

MPRI – Cours 2-12-2



F. Morain



Lecture II: Primality

2011/09/20

I. Compositeness tests

Z) Definition and classification.

A) Fermat.

B) Euler-Solovay/Strassen.

C) Artjuhov-Miller-Rabin.

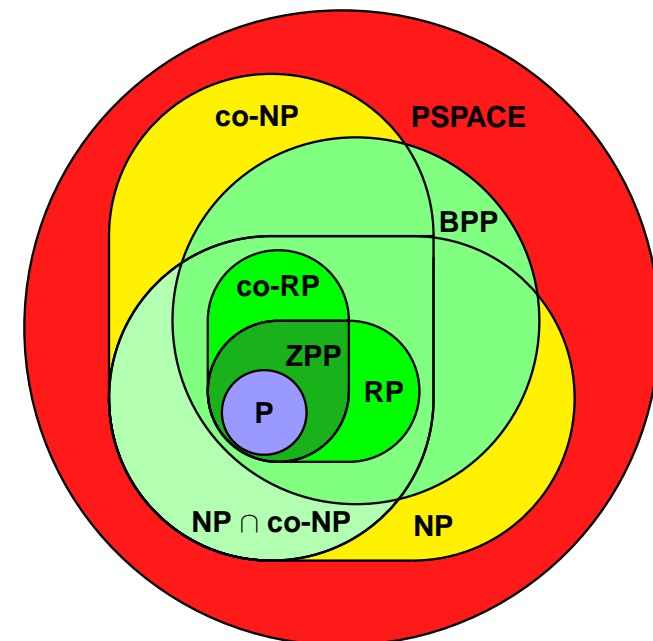
D) Other tests.

Plan

I. Compositeness tests.

II. Primality tests.

Z) Definition and classification



My view of randomized algorithms

Def. A **Monte Carlo algorithm** for deciding that $X \in \mathbb{A}$ returns **yes** or **I don't know**:

$$\text{Proba}(\text{"yes"} \mid X \notin \mathbb{A}) = 0$$

$$\text{Proba}(\text{"I don't know"} \mid X \in \mathbb{A}) \leq 1 - \delta, \text{ for absolute } 0 < \delta < 1.$$

Def. A decision problem is in **RP** if there exists a polynomial time Monte Carlo algorithm that solves it.

Rem. \neq error on the answer; or a failure in the computer.

Def. A **Las Vegas algorithm** answers **yes**, **no** or **I don't know**:

$$\text{Proba}(\text{"yes"} \mid X \notin \mathbb{A}) = 0, \quad \text{Proba}(\text{"no"} \mid X \in \mathbb{A}) = 0$$

$$\text{Proba}(\text{"I don't know"} \mid X \in \mathbb{A}) \leq 1 - \delta.$$

Def. $\text{ZPP} = \text{RP} \cap \text{co-RP}$.

Compositeness test: deciding that N is composite.

Primality test: deciding that N is prime.

The test

function isComposite(N)

1. Choose a at random in $\mathbb{Z}/N\mathbb{Z} - \{0\}$.
2. Compute $g = \text{gcd}(a, N)$; if $g > 1$, **then** return (**yes**, $g \mid N$).
3. if $a^{N-1} \not\equiv 1 \pmod{N}$, **then** return (**yes**, a)
otherwise return **I don't know**.

Cost. $O((\log N)M(\log N))$; typically $O((\log N)^3)$, asymptotically $\tilde{O}((\log N)^2)$.

Prop. $\text{Proba}(\text{"I don't know"}) = P(N)/(N-1)$.

Proof. Probability of yes is:

$$\left(1 - \frac{\varphi(N)}{N-1}\right) + \frac{\varphi(N)}{N-1} \left(1 - \frac{P(N)}{\varphi(N)}\right). \square$$

Rem. if N is prime, proba is 1...!

A) Fermat

Idea: if $\text{gcd}(a, N) = 1$, then $a^{N-1} \equiv 1 \pmod{N}$.

But: $2^{340} \equiv 1 \pmod{341}$: **pseudoprime** to base 2 (psp-2).

Thm. There exists an infinite number of psp-2 numbers.

Thm. (Pomerance) For $x \geq x_0$: $x^{2/7} \ll P_2(x) \ll xL(x)^{-1/2}$ with $L(x) = \exp\{\log x \log \log x / \log \log x\}$.

Def. $P(N) = \#\{a \in (\mathbb{Z}/N\mathbb{Z})^*, a^{N-1} \equiv 1 \pmod{N}\}$.

Thm. If $N = \prod_i p_i^{\alpha_i}$, $P(N) = \prod_i \text{gcd}(p_i - 1, N - 1)$.

Proof: ■

Carmichael numbers

Def. composite N s.t. $P(N) = \varphi(N)$.

Ex. 541.

Rem. $P(N)/(N-1) = \varphi(N)/(N-1)$ close to 1.

Thm.(Alford, Granville, Pomerance, 1992) There are infinitely many Carmichael numbers.

More properties of Carmichael numbers:

1. N is squarefree.
2. For all $p \mid N$, $p-1 \mid N-1$ (equivalently $\lambda(N) \mid N-1$).
3. N has at least three prime factors.

B) Euler and Solovay-Strassen

Idea: (Euler) if N is prime and $\gcd(a, N) = 1$, then $a^{(N-1)/2} \equiv \left(\frac{a}{N}\right) \pmod{N}$.

Pb: $2^{(1105-1)/2} \equiv \left(\frac{2}{1105}\right) \pmod{1105}$; this is an Euler pseudoprime to base 2 (epsp-2). There are an infinite number of them.

Prop. $E_2(x) \leq P_2(x)$.

Def. $\mathcal{E}(N) = \{a \in (\mathbb{Z}/N\mathbb{Z})^*, a^{(N-1)/2} \equiv \left(\frac{a}{N}\right) \pmod{N}\}$; $E(N) = \#\mathcal{E}(N)$.

Prop. $\mathcal{E}(N)$ is proper subgroup of $(\mathbb{Z}/N\mathbb{Z})^*$.

Coro. $E(N)/\varphi(N) \leq 1/2$.

The test

function isComposite2(N)

1. Choose a at random in $\mathbb{Z}/N\mathbb{Z} - \{0\}$.
2. Compute $g = \gcd(a, N)$; if $g > 1$, then return (yes, $g \mid N$).
3. If $a^{(N-1)/2} \not\equiv \left(\frac{a}{N}\right) \pmod{N}$ then return (yes, a)
else return I don't know.

Prop. $\text{Proba}(\text{"I don't know"}) = E(N)/(N-1) \leq 1/2$.

Coro. isComposite? \in RP (hence isPrime? \in co-RP).

Miller (1975): $a = 2, 3, \dots$; Ankeny–Montgomery–Lenstra–Bach: if an adequate Riemann hypothesis is true, then the smallest witness is $< 2(\log N)^2$, yielding a deterministic $O((\log N)^3 M(\log N))$ algorithm.

The exact value of $E(N)$

Thm. (Monier) Write $N = \prod_{i=1}^k p_i^{\alpha_i}$ where p_i are distinct odd primes, $\alpha_i \geq 1$. Write $N = 1 + 2^s t$ with t odd and $p_i = 1 + 2^{s_i} t_i$ with t_i odd. Assume $s_1 \leq s_2 \leq \dots \leq s_k$ and put $T_i = \gcd(t, t_i)$, $n_i = \gcd((N-1)/2, p_i - 1)$ and $\mathcal{N} = \prod_i n_i$. Then

$$E(N) = \delta(N) \mathcal{N}$$

where

$$\delta(N) = \begin{cases} 2 & \text{if } s = s_1 \\ 1/2 & \text{if } \exists i, \alpha_i \text{ odd and } s_i < s \\ 1 & \text{otherwise.} \end{cases}$$

Proof. exercise.

C) Artjuhov-Miller-Rabin

N being odd, write $N - 1 = 2^s t$ with $s \geq 1$ and odd t .

$$a^{N-1} - 1 = (a^t - 1)(a^{2t} + 1) \dots (a^{2^{s-1}t} + 1)$$

$$(AMR_a) : a^t \equiv 1 \pmod{N} \text{ or } \exists j, 0 \leq j < s, a^{2^j t} \equiv -1 \pmod{N}.$$

Pb: $N = 2047 = 23 \times 89$ is s.t. $N - 1 = 2 \times 1023$ and $2^{(N-1)/2} \equiv 1 \pmod{N}$: strong-pseudoprime to base 2 (spsp-2).

Thm. spsp- $a \Rightarrow$ epsp- a .

Def. $F(N) = \#\{a \in (\mathbb{Z}/N\mathbb{Z})^*, (AMR_a) \text{ is satisfied}\}$.

Thm. (Monier)

$$F(N) = \left[1 + \frac{2^{ks_1} - 1}{2^k - 1} \right] \prod_{i=1}^k T_i.$$

Coro. $F(N)/(N-1) \leq 1/4$.

The test

function isComposite3(N)

1. Choose a at random in $\mathbb{Z}/N\mathbb{Z} - \{0\}$.
2. Compute $g = \gcd(a, N)$; **if** $g > 1$, **then** return (yes, $g \mid N$).
3. **If** (AMR_a) **then** return (yes, a)
else return I don't know.

Prop. Proba("I don't know") = $F(N)/(N - 1)$.

Building primes?

function randomProbablePrime(b)

repeat

choose odd N at random in $[2^{b-1}, 2^b[$
until N passes k tests.

$p_{b,k} = \text{Proba}(X = N \text{ is composite} \mid Y_k = N \text{ passes } k \text{ tests}) = ?$

Rem. What we know is

$\text{Proba}(Y_k = N \text{ passes } k \text{ tests} \mid X = N \text{ is composite}) \leq (1/4)^k$.

Thm. (Burthe, 1996) $\forall b \geq 2, \forall k \geq 1, p_{b,k} \leq 4^{-k}$.

Numerical tables

x	$P_2(x)$	$E_2(x)$	$F_2(x)$	$C(x)$	$\pi(x)$
10^4	22	12	5	7	1229
10^5	78	36	16	16	9592
10^6	245	114	46	43	78498
10^7	750	375	162	105	664579
10^8	2057	1071	488	255	5761455
10^9	5597	2939	1282	646	50847534
10^{10}	14884	7706	3291	1547	455052511
25×10^9	21853	11347	4842	2163	1091987405
10^{11}	38975	20417	8607	3605	4118054813
10^{12}	101629	53332	22407	8241	37607912018
10^{13}	264239	139597	58897	19279	346065536839
10^{14}				44706	3204941750802
10^{15}				105212	29844570422669
10^{16}				246683	279238341033925

Other tests

Goal: reduce the non-answer probability while keeping the computations fast.

- Algebraic extensions: Lucas (degree 2), Adams & Shanks, Gurak.
- Elliptic curves: Gordon.
- Combinations of the preceding: no examples known of $\text{spsp-}a$ and Lucas pseudoprime, for instance.
- Frobenius pseudoprimes à la Grantham: $\leq 1/7710$. Cf. also Zhang.

II. Primality tests

A) Fermat

B) En route for **P**.

C) Agrawal, Kayal, Saxena.

Example of use

Hyp. We know how to find all prime factors < 20 .

$$\begin{aligned} N_0 &= 100003, & N_0 - 1 &= 2 \times 3 \times 7 \times N_1, \\ N_1 &= 2381, & N_1 - 1 &= 2^2 \times 5 \times 7 \times 17 \end{aligned}$$

p	2	5	7	17
$3^{(N_1-1)/p} \bmod N_1$	2380	1347	1944	949

$\Rightarrow N_1$ is prime

$$s = N_1 > \sqrt{N_0}$$

$$2^{N_0-1} \equiv 1 \pmod{N_0}, \gcd(2^{(N_0-1)/N_1} - 1, N_0) = 1$$

$\Rightarrow N_0$ is prime

Rem. We have got a (recursive) primality proof of depth $O(\log N)$.

A) Fermat

Thm. N is prime if and only if $(\mathbb{Z}/N\mathbb{Z})^*$ is cyclic of order $N - 1$:

$$\left. \begin{aligned} a^{N-1} &\equiv 1 \pmod{N} \\ \forall p \mid N-1, a^{\frac{N-1}{p}} &\not\equiv 1 \pmod{N} \end{aligned} \right\} \Rightarrow N \text{ is prime}$$

Certificate: $(N, \{p \mid N-1\}, a) \Rightarrow \text{isPrime?} \in \text{NP}$.

Thm. (Pocklington, 1914) Let s s.t. $s \mid N-1$

$$\left. \begin{aligned} a^{N-1} &\equiv 1 \pmod{N} \\ \forall q \text{ prime} \mid s, \gcd(a^{\frac{N-1}{q}} - 1, N) &= 1 \end{aligned} \right\} \Rightarrow \forall p \mid N, p \equiv 1 \pmod{s}$$

Coro. $s > \sqrt{N} \Rightarrow N$ is prime.

Rem. factorisation is not polynomial time in the classical world (see later), but polynomial quantic; search for a is not either (except if Riemann is true or randomized approach).

The $N + 1$ test

For a_0 and a_1 integers, let:

$$A_N = A_N(a_0, a_1) = \mathbb{Z}/N\mathbb{Z}[T]/(T^2 + a_1T + a_0)$$

and $\Delta = a_1^2 - 4a_0$.

Elements of A_N are $u + v\alpha$ with u, v dans $\mathbb{Z}/N\mathbb{Z}$, computations made using $\alpha^2 = -a_1\alpha - a_0$.

Thm. Let p be a prime $\nmid \Delta$.

- if $(\Delta/p) = -1$, then $A_p \sim \mathbb{F}_{p^2}$;
- if $(\Delta/p) = +1$, then $A_p \sim \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

Proof : If $(\Delta/p) = -1$, $T^2 + a_1T + a_0$ is irreducible, hence we recover the classical construction of \mathbb{F}_{p^2} .

If $(\Delta/p) = +1$, $T^2 + a_1T + a_0 = (T - u)(T - v)$ with $u \not\equiv v \pmod{p}$.

Therefore

$$A_p \sim (\mathbb{Z}/p\mathbb{Z})[T]/(T - u) \times (\mathbb{Z}/p\mathbb{Z})[T]/(T - v) \sim \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}. \square$$

Thm. Let N be an odd integer. Assume that we found a_0, a_1 s.t. $\Delta = a_1^2 - 4a_0$ satisfies $(\Delta/N) = -1$. Write $N + 1 = \prod_i q_i^{\beta_i}$. Suppose we have found $\theta \in A_N = A_N(a_0, a_1)$ s.t.

$$\theta^{N+1} = 1 \text{ in } A_N,$$

and for all i :

$$\theta^{(N+1)/q_i} = u_i + v_i\alpha \text{ with } (u_i - 1, v_i, N) = 1.$$

Then N is prime.

Proof : assume N is composite and let $p \mid N$ with $p \leq \sqrt{N}$.

Reduce $A_N \bmod p$ towards A_p :

$$\tau = \theta \bmod p = (u \bmod p) + (v \bmod p)\alpha.$$

We get

$$\tau^{N+1} = 1 \text{ in } A_p,$$

and

$$\tau^{(N+1)/q_i} \neq 1 \text{ in } A_p$$

which proves τ has ordre $N + 1$ in $(A_p)^*$.

Hence $N + 1 \leq \#A_p = p^2$, contradiction. \square

Remarks

- Pocklington-like theorems exist.
- Deduce from this the degree 2 pseudoprimes.
- All this can be reformulated in terms of Lucas sequences (bouhhhh!).
- **Lucas-Lehmer**: $M_m = 2^m - 1$ is prime iff for $L_0 = 4$, $L_{n+1} = L_n^2 - 2 \bmod M_m$, one has $L_{m-2} = 0$ [using $\sqrt{3}$].
 \Rightarrow largest known primes, e.g., $M_{43112609}$ with 12, 978, 189 decimal digits.

Lower bound (?) for primality proving algorithms:
 $O((\log N)M(M_p))$ (super fast arithmetic!).

Choosing θ : using $\bar{\alpha} = -a_1 - \alpha$ (conjugate), enough to choose

$$\theta = \frac{\alpha + m}{\bar{\alpha} + m} = \frac{(m^2 - a_0) + (2m - a_1)\alpha}{m(m - a_1) + a_0}$$

for varying m .

Ex. Consider $N = 101$; $N + 1 = 2 \times 3 \times 17$. Take $a_1 = -2$, $a_0 = -1$, $\Delta = 8$ and $(\frac{8}{101}) = (\frac{2}{101}) = -1$. Take $\theta = 1 + 2\alpha$ (using $m = 1$)

$$\theta^{102} = 1, \theta^{102/2} = 100, \gcd(100 - 1, N) = 1,$$

$$\theta^{102/3} = 47T + 3, \gcd(3 - 1, 47, N) = 1,$$

$$\theta^{102/17} = 23T + 85, \gcd(85 - 1, 23, N) = 1.$$

B) En route for P

- Gauss and Jacobi sums: L. Adleman, C. Pomerance, S. Rumely (1980, 1983); H. Cohen, H. W. Lenstra, Jr (1981 – 1984); H. Cohen, A. K. Lenstra (1982, 1987). W. Bosma & M.-P. van der Hulst (1990); P. Mihăilescu (1998). **deterministic** $O((\log N)^{c_1 \log \log \log N})$.
- almost **RP**: Goldwasser and Kilian using elliptic curves (1986); practical algorithm by Atkin (1986; later FM). See Smith's part.
- **RP**: Adleman and Huang using hyperelliptic curves (1986ff). See Smith's part.

First idea: (Agrawal, Biswas – 1999)

Prop. N is prime iff $P(X) = (X + 1)^N - X^N - 1 \equiv 0 \pmod{N}$.

In practice: choose $Q(X) \in \mathbb{Z}/N\mathbb{Z}[X]$ at random of degree $O(\log N)$. If

$$(X + 1)^N \not\equiv X^N + 1 \pmod{(Q(X), N)}$$

then N is composite.

The probability of failure is bounded by $1 - 1/(4 \log N)$.

Conjecture: If N is composite, there exists $1 \leq r \leq \log N$ s.t. $P(X)$ is not divisible by $X^r - 1$ modulo N .

Analysis

Cost: s computations of X^N modulo $(X^r - 1, N)$; one computation costs $O(\log N)$ products of degree r polynomials, hence:

$$O(s(\log N)M_P(r)M(\log N)).$$

Prop. If $s = \lfloor 2\lfloor \sqrt{r} \rfloor \log N / \log 2 \rfloor + 1$ and $q \geq 2s$, then

$$\binom{q - 1 + s}{s} > N^{2\lfloor \sqrt{r} \rfloor}.$$

Proof:

$$\binom{q - 1 + s}{s} > (q/s)^s \geq 2^s > N^{2\lfloor \sqrt{r} \rfloor}.$$

Coro. $O((\log N)^{2r^{1/2}}M_P(r)M(\log N))$.

Analytical number theory: we can find $r = (\log N)^{2/(2\delta-1)}$ for $\delta \in]0.5, 0.676]$.

Thm. Let N, s be integers, r a prime number and $q = P(r - 1)$. If:

(0)

$$\binom{q - 1 + s}{s} > N^{2\lfloor \sqrt{r} \rfloor};$$

(i) $N \neq M^k, k > 1$;

(ii) N has no prime factor $\leq s$;

(iii) $N^{(r-1)/q} \pmod{r} \notin \{0, 1\}$;

(iv) $\forall a, 1 \leq a \leq s, (X - a)^N \equiv X^N - a \pmod{(X^r - 1, N)}$;

then N is prime.

For a proof, see FM's Bourbaki article.

At last...

Using $r = (\log N)^{2/(2\delta-1)}$.

Coro. There exists a deterministic primality proving algorithm whose running time is

$$O\left((\log N)^{(8\delta+1)/(2\delta-1)}\right)$$

using $M_P(r) = r^2, M(\log N) = (\log N)^2$; and

$$\tilde{O}\left((\log N)^{6\delta/(2\delta-1)}\right)$$

with $M_P(r) = \tilde{O}(r), M(\log N) = \tilde{O}(\log N)$.

Proof:

$$L^2 r^{1/2} M_P(r) M(\log N) = L^{2+2/(2\delta-1)+4/(2\delta-1)+2} \square$$

Ex. (AKS original) $\delta = 2/3$: 19, 12; $\delta = 1$ (Sophie Germain): 9, 6.

Rem. Jacobi $O((\log N)^c \log \log \log N)$, ECPP $\tilde{O}((\log N)^4)$.

Rem. Non effective.

What next?

- cf. D. Bernstein homepage for more on the history of improvements to the basic test.
- Including: **H. W. Lenstra, Jr.** ($\tilde{O}_{\text{eff}}((\log N)^{12})$ or $\tilde{O}((\log N)^8)$), **S. David**.
- Cleaner version of **AKS**: $\tilde{O}_{\text{eff}}((\log N)^{10.5})$ or $\tilde{O}((\log N)^{7.5})$.
- **H. W. Lenstra, C. Pomerance** : $\tilde{O}_{\text{eff}}((\log N)^6)$.
- **P. Berrizbeitia / Q. Cheng** :
Let r prime s.t. $r^\alpha \parallel N - 1$, $r \geq \log^2 N$; $1 < a < N$ s.t.
 $a^{r^\alpha} \equiv 1 \pmod N$, $\gcd(a^{r^{\alpha-1}} - 1, N) = 1$,
 $(X + 1)^N = X^N + 1 \pmod{(X^r - a, N)}$, then N is prime. Heuristic complexity would be $\tilde{O}((\log N)^4)$ for these numbers.
- **D. Bernstein, P. Mihăilescu**: use $e \mid N^d - 1$; inject cyclotomic ideas, $\tilde{O}((\log N)^4)$.

What's left to be done?

Open questions:

- Is $\tilde{O}((\log N)^4)$ the best running time for all numbers?
Compare: $\tilde{O}((\log N)^2)$ for Fermat or Mersenne numbers.
Claim (Lukes, Patterson, Williams): $\tilde{O}((\log N)^3)$ under GRH? (pseudosquares or pseudocubes).
- Combination of tests?

V. Conclusions for primality

Which algorithm?

- **easy to understand / implement, fast**: compositeness tests;
- **fast, proven**: Jacobi;
- **fast, heuristic**: ECPP;
- **certificate**: ECPP;
- **deterministic polynomial**: AKS.

D. Bernstein has an AKS example for $2^{1024} + 643$ (13 hours on 800 MHz PC, 200 Mb memory).

To be compared to FASTECPP:

14/07/03: FM, **7000dd** with mpifastECPP.

19/08/03: J. Franke, T. Kleinjung, T. Wirth, **10000dd**.

06/06: FM, **20,562 dd** with mpifastECPP.