# IMPROVED LOW-DENSITY SUBSET SUM ALGORITHMS

Matthijs J. Coster, Antoine Joux,
Brian A. LaMacchia, Andrew M. Odlyzko,
Claus-Peter Schnorr and Jacques Stern

**Abstract.** The general subset sum problem is NP-complete. However, there are two algorithms, one due to Brickell and the other to Lagarias and Odlyzko, which in polynomial time solve almost all subset sum problems of sufficiently low density. Both methods rely on basis reduction algorithms to find short non-zero vectors in special lattices. The Lagarias-Odlyzko algorithm would solve almost all subset sum problems of density $< 0.6463\ldots$ in polynomial time if it could invoke a polynomial-time algorithm for finding the shortest non-zero vector in a lattice. This paper presents two modifications of that algorithm, either one of which would solve almost all problems of density $< 0.9408\ldots$ if it could find shortest non-zero vectors in lattices. These modifications also yield dramatic improvements in practice when they are combined with known lattice basis reduction algorithms.
**Key words.** subset sum problems; knapsack cryptosystems; lattices; lattice basis reduction.
**Subject classifications.** 11Y16.

## 1. Introduction

The *knapsack* or *subset sum* problem is to find, given positive integers $a_1, \ldots, a_n$ (the weights) and $s$, some subset of the $a_i$ that sum to $s$, or equivalently to find variables $e_1, \ldots, e_n$, with $e_i \in \{0, 1\}$, such that

$$\sum_{i=1}^{n} e_i a_i = s. \tag{1.1}$$

This problem is known to be NP-complete [10] (in its feasibility recognition form), and so is thought to be very hard in general. This has led to the invention of several public-key cryptosystems based on the knapsack problem.

Almost all of these have been broken by now, however. (See [2, 3, 6, 17] for surveys of this field.) Most of the attacks exploited specific constructions of the relevant cryptosystems. In addition, two algorithms have been proposed, one by Brickell [1] and the other by Lagarias and Odlyzko [13] which show that almost all low-density subset sum problems can be solved in polynomial time. The *density* of a set of weights $a_1, \ldots, a_n$ is defined by

$$d = \frac{n}{\log_2 \max\limits_{1 \le i \le n} a_i}. \tag{1.2}$$

The interesting case is $d \le 1$, since for $d > 1$ there will in general be many subsets of weights with the same sum, and so such sets of weights could not be used for transmitting information. The Brickell and Lagarias-Odlyzko algorithms solve almost all subset sum problems with $d$ sufficiently small.

Both the Brickell and Lagarias-Odlyzko algorithms reduce the subset sum problem to that of finding a short vector in a lattice. The exact complexity of finding short vectors in lattices is not known, and expert opinion appears to be divided as to whether this problem is polynomial or not. At the moment, the best known polynomial time method in this area is the $L^3$ lattice basis reduction algorithm of Lenstra, Lenstra, and Lovász [15], which is only guaranteed to find a non-zero vector in an $n$-dimensional lattice that is at most an exponential times the length of the shortest non-zero vector in that lattice. If one uses that algorithm, the Lagarias-Odlyzko method can be shown rigorously to solve almost all subset sum problems of density $< c/n$ for large $n$ and for a fixed constant $c$, as is done in [13]. (See [8] for a simplified analysis of the algorithm.) Using more recent algorithms of Schnorr [21], one can improve the cutoff bound to $c'/n$ for arbitrarily small constants $c' > 0$, but at the cost of increasing the degree of the polynomial that bounds the running time.

Finding short vectors in lattices may be very hard in general. On the other hand, published algorithms, such as the $L^3$ one, perform much better in practice than is guaranteed by their worst case bounds, especially when they are modified [13, 14, 19, 22], and new algorithms are being invented [20, 21, 23]. Thus it is possible that on average, the problem of finding short vectors in lattices is easy, even if it is hard in the worst case. Therefore it seems worthwhile to separate the issues of efficiency of lattice basis reduction algorithms from the question of how well the subset sum problem can be reduced to that of finding a short vector in a lattice. (Note that Paz and Schnorr [18] have shown that the general problem of finding the shortest non-zero vector in a lattice is reducible to that of solving some subset sum problem, but with some loss of efficiency.)

Consider a *lattice oracle* that, given a basis for a lattice, with high probability yields in polynomial time the shortest non-zero vector in that lattice. We do not know how to construct such an oracle, but it might be possible to do so, and in any case in relatively low dimensions, known polynomial time algorithms act like such an oracle. The analysis of [13] showed that availability of such an oracle would let the Lagarias-Odlyzko algorithm solve almost all subset sum problems of density $< 0.6463\ldots$, but not higher than that. (Similar analyses are not available for the Brickell algorithm [1], although it seems to require even lower densities. See also [9].)

In this note we analyze two simple modifications of the part of the Lagarias-Odlyzko algorithm that reduces the subset sum problem to a short vector in a lattice problem. We show that with either of these modifications, a single call to a lattice oracle would lead to polynomial time solutions of almost all problems of density $< 0.9408\ldots$. Empirical tests show that these modifications also lead to dramatic improvements in the performance of practical algorithms. We present some results on this in Section 5. More data and fuller comparisons are given in [14].

In Section 2 we derive the Lagarias-Odlyzko bound using the approach in [8]. We show in Section 3 that this bound may be increased to $0.9408\ldots$ using a simple modification of the Lagarias-Odlyzko attack. Section 4 sketches the other modification, which appears to be quite different, but which yields the same bound, and its analysis reduces to essentially the same lattice point counting problem. Finally, Section 5 discusses possible improvements on the new bound and practical results.

This paper is based on the results of two independent investigations of the same problem. The modification of the Lagarias-Odlyzko attack described in Section 3 is due to Coster, LaMacchia, Odlyzko and Schnorr, and an extended abstract of it appears in [5]. The other modification, outlined in Section 4, is due to Joux and Stern, and was presented earlier in [12].

## 2. Previous results

In [13], Lagarias and Odlyzko show that if the density is bounded by $0.6463\ldots$, the lattice oracle is guaranteed to find the solution vector with high probability. This section derives the $0.6463\ldots$ bound using simpler techniques due to Frieze [8]. Our presentation differs from that of [8] in a few technical details.

Let $A$ be a positive integer and let $a_1, \ldots, a_n$ be random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $\boldsymbol{e} = (e_1, \ldots, e_n) \in \{0,1\}^n$, $\boldsymbol{e} \neq (0,0,\ldots,0)$ depending only on $n$, be fixed and let

$$s = \sum_{i=1}^{n} e_i a_i, \qquad\qquad t = \sum_{i=1}^{n} a_i.$$

We may assume that $s \geq t/n$, since if $s < t/n$ any $a_i \geq t/n$ cannot be in the subset, and may be removed from consideration. Similarly, $s \leq (1-(1/n))t$, otherwise any $a_i \geq t/n$ must be in the subset. Thus,

$$\frac{1}{n}t \leq s \leq \frac{n-1}{n}t. \tag{2.1}$$

We recall the Lagarias-Odlyzko attack on low-density subset sum problems. Define the vectors $\boldsymbol{b_1}, \ldots, \boldsymbol{b_{n+1}}$ as follows:

$$\begin{aligned} \boldsymbol{b_1} &= (1, 0, \ldots, 0, Na_1), \\ \boldsymbol{b_2} &= (0, 1, \ldots, 0, Na_2), \\ &\vdots \\ \boldsymbol{b_n} &= (0, 0, \ldots, 1, Na_n), \\ \boldsymbol{b_{n+1}} &= (0, 0, \ldots, 0, Ns), \end{aligned}$$

where $N$ is a positive integer which will be chosen later. Let $L$ be the lattice spanned by the vectors $\boldsymbol{b_1}, \ldots, \boldsymbol{b_{n+1}}$ (i.e. $L = \{\sum_{i=1}^{n+1} z_i \boldsymbol{b_i} \colon z_i \in \mathbb{Z} \quad \text{for } 1 \leq i \leq n+1\}$).

Notice that the solution vector $\hat{\boldsymbol{e}} = (e_1, \ldots, e_n, 0)$ is in $L$. Following the proof in [8] we are interested in vectors $\hat{\boldsymbol{x}} = (x_1, x_2, \ldots, x_{n+1})$ which satisfy:

$$\begin{aligned} \|\hat{\boldsymbol{x}}\| &\leq \|\hat{\boldsymbol{e}}\|, \\ \hat{\boldsymbol{x}} &\in L, \\ \hat{\boldsymbol{x}} &\notin \{\boldsymbol{0}, \hat{\boldsymbol{e}}, -\hat{\boldsymbol{e}}\}. \end{aligned} \tag{2.2}$$

We may assume that

$$\sum_{i=1}^{n} e_i \leq \tfrac{1}{2}n, \tag{2.3}$$

(i.e. the subset contains at most one-half of the $a_i$'s). If $\sum_{i=1}^{n} e_i > \frac{1}{2}n$, we may replace $s$ by $t-s$, $\boldsymbol{b_{n+1}}$ by $\boldsymbol{b'_{n+1}} = (0, \ldots, 0, N(t-s))$, and $\hat{\boldsymbol{e}}$ by $\hat{\boldsymbol{e}}' = (1-e_1, 1-e_2, \ldots, 1-e_n, 0)$. Solving this problem is equivalent to solving the given problem, $\sum_{i=1}^{n}(1-e_i) \leq \frac{1}{2}n$, and $s' = t-s \geq t/n$. (To be fully rigorous, we actually apply the basic method to two problems, at least one of which is covered by the condition $\sum_{i=1}^{n} e_i \le \frac{1}{2}n$, and our analysis below applies to this case.)

Choose $N > \sqrt{n}$. It is clear that $\hat{\boldsymbol{x}}$ satisfies Equation (2.2) only if $x_{n+1} = 0$. (Otherwise, $\|\hat{\boldsymbol{x}}\| \ge |x_{n+1}| \ge N > \sqrt{n} \ge \|\hat{\boldsymbol{e}}\|$, which contradicts Equation (2.2).) Let $y$ be defined by

$$ys = \sum_{i=1}^{n} x_i a_i, \tag{2.4}$$

and deduce that

$$|y|s = \left|\sum_{i=1}^{n} x_i a_i\right| \le \|\hat{\boldsymbol{x}}\| \left|\sum_{i=1}^{n} a_i\right| \le t\sqrt{\tfrac{1}{2}n}. \tag{2.5}$$

Hence, using Equation (2.1) above,

$$|y| \le n\sqrt{\tfrac{1}{2}n}. \tag{2.6}$$

Note that since $-y$ is the coefficient of $\boldsymbol{b}_{n+1}$ in the expansion of $\hat{\boldsymbol{x}}$ in terms of the basis vectors, $y \in \mathbb{Z}$.

We will show that the probability $P$ — that a lattice $L$ contains a short vector which satisfies Equation (2.2) — is:

$$\begin{aligned} P &= \Pr(\exists\ \hat{\boldsymbol{x}} \text{ which satisfies Equation (2.2)}) \\ &\le n\left(2n\sqrt{\tfrac{1}{2}n}+1\right)\frac{2^{c_0 n}}{A}, \quad \text{for } c_0 = 1.54724\ldots \end{aligned} \tag{2.7}$$

This implies that, if $A = 2^{cn}$ with $c > c_0$, $\lim\limits_{n\to\infty} P = 0$. If the density of a subset sum problem is less than $0.6463\ldots$, then

$$\begin{aligned} \frac{n}{\log_2 \max\limits_{1\le i\le n} a_i} < 0.6463\ldots &\implies \max_{1\le i\le n} a_i > 2^{n/0.6463\ldots} \\ &\implies A > 2^{c_0 n}. \end{aligned}$$

Thus, all subset sum problems with density $< 0.6463\ldots$ could be solved in polynomial time, given the existence of a lattice oracle.

We will now prove Equation (2.7). Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ denote an element of $\mathbb{Z}^n$. (Note that if $\hat{\boldsymbol{x}} = (x_1, \ldots, x_n, 0)$, then $\|\hat{\boldsymbol{x}}\| = \|\boldsymbol{x}\|$ and as a special case we have $\|\hat{\boldsymbol{e}}\| = \|\boldsymbol{e}\|$.) First we estimate the probability $P$ by

$$\begin{aligned}
P &\le \Pr(\exists\ \boldsymbol{x}, y \text{ s.t. } \|\boldsymbol{x}\| \le \|\boldsymbol{e}\|, |y| \le n\sqrt{\tfrac{1}{2}n}, \boldsymbol{x} \notin \{\boldsymbol{0}, \boldsymbol{e}, -\boldsymbol{e}\}, \sum_{i=1}^{n} x_i a_i = ys), \\
&\le \Pr\left(\sum_{i=1}^{n} a_i x_i = ys\colon 0 < \|\boldsymbol{x}\| \le \|\boldsymbol{e}\|,\ |y| \le n\sqrt{\tfrac{1}{2}n},\ \boldsymbol{x} \notin \{\boldsymbol{0}, \boldsymbol{e}, -\boldsymbol{e}\}\right) \\
&\quad \cdot |\{\boldsymbol{x}\colon \|\boldsymbol{x}\| \le \|\boldsymbol{e}\|\}| \cdot \left|\left\{y\colon |y| \le n\sqrt{\tfrac{1}{2}n}\right\}\right|. \qquad (2.8)
\end{aligned}$$

We have to estimate three factors in the right side of Equation (2.8). For the first factor of Equation (2.8) we may rewrite $\sum_{i=1}^{n} a_i x_i = ys$ as:

$$\sum_{i=1}^{n} a_i z_i = 0, \quad \text{where } z_i = x_i - y e_i.$$

Since $\boldsymbol{x}$ is non-zero and $\|\boldsymbol{x}\| \le \|\boldsymbol{e}\|$, we have $\boldsymbol{z} = (z_1, \ldots, z_n) \ne \boldsymbol{0}$, and so we may assume without loss of generality (by increasing the bound for the probability by a factor of at most $n$) that $z_1 \ne 0$. If $z'$ is defined as $-(\sum_{i=2}^{n} a_i z_i / z_1)$, then

$$\begin{aligned}
\Pr\left(\sum_{i=1}^{n} a_i z_i = 0\right) &= \Pr(a_1 = z'), \\
&= \sum_{j=1}^{A} \Pr(a_1 = z' | z' = j) \Pr(z' = j), \\
&= \sum_{j=1}^{A} \Pr(a_1 = z') \Pr(z' = j), \quad (a_1 \text{ and } j \text{ are independent}), \\
&= \sum_{j=1}^{A} \frac{1}{A} \Pr(z' = j), \\
&\le \frac{1}{A}.
\end{aligned}$$

Now we consider the second factor of Equation (2.8). From [13] (which borrowed the technique from a preliminary version of [16]) we know that

$$|\{\boldsymbol{x}\colon \|\boldsymbol{x}\| \le \|\boldsymbol{e}\|\}| \le \left|\left\{\boldsymbol{x}\colon \|\boldsymbol{x}\| \le \sqrt{\tfrac{1}{2}n}\right\}\right| \le 2^{c_0 n}, \quad \text{where } c_0 = 1.54724\ldots \qquad (2.9)$$

It is clear that the last factor of Equation (2.8) can be estimated by $2n\sqrt{\frac{1}{2}n} + 1$. This proves Equation (2.7).

## 3. A new, improved bound on the density

The main result of this note is an improvement in the maximum density of subset sum problems which can "almost always" be solved:

THEOREM 3.1. *Let $A$ be a positive integer, and let $a_1, \ldots, a_n$ be random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $\boldsymbol{e} = (e_1, \ldots, e_n) \in \{0,1\}^n$ be arbitrary, and let $s = \sum_{i=1}^{n} e_i a_i$. If the density $d < 0.9408\ldots$, then the subset sum problem defined by $a_1, \ldots, a_n$ and $s$ may "almost always" be solved in polynomial time with a single call to a lattice oracle.*

PROOF. We need to make only minor changes to the proof presented in Section 2. As above, $A$ is a fixed positive integer and $a_1, \ldots, a_n$ are random integers with $0 < a_i \leq A$ for $1 \leq i \leq n$. Let $\boldsymbol{e} = (e_1, \ldots, e_n) \in \{0,1\}^n$ be fixed, let $s = \sum_{i=1}^{n} e_i a_i$, and let $t = \sum_{i=1}^{n} a_i$. Vectors $\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}$ are defined as in Section 2. Vector $\boldsymbol{b_{n+1}}$ is replaced, however, by

$$\boldsymbol{b'_{n+1}} = \left(\tfrac{1}{2}, \tfrac{1}{2}, \ldots, \tfrac{1}{2}, Ns\right).$$

Let $L'$ be the lattice spanned by the vectors $\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}, \boldsymbol{b'_{n+1}}$.

In Section 2, we knew that the vector $\hat{\boldsymbol{e}} = (e_1, \ldots, e_n, 0)$ was in the lattice $L$. Notice that the new lattice $L'$ does not contain $\hat{\boldsymbol{e}}$ but instead contains the vector $\hat{\boldsymbol{e}}'$:

$$\hat{\boldsymbol{e}}' = (e'_1, \ldots, e'_n, 0), \quad \text{where } e'_i = e_i - \tfrac{1}{2}\ .$$

Since $e_i \in \{0,1\}$ for $1 \leq i \leq n$, we know that $e'_i \in \{-\frac{1}{2}, \frac{1}{2}\}$ for $1 \leq i \leq n$. Notice that $\|\hat{\boldsymbol{e}}'\|^2 \leq \frac{1}{4}n$ independent of the number of $e_i$'s which are equal to 1.

Again, we are interested in the number of vectors $\hat{\boldsymbol{x}}$ which satisfy conditions similar to Equation (2.2):

$$\begin{aligned} &\|\hat{\boldsymbol{x}}\| \leq \|\hat{\boldsymbol{e}}'\|, \\ &\hat{\boldsymbol{x}} \in L', \\ &\hat{\boldsymbol{x}} \notin \{\boldsymbol{0}, \hat{\boldsymbol{e}}', -\hat{\boldsymbol{e}}'\}. \end{aligned} \tag{3.1}$$

Setting $N > \frac{1}{2}\sqrt{n}$ implies that $x_{n+1} = 0$ for any $\hat{\boldsymbol{x}}$ which satisfies Equation (3.1). Suppose that $\hat{\boldsymbol{x}} = \sum_{i=1}^{n} y_i \boldsymbol{b_i} + y\boldsymbol{b'_{n+1}}$ satisfies Equation (3.1), then we can express $x_i$ in terms of $y_i$ and $y$ in the following way

$$x_i = y_i + \tfrac{1}{2}y, \quad \text{for } 1 \leq i \leq n,$$

$$0 = x_{n+1} = N \cdot \left\{ \sum_{i=1}^{n} a_i y_i + ys \right\}.$$

This implies that

$$\sum_{i=1}^{n} a_i y_i = -ys.$$

Therefore, Equation (2.4) can be replaced by:

$$\sum_{i=1}^{n} x_i a_i = \tfrac{1}{2} y(t-2s), \tag{3.2}$$

since $(\sum_{i=1}^{n} \boldsymbol{b_i}) - 2\boldsymbol{b'_{n+1}} = (0, 0, \ldots, 0, N(t-2s))$.

We now establish a bound on the size of $|y|$. From above,

$$\begin{aligned} |y(t-2s)| &= 2\left|\sum_{i=1}^{n} x_i a_i\right| \\ &\leq n\alpha\sqrt{n}, \quad \text{where } \alpha = \max_{1\leq i\leq n} a_i. \end{aligned} \tag{3.3}$$

If $|t-2s| \geq \frac{1}{2}\alpha$, then $|y||t-2s| \geq \frac{1}{2}|y|\alpha$, and

$$|y| \leq 2n\sqrt{n}, \tag{3.4}$$

by Equation (3.3). If $|t-2s| < \frac{1}{2}\alpha$, then we can solve two problems: one where $\alpha$ is assumed to be part of the subset which sums to $s$, and one where $\alpha$ is assumed to be part of the subset which sums to $t-s$. In the first case, the new problem has $s' = s-\alpha, t' = t-\alpha$, and

$$|t'-2s'| = |t-\alpha-2s+2\alpha| = |t-2s+\alpha| \geq \tfrac{1}{2}\alpha. \tag{3.5}$$

For the second case, the new problem has $s' = s, t' = t-\alpha$, and

$$|t'-2s'| = |t-2s-\alpha| \geq \tfrac{1}{2}\alpha. \tag{3.6}$$

Thus we may always assume $|t-2s| \geq \frac{1}{2}\alpha$ and that the bound in Equation (3.4) holds.

We may now calculate the bound on probability $P$ that there exists a vector $\hat{\boldsymbol{x}}$ which satisfies Equation (3.1). We now let $\boldsymbol{x} = (x_1, \ldots, x_n)$ be any vector such that $2\,\boldsymbol{x} \in \mathbb{Z}^n$. We obtain the following bound, similar to Equation (2.8):

$$P \leq \Pr\left(\sum_{i=1}^{n} a_i x_i = \tfrac{1}{2}y(t-2s)\right) \cdot \left|\left\{\boldsymbol{x} \colon \|\boldsymbol{x}\| \leq \tfrac{1}{2}\sqrt{n}\right\}\right| \cdot \left(4n\sqrt{n}+1\right). \tag{3.7}$$

As in Section 2, $\Pr(\sum_{i=1}^{n} a_i x_i = \frac{1}{2}y(t-2s)) \leq n/A$. The extra factor of $n$ in the estimate comes from assuming that the vector $\boldsymbol{x}$ has $x_1 \neq 0$. To estimate the number of vectors $\boldsymbol{x}$ with $\|\boldsymbol{x}\| \leq \frac{1}{2}\sqrt{n}$, we again use the technique in [13, 16], but in a more complicated way. The number of $\boldsymbol{x}$ with $\|\boldsymbol{x}\| \leq \sqrt{n}/2$ is bounded above by

$$\begin{aligned} &\left|\left\{\boldsymbol{w} = (w_1, \ldots, w_n) : w_i \in \mathbb{Z} \quad \text{for all } i, \|\boldsymbol{w}\| \leq \tfrac{1}{2}\sqrt{n}\right\}\right| \\ &+ \left|\left\{\boldsymbol{w} = (w_1, \ldots, w_n) : w_i \in \mathbb{Z} \quad \text{for all } i, \|\boldsymbol{w} - (\tfrac{1}{2}, \tfrac{1}{2}, \ldots, \tfrac{1}{2})\| \leq \tfrac{1}{2}\sqrt{n}\right\}\right|. \quad (3.8) \end{aligned}$$

In [16] it is shown that for $n$ sufficiently large, the second summand in Equation (3.8) above is smaller than the first summand by a factor that is exponential in $n$. In any case, the second summand equals $2^n$. By the method of [13, 16], the first summand is bounded, for every $u > 0$, by

$$2^{(\log_2 e)\delta(u)n},$$

where

$$\delta(u) = \tfrac{1}{4}u + \ln\theta(e^{-u}), \quad \text{for } \theta(z) = 1 + 2\sum_{k=1}^{\infty} z^{k^2}.$$

Numerically, we may calculate the minimum value of $\delta(u)$, and obtain

$$\delta(u) \geq \delta(u_0) = 0.7367\ldots, \qquad \text{for } u_0 = 1.8132\ldots$$

Thus, for large $n$, we have

$$\left|\left\{\boldsymbol{x} \colon \|\boldsymbol{x}\| \leq \tfrac{1}{2}\sqrt{n}\right\}\right| \leq 2^{c_0' n}, \qquad \text{for } c_0' = 1.0628\ldots,$$

$$P \leq n\left(4n\sqrt{n} + 1\right)\frac{2^{c_0' n}}{A}.$$

Thus, any subset sum problem with density $d < 1/c_0' = 0.9408\ldots$ may be solved in polynomial time, given the existence of a lattice oracle. $\square$

## 4. Another improvement on the critical density

The preceding section showed one way to improve on the critical density below which a lattice oracle would enable one to solve most subset sum problems. This was achieved by replacing the lattice $L$ in the Lagarias-Odlyzko attack by the lattice $L'$. Here we sketch how a comparable increase in the critical density can be accomplished by using a lattice $L''$, which is very different. The lattice $L''$ is generated by the following vectors in $\mathbb{R}^{n+2}$:

$$\begin{aligned}
\boldsymbol{b_1} &= (n+1, -1, -1, \ldots, -1, Na_1), \\
\boldsymbol{b_2} &= (-1, n+1, -1, \ldots, -1, Na_2), \\
&\vdots \\
\boldsymbol{b_n} &= (-1, \ldots, -1, n+1, -1, Na_n), \\
\boldsymbol{b_{n+1}} &= (-1, \ldots, -1, -1, n+1, -Ns),
\end{aligned}$$

where $N$ is a large positive integer ($N \geq n^2$, say).

If we consider

$$\boldsymbol{w} = \sum_{i=1}^{n+1} x_i \boldsymbol{b_i} = (w_1, \ldots, w_{n+2}), \tag{4.1}$$

then for $1 \leq j \leq n+1$,

$$w_j = (n+1)x_j - \sum_{\substack{i=1 \\ i \neq j}}^{n+1} x_i, \tag{4.2}$$

while

$$w_{n+2} = N \left( \sum_{i=1}^{n} x_i a_i - x_{n+1} s \right). \tag{4.3}$$

Simple manipulation yields

$$\|\boldsymbol{w}\|^2 = (n+2)^2 \sum_{i=1}^{n+1} x_i^2 - (n+3) \left( \sum_{i=1}^{n+1} x_i \right)^2 + N^2 \cdot E^2, \tag{4.4}$$

where

$$E = \sum_{i=1}^{n} x_i a_i - x_{n+1} s. \tag{4.5}$$

If $x_i = e_i$ for $1 \leq i \leq n$, $x_{n+1} = 1$, then $E = 0$ and $\|\boldsymbol{w}\|^2$ is bounded by approximately $n^3/4$ (and even less if the $e_i$ are mostly 1's or mostly 0's). We next indicate how to show that most of the time there will be no shorter vectors. If $E \neq 0$, then $\|\boldsymbol{w}\|^2 \geq N^2 \geq n^4$, so we only have to worry about the case $E = 0$, and by the method of the preceding two sections, it suffices to bound the number of $\boldsymbol{x} \in \mathbb{Z}^{n+1}$ such that

$$F(\boldsymbol{x}) = (n+2)^2 \sum_{i=1}^{n+1} x_i^2 - (n+3) \left( \sum_{i=1}^{n+1} x_i \right)^2 \tag{4.6}$$

satisfies $F(\boldsymbol{x}) \leq n^3/4$. Now

$$F(\boldsymbol{x}) = \sum_{i=1}^{n+1} x_i^2 + (n+3) \sum_{1 \leq i < j \leq n+1} (x_i - x_j)^2. \tag{4.7}$$

Suppose that

$$t = \frac{1}{n+1} \sum_{i=1}^{n+1} x_i.$$

Then

$$F(\boldsymbol{x}) = \sum_{i=1}^{n+1} x_i^2 + (n+1)(n+3) \sum_{i=1}^{n+1} (x_i - t)^2.$$

Therefore if $F(\boldsymbol{x}) \lesssim n^3/4$, then

$$\sum_{i=1}^{n+1} (x_i - t)^2 \lesssim n/4,$$

and so $\boldsymbol{x}$ lies in a sphere of radius $\frac{1}{2}\sqrt{n}$ with center at $(t, t, \ldots, t)$, and the bounds quoted before for the number of lattice points in such a sphere apply. It remains to show that not too many different values of $t$ can occur. If $F(\boldsymbol{x}) \lesssim n^3/4$, then clearly $\sum_{i=1}^{n+1} x_i^2 \lesssim n^3/4$, and so by the Cauchy-Schwartz inequality,

$$\left( \sum_{i=1}^{n} x_i \right)^2 \leq (n+1) \sum_{i=1}^{n+1} x_i^2 \lesssim n^4/4,$$

so $|t| \lesssim n/2$. Furthermore, $(n+1)t \in \mathbb{Z}$, so we have $\lesssim n^2$ different values of $t$, and this yields the desired bound for the number of $\boldsymbol{x} \in \mathbb{Z}^{n+1}$ with $F(\boldsymbol{x}) \lesssim n^3/4$.

The critical density for this method is exactly the same as for the method of Section 3 since both depend on the number of lattice points in any sphere in $\mathbb{R}^n$ (for the method of Section 3) or $\mathbb{R}^{n+1}$ (for the method of this section) that have radius approximately $\frac{1}{2}\sqrt{n}$ being smaller than $A$. However, the lattice $L''$ used in this section is very different from the lattice $L'$ of Section 3.

The main reason $L$ performs so much more poorly than lattices $L'$ and $L''$ is that it contains many short vectors in which some of the first $n$ coordinates are $-1$. In lattice $L'$, corresponding vectors are not all that short, since distance is in effect measured from a vector with coordinates mostly $\frac{1}{2}$, so a $-1$ contributes much more to the length than a 0 or 1. In lattice $L''$, a similar effect is attained by arranging the distance to contain the term $F(\boldsymbol{x})$ of Equation (4.7); the last sum penalizes vectors $\boldsymbol{x}$ with the $x_i$ far from their mean.

## 5. Discussion

The analysis of Section 3 shows that it is possible to improve the density bound from 0.6463... to 0.9408... by modifying one vector in the lattice basis. We now consider the possibilities of improving on this bound.

Solving subset sum problems with basis reduction is closely connected to lattice covering problems. In particular, we want to cover the vertices of the $n$-cube (representing the possible $\boldsymbol{e}$ solution vectors) with a polynomial number of $n$-spheres of radius $\sqrt{\alpha n}$. Lagarias and Odlyzko showed that it was possible to cover the $n$-cube with two $n$-spheres of radius $\sqrt{\frac{1}{2}n}$. The two spheres (centered at $(0,0,\ldots,0)$ and $(1,1,\ldots,1)$) correspond to the two basis reduction problems which must be solved for any given subset sum problem. Our analysis in Section 3 uses one $n$-sphere of radius $\frac{1}{2}\sqrt{n}$ centered at $(\frac{1}{2},\frac{1}{2},\ldots,\frac{1}{2})$ to cover all the points.

One way to improve the bound presented above would be to show that it is possible to cover the vertices of the $n$-cube with a polynomial number of $n$-spheres of radius $\sqrt{\alpha n}$ with $\alpha < \frac{1}{4}$. We show that this is not possible, and that the asymptotic bound of 0.9408... cannot be improved in this way. The following proposition shows that any $n$-sphere of radius $\sqrt{\alpha n}$ with $\alpha < \frac{1}{4}$ can cover only an exponentially small fraction of the vertices of the $n$-cube. Thus, no polynomial collection of such spheres can satisfy our requirements.

PROPOSITION 5.1. *Any sphere of radius $\sqrt{\alpha n}$, $\alpha < \frac{1}{4}$, in $\mathbb{R}^n$ contains at most $(2-\delta)^n$ points of $\{0,1\}^n$, for some $\delta = \delta(\alpha) > 0$.*

PROOF. Suppose that the $n$-sphere is centered at the point $\boldsymbol{c} = (c_1,\ldots,c_n)$. We are interested in the number of points $\boldsymbol{e} \in \{0,1\}^n$ for which $\|\boldsymbol{c}-\boldsymbol{e}\|^2 \le \alpha n$. Using the upper bound technique of [16], we show that $N$, the number of points in $\{0,1\}^n$ inside the sphere, is bounded by

$$N \le e^{\alpha n} \prod_{i=1}^{n} (e^{-c_i^2} + e^{-(c_i-1)^2}). \tag{5.1}$$

If the point $\boldsymbol{e} = (e_1, \ldots, e_n)$ is inside the sphere, then $\|\boldsymbol{c}-\boldsymbol{e}\|^2 = \sum_{i=1}^n (c_i - e_i)^2 \leq \alpha n$, and after expanding the right side, Equation (5.1) contains a term of the form

$$\exp\left(\alpha n - \sum_{i=1}^{n} (c_i - e_i)^2\right) \geq 1,$$

for each such point $\boldsymbol{e}$, which proves Equation (5.1) since all terms in the expansion are nonnegative.

Since the terms in the product in Equation (5.1) are independent, we know that the value of $N$ is bounded by

$$e^{\alpha n} \max_{\boldsymbol{c} \in \mathbb{R}^n} \prod_{i=1}^{n} \left(e^{-c_i^2} + e^{-(c_i-1)^2}\right) \leq e^{\alpha n}(2e^{-1/4})^n.$$

(It is easy to show that the maximum value of $f(z) = e^{-z^2} + e^{-(z-1)^2}$ is $2e^{-1/4}$.) Thus,

$$\begin{aligned} N \leq e^{n\alpha} 2^n e^{(-1/4)n} &= 2^n e^{n(\alpha - 1/4)} \\ &= (2 - \delta(\alpha))^n, \quad \text{for } \delta(\alpha) = 2(1 - e^{\alpha - 1/4}) \end{aligned}$$

For all $\alpha < \frac{1}{4}$, $\delta(\alpha) > 0$, which proves the proposition. $\square$

As $n \to \infty$, any $n$-sphere with radius $\sqrt{\alpha n}$, $\alpha < \frac{1}{4}$, will contain at most $(2 - \delta(\alpha))^n$ points in $\{0,1\}^n$. Thus, any polynomial-sized collection of spheres cannot contain all the points in $\{0,1\}^n$. Thus we cannot hope to asymptotically improve the $0.9408\ldots$ bound by reducing a polynomial number of bases with different $\boldsymbol{b_{n+1}}$ vectors. However, for small dimensions it might be possible to improve the bound, even though any such advantage will disappear as $n$ grows.

In cases where the subset sum problem (Equation 1.1) to be solved is known to have $\sum e_i$ small (as occurs in some knapsack cryptosystems, such as the Chor-Rivest one [4], which has still not been broken), it is possible to again improve on the results of [13] by the approach of Section 3. For example, if we know that

$$\sum_{i=1}^{n} e_i = \beta n,$$

we can replace the vector $\boldsymbol{b_{n+1}}$ in the basis of $L$ by

$$\boldsymbol{b''_{n+1}} = (\beta, \beta, \ldots, \beta, Ns),$$

and then the lattice $L$ will contain a vector of length $\sqrt{n\beta(1-\beta)}$, and our analysis shows that in this case it then becomes possible to solve most problems with even smaller weights $a_i$. The density bound for the approach in Section 4 is similarly improved if it is known that $\sum e_i = \beta n$; no modification of lattice $L''$ is required to take advantage of this additional information. However, it appears that there are choices for the parameters in the Chor-Rivest knapsack which yield subset sum problems with densities above even these improved bounds. Thus asymptotically our algorithms do not threaten the security of this cryptosystem. On the other hand, for moderate sizes of the problem (such as a challenge version of the Chor-Rivest knapsack with $n = 103$ that was constructed by B. Chor) solutions can be found with nonnegligible probability. Thus to obtain secure cryptosystems, one has to use very large values of the basic parameters, which make this scheme less attractive.

When we consider the $L_\infty$ or sup-norm,

$$\|(x_1, \ldots, x_n)\|_\infty = \max_{1 \le j \le n} |x_j|,$$

then we find that the vector $\hat{\mathbf{e}}'$ has norm $1/2$. Therefore, we can solve all subset sum problems of any density if we have a lattice oracle for the sup-norm, as was pointed out by Michael Kaib.

The general sup-norm shortest vector problem is known to be NP-complete [7]; the complexity of the square-norm shortest vector problem is an open problem. That a sup-norm lattice oracle yields a better density bound than a square-norm lattice oracle suggests that the shortest vector problem for the sup-norm might be harder than for the square-norm.

The discussion above dealt with the approach of Section 3 to improving the Lagarias-Odlyzko algorithm, and showed that the simplest idea for improving on it further does not work. We do not see any way to improve on either that method or the one in Section 4.

Sections 3 and 5 presented theoretical results that assume the availability of an efficient method for finding the shortest non-zero vector in a lattice. When one uses known algorithms for lattice basis reduction, applying them to lattice $L'$ instead of lattice $L$ also yields dramatic improvements, although the results are not as good as they would be in the presence of a lattice oracle. For example, Table 1 presents the comparison obtained in one particular set of experiments. The lattices used were not exactly $L$ and $L'$, and the reduction algorithm used a combination of ideas from several sources. More extensive data sets and details of the computations are presented in [14]. For each entry in Table 1, $n$ denotes the number of items, and $b$ the number of bits (chosen at random)

| $n$ | $b$ | $L$ | $L'$ |
|---|---|---|---|
| 50 | 50 | 0.05 | 1.00 |
| 50 | 60 | 0.55 | 1.00 |
| 50 | 75 | 1.00 | – |
| 66 | 76 | – | 0.25 |
| 66 | 84 | – | 0.80 |
| 66 | 92 | – | 0.95 |
| 66 | 100 | – | 1.00 |
| 66 | 104 | 0.30 | 1.00 |
| 66 | 108 | 0.55 | 1.00 |
| 66 | 112 | 0.60 | 1.00 |
| 66 | 116 | 1.00 | – |

Table 1: Fraction of random subset sum problems solved by a particular reduction algorithm applied to bases $L$ and $L'$, respectively.

for each item. For each $(n, b)$ combination, 20 problems were attempted, where in each case $e_i = 1$ for exactly $n/2$ of the items. The entries for the $L$ and $L'$ column indicate what fraction of the 20 problems were solved in each case. It would be of interest to obtain similar comparisons for implementations of other algorithms, such as that of [11]. Combining the improved lattice $L'$ of this paper with variants of the algorithms of [20] leads to solutions of subset sum problems of even higher densities, as is shown in [22].

## Acknowledgements

M. Coster's visit to AT&T Bell Laboratories was supported by a Fulbright scholarship. A. Joux is supported by DGA.

## References

[1] E. F. Brickell, Solving low density knapsacks, in *Advances in Cryptology, Proceedings of Crypto '83*, Plenum Press, New York, 1984, 25-37.

[2] E. F. Brickell, The cryptanalysis of knapsack cryptosystems, in *Applications of Discrete Mathematics*, R. D. Ringeisen and F. S. Roberts, eds., SIAM, 1988, 3-23.

[3] E. F. Brickell and A. M. Odlyzko, Cryptanalysis: a survey of recent results, *Proc. IEEE* **76** (1988), 578-593.

[4] B. Chor and R. Rivest, A knapsack-type public key cryptosystem based on arithmetic in finite fields, *IEEE Trans. Information Theory* **IT-34** (1988), 901-909.

[5] M. J. Coster, B. A. LaMacchia, A. M. Odlyzko and C.-P. Schnorr, An improved low-density subset sum algorithm, in *Advances in Cryptology: Proceedings of Eurocrypt '91*, D. W. Davies, ed., *Lecture Notes in Computer Science* **547**, Springer-Verlag, New York, 1991, 54-67.

[6] Y. Desmedt, What happened with knapsack cryptographic schemes?, in *Performance Limits in Communication, Theory and Practice*, J. K. Skwirzynski, ed., Kluwer, Boston, 1988, 113-134.

[7] P. van Emde Boas, *Another NP-complete partition problem and the complexity of computing short vectors in a lattice*, Rept. 81-04, Dept. of Mathematics, Univ. of Amsterdam, 1981.

[8] A. M. Frieze, On the Lagarias-Odlyzko algorithm for the subset sum problem, *SIAM J. Comput.* **15(2)** (1986), 536-539.

[9] M. L. Furst and R. Kannan, Succinct certificates for almost all subset sum problems, *SIAM J. Comput.* **18** (1989), 550-558.

[10] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, 1979.

[11] J. Håstad, B. Just, J. C. Lagarias, and C.-P. Schnorr, Polynomial time algorithms for finding integer relations among real numbers, *SIAM J. Comput.* **18(5)** (1989), 859-881.

[12] A. Joux and J. Stern, Improving the critical density of the Lagarias-Odlyzko attack against subset sum problems, *Proceedings of Fundamentals of Computation Theory '91*, L. Budach, ed., *Lecture Notes in Computer Science* **529**, Springer-Verlag, New York, 1991, 258-264.

[13] J. C. Lagarias and A. M. Odlyzko, Solving low-density subset sum problems, *J. Assoc. Comp. Mach.* **32(1)** (1985), 229-246.

[14] B. A. LaMacchia, *Basis Reduction Algorithms and Subset Sum Problems*, SM Thesis, Dept. of Elect. Eng. and Comp. Sci., Massachusetts Institute of Technology, Cambridge, MA, 1991. Also available as AI Technical Report 1283, MIT Artificial Intelligence Laboratory, Cambridge, MA, 1991.

[15] A. K. Lenstra, H. W. Lenstra, and L. Lovász, Factoring polynomials with rational coefficients, *Math. Ann.* **261** (1982), 515-534.

[16] J. E. Mazo and A. M. Odlyzko, Lattice points in high-dimensional spheres, *Monatsh. Math.* **110** (1990), 47-61.

[17] A. M. Odlyzko, The rise and fall of knapsack cryptosystems, in *Cryptology and Computational Number Theory*, C. Pomerance, ed., *Proc. Symp. Appl. Math.* **42**, Amer. Math. Soc., Providence, 1990, 75-88.

[18] A. Paz and C.-P. Schnorr, Approximating integer lattices by lattices with cyclic factor groups, in *Automata, Languages, and Programming: $14^{th}$ ICALP*, *Lecture Notes in Computer Science* **267**, Springer-Verlag, New York, 1987, 386-393.

[19] S. Radziszowski and D. Kreher, Solving subset sum problems with the $L^3$ algorithm, *J. Combin. Math. Combin. Comput.* **3** (1988), 49-63.

[20] C.-P. Schnorr, A hierarchy of polynomial time lattice basis reduction algorithms, *Theoretical Computer Science* **53** (1987), 201-224.

[21] C.-P. Schnorr, A more efficient algorithm for lattice basis reduction, *J. Algorithms* **9** (1988), 47-62.

[22] C.-P. Schnorr and M. Euchner, Lattice Basis Reduction: Improved Practical Algorithms and Solving Subset Sum Problems, in *Proceedings of Fundamentals of Computation Theory '91*, L. Budach, ed., *Lecture Notes in Computer Science* **529**, Springer-Verlag, New York, 1991, 68-85.

[23] M. Seysen, Simultaneous reduction of a lattice basis and its reciprocal basis, *Combinatorica*, to appear.

Manuscript received 2 August 1991

Matthijs J. Coster
AT&T Bell Laboratories
Murray Hill, NJ 07974
USA

Antoine Joux
GRECC/DMI
Ecole Normale Supérieure
45 rue d'Ulm
75230 PARIS Cedex 05
FRANCE
joux@frulm63.bitnet

Brian A. LaMacchia
AT&T Bell Laboratories
Murray Hill, NJ 07974
USA

Andrew M. Odlyzko
AT&T Bell Laboratories
Murray Hill, NJ 07974
USA
amo@research.att.com

Claus-Peter Schnorr
Universität Frankfurt
Fachbereich Mathematik/Informatik
Postfach 11 19 32
6000 Frankfurt am Main
GERMANY
schnorr@informatik.uni-frankfurt.de

Jacques Stern
GRECC/DMI
Ecole Normale Supérieure
45 rue d'Ulm
75230 PARIS Cedex 05
FRANCE
stern@dmi.ens.fr

Current address of M. Coster:
Department of Economics
Room B-837
University of Tilburg
P.O. Box 90153
5000 LE Tilburg
The NETHERLANDS
coster@kub.nl

Current address of B. LaMacchia:
MIT AI Laboratory
545 Technology Square
Cambridge, MA 02139
USA
bal@mit.edu