre-image of this m-th row has a zero $m+1$ column entry. Thus since $\mathcal{T}$ has no zero-divisors, the $\mathbf{e}_{m+1}$ must be zero in the first place. Now if $\psi(\mathbf{f}_{m-1})$ had a non-zero $\mathbf{e}_{m+1}$ component, so would $\psi(a_{m-1,m-1}\mathbf{f}_{m-1} + a_{m-1,m}\mathbf{f}_m)$, that is, the image under ψ of the next-to-last row of A^u. But this image is explicitly known as the $m-1$ row of $U \cdot T$ and thus has no $\mathbf{e}_{m+1}$ component. Proceeding up the rows of A^u in this manner makes it clear that $\mathbf{e}_{m+1}$ has no pre-image under ψ, the element y does not exist, so ϕ cannot be an isomorphism if $n > m$.

Remark 4. The "advantage" of this proof is that it uses only generalized row operations and no change of basis, but it works easily only when zero-divisors are not present. We now come to a result slightly more involved than the others, and where some details of the proof have been left to the reader.

Proposition 11. *For any P.I.R. $\mathcal{T}$, a f.g. torsion module M has a unique reduced Smith form.*

Proof sketch. The existence is Theorem 1. As in Proposition 9, we start with Smith vectors $a_1 \mid \alpha_2 \mid \cdots \mid a_s$ and $b_1 \mid b_2 \mid \cdots \mid b_t$ where $s > t$ and we can assume that $b_1 = f_1 \cdot a_1$. Thus there is an isomorphism

$$\phi : \frac{\mathcal{T}}{(a_1)} \oplus \cdots \oplus \frac{\mathcal{T}}{(a_s)} \to \frac{\mathcal{T}}{(b_1)} \oplus \cdots \oplus \frac{\mathcal{T}}{(b_t)}.$$

In the manner of Lemma D, we examine the injective mapping

$$\psi_1 : V = \frac{\mathcal{T}}{(a_1)} \to \frac{\mathcal{T}}{(b_1)} \oplus \cdots \oplus \frac{\mathcal{T}}{(b_t)}.$$

The image of the unity 1 under ψ_1 appears in the j-th summand as v_j. Since the ring $\mathcal{T}$ is not necessarily a domain, we need to work with the coordinates of its elements according to Proposition 6. Each coordinate, say v_i^α lies either in a domain (P.I.D.) or a nilpotent (NP) ring.

Assume first of all that all of the α-rings involved are domains. Then not all the coordinates a_1^α can be units, for if so, the Smith form $\vec{a}$ would not be reduced. Pick a specific coordinate α where a_1 is not a unit. We now concentrate on the α-coordinates of $\vec{a}$ and $\vec{b}$ as we form diagonal presentations of the cokernels of the successive injections induced by ϕ, starting with ψ_1 and successively to $\psi_i : \frac{\mathcal{T}}{(a_i)} \to \operatorname{coker}(\psi_{i-1})$, $i = 2, \ldots, s$. Use $\{c_j\}$ to denote generators of those ideals that arise from $\{b_j\}$ after cokernel operation has been performed a succession of times. Note that if $a_i^\alpha \neq 0$, and $c_j^\alpha = 0$, then the image component $v_j^{i,\alpha} = \psi_i(1)$, which has vector index j and ring coordinate α, must be zero. This follows from the homomorphism property of ψ_i since the α coordinate of $a_i \cdot 1$ is zero but cannot be in α coordinate of $\frac{\mathcal{T}}{(c_j)}$ since that coordinate ring is a domain. Furthermore and by contrast, when

$a_i^\alpha = 0$, the element $\psi_i(1)$ must have a non-zero component in α-coordinate in $\frac{\mathcal{T}}{(c_j)}$, where in fact $c_j^\alpha = 0$. If this were not the case, consider the element c, the product of all those c_j in which $\psi_i(1)$ has a non-zero image component. All of the α-coordinates of these particular $\{c_j\}$ are non-zero, so by the domain property of this α coordinate ring, we have $c^\alpha \neq 0$. But $c \cdot \psi_i(1) = 0$ by construction, so by the injectivity property, $c = 0$ in $\frac{\mathcal{T}}{(a_i)}$. This is a contradiction since $c^\alpha \neq 0$ in $\mathcal{T}^\alpha$, and since $a^\alpha = 0$ by assumption, there is no "modulo" taken at all.

The result of these arguments is that for coordinate α, the number of 0's on the left in $\vec{a}$ is the same as the number on the right in $\vec{c}$. Each 0 α-coordinate is eliminated by the cokernel operation, so we have non-zero entries of $\psi_i(1)$ in the torsion-free entries of $\vec{c}$. On the right hand side, renaming the vector components using generalized column operations, forming the cokernel turns exactly one torsion-free component into a new torsion entry, say $\frac{\mathcal{T}}{(d_k)}$. Thus there are at least as many zeros on the right as on the left, and by a similar argument on the inverse ϕ^{-1}, the number of zeros must be equal.

We are using a more refined "invariant" than was needed for the proof in Proposition 9, the P.I.D. case. There we directly employed the capacity $\chi(M)$. Here we want to look at the capacity in each α-coordinate, but keep it non-zero by leaving out the zero vector components of $\vec{c}$ in the product. This works since for each α, the number of zeros on both sides is the same. To finish the argument, we get this "refined" $\chi^\alpha(\vec{a})$ and $\chi^\alpha(\vec{b})$ as equal, based on Lemma D (or Proposition 8) and the isomorphism property. Next examine $a_1 \cdot M$. We easily obtain a new presentation of this module both from $\vec{a}$ and $\vec{b}$, through dividing the components by a_1. On the left in α-coordinate there are fewer non-zero components b_j^α than non-zero a_i^α. Thus the total factors a_1^α divided out on the right is fewer, but χ of both sides must remain the same. This leads as in Proposition 9 to a_1^α being revealed as a unit. This however contradicts the choice of α, precisely without leading unit components a_1.

The general case is where both α P.I.D.'s and NP rings can occur. The remaining argument then deals with an NP ring $\mathcal{T}^\alpha$ where $\vec{a}^\alpha$ has no leading units. If σ is the nilpotent generator for $\mathcal{T}^\alpha$, the vectors $\vec{a}$ and $\vec{b}$ are just finite sequences of increasing powers of σ. Again consider just the α-coordinate. Here the classical capacity χ is too weak an invariant to characterize the possible vectors. It is shown that each time the cokernel of ϕ_i is taken, the *total* power of σ on the right hand side $\vec{c}$ goes down by k, where $a_i = \sigma^k$, up to multiplicative unit. Any zero entry is treated the same, namely as σ^r where r is the power of nilpotency. This is consistent with selecting the "minimal quotient element" when dividing. Similar to before, instead of χ, the "total power" invariant on both sides must be the same. But since $\vec{a}$ is longer than $\vec{b}$, when we form $a_1 \cdot M$, we get two presentations with the same lengths as before, but which do not have the same total power, since σ^k is removed more times from one side than the other. This contradiction finally establishes that the diagonal presentations of M denoted $\vec{a}$ and $\vec{b}$ had to have the same length, and by the dependent argument in fact do give identical Smith forms.

We restate this result as a theorem, but note its generalization in Proposition 14. Consider f.g. torsion modules.

Theorem 2. *Over a P.I.R. the (reduced) Smith form depends only on module type (isomorphism class).*

Comment. This result depends on Proposition 8 and thus essentially on Lemmas

C and D. Again the Smith form is defined as a consequence of Theorem 1.

Proposition 12. *Any P.I.R. has the invariant basis property.*

Proof. Although the argument appears familiar in view of the above, use of coordinates is not necessary. We have a situation with $\vec{a}$ and $\vec{b}$ as before, but all $a_i = b_j = 0$. The first component $\mathcal{T}$ on the left maps to $(v_1, \ldots, v_t)$ in the free direct sum of t components on the right. Renaming generators leads to this module being written

$$\frac{\mathcal{T}}{(d_1)} \oplus \mathcal{T} \oplus \cdots \oplus \mathcal{T}, \tag{11}$$

where $d_1 = \gcd(v_1, \ldots, v_t)$ (some of these entries may be zero) using generalized column operations. Thus reduction of one free term from the left leads to reduction of a free term from the right. We then show as in the last Proposition that a free term must map its generator into a vector with at least one non-zero component in a free (non-cyclic) term on the RHS. Therefore the number of components on the left was not larger than on the right, and by symmetry they are equal. Thus the size of a basis is well-defined.

We conclude with a general method to demonstrate the uniqueness of the Smith form. We adhere to our "elementary" constraints, but will use the $\mathcal{R}[X]$ construction and the method of Vasconcelos and Strook, see [18]. No "principal" or "Hermite" requirement is made on the ground ring $\mathcal{R}$ in the rest of the Section.

Lemma R. *Let A be an $n \times n$ matrix over a commutative ring $\mathcal{R}$, seen as a linear mapping ϕ of $M = \bigoplus_{i=1}^{n} \mathcal{R}$ to itself by means of*

$$\phi(\mathbf{e}_i) = \sum_{j=1}^{n} a_{ji}\mathbf{e}_j. \tag{11}$$

Then if ϕ is surjective, ϕ is injective. Equivalently, if A has a left inverse B such that $B \cdot A = I_{n\times n}$ then A has an $n \times n$ right inverse C, with $A \cdot C = I$.

Proof. With a standard basis $\{\mathbf{e}_i\}$ of $M = \mathcal{R}^n$, we have M as an $\mathcal{R}[X]$ module by the action of A in the usual way. The surjectivity of ϕ implies the module equation, for some $\beta_{ij} \in \mathcal{R}$,

$$-\mathbf{e}_i = \beta_{i\,1} X\mathbf{e}_1 + \cdots + \beta_{i\,n} X\mathbf{e}_n,$$

for $a \leq i \leq n$. These equations lead to a square matrix of relations with $\beta_{i\,i}X + 1$ on the diagonal and $\beta_{i\,j}X$ off the diagonal. Thus Proposition 2, Nakayama's Lemma, applies and we have an annihilator $\theta(X)$ of M which is $\equiv 1\ (X)$. Now suppose that $u \in M$ is in the kernel of the operator A. Thus if $\theta(X) = \omega(X) \cdot X + 1$, we see that $0 = \theta \cdot u = u$, so the kernel of ϕ vanishes and the other conclusions follow.

Corollary 1. *If for $n \times n$ matrices A, B over a commutative ring $\mathcal{R}$ we have $B \cdot A = I$, then also*

$$A \cdot B = I.$$

Proof. By Lemma R there is a right inverse C for A, hence $C = (BA)C = B(AC) = B$.

If the classical adjoint construction is available, it is easy to show that $B \cdot A = I$ implies that A has a right inverse. Indeed, $ABA \cdot \mathrm{adj}(A) = AB$ up to a unit, since the determinant is a unit. But by hypothesis, this will also itself equal a unit, substituting I for $B \cdot A$.

Proposition 13. *Any commutative ring $\mathcal{R}$ has the invariant basis property. Furthermore, if a basis has s elements, then any generating (spanning) subset of the module has at least s elements.*

Proof. Suppose M as above had two bases, generating sets $\{\mathbf{e}_i\}$, $i = 1, \ldots, s$, and $\{\mathbf{f}_j\}$, $j = 1, \ldots, t$, which are linearly independent. If $s < t$, so the second set is more numerous than the first, form a consistent linear mapping $\zeta : M \to M$ by definition on the basis elements $\mathbf{f}_j$, $1 \leq j \leq t$, $\zeta(\mathbf{f}_j) = \mathbf{e}_j$ for $j \leq s < t$ and $\zeta(\mathbf{f}_j) = 0$ for $s < j \leq t$. This mapping is surjective and by Lemma R, must be injective, but $\zeta(\mathbf{f}_t) = 0$ so this is absurd. Thus the cardinality of basis for a finite power of $\mathcal{R}$ is uniquely defined. For the final part, note that in the above argument we did not use the linear independence of $\{\mathbf{e}_i\}$, only its spanning property.

Proposition 14. *If a f.g. torsion module over a commutative ring has a Smith form decomposition, this is the only one such.*

Proof. In previous work the difficult part was to show that any two reduced Smith forms have the same number of diagonal entries. Suppose that $S_1 \supset \cdots \supset S_s \neq (0)$ are proper, non-zero ideals of $\mathcal{R}$ leading to a Smith form

$$M \simeq \frac{\mathcal{R}}{S_1} \oplus \cdots \oplus \frac{\mathcal{R}}{S_s}. \tag{13}$$

For any module N over $\mathcal{R}$ and an ideal $\mathcal{I} \subset \mathcal{R}$ we may form the $\mathcal{R}$-module $\tilde{N} = \frac{N}{\mathcal{I}}$ by introducing the general relation $\tilde{r} \cdot n \equiv 0$ for $\tilde{r} \in \mathcal{I}$, $n \in N$. For the module M take $\mathcal{I} = S_1$. Since all $S_i \subset S_1$ we obtain

$$\tilde{M} = \frac{\mathcal{R}}{S_1} \oplus \cdots \oplus \frac{\mathcal{R}}{S_1},$$

with s summands, where all the ideals forming quotients have been replaced by their sum S_1, and may form the projection $\pi : M \to \tilde{M}$. Now *any* generating set for M over $\mathcal{R}$ leads to a generating set for $\tilde{M}$ by projection. We may regard $\tilde{M}$ as constructed, as a free module over $\tilde{\mathcal{R}} \simeq \frac{\mathcal{R}}{S_1}$, and also we obtain a set of generators for this free module. But the standard basis for $\tilde{M} \simeq \tilde{\mathcal{R}}^s$ over $\tilde{\mathcal{R}}$ has cardinality s, so by Proposition 13, any such generating set must have at least s elements. If a *shorter* Smith decomposition existed, of length t, with ideals T_j in place of S_i in (13), and $1 \leq j \leq t < s$, we get one generator for M from each summand, hence have t generators that project to a generating set for $\tilde{M}$ over $\mathcal{R}$, and to the free module $\tilde{M}$ over $\tilde{\mathcal{R}}$. But this gives a contradiction. Now one can apply the "Kaplansky argument" as in the proof of Proposition 9, to finish the proof of Proposition 14.

Appendix A: Formulas in a Bézoutian Ring

We list with proof some useful formulas involving gcd, lcm and so forth, culminating in Proposition 15 which was key to establishing Theorem 2 on the invariance of the Smith form. The proofs are carried out with the same philosophy we adopted for the algorithms and theorems given above. That is, we attempt to maintain a principle of strict construction of entities used in a proof. Given a pair of elements of the Euclidean domain $\mathcal{R}$, we have the Euclidean algorithm for their gcd which is the epitome of such an algorithmic construction. In treating a more general ring such as a Bézoutian ring, we consider and deem that to find a principal generator of the span of two elements, is to evaluate a computationally simple function at the pair of elements. In talking about a matrix M, all matrices in subsequent discussion about invariants were no larger than M.

Consider the Fundamental Theorem of Arithmetic. It is highly non-constructive in that even over the integers there is no known way to find the factorization much better than by a search, where the computational cost is at least the order of a high degree polynomial, in the number of digits. If $\mathcal{R}$ is a polynomial ring over an infinite field, this factorization is not possible using algebraic methods at all, in general. Those working in the foundations of different areas of mathematics are certainly motivated to seek proofs of conceptual simplicity, which may sometimes be longer than sophisticated proofs. One reason is to present proofs of sufficient rigor to be affirmed by an automated proof-checker. In the history of mathematics, the "elementary" proof, one that did not incorporate too many advanced results from outside the field of the theorem itself, was sought after. The Prime Number Theorem itself is the best known example of this.

The formulas presented below are meant to express an identity of principal ideals. That is, if we write

$$\Phi_1(x, y, z) \sim \Phi_2(x, y, a),$$

we should have two computable expressions, each of which may not even be uniquely defined as a ring element, but whose generated ideal is well-defined and equal to the other ideal. For example, if $\mathcal{R}$ is not a domain, $z = \gcd(a, b)$ is not well-defined, even up to a unit multiplier, but (z) is well-defined. When $a = a_1\Delta$, $b = B_1\Delta$, $\Delta = fa + gb$, the function $\Phi_1(a, b) = f$ does not define a unique ideal in general, but as we saw in Proposition 4, the expressions Δ and $f \cdot g\,\Delta$ are well-defined in this sense, that the ideal generated *is* independent of the choice of Δ, f, g. In what follows all Roman letters are elements of $\mathcal{R}^+$ (non-zero elements). Expressions such as $\gcd(0, a)$ are consistently given the value a.

Fact 3. *For $a, b, c \in \mathcal{R}$ we have $((a, b), c) \sim (a, (b, c))$ and $[[a, b], c] \sim [a, [b, c]]$.*

Proof. For the first part, the principal ideal expressed either way is the sum of the three ideals $(a), (b)$ and (c). For the second part, we have two ways of expressing a principal generator of the intersection of the three ideals. More generally, an expression of k given variables with parentheses placed consistently represents a divisor of all the variables, which is a multiple of any other such divisor. Thus we may write

$$(a_1, \ldots, a_k),$$

which represents (uniquely up to association) an element $x \in \mathcal{R}$ such that $x \mid a_j$, $1 \leq j \leq k$, and for some $\{\alpha_j \in \mathcal{R}\}$, we have $x = \alpha_1 a_1 + \cdots + \alpha_k a_k$.

The following Facts are also simple consequences of the set-theoretic definition of lcm and gcd.

Fact 4. *We have*

$$\operatorname{lcm}[A_1, \ldots, A_n] \sim [A_1\, [\ldots [A_{n-1}, A_n] \ldots] \sim [\ldots [A_1, A_2] \ldots]\, A_n].$$

This states that one may compute an lcm of a vector of elements through successively finding the principal generator of an intersection of two principal ideals, according to some consistent placement of the parentheses.

Fact 5. *For $a, b, z \in \mathcal{R}$, $(az, bz) \sim z(a, b)$.*

Proof. This follows since the span of az and bz is the same as z times the span of a and b.

The following two dual formulas reiterate the distributive law of the lattice of principal integral ideals of a Bézoutian ring.

Fact 6. *For all $a, b, c \in \mathcal{R}^+$,*

$$\begin{aligned} (a, [b, c]) &\sim [(a, b), (a, c)] \\ [a, (b, c)] &\sim ([a, b],\, [a, c]). \end{aligned} \tag{14}$$

Proof. This follows from Proposition 5, and the explicit formulas in Definition 1.

Fact 7. *For $a, b_1, \ldots, b_k \in \mathcal{R}$, we have*

$$(a, \operatorname{lcm}[b_1, \ldots, b_k]) \sim \operatorname{lcm}[(a, b_1), \ldots, (a, b_k)].$$

Proof. By Fact 4 we may rewrite the lcm in the LHS as a sequence of least common multiples of pairs of elements, nested to the right. That is, the LHS equals $(a, [b_1, [b_2, [\ldots, [b_{k-1}, b_k] \ldots])$. This expression by use of Fact 6 now equals $[(a, b_1), (a, [b_2, \ldots, b_k])]$. This process can be repeated a finite number of times on the inner gcd so eventually we obtain a nested sequence of brackets of the form $[(a, b_1), [(a, b_2), [\ldots, [(a, b_{k-1}, (a, b_k)] \ldots]$ which by Fact 4 is equal (associate) to the RHS of the conclusion.

Fact 8. *For elements $a, b, c \in \mathcal{R}$,*

$$a \cdot \left(\frac{b}{(a, b)}, \frac{c}{(a, c)} \right) \sim a \cdot \frac{(b, c)}{(a, b, c)}.$$

Proof. This follows from the second part of Fact 6, $([a, b], [a, c]) \sim [a, (b, c)]$ where we rewrite $[a, b] = \frac{ab}{(a, b)}$ and similarly, and by Facts 3 and 5.

The next formula is included for completeness but is not used in the sequel, though it will generate a "max-min" formula in Appendix B.

Fact 9.

$$A \cdot \left(\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right) \sim A \cdot \frac{(B_1, \ldots, B_k)}{(A, B_1, \ldots, B_k)}.$$

Proof. A suitable ground case is just Fact 8. For $k > 2$ we have

$$\left(\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right) \sim \left(\frac{B_1}{(A, B_1)}, (\frac{B_2}{(A, B_2)}, \ldots, \frac{B_k}{(A, B_k)}) \right)$$

from Fact 3, which in turn equals, using an induction hypothesis for smaller k, $\left(\frac{B_1}{(A,B_1)}, \frac{(B_2,\ldots,B_k)}{(A,(B_2,\ldots,B_k))} \right)$. Letting $a = A$, $b = B_1$, $c = (B_2, \ldots, B_k)$ and applying Fact 8 again yields

$$\frac{(B_1, (B_2, \ldots, B_k))}{(A, (B_1, (B_2, \ldots, B_k)))},$$

which equals the desired RHS in the statement of Fact 9.

Fact 10. *We have the following Product Formula for least common multiple.*

$$A \cdot [B_1, \ldots, B_k] \sim A \cdot \left[\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right] \cdot [(A, B_1), \ldots, (A, B_k)].$$

Proof. The statement is true if $k = 1$, we assume also that it holds for values less than our given k. Taking brackets nested to the right, multiplying top and bottom by (A, B_1), and expanding $[B_2, \ldots, B_k]$ according to the induction hypothesis yields

$$\begin{aligned}[B_1, \ldots, B_k] &\sim \frac{B_1 \cdot [B_2, \ldots, B_k]}{(B_1, [B_2 \ldots B_k])} \\ &\sim \frac{B_1}{(A, B_1)} \cdot (A, B_1) \left[\frac{B_2}{(A, B_2)}, \ldots, \frac{B_k}{(A, B_k)} \right] \cdot \frac{[(A, B_2), \ldots, (A, B_k)]}{(B_1, [B_2 \ldots B_k])}.\end{aligned}$$

Using definitional formulas for the lcm yields

$$\begin{aligned}&\left[\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right] \cdot \left(\frac{B_1}{(A, B_1)}, \left[\frac{B_2}{(A, B_2)}, \ldots, \frac{B_k}{(A, B_k)} \right] \right) \\ &\qquad \cdot \frac{[(A, B_1) \ldots (A, B_k)] \cdot ((A, B_1), [(A, B_2) \ldots (A, B_k)])}{(B_1, [B_2 \ldots B_k])}.\end{aligned}$$

Comparing this expression with the desired RHS of the conclusion, we need only prove the following formula.

$$\begin{aligned}(15) \quad &A \cdot \left(\frac{B_1}{(A, B_1)}, \left[\frac{B_2}{(A, B_2)}, \ldots, \frac{B_k}{(A, B_k)} \right] \right) \\ &\qquad \cdot ((A, B_1), [(A, B_2) \ldots (A, B_k)]) \sim A \cdot (B_1, [B_2 \ldots B_k]).\end{aligned}$$

Using Fact 7, the LHS of (15) is associate to

$$\begin{aligned}&A \cdot \left[(\frac{B_1}{(A, B_1)}, \frac{B_2}{(A, B_2)}), \ldots, (\frac{B_1}{(A, B_1)}, \frac{B_k}{(A, B_k)}) \right] \\ &\qquad \cdot [(A, B_1, B_2), (A, B_1, B_3), \ldots, (A, B_1, B_k)],\end{aligned}$$

and now applying Fact 8 in the first lcm factor and Fact 3 in the second factor gives

$$A \cdot \left[\frac{(B_1, B_2)}{(A, (B_1, B_2))}, \ldots, \frac{(B_1, B_k)}{(A, (B_1, B_k))} \right] [(A, (B_1, B_2)) \ldots (A, (B_1, B_k))].$$

Setting $F_1 = (B_1, B_2), \ldots, F_{k-1} = (B_1, B_k)$ we see that by the present Fact 10 for $k-1$, this last expression equals

$$A \cdot [F_1, \ldots, F_{k-1}] \sim A \cdot [(B_1, B_2), (B_1, B_3), \ldots, (B_1, B_k)],$$

which by Fact 7 is equal to the RHS of (15). This is what was required to complete the proof of Fact 10.

Proposition 15.

$$A \cdot \left[\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right] \sim [A, B_1, \ldots, B_k].$$

Proof. By Fact 8 we obtain

$$A \cdot \left[\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right] \sim A \cdot \left[\frac{B_1}{(A, B_1)}, \ldots, \frac{B_k}{(A, B_k)} \right] \frac{[(A, B_1) \ldots (A, B_k)]}{(A, [B_1 \ldots B_k])}.$$

Utilizing Fact 10 and the formula $[x, y](x, y) = xy$, this yields immediately

$$\frac{A \cdot [B_1, \ldots, B_k]}{(A, [B_1, \ldots, B_k])} \sim [A, B_1, \ldots, B_k],$$

which completes the proof of Proposition 15 and Theorem 2.

Appendix B: Max-min formulas for functions with values in an ordered group

Let $\mathcal{X}$ be a set and $\mathcal{C}$ be a subgroup of the additive group of real numbers $\mathbb{R}$. Then functions $f : \mathcal{X} \to \mathcal{C}$ have an abelian group structure and also a lattice structure inherited from the ordering. For example, if $\mathcal{X}$ is the discrete set $\{1, 2, \ldots, k\}$, such a function is a vector

$$(v_1, \ldots, v_k)$$

with $v_i \in \mathcal{C} \subset \mathbb{R}$. The vectors may be added, subtracted and compared to produce $\vec{w} = \max(\vec{u}, \vec{v})$, where $w_i = \sup(u_i, v_i)$.

In greater generality we have for $t \in \mathcal{X}$

$$f \wedge g(t) = \max\{f(t), g(t)\}$$
$$f \vee g(t) = \min\{f(t), g(t)\},$$

or

$$f \wedge g = \frac{1}{2}(f + g) + \frac{1}{2}|f - g|$$
$$f \vee g = \frac{1}{2}(f + g) - \frac{1}{2}|f - g|.$$

We set down some formulas in this "lattice-ordered group", that correspond to the Facts of Appendix A. The formulas of Appendix A apply in particular to the domain $\mathbb{Z}$, so allowing ourselves finally to write out gcd and lcm explicitly given the primary factorizations, we consider the vectors of exponents. For example, if $a = 5^3 7^8 11^5$, $b = 5^4 7^5 11^2$, the formulas $\gcd(a, b) = 5^3 7^5 11^2$, $\mathrm{lcm}(a, b) = 5^4 7^8 11^5$ are the same as $u \vee v = (3, 5, 2)$, $u \wedge v = (4, 8, 5)$, considering only those three exponents. This correspondence between results on integers, and on vectors of exponents, is convincing for the case where the exponents are natural numbers. Then the formulas are not always defined, but are valid when defined (when one can perform subtraction). There is no reason they should not hold more generally, for any ordered abelian group $\mathcal{C}$. Of course it is rather routine to prove the results through axioms (where some form of lattice distributive law will enter in), or in the concrete case of the lattice of functions.

Formula 1. *(After Fact 8) For vectors u, v, w we have*

$$\min(v - \min(u, v), w - \min(u, w)) = \min(v, w) - \min(u, v, w).$$

Formula 2. *(After Fact 9) For functions $g, f_1, \ldots, f_k : \mathcal{X} \to \mathcal{C}$*

$$\min(f_1 - \min(g, f_1), f_2 - \min(g, f_2), \ldots, f_k - \min(g, f_k)) \\ = \min(f_1, \ldots, f_k) - \min(g, f_1, \ldots, f_k).$$

Formula 3. *(After Fact 10) As above we have,*

$$\max(f_1, \ldots, f_k) = \\ \max(f_1 - \min(g, f_1), \ldots, f_k - \min(g, f_k)) + \max(\min(g, f_1), \ldots, \min(g, f_k)).$$

Formula 4. *(After Proposition 15) With same assumptions we have,*

$$g + \max\{f_1 - \min(g, f_1), \ldots, f_k - \min(g, f_k)\} = \max\{g, f_1, \ldots, f_k\}.$$

References

[1] M.F. Atiyah & I.G. MacDonald, *Introduction to Commutative Algebra*, Addison-Wesley, Reading, MA, 1969.

[2] G. Birkhoff, *Lattice Theory (3rd ed.)*, American Mathematical Society, Providence, 1967.

[3] W.C. Brown, *Matrices over Commutative Rings*, Marcel Dekker, New York, 1993.

[4] R.W. Butler and J.A. Sjogren, *A PVS Graph Theory Library*, NASA/TM-1998-206923, February 1998.

[5] P.M. Cohn, *Free Rings and their Relations, 2nd Edition*, Academic Press, London, 1985.

[6] *www.jon-arny.com* .

[7] L. Gillman and M. Henriksen, *Some Remarks about Elementary Divisor Rings*, Trans. Amer. Math. Soc. **82** (1956), 362–365.

[8] A.M.W. Glass, *Partially ordered groups*, World Scientific, Singapore, 1999.

[9] G.A. Gratzer, *Lattice theory; first concepts and distributive lattices*, W.H. Freeman, San Francisco, 1971.

[10] R.M. Guralnick, L.S. Levy & C. Odenthal, *Elementary divisor theorem for noncommutative PID's*, Proc. Amer. Math. Soc. **103** (1988), 1003–1011.

[11] Nathan Jacobson, *The Theory of Rings*, American Math. Soc., New York, 1943.

[12] Paul Jaffard, *Les systèmes d'idéaux*, Dunod – Travaux et Recherches Mathématiques IV, Paris, 1960.

[13] Thomas Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[14] Irving Kaplansky, *Elementary divisors and modules*, Trans. Amer. Math. Soc. **66** (1949), 464–491.

[15] M.E. Keating, *A First Course in Module Theory*, Imperial College Press, London, 1998.

[16] T.Y. Lam, *Exercises in Classical Ring Theory*, Springer-Verlag, New York, 1995.

[17] M.D. Larsen, W.J. Lewis & T.S. Shores, *Elementary Divisor Rings and Finitely Presented Modules*, Trans. Amer. Math. Soc. **187** (1974), no. 1, 231–248.

[18] H. Matsumura, *Commutative Algebra*, Benjamin/Cummings, Reading, MA, 1980.

[19] N.H. McCoy, *The Theory of Rings*, Macmillan, New York, 1973.

[20] Wm. McCune, *Otter and first-order theories* (February 2002), Argonne National Laboratory.

[21] C. Moler and C. Van Loan, *Nineteen Dubious Ways to Compute the Exponential of a Matrix*, SIAM Review **20** (1978), no. 4, 801–836.

[22] *Otter: An Automated Deduction System*, www-unix.mcs.anl.gov/AR/otter .

[23] B.L. van der Waerden, *Algebra II*, Springer-Verlag, Berlin, 1967.

AROSR/NM
4015 Wilson Blvd. Suite 713
Ballston Station, Virginia
22203-1954
E-mail address: jon.sjogren@afosr.af.mil