Quantum algorithms (CO 781, Winter 2008)
Prof. Andrew Childs, University of Waterloo

LECTURE 4: Hallgren's algorithm for solving Pell's equation

In this and the next lecture, we will explore a final application of the quantum Fourier transform over abelian groups, namely an algorithm discovered by Hallgren for solving a quadratic diophantine equation known as *Pell's equation*. This algorithm is interesting for at least two reasons. First, it gives an application of quantum algorithms to a new area of mathematics, algebraic number theory (and indeed, subsequent work has shown that quantum computers can also efficiently solve other problems in this area). Second, it extends the solution of the abelian HSP to the case of an infinite group, namely the real numbers.

There are two main parts to the quantum algorithm for solving Pell's equation. First, we define a periodic function whose period encodes the solution to the problem. To define this function, we must introduce some notions from algebraic number theory. Second, we show how to find the period of a black-box function defined over the real numbers even when the period is irrational.

Pell's equation Given a squarefree integer d (i.e., an integer not divisible by any perfect square), the Diophantine equation

$$x^2 - dy^2 = 1 \tag{1}$$

is known as *Pell's equation*. This appellation provides a nice example of Stigler's Law of Eponomy in action, as Pell had nothing whatsoever to do with the equation. The misattribution is apparently due to Euler, who confused Pell with a contemporary, Brouncker, who had actually worked on the equation. In fact, Pell's equation was studied in ancient India, where (inefficient) methods for solving it were developed about a century before Pell. (Indeed, Lenstra suggests that most likely, Pell was named after the equation.)

The left hand side of Pell's equation can be factored as

$$x^2 - dy^2 = (x + y\sqrt{d})(x - y\sqrt{d}). \tag{2}$$

Note that a solution of the equation $(x, y) \in \mathbb{Z}^2$ can be encoded uniquely as the real number $x+y\sqrt{d}$: since $\sqrt{d}$ is irrational, $x + y\sqrt{d} = w + z\sqrt{d}$ if and only if $(x, y) = (w, z)$. (Proof: $\frac{x-w}{z-y} = \sqrt{d}$.) Thus we can also refer to the number $x + y\sqrt{d}$ as a solution of Pell's equation.

There is clearly no loss of generality in restricting our attention to *positive* solutions of the equation, namely those for which $x > 0$ and $y > 0$. It is straightforward to show that if $x_1 + y_1\sqrt{d}$ is a positive solution, then $(x_1 + y_1\sqrt{d})^n$ is also a positive solution for any $n \in \mathbb{N}$. In fact, one can show that *all* positive solutions are obtained in this way, where $x_1 + y_1\sqrt{d}$ is the *fundamental solution*, the smallest positive solution of the equation. Thus, even though Pell's equation has an infinite number of solutions, we can in a sense find them all by finding the fundamental solution.

Some examples of fundamental solutions for various values of d are shown in the following table. Notice that while the size of the fundamental solution generally increases with increasing d, the behavior is far from monotonic: for example, x_1 has 44 decimal digits when $d = 6009$, but only 11 decimal digits when $d = 6013$. But it is possible for the solutions to be very large—the size of $x_1 + y_1\sqrt{d}$ is only upper bounded by $2^{O(\sqrt{d}\log d)}$. Thus it is not even possible to *write down* the fundamental solution with $\mathrm{poly}(\log d)$ bits.

d	x_1	y_1
2	3	2
3	2	1
5	9	4
⋮		
13	649	180
14	15	4
⋮		
6009	1316340106327253158925944695105994738840139 75 $\approx 1.3 \times 10^{44}$	1698114661157803451688949237883146576681644 $\approx 1.6 \times 10^{42}$
6013	40929908599	527831340
⋮		

To get around this difficulty, we define the *regulator* of the fundamental solution,

$$R := \ln(x_1 + y_1\sqrt{d}). \tag{3}$$

Since $R = O(\sqrt{d}\log d)$, we can write down $\lceil R \rceil$ using $O(\log d)$ bits. Now R is an irrational number, so determining only its integer part may seem unsatisfactory. But in fact, given the integer part of R, there is a classical algorithm to compute n digits of R in time $\mathrm{poly}(\log d, n)$. Thus it suffices to give an algorithm that finds the integer part of R in time $\mathrm{poly}(\log d)$. The best known classical algorithm for this problem takes time $2^{O(\sqrt{\log d \log\log d})}$ assuming the generalized Riemann hypothesis, or time $O(d^{1/4}\,\mathrm{poly}(\log d))$ with no such assumptions.

A bit of algebraic number theory As mentioned above, there are two main parts to the quantum algorithm for Pell's equation: first, the definition of a periodic function over the reals whose period encodes the regulator, and second, a solution of the period-finding problem in the case where the period might be irrational. We will start by showing how to define the periodic function. To do this, we need to introduce some concepts from algebraic number theory.

Given a squarefree positive integer d, the *quadratic number field* $\mathbb{Q}[\sqrt{d}]$ is defined as

$$\mathbb{Q}[\sqrt{d}] := \{x + y\sqrt{d} : x, y \in \mathbb{Q}\}. \tag{4}$$

You can easily check that this is a field with the usual addition and multiplication operations. We can also define an operation called *conjugation*, defined by

$$\overline{x + y\sqrt{d}} := x - y\sqrt{d}. \tag{5}$$

You can easily check that conjugation of elements of $\mathbb{Q}[\sqrt{d}]$ has many of the same properties as complex conjugation, and indeed $\mathbb{Q}[\sqrt{d}]$ behaves in many respects like $\mathbb{C}$, with $\sqrt{d}$ taking the place of the imaginary unit $\mathrm{i} = \sqrt{-1}$. Defining the ring $\mathbb{Z}[\sqrt{d}] \subset \mathbb{Q}[\sqrt{d}]$ as

$$\mathbb{Z}[\sqrt{d}] := \{x + y\sqrt{d} : x, y \in \mathbb{Z}\}, \tag{6}$$

we see that solutions of Pell's equation correspond to $\xi \in \mathbb{Z}[\sqrt{d}]$ satisfying $\xi\bar{\xi} = 1$.

Notice that any solution of Pell's equation, $\xi \in \mathbb{Z}[\sqrt{d}]$, has the property that its multiplicative inverse over $\mathbb{Q}[\sqrt{d}]$, $\xi^{-1} = \bar{\xi}/\xi\bar{\xi} = \bar{\xi}$, is also an element of $\mathbb{Z}[\sqrt{d}]$. In general, an element of a ring with an inverse that is also an element of the ring is called a *unit*. In $\mathbb{Z}$, the only units are ± 1, but in other rings it is possible to have more units. It should not be a surprise that the set of units of $\mathbb{Z}[\sqrt{d}]$ is closely related to the set of solutions of Pell's equation. Specifically, we have

Proposition. *$\xi = x + y\sqrt{d}$ is a unit in $\mathbb{Z}[\sqrt{d}]$ if and only if $\xi\bar{\xi} = x^2 - dy^2 = \pm 1$.*

Proof. We have

$$\xi^{-1} = \frac{\bar{\xi}}{\xi\bar{\xi}} = \frac{x - y\sqrt{d}}{x^2 - dy^2}. \tag{7}$$

If $x^2 - dy^2 = \pm 1$, then clearly $\xi^{-1} \in \mathbb{Z}[\sqrt{d}]$. Conversely, if $\xi^{-1} \in \mathbb{Z}[\sqrt{d}]$, then so is

$$\xi^{-1}\overline{\xi^{-1}} = \frac{(x - y\sqrt{d})(x + y\sqrt{d})}{(x^2 - dy^2)^2} = \frac{1}{x^2 - dy^2}, \tag{8}$$

which shows that $x^2 - dy^2 = \pm 1$. □

It is not hard to show that the set of all units in $\mathbb{Z}[k]$ is given by $\{\pm\epsilon_1^n : n \in \mathbb{Z}\}$, where ϵ_1 is the *fundamental unit*, the smallest unit greater than 1. The proof is essentially the same as the proof that all solutions of Pell's equation are powers of the fundamental solution.

If we can find ϵ_1, then it is straightforward to find all the solutions of Pell's equation. If $\epsilon_1 = x + y\sqrt{d}$ has $x^2 - dy^2 = +1$, then the units are precisely the solutions of Pell's equation. On the other hand, if $x^2 - dy^2 = -1$, then $\epsilon_2 := \epsilon_1^2$ satisfies $\epsilon_2\bar{\epsilon}_2 = \epsilon_1^2\bar{\epsilon}_1^2 = (-1)^2 = 1$; in this case the solutions of Pell's equation are $\{\pm\epsilon_1^{2n} : n \in \mathbb{Z}\}$. Thus our goal is to find ϵ_1. Just as in our discussion of the solutions to Pell's equation, ϵ_1 is too large to write down, so instead we will compute the *regulator of the fundamental unit*, $\mathcal{R} := \ln \epsilon_1$.

To define a periodic function that encodes $\mathcal{R}$, we need to introduce the concept of an *ideal* of a ring (and more specifically, a *principal ideal*). For any ring R, we say that $I \subseteq R$ is an ideal if it is closed under integer linear combinations and under multiplication by arbitrary elements of of R. For example, $2\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

We say that an ideal is *principal* if it is generated by a single element of the ring, i.e., if it is of the form αR for some $\alpha \in R$. In the example above, $2\mathbb{Z}$ is a principal ideal. (Not all ideals are principal; for example, consider $x\mathbb{Z}[x, y] + y\mathbb{Z}[x, y] \subseteq \mathbb{Z}[x, y]$, an ideal in the ring of polynomials in x, y with integer coefficients.)

A periodic function for the units of $\mathbb{Z}[\sqrt{d}]$ Principal ideals are useful because the function mapping the ring element $\xi \in \mathbb{Z}[\sqrt{d}]$ to the principal ideal ξR is periodic, and its periodicity corresponds to the units of $\mathbb{Z}[\sqrt{d}]$. Specifically, we have

Proposition. *$\xi\mathbb{Z}[\sqrt{d}] = \zeta\mathbb{Z}[\sqrt{d}]$ if and only if $\xi = \zeta\epsilon$ where ϵ is a unit in $\mathbb{Z}[\sqrt{d}]$.*

Proof. If ϵ is a unit, then $\xi\mathbb{Z}[\sqrt{d}] = \zeta\epsilon\mathbb{Z}[\sqrt{d}] = \zeta\mathbb{Z}[\sqrt{d}]$ since $\epsilon\mathbb{Z}[\sqrt{d}] = \mathbb{Z}[\sqrt{d}]$ by the definition of a unit. Conversely, suppose that $\xi\mathbb{Z}[\sqrt{d}] = \zeta\mathbb{Z}[\sqrt{d}]$. Since $1 \in \mathbb{Z}[\sqrt{d}]$, $\xi \in \xi\mathbb{Z}[\sqrt{d}] = \zeta\mathbb{Z}[\sqrt{d}]$, so there is some $\mu \in \mathbb{Z}[\sqrt{d}]$ satisfying $\xi = \zeta\mu$. Similarly, $\zeta \in \zeta\mathbb{Z}[\sqrt{d}] = \xi\mathbb{Z}[\sqrt{d}]$, so there is some $\nu \in \mathbb{Z}[\sqrt{d}]$ satisfying $\zeta = \xi\nu$. Thus we have $\xi = \zeta\mu = \xi\nu\mu$. This shows that $\nu\mu = 1$, so μ and ν are units (indeed, $\nu = \mu^{-1}$). □

Thus the function $g(\xi) = \xi\mathbb{Z}[\sqrt{d}]$ is (multiplicatively) periodic with period ϵ_1. In other words, letting $\xi = e^z$, the function

$$h(z) = e^z\mathbb{Z}[\sqrt{d}] \tag{9}$$

is (additively) periodic with period $\mathcal{R}$. However, we cannot simply use this function since it is not possible to succinctly represent the values it takes.

To define a more suitable periodic function, Hallgren uses the concept of a *reduced* ideal, and a way of measuring the distance between principal ideals. The definition of a reduced ideal is rather technical, and we will not go into the details. For our purposes, it is sufficient to note that there are only finitely many reduced principal ideals, and in fact only $O(d)$ of them, so we can represent a reduced principal ideal using $\mathrm{poly}(\log d)$ bits.

Hallgren also uses a function that measures the distance of any principal ideal from the *unit ideal*, $\mathbb{Z}[\sqrt{d}]$. This function is defined as

$$\delta(\xi\mathbb{Z}[\sqrt{d}]) := \ln\left|\frac{\xi}{\bar{\xi}}\right| \bmod \mathcal{R}. \tag{10}$$

Notice that the unit ideal has distance $\delta(1\mathbb{Z}[\sqrt{d}]) = \ln|1/1| \bmod \mathcal{R} = 0$, as required. Furthermore, the distance function does not depend on which generator we choose to represent an ideal, since (by the above proposition) two equivalent ideals have generators that differ by some unit ϵ, and

$$\delta(\epsilon\mathbb{Z}[\sqrt{d}]) = \ln\left|\frac{\epsilon}{\bar{\epsilon}}\right| \bmod \mathcal{R} = \ln\left|\frac{\epsilon}{\epsilon^{-1}}\right| \bmod \mathcal{R} = \ln|\epsilon^2| \bmod \mathcal{R} = 2\ln|\epsilon| \bmod \mathcal{R} = 0. \tag{11}$$

With this definition of distance, one can show that the reduced ideals are not too far apart, so that there is a reduced ideal close to any non-reduced ideal.

The periodic function used in Hallgren's algorithm, $f(z)$, is defined as the reduced principal ideal whose distance from the unit ideal is maximal among all reduced principal ideals of distance at most z (together with the distance from z, to ensure that the function is one-to-one within each period). In other words, we select the reduced principal ideal "to the left of or at z".

This function is periodic with period $\mathcal{R}$, and can be computed in time $\mathrm{poly}(\log d)$. Thus it remains to show how to perform period finding when the period of the function might be irrational.