

REDUCTIONS BETWEEN SHORT VECTOR PROBLEMS AND SIMULTANEOUS APPROXIMATION

DANIEL E. MARTIN

ABSTRACT. In 1982, Lagarias showed that solving the approximate Shortest Vector Problem also solves the problem of finding simultaneous Diophantine approximations [21]. Here we provide a deterministic, dimension-preserving reduction in the reverse direction. It has polynomial time and space complexity, and it is gap-preserving under the appropriate norms. We also give an alternative to the Lagarias algorithm by first reducing the version of simultaneous approximation in [21] to one with no explicit range in which a solution is sought.

1. INTRODUCTION

Our primary result is to show that a short vector problem reduces deterministically and with polynomial complexity to a single simultaneous approximation problem as presented in the definitions below. We use \min^\times to mean nonzero minimum, $\{\mathbf{x}\} \in (-1/2, 1/2]^n$ to denote the fractional part of $\mathbf{x} \in \mathbb{R}^n$, and $[x]$ to denote the set $\{1, \dots, [x]\}$ for $x \in \mathbb{R}$.

Definition 1.1. A *short vector problem* takes input $\alpha \in [1, \infty)$ and nonsingular $M \in M_n(\mathbb{Z})$. A valid output is $\mathbf{q}_0 \in \mathbb{Z}^n$ with $0 < \|M\mathbf{q}_0\| \leq \alpha \min_{\mathbf{q} \in \mathbb{Z}^n}^\times \|M\mathbf{q}\|$. Let SVP denote an oracle for such a problem.

Definition 1.2. A *good Diophantine approximation problem* takes input $\alpha, N \in [1, \infty)$ and $\mathbf{x} \in \mathbb{Q}^n$. A valid output is $q_0 \in [\alpha N]$ with $\|\{q_0\mathbf{x}\}\| \leq \alpha \min_{q \in [N]} \|\{q\mathbf{x}\}\|$. Let GDA denote an oracle for such a problem.

Our reduction asserts that if we can find short vectors in a very restricted family of lattices then we can find them in general—behind a good Diophantine approximation problem is the lattice generated by \mathbb{Z}^n and just one additional vector, \mathbf{x} .

Literature more commonly refers to a short vector problem as a *Shortest Vector Problem* when $\alpha = 1$ and an *approximate Shortest Vector Problem* otherwise (often unrestricted to sublattices of \mathbb{Z}^n , though we have lost no generality). A brief exposition can be found in [26]. See [14] or [24] for a more comprehensive overview, [27] for a focus on cryptographic applications, [19] for a summary of hardness results, and [6] for relevance and potential applications to post-quantum cryptography.

Regarding simultaneous approximation, Brentjes highlights several algorithms in [7]. For a sample of applications to attacking clique and knapsack-type problems see [13], [20], and [31]. Examples of cryptosystems built on the hardness of simultaneous approximation are [2], [4], and [16]. This version is taken from [9] and [29].

Date: June 4, 2020.

2010 Mathematics Subject Classification. Primary: 52C07, 11H06, 68W25.

Key words and phrases. lattice reduction, Shortest Vector Problem, simultaneous Diophantine approximation.

The reduction, given in Algorithm 3, preserves the gap α when the ℓ_∞ -norm is used for both problems. This means the short vector problem defined by α and M is solved by calling $\text{GDA}(\alpha, \mathbf{x}, N)$ for some $\mathbf{x} \in \mathbb{Q}^n$ and $N \in \mathbb{R}$. It is the reverse of Lagarias’ 1982 reduction from good Diophantine approximation to SVP. (See Theorem B in [21], which refers to the problem as *good simultaneous approximation*. We borrow its name from [9] and [29].) Though there is an important contextual distinction: [21] relates simultaneous approximation under the ℓ_∞ -norm to lattice reduction under the ℓ_2 -norm, whereas *all reductions in this paper assume a consistent norm*.

Under Lagarias’ setup (the most common)—the ℓ_∞ -norm for GDA and the ℓ_2 -norm for SVP—we are not the first to go the reverse direction. In a seminar posted online from July 1, 2019, Agrawal presented an algorithm achieving this reduction which was complete less some minor details [1]. Tersely stated, he takes an upper triangular basis for a sublattice of \mathbb{Z}^n and transforms it inductively, using integer combinations and rigid rotations with two basis vectors at a time, into a lattice (a rotated copy of the original) whose short vectors can be found via simultaneous approximation. The short vector problem defined by α and M gets reduced to $\text{GDA}(\alpha/\sqrt{2n}, \mathbf{x}, N)$, called multiple times in order to account for the unknown minimal vector length which is used to determine \mathbf{x} .

In contrast, the reduction here takes a completely different approach. It finds a sublattice which is nearly scaled orthonormal, so that only one additional vector is needed to generate the original lattice. This extra vector is the input for GDA. We note that when switching between norms, our reduction is also not gap-preserving. To use Algorithm 3 to solve a short vector problem with respect to the ℓ_2 -norm via GDA with respect to the ℓ_∞ -norm, the latter must be executed with the parameter α/\sqrt{n} to account for the maximum ratio of nonzero norms $\|\mathbf{q}\|_2/\|\mathbf{q}\|_\infty$.

The relationship between the two problems in Definitions 1.1 and 1.2 will be studied through the following intermediary.

Definition 1.3. A *simultaneous approximation problem* takes input $\alpha \in [1, \infty)$ and $\mathbf{x} \in \mathbb{Q}^n$. A valid output is $q_0 \in \mathbb{Z}$ with $0 < \|\{q_0\mathbf{x}\}\| \leq \alpha \min_{q \in \mathbb{Z}} \|\{q\mathbf{x}\}\|$. Let SAP denote an oracle for such a problem.

This problem prohibits only the trivial solution, the least common denominator of \mathbf{x} ’s entries, while “ N ” in a good Diophantine approximation problem is generally more restrictive.

Section 2 explores the relationship between the two versions of simultaneous approximation given in Definitions 1.2 and 1.3. Among the results, only Proposition 2.1 in Subsection 2.1 is required to verify the final reduction of a short vector problem to either version of simultaneous approximation. Subsection 2.2 contains Algorithm 1. It reduces a good Diophantine approximation problem to polynomially many SAP calls, each executed with the parameter $\alpha/3.06$. So while this reduction is not gap-preserving, the inflation is independent of the input.

Section 3 reduces both versions of simultaneous approximation to SVP. It begins with Algorithm 2, which solves Definition 1.3’s version. We remark at the end of Subsection 3.1 how this reduction adapts to the inhomogeneous forms of these problems, meaning the search for $q_0 \in \mathbb{Z}$ or $\mathbf{q}_0 \in \mathbb{Z}^n$ that makes $q_0\mathbf{x} - \mathbf{y}$ or $M\mathbf{q}_0 - \mathbf{y}$ small for some $\mathbf{y} \in \mathbb{Q}^n$. (In this case the latter is known as the *approximate Closest Vector Problem*. See chapter 18 of [14], for example.) Then Subsection 3.2 combines

Algorithms 1 and 2 to solve Definition 1.2’s version of simultaneous approximation using SVP. This is our alternative to the Lagarias reduction.

Finally, Algorithm 3 in Section 4 reduces a short vector problem to GDA or SAP. It also adapts to the inhomogeneous versions of SVP and SAP (not GDA, as mentioned at the end of Subsection 4.3). In Corollary 4.9 we observe that Algorithm 3 facilitates a simpler proof that GDA is NP-hard under an appropriate bound on α , a result first obtained in [9]. Then we combine Algorithms 2 and 3 in Subsection 4.2 to solve a simultaneous approximation problem with GDA. In particular, we give all six reductions among the defined problems, as shown in the diagram below.

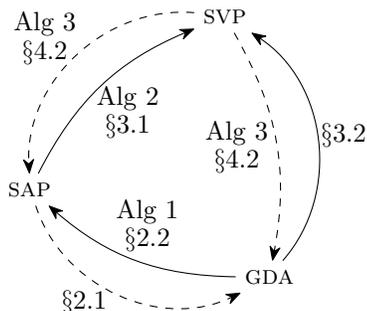


FIGURE 1. Algorithm and subsection number of reductions.

The two reductions in Figure 1 without algorithm numbers are achieved by following the two arrows that combine to give the same source and target. *Dashed arrows indicate a norm restriction. Each must be executed under either the ℓ_1 , ℓ_2 , or ℓ_∞ -norm.* However, we point out in Subsection 4.3 how the restriction can be alleviated to any ℓ_p -norm provided we accept additional gap inflation by a constant arbitrarily close to 1.

The results are summarized by the following table. It uses m and d to denote the maximal magnitude among input integers and the least common denominator of the input vector, respectively. The matrix or vector dimension is n , and p defines the norm. Trivial cases that cause logarithms to equal 0 are ignored. Column descriptions follow.

The results are summarized by the following table. It uses m and d to denote the maximal magnitude among input integers and the least common denominator of the input vector, respectively. The matrix or vector dimension is n , and p defines the norm. Trivial cases that cause logarithms to equal 0 are ignored. Column descriptions follow.

Reduction	Operations	Integers	Inflation	Calls
GDA \rightarrow SAP	$n \log m$	$n \log m$	3.06	$\lceil \log_2 d / \alpha N \rceil$
SAP \rightarrow SVP	$(n + \log m)^2$	$n \log m$	1	1
GDA \rightarrow SVP	$(n + \log m)^2$	$n \log m$	3.06	$\lceil \log_2 d / \alpha N \rceil$
SVP \rightarrow GDA	$n^4 \log mn$	$n^4 \log mn$	$n^{1/p}$	1
SVP \rightarrow SAP	$n^4 \log mn$	$n^4 \log mn$	1	1
SAP \rightarrow GDA	$n^5 \log m$	$n^5 \log m$	$n^{1/p}$	1

TABLE 1. Summary of reduction complexities and gap inflations.

Operations: Big- O bound on the number of arithmetic operations per oracle call.

Integers: Big- O bound on the length of integers used throughout the reduction.

Inflation: Maximum gap inflation. For example, to solve a good Diophantine approximation problem with some α using Algorithm 1, SAP is called with $\alpha/3.06$.

Calls: Upper bound on the number of required calls to the oracle.

2. VERSIONS OF SIMULTANEOUS APPROXIMATION

2.1. SAP to GDA. Rather than give a complete reduction from a simultaneous approximation problem to GDA, which is postponed until the end of Subsection 4.2, the purpose of this subsection is to observe a condition on the input that makes these two versions of simultaneous approximation nearly equivalent.

Proposition 2.1. *Suppose the i^{th} coordinate of \mathbf{x} is of the form $x_i = 1/d$, where $d \in \mathbb{N}$ makes $d\mathbf{x} \in \mathbb{Z}^n$. Under an ℓ_p -norm, $\text{GDA}(\alpha, \mathbf{x}, N)$ solves the simultaneous approximation problem defined by $\alpha n^{1/p}$ and \mathbf{x} with $N = d/2\alpha$.*

Proof. Let $q_{\min} \in [d/2]$ be such that $\|\{q_{\min}\mathbf{x}\}\|$ is the nonzero minimum. A vector's fractional part is in $(-1/2, 1/2]^n$, making its length at most $n^{1/p}/2$. So we may assume that $\|\{q_{\min}\mathbf{x}\}\| < 1/2\alpha$, since otherwise every integer in $[N] = [d/2\alpha]$ solves the simultaneous approximation problem defined by $\alpha n^{1/p}$ and \mathbf{x} .

Under an ℓ_p -norm, $\|\{q_{\min}\mathbf{x}\}\|$ is an upper bound for its i^{th} coordinate, q_{\min}/d . Combined with the assumption $\|\{q_{\min}\mathbf{x}\}\| < 1/2\alpha$, this gives $q_{\min} \in [d/2\alpha] = [N]$, which implies $\min_{q \in [N]} \|\{q\mathbf{x}\}\| = \min_{q \in \mathbb{Z}}^{\times} \|\{q\mathbf{x}\}\|$. And because $\alpha N < d$, it is guaranteed that $\text{GDA}(\alpha, \mathbf{x}, N)$ is not a multiple of d . \square

Note that without an assumption on \mathbf{x} like the one used in this proposition, there is no natural choice for N that makes GDA solve a simultaneous approximation problem. If we set it too small, say with $N < d/2$, then $\min_{q \in [N]} \|\{q\mathbf{x}\}\|$ may be unacceptably larger than $\min_{q \in \mathbb{Z}}^{\times} \|\{q\mathbf{x}\}\|$, potentially making GDA's approximation poor. If we set it too large, say with $N \geq d/\alpha$, then GDA may return d , which is not a valid output for the initial simultaneous approximation problem.

To get around this, our strategy is to first reduce a simultaneous approximation problem to SVP with Algorithm 2. Then in Algorithm 3, which reduces a short vector problem to SAP, we are careful to produce an input vector for the oracle that satisfies the hypothesis of Proposition 2.1 in order to admit GDA.

2.2. GDA to SAP. Let d continue to denote least common denominator of \mathbf{x} . The problem faced in this reduction is that outputs for a good Diophantine approximation problem are bounded by αN , which may be smaller than $d/2$. This leaves no guarantee that $\text{SAP}(\alpha, \mathbf{x})$, call this integer $d_1 \in [d/2]$, is a solution. But knowing that \mathbf{x} is very near a rational vector \mathbf{x}_1 with least common denominator d_1 allows us to call SAP again, now on \mathbf{x}_1 to get $d_2 \in [d_1/2]$. This is the least common denominator of some \mathbf{x}_2 near \mathbf{x}_1 , and we continue in this fashion until the bound αN has been reached. To facilitate this, we adopt the convention that modular reduction returns an integer with magnitude at most half the modulus.

Algorithm 1: A reduction from a good Diophantine approximation problem to multiple calls to SAP under a consistent norm.

input: $\alpha, N \in [1, \infty)$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Q}^n$

output: $q_0 \in [\alpha N]$ with $\|\{q_0\mathbf{x}\}\| \leq \alpha \min_{q \in [N]} \|\{q\mathbf{x}\}\|$

1: $d \leftarrow \text{lcd}(x_1, \dots, x_n) > 0$

2: **while** $d > \alpha N$ **do**

3: $d \leftarrow |\text{SAP}(\alpha/3.06, \mathbf{x}) \bmod d|$ \triangleright good, but large approximation

4: $\mathbf{x} \leftarrow \mathbf{x} - \{\mathbf{d}\mathbf{x}\}/d$ \triangleright now $\text{lcd}(\mathbf{x}) = d$, at most half of

5: **return** d the previous iteration's lcd

Proposition 2.2. *The output of Algorithm 1 solves the initial good Diophantine approximation problem.*

Proof. Let d_i and \mathbf{x}_i denote the values of “ d ” and “ \mathbf{x} ” after i **while** loop iterations have been completed. In particular, d_0 and \mathbf{x}_0 are defined by the input. Also let $I + 1$ be the total number of iterations executed, so the output is d_{I+1} .

The triangle inequality gives

$$(2.1) \quad \|\{d_{I+1}\mathbf{x}\}\| \leq \|\{d_{I+1}\mathbf{x}_I\}\| + d_{I+1} \sum_{i=1}^I \|\mathbf{x}_i - \mathbf{x}_{i-1}\|.$$

With $\lambda_i = \min_{q \in [N]} \|\{q\mathbf{x}_i\}\|$, the choice of d_{I+1} bounds the first summand by $\alpha\lambda_I/c$, where $c = 3.06$ in the algorithm but is left undetermined for now. Similarly, the choice of $d_i = \text{SAP}(\alpha/c, \mathbf{x}_{i-1})$ and the fact that $d_{i-1} > \alpha N \geq N$ make

$$(2.2) \quad \|\mathbf{x}_i - \mathbf{x}_{i-1}\| = \frac{\|\{d_i\mathbf{x}_{i-1}\}\|}{d_i} \leq \frac{\alpha \min_{q \in \mathbb{Z}} \|\{q\mathbf{x}_{i-1}\}\|}{cd_i} \leq \frac{\alpha\lambda_{i-1}}{cd_i}.$$

So to bound (2.1) it must be checked that the λ_i 's are not too large. To this end, fix some $i \leq I$ and let $q_{\min} \in [N]$ satisfy $\|\{q_{\min}\mathbf{x}_{i-1}\}\| = \lambda_{i-1}$. Then we have the following upper bound on λ_i , where the three inequalities are due to the triangle inequality, inequality (2.2), and $q_{\min} \leq N < d_I/\alpha \leq d_i/2^{I-i}\alpha$, respectively.

$$\|\{q_{\min}\mathbf{x}_i\}\| \leq \lambda_{i-1} + q_{\min}\|\mathbf{x}_i - \mathbf{x}_{i-1}\| \leq \lambda_{i-1} \left(1 + \frac{\alpha q_{\min}}{cd_i}\right) < \lambda_{i-1} \left(1 + \frac{1}{2^{I-i}c}\right).$$

Inductively, this gives

$$(2.3) \quad \lambda_i < \lambda_0 \prod_{j=1}^i \left(1 + \frac{1}{2^{I-j}c}\right).$$

Now the three numbered inequalities above can be combined to get

$$\|\{d_{I+1}\mathbf{x}\}\| \leq \frac{\alpha d_{I+1}}{c} \sum_{i=0}^I \frac{\lambda_i}{d_{i+1}} \leq \frac{\alpha}{c} \sum_{i=0}^I \frac{\lambda_i}{2^{I-i}} \leq \frac{\alpha\lambda_0}{c} \sum_{i=0}^I \frac{1}{2^{I-i}} \prod_{j=1}^i \left(1 + \frac{1}{2^{I-j}c}\right).$$

Thus the output approximation quality, $\|\{d_{I+1}\mathbf{x}\}\|$, is at most $\alpha \min_{q \in [N]} \|\{q\mathbf{x}\}\| = \alpha\lambda_0$ provided c satisfies

$$1 \geq \frac{1}{c} \sum_{i=0}^{\infty} \frac{1}{2^i} \prod_{j=i}^{\infty} \left(1 + \frac{1}{2^j c}\right).$$

This justifies our choice of $c = 3.06$ in line 3. \square

Proposition 2.3. *Let $m > 1$ be the maximum magnitude among integers defining \mathbf{x} , and let $d > 1$ be its least common denominator. The reduction in Algorithm 1 requires an initial $O(n \log m)$ operations plus $O(n)$ operations for each call to SAP, of which there are at most $\lceil \log_2(d/\alpha N) \rceil$, on integers of length $O(n \log m)$.*

Proof. Repeated applying the Euclidean algorithm computes d with $O(n \log m)$ operations on integers of length $O(n \log m)$. Modular reduction in line 3 decreases each successive least common denominator by at least a factor of $1/2$. This bounds the number of **while** loop iterations by $\lceil \log_2(d/\alpha N) \rceil$. \square

3. REDUCING TO SVP

First we restrict attention to Definition 1.3's version of simultaneous approximation (SAP) in Algorithm 2. Then we will compare the combination with Algorithm 1 to Lagarias' reduction in [21] from a good Diophantine approximation problem.

3.1. SAP to SVP. Here we replace the $n + 1$ vectors associated to simultaneous approximation, namely \mathbf{x} and a basis for \mathbb{Z}^n , with n vectors generating the same lattice. There are algorithms for which this a byproduct, like Pohst's modified (to account for linearly dependent vector inputs) LLL algorithm [23] or Kannan and Bachem's Hermite normal form algorithm [18]. But as a consequence of achieving additional basis properties, they are more complicated and require more operations than necessary. We briefly present an alternative because the improved time complexity is relevant to the next subsection.

Algorithm 2: A gap-preserving reduction from a simultaneous approximation problem to one call to SVP under a consistent norm.

input: $\alpha \in [1, \infty)$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{Q}^n$

output: $q_0 \in \mathbb{Z}$ with $0 < \|q_0 \mathbf{x}\| \leq \alpha \min_{q \in \mathbb{Z}} \|q \mathbf{x}\|$

1: $d \leftarrow \text{lcd}(x_1, \dots, x_n)$

2: $x_n \leftarrow x_n + a$ with a an integer that makes $\text{gcd}(dx_1, \dots, dx_{n-1}, d(x_n + a)) = 1$ ▷ make sure $d\mathbf{x}$ extends to a basis for \mathbb{Z}^n

3: $M \leftarrow \text{SL}_n(\mathbb{Z})$ matrix with first column $d\mathbf{x}$

4: $M \leftarrow M$ with last $n-1$ columns scaled by d ▷ generates scaled original lattice

5: **return** $\text{SVP}(\alpha, M)_1$ ▷ first coordinate is a solution

Proposition 3.1. *The output of Algorithm 2 solves the initial simultaneous approximation problem.*

Proof. First note that a in line 2 exists. As d is the *least* common denominator, $\text{gcd}(dx_1, \dots, dx_n)$ and d are coprime. So take a to be divisible by those primes which divide $\text{gcd}(dx_1, \dots, dx_{n-1})$ but not dx_n . Also, since a is an integer, the new value of \mathbf{x} defines the same simultaneous approximation problem as the input.

Coprime entries means \mathbf{x} extends to some $M \in \text{SL}_n(\mathbb{Z})$. (One method is mentioned in the next proof.) Now, the columns of dM generate $d\mathbb{Z}^n$, so the same is true if we only scale the last $n-1$ columns by d . In particular, the columns of the new M in line 4 generate $d\mathbf{x}$ and $d\mathbb{Z}^n$, which in turn generate each column. Thus M defines a basis for the original simultaneous approximation lattice scaled by d .

Finally, the last $n-1$ columns of M are vectors in $d\mathbb{Z}^n$, so that $M\text{SVP}(\alpha, M) \equiv \text{SVP}(\alpha, M)_1 d\mathbf{x} \pmod{d\mathbb{Z}^n}$. This verifies that $\text{SVP}(\alpha, M)_1$ is the integer we seek. \square

Proposition 3.2. *Let $m > 1$ be the maximum magnitude among integers defining \mathbf{x} . The reduction in Algorithm 2 requires $O((n + \log m)^2)$ operations on integers of length $O(n \log m)$.*

Proof. As with Algorithm 1, line 1 requires $O(n \log m)$ operations on and resulting in integers of length $O(n \log m)$.

Skipping line 2 for now, the i^{th} column (for $i \geq 2$) of M in line 3 can be set to

$$\left(\frac{b_1 dx_1}{\text{gcd}(dx_1, \dots, dx_{i-1})}, \dots, \frac{b_1 dx_{i-1}}{\text{gcd}(dx_1, \dots, dx_{i-1})}, b_2, 0, \dots, 0 \right),$$

where $b_2 \text{gcd}(dx_1, \dots, dx_{i-1}) - (-1)^i b_1 dx_i = \text{gcd}(dx_1, \dots, dx_i)$. The determinant of the top, left $i \times i$ minor is then $\text{gcd}(dx_1, \dots, dx_i)$ by induction. To find b_1 and b_2 we execute the Euclidean algorithm on $\text{gcd}(d_i x_1, \dots, d_i x_{i-1})$ and $d_i x_i$, where $d_i = \text{lcd}(x_1, \dots, x_i)$. But $\text{gcd}(d_i x_1, \dots, d_i x_{i-1})$ is at most m times $\text{gcd}(d_{i-1} x_1, \dots, d_{i-1} x_{i-1})$,

which divides the greatest common divisor of the numerators of x_1, \dots, x_i . So for each i the Euclidean algorithm needs $O(\log m)$ operations.

Before computing the last column of M , we find a in line 2 to ensure a determinant of 1. As discussed in the last proof, we can start with $a = \gcd(dx_1, \dots, dx_{n-1})$ and replace it with $a/\gcd(a, dx_n)$ until nothing changes. This requires $O(\log a) = O(\log m)$ executions of the Euclidean algorithm, each taking $O(\log m)$ operations.

Scaling all but the first column by d in line 4 takes $n^2 - n$ operations. \square

We remark that this algorithm adapts to inhomogeneous forms of these problems. To find $q_0 \in \mathbb{Z}$ with $0 < \|q_0 \mathbf{x} - \mathbf{y}\| \leq \alpha \min_{q \in \mathbb{Z}} \|q \mathbf{x} - \mathbf{y}\|$ when $q \mathbf{x} - \mathbf{y} \in \mathbb{Z}^n$ has no solution, we can perform the same reduction and finish by calling an oracle which solves the approximate Closest Vector Problem defined by α , M , and $d\mathbf{y}$.

3.2. GDA to SVP. Combining Algorithms 1 and 2 gives an alternative to the Lagarias reduction from good Diophantine approximation to SVP in [21]. We execute Algorithm 1, but use Algorithm 2 to compute $\text{SAP}(\alpha/3.06, \mathbf{x})$ in line 3. By Proposition 2.3, this requires at most $\lceil \log_2(d/\alpha N) \rceil$ calls to SVP. And Proposition 3.2 states that each call requires $O((n + \log m)^2)$ operations on integers of length $O(n \log m)$.

Recall that switching from ℓ_2 to ℓ_∞ decreases a nonzero norm by at most a factor of $1/\sqrt{n}$. In particular, by executing this combination of Algorithms 1 and 2 with respect to the ℓ_2 -norm, we get an ℓ_∞ solution to the initial good Diophantine approximation problem provided we use $\alpha/3.06\sqrt{n}$ for SVP.

Lagarias achieves this reduction with the now well-known trick from [22] of reducing the lattice generated by \mathbb{Z}^n and \mathbf{x} , bumped up a dimension by putting 0 in every $(n + 1)$ th coordinate but \mathbf{x} 's. The ideal value for the last coordinate of \mathbf{x} , which is guessed at using $\lfloor n + \log_2 dN \rfloor$ calls of the form $\text{SVP}(\alpha/\sqrt{5n}, M)$ for varying M , is $\min_{q \in [N]} \|q \mathbf{x}\|/N$. (The gap inflation approaches \sqrt{n} as our guesses get better.) The Lagarias reduction requires an initial $O(n \log m)$ arithmetic operations to compute the least common denominator, d , then only one additional operation per call. The integers involved have input length $O(\log m^n N)$.

Whether the benefit of fewer calls to SVP outweighs the increased operations per call depends on the complexity of the oracle. Ours is an asymptotic improvement when the number of operations performed by SVP exceeds $O((n + \log m)^2)$.

4. REDUCING TO GDA OR SAP

We begin again by focusing first on the reduction to SAP.

4.1. Intuition. Consider an input matrix $M \in M_n(\mathbb{Z})$ for a short vector problem. Let $d = \det M$, and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ denote the standard basis vectors for \mathbb{Z}^n . If there were one vector, call it $\mathbf{b} \in \mathbb{Z}^n$, for which the set $\{M\mathbf{b}, d\mathbf{e}_1, \dots, d\mathbf{e}_n\}$ generated the columns of M , our reduction would just amount to finding it. This is exactly the setup for simultaneous approximation: $n + 1$ vectors, n of which are scaled orthonormal. A solution could be obtained by doing simultaneous approximation on $M\mathbf{b}/d$, scaling the resulting short vector by d , and applying M^{-1} (to comply with Definition 1.1). Unfortunately, unless $n \leq 2$ or $d = \pm 1$, such a \mathbf{b} does not exist. Indeed, the adjugate matrix, $\text{adj } M = d M^{-1}$, has at most rank 1 over $\mathbb{Z}/p\mathbb{Z}$ for a prime p dividing d . So at least $n - 1$ additional vectors are required to have full rank modulo p , a prerequisite to having full rank over \mathbb{Q} . But asking that $M\mathbf{b}$ generate the columns of M alongside $d\mathbf{e}_1, \dots, d\mathbf{e}_n$ is equivalent to asking that \mathbf{b} generate \mathbb{Z}^n alongside the columns of $\text{adj } M$.

What mattered is the matrix with columns $d\mathbf{e}_1, \dots, d\mathbf{e}_n$ being scaled orthonormal. As such, multiplying by it or its inverse has no effect on a vector's relative length. So we plan to find a different set of n column vectors—a set for which just one additional $M\mathbf{b}$ is needed to generate the original lattice—which is nearly scaled orthonormal, making the effect of its corresponding matrix multiplication on α negligible. The initial short vector problem becomes a search for an integer combination of $M\mathbf{b}$ and these columns, say $\mathbf{c}_1, \dots, \mathbf{c}_n$. We can therefore solve the simultaneous approximation problem defined by α and $[\mathbf{c}_1 \cdots \mathbf{c}_n]^{-1}M\mathbf{b}$. This works as long as multiplying by $[\mathbf{c}_1 \cdots \mathbf{c}_n]$ changes the ratio between the lengths of the shortest vector and our output by less than whatever is afforded by the fact that lattice norms form a discrete set.

An arbitrary lattice may have all of its scaled orthonormal sublattices contained in $d\mathbb{Z}^n$. So as candidates for the matrix $[\mathbf{c}_1 \cdots \mathbf{c}_n]$, we look for something of the form $cd\text{Id} + MA = M(c\text{adj}M + A)$ for some $c \in \mathbb{Z}$ and $A \in M_n(\mathbb{Z})$. If the entries of A are sufficiently small, then multiplication by this matrix has a similar effect on relative vector norms as multiplying by $cd\text{Id}$, which has no effect.

We will tailor our choice of c and A so that a coordinate of the simultaneous approximation vector, $(c\text{adj}M + A)^{-1}\mathbf{b}$, is $1/\det(c\text{adj}M + A)$. This admits Proposition 2.1 and hence GDA.

4.2. SVP to GDA or SAP. Algorithm 3 uses the following.

Notation 4.1. For polynomials $f_1 = \sum_i f_{1,i}x^i$ and $f_2 = \sum_i f_{2,i}x^i$ with maximum degree d , let $C(f_1, f_2)$ denote the matrix of their coefficients,

$$\begin{bmatrix} f_{1,d} & & 0 & f_{2,d} & & 0 \\ \vdots & \ddots & & \vdots & \ddots & \\ f_{1,1} & \cdots & f_{1,d} & f_{2,1} & \cdots & f_{2,d} \\ f_{1,0} & \cdots & f_{1,d-1} & f_{2,0} & \cdots & f_{2,d-1} \\ & \ddots & \vdots & & \ddots & \vdots \\ 0 & & f_{1,0} & 0 & & f_{2,0} \end{bmatrix}.$$

The matrix above can determine when f_1 and f_2 are coprime over $\mathbb{Q}(x)$ in lieu of polynomial long division, where coefficient growth is exponential without complicated mitigations as in [8]. We demonstrate this now to give some clarity to the meaning behind lines 5 and 6 of Algorithm 3.

Lemma 4.2. *Let $f_1, f_2 \in \mathbb{Z}[x]$, not both constant. As an ideal in $\mathbb{Z}[x]$, (f_1, f_2) contains $\det C(f_1, f_2)$, which is nonzero if and only if f_1 and f_2 have no common root in the algebraic closure of \mathbb{Q} .*

Proof. Let $d = \max(\deg f_1, \deg f_2)$. Consider the vector in \mathbb{Z}^{2d} whose only (perhaps) nonzero entry is $\det C(f_1, f_2)$ in the last coordinate. This is the image under $C(f_1, f_2)$ of some nonzero integer vector. We can split the entries of this vector down the middle to get coefficients for $g_1, g_2 \in \mathbb{Z}[x]$ that have degree at most $d-1$ and satisfy $\det C(f_1, f_2) = f_1g_1 + f_2g_2 \in (f_1, f_2)$.

Plugging a common root of f_1 and f_2 into this last equation, should one exist, shows that $\det C(f_1, f_2) = 0$. Conversely, suppose $f_1g_1 + f_2g_2 = 0$ and that $\deg f_1 = d \geq 1$. Then g_2 must be nonzero to avoid the same being true of g_1 , contradicting our choice of nonzero coefficient vector. But g_2 has degree at most $d-1$. So $f_1g_1 = -f_2g_2$ implies that at least one of f_1 's d roots must be shared by f_2 . \square

Notation 4.3. For a matrix M , let $M_{i,j}$ denote the entry in its i^{th} row and j^{th} column, and let \check{M}_i denote its top, left $i \times i$ minor.

Line 1 of the next algorithm requires knowing the position of a nonzero entry in the input matrix, and line 8 requires knowing the maximum magnitude among entries. For notational convenience, we assume that $M_{n,1}$ is the nonzero maximum.

Algorithm 3: A reduction from a short vector problem with $n \geq 2$ to one call to SAP (gap-preserving) or GDA under a consistent ℓ_p -norm with $p \in \{1, 2, \infty\}$.

input: $a \geq b \in \mathbb{N}$ ($\alpha = a/b$), $M \in M_n(\mathbb{Z})$ with $0 \neq \det M$ and $M_{n,1} = \max_{i,j} |M_{i,j}|$
output: $\mathbf{q}_0 \in \mathbb{Z}^n$ with $0 < \|M\mathbf{q}_0\| \leq \alpha \min_{\mathbf{q} \in \mathbb{Z}^n}^\times \|M\mathbf{q}\|$

- 1: $p \leftarrow$ least prime not dividing $M_{n,1} \det M$
- 2: $M \leftarrow x \operatorname{adj} M + p \operatorname{Id}$ $\triangleright M = M(x)$ has linear polynomial entries
- 3: **for** $i \leftarrow 2$ **to** n **do** \triangleright need not compute determinant to test each j —see Theorem 4.8
- 4: $M_{i,1} \leftarrow M_{i,1} + p$
- 5: $M_{i,i-1} \leftarrow M_{i,i-1} + p^j$ with $j > 0$ minimal \triangleright need not compute determinant to test each j —see Theorem 4.8
so $\det C((\operatorname{adj} \check{M}_i)_{i,1}, (\operatorname{adj} \check{M}_i)_{i,2}) \neq 0$
- 6: $c \leftarrow \det C((\operatorname{adj} M)_{n,1}, (\operatorname{adj} M)_{n,2})$
- 7: $c \leftarrow c/p^j$ with j maximal or $p+1$ if $|c| = p^j$ \triangleright make c coprime to p
- 8: $M \leftarrow M(c^j)$ with $j = \lceil \log_{|c|} a^2 (2M_{n,1}n)^{3n} \rceil$ \triangleright substitute for x so $M \in M_n(\mathbb{Z})$
- 9: $b_1, b_2 \leftarrow$ integers with $|b_1|$ minimal so $1 = b_1(\operatorname{adj} M)_{n,1} + b_2(\operatorname{adj} M)_{n,2}$ \triangleright that these exist guarantees $M\mathbf{x}$ (line 10) and M generate \mathbb{Z}^n
- 10: $\mathbf{x} \leftarrow M^{-1}(b_1, b_2, 0, \dots, 0)$
- 11: $q_0 \leftarrow$ SAP(α, \mathbf{x}) or GDA($\alpha/n^{1/p}, \mathbf{x}, N$) with $N = n^{1/p} \det M/2\alpha$ \triangleright GDA works since $x_n = 1/\det M$, recall Proposition 2.1
- 12: **return** $M\{q_0\mathbf{x}\}$

Let us show that the **for** loop can always be completed.

Lemma 4.4. *For $i = 2, \dots, n$, there is some $j \leq 2i - 2$ satisfying the criterion of line 5 in the **for** loop iteration corresponding to i .*

Proof. When $i = 2$ we are asked to find j for which the linear polynomials $M_{1,1}$ and $M_{2,1} + p^j$ do not share a root (by Lemma 4.2). The constant term of $M_{1,1}$ is p by line 2, meaning it has at most one root. So asking that $j \leq 2i - 2 = 2$ gives enough space to avoid the at-most-one value of j that fails. Now suppose $i \geq 3$ and that the claim holds for $i - 1$. Let M be its value after line 4 in the **for** loop iteration corresponding to i , and let $f_1 = (\operatorname{adj} \check{M}_{i-1})_{i-1,1}$ and $f_2 = (\operatorname{adj} \check{M}_{i-1})_{i-1,2}$.

By assumption there are $g_1, g_2 \in \mathbb{Z}[x]$ with $g_1 f_1 + g_2 f_2 = \det C(f_1, f_2) \neq 0$. Fix an integer j , and let $h_1 = (\operatorname{adj} \check{M}_i)_{i,1} - p^j f_1$ and $h_2 = (\operatorname{adj} \check{M}_i)_{i,2} - p^j f_2$, the polynomials we hope to make coprime with the appropriate choice of j . We have

$$\begin{bmatrix} f_2 & -f_1 \\ g_1 & g_2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} f_2(\operatorname{adj} \check{M}_i)_{i,1} - f_1(\operatorname{adj} \check{M}_i)_{i,2} \\ g_1(\operatorname{adj} \check{M}_i)_{i,1} + g_2(\operatorname{adj} \check{M}_i)_{i,2} - p^j \det C(f_1, f_2) \end{bmatrix}.$$

In the column on the right, where we now focus our attention, p^j has been isolated.

For each root of the top polynomial, there is at most one value of j that makes it a root of the bottom. Thus it suffices to show that $f_2(\operatorname{adj} \check{M}_i)_{i,1} - f_1(\operatorname{adj} \check{M}_i)_{i,2}$ is not the zero polynomial. Then its degree, which is at most $2i - 3$, bounds how many values of j can make the right-side polynomials share a root. As this occurs whenever h_1 and h_2 share a root, Lemma 4.2 would complete the proof.

To show that $f_2(\text{adj } \check{M}_i)_{i,1} - f_1(\text{adj } \check{M}_i)_{i,2}$ is nonzero, we compute its constant term from the following matrix:

$$(4.1) \quad \begin{bmatrix} p & 0 & \cdots & 0 & 0 & 0 \\ p + p^{j_2} & p & & & 0 & 0 \\ p & p^{j_3} & & & & 0 \\ \vdots & & \ddots & & & \vdots \\ p & 0 & & p^{j_{i-1}} & p & 0 \\ p & 0 & \cdots & 0 & p^j & p \end{bmatrix}.$$

These are the constants in \check{M}_i after adding p^j in the $i, i-1$ position—the main diagonal comes from line 2, the first column comes from line 4, and the second diagonal comes from line 5. To compute h_1 or h_2 , we use cofactor expansion along the bottom row after deleting the last column and the first or second row. The $(i-2) \times (i-2)$ minor determinants that are multiplied by the bottom row constant p_j are exactly f_1 and f_2 up to a sign. What remains sums to $(\text{adj } \check{M}_i)_{i,1}$ and $(\text{adj } \check{M}_i)_{i,2}$. So the constant terms of $(\text{adj } \check{M}_i)_{i,1}$, $(\text{adj } \check{M}_i)_{i,2}$, and f_2 are p^{i-1} , 0, and p to the power $1 + j_3 + \cdots + j_{i-1}$, respectively. This makes p to the power $i + j_3 + \cdots + j_{i-1}$ the constant term of $f_2(\text{adj } \check{M}_i)_{i,1} - f_1(\text{adj } \check{M}_i)_{i,2}$. \square

We remark that by using a large integer instead of x in line 2, the **for** loop could successively make pairs of integers coprime rather than polynomials. Then the Euclidean algorithm could test j in line 5; determinants involving polynomial entries need not be computed. We might expect such an algorithm to require $O(n^3 \log M_{n,1}n)$ operations (this uses that the average ratio with Euler's phi function, $\varphi(n)/n$, is a positive constant), but the provable worst case is bad. The best current asymptotic upper bound on the size of the interval that must be sieved or otherwise searched to find j is due to Iwaniec [17]. It only limits the algorithm to $O(n^7 \log M_{n,1}n)$ operations. We favored the polynomial approach because of an easier bound on j (Lemma 4.4) and a better provable worst case (Theorem 4.8).

The next lemma allows the vector in line 10 to pass as “**b**” from Subsection 4.1.

Lemma 4.5. *With M denoting its value in line 9, $\gcd((\text{adj } M)_{n,1}, (\text{adj } M)_{n,2}) = 1$.*

Proof. By Lemma 4.2, it suffices to prove $\gcd((\text{adj } M)_{n,1}, (\text{adj } M)_{n,2}, c) = 1$ with c as in line 6. Now let c' be c/p^j or $p+1$ as in line 7. Recall the constant terms displayed in (4.1), which show that $(\text{adj } M)_{n,2}$ is a power of p modulo c' . This implies $\gcd((\text{adj } M)_{n,1}, (\text{adj } M)_{n,2}, c)$ is a power of p since $p \nmid c'$. But the constants added throughout the **for** loop are multiples of p . So before substituting for x , only the leading term of $(\text{adj } M)_{n,1}$ might have been nonzero modulo p . With M now the original input matrix, this leading term is $M_{n,1} \det M^{n-2} x^{n-1}$. By line 1 this is coprime to p whenever the same is true of the integer substituted for x . \square

Lemma 4.6. *Let M be the input matrix, let c^j be as in line 8, and let A be such that $c^j \text{adj } M + A$ is the value of “ M ” in line 9. Then $\|MA\|_{\text{op}} < (2nM_{n,1})^{3n}/5n$ under any ℓ_p -norm.*

Proof. The operator norm is $\max_{\|\mathbf{u}\|=1} \|MA\mathbf{u}\|$. Using $\|\mathbf{u}\|_{\infty} \leq 1$ gives

$$(4.2) \quad \|MA\mathbf{u}\| \leq n \|MA\mathbf{u}\|_{\infty} \leq n^2 \max_{i,j \in [n]} |(MA)_{i,j}|.$$

We refer back to (4.1), which displays the entries of A when $i = n$. Lemma 4.4 says

$j_i \leq 2i - 2$, so the entries of MA are bounded in magnitude by

$$(4.3) \quad \max_{i,j \in [n]} |M_{i,j}| \max(np + p^2, p + p^{2n-2}) \leq 2M_{n,1}p^{2n-2} \leq 2M_{n,1}^3p^{2n-2}.$$

(Recall that $n \geq 2$ for this inequality.) Here $np + p^2$ comes from the first column of A , and $p + p^{2n-2}$ comes from the $(n-1)^{\text{th}}$ column.

Now we turn to the size of p . If $x \in \mathbb{R}$ is such that $x\#$, the product of primes not exceeding x , is larger than $M_{n,1}|\det M|$, then it must be that $p < x$. Rosser and Schoenfeld's lower bound on Chebyshev's theta function, $\vartheta(x) = \sum_{p \leq x} \log p$, gives $\vartheta(x) > 0.231x$ when $x \geq 2$ [28]. For the determinant we use Hadamard's bound: $|\det M| \leq (M_{n,1}\sqrt{n})^n$ [15]. So take $x = (\log M_{n,1}^3 n^n)/0.462$ (note that $x \geq 2$ even when $n = 2$ and $M_{n,1} = 1$, allowing for the Rosser, Schoenfeld bound) to get

$$\log x\# = \vartheta(x) > 0.231x = \frac{1}{2} \log M_{n,1}^{3n} n^n \geq \log M_{n,1}^{n+1} n^{n/2} \geq \log M_{n,1} |\det M|.$$

Combining $p < x$ with (4.2) and (4.3) gives $\|MA\|_{\text{op}} < 2M_{n,1}^3 n^2 x^{2n-2}$. We must show that this is less than the stated bound of $(2nM_{n,1})^{3n}/5n$. To do this, raise both expressions to the power $1/(n-1)$ and use $(5/4)^{1/(n-1)} \leq 5/4$. This simplifies the desired inequality to $(\log M_{n,1}^3 n)^2 < 1.366M_{n,1}^3 n$, which is true. \square

Theorem 4.7. *Under the ℓ_1 , ℓ_2 , or ℓ_∞ -norm, the output of Algorithm 3 solves the initial short vector problem.*

Proof. There are two parts to the proof: 1) showing that the algorithm replaces the columns of M with $n+1$ vectors that define the same lattice, n of them being “nearly” scaled orthonormal, and 2) showing that nearly scaled orthonormal is as good as being scaled orthonormal. Throughout the proof, let M be the input matrix, let c^j be as in line 8, let M' be the value of “ M ” in line 9, and let $A = M' - c^j \text{adj } M$ be the matrix of constants added throughout the **for** loop (as used in Lemma 4.6 and as shown in (4.1) when $i = n$).

For part 1), with $\mathbf{b} = (b_1, b_2, 0, \dots, 0)$ as in line 10, Lemma 4.5 gives

$$\mathbf{x} = M'^{-1}\mathbf{b} = \frac{(x_1, x_2, \dots, 1)}{\det M'}.$$

By Cramer's rule [10], the 1 in the last coordinate is the determinant after replacing the last column of M' by \mathbf{b} , so that these n columns generate \mathbb{Z}^n . This in turn shows that the columns of MM' and $M\mathbf{b}$ generate the input lattice. Also note, by Proposition 2.1, that the 1 in the last coordinate of $\det M' \cdot \mathbf{x}$ allows for GDA in place of SAP with N set to $n^{1/p} \det M'/2\alpha$ and the gap scaled by $1/n^{1/p}$.

Instead of finding a short integer combination of $M\mathbf{b}$ and the columns of

$$(4.4) \quad MM' = c^j \det M \text{Id} + MA,$$

Algorithm 3 uses $(MM')^{-1}(M\mathbf{b}) = \mathbf{x}$ and the columns of $(MM')^{-1}(MM') = \text{Id}$. Then $MM'\{q_0\mathbf{x}\}$ is proposed as a short vector. It is indeed an element of the original lattice as the coordinates of $M'\{q_0\mathbf{x}\} \equiv q_0\mathbf{b} \pmod{\mathbb{Z}^n}$ are all integers. But it must be checked is that $MM'\{q_0\mathbf{x}\}$ is short whenever $\{q_0\mathbf{x}\}$ is. Part 2) of the proof is to make precise the insignificance of the second matrix summand, MA , in (4.4). We begin by computing how much multiplication by the full matrix in (4.4) is allowed to inflate the gap without invalidating the output of GDA or SAP.

By Minkowski's theorem [25], the magnitude of the shortest vector in the original lattice with respect to the ℓ_∞ -norm is not more than $|\det M|^{1/n}$. So under an ℓ_p -norm with $p \in \mathbb{N}$, the shortest vector has some magnitude, say λ ,

with $(n^{1/p}|\det M|^{1/n})^p \geq \lambda^p \in \mathbb{Z}$. In particular, $n|\det M|^{2/n} \geq \lambda^2 \in \mathbb{Z}$ when $p \in \{1, 2, \infty\}$. Now, if $\mathbf{q} \in \mathbb{Z}^n$ is such that $\|M\mathbf{q}\|^2 < (a^2\lambda^2 + 1)/b^2$, then it must be that $\|M\mathbf{q}\| \leq a\lambda/b$ since there are no integers strictly between $(a\lambda/b)^2$ and $(a^2\lambda^2 + 1)/b^2$. Thus multiplication by MM' must inflate the gap between the norms of our output vector and the shortest vector by less than

$$(4.5) \quad \frac{\sqrt{a^2\lambda^2 + 1}}{b\alpha\lambda} = \frac{\sqrt{a^2\lambda^2 + 1}}{a\lambda} \geq \frac{\sqrt{a^2n|\det M|^{2/n} + 1}}{a\sqrt{n}|\det M|^{1/n}}.$$

Scaling does not affect the ratio of vector norms, so to determine the effect of multiplication by (4.4) it suffices to consider the matrix

$$(4.6) \quad \text{Id} + MA/c^j \det M$$

instead. If \mathbf{q}_{\min} is a shortest nonzero vector in the simultaneous approximation lattice generated by \mathbb{Z}^n and \mathbf{x} , then the shortest vector after applying (4.6) to this lattice has norm at least $(1 - \|MA\|_{\text{op}}/|c^j \det M|)\|\mathbf{q}_{\min}\|$. Similarly, the vector $\{q_0\mathbf{x}\}$ obtained using q_0 from line 11 increases in norm by at most a factor of $(1 + \|MA\|_{\text{op}}/|c^j \det M|)$. Combining this with our conclusion regarding (4.5) shows that it suffices to verify that the following inequality holds:

$$(4.7) \quad \frac{1 + \|MA\|_{\text{op}}/|c^j \det M|}{1 - \|MA\|_{\text{op}}/|c^j \det M|} \leq \frac{\sqrt{a^2n|\det M|^{2/n} + 1}}{a\sqrt{n}|\det M|^{1/n}}.$$

Now solve for $|c^j|$ to get a lower bound of

$$\frac{\sqrt{a^2n|\det M|^{2/n} + 1} + a\sqrt{n}|\det M|^{1/n}}{\sqrt{a^2n|\det M|^{2/n} + 1} - a\sqrt{n}|\det M|^{1/n}} \cdot \frac{\|MA\|_{\text{op}}}{|\det M|} < \frac{(5a^2n|\det M|^{2/n})\|MA\|_{\text{op}}}{|\det M|}.$$

Ignoring the powers of $|\det M|$ since $2/n \leq 1$, we see that j in line 8 is chosen to make the bound above agree exactly with Lemma 4.6. \square

Theorem 4.8. *Let $m = \max(a^{1/n^3}, M_{n,1})$. The reduction in Algorithm 3 requires $O(n^4 \log mn)$ operations on integers of length $O(n^4 \log mn)$.*

Proof. We will use that finding determinants, adjugates, inverses, or characteristic polynomials of $n \times n$ matrices with entry magnitudes bounded by m requires $O(n^3)$ operations on integers of length $O(n \log mn)$. For example, see Danilevskii's method for the characteristic polynomial [11] and the Bareiss algorithm for the others [5]. Note that we may then compute determinants of matrices with linear polynomial entries in $O(n^3)$ operations provided the matrix of linear terms or the matrix of constant terms is invertible.

In the proof of Lemma 4.6 we showed that the prime from line 1 is less than $(\log M_{n,1}^3 n^n)/0.462$. So finding it does not contribute to asymptotic complexity.

Now consider the **for** loop, where we must avoid recomputing the determinant in line 5 for each value of j in order to meet the prescribed bounds.

Let $i \geq 3$ and fix some notation: M is its value after line 4, $f_1 = (\text{adj } \tilde{M}_{i-1})_{i-1,1}$ and $f_2 = (\text{adj } \tilde{M}_{i-1})_{i-1,2}$, g_1 and g_2 have degree at most $i-3$ and $f_1 g_1 + f_2 g_2 = \det C(f_1, f_2) \neq 0$, and for some j , $h_1 = (\text{adj } \tilde{M}_i)_{i,1} - p^j f_1$ and $h_2 = (\text{adj } \tilde{M}_i)_{i,2} - p^j f_2$. Note for computing $(\text{adj } \tilde{M}_i)_{i,2}$ that the constant term matrix is not invertible, which may also be true of the linear term matrix. Because this complicates combining the Bareiss and Danilevskii algorithms, we could find $(\text{adj } \tilde{M}_i)_{i,2}$ indirectly by computing h_2 for two values of j that produce an invertible constant term matrix (recall from (4.1) that f_2 has nonzero constant term), then solve for it.

Call the polynomials in the resulting column vector below h'_1 and h'_2 :

$$(4.8) \quad \begin{bmatrix} f_2 + p^j g_1 x^{2i-3} & -f_1 + p^j g_2 x^{2i-3} \\ g_1 & g_2 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ = \begin{bmatrix} f_2(\text{adj } \tilde{M}_i)_{i,1} - f_1(\text{adj } \tilde{M}_i)_{i,2} - p^j \det C(f_1, f_2) x^{2i-3} \\ g_1(\text{adj } \tilde{M}_i)_{i,1} + g_2(\text{adj } \tilde{M}_i)_{i,2} - p^j \det C(f_1, f_2) \end{bmatrix}.$$

Note that if j makes h'_1 and h'_2 avoid a common root, then it does so for h_1 and h_2 .

View $C(h'_1, h'_2)$ as a matrix with linear polynomial entries where p^j is the variable. This variable only appears in the leading term of h'_1 and the constant term of h'_2 . So p^j only occurs on the main diagonal of $C(h'_1, h'_2)$, where its coefficient is nonzero. In particular, the polynomial $\det C(h'_1, h'_2)$ can be found in $O(n^3)$ operations. Substituting different values of p^j into this polynomial until one is nonzero avoids repeatedly finding determinants. And note that we still need only test up to $j = 2i - 2$ as stated in Lemma 4.4 because the determinant of the matrix in (4.8) is a constant (a unit in $\mathbb{Q}(x)$). Thus each **for** loop iteration requires $O(n^3)$ operations.

The integers composing the linear polynomial matrix entries that begin each **for** loop iteration are small powers of $p = O(n \log M_{n,1}n)$ and entries in the adjugate of the input matrix, M . By Hadamard's bound they are thus $O(n \log M_{n,1}n)$ in length. Hadamard's bound also applies to the coefficients of $(\text{adj } \tilde{M}_i)_{i,1}$ and $(\text{adj } \tilde{M}_i)_{i,2}$, making their lengths $O(n^2 \log M_{n,1}n)$. And it applies again to make $\det C((\text{adj } \tilde{M}_i)_{i,1}, (\text{adj } \tilde{M}_i)_{i,2})$ have length $O(n^3 \log M_{n,1}n)$. This is our bound on the length of c in line 6 and hence the length of c in line 7. The length of c^j in line 8 is then $O(\max(\log a^2(2M_{n,1}n)^{3n}, \log |c|)) = O(n^3 \log mn)$, with the maximum accommodating the ceiling function. Then a final application of Hadamard's bound for lines 9 and 10 makes integer lengths $O(n^4 \log mn)$. This is therefore a bound on the number of operations required by the Euclidean algorithm in line 9. \square

In [12], Dinur proves the NP-hardness of short vector problems under the ℓ_∞ -norm when $\alpha = n^{c/\log \log n}$ for some $c > 0$ by giving a direct reduction from the Boolean satisfiability problem (SAT). As a consequence, Theorems 4.7 and 4.8 prove the same for both good Diophantine approximation and simultaneous approximation problems. (There is no gap inflation for GDA in line 11 under the ℓ_∞ -norm.)

Corollary 4.9. *Good Diophantine approximation and simultaneous approximation problems are NP-hard under the ℓ_∞ -norm with $\alpha = n^{c/\log \log n}$ for some $c > 0$. \square*

This result is known for good Diophantine approximation [9], though the reduction $\text{SAT} \rightarrow \text{SVP} \rightarrow \text{GDA}$ completed here is simpler. In [9], Chen and Meng adapt the work of Dinur as well as Rössner and Seifert [30] to reduce SAT to finding short integer vectors that solve a homogeneous system of linear equations (HLS). It works by first reducing to the problem of finding pseudo-labels for a regular bipartite graph (PSL) with an algorithm in [3]. From HLS, [30] is further employed to reduce the number of equations in the system to one (SIR), which is then reduced to GDA in a different paper of Rössner and Seifert [29]. Each of the reductions, $\text{SAT} \rightarrow \text{PSL} \rightarrow \text{HLS} \rightarrow \text{SIR} \rightarrow \text{GDA}$, is gap-preserving under the ℓ_∞ -norm.

Short vector problems are only known to be NP-hard under the ℓ_∞ -norm. But there are other hardness results under a general ℓ_p -norm for which Theorems 4.7 and 4.8 can be considered complementary. See [19] for an exposition.

Another corollary is the reduction from a simultaneous approximation problem to GDA, giving the final row of Table 1. By Proposition 3.2, Algorithm 2 results in one

call to SVP with integers of length $O(n \log m)$, where we can take m to be the maximum magnitude among a^{1/n^4} (still $\alpha = a/b$) and the integers defining \mathbf{x} . Then Theorem 4.8 implies the reduction to GDA requires $O(n^4 \log m^n n) = O(n^5 \log m)$ (absorbing the operations required by Algorithm 2) on integers of length $O(n^5 \log m)$.

4.3. Further discussion. The last algorithm was restricted to an ℓ_p -norm for $p \in \{1, 2, \infty\}$. So we will discuss what happens with a more general approach.

Multiplication by MM' , shown in (4.6), may change the gap between the length of the shortest vector in the simultaneous approximation lattice and that of the vector output by GDA or SAP. That this potential inflation does not invalidate our output relies on the set of vector norms being discrete and α being rational—facts that were exploited to produce the expression in (4.5). The idea behind that paragraph is to find a nonempty interval $(\alpha\lambda, \alpha'\lambda)$, where $\lambda = \min_{\mathbf{q} \in \mathbb{Z}^n} \|M\mathbf{q}\|$, that contains no norms from the lattice defined by M (or even \mathbb{Z}^n for the interval tacitly given in the proof). This creates admissible inflation, α'/α , which is (4.5).

The purpose of restricting to ℓ_1 , ℓ_2 , or ℓ_∞ is to facilitate finding this interval. Knowing that $(b\alpha\lambda)^2 \in \mathbb{Z}$ for some $b \in \mathbb{Z}$ simplifies the search for α' . The same is true for any ℓ_p -norm with $p \in \mathbb{N}$. But the immediate analogs of (4.5), (4.6), and (4.7) lead to a replacement for the very last bound used in the proof of the form

$$\frac{(5pa^pn |\det M|^{p/n}) \|MA\|_{\text{op}}}{2 |\det M|}.$$

This makes the number of operations needed to execute line 9 depend exponentially on the input length $\log p$. We have not taken into account, however, the possibility of a nontrivial lower bound for the difference between large consecutive integers which are sums of n perfect p^{th} powers. Such a bound would allow for a longer interval, $(\alpha\lambda, \alpha'\lambda)$, that provably contains no lattice norms.

These arguments are all in effort to preserve the gap when reducing to SAP or, when $p = \infty$, GDA. The situation clarifies if a small amount of inflation is allowed. To solve a short vector problem with gap α using SAP with gap $\alpha' < \alpha$, the inequality (4.7) becomes

$$\frac{1 + \|MA\|_{\text{op}}/|c^j \det M|}{1 - \|MA\|_{\text{op}}/|c^j \det M|} \leq \frac{\alpha}{\alpha'}.$$

We still need to substitute of power of c for x in line 8 for the purpose of Lemma 4.5. Given these two constraints, it is sufficient in line 8 to take $M \leftarrow M(c^j)$ for

$$j = \left\lceil \log_{|c|} \frac{(\alpha + \alpha') \|MA\|_{\text{op}}}{(\alpha - \alpha') |\det M|} \right\rceil,$$

which can be made explicit with Lemma 4.6. There is no need to insist that α is rational or impose restrictions on $p \in [1, \infty]$ defining the norm.

As a final note, the reduction to SAP adapts to the inhomogeneous forms of these problems while the reduction to GDA does not. If $\mathbf{y} \in \mathbb{Z}^n$ (or $\mathbf{y} \in \mathbb{R}^n$ if we do not intend to preserve the gap as discussed above), then the reduction can end by solving the simultaneous approximation problem of finding $q_0 \in \mathbb{Z}$ with $\|\{q_0 \mathbf{x} - (MM')^{-1} \mathbf{y}\}\| \leq \alpha \min_{q \in \mathbb{Z}} \|\{q \mathbf{x} - (MM')^{-1} \mathbf{y}\}\|$, using the matrix from (4.6). But unless we know that the first coordinate of $(MM')^{-1} \mathbf{y}$ is an integer, there is no clear modification to Proposition 2.1 that permits the use of GDA.

REFERENCES

- [1] M. Agrawal. Simultaneous Diophantine approximation and short lattice vectors. <https://www.youtube.com/watch?v=7SGCXbim6Ug>, 2019. Accessed: 2019-12-01.
- [2] F. Armknecht, C. Elsner, and M. Schmidt. Using the inhomogeneous simultaneous approximation problem for cryptographic design. In *Africacrypt*, pages 242–259. Springer, 2011.
- [3] S. Arora, L. Babai, J. Stern, and E. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *J. Comput. System Sci.*, 54(2):317–331, 1997.
- [4] W. Baocang and H. Yupu. Public key cryptosystem based on two cryptographic assumptions. *IEE Proc. Comms.*, 152(6):861–865, 2005.
- [5] E. H. Bareiss. Sylvester’s identity and multistep integer-preserving Gaussian elimination. *Math. Comp.*, 22(103):565–578, 1968.
- [6] D. J. Bernstein and T. Lange. Post-quantum cryptography. *Nature*, 549(7671):188–194, 2017.
- [7] A. J. Brentjes. Multi-dimensional continued fraction algorithms. *MC Tracts*, 1981.
- [8] W. S. Brown. On Euclid’s algorithm and the computation of polynomial greatest common divisors. *J. ACM*, 18(4):478–504, 1971.
- [9] W. Chen and J. Meng. An improved lower bound for approximating Shortest Integer Relation in ℓ_∞ -norm (SIR $_\infty$). *Inform. Process. Lett.*, 101(4):174–179, 2007.
- [10] G. Cramer. *Introduction à l’analyse des lignes courbes algébriques*. Cramer & Cl. Philibert, 1750.
- [11] A. Danilevskii. On the numerical solution of the secular equation. *Mat. Sb.*, 44(2):169–172, 1937.
- [12] I. Dinur. Approximating SVP $_\infty$ to within almost-polynomial factors is NP-hard. *Theor. Comput. Sci.*, 285(1):55–71, 2002.
- [13] A. Frank and É. Tardos. An application of simultaneous Diophantine approximation in combinatorial optimization. *Combinatorica*, 7(1):49–65, 1987.
- [14] S. D. Galbraith. *Mathematics of public key cryptography*. Cambridge University Press, 2012.
- [15] J. S. Hadamard. Résolution d’une question relative aux déterminants. *B. Sc. Math.*, 2:240–246, 1893.
- [16] H. Inoue, S. Kamada, and K. Naito. Simultaneous approximation problems of p -adic numbers and p -adic knapsack cryptosystems—Alice in p -adic numberland. *p-Adic Numbers Ultrametric Anal. Appl.*, 8(4):312–324, 2016.
- [17] H. Iwaniec. On the problem of Jacobsthal. *Demonstr. Math.*, 11(1):225–232, 1978.
- [18] R. Kannan and A. Bachem. Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matrix. *SIAM J. Comput.*, 8(4):499–507, 1979.
- [19] R. Kumar and D. Sivakumar. Complexity of SVP. *SIGACT News*, 32(3):40–52, 2001.
- [20] J. C. Lagarias. Knapsack public key cryptosystems and Diophantine approximation. In *Eurocrypt*, pages 3–23. Springer, 1984.
- [21] J. C. Lagarias. The computational complexity of simultaneous Diophantine approximation problems. *SIAM J. Comput.*, 14(1):196–209, 1985.
- [22] H. W. Lenstra, A. K. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Math. Ann.*, 264(4):515–534, 1982.
- [23] M. Pohst. A modification of the LLL reduction algorithm. *J. Sym. Comp.*, 4(1):123–127, 1987.
- [24] D. Micciancio and S. Goldwasser. *Complexity of lattice problems: a cryptographic perspective*, volume 671. Springer Science & Business Media, 2012.
- [25] H. Minkowski. *Geometrie der zahlen*, volume 40. R. G. Teubner: Leipzig/Berlin, 1910.
- [26] P. Nguyen. Lattice reduction algorithms: Theory and practice. In *Eurocrypt*, pages 2–6. Springer, 2011.
- [27] C. Peikert. A decade of lattice cryptography. *Found. Trends Theor. Comput. Sci.*, 10(4):283–424, 2016.
- [28] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.*, 6(1):64–94, 1962.
- [29] C. Rössner and J.-P. Seifert. Approximating good simultaneous Diophantine approximations is almost NP-hard. In *MFCS*, pages 494–505. Springer, 1996.
- [30] C. Rössner and J.-P. Seifert. On the hardness of approximating shortest integer relations among rational numbers. *Theor. Comput. Sci.*, 209(1-2):287–297, 1998.
- [31] A. Shamir. A polynomial time algorithm for breaking the basic Merkle-Hellman cryptosystem. In *FOCS*, pages 145–152. IEEE, 1982.

UNIVERSITY OF COLORADO, BOULDER, CO, UNITED STATES
 Email address: daniel.e.martin@colorado.edu