INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

A Simple Approximation Algorithm in $\mathbf{Z}[e^{2\pi i/8}]$

M.A. Shokrollahi and V. Stemann

TR-96-032

August 1996

Abstract

We describe a very simple and efficient new algorithm for the approximation of complex numbers by algebraic integers in $\mathbb{Z}[e^{2\pi i/8}]$ whose coeffcients with respect to the usual basis are bounded in absolute value by a given integer M. Its main idea is the use of a novel signature technique. An important application is the reduction of dynamic range requirements for residue number system implementations of the discrete Fourier transform. The algorithm uses at most $10\log(M)$ arithmetic steps and $2.4\log(M)$ additional memroy. It yields approximations within a distance of at most $3.42/M$. Several examples are included which show that the algorithm is very fast in practice. For instance, 50000 complex approximations take less than 0.7 seconds on a SPARC-5.

1 Introduction

In a pioneering paper Cozzens and Finkelstein [1] suggest to use cyclotomic integers in the ring $\mathbb{Z}[\zeta] := \left\{\alpha_0 + \alpha_1\zeta + \ldots + \alpha_{2^{n-1}-1}\zeta^{2^{n-1}-1} \;\middle|\; \alpha_i \in \mathbb{Z}\right\}$, $\zeta = e^{2\pi i/2^n}$, to approximate the input as well as program constants for residue number system processing of functions such as the Discrete Fourier Transform. The reason why this strategy is superior to conventional scaling is that $\mathbb{Z}[\zeta]$ is *dense* in $\mathbb{C}$ for $n \geq 3$. Hence for all $z \in \mathbb{C}$ and all $\varepsilon > 0$ there is an approximation $a \in \mathbb{Z}[\zeta]$ of z, such that $|a - z| \leq \varepsilon$.

The question arises how to approximate a complex number by an element of $\mathbb{Z}[\zeta]$. We will discuss this problem in detail for the case $\zeta = e^{2\pi i/8}$. To respect dynamic range requirements, the algorithm has to have an additional input $M \in \mathbb{N}$ and should output approximations in the set $\mathbb{Z}[\zeta]_M$ which is defined as the set of linear combinations of powers of ζ with integer coefficients bounded in absolute value by M.

In [1] Cozzens and Finkelstein present an algorithm whose basic ingredient is exhaustive search, and hence, is impractical for larger values of M.

Games [2, 3] develops a greedy algorithm, a rough sketch of which is as follows: in a first step a small element ε of $\mathbb{Z}[\zeta]_M$ is found. Then the algorithm starts at the origin and successively improves the approximation by adding an appropriate element of the form $\zeta^i\varepsilon$ without violating the bound on the coefficients. He shows that an extension of this algorithm achieves an approximation error $\leq |\varepsilon|/\sqrt{2-\sqrt{2}}$. Finally he gives an explicit algorithm, based on continued fractions, to find the shortest vector. The final result is an algorithm with worst case approximation error of $O(1/M)$. As the algorithm needs to compute the absolute values of $\zeta^i\varepsilon$ at each step, and to compare it with the number to be approximated, it may not be suitable for real time applications.

Marcellin and Fischer [4] suggest the use of tables of size $O(M)$ to reduce the amount of computations needed in the above algorithm. However, this might be impractical for large M as well.

We present a very simple algorithm bases on a signature technique to solve the approximation problem. For appropriate choices of M the algorithm approximates a complex number in the unit circle with an error of less than $3.42/M$, needs at most $10\log(M)$ arithmetic operations, and uses a table of size $2.4\log(M)$. For arbitrary M the bound on the approximation error has to be multiplied by an additional factor of $1+\sqrt{2}$. Due to its simple structure, our algorithm is also suitable for real time computations.

The new and simple idea of the algorithm is explained in the following. In a first step we reduce the complex approximation problem to the approximation problem of real numbers in the interval $[0,1]$ by elements of $\mathbb{Z}[\sqrt{2}]_M$ which is the set of all $\alpha_0 + \alpha_1\sqrt{2}$ such that $|\alpha_0|, |\alpha_1| \leq M$. This is done by separately approximating the real and imaginary part. This only gives an additional factor of two in the bound on the coefficients of the final approximation. The transition to real approximations allows the use of a signature technique: if $a = \alpha_0 + \alpha_1\sqrt{2} \in \mathbb{Z}[\sqrt{2}]_M$ is a number in the interval $[0,1]$, then α_0 and α_1 have opposite signs. Assume that a has signature $(-,+)$, i.e., $\alpha_0 \leq 0$ and $\alpha_1 \geq 0$ then $a + b \in \mathbb{Z}[\sqrt{2}]_M$ for any $b \in \mathbb{Z}[\sqrt{2}]_M$ with signature $(+,-)$. Hence, if ε_1 and ε_2 are small positive elements of opposite signature, then we can approximate any given real number in $[0,1]$ by starting with 0 and adding to the current approximation one of the ε_i according to their signature. There is a tradeoff between the approximation error obtained

and the running time of this algorithm. To remedy the situation we compute for a sequence $P_1 \le P_2 \le \cdots \le P_\ell \le M$ decreasing elements $\varepsilon_k \in \mathbb{Z}[\sqrt{2}]_{P_k}$ such that two consecutive ones have opposite signature, and use these elements to obtain aproximations in $\mathbb{Z}[\sqrt{2}]_{P_k}$. This amounts in the storage of ℓ elements. We will apply Games' idea [2] to use convergents of the continued fractione expansion of $\sqrt{2}$ to construct the elements ε_k.

Well-known facts about continued fractions are reviewed in the next section. We describe and analyze the algorithm in detail in Section 3. Results of our implementation are given in the final section.

2 Preliminaries

2.1 Reduction to the real case

We first show how to reduce the approximation problem to the real case. Let $\mathbb{Z}[\sqrt{2}]_M := \{\alpha_0 + \alpha_1\sqrt{2} \mid \alpha_0, \alpha_1 \in \mathbb{Z}, |\alpha_0|, |\alpha_1| \le M\}$.

Lemma 1. *Let $z \in \mathbb{C}$, and $a := \alpha_0 + \alpha_1\sqrt{2} \in \mathbb{Z}[\sqrt{2}]_{M/2}$, $b := \beta_0 + \beta_1\sqrt{2} \in \mathbb{Z}[\sqrt{2}]_{M/2}$ be such that $|a - \mathrm{Re}(z)| \le \delta$ and $|b - \mathrm{Im}(z)| \le \delta$. Then $a + ib \in \mathbb{Z}[e^{2\pi i/8}]_M$ is an approximation of ω with error $\le \sqrt{2}\delta$.*

Proof. Let $\zeta = e^{2\pi i/8}$. Then $a + ib = \alpha_0 + (\alpha_1 + \beta_1)\zeta + \beta_0\zeta^2 + (\beta_1 - \alpha_1)\zeta^3$. Note that $|\alpha_1|, |\beta_1| \le M/2$. Therefore, the coefficients of $a + ib$ are bounded by M. The bound for the error follows easily. □

2.2 Convergents

Let (P_n) and (Q_n) be sequences of integers defined by

$$\begin{array}{ll} P_0 = 1,\ P_1 = 1,\ \forall\, n \ge 2: & P_n = P_{n-1} + 2Q_{n-1}, \\ Q_0 = 0,\ Q_1 = 1,\ \forall\, n \ge 2: & Q_n = Q_{n-1} + P_{n-2}. \end{array} \tag{1}$$

The quotients $P_1/Q_1, P_2/Q_2, \ldots$ are called the *convergents* of the continued fraction expansion of $\sqrt{2}$. The following (well-known) results will be useful in analyzing our approximation algorithm.

Lemma 2. *The following assertions hold for $k \ge 1$:*

(a) $(-1 + \sqrt{2})^k = (-1)^k(P_k - Q_k\sqrt{2})$.

(b) $1/(3Q_k) \le |P_k - Q_k\sqrt{2}| \le 1/(2Q_k)$.

(c) $Q_k = \left((1 + \sqrt{2})^k - (1 - \sqrt{2})^k\right)/(2\sqrt{2})$, $P_k = \left((1 + \sqrt{2})^k + (1 - \sqrt{2})^k\right)/2$.

(d) $(1 + \sqrt{2})^k/3 \le Q_k \le (1 + \sqrt{2})^{k-1}$, $P_k \le (1 + \sqrt{2})^k$.

Precomputation: $\varepsilon_0, \ldots, \varepsilon_\ell$ where $\varepsilon_i = (-1+\sqrt{2})^i$.
Input: $\omega \in [0,1]$.
Output: $a \in \mathbb{Z}[\sqrt{2}]_M$, $0 \le \omega - a \le \varepsilon_{\ell-1}$.
$a := 0$
for $k = 1$ **to** ℓ **do**
  while $a = \alpha_0 + \alpha_1\sqrt{2} < \omega$ **do**
    if $(-1)^k\alpha_0 \ge 0$ **then** $a := a + \varepsilon_{k-1}$
      else $a := a + \varepsilon_k$

Figure 1: Approximation By Signatures (**ABS**) for $M = P_\ell$

PROOF. (a) This is obvious from the recursion formulas. (b) The right hand inequality is well known, see [5]. For the left hand inequality observe first that $(-1-\sqrt{2})^k = (-1)^k(P_k + Q_k\sqrt{2})$. Hence, by (a) we get $P_k^2 - 2Q_k^2 = (-1)^k$. The assumption $|P_k - Q_k\sqrt{2}| < \frac{1}{3Q_k}$ would yield the contradiction

$$\begin{aligned} 1 &= |P_k^2 - 2Q_k^2| = |P_k - Q_k\sqrt{2}||P_k + Q_k\sqrt{2}| \\ &< \frac{1}{3Q_k} Q_k \left(2\sqrt{2} + \frac{1}{3Q_k}\right) < \frac{2\sqrt{2}}{3} + \frac{1}{9Q_k} < 1 \end{aligned}$$

for $n \ge 2$. For $n = 1$ the assertion can be verified directly.
(c)+(d) The formulas for Q_k and P_k follow from (1). The inequalities can be easily derived from these formulas. □

3 The Algorithm

For $k \ge 1$ let $\varepsilon_k = (-1+\sqrt{2})^k = (-1)^k(P_k - Q_k\sqrt{2})$. We will use this sequence of numbers to improve our current approximation at each step. The following lemma is the basis of our approximation algorithm.

Lemma 3. *Let $k \ge 1$, $M \ge P_k$, $a = \alpha_0 + \alpha_1\sqrt{2} \in \mathbb{Z}[\sqrt{2}]_M \cap [0,1]$, and $\varepsilon_i := (-1+\sqrt{2})^k$. If $(-1)^k\alpha_0 \ge 0$, then $a + \varepsilon_{k-1} \in \mathbb{Z}[\sqrt{2}]_M$, and if $(-1)^k\alpha_0 < 0$, then $a + \varepsilon_k \in \mathbb{Z}[\sqrt{2}]_M$.*

PROOF. Since $a \in [0,1]$, $\alpha_0 \le 0$ is equivalent to $\alpha_1 \ge 0$. Suppose that $(-1)^k\alpha_0 \ge 0$. Then $\alpha_0(-1)^{k-1}P_{k-1} < 0$, which shows that $|\alpha_0 + (-1)^{k-1}P_{k-1}| < \max\{|\alpha_0|, P_{k-1}\} \le M$. Similarly, $|\alpha_1 - (-1)^{k-1}Q_{k-1}| \le M$. The case $(-1)^k\alpha_0 < 0$ is handled analogously. □

The algorithm *Approximation By Signature* **ABS** is given as pseudocode in Figure 1 and works as follows: starting with $a = 0$ we run through all k between 1 and ℓ and improve our approximation by adding to a the number ε_k or ε_{k-1} depending on the sign of the first coefficient of a. As we are adding up positive numbers, the algorithm will eventually terminate.

Theorem 4. *Algorithm* **ABS** *computes a with $0 \le \omega - a \le (1+\sqrt{2})/P_\ell$ in at most 3ℓ iterations.*

PROOF. Let $M := P_\ell$. Lemma 3 assures that $a \in \mathbb{Z}[\sqrt{2}]_M$. The approximation error at step k can be upper bounded by ε_{k-1} which is less than $(1+\sqrt{2})/P_k$ by Lemma 2(d).

Let us study how many times the inner loop is performed at step $k > 1$. First, notice that for each k the algorithm produces an approximation $a_k \in \mathbb{Z}[\sqrt{2}]_{P_k}$ of ω. Furthermore, $a_k = a_{k-1} + e_0\varepsilon_k + e_1\varepsilon_{k-1}$ for some nonnegative integers e_0 and e_1. The number of times the inner loop is performed at step k is then $e_0 + e_1$. Since $\omega - a_{k-1} \leq \varepsilon_{k-2}$ and $\omega - a_k \geq 0$, we obtain $e_0(-1+\sqrt{2})^2 + e_1(-1+\sqrt{2}) < 1$. This gives $e_1 \leq 2$, $e_0 \leq 2$, and rules out the case $e_0 = e_1 = 2$. Hence, $e_0 + e_1 \leq 3$.

At step 1 the loop is performed at most $\omega/\varepsilon_1 < 3$ times. This implies the assertion. □

For $M = P_\ell$ the error bound given in algorithm **ABS** can be improved to $(1+\sqrt{2})/(2M)$ if we don't require that the approximation be less than ω. Using Lemma 1 this yields an error of $2(1+\sqrt{2})/(\sqrt{2}M) < 3.42/M$ for the approximation of complex numbers. If M is an arbitrary integer, we have to multiply all these bounds by an additional factor of $(1+\sqrt{2})$.

We can also analyze the number of arithmetic operations of our algorithm: since it uses 4 operations in each run of the inner loop (two for the addition of the elements, one for the computation of the value as explained in the next section, and one for computing the distance to ω), we obtain the upper bound of $12\log_{1+\sqrt{2}}(M) < 10\log(M)$ given in the introduction.

4 Implementation Results and Conclusions

In this section, we will report on our implementations of the algorithm **ABS**. We designed a data type for the elements of $\mathbb{Z}[\sqrt{2}]$ consisting of two integer coefficients and a value (IEEE 64-bit double format) corresponding to the value of the integer as a real number. This reduced computing the value of an approximation to floating point addition.

Table 1 summarizes the result of computations on random numbers. The first column corresponds to M, the second contains the theoretical upper bound on the number of iterations as given in Theorem 4, and the third column gives the maximum number of iterations encountered for 1000 runs on random numbers in $[0,1]$. The fourth column shows the upper bound on the worst case error which is equal to $3(1+\sqrt{2})^2/(2M)$, and finally, the last column shows the maximum error encountered during the approximation of the random numbers.

In Table 2 we have given our approximations of a primitive 1024-th root of unity. Use of scaling can reduce the error even further: the idea is to approximate complex numbers by elements from $\mathbb{Z}[e^{2\pi i/8}]_M/L$ for some large number L. Using this technique, we get, e.g., the approximation

$$\frac{1}{47321}(8120 - 2856e^{2\pi i/8} - 4080e^{4\pi i/8} + 8626e^{6\pi i/8})$$

with an error less than 10^{-9}.

Our algorithm runs very fast in practice. For instance, 100000 approximations of random numbers in the interval $[0,1]$ took 0.7 seconds on a SPARC-5. Due to its simple form, it is well suited for real time applications when combined with scaling methods.

Table 1: Number of iterations and observed error for **ABS**

M	comp.	max.	comp. error	max. error
3	3	3	0.4142135624	0.3565025624
7	6	6	0.1715728753	0.1209078753
17	9	9	0.0710678119	0.0217268119
41	12	11	0.0294372515	0.0103675030
99	15	13	0.0121933088	0.0060718692
239	18	15	0.0050506339	0.0047838281
577	21	17	0.0020920411	0.0005971942
1393	24	20	0.0008665518	0.0008221857
3363	27	23	0.0003589375	0.0003230196
8119	30	27	0.0001486768	0.0001285928
19601	33	29	0.0000615839	0.0000583643
47321	36	35	0.0000255089	0.0000174464
114243	39	18	0.0000105661	0.0000103679
275807	42	26	0.0000043766	0.0000042783
665857	45	23	0.0000018129	0.0000016323
1607521	48	32	0.0000007509	0.0000007482
3880899	51	34	0.0000003110	0.0000002396
9369319	54	28	0.0000001288	0.0000001040
22619537	57	33	0.0000000534	0.0000000540
54608393	60	43	0.0000000221	0.0000000190
131836323	63	43	0.0000000092	0.0000000130

Table 2: Approximation of $e^{2\pi i/1024}$

M	coefficients				error
3	0	0	0	0	1.0000000000000
7	2	-1	0	1	0.4142401836883
17	-2	2	0	-2	0.1716637449664
41	8	-5	0	5	0.0713134465244
99	-16	12	0	-12	0.0300515043906
239	42	-29	0	29	0.0136333100745
577	-98	239	-239	99	0.0064553678835
1393	240	-239	99	99	0.0023400844427
3363	-576	1323	-1294	507	0.0011163208005
8119	1394	-1463	676	507	0.0004043586918
19601	-3362	5263	-4080	507	0.0001301259489
47321	8120	-2856	-4080	8626	0.0000435838918
114243	-19600	16745	-4080	-10975	0.0000107648845
275807	94643	-144819	110163	-10975	0.0000046713682
665857	-181164	245231	-165644	-10975	0.0000023025240
1607521	-181164	-225601	500213	-481807	0.0000015774491
3880899	-181164	440256	-441451	184050	0.0000006593521
9369319	2092214	-3911475	3439448	-952639	0.0000003123136
22619537	-7277105	6594533	-2048972	-3696849	0.0000001248819
54608393	-7277105	-2774786	11201246	-13066168	0.0000000485207
131836323	-7277105	19844751	-20787610	9553369	0.0000000113069

References

[1] J. H. Cozzens and L. A. Finkelstein. Computing the discrete Fourier transform using residue number systems in a ring of algebraic integers. *IEEE Transactions on Information Theory*, 31(5):580–588, 1985.

[2] R. A. Games. Complex approximations using algebraic integers. *IEEE Transactions on Information Theory*, 31(5):565–579, 1985.

[3] R. A. Games. An algorithm for complex approximation in $\mathbb{Z}[e^{2\pi i/8}]$. *IEEE Transactions on Information Theory*, 32(4):603–607, 1986.

[4] M. W. Marcellin and Th. R. Fischer. Encoding algorithms for complex approximation in $\mathbb{Z}[e^{2\pi i/8}]$. *IEEE Transactions on Information Theory*, 35(5):1133–1136, September 1989.

[5] I. Niven and H. S. Zuckerman. *An Introduction to the Theory of Numbers*. Wiley, New York, 4th edition, 1980.