Cent. Eur. J. Math. • 6(3) • 2008 • 482-487
DOI: 10.2478/s11533-008-0038-4

VERSITA

Central European Journal of **Mathematics**

Rational points on the unit sphere

Research Article

Eric Schmutz*

Mathematics Department, Drexel University, Philadelphia, Pennsylvania, 19104

Received 21 January 2008; accepted 1 June 2008

Abstract: It is known that the unit sphere, centered at the origin in $\mathbb{R}^n$, has a dense set of points with rational coordinates. We give an elementary proof of this fact that includes explicit bounds on the complexity of the coordinates: for every point $\mathbf{v}$ on the unit sphere in $\mathbb{R}^n$, and every $\epsilon > 0$, there is a point $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ such that:

- $||\mathbf{r} - \mathbf{v}||_\infty < \epsilon$.
- $\mathbf{r}$ is also a point on the unit sphere; $\sum r_i^2 = 1$.
- $\mathbf{r}$ has rational coordinates; $r_i = \frac{a_i}{b_i}$ for some integers a_i, b_i.
- for all i, $0 \leq |a_i| \leq b_i \leq (\frac{32^{1/2}\lceil \log_2 n \rceil}{\epsilon})^{2\lceil \log_2 n \rceil}$.

One consequence of this result is a relatively simple and quantitative proof of the fact that the rational orthogonal group $O(n, \mathbb{Q})$ is dense in $O(n, \mathbb{R})$ with the topology induced by Frobenius' matrix norm. Unitary matrices in $U(n, \mathbb{C})$ can likewise be approximated by matrices in $U(n, \mathbb{Q}(i))$.

MSC: 11J99,14G05

Keywords: Diophantine approximation • orthogonal group • unitary group • rational points • unit sphere

1. Introduction

Suppose that M is a manifold embedded in $\mathbb{R}^n$, and that $\mathbf{v}$ is a point on M. Does M have points with rational coordinates that are as close as we like to $\mathbf{v}$? The answer may depend on the embedding; it is not an intrinsic property of M. If the rational points are dense, what is the tradeoff between the closeness of approximation and the sizes of the denominators that the rational coordinates must have? In general, such problems can be very difficult. This paper discusses two closely related special cases that are relatively easy: unit spheres centered at the origin, and real orthogonal matrices. See [13] for a deep algebraic treatment of weak approximation in algebraic varieties that is relevant to these examples. The methods of this paper are comparatively elementary, and the metric theory is not discussed. In other words, this

* *E-mail: Eric.Jonathan.Schmutz@drexel.edu*

paper provides crude but simple bounds that hold for all points v, rather than sophisticated bounds that hold for typical points. Problems of the latter type are addressed by the "Metric Diophantine Approximation on Manifolds" literature. (See, for example, [2], [1], [6].)

2. Rational points on the sphere

Although $\mathbb{Q}^2$ is dense in $\mathbb{R}^2$, it is not true in general that a circle in the plane has a dense set of rational points. For example, if $\xi = \sqrt[3]{2}$ (or any other real number that satisfies no quadratic polynomial in $\mathbb{Z}[x]$), then a unit circle centered at $(\xi, 0)$ has no rational points on it. Let S_ρ^{n-1} be the set of points in $\mathbb{R}^n$ for which the Euclidean distance from the origin is ρ. Humke and Krajewski characterized the radii ρ for which $S_\rho^1 \bigcap \mathbb{Q}^2$ is dense in S_ρ^1. In particular, the unit circle, centered at the origin in $\mathbb{R}^2$, does have a dense set of points with rational coordinates [5]. It is also known that, for all n, the unit sphere S_1^n has a dense set of rational points. (I thank Andy Hicks and Greg Naber for pointing out that the inverse of the stereographic projection takes rational points to rational points.) This paper provides a different proof that is "quantitative"in the sense that that it takes into account the complexity of the rational coordinates. If $r = \frac{a}{b}$ for some coprime integers a, b define the denominator $D(r)$ of r to be $|b|$. If we restrict the size of the denominator, it is apparently harder to get a good approximation. Our main interest is this tradeoff between closeness of approximation and size of the denominator.
We begin with the special case where the dimension of the ambient space is a power of two.

Lemma 2.1.

Suppose $0 < \gamma < .02$, *and suppose* $\alpha_1, \alpha_2, \ldots, \alpha_M$ *is a sequence of* $M = 2^m$ *real numbers such that* $\alpha_1^2 + \alpha_2^2 + \ldots \alpha_M^2 = 1$. *Then there are rational numbers* $r_i = \frac{a_i}{b_i}, i = 1, 2, \ldots, M$ *such that*

- $\sum_{i=1}^{M} r_i^2 = 1$, *and*
- *for all* i, $0 < b_i \leq (\frac{2}{\gamma^2})^m$, *and*
- *for all* i, $|r_i - \alpha_i| \leq 4\gamma m$.

Proof. The proof is by induction on m. The base case $m = 0$ is trivial since necessarily $\alpha_1^2 = 1$ and we can take $r_1 = \alpha_1$. Let $m > 0$, and assume the inductive hypothesis.
Let $\sigma_1 = \left(\sum_{i=1}^{M/2} \alpha_i^2\right)^{1/2}$, and let $\sigma_2 = \left(\sum_{i=\frac{M}{2}+1}^{M} \alpha_i^2\right)^{1/2}$. Note that

$$\sigma_1^2 + \sigma_2^2 = 1. \tag{1}$$

Without loss of generality, assume $\sigma_1 \geq \sigma_2$. This and (1) imply that

$$\sigma_2 \leq \frac{1}{\sqrt{2}}. \tag{2}$$

If $\sigma_2 = 0$, then we can put $r_i = 0$ for $i > \frac{M}{2}$, and apply the inductive hypothesis to $\alpha_1, \alpha_2, \ldots, \alpha_{M/2}$. We therefore assume that $\sigma_2 > 0$.
Adapting an argument of Humke and Krajewski [5], define $f(y) = \frac{2y}{1+y^2}$, and let y_* be the unique solution in $(0, 1)$ to the equation $f(y) = \sigma_2$. Thus $y_* = \sigma_2^{-1} - \sqrt{\sigma_2^{-2} - 1}$. By a well known theorem on Diophantine approximation [3], we can choose coprime integers k, ℓ such that $0 < \ell \leq \frac{1}{\gamma}$ and

$$|\frac{k}{\ell} - y_*| \leq \frac{\gamma}{\ell}. \tag{3}$$

With this choice of k and ℓ, define

$$R_2 = \frac{2k\ell}{k^2+\ell^2} = f(\frac{k}{\ell}) \tag{4}$$

and

$$R_1 = \sqrt{1-R_2^2} = \frac{\ell^2-k^2}{k^2+\ell^2}. \tag{5}$$

Ultimately R_1 and R_2 will serve as rational approximations for σ_1 and σ_2 respectively. But first we use them to define the rational numbers r_i. By the inductive hypothesis, applied to the numbers $\frac{\alpha_i}{\sigma_1}$, we can choose rational numbers $q_i, i = 1, 2, \ldots, \frac{M}{2}$ such that

$$\sum_{i=1}^{M/2} q_i^2 = 1 \tag{6}$$

$$\left|q_i - \frac{\alpha_i}{\sigma_1}\right| \le 4\gamma(m-1) \qquad \left(\text{for } i = 1, 2, \ldots, 2^{m-1}\right) \tag{7}$$

$$D(q_i) \le (\frac{2}{\gamma^2})^{m-1} \qquad \left(\text{for } i = 1, 2, \ldots, 2^{m-1}\right). \tag{8}$$

We can likewise choose rational numbers q_i, for $2^{m-1} < i \le 2^m$, such that

$$\sum_{i=2^{m-1}+1}^{M} q_i^2 = 1 \tag{9}$$

$$\left|q_i - \frac{\alpha_i}{\sigma_1}\right| \le 4\gamma(m-1) \qquad \left(\text{for } 2^{n-1} < i \le 2^m\right) \tag{10}$$

$$D(q_i) \le (\frac{2}{\gamma^2})^{m-1} \qquad \left(\text{for } 2^{m-1} < i \le 2^m\right). \tag{11}$$

For $i \le \frac{M}{2}$, define $r_i = R_1 q_i$. Similarly, for $i > \frac{M}{2}$, define $r_i = R_2 q_i$. Clearly $r_1, r_2, \ldots, r_M$ are rational. We must verify that these numbers satisfy the conditions stated in the lemma.
The first condition is that $(r_1, r_2, \ldots, r_M)$ is a point on the unit sphere. By (6) and(9),

$$\sum_{i=1}^{M} r_i^2 = R_1^2 \sum_{i=1}^{M/2} q_i^2 + R_2^2 \sum_{M/2+1}^{M} q_i^2 = R_1^2 + R_2^2 = 1. \tag{12}$$

The second condition is a bound on the denominators of the rational approximations. Note that, for all i, $D(r_i) \le (k^2+\ell^2)D(q_i) \le \frac{2}{\gamma^2} D(q_i)$. By (8) and (11), $D(q_i) \le \left(\frac{2}{\gamma^2}\right)^{m-1}$. Thus, for all i,

$$D(r_i) \le (\frac{2}{\gamma^2})^m. \tag{13}$$

For the third condition, begin with the observation that, for all $i > \frac{M}{2}$,

$$|r_i - \alpha_i| = |R_2 q_i - \sigma_2 q_i + \sigma_2 q_i - \alpha_i|$$

$$\le |q_i||R_2 - \sigma_2| + |\sigma_2||q_i - \frac{\alpha_i}{\sigma_2}|.$$

It is clear from (1) and (9) that $\sigma_2 \le 1$ and $|q_i| \le 1$. It follows by (10) that

$$|r_i - \alpha_i| \le |R_2 - \sigma_2| + 4\gamma(m-1) = |f(\frac{k}{\ell}) - f(y_*)| + 4\gamma(m-1). \tag{14}$$

By the mean value theorem, $|f(x) - f(y)| \le 2|x - y|$ for all $x, y \in (0, 1)$. Hence

$$|R_2 - \sigma_2| = |f(\frac{k}{\ell}) - f(y_*)| \le 2|\frac{k}{\ell} - y_*| \le \frac{2\gamma}{\ell} \le 2\gamma. \tag{15}$$

Putting this back into (14), we get, for all $i > \frac{M}{2}$,

$$|r_i - \alpha_i| \le 2\gamma + 4\gamma(m-1) < 4\gamma m.$$

Similarly, for $i \le 2^{m-1}$,

$$|r_i - \alpha_i| \le |R_1 - \sigma_1| + |q_i - \frac{\alpha_i}{\sigma_1}| \le |R_1 - \sigma_1| + 4\gamma(m-1). \tag{16}$$

Let $g(x) = \sqrt{1-x^2}$, so that,for some ξ between R_2 and σ_2, we have

$$|R_1 - \sigma_1| = |g(R_2) - g(\sigma_2)| = |g'(\xi)||R_2 - \sigma_2|.$$

Note that $|g'(x)| = x(1-x^2)^{-1/2}$ is increasing on $(0, 1)$. By (15) and (2), we have $R_2 \le \sigma_2 + 2\gamma \le \frac{1}{\sqrt{2}} + 2(.02) < .75$. Hence $|g'(\xi)| \le |g'(.75)| < 2$, and

$$|R_1 - \sigma_1| \le 2|R_2 - \sigma_2| + 4\gamma(m-1) \le 4\gamma m.$$

□

Now we can proceed to the general case of numbers that are not necessarily powers of two.

Theorem 2.1.
Suppose $\epsilon \in (0, .08)$, and suppose $\alpha_1, \alpha_2, \ldots, \alpha_n$ is a sequence of n real numbers such that $\alpha_1^2 + \alpha_2^2 + \ldots \alpha_n^2 = 1$. Then there are rational numbers $r_i = \frac{a_i}{b_i}, i = 1, 2, \ldots, m$ such that

- $\sum_{i=1}^{m} r_i^2 = 1$, *and*
- *for all i, $0 < b_i \le (\frac{32\lceil \log_2 n \rceil^2}{\epsilon^2})^{\lceil \log_2 n \rceil}$, and*
- *for all i, $|r_i - \alpha_i| < \epsilon$.*

Proof. Let $m = \lceil \log_2 n \rceil$. If n is not a power of two,then "pad with zeroes", i.e. define $\alpha_i = 0$ for $n < i \le 2^m$. Then Lemma 2.1 is directly applicable with $\gamma = \epsilon/4m$. □

A final comment is that there was a good reason to consider powers of two first. The advantage of working with $M = 2^m$ first is that the depth of the recursion is only $O(\log M)$, and consequently we get much better bounds for the denominators.

3. Approximating orthogonal matrices

First we fix some notation. For any $n \times n$ matrix $A = (a_{i,j})_{1 \le i,j \le n}$, let $||A||_2 = \sqrt{\sum_{i,j} a_{i,j}^2}$. Also let $||A||_\infty = \max_{i,j} |a_{i,j}|$. Similarly, for a vector $\mathbf{v}$, let $||\mathbf{v}||_\infty$ be the maximum of the components' magnitudes. For any subset $F \subseteq \mathbb{R}$, let $O(n, F)$ be the set of all $n \times n$ matrices A, with entries in F, for which the columns are orthonormal with respect to the standard inner product, i.e. for which $A^t = A^{-1}$.
It is apparently known that $O(n, \mathbb{Q})$ is dense in $O(n, \mathbb{R})$. The case $n = 3$ has been studied in some detail because of physics and engineering applications. See, for example, [10]. I do not know a suitable reference for general n, but Margulis [7] credits Platonov [11] with a proof that $SO(n, \mathbb{Z}[\frac{1}{5}])$ is dense in $SO(n, \mathbb{R})$. I cannot read [11], but Chapter

7 of [13] is relevant. Platonov's proof is not very accessible, and it is a non-trivial matter for a general mathematical reader to sort through the topologies. See [8], [9]. Theorem 3.1 below is a relatively simple proof that $O(n,\mathbb{Q})$ is dense in $O(n,\mathbb{R})$ in the topology induced by Frobenius' norm $||\cdot||_2$ (i.e. the subspace topology inherited from $\mathbb{R}^{n^2}$ when we regard $n\times n$ matrices as vectors in $\mathbb{R}^{n^2}$). It provides crude but explicit bounds on the denominators of the approximating matrices' entries. I do not know whether such bounds can be deduced from Platonov's work.

Theorem 3.1.
For any $n\times n$ real orthogonal matrix $T\in O(n,\mathbb{R})$, and any $\delta>0$, there is a rational orthogonal matrix $A\in O(n,\mathbb{Q})$ such that

- $||T-A||_2<\delta$.
- *Each entry of A has denominator less than* $(\frac{16\sqrt{2}n^2\lceil\log_2 n\rceil}{\delta})^{2n^2\lceil\log_2 n\rceil}$.

Proof. For any unit vector $\mathbf{u}$ (represented as an $n\times 1$ matrix), let $H_{\mathbf{u}}$ be the corresponding Householder matrix, i.e.

$$H_{\mathbf{u}}=I-2\mathbf{u}\mathbf{u}^t. \tag{17}$$

It is well known [14], [4] that, for some $h\le n$ and some unit vectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_h$, we have

$$T=\prod_{k=1}^{h}H_{\mathbf{u}_k}. \tag{18}$$

Let $\beta=(\frac{16\sqrt{2}n^2\lceil\log_2 n\rceil}{\delta})^{2\lceil\log_2 n\rceil}$. By Theorem 2.1 (with $\epsilon=\frac{\delta}{4n^2}$), we can, for each k, choose a unit vector $\mathbf{a}_k$ such that:

$$\mathbf{a}_k \text{ has rational coordinates} \tag{19}$$

$$||\mathbf{a}_k-\mathbf{u}_k||_\infty<\frac{\delta}{4n^2} \tag{20}$$

$$\text{each coordinate of } \mathbf{a}_k \text{ has denominator less than } \beta. \tag{21}$$

For this choice of unit vectors $\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_h$, let

$$A=\prod_{k=1}^{h}H_{\mathbf{a}_k}. \tag{22}$$

Clearly A and has rational coefficients. We must show that it is a good approximation for T. Beginning with (18) and (22), we can add and subtract a term to get

$$\begin{aligned}\Big|\Big|A-T\Big|\Big|_2&=\Big|\Big|H_{\mathbf{a}_1}(\prod_{k=2}^{h}H_{\mathbf{a}_k}-\prod_{k=2}^{h}H_{\mathbf{u}_k})+(H_{\mathbf{a}_1}-H_{\mathbf{u}_1})(\prod_{k=2}^{h}H_{\mathbf{u}_k})\Big|\Big|_2\\&\le\Big|\Big|H_{\mathbf{a}_1}(\prod_{k=2}^{h}H_{\mathbf{a}_k}-\prod_{k=2}^{h}H_{\mathbf{u}_k})\Big|\Big|_2+\Big|\Big|(H_{\mathbf{a}_1}-H_{\mathbf{u}_1})(\prod_{k=2}^{h}H_{\mathbf{u}_k})\Big|\Big|_2\end{aligned} \tag{23}$$

Because a Householder matrix is orthogonal, we have $\Big|\Big|H_{\mathbf{v}}B\Big|\Big|_2=\Big|\Big|B\Big|\Big|_2=\Big|\Big|BH_{\mathbf{v}}\Big|\Big|_2$ for any unit vector $\mathbf{v}$ and any $n\times n$ matrix B. Applying this repeatedly to (23), we get

$$\Big|\Big|A-T\Big|\Big|_2\le\Big|\Big|\prod_{k=2}^{h}H_{\mathbf{a}_k}-\prod_{k=2}^{h}H_{\mathbf{a}_u}\Big|\Big|_2+\Big|\Big|H_{\mathbf{a}_1}-H_{\mathbf{u}_1}\Big|\Big|_2.$$

The same argument can be applied again to the first term on the right. Thus, by iterating, we get

$$\Big|\Big|A-T\Big|\Big|_2 = \Big|\Big|\prod_{k=1}^{h} H_{\mathbf{a}_k} - \prod_{k=1}^{h} H_{\mathbf{u}_k}\Big|\Big|_2 \le \sum_{k=1}^{h}\Big|\Big|H_{\mathbf{a}_k} - H_{\mathbf{u}_k}\Big|\Big|_2. \tag{24}$$

For any $i \le n$, let $\mathbf{u}_k(i)$ and $\mathbf{a}_k(i)$ respectively be the i'th coordinates of the unit vectors $\mathbf{u}_k$ and $\mathbf{a}_k$. Then entry i,j of $H_{\mathbf{a}_k} - H_{\mathbf{u}_k}$ is

$$2\mathbf{u}_k(i)\Big(\mathbf{u}_k(j) - \mathbf{a}_k(j)\Big) + 2\mathbf{a}_k(j)\Big(\mathbf{u}_k(i) - \mathbf{a}_k(i)\Big).$$

Combining this with (20), we get

$$||H_{\mathbf{a}_k} - H_{\mathbf{u}_k}||_2 \le n||H_{\mathbf{a}_k} - H_{\mathbf{u}_k}||_\infty \le 4n||\mathbf{a}_k - \mathbf{u}_k||_\infty \le \frac{\delta}{n}. \tag{25}$$

Putting (25) back into (24), we get the half of the theorem: $||T-A||_2 < \delta$.

We still need to bound the sizes of the entries of the approximating matrix $A = \prod_{i=1}^{h} H_{\mathbf{a}_i}$. The h vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_h$ have collectively at most $hn \le n^2$ rational coordinates, each of which has denominator less than β. Therefore β^{n^2} is an upper bound for the least common multiple of the denominators of the coordinates, and consequently each entry of A has denominator less than β^{n^2}. □

Essentially the same argument can be carried out for unitary matrices. Any unitary matrix can $U \in U(n,\mathbb{C})$ can be written as a product of Householder matrices of the form $H = I - 2\mathbf{u}\mathbf{u}^*$, where $\mathbf{u}$ is a unit vector in $\mathbb{C}^n$ [14]. By identifying a unit vector $u \in \mathbb{C}^n$ with a unit vector in $\mathbb{R}^{2n}$, we can choose a unit vector $\mathbf{a}$ in $\mathbb{Q}(i)^n$ so that $||\mathbf{u}-\mathbf{a}||$ is is as small as we like.

References

[1] Beresnevich V.V., Bernik V.I., Kleinbock D.Y., Margulis G.A., Metric diophantine approximation: the Khintchine-Groshev theorem for nondegenerate manifolds, Mosc. Math. J., 2002, 2, 203–225

[2] Bernik V.I., Dodson M.M., Metric diophantine approximation on manifolds, Cambridge University Press, Cambridge, 1999

[3] Hardy G.H., Wright E.M., An introduction to the theory of numbers, 5th ed., Oxford University Press, Oxford, 1983

[4] Householder A., Unitary triangularization of a nonsymmetric matrix, J. ACM, 1958, 5, 339–342

[5] Humke P.D., Krajewski L.L., A characterization of circles which contain rational points, Amer. Math. Monthly, 1979, 86, 287–290

[6] Kleinbock D.Y., Margulis G.A., Flows on homogeneous spaces and diophantine approximation on manifolds, Ann. of Math.(2), 1998, 148, 339–360

[7] Margulis G.A., Some remarks on invariant means, Monatsh. Math., 1980, 90, 233-235

[8] Mazur B., The topology of rational points, Experiment. Math., 1992, 1, 35–45

[9] Mazur B., Speculations about the topology of rational points: an update, Astérisque, 1995, 228, 165–182

[10] Milenkovic V.J., Milenkovic V., Rational orthogonal approximations to orthogonal matrices, Comput. Geom., 1997, 7, 25–35

[11] Platonov V.P., The problem of strong approximation and the Kneser-Tits hypothesis for algebraic groups, Izv. Akad. Nauk SSSR Ser. Mat., 1969, 33, 1211–1219 (in Russian)

[12] Platonov V.P., A supplement to the paper "The problem of strong approximation and the Kneser-Tits hypothesis for algebraic groups", Izv. Akad. Nauk SSSR Ser. Mat., 1970, 34, 775–777 (in Russian)

[13] Platonov V.P., Rapinchuk A., Algebraic groups and number theory, Academic Press, Boston, 1994

[14] Uhlig F., Constructive ways for generating (generalized) real orthogonal matrices as products of (generalized) symmetries, Linear Algebra Appl., 2001, 332/334, 459–467