A DICHOTOMY PRINCIPLE FOR REPRESENTATIONS OF PRIMES BY QUADRATIC FORMS

M. G. MAHMOUDI

Abstract. The quadratic value theorem provides a dichotomy principle for the representation of an irreducible polynomial p in the ring $A = F[X_1, \cdots, X_n]$ where F is a field, by an anisotropic quadratic form q defined over F: either a scalar multiple of p is multiplicatively generated by the values of q over A, or q is anisotropic over the residue field of A at p. We examine the validity of this principle in wider contexts.

Mathematics Subject Classification: 11E04, 11E25.

1. Introduction

By the classical theorem in arithmetic on the sums of two squares, the positive integers which can be written as a sum of two squares are precisely those whose prime factors of the form $4k + 3$ appear an even number of times in their prime factorization. A counterpart of this result for polynomial ring $F[X]$ where F is a field of characteristic different from 2 is also known. It turns out that a polynomial $g \in F[X]$ is a sum of squares of two polynomials, if the leading coefficient of g is a sum of squares of two elements of F and for every irreducible factor p of g for which -1 is not a square in the field $F[X]/(p(X))$, the exponent of p in the prime factorization of g is even (see [4, Thm 1, 2]). A multivariable version of the previous result is known as "*quadratic value theorem*" (see [2, 18.3]). This theorem states that if q is an arbitrary anisotropic quadratic form over a field F of arbitrary characteristic and $g \in F[T] = F[t_1, \cdots, t_n]$ is a nonzero polynomial then a scalar multiple of g lies in the group generated by the nonzero values represented by q over the field $F(T) = F(t_1, \cdots, t_n)$ if and only if q is isotropic over the residue field at p for each irreducible divisor p occurring to an odd power in the factorization of g. In the case where the set of all nonzero values represented by q form a group (e.g., if q is a Pfister form), this theorem indicates that a scalar multiple of a nonzero polynomial $g \in F[T]$ is represented by q if and only if q is isotropic for each irreducible divisor p occurring to an odd power in the factorization of g. Note that the quadratic value theorem, together with Cassels-Pfister theorem (which guarantees a polynomial solution $f_1, \cdots, f_n \in F[X]$ of $a_1 f_1(X)^2 + \cdots + a_n f_n(X)^2 = g(X)$, $a_1, \cdots, a_n \in F$, if there exists a rational function solution $f_1, \cdots, f_n \in F(X)$), imply the above polynomial version of the two squares theorem.

Two main important steps in the proof of these types of theorems are, first, to determine the conditions under which a prime or irreducible element p can be represented by a certain quadratic form, secondly, to determine the conditions under which p divides a nonzero value represented by that form. A look at the above theorems reveals that there always exists a dichotomy between the primes which can be represented by the corresponding form (or a scalar multiple of them lies in the group generated by the nonzero values represented by that form) and those primes p for which the corresponding form is anisotropic over the residue field at that prime. For example, the classical theorem on sums of two squares in

arithmetic shows that a prime p can be either written as a sum of two squares, i.e., $p = x^2 + y^2$ for some integers x and y, or $p \nmid x^2 + y^2$ unless $p|x$ and $p|y$. In the language of quadratic forms, this means that the binary quadratic form $q = X^2 + Y^2$ is either anisotropic over $\mathbb{F}_p$ or p is represented by q over $\mathbb{Z}$ (see the precise definition of anisotropic forms in §2). We have a similar dichotomy for the polynomial version of the two squares theorem and the quadratic value theorem.

The quadratic value theorem provides a very helpful principle which can be used to give a unified treatment of many theorems in the theory of quadratic forms and some related results on sums of squares. For example, the well-known Artin-Springer theorem [13] which states that an anisotropic quadratic form remains anisotropic after a scalar extension of odd degree, is a quasi-immediate corollary of it. Also, another result which has a similar esprit (see [12]) stating that if F is an ordered field and $f(X)$ is an irreducible polynomial on F which changes sign on F then the ordering of F can be extended to $F[X]/(f(X))$ and more generally a substitution principle, can be easily deduced from the quadratic value theorem (see [2, 18.9] and (6.3)).

Particular cases of the quadratic value theorem or very related results can be found in [8, Ch. X, 2.13], [10, Ch. 7, 1.7], [6], [7], [11, Ch. 2, 8.6] and [5, §5]. The proof of many of the above results is reminiscent of the proof of Artin-Springer theorem. Primes of the form $4k + 3$ in the theorem on sums of two squares play a similar role as the polynomials of odd degree in the Artin-Springer theorem. According to [10, p. 99], the main idea of these proofs "comes from an old theorem of Legendre on ternary quadratic forms over $\mathbb{Z}$."

A first step, used in the proof of many of the above results, is to apply a suitable division algorithm. Roughly speaking, if in an integral domain A, a prime p *of type* $4k + 3$ (or a polynomial of odd degree, depending on the context) divides an expression like $a_1\beta_1^2 + \cdots + a_n\beta_n^2$, then the technique consists in estimating the *magnitude* of $a_1\gamma_1^2 + \cdots + a_n\gamma_n^2$ where $\gamma_1, \cdots, \gamma_n$ are the remainders of the divisions of $\beta_1, \cdots, \beta_n$ by p. The relation $p|a_1\gamma_1^2 + \cdots + a_n\gamma_n^2$ leads to a contradiction, either by an infinite descent or by finding *a smaller* prime *of type* $4k + 3$ dividing $a_1\gamma_1^2 + \cdots + a_n\gamma_n^2$ and finishing by induction. This is actually what happens in the classical proof of Artin-Springer theorem where A is the polynomial ring $F[X]$ over a field F and p is an irreducible polynomial odd degree.

Unfortunately the ordinary euclidean division algorithm is not always available. Already for the case where $A = R[X]$ where R is an integral domain but not a field, one does not have such an algorithm with respect to the ordinary degree map. Nevertheless, we have a division algorithm in a weak sense: for every polynomial $f(X), g(X) \in R[X]$ there exist a nonzero scalar $\alpha \in R$ and the polynomials $q(X), r(X) \in R[X]$ such that $\alpha f(X) = g(X)q(X) + r(X)$ with $\deg(r(X)) + 1 \leq \deg(g(X))$. In this situation, the constant polynomials, i.e., the elements of R such as α, are *infinitesimal*, in the sense that an inequality like $\deg(a(X)) \leq \deg(b(X))$ does not change if one multiplies $a(X)$ by the elements of R. Also, the degree function satisfies the *ultrametric* inequality $\deg(f(X)+g(X)) \leq \max\{\deg(f(X)), \deg(g(X))\}$. It was perhaps Hurwitz, who for the first time, implicitly used such kinds of division algorithms in the context of number fields in [3]. Some of their number theoretics applications in connection with euclidean number fields appeared several decades later in [9]. For the polynomial ring $F[X_1, \cdots, X_n]$ where F is a field, this division algorithm has been implicitly used in the proofs of a norm principle in [7] and in the quadratic value theorem in [2, 18.3].

In this work, we apply and further exploit this machinery for more general weak division algorithms. Inspired by the division algorithm mentioned above, we consider an (ϵ, δ, m, n)-division algorithm in an arbitrary integral domain A, with a

degree like homomorphism $d : A \to (\Gamma, +)$ where Γ is a commutative positive totally ordered semigroup such that for every $a, b \in A$ with $b \neq 0$, there exist $q, r \in A$ such that $\alpha a = bq + r$ and α comes from a subset S_b of the nonzero elements of A depending only on b such that $d(\alpha_1 \cdots \alpha_n) + d(r) + \delta \leq d(b)$ for every $\alpha_1, \cdots, \alpha_n \in S_b$ (this intuitively encodes the "infinitesimal" nature of the elements of S_b) and $d(a_1 + \cdots + a_m) \leq \max\{d(a_1), \cdots, d(a_m)\} + \epsilon$ for every $a_1, \cdots, a_m \in A$ (this interprets the "ultrametric" nature of d).

In such circumstances, we state and prove in (4.3), a general dichotomy principle which shows that under reasonable assumptions, for a quadratic form $q(X_1, \cdots, X_n)$ with coefficients in a unique factorization domain A and a prime $p \in A$, one has two alternatives: either q is anisotropic over the residue field of A at p or there exists an invertible element $c \in A$ such that cp belongs to the group generated by the nonzero values represented by the form q in the multiplicative group of the field of fractions of A. In order to obtain truly number theoretic applications of this result, we also need a corresponding Cassels-Pfister type theorem. This Cassels-Pfister theorem is proved in (5.1).

In the last section of this paper, by considering different weak division algorithms (e.g., the binary division algorithm in $\mathbb{Z}$), we present some applications of the dichotomy principle given in this work.

2. Notation and Terminology

Unless otherwise stated, all fields considered in this work are of arbitrary characteristic. Let A be an integral domain. A *quadratic form* on A is a homogeneous polynomial $q(X_1, \cdots, X_n) \in A[X_1, \cdots X_n]$ of degree 2. The set of all nonzero values represented by $q = q(X_1, \cdots, X_n)$ over A is denoted by $D_A(q)$ or simply by $D(q)$ if no confusion arises, in other words

$$D_A(q) = \{q(\alpha) \neq 0 \mid \alpha \in A^n\}.$$

The group generated by $D(q)$ in the field of fractions of A is denoted by $\langle D(q) \rangle$. The bilinear form associated to q is denoted by B, i.e.,

$$B(X, Y) = q(X + Y) - q(X) - q(Y), \ X, Y \in A^n. \tag{1}$$

The A-submodule $\mathrm{rad}(B) = \{v \in A^n : B(v, w) = 0 \text{ for all } w \in A^n\}$ of A^n is called the *radical* of B. The standard *Gram* matrix of $q = \sum_{i<j} a_{ij} X_i X_j$ is the matrix $G = [b_{ij}]$ where $b_{ii} = 2a_{ii}$ and $b_{ij} = a_{ij}$ for $1 \leq i \neq j \leq n$. It follows that $u^t G u = 2q(u)$ for every column vector $u \in A^n$. A quadratic form q is called *nondegenerate* if the associated bilinear form B is nondegenerate. Alternatively q is nondegenerate if the standard Gram matrix of q is invertible (or equivalently $\mathrm{rad}(B) = 0$). The form q is said to be *isotropic*, if there exists a vector $u \in A^n \backslash \mathrm{rad}(B)$ such that $q(u) = 0$; such a vector u is called an *isotropic vector* for q. Hence, if q is nondegenerate, then q being isotropic simply means that there exists a nonzero vector $u \in A^n$ such that $q(u) = 0$. The form q is called *anisotropic* if q is not isotropic, in other words $q(u) = 0$ implies that $u \in \mathrm{rad}(B)$. In this sense, the form $X^2 + Y^2$ is anisotropic over $\mathbb{F}_2$.

For an integral domain A and a prime element p (resp. prime ideal P) of A, the *residue field* at p (resp. at P) is defined as the field of fractions of the integral domain $A/(p)$ (resp. A/P) and is denoted by $A(p)$ (resp. $A(P)$).

We recall that an ordered semigroup $(S, \leq)$ is called *positive* if for every $x, y \in S$ we have $x \leq xy$ and $x \leq yx$. Let $\mathbb{N}_0$ denote the set of nonnegative integers.

For a real number x, the largest integer less than or equal to x and the smallest integer greater than or equal to x are respectively denoted by $\lfloor x \rfloor$ and $\lceil x \rceil$. The fractional part of a positive real number x is denoted by $\{x\}$, i.e., $\{x\} = x - \lfloor x \rfloor$.

3. A weak division algorithm

Consider a commutative totally ordered positive semigroup $(\Gamma, +, \leq)$. We consider two extra symbols $-\infty, +\infty$ with the following formal properties $\pm\infty + \gamma = \pm\infty$, $\pm\infty + (\pm\infty) = \pm\infty$ and $-\infty < \gamma < +\infty$ for all $\gamma \in \Gamma$. Let m, n be two positive integers and $\epsilon, \delta \in \Gamma \cup \{\pm\infty\}$. We say that an integral domain A has an (ϵ, δ, m, n)-division algorithm with respect to a degree-like function $d : A \to \Gamma \cup \{-\infty\}$ if the following conditions are satisfied:
(i) $d(a) = -\infty$ if and only if $a = 0$.
(ii) $d(ab) = d(a) + d(b)$ for all $a, b \in A$.
(iii) $d(a_1 + a_2 + \cdots + a_m) \leq \max\{d(a_1), d(a_2), \cdots, d(a_m)\} + \epsilon$ for all $a_i \in A$.
(iv) For every $b \in A$ with $b \neq 0$, there exists a subset S_b of nonzero elements of A with $1 \in S_b$ such that for every $a \in A$ there exists $q, r \in A$ and $\alpha \in S_b$ with $\alpha a = qb + r$ and $d(\alpha_1\alpha_2 \cdots \alpha_n) + d(r) + \delta \leq d(b)$ for all $\alpha_i \in S_b$, $i = 1, \cdots, n$.

We say that d is *δ-strict*, if in the condition (iv) above, $d(\alpha_1\alpha_2 \cdots \alpha_n) + d(r) + \delta < d(b)$. We say that d is *ϵ-strict* if in the condition (iii) above we have the strict inequality $d(a_1 + a_2 + \cdots + a_m) < \max\{d(a_1), d(a_2), \cdots, d(a_m)\} + \epsilon$ if for some i, j we have $d(a_i) \neq d(a_j)$.

Lemma 3.1. (a) *For every invertible element u of A, the element $e = d(u)$ is the minimum element of the image of d.*
(b) *We have $e + e = e$.*
(c) *If Γ is a cancellative semigroup then e is the neutral element of Γ.*

Proof. (a) Since Γ is a positive semigroup, we have $d(u) \leq d(u) + d(u^{-1}a) = d(a)$ for every $a \in A$.
(b) We have $e \leq e + e$ by the positiveness of Γ. On the other hand, if u is an invertible element of A we have $e + e = d(u) + d(u) = d(u^2) \leq e$ since u^2 is invertible. Hence, $e + e = e$.
(c) This follows from the general fact that in a cancellative semigroup an element e is the neutral element if and only if e is idempotent. □

Example 3.2. (a) For every positive integers n, m, the absolute value function $d : \mathbb{Z} \to (\mathbb{N}_0, +) \cup \{-\infty\}$, $d(x) = |x|$, $x \neq 0$ and $d(0) = -\infty$ gives a $(+\infty, 1, m, n)$-division algorithm. In this case for every nonzero integer b, we can take $S_b = \{1\}$. This is the classical euclidean division algorithm.
(b) If $A[X]$ is the polynomial ring over an arbitrary integral domain A, then the ordinary degree function $d : A[X] \to (\mathbb{N}_0, +) \cup \{-\infty\}$ gives a $(0, 1, m, n)$-division algorithm for every positive integers n, m. In this case for every polynomial $b \in A[X]$, one can take $S_b = A \backslash \{0\}$. If A is a field, then we have the ordinary division algorithm of polynomials and one can take $S_b = \{1\}$.

Lemma 3.3. *Let c be a real number with $c > 1$. Then, for every real number x, there exists an integer $k \in \{1, 2, \cdots, \lceil c - 1 \rceil\}$ and an integer q such that $|kx - q| \leq 1/c$. This inequality is strict unless there exists an integer $k \in \{1, 2, \cdots, \lceil c - 1 \rceil\}$ such that $\{kx\} = \frac{c-1}{c}$.*

Proof. First consider the case where $1 < c \leq 2$. We should prove that there exists an integer q such that $|x - q| \leq 1/c$. In fact there exists an integer q such that $|x - q| \leq 1/2 \leq 1/c$ and the proof is complete.

Now consider the case where $c > 2$. If for some $k \in \{1, \cdots, \lceil c - 1 \rceil\}$, the quantity $\{kx\}$ belongs to one of the intervals $[0, 1/c[$ or $[(c-1)/c, 1[$ then for $q = 0$ or $q = 1$ we obtain $|kx - q| \leq 1/c$ and the proof is complete (note that this inequality is strict unless $\{kx\} = \frac{c-1}{c}$). Otherwise, for every $k \in \{1, \cdots, \lceil c - 1 \rceil\}$ we have $\{kx\} \in [1/c, (c-1)/c[$. Since $[1/c, (c-1)/c[$ is the union of the intervals

$I_1 \cup \cdots I_{\lceil c-1 \rceil -1}$ where

$$I_i = [\frac{1}{c} + (i-1)\frac{c-2}{c(\lceil c-1 \rceil -1)}, \frac{1}{c} + i\frac{c-2}{c(\lceil c-1 \rceil -1)}[,$$

by pigeonhole principle there exists an interval I_i for some i with $1 \leq i \leq \lceil c-1 \rceil -1$ such that $\{kx\} \in I_i$. It follows that for some $m, n \in \{1, \cdots, \lceil c-1 \rceil\}$ with $m > n$ we have $\{mx\}, \{nx\} \in I_i$. This implies that $|\{mx\} - \{nx\}| < \text{length}(I_i) = \frac{c-2}{c(\lceil c-1 \rceil -1)} \leq 1/c$. Now if we put $q = \lfloor mx \rfloor - \lfloor nx \rfloor$ we obtain $|(m-n)x - q| = |\{mx\} - \{nx\}| < 1/c$. □

In the below lemma we use the convention $\log_c(0) = -\infty$ for a real number $c > 1$.

Lemma 3.4. *Let $c > 1$ be a real number and m, n be positive integers. Then $\mathbb{Z}$ has a $(\log_c m, 1 - n\log_c(\lceil c-1 \rceil), m, n)$-division algorithm with respect to the function $d : \mathbb{Z} \to \mathbb{R}_{\geq 0} \cup \{-\infty\}$ given by $d(x) = \log_c |x|$.*

Proof. We should check the validity of (i), (ii), (iii) and (iv). We obviously have (i) and (ii). To prove (iii), consider $a_1, \cdots, a_m \in \mathbb{Z}$ with $a_i \neq 0$ for all i. Choose i with $1 \leq i \leq m$ such that

$$|a_i| = \max\{|a_1|, |a_2|, \cdots, |a_m|\}.$$

We obtain $d(a_1 + \cdots + a_m) = \log_c(|a_1 + a_2 + \cdots + a_m|) \leq \log_c(|a_1| + |a_2| + \cdots + |a_m|) \leq \log_c(m|a_i|) = \log_c(|a_i|) + \log_c m$ and the proof of (iii) is finished.

To prove (iv), consider a nonzero $b \in \mathbb{Z}$. Take the set $S_b = \{1, \cdots, \lceil c-1 \rceil\}$ (here S_b does not depend on b). Consider an element $a \in \mathbb{Z}$. First note that since $d(-1) = 0$, we have $d(z) = d(-z)$ for every $z \in \mathbb{Z}$, hence proving the result for $-a$ or $-b$ instead of a or b are equivalent. Thus, without loss of generality we may assume that both a and b are positive. Also, note that if $x = a/b$ is an integer then by taking $\alpha = 1$ and $r = 0$, the validity of (iv) is established. Hence, we may assume that x is not integer. According to (3.3), there exists an integer q and an element $\alpha \in \{1, \cdots, \lceil c-1 \rceil\}$ such that $|\alpha x - q| \leq 1/c$ hence

$$d(\alpha x - q) = \log_c(|\alpha x - q|) \leq \log_c(1/c) = -1 \tag{2}$$

We have $\alpha x - q = (\alpha a - qb)/b$. Putting $r := \alpha a - bq \in \mathbb{Z}$, we obtain $\alpha a = qb + r$. It remains to show that $d(\alpha_1 \alpha_2 \cdots \alpha_n) + d(r) + \delta \leq d(b)$ or equivalently $\log_c(\alpha_1 \alpha_2 \cdots \alpha_n) + \log_c(|r/b|) + \delta \leq 0$ for every $\alpha_1, \cdots, \alpha_n \in S_b$. But this inequality can be proved as follows: $\log_c(\alpha_1 \cdots \alpha_n) + \log_c(|r/b|) + \delta = \log_c(\alpha_1 \cdots \alpha_n) + \log_c(|\alpha x - q|) + \delta \leq n\log_c(\lceil c-1 \rceil) - 1 + \delta = 0$ by (2) and the fact that $\alpha_i \leq \lceil c-1 \rceil$ for $i = 1, \cdots, n$. □

4. A dichotomy principle

Definition 4.1. Let $(\Gamma, +)$ be an abelian torsion-free group. We say that a quadratic form $q(X_1, \cdots, X_n) = \sum_{i \leq j} a_{ij} X_i X_j \in A[X_1, \cdots, X_n]$ is Γ-*equilibrated* if every solution $(\gamma_1, \cdots, \gamma_n) \in \Gamma^n$ of the equations $\gamma_i + \gamma_j = \gamma_{i'} + \gamma_{j'} = \gamma_{i''} + \gamma_{j''} = \cdots = \gamma$ where $\gamma_1, \cdots, \gamma_n \in \Gamma$ and $\gamma \in G$ is a constant and $(i, j), (i', j'), (i'', j''), \cdots$ are all indices for which $a_{ij}, a_{i'j'}, a_{i''j''}, \cdots$ are nonzero, is of the form $\gamma_1 = \gamma_2 = \cdots = \gamma_n$.

Example 4.2. Every diagonal quadratic form $q = a_1 X_1^2 + \cdots + a_n X^2$, $a_i \in A$ is $\mathbb{Z}$-equilibrated since the equations $2\gamma_1 = \cdots = 2\gamma_n = \gamma$ imply that $\gamma_1 = \cdots = \gamma_n$. The two dimensional form $q(X_1, X_2) = X_1 X_2$ is not $\mathbb{Z}$-equilibrated since the equation $\gamma_1 + \gamma_2 = \gamma$ does not imply that $\gamma_1 = \gamma_2$.

Theorem 4.3. *Let A be a unique factorization domain. Assume that A has an $(\epsilon, \delta, m, n-1)$-division algorithm with respect to a degree-like function $d : A \to \Gamma \cup \{-\infty\}$ where Γ is a commutative positive totally ordered cancellative semigroup. Assume further that the image of d has the least element property and let e be the minimum element of the image of d and suppose that $\delta > e$.*

Let $q(X_1, \cdots, X_n) = \sum_{i \leq j} a_{ij} X_i X_j \in A[X_1, \cdots, X_n]$ be a quadratic form with m nonzero terms and let $\mu = \max\{d(a_{ij}) : 1 \leq i, j \leq n\}$. Assume that for all primes p for which q is degenerate over $A(p)$, there exists an invertible element $c \in A$ such that $pc \in \langle D(q) \rangle$. Consider the following situations:
(v) $2\delta > \mu + \epsilon$.
(vi) $2\delta \geq \mu + \epsilon$ *and d is δ-strict.*
(vii) $\mu = e$, $2\delta \geq \epsilon$, *q is Γ-equilibrated, d is ϵ-strict, and for every element $a \in A$ with $d(a) \neq e$ there exists a prime element in A, depending only on $d(a)$, which divides a.*
Assume that in the situations (v) *and* (vi) *for all primes p of degree e (if any) and in the situation* (vii) *for all primes p of degree $\leq \epsilon$ (if any) one of the following alternatives holds: either q is anisotropic over $A(p)$ or there exists an invertible element $c \in A$ such that $pc \in \langle D_A(q) \rangle$. Then the same alternatives hold for all primes p in all situations* (v), (vi) *and* (vii).

Proof. By (3.1), e is necessarily the neutral element of Γ. Since the image of d is well-ordered, we can proceed by induction on the degree of p. Assume that the assertion is true for all primes of degree less than γ, for some γ in the image of d. It is enough to show that if p is a prime element of degree γ such that q is nondegenerate over $A(p)$ and q is isotropic over $A(p)$ then there exists an invertible element $c \in A$ such that $pc \in \langle D(q) \rangle$. By the hypotheses, this assertion is true if q is degenerate over $A(p)$. Hence, we may that q is nondegenerate over $A(p)$. We may also assume that $d(p) > e$ (resp. $d(p) > \epsilon$) in the situations (v) and (vi) (resp. (vii)), since otherwise the conclusion is guaranteed by the hypotheses. Let $\xi = (\alpha_1, \cdots, \alpha_n) \in A(p)^n$ be an isotropic vector for q. Hence, $q(\xi) = 0$ and $\xi \notin \mathrm{rad}(B_{A(p)}) = \{0\}$. It follows that there exist elements $f_1, \cdots, f_n, h \in A$ such that at least one of f_i's is nonzero modulo p and

$$q(f_1, \cdots, f_n) = ph. \tag{3}$$

Since A has an $(\epsilon, \delta, m, n-1)$-division algorithm with respect to d, there exists a subset S_p of nonzero elements of A and an element $\alpha_i \in S_p$ such that

$$\alpha_i f_i = q_i p + r_i, \ i = 1, \cdots, n, \tag{4}$$

where $q_i, r_i \in A$ and for every $\beta_1, \cdots, \beta_{n-1} \in S_p$

$$d(\beta_1 \cdots \beta_{n-1}) + d(r_i) + \delta \leq d(p), \ i = 1, \cdots, n. \tag{5}$$

We claim that for every $\beta \in S_b$ we have $p \nmid \beta$. In fact, if for some $\beta \in S_p$ we have $p|\beta$, then (5) implies that $d(p) + \delta \leq d(p)$, hence $d(p) + \delta = d(p)$ since Γ is a positive semigroup. This implies that $\delta = e$, because Γ is cancellative, this contradicts our assumption that $\delta > e$. In particular, we have $p \nmid \alpha_i$ for all $i = 1, \cdots, n$. We now claim that at least one of r_i's is nonzero. In fact, if $r_i = 0$ for all $i = 1, \cdots, n$ then the relation (4) together with the fact that $p \nmid \alpha_i$ imply that $p|f_i$ for all i. This contradicts our assumption that at least one of f_i's is nonzero modulo p. Now, by multiplying both sides of (3) by $\alpha_1^2 \cdots \alpha_n^2$ and using (4), we obtain

$$q(f_1', \cdots, f_n') = ph', \text{ for some } h' \in A, \tag{6}$$

where $f_1' = \alpha_2 \alpha_3 \cdots \alpha_n r_1, \cdots, f_n' = \alpha_1 \alpha_2 \cdots \alpha_{n-1} r_n$. Since

$$\alpha_1, \alpha_2, \cdots, \alpha_n \in S_p,$$

using (5), we obtain $d(f_i') + \delta \leq d(p)$ for every $i = 1, \cdots, n$. Hence, from the beginning we may assume that in the relation (3), the following additional condition holds

$$d(f_i) + \delta \leq d(p). \tag{7}$$

We may assume that the terms $f_1, \cdots, f_n$ do not have a common prime divisor; in fact if a prime element $p_1 \in A$ divides all f_i's then by (3) we obtain $p_1^2 | ph$. As we have mentioned, at least one of f_i's is nonzero modulo p, hence p and p_1 are not associated primes, it follows that $p_1^2 | h$ and we may cancel p_1^2 from both sides of (3).

Since we have an $(\epsilon, \delta, m, n-1)$-division algorithm with respect to d, we obtain

$$d(q(f_1, \cdots, f_n)) \leq \max\{d(a_{ij}) + d(f_i f_j)) : a_{ij} \neq 0\} + \epsilon \tag{8}$$

By adding 2δ to both sides of (8) and using (7) and the assumption $d(a_{ij}) \leq \mu$ for $a_{ij} \neq 0$ we obtain:

$$d(q(f_1, \cdots, f_n)) + 2\delta \leq \mu + 2d(p) + \epsilon \tag{9}$$

In the situation (v) or (vi), at least one of the inequalities $2\delta \geq \mu + \epsilon$ or (5) (hence (7)) is strict. Note that since (9) is a consequence of (7), the inequality (9) is strict too. Hence, in both cases, from (9) we obtain the strict inequality

$$d(q(f_1, \cdots, f_n)) < 2d(p). \tag{10}$$

We claim that in the situation (vii), this inequality holds as well. First note that the condition $\mu = e$ implies that $d(a_{ij}) = e$ for all nonzero a_{ij}'s. To prove (10), note that as mentioned at the beginning of the proof, we may assume that $d(p) > \epsilon$ in the situation (vii). If for some $i < j$ we have $d(f_i) \neq d(f_j)$ then as q is Γ-equilibrated, there exist two pairs (r, s) and (r', s') with $r \leq s$ and $r' \leq s'$ such that $a_{rs} \neq 0$ and $a_{r's'} \neq 0$ and $d(f_r f_s) \neq d(f_{r'} f_{s'})$. This relation together with the hypothesis in (vii) that d is ϵ-strict, imply that the inequality (8) is strict too. Consequently, (9) is strict and we obtain (10). If for all $i < j$, $d(f_i) = d(f_j)$, then by the last hypothesis in (vii), either $f_1, \cdots, f_n$ have a common prime divisor or $d(f_i) = e$ for $i = 1, \cdots, n$. But since f_i's do not have a common prime divisor, it follows that the first case cannot occur. Hence, $d(f_i) = e$, for all i. Now, using (3) and (8) we obtain $d(p) \leq d(q(f_1, \cdots, f_n)) \leq e + \epsilon = \epsilon$ which is a contradiction. Thus, we have shown the validity of (10) in all situations (v), (vi) and (vii).

The relation (10) together with (3) imply that

$$d(h) < d(p). \tag{11}$$

First consider the case where h is a non-invertible. Let $h = p_1 \cdots p_r$ be a decomposition of h into prime elements p_i's. By (11) we have $d(p_i) < d(p)$ for $i = 1, \cdots, r$. We claim that there exists an invertible element $c_i \in A$ such that $p_i c_i \in \langle D(q) \rangle$ for $i = 1, \cdots, r$. The correctness of this claim is guaranteed by the hypotheses if q is degenerate over $A(p_i)$. If q is nondegenerate over $A(p_i)$ then considering the relation (3) modulo p_i (and remarking that $f_1, \cdots, f_n$ do not have a common prime divisor, hence at least one of them is nonzero modulo p_i) we obtain that $q(X_1, \cdots, X_n)$ is isotropic over $A(p_i)$. By induction, the second alternative holds and the proof of the claim is completed. Hence, for every i we have $p_i c_i \in \langle D(q) \rangle$. It follows that $hc_1 \cdots c_r \in \langle D(q) \rangle$ and the relation (3) implies that $pc_1 \cdots c_r \in \langle D(q) \rangle$ as desired.

Finally, consider the case where $c = h$ is invertible. The relation (3) implies that $pc \in \langle D(q) \rangle$ and the proof is complete. □

Corollary 4.4. *With the same hypotheses as in* (4.3)*, assume further that q represents* 1*, then for a nonzero element $f \in A$ the following conditions are equivalent:*
(a) *There exists an invertible element $f_0 \in A$ such that $f_0 f \in \langle D(q) \rangle$.*

(b) *The form q is isotropic over $A(p)$ for every prime divisor p of f occurring to an odd power in the factorization of f.*

Proof. (a)⇒(b): We may assume that f is not invertible; Let p a prime divisor of f occurring to an odd power in the factorization of f. By (a) there exist anisotropic vectors $v_1, \cdots, v_r, u_1, \cdots, u_s \in A^n$ such that

$$f_0 f q(u_1) \cdots q(u_s) = q(v_1) \cdots q(v_r).$$

We can write $u_i = p^{\alpha_i} u'_i$ and $v_j = p^{\beta_j} v'_j$ for some nonnegative $\alpha_i, \beta_j \in \mathbb{Z}$ such that at least one entry of u'_i and v'_j is not divisible by p. From the above relation we obtain

$$p^{2\alpha} f_0 f q(u'_1) \cdots q(u'_s) = p^{2\beta} q(v'_1) \cdots q(v'_r),$$

where $\alpha = \alpha_1 + \cdots + \alpha_s$ and $\beta = \beta_1 + \cdots + \beta_r$. Since p occurs to an odd power in a factorization of f, the above relation implies that either there exists an index i such that $p|q(v'_i)$ or there exists an index j such that $p|q(u'_j)$. In both cases, q becomes isotropic over $A(p)$.
(b)⇒(a) If f is invertible, then by taking $f_0 = f^{-1}$, (a) holds. Otherwise, it suffices to prove that for every prime divisor p of f occurring to an odd power in the factorization of f, there exists an invertible element c such that $pc \in \langle D(q) \rangle$, but this is exactly the conclusion of (4.3). □

5. A Cassels-Pfister's representation theorem

Theorem 5.1. *Let A be an integral domain. Assume that A has an (ϵ, δ, m, n)-division algorithm with respect to a degree-like function $d : A \to \Gamma \cup \{-\infty\}$ where Γ is a commutative positive totally ordered cancellative semigroup. Assume further that the image of d has the least element property and let e is the minimum element of the image of d and suppose that $\delta > e$ and $d^{-1}(e)$ consists only of invertible elements. Let $q(X_1, \cdots, X_n) = \sum_{i \leq j} a_{ij} X_i X_j \in A[X_1, \cdots, X_n]$ be an anisotropic quadratic form with m nonzero terms and let $\mu = \max\{d(a_{ij}) : 1 \leq i, j \leq n\}$. Assume that one of the following conditions hold*
(v) $2\delta > \mu + \epsilon$.
(vi) $2\delta \geq \mu + \epsilon$ *and d is δ-strict.*
In this case, if an element $b \in A$ is represented by q over the field of fractions K of A then b is already represented by q over A.

Proof. As we have mentioned in (3.1), e is the neutral element of Γ. Since b is represented by q over the field of fractions of A, there exist $f_0, f_1, \cdots, f_n \in A$ with $f_0 \neq 0$ such that

$$q(\frac{f_1}{f_0}, \cdots, \frac{f_n}{f_0}) = b. \tag{12}$$

We proceed by induction on $d(f_0)$. If $d(f_0) = e$, the hypothesis on $d^{-1}(e)$ implies that f_0 is invertible, hence $f_i/f_0 \in A$ for $i = 1, \cdots, n$ and the proof is complete. We may assume that $d(f_0) > e$ and $d(f_0)$ is minimal (with $f_0 \neq 0$) among all representations of the form (12). The relation (12) leads to

$$q(f_1, \cdots, f_n) - b f_0^2 = 0. \tag{13}$$

Thus, if we put $Q(X_1, \cdots, X_{n+1}) = q(X_1, \cdots, X_n) - bX_{n+1}^2$, for $u = (f_1, \cdots, f_n, f_0)$ we have $Q(u) = 0$. We may assume that $f_1, \cdots, f_n, f_0$ do not have a non-invertible common divisor, otherwise we can cancel it from both sides of (13). As we have an (ϵ, δ, m, n)-division algorithm with respect to d, there exists a subset S_{f_0} of nonzero elements of A such that

$$\alpha_i f_i = g_i f_0 + r_i, \tag{14}$$

where $r_i, g_i \in A$, $\alpha_i \in S_{f_0}$ and for every $\beta_1, \cdots, \beta_n \in S_{f_0}$ we have

$$d(\beta_1 \cdots \beta_n) + d(r_i) + \delta \leq d(f_0). \tag{15}$$

In particular, we can take $r_0 = 0$ and $\alpha_0 = g_0 = 1$. Consider the vector

$$v = (\frac{g_1}{\alpha_1}, \cdots, \frac{g_n}{\alpha_n}, \frac{g_0}{\alpha_0}) = (\frac{g_1}{\alpha_1}, \cdots, \frac{g_n}{\alpha_n}, 1) \in K^{n+1}. \tag{16}$$

We have $Q(v) \neq 0$, since we would otherwise obtain

$$q(\frac{\alpha_2 \cdots \alpha_n g_1}{\alpha_1 \cdots \alpha_n}, \cdots, \frac{\alpha_1 \cdots \alpha_{n-1} g_n}{\alpha_1 \cdots \alpha_n}) = b$$

which contradicts the minimality of the degree of f_0 (note that by (15) we have $d(\alpha_1 \cdots \alpha_n) < d(\alpha_1 \cdots \alpha_n) + \delta \leq d(f_0)$ since $\delta > e$). The map $\tau_v : K^{n+1} \to K^{n+1}$ defined by

$$\tau_v(x) = x - \frac{B(x,v)}{Q(v)} v, \tag{17}$$

where B is the bilinear form associated to Q as defined in (1), is an isometry (c.f., [2, 7.2]). If $\mathrm{char}(K) \neq 2$ (resp. $\mathrm{char}(K) = 2$), the map τ_v is usually called the *reflection* (resp. *orthogonal transvection*) along v. Since $Q(u) = 0$ we obtain $Q(\tau_v(u)) = 0$, in other words

$$Q(u - \frac{B(u,v)}{Q(v)} v) = 0.$$

Multiplying this relation by $Q(v)^2$ we obtain $Q(Q(v)u - B(u,v)v) = 0$. Put

$$w = Q(v)u - B(u,v)v. \tag{18}$$

Hence $Q(w) = 0$. We can write $w = (w_1, \cdots, w_n, w_0)$ where

$$w_i = Q(v) f_i - B(u,v) \frac{g_i}{\alpha_i}.$$

We try to simplify the last entry w_0 of w. We have

$$\begin{aligned} w_0 &= Q(v) f_0 - B(u,v) \frac{g_0}{\alpha_0} = Q(v) f_0 - B(u,v) \\ &= f_0 (Q(v) - B(u,v) \frac{1}{f_0}) = f_0 Q(v - \frac{1}{f_0} u) = \frac{1}{f_0} Q(f_0 v - u) \end{aligned} \tag{19}$$

The i-th entry of $f_0 v - u$ is equal to $f_0 \frac{g_i}{\alpha_i} - f_i = \frac{1}{\alpha_i}(f_0 g_i - \alpha_i f_i) = -\frac{r_i}{\alpha_i}$ by (14), hence

$$f_0 v - u = (-\frac{r_1}{\alpha_1}, \cdots, -\frac{r_n}{\alpha_n}, -\frac{r_0}{\alpha_0}) = (-\frac{r_1}{\alpha_1}, \cdots, -\frac{r_n}{\alpha_n}, 0) \tag{20}$$

As $Q(w) = 0$, we obtain $Q(w') = 0$ where $w' = \alpha_1 \cdots \alpha_n w \in A^{n+1}$. We can write $w' = (w'_1, \cdots, w'_n, w'_0)$ where $w'_i = \alpha_1 \cdots \alpha_n w_i$. Thus,

$$q(w'_1, \cdots w'_n) - b {w'_0}^2 = 0 \tag{21}$$

Using (19) and (20) we obtain

$$\begin{aligned} w'_0 = \alpha_1 \cdots \alpha_n w_0 &= (\alpha_1 \cdots \alpha_n) \frac{1}{f_0} Q(f_0 v - u) \\ &= (\alpha_1 \cdots \alpha_n) \frac{1}{f_0} Q(-\frac{r_1}{\alpha_1}, \cdots, -\frac{r_n}{\alpha_n}, 0) \end{aligned} \tag{22}$$

Now by setting

$$(f'_1, \cdots, f'_n, f'_0) = (-\alpha_0 \alpha_2 \cdots \alpha_n r_1, \cdots, -\alpha_0 \alpha_1 \cdots \alpha_{n-1} r_n, 0) \tag{23}$$

and multiplying (22) by $\alpha_1 \cdots \alpha_n f_0$ we obtain

$$\alpha_1 \cdots \alpha_n f_0 w'_0 = Q(f'_1, \cdots, f'_n, f'_0) = Q(f'_1, \cdots, f'_n, 0) = q(f'_1, \cdots, f'_n) \tag{24}$$

Assume that $d(a_{ij}f'_if'_j)$ is maximum among all terms of

$$q(f'_1,\cdots,f'_n)=\sum_{i<j}a_{ij}f'_if'_j$$

for some i,j. The relations (23), (24) and the fact that there exists an (ϵ,δ,m,n) division algorithm with respect to d imply that

$$\begin{aligned} d(w'_0)+d(f_0)+d(\alpha_1\cdots\alpha_n) &= q(f'_1,\cdots,f'_n)\\ &\le \max\{d(a_{ij})+d(f'_if'_j)\}+\epsilon\\ &= d(a_{ij})+d(f'_i)+d(f'_j)+\epsilon\\ &\le \mu+2d(\alpha_1\cdots\alpha_n)+d(r_i)+d(r_j)+\epsilon. \end{aligned} \tag{25}$$

Adding 2δ to all sides of (25) and thanks to (15) we obtain:

$$\begin{aligned} d(w'_0)+d(f_0)+2\delta &\le d(w'_0)+d(f_0)+2\delta+d(\alpha_1\cdots\alpha_n)\\ &\le \mu+2d(\alpha_1\cdots\alpha_n)+d(r_i)+d(r_j)+\epsilon+2\delta\\ &\le \mu+2d(f_0)+\epsilon, \end{aligned} \tag{26}$$

hence

$$d(w'_0)+2\delta\le\mu+d(f_0)+\epsilon. \tag{27}$$

In the case (v), we have $2\delta>\mu+\epsilon$, hence (27) leads to the inequality $d(w'_0)<d(f_0)$. In the case (vi), the inequality (15) is strict, hence the second inequality in (26) is strict too. Hence, the inequality (27) which is an implication of (26) is strict. Thus, in this case we also have the inequality $d(w'_0)+2\delta<\mu+d(f_0)+\epsilon$. This together with the hypothesis $2\delta\ge\mu+\epsilon$ in (vi) imply that $d(w'_0)<d(f_0)$. Thus, in both cases we obtain the inequality $d(w'_0)<d(f_0)$. Also, note that w'_0 is nonzero; since otherwise (21) and the fact that q is anisotropic implies that $w'_1=\cdots=w'_n=0$. We obtain then $w_1=\cdots=w_n$. Hence, $w=(w_1,\cdots,w_n,w_0)=0$. Now the definition of w given in (18) implies that u and v are linearly independent. Since $Q(u)=0$, we obtain $Q(v)=0$. This is a contradiction, because we have shown that $Q(v)\neq 0$. Hence, we have proved that w'_0 is nonzero and $d(w'_0)<d(f_0)$. This together with (21) contradict our assumption of the minimality of $d(f_0)$. □

6. Some applications

Example 6.1. (Cassels-Pfister's Theorem, [10, 2.3]) Consider the polynomial ring $F[t]$ over a field F. Let $q(X_1,\cdots,X_n)=\sum a_{ij}(t)X_iX_j$ be an anisotropic quadratic form where $a_{ij}(t)\in F[t]$ is of degree at most 1 for every i and j. Let $f(t)\in F[t]$ be a polynomial which is represented by q over the field of fractions $F(t)$ of $F[t]$. Then $f(t)$ is already represented by q over $F[t]$.

In this situation, we use the ordinary degree function $d:F[X]\to\mathbb{N}_0$ mentioned in (3.2)(b). The function d induces a $(0,1,m,n)$-division algorithm for every m,n. In particular, we can take n (resp. m) as the number of variables (resp. the number of nonzero terms) of the form q. With the notations of (5.1) we have $\mu\le 1$. Thus, the condition $2\delta>\mu+\epsilon$ is satisfied and the conclusion holds. Note that the hypothesis on $d^{-1}(e)$ in (5.1) prevents to replace the field F with an arbitrary unique factorization domain.

Example 6.2. (Quadratic value theorem, [2, Thm. 18.3]) Let F be a field and let $F[T]=F[t_1,\cdots,t_r]$. Consider a nondegenerate anisotropic quadratic form $q(X_1,\cdots,X_n)=\sum a_{ij}X_iX_j$ where $a_{ij}\in F$. Let p be a prime element of $F[T]$. Then either $q(X_1,\cdots,X_n)$ remains anisotropic over $F[T](p)$ or there exists a nonzero element $c\in F$ such that $cp\in\langle D_{F[T]}(q)\rangle$.

We prove this result by induction on r. For $r = 0$, the conclusion holds since $F[T] = F$ does not contain any prime element. Hence, let assume that $r \geq 1$. Put $A = F[t_1, \cdots, t_{r-1}]$, hence $F[T] = A[t]$. According to (3.2)(b), we have a $(0, 1, m, n-1)$ division algorithm on $A[t]$ with respect to the usual degree function $d : A[t] \to \mathbb{N}_0$, where m denotes the number of nonzero terms of q. We have $\mu = 0$, hence the condition $2\delta \geq \mu + \epsilon$ is satisfied. In order to apply (4.3), it suffices to check that for every prime $p \in A[t]$ of order $e = 0$ (hence $p \in A$) either q is anisotropic over $A[t](p)$ or there exists an invertible element $c \in A[t]$ such that $cp \in \langle D_{A[t]}(q)\rangle$. We have a ring isomorphism $A[t]/(p) \simeq A(p)[t]$ since $p \in A$, hence we have a field isomorphism $A[t](p) \simeq A(p)(t)$. If q is isotropic over $A[t](p)$, the relation $A[t](p) \simeq A(p)(t)$ implies that q is isotropic over $A(p) = F[t_1, \cdots, t_r](p)$. By induction there exists a nonzero $c \in F$ such that $cp \in \langle D_{F[t_1,\cdots,t_{r-1}]}(q)\rangle \subseteq \langle D_{F[t_1,\cdots,t_r]}(q)\rangle$ and the proof is complete. Thus, we have checked the validity of the result for all primes of degree $\leq e = 0$. The conclusion now follows from (5.1).

Example 6.3. (Compare [2, 18.9]) Let F be a field and let $F[T] = F[t_1, \cdots, t_r]$. Consider a nondegenerate anisotropic quadratic form $q(X_1, \cdots, X_n) = \sum a_{ij} X_i X_j$ where $a_{ij} \in F$. Let $p(T)$ be an irreducible polynomial in $F[T]$. Suppose that there exists an element $c \in F^r$ such that $p(c) \notin \langle D_{F[T]}(q)\rangle$. Then q is anisotropic over $F[T](p(T))$.

By (6.2) if q is not anisotropic over $F[t](p(T))$, there exists an invertible scalar $\lambda \in F$ such that $\lambda p(T) \in \langle D_{F[T]}(q)\rangle$. It follows that there exist polynomials $f_i(T) \in D_{F[T]}(q)$ and $g_j(T) \in D_{F[T]}(q)$ for $i = 1, \cdots, n$ and $j = 1, \cdots, m$, such that $f_1(T) \cdots f_n(T) p(T) = g_1(T) \cdots g_m(T)$. Putting $T := c$ in this identity, implies that $p(c) \in \langle D_{F[T]}(q)\rangle$, which is a contradiction.

Example 6.4. (Multivariable version of Artin-Springer's theorem) With the same notation as in (6.3), if the degree of p with respect to one of its variables $t \in \{t_1, \cdots, t_r\}$ is odd then q remains anisotropic over $F[T](p)$.

It is enough to compare the degrees of both sides of $f_1(T) \cdots f_n(T) p(T) = g_1(T) \cdots g_m(T)$ with respect to the corresponding variable t.

Example 6.5. (Two squares theorem)
(a) If an integer is a sum of two rational squares, then it can be written as a sum of two integral squares.
(b) Every prime number of the form $4k + 1$ can be written as a sum of squares of two integers.
(c) If p is a prime number of the form $4k + 3$ and $p | a^2 + b^2$ for some integers a and b then p divides both a and b.

Consider the quadratic form $q(X_1, X_2) = X_1^2 + X_2^2$. Here, we have $n = m = 2$. By (3.4) for $c = 2$, the ring $\mathbb{Z}$ has an $(\epsilon, \delta, m, n) = (1, 1, 2, 2)$-division algorithm with respect to the degree function $d(x) = \log_2 |x|$. Note that the image of d is a discrete, hence a well-ordered subset of nonnegative reals. Hence, for $A = \mathbb{Z}$, $\Gamma = (\mathbb{R}_{\geq 0}, +)$, we can apply (4.3) if the respective hypotheses are satisfied. We have $\mu = 0$, hence $2\delta > \epsilon + \mu$ is satisfied and the conclusion of (a) follows from (4.3).

Again using (3.4) for $c = 2$, the ring $\mathbb{Z}$ has an $(\epsilon, \delta, m, n-1) = (1, 1, 2, 1)$-division algorithm with respect to the degree function $d(x) = \log_2 |x|$. Also, q is nondegenerate over all $\mathbb{F}_p$ for all primes p. In this situation we have $\mu = 0$, hence $2\delta > \mu + \epsilon$ holds. Consider a prime number p of the form $4k + 1$. According to (4.3), either q remains anisotropic over $\mathbb{F}_p$ or pc is a sum of two squares in $\mathbb{Q}$ for some $c \in \{1, -1\}$. The former case is ruled out (for instance, one has $p | (\frac{p-1}{2}!)^2 + 1$ by Wilson's theorem). In the latter case, $c = -1$ is rules out since q is a positive definite form. Hence, p is a sum of two squares in $\mathbb{Q}$. Now (6.5) (a) implies that p is a sum of two integral squares.

(2) If p does not divide a or b then q becomes isotropic over $\mathbb{F}_p$. By (4.3) we deduce that p is represented by q over $\mathbb{Q}$, hence over $\mathbb{Z}$. Now a congruence argument modulo 4 yields a contradiction.

Example 6.6. (a) A nonnegative integer can be written, in the form x^2+xy+y^2 for some $x, y \in \mathbb{Q}$ if and only if it can be written in the same form for some $x, y \in \mathbb{Z}$. (b) A nonnegative prime $p \neq 3$ can be written in the form x^2+xy+y^2 for some $x, y \in \mathbb{Z}$ if and only if p of the form $3k+1$.

Consider the quadratic form $q(X_1, X_2) = X_1^2 + X_1X_2 + X_2^2$ over $\mathbb{Z}$. Here we have $n = 2$ and $m = 3$. According to (3.4), for $c = 2$ we have an $(\epsilon, \delta, m, n) = (\log_2 3, 1, 3, 2)$ on $\mathbb{Z}$ with respect to the degree function $d : \mathbb{Z} \to \mathbb{R}_{\geq 0}$ given by $d(x) = \log_2 |x|$. Here we have $\mu = 0$, hence the condition $2\delta > \mu + \epsilon$ is satisfied and the conclusion of (a) follows from (5.1).

To prove (b), it is enough by (a) to show that $p \in D_{\mathbb{Q}}(q)$ if and only if p is of the form $3k+1$. We first show that $\langle D_{\mathbb{Z}}(q)\rangle = D_{\mathbb{Q}}(q)$. To see this note that since the discriminant of q is 3, the form q is equivalent to the form $q'(X_1, X_2) = X_1^2 + 3X_2^2$ over $\mathbb{Q}$. Hence, $\langle D_{\mathbb{Z}}(q)\rangle = \langle D_{\mathbb{Z}}(q')\rangle = D_{\mathbb{Q}}(q') = D_{\mathbb{Q}}(q)$ (in fact, more generally this argument shows that if a quadratic form q defined over an integral domain A is equivalent to a multiplicative form over the field of fractions K of A then $\langle D_A(q)\rangle = D_K(q)$). Hence, it is enough to prove that $p \in \langle D_{\mathbb{Z}}(q)\rangle$ if and only if p is of the form $3k+1$. Again using (3.4), for $c = 2$ we have an $(\epsilon, \delta, m, n-1) = (\log_2 3, 1, 3, 1)$ division algorithm with respect to the degree function d defined above. Note that $p = 3$ is the only prime number for which q is nondegenerate over $\mathbb{F}_p$, by the way $3 \in D_{\mathbb{Q}}(q)$. By (4.3), for every prime p, either q is anisotropic over $\mathbb{F}_p$ or $p \in \langle D_{\mathbb{Z}}(q)\rangle$. But since the discriminant of q is 3, it is anisotropic precisely when the Legendre symbol $(-3/p)$ is nontrivial or equivalently p is not of the form $3k+1$.

Example 6.7. (L. Aubry, 1912, [1, p. 273]) Let n be an integer which is a sum of three rational squares. Then n is a sum of three integral squares.

Consider the quadratic form $q(X_1, X_2, X_3) = X_1^2+X_2^2+X_3^2$. We have $m = n = 3$. According to (3.4), for $c = 2$ there exists a $(\epsilon, \delta, m, n) = (\log_2(3), 1, 3, 3)$-division algorithm with respect to the degree function $d(x) = \log_2(|x|)$ on $\mathbb{Z}$. Here we have $\mu = 0$ and the condition $2\delta > \mu + \epsilon$ holds. According to (5.1), we conclude that n is a sum of three integral squares.

Example 6.8. (Weak version of the four squares theorem) Every prime number p can be written as a sum of four rational squares.

Consider the quadratic form $q(X_1, X_2, X_3, X_4) = X_1^2 + X_2^2 + X_3^2 + X_4^2$. Here we have $m = n = 4$. We use the same notation as in the proof of (6.5). According to (3.4), for $c = 2$ the ring $\mathbb{Z}$ has a $(2, 1, 4, 3)$-division algorithm with respect to the degree-like function $d(x) = \log_2 |x|$. Here $\mu = 0$ and $2\delta = \mu + \epsilon$ and we apply (4.3) (vii). It follows that q remains anisotropic over $\mathbb{F}_p$ or p is represented by q over $\mathbb{Q}$. The first case is ruled out since every form of dimension ≥ 3 over a finite field is isotropic. Hence, p is a sum of four squares in $\mathbb{Q}$.

References

[1] L. E. Dickson, *History of the theory of numbers. Vol. II: Diophantine analysis*, Chelsea Publishing Co., New York, 1966.

[2] R. Elman, N. Karpenko, and A. Merkurjev, *The algebraic and geometric theory of quadratic forms*, American Mathematical Society Colloquium Publications, vol. 56, American Mathematical Society, Providence, RI, 2008.

[3] A. Hurwitz, *Der Euklidische Divisionssatz in einem endlichen algebraischen Zahlkörper.*, Math. Z. **3** (1919), 123–126.

[4] J.-R. Joly, *Théorème des deux carrés dans un anneau de polynômes*, (1971), 113–117. Bull. Soc. Math. France, Mém. 25.

[5] I. Kersten, *Metrische vektorräume und quadratische formen*, Sonderforschungsbereich 343, 1996.

[6] M. Knebusch, *Ein Satz über die Werte von quadratischen Formen über Körpern*, Invent. Math. **12** (1971), 300–303.

[7] ———, *Specialization of quadratic and symmetric bilinear forms, and a norm theorem*, Acta Arith. **24** (1973), 279–299, Collection of articles dedicated to Carl Ludwig Siegel on the occasion of his seventy-fifth birthday, III.

[8] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics, vol. 67, American Mathematical Society, Providence, RI, 2005.

[9] H. W. Lenstra, Jr., *Euclidean number fields of large degree*, Invent. Math. **38** (1976/77), no. 3, 237–254.

[10] A. Pfister, *Quadratic forms with applications to algebraic geometry and topology*, London Mathematical Society Lecture Note Series, vol. 217, Cambridge University Press, Cambridge, 1995.

[11] W. Scharlau, *Quadratic and Hermitian forms*, Grundlehren der Mathematischen Wissenschaften, vol. 270, Springer-Verlag, Berlin, 1985.

[12] J.-P. Serre, *Extensions de corps ordonnés*, C. R. Acad. Sci. Paris **229** (1949), 576–577.

[13] T. A. Springer, *Sur les formes quadratiques d'indice zéro*, C. R. Acad. Sci. Paris **234** (1952), 1517–1519.

M. G. Mahmoudi, mmahmoudi@sharif.ir, Department of Mathematical Sciences, Sharif University of Technology, P. O. Box 11155-9415, Tehran, Iran. Fax: (+98) (21) 6616-5117