Thesis Proposal

Gregory Minton

December 8, 2011

Thesis Title: Topics and Algorithms in the Theory of Quadratic Forms

Advisor: Abhinav Kumar

The theory of quadratic forms is a rich subject lying at the intersection of number theory, algebraic geometry, and discrete geometry. This confluence results from the fact that a variety of questions, from sphere packing in Euclidean space to certain local-global principles in arithmetic geometry, can be phrased naturally in the language of lattices and quadratic forms. In recognition of the diversity of techniques and applications admitted by this field, we propose to study a collection of problems instead of just one. By doing so, we intend to become well-versed in the varied branches of the discipline and thus be prepared to make fruitful connections between them.

This proposal is divided into three sections, reflecting the high-level classification of the field into geometric, arithmetic, and algebraic theories. In each section we propose a few specific questions that seem worthy of study, highlighting both smaller problems to start with and possible extensions for further work. For the final thesis, we intend to address at least one question in each of the three categories.

1 Geometric Theory

The geometric theory of lattices in Euclidean space is probably the oldest branch of the theory of quadratic forms. It goes back at least as far as Thomas Harriot, mathematical assistant to Sir Walter Raleigh, who was tasked with determining the best way to stack cannonballs on the deck of a ship [25]. Harriot communicated with Kepler concerning this question, leading Kepler to state what we now call Kepler’s conjecture, asserting that the common packing of cannonballs (or oranges, etc.) is in fact optimal [30].

Kepler’s conjecture remained a conjecture for almost four centuries, until a heavily computational proof was presented by Hales [24]. (This work has the distinction of having been published in the Annals of Mathematics despite the referees only being “99% certain” of its correctness, due to the overwhelming computation involved.) More generally, one can ask the corresponding question in any dimension $n$: in Euclidean space $\mathbb{R}^n$, what is the optimal density$^1$ of a sphere packing (i.e. a configuration of congruent spheres which intersect only at points of tangency)?

To date, this question has only been resolved in dimensions $n \leq 3$ (with the $n = 3$ case due to Hales)$^2$. However, if we limit our attention to lattice packings, i.e. sphere packings where the centers of the spheres form a lattice in $\mathbb{R}^n$, then the answer is known in six more dimensions. The optimal density of a lattice packing is called the lattice packing constant. This constant is equivalent to (but not equal to) the Hermite constant, a quantity also of importance in the geometry of numbers and in the arithmetic theory of lattices.

Definition. Given a lattice $\Lambda \subset \mathbb{R}^n$, its norm $||\Lambda||$ is the smallest (squared) norm of a nonzero element of $\Lambda$. Its determinant $\det(\Lambda)$ is the determinant of any Gram matrix associated to $\Lambda$. The Hermite constant

\footnote{There is some technicality involved in making this question well-defined, as arbitrary configurations need not have a density.}

\footnote{However, in dimensions $n = 8$ and $n = 24$, the optimal packing constant is known up to a very small numerical error factor [11].}
\( \gamma_n \) in dimension \( n \) is
\[
\gamma_n = \sup_{\Lambda \subset \mathbb{R}^n} \frac{||\Lambda||}{\det(\Lambda)^{1/n}},
\]
where the supremum is taken over all (full-dimensional) lattices in \( \mathbb{R}^n \).

**Remark.** It is a theorem that the supremum in the definition of \( \gamma_n \) is always achieved. Using modern machinery, this result is understood as an immediate corollary of Mahler’s compactness theorem [39].

Figure 1 shows the dimensions in which we provably know the Hermite constant.

<table>
<thead>
<tr>
<th>Dimension ( n )</th>
<th>Proven by</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(trivial)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Lagrange</td>
<td>1773</td>
</tr>
<tr>
<td>3</td>
<td>Gauss</td>
<td>1840</td>
</tr>
<tr>
<td>4</td>
<td>Korkin and Zolotarev</td>
<td>1877</td>
</tr>
<tr>
<td>5</td>
<td>Korkin and Zolotarev</td>
<td>1877</td>
</tr>
<tr>
<td>6</td>
<td>Blichfeldt</td>
<td>1934</td>
</tr>
<tr>
<td>7</td>
<td>Blichfeldt</td>
<td>1934</td>
</tr>
<tr>
<td>8</td>
<td>Blichfeldt</td>
<td>1934</td>
</tr>
<tr>
<td>24</td>
<td>Cohn and Kumar</td>
<td>2004</td>
</tr>
</tbody>
</table>

Figure 1: Dimensions in which the Hermite constant \( \gamma_n \) is known [57, pg. 29].

The technique used by Blichfeldt to determine the Hermite constant for \( n \leq 8 \) involves studying the domain of Korkin-Zolotarev reduced forms [4].

**Definition.** Given a quadratic form \( q(x_1, \ldots, x_n) \), its **Lagrange expansion** is the unique decomposition
\[
q(x_1, \ldots, x_n) = \sum_{i=1}^{n} A_i \left( x_i - \sum_{j=i+1}^{n} \alpha_{ij} x_j \right)^2.
\]

The coefficients \( A_i \) are the **outer coefficients** of \( q \), and the \( \alpha_{ij} \) are the **inner coefficients**. The (positive definite) quadratic form \( q \) is **Korkin-Zolotarev reduced** if \( |\alpha_{ij}| \leq 1/2 \) for all \( i, j \); \( ||q|| = A_1 \), i.e. the first basis vector is shortest; and \( \sum_{i=2}^{n} A_i (x_i - \sum_{j>i} \alpha_{ij} x_j)^2 \) is (recursively) Korkin-Zolotarev reduced.

**Remark.** Those familiar with matrix decompositions will recognize the Lagrange expansion as being closely related to the Cholesky decomposition of the Gram matrix.

**Remark.** As stated, Korkin-Zolotarev reduction is not actually a reduction in the strict sense of providing a fundamental domain for
\[
O^+ (\mathbb{R}^n) \backslash \text{SL}_n (\mathbb{R}) / \text{SL}_n (\mathbb{Z}).
\]
To remedy this, the usual definition imposes the additional condition that \( \alpha_{i,i+1} \geq 0 \) for \( i = 1, \ldots, n-1 \). However, this is unimportant for our purposes.

**Remark.** As defined, the first condition for Korkin-Zolotarev reduction involves infinitely many inequalities
\[
q(1, 0, \ldots, 0) \leq q(v) \quad , \quad v \in \mathbb{Z}^n \backslash \{0\}.
\]
However, it is known that there is a finite set of \( v \in \mathbb{Z}^n \) which suffices [46]. In fact, there is a unique minimal set, called the “Novikova set” in [48].

The main advantage of working with the Korkin-Zolotarev reduction domain is that the two quantities relevant for the Hermite constant, the norm of a lattice and its determinant, are expressible in terms of the \( n \) outer coefficients only. Thus, if we can prove inequalities valid for the outer coefficients, then this translates into inequalities concerning the Hermite constant. An additional advantage of this reduction domain is that it is recursive in nature, so inequalities valid in dimension \( m \) are also valid for any contiguous block of \( m \) outer coefficients in dimension \( n \geq m \). Hence there is some hope of bounding the Hermite constant in large dimensions by working only in smaller dimensions.

Blichfeldt was able to implement this program to give sharp bounds on \( \gamma_n \) for \( n \leq 8 \). However,
Blichfeldt’s proof is quite difficult . . . no simple proof of the optimality of $E_7$ is known, nor has Blichfeldt’s work been extended to higher dimensions. [14, pg. 12]

One interesting approach to this problem has recently been proposed by Pendavingh and van Zwam in [48]. Their idea is as follows. In terms of the outer coefficients of a quadratic form, if we fix $A_1 = 1$, then determining the Hermite constant amounts to minimizing the determinant, $A_1 \cdots A_n$. This is a concave function, so its minimum over a domain is unchanged if we take the convex hull. This observation shows that linear inequalities in the space of outer coefficients suffice to determine the Hermite constant. Building on this, Pendavingh and van Zwam treat the problem of determining linear inequalities as an optimization problem in mathematical programming, which they numerically solve using a branch-and-bound procedure applied to a semidefinite relaxation. Proceeding in this way, they obtain three new inequalities in the space of outer coefficients.

**Theorem** ([48, Theorems 6.1 and 6.2]). The following inequalities hold in the space of outer coefficients of Korkin-Zolotarev reduced quadratic forms
\begin{align*}
-25A_1 - 36A_2 + 48A_3 + 40A_4 &\geq -7 \cdot 10^{-6}A_4; \\
-5A_1 + 2A_2 + 8A_3 &\geq -3 \cdot 10^{-4}A_3; \\
-4A_1 - 3A_2 + 4A_4 + 8A_5 &\geq -5 \cdot 10^{-5}A_5.
\end{align*}

Pendavingh and van Zwam conjecture, of course, that the above inequalities remain valid with zero on the right-hand side. If this is true for just the first inequality, then the convex hull of $(A_1, A_2, A_3, A_4)$ is fully determined. If this is true for the other two inequalities, then these techniques determine the Hermite constant up to dimension 8, reproducing Blichfeldt’s result in a more programmatic fashion.

We are now in a position to state our proposed original work. The first problem is obvious in light of our last remark.

**Problem 1.** Prove the conjectured exact analogues of the inequalities in [48].

We believe Problem 1 to be tractable. The branch-and-bound semidefinite programming framework in [48] is unable to prove exact inequalities, due to a fundamental lack of exactness in the semidefinite relaxation. However, that framework does allow us to prove the inequality away from small neighborhoods around points where the inequality is strict. In these small neighborhoods, we propose to use bounds on the underlying (nonconvex) quadratic program to prove that the identified points of equality are in fact local optima. If we obtain concrete bounds on the radius in which these points are optima, and extend the semidefinite bounds to cover the complement of these small balls, then we will have a proof of exactness. This approach also has the collateral benefit of identifying all of the points at which equality is achieved. These points are lattices which are vertices in the space of outer coefficients, and it is reasonable to think that at least some of them will be geometrically interesting.

This same semidefinite framework is applied in [48] to determining Novikova sets, the finite sets that determine Korkin-Zolotarev reduction. The inherent inexactness of this procedure prevents the determination of whether some vectors are necessary or not [66]. Thus, if Problem 1 is successful, a natural follow-up would be the following.

**Problem 2.** Determine the Novikova set in dimension 5 (and higher!).

A more ambitious problem is the following.

**Problem 3.** Prove more inequalities in the space of outer coefficients of Korkin-Zolotarev reduced forms. Specifically, find enough inequalities to determine $\gamma_9$ and $\gamma_{10}$.

The inequalities known so far, even the conjectured exact forms, only give the bound $\gamma_9 \leq 1024$. This is worse than the Cohn-Elkies bound for sphere packings, and it is twice the current lower bound (which we expect to the sharp) [48]. Using points which are known to be valid outer coefficients, we have shown that five-dimensional inequalities do not suffice to determine $\gamma_9$; thus higher dimensional inequalities are necessary to satisfy Problem 3.

Besides just being the first unresolved cases, the dimensions $n = 9$ and $n = 10$ exhibit very interesting behavior. In nine dimensions, the best known packing constant is achieved by a “floating” family. That is, there is a one-parameter of continuous varying packings, all of which achieve the same packing constant.
However, despite all of this “room”, it is not possible to locally improve the packing density! Conway and Sloane refer to this as the “fluid diamond packing” [15]. However, we do not currently know that this is the optimal packing density, even amongst lattice packings. Establishing this fact would put the curiosity of floating packings on much firmer ground.

In ten dimensions we see a new phenomenon: the best known packing is non-lattice! This best packing, called the Best packing, is obtained by applying Conway and Sloane’s “Construction A” [14, pg. 137] to Marc Best’s nonlinear binary code of length 10 [3]. The associated packing constant is better (by about 8%) than the best known lattice packing. If the best known lattice packing is in fact optimal, then determining $\gamma_{10}$ (or at least upper bounding it sufficiently well) would prove that, in at least one dimension, non-lattice packings beat lattice packings. This would be a very interesting result, as we expect this phenomenon to hold in most (all?) sufficiently large dimensions, but there are no proven separation results! In fact, even our current bounds on the packing constant in high dimensions fail to distinguish lattice and non-lattice packings: the upper bounds apply to all packings, while the lower bounds are obtained from lattices. We find the idea of establishing some proven difference between lattice and non-lattice packings to be quite compelling. (Incidentally, we note that there is some expectation that periodic, if not lattice, packings are always optimal. This conjecture is attributed to Zassenhaus [23].)

If Problem 3 proves too difficult, one alternative goal would be to simply determine the convex hull of the outer coefficients in dimensions higher than 4. (Recall that Problem 1, if completed, would establish the convex hull in dimension 4.) It is known that, even assuming the exact form of the two five-dimensional inequalities in [48], the five-dimensional convex hull is still not fully determined. Understanding this domain, and in particular studying its vertices, might give valuable insight into the geometry of Korkin-Zolotarev reduced forms.

Another possible goal in this vein would be to more directly extend the work of Blichfeldt. While this seems quite difficult (as indicated in the quote above), Watson [70] and Vetchinkin [69] have both reviewed, verified, and in some ways extended Blichfeldt’s work. The idea here, at the highest level, is that inequalities (perhaps nonlinear) might be provable by more theoretical considerations than the brute force approach of mathematical programming. However, one difficulty with this (and with Problem 3) is that in some sense the extremely high packing density of the $E_8$ root lattice translates into relative ease of bounding the packing constant in dimensions $n \leq 8$. The packing constant in higher dimensions, rather than being an “easy” consequence of $E_8$, appears to be more intricately determined.

At a high level, the basic formula we have been following so far is to choose a method of reduction for quadratic forms, prove inequalities valid for that reduction domain, and deduce bounds on quantities of interest (i.e. the Hermite constant). In principle there may be another reduction method which is better suited for this than that of Korkin and Zolotarev. For instance, there is the Venkov family of reduction domains ([68]; c.f. [64], [56], [65]). This has the advantage of being determined by finitely many linear inequalities in the entries of the Gram matrix. We have not explored Venkov reduction in any detail; at first glance it appears ill-suited to our needs, as a Venkov-reduced quadratic form may not have the first basis vector being the shortest, but perhaps there is some application in computational geometry to be found.

In a list of reduction theories, we would be remiss if we did not discuss Minkowski reduction. This well-known sense of reduction is defined “greedily”, as follows.

**Definition.** A quadratic form is **Minkowski reduced** if, for each $i = 1, 2, \ldots, n$, the $i$th basis vector $e_i$ is a shortest vector amongst all $v \in \mathbb{Z}$ such that $\{e_1, \ldots, e_{i-1}, v\}$ may be extended to a basis of $\mathbb{Z}^n$.

**Remark.** As with Korkin-Zolotarev reduction, a sign condition is needed to make this a “true” notion of reduction.

As with Korkin-Zolotarev reduction, it is known that finitely many inequalities suffice to determine the domain of Minkowski reduction. Also, as with Venkov reduction, these inequalities are linear in the Gram matrix. Thus, Minkowski reduction has two distinct advantages for our purposes: the lattice norm is always achieved by the first basis vector, and the domain is a polytope! A result of Minkowski tells us that the determinant is (essentially) a concave function of the Gram matrix [42], so to determine the Hermite constant it suffices to enumerate the extreme rays of the polytope. Unfortunately, as the dimension grows, the number of extreme rays grows quite quickly, making this approach infeasible. (In dimension 5, there are already 15971 extreme rays, and the number grows quickly from there [57, pg. 41].)
Despite this, there are still interesting questions regarding the domain of Minkowski reduced quadratic forms. For instance, while it is infeasible to enumerate all of the extreme rays, it nonetheless may be possible to enumerate all of the relevant rays, by a neighbor search which does not follow directions in which the determinant is bounded below by the current optimum. We do not know currently know whether this approach is tractable, but it is easy enough to try.

**Problem 4. Implement this strategy for searching through the relevant extreme rays of Minkowski reduction.**

Setting aside the issue of determining Hermite constants, the first, most obvious, problem regarding Minkowski reduction is determining the defining vectors, i.e. the finite set of vectors which define the polytope. These vectors are known up to dimension 7, with \( n \leq 6 \) due to Minkowski [43] and \( n = 7 \) due to Tammela [63]. (Some redundancies in their lists are noted in [57, pg. 41].) Thus, the following problem naturally presents itself.

**Problem 5. Determine the defining vectors for Minkowski reduction in dimension 8.**

Thanks to the polytopal nature of Minkowski reduction, testing if a given vector is necessary amounts to optimizing a single linear program. Thus one could immediately approach Problem 4 computationally. The obstruction is in producing a sufficiently small sufficient list of vectors. The typical approach, due to van der Waerden, involves bounding the diagonal entries \( q_{ii} \) of a Minkowski reduced Gram matrix in terms of the successive minima \( \lambda_i \). The current bound, unimproved since [65], is

\[
q_{ii} \leq \Delta_i \lambda_i, \quad \Delta_i = \max \left\{ 1, \left( \frac{5}{4} \right)^{i-4} \right\}.
\]

A better bound would translate into a smaller set of vectors which must be tested. It is conjectured in [57, pg. 42] that the factor \( \Delta_i \) may be improved to \( \max\{1, i/4\} \). (This improvement, if valid, is optimal.) This would be a substantial result, and so we explicitly state it as a problem.

**Problem 6. Prove (or disprove) the conjecture \( q_{ii} \leq \max\{1, i/4\} \cdot \lambda_i \).**

As a final note regarding Minkowski reduction, we observe that there may be interesting results obtainable by applying the techniques of eigenvalue optimization. The motivation for this statement is the following result, which generalizes Minkowski’s concavity theorem.

**Theorem ([38, Theorem 3.1]).** Symmetric convex functions of the eigenvalues of a symmetric matrix are convex functions of the matrix entries.

Thus, using convex programming, we may minimize any convex symmetric function of the eigenvalues of the Gram matrix. The class of convex symmetric functions includes the linear functionals \((a_1, \ldots, a_n) \cdot (\lambda_1, \ldots, \lambda_n)\), where \( \lambda_1 \leq \cdots \leq \lambda_n \) are the eigenvalues in ascending order and the coefficients \( a_i \) are also in ascending order.

Unfortunately, optimization over the class of convex symmetric functions cannot give us enough information to minimize the product of the eigenvalues (i.e. the determinant). We have been unable to overcome this difficulty, and so we do not state an explicit problem along these lines.

Finally, this discussion of the computational geometry of quadratic forms would be incomplete without mention of Voronoi’s algorithm. This is an algorithm to enumerate all of the perfect forms of a given dimension; it is known that the Hermite constant must be achieved by a perfect form (Theorem 3.4.6 in [40]), so this enumeration is sufficient for the purpose of determining \( \gamma_n \). (It is also known that all perfect forms are extreme rays for Minkowski reduction, so that the enumeration task assumed by Voronoi’s algorithm is easier than Minkowski’s.) Voronoi’s algorithm has been successfully run up to dimension \( n = 8 \), where there are exactly 10916 perfect forms [59]. This task required significant computing time, even after crucial speedups (e.g. exploitation of symmetry), and so it is not expected to be feasible in the near-term to extend this to \( n = 9 \).

However, there are other opportunities to apply Voronoi’s algorithm. For instance, there is a variant for periodic packings [57, pg. 77]. Another direction is to study lattices with additional structure, e.g. lattices over the Gaussian, Eisenstein, or Hurwitz integers. (More generally, one might study lattices over
any (maximal) order.) The definition of the Hermite constant naturally extends to these settings, and its determination is a fundamental problem. This question has been considered before, e.g. in [67], but we believe that there still remains low-hanging computational fruit. Thus, we close this section with the following open-ended problem.

**Problem 7.** Compute the Hermite constant for lattices over interesting maximal orders in new dimensions.

### 2 Arithmetic Theory

The arithmetic theory of quadratic forms has its roots at least as far back as Diophantus, who appears to have been aware, at least empirically, of the fact that every positive integer is the sum of at most four squares [27, pg. 110]. Sums of squares were later studied by Fermat, and eventually the empirical conjectures were proven by Lagrange and Legendre; these proofs may be called the beginning of the rigorous arithmetic theory. Nowadays the theory has grown far beyond these results, but even so there remain open questions about integral quadratic forms. In this proposal we focus mainly on the following problem.

**Problem 8.** Give an efficient algorithm for computing an explicit integral equivalence between (equivalent) integral quadratic forms.

The status of Problem 8 depends heavily on whether definite or indefinite forms are under consideration. For definite forms, there exist algorithms, and even implementations, which are satisfactory in many cases. For instance, the computer algebra systems MAGMA and SAGE both support such computations “out of the box”. (These algorithms are not efficient in the sense of running in polynomial time. To the author’s knowledge, there is no expectation of a polynomial time equivalence algorithm for definite forms, but neither is there a proof of hardness. This could be an avenue for future work, or at least future literature search.)

When the forms under consideration are binary (but possibly indefinite), there is again a well-developed theory; see e.g. [7, Ch. 14]. This theory does not appear to have been implemented in any widely-available computer code, but there is no great barrier to doing so.

The remaining case is that of indefinite forms in at least three variables. In such cases we have the theory of the spinor genus, which at this point is classical. The basic result of this theory is the following.

**Theorem** ([7, Theorem 11.1.4]). Two indefinite forms of rank at least 3 are in the same spinor genus iff they are integrally equivalent.

We will not define the spinor genus here, and instead just note that it is an invariant of an integral quadratic form which is essentially local. In fact, quite often a genus of quadratic forms consists of a single spinor genus [7, Theorem 11.1.3]. Testing quadratic forms for local equivalence (i.e. for being in the same genus) is easy, and code for this may be found in the usual packages. Thus, in many cases, it is easy to determine if two quadratic forms are integrally equivalent. However, this leaves open the specific question we stated in Problem 8, that of computing an explicit equivalence!

There does not seem to have been much progress on this problem. The standard references have little to offer; they just note that we could simply search over all matrices until we find one that works:

From a purely logical point of view the second question [computing an equivalence] is a mere rhetorical flourish once we have a decision procedure for the first question [determining equivalence]. [7, pg. 327]

There do not seem to be good algorithms for the remaining problems. ... Logically there is no difficulty: the proofs of the theorems on genus and spinor genus ... are at bottom computationally effective. For example ... if two forms are known to be in the same genus, we can in principle search through all rational matrices until a rational equivalence of denominator $r$ prime to $2d$ is found. [14, pg. 404]

Needless to say, this logical “solution” is not sufficient for real-world computations.

As the second quote indicates, the proof of the spinor genus theorem is in principle effective. Thus one approach to Problem 8 is to follow the standard proof, implementing each step. The crucial result needed for the proof of the main spinor genus theorem is the following:
**Theorem** ([7, Theorem 9.1.5]). Let \( q \) be a regular indefinite integral quadratic form in \( n \geq 4 \) variables, and let \( a \neq 0 \) be an integer which is represented by \( q \) everywhere locally. Then \( q \) integrally represents \( a \).

Again, this theorem has a proof which is potentially effective. The problem is that the integers involved become prohibitively large. If one could improve the implementation aspects of the proof, this might translate into an effective algorithmic form of the spinor genus theorem. This motivates the following subproblem of Problem 8.

**Problem 9.** Given an indefinite quadratic form in at least four variables and an integer which it represents, give an algorithm to efficiently compute an explicit representation.

Of course, in proposing a successful resolution to Problem 8 (resp., Problem 9), we are implicitly assuming that there exists a “small” (polynomial-sized, say) equivalence (resp., representation). Existence of small solutions is another aspect of the problem which can be studied. This has received attention; see e.g. [5] for the representation problem and [20] for a partial result towards the equivalence problem. In particular, we now know that (generally) there do exist polynomial-sized solutions. The known bounds are quite large, however, so it seems likely that improvements are possible.

Problem 8, and even Problem 9, is rather ambitious, so we also propose a simpler problem to start with. This is the following:

**Problem 10.** Give an efficient algorithm to determine if two quadratic forms are in the same spinor genus.

By the spinor genus theorem, for indefinite forms in at least three variables, Problem 10 is the same as deciding integral equivalence. It is the natural precursor to Problem 8; it is also the problem that our discussion above seems to indicate is solved. The situation, though, is more complicated than that. While the theory exists for determining spinor equivalence, it does not seem to have been implemented in any public computer package.\(^3\) This gap between algorithm and implementation is what Problem 10 proposes to close.

Following standard theory, one tests for spinor equivalence of two quadratic forms \( q_1 \) and \( q_2 \) (of common determinant \( d \)) using the following steps:

- Realize \( q_1 \) and \( q_2 \) as lattices in the same \( \mathbb{Q} \)-quadratic space which are \( p \)-adically equivalent for all \( p \) dividing \( 2d \):
  - Compute a rational equivalence between \( q_1 \) and \( q_2 \)
  - Use the local equivalence of \( q_1 \) and \( q_2 \) to make the rational equivalence \( p \)-adically integral for all \( p \mid 2d \)
- Compute the intersection index of these two lattices (the index of their intersection in each)
- Test if this intersection index is in the “spinor kernel”

Following [7, Theorem 9.1.4], the second part of the first step is an easy, local procedure. The second step is simple, and the last step is just a matter of computing the spinor kernel: enough details to do this are given in [7], and it is made explicit in [14]. Thus the only obstruction to implementing this procedure is the first step, computing the rational equivalence. In recognition of this, we explicitly state the following problem.

**Problem 11.** Given two (equivalent) rational quadratic forms, give an efficient algorithm to compute an explicit rational equivalence.

Problem 11 may be restated as “making the Weak Hasse Principle effective.” In [14], Conway and Sloane give a proof of the Weak Hasse Principle which is already effective, and following their approach is what we propose to do for Problem 11. The details of this are presented in the note [44]. That note describes the implementation, but does not go into any depth concerning the complexity of the algorithm; additional work along these lines is warranted. A promising reference for the theoretical complexity of these problems is [26].

\(^3\)Again, the situation is different for definite forms. In this case, there does exist available computer code, e.g. in Magma.
An interesting subproblem which arose in our implementation is that of expressing a prime (or twice a prime, in the case \( p \equiv 7 \mod 8 \)) as a sum of three squares. This problem appears to have evaded a satisfactory answer in the literature. The related problems of representations by two and four squares have excellent algorithms: for two squares one may essentially use the Euclidean algorithm in \( \mathbb{Z} \sqrt{-1} \), and there are related algorithms for four squares (see e.g. [6]). However, the three squares case lacks the multiplicative structure of its two and four square cousins, which presumably is responsible for our lack of a simple algorithm. We discuss the situation and give a new algorithm in [44], but because of the fundamental nature of the question we repeat it here as an open problem.

**Problem 12.** Give an efficient algorithm which, given a positive integer \( n \) not of the form \( 4^k(8k+7) \), returns an expression of \( n \) as the sum of three squares.

In closing this section, we note that there are other approaches to Problem 11. For instance, if one had an algorithm for computing isotropic vectors of rational quadratic forms, this could be harnessed to give a solution to Problem 11. This corresponds to making the Strong Hasse Principle effective, instead of just the Weak Principle.

This problem has received some attention in the literature. There appears to be a fundamental difference between finding isotropic vectors for ternary forms and finding isotropic vectors for quaternary forms. (One can handle higher dimensional forms by reduction to these cases.) For ternary forms, two works of particular note are that of Cremona and Rusin [17] and that of Simon [61]. Cremona and Rusin first diagonalize the quadratic form and then reduce its coefficients; this approach appears to offer the best algorithm for quadratic forms which are already diagonal. An implementation may be found in the mwrank package [16], which is wrapped for use in SAGE. Simon’s approach works with a nondiagonal Gram matrix directly, and iteratively reduces the determinant one prime at a time. This appears to offer the best algorithm for nondiagonal forms, and a PARI implementation is available from the author. Simon also discusses algorithms for the quaternary (and higher) case in [60]. (A recent student of his has given some improvements on this, but they appear to be unpublished as yet.) Implementations of Simon’s algorithms, due to Mark Watkins, may be found in recent versions of Magma.

### 3 Algebraic Theory

The algebraic theory of quadratic forms, as understood today, began with Witt’s work in [71] and was revitalized by Pfister’s use of multiplicative forms in [53]. The algebraic theory is primarily concerned with the behavior of quadratic forms over fields, but in the last few decades there has also been work on the quadratic invariants of general commutative rings. The bulk of this section consists of problems of this type. We thank David Leep for suggesting Problems 13, 16, and 17.

Two of the most fundamental quadratic invariants of a ring are the level and the Pythagoras number.

**Definition.** For a commutative ring \( R \) (with identity), the level \( s(R) \) is the smallest number \( n \) such that there exist \( x_1, \ldots, x_n \in R \) with \( x_1^2 + \cdots + x_n^2 = -1 \). (If there is no sum-of-squares decomposition for \(-1\) in \( R \), then \( s(R) = \infty \) and \( R \) is said to be semireal.) The Pythagoras number \( P(R) \) is the smallest number \( n \) such that any element of \( R \) which is a sum of squares is already a sum of at most \( n \) squares.

We will return to the level below, but for now we focus on the Pythagoras number. Consider polynomial algebras over the real numbers. Clearly \( P(\mathbb{R}) = 1 \), and an easy argument shows that \( P(\mathbb{R}[x]) = 2 \). However, it is known that \( P(\mathbb{R}[x, y]) = \infty \) [8, Theorem 4.1]. This means that there is a sequence of bivariate polynomials, all of which are sums of squares, such that the number of squares needed to sum to the given polynomials grows to infinity. The polynomials implicit in the proof of this theorem have quite large degree; we are naturally led to ask if this is necessary. This motivates the following problem.

**Definition.** Given a subset \( S \) of a ring \( R \), the (restricted) Pythagoras number of \( S \) is the smallest number \( n \) such that any element of \( S \) which is a sum of squares in \( R \) is already a sum of at most \( n \) squares in \( R \).

**Problem 13.** For \( n, m \geq 1 \), let \( \mathbb{R}[x_1, \ldots, x_n]_m \) be the subspace consisting of homogeneous polynomials of degree \( m \). Determine \( P(n, m) := P(\mathbb{R}[x_1, \ldots, x_n]_m) \).
This problem is studied in some detail in [9]; we have chosen our notation to follow theirs. Note in particular that we are restricting to homogeneous polynomials, but dehomogenizing shows that our problem is equivalent to the inhomogeneous case in one fewer variable. Thus our motivating study of \( \mathbb{R}[x, y] \) corresponds to the \( n = 3 \) case of Problem 13.

A priori it is not obvious that \( P(n, m) \) is always finite, but there is an easy argument showing that \( P(n, m) \) is upper bounded by \( \left( \frac{m/2+n-1}{n-1} \right) \), the number of monomials of degree \( m/2 \). More refined arguments lead to the following result.

**Theorem** ([9, Main Theorem]). For any \( n, m \geq 1 \),

\[
\frac{2e + 1 - \sqrt{(2e + 1)^2 - 8a}}{2} \leq P(n, m) \leq \frac{\sqrt{1 + 8a} - 1}{2}
\]

where \( a = \binom{n+m-1}{n-1} \) and \( e = \binom{n+m-2}{n-1} \).

The Main Theorem in [9] actually gives a family of bounds determined by a choice of “cage”; by careful selection of which cage to use, an improved lower bound is obtained in [36]. However, as a general upper bound, the above theorem does not appear to be have been beaten.

In the special case of \( m = 2 \), we know from the theory of quadratic forms that \( P(n, 2) = n \). The bounds of the theorem above suffice to determine this. In the special case of \( n = 2 \), where we have \( P(2, m) = 2 \), the lower bound is correct but the upper bound is not; it grows as \( m^{1/2} \). More generally, these bounds suffice to determine the asymptotic order of \( P(n, m) \) when \( m \) is fixed and \( n \) grows (it is \( \eta^{m/2} \)), but they do not determine the order when \( n \) is fixed and \( m \) grows. This is the limit in which we would like improved bounds.

It is worth noting that there is one (and only one) additional case in which we know \( P(n, m) \): this is \( P(3, 4) \), which equals 3 by a result of Hilbert [29]. (This case is probably better known as being one of two exceptional cases where every positive semidefinite form is a sum of squares. What is less well known is that Hilbert showed that 3 squares suffice.) This result still does not have a simple proof; Hilbert used an argument in algebraic geometry (see [54] and [62] for expositions of this proof), and there is a newer algebro-geometric proof available [49], but work on an elementary approach is still underway [51]. (We should remark that there is an elementary proof showing that positive semidefinite ternary quartic forms are sums of at most five squares [10].) Our bounds on \( P(n, m) \) do not suffice to determine \( P(3, 4) \); the upper bound obtained from the above theorem is 5.

The \( n = 3 \) case is worth more attention. In this case, the lower bound in the above theorem converges to 4 as \( m \to \infty \), while the upper bound is exactly \( m + 1 \). The improved lower bounds in [36] are still bounded by 4, so there is no substantial change there. Because \( P(\mathbb{R}[x, y]) = \infty \), we know that \( \lim \sup P(3, m) = \infty \), and so the lower bounds are certainly not sharp. As for the upper bound, unpublished work of David Leep (using quadratic form theory over \( \mathbb{R}(x) \)) shows that \( P(3, m) \leq \frac{m}{2} + 2 \) [35]. However, even this improved bound is not correct in the case \( P(3, 4) \). Setting aside the goal of exactly determining \( P(3, m) \), it would still be of interest to know its asymptotics, and so we pose the following subproblem of Problem 13.

**Problem 14.** Determine the asymptotic growth of \( P(3, m) \) as \( m \to \infty \).

The main approach to proving upper bounds for \( P(n, m) \) seems to be the so-called “Gram matrix method”. This consists of thinking of the real vector space of degree-(\( m/2 \)) forms as having coordinates indexed by the monomials, so that a polynomial of degree \( m/2 \) is represented by a vector of length \( \binom{m/2+n-1}{n-1} \) and its square is represented by the outer product of that vector with itself. Thus sum-of-square polynomials of degree \( m \) are represented as symmetric, positive semidefinite matrices of dimension \( \binom{m/2+n-1}{n-1} \). Of course, generally there is more than one matrix representation for a given polynomial; the family of such representations forms an affine subspace in the space of symmetric matrices. To determine the length of a polynomial (i.e. the minimum number of squares needed to express it), we just need to find the minimum rank of a positive semidefinite matrix in the given affine subspace.

Using the following easy (but independently obtained) result, we were able to reproduce the upper bound in [9, Main Theorem].

**Theorem.** Let \( S_n \) be the space of symmetric \( n \times n \) real matrices, and let \( S_n^+ \subset S_n \) be the cone of positive semidefinite matrices. Let \( A \subset S_n \) be an affine subspace. If \( A \cap S_n^+ \neq \emptyset \) and \( \text{codim} A < \binom{r+1}{2} \), then \( A \cap S_n^+ \) contains a matrix of rank less than \( r \).
We do not know if this theorem is best possible, and so we propose the following question for study.

**Problem 15.** Given integers $1 \leq r \leq n$, determine the largest integer $d$ such that the following holds: any affine subspace in $S_n$ which intersects $S_n^+$ and has codimension less than $d$ must contain a positive semidefinite matrix of rank less than $r$.

This question fits into a body of problems studying spaces of matrices with large rank. It is interesting to note that the symmetric, semidefinite condition significantly changes the character of this particular question; while $\dim A > 0$ is enough to guarantee a singular matrix in this framework, there are affine subspaces of $\text{Mat}_n(\mathbb{R})$ contained in $\text{GL}_n(\mathbb{R})$ with $\dim A = n$. (Even more interestingly, there is a classification of such spaces over a general field which depends only on the quadratic properties of that field.) This theory is relatively new; a reference is [47]. The theory of linear subspaces of invertible matrices is older; for this the primary reference seems to be [1].

Setting aside the goal of general bounds, given a specific sum-of-squares form, one may interpret the Gram matrix method as giving a computational specification of the length of that form. While the problem of rank minimization in an affine space is NP hard [55], one might still hope to compute some small examples. One approach to this problem is to apply general solvers for systems of real polynomial equations to the system of $r \times r$ minors (which all must vanish for the matrix to have rank less than $r$). This idea, using the computer package REALSOLVING, is presented in [50]. (In that paper the authors demonstrate an application in the $n = 3$, $m = 4$ case.) A newer version of this package, RAGLIB, may be better suited for future efforts in this direction [21]. We think that this idea has promise for experimentally determining new lower bounds in some simple cases, which may help to resolve questions such as Problem 14.

Our next question concerns the Pythagoras number of power series rings. The most general statement would be to determine $P(R[[x]])$, preferably in terms of quadratic invariants of the base ring $R$. This question is likely far too ambitious, so we instead start with the following limited case.

**Problem 16.** Let $R$ be a real principal ideal domain. Determine $P(R[[x]])$ (in terms of invariants of $R$).

The motivation for Problem 16 is the case $R = \mathbb{Z}$. In this case we have $P(\mathbb{Z}[[x]]) = 5$, which seems to have been first proven by Liese [37]. This work is not readily available; an independent proof has been provided by Leep [34] and reproduced by the author. This proof relies on the following modified invariant of $\mathbb{Z}$.

**Definition.** For a ring $R$, the unimodular Pythagoras number of $R$ is the smallest number $n$ such that any nonzero sum of squares $x$ in $R$ is a sum of $n$ squares, $x = a_1^2 + \cdots + a_n^2$, such that $(a_1, \ldots, a_n)$ is the unit ideal in $R$.

The unimodular Pythagoras number of $\mathbb{Z}$ is readily seen to be 5, by simply taking 1 as one of the summands and then decomposing the remainder using at most 4 squares. With this in hand, the argument showing that $P(\mathbb{Z}[[x]]) = 5$ is straightforward. Generally speaking, the unimodular Pythagoras number of $R$, when finite, determines $P(R[[x]])$. However, we have very examples few of this besides $\mathbb{Z}$. One of the few is $\mathbb{Z}[\frac{1+\sqrt{5}}{2}]$, in which every totally positive integer is a sum of squares of units [58]. This lack of examples naturally leads to the following problem, which we state in a restricted setting in the hope of making it more tractable.

**Problem 17.** Compute the unimodular Pythagoras number of the rings of integers in real quadratic number fields.

It is not always possible to compute $P(R[[x]])$ via the unimodular Pythagoras number of $R$, though. For instance, it is possible for the unimodular Pythagoras number to be infinite despite the Pythagoras number being finite. A simple example is provided by $R = \mathbb{R}[[t]]$; $t^2$ is a square but it is not a sum of unimodular squares. (In principle it is also possible to have a ring with finite Pythagoras number in which every sum of squares is a sum of unimodular squares, but the unimodular Pythagoras number is infinite. We do not know of any examples of such a ring, but we suspect they exist.) In the case $R = \mathbb{R}[[t]]$ it is known that $P(R) = 1$ [8, Theorem 5.1(1)]; $P(R[[x]]) = 2$ [8, Corollary 5.14]; and $P(R[[x]][[y]]) = \infty$ [8, Corollary 6.7]. The most interesting of these results is $P(R[[x]]) = 2$, and it is obtained using the Weierstrass Preparation...
Theorem. This depends rather crucially on the structure of this particular ring, so we should not expect it to generalize to a full solution to Problem 16.

Our next family of problems concerns the quadratic invariants of generic algebras. The prototypical example of such a result is the following.

**Definition.** Given a ring $R$ and a positive integer $n$,

$$A_n(R) := R[X_1, \ldots, X_n]/(1 + X_1^2 + \cdots + X_n^2).$$

**Theorem** ([18, pg. 847]). For any $n \geq 1$, $s(A_n(\mathbb{R})) = n$.

This theorem settled the open question of whether there exist rings with any prescribed level. (The corresponding question for fields has a negative answer; Pfister’s work shows that the level of a field is either infinite or a power of two [53,].) The proof in [18] is topological in nature, appealing to the Borsuk-Ulam Theorem from algebraic topology. Using the “homogeneous 2-Nullstellensatz” (see [52, Ch. 4]) it is possible to give an algebraic proof of this theorem (c.f. [31],[2]), but nonetheless all known proofs are decidedly topological in flavor. For more discussion of the interplay between the algebraic and topological approaches to this theorem, see [45].

More generally, Dai and Lam present in [19] a connection between the “level” of a topological space with involution and the algebraic level of an associated “coordinate ring”. (The notion of the “level” of a space had previously been defined in topological circles as the “coindex” [12],[13]. In addition to those sources, some very accessible applications of the Borsuk-Ulam Theorem to coindex computations are given in [41].) Using this topological theory, Dai and Lam obtain the following results.

**Definition.** For a commutative ring $R$ (with identity), the *sublevel* $\sigma(R)$ is the smallest number $n$ such that there exist $x_1, \ldots, x_{n+1} \in R$ which generate the unit ideal and satisfy $x_1^2 + \cdots + x_{n+1}^2 = 0$. (That is, the sublevel is the smallest $n$ such that $(n+1)/1$ is isotropic.)

**Remark.** For any ring $R$, clearly $\sigma(R) \leq s(R)$, as we can just take $x_{n+1} = 1$. If 2 is a unit in $R$ then it is easy to see that $s(R) \leq \sigma(R) + 1$ [19, Proposition 5.5]; moreover, if $s(R) \in \{1, 2, 4, 8\}$, then the bilinear composition laws imply $\sigma(R) = s(R)$ [19, Proposition 5.9]. Thus, for rings $R$ in which 2 is a unit (for instance, in real algebras) we always have $\sigma(R) \in \{s(R) - 1, s(R)\}$, and the first option is only possible if $s(R) \not\in \{1, 2, 4, 8\}$.

**Theorem** ([19, Corollary 5.12]). For any $n \geq 1$, $\sigma(A_n(\mathbb{R})) = n$.

**Theorem** ([19, pp. 417–418]). Given a ring $R$ and a positive integer $n$,

$$B_n(R) := \frac{R[x_1, \ldots, x_{n+1}, y_1, \ldots, y_{n+1}]}{(x_1^2 + \cdots + x_{n+1}^2, 1 - x_1 y_1 - \cdots - x_{n+1} y_{n+1})}.$$

For any $n \geq 1$, $\sigma(B_n(\mathbb{R})) = n$. If $n \in \{1, 3, 7\}$ then $s(B_n(\mathbb{R})) = n$; otherwise $s(B_n(\mathbb{R})) = n + 1$.

The first theorem amounts to the statement that $n(1)$ is anisotropic over $A_n(\mathbb{R})$. This is proven using a stronger form of the Borsuk-Ulam Theorem which, despite being easy to formulate as an algebraic statement, does not have an algebraic proof at present. The second theorem uses much higher-powered topological results involving Stiefel manifolds. Thus, while we propose below to provide algebraic proofs for both, the first is likely more accessible.

**Problem 18.** Give algebraic proofs for the values of $\sigma(A_n(\mathbb{R}))$, $\sigma(B_n(\mathbb{R}))$, and $s(B_n(\mathbb{R}))$.

In addition to results for which we only have algebraic proofs, there are several questions concerning generic rings which remains conjectures. One of the simplest of these is the following.

**Problem 19.** Prove or disprove: for any ring $R$ and positive integer $n$, $s(A_n(R)) = \min\{s(R), n\}$.\footnote{In more detail, the result $s(A_n(\mathbb{R})) = n$ just uses the fact that odd maps on spheres are not null-homotopic. The result $\sigma(A_n(\mathbb{R})) = n$ uses the stronger fact that odd and even maps are not homotopic.}
Problem 19 is called the “Level Conjecture” in [19, pg. 419]. The conjecture is borne out in all known examples, which include two interesting families of rings $R$:

- semireal rings [32, Theorem XIII.4.41]; and
- fields [2].

The argument for semireal rings is a straightforward reduction to the real case, which as previously discussed is proved topologically. The argument for fields is a clever algebraic trick that passes to a 2-field and then applies the homogeneous Nullstellensatz for such fields. The homogeneous Nullstellensatz for real-closed fields, a special case of 2-fields, is essentially the algebraic Borsuk-Ulam Theorem; this is what we use for the algebraic proof of $s(A_n(R)) = n$. However, the theorem for general 2-fields is stronger. For expositions of these arguments, see [45].

Several more open problems are mentioned in [19], including the “Tensor Conjecture” $s(A \otimes_{\mathbb{R}} B) = \min\{s(A), s(B)\}$ (for $\mathbb{R}$-algebras $A, B$). However, we find the Level Conjecture to be the most compelling of these open problems, in part because of the partial progress already made and in part because it seems to live at the interface of algebra and topology.

So far we have been discussing problems involving the quadratic invariants of rings, but, as mentioned in the section introduction, the algebraic theory of quadratic forms is largely concerned with fields. Many active areas of research in this theory concern function fields of quadrics, but these require heavy machinery from algebraic geometry. For this reason, we prefer to stick to problems that are more traditionally algebraic. One source of such problems is the interplay between quadratic forms theory and field theory. For instance, the notion of $C_i$ fields is of great use in computing the $u$-invariants of certain fields.

**Definition.** A field $F$ is $C_i$ if any form over $F$, of degree $d$ in $n > d^i$ variables, has a nontrivial zero in $F$.

**Definition.** The $u$-invariant of a field $F$, $u(F)$, is the largest $n$ such that there exists an anisotropic $F$-quadratic form of rank $n$.

It is immediate from the definitions that a $C_i$ field has $u$-invariant at most $2^i$. In many cases, e.g. finite fields ($C_1$) and iterated Laurent series fields $\mathbb{C}((x_1)) \cdots ((x_i))$ ($C_i$), this turns out to be sharp. In other cases we do not know that fields are $C_i$, or perhaps we know them not to be, but from the perspective of the $u$-invariant they still behave as such. One example of this is $p$-adic fields, e.g. $\mathbb{Q}_p$. These are known to not be $C_2$ (at least for $p = 2$), but in many ways, including the $u$-invariant, they behave like $C_2$ fields [28]. (This even includes the property that $u(\mathbb{Q}_p(x_1, \ldots, x_n)) = 2^{2^{2n}}$, thanks to a remarkable argument of Leep [33].) There are a variety of related open problems, which are nicely summarized in [28].

For this proposal we look instead at a different example of a field which seems to behave like it is $C_2$: $\mathbb{C}((x, y))$. (Note that this is not the same field as $\mathbb{C}((x))(y)$, which is $C_2$ by the standard theory.) It is not known whether this field is $C_2$ [22, pg. 36], but we do have the following result.

**Theorem** ([8, Theorem 5.16]). Let $k$ be an algebraically closed field of characteristic 0 and define $F = k((x, y))$. Then $F$ is “$C_2$ for diagonal forms”, in the sense that any diagonal form $f_1 x_1^2 + \cdots + f_n x_n^2$ over $F$ with $n > d^2$ variables has a nontrivial zero.

The proof of this theorem is a fairly direct application of the Weierstrass Preparation Theorem. It only works when $\text{char} k$ is prime to the degree $d$ of the form, which is why we have stated the result for $\text{char} k = 0$. Unfortunately, this condition is omitted in the original source [8, pg. 69] (and the omission is reprinted in [32, pg. 481]), where the theorem is asserted to hold for all algebraically closed fields $k$. Of course, there is no difficulty in deducing that the $u$-invariant is 4, as both sources have the implicit, running assumption that the characteristic is not 2. However, the field theory question is still wide open. Thus, we close this proposal with the following problem.

**Problem 20.** Let $k$ be an algebraically closed field (of any characteristic). Determine if $k((x, y))$ is a $C_2$ field, or at least a $C_2$ field for diagonal forms.
References


[22] Greenberg, M., Lectures on Forms in Many Variables.

[23] Gruber, P.M., Convex and Discrete Geometry.


13


[32] Lam, T.Y., Introduction to Quadratic Forms over Fields.


[34] Leep, D.B., preprint.


[37] Liese, H., “Quadratsummen in \( \mathbb{Z} \) und \( \mathbb{Z}[[x]] \).” Staatsexamens-Arbeit (1975), unpublished.


[40] Martinet, J., Perfect Lattices in Euclidean Spaces.


[52] Pfister, A., Quadratic Forms with Applications to Algebraic Geometry and Topology.


