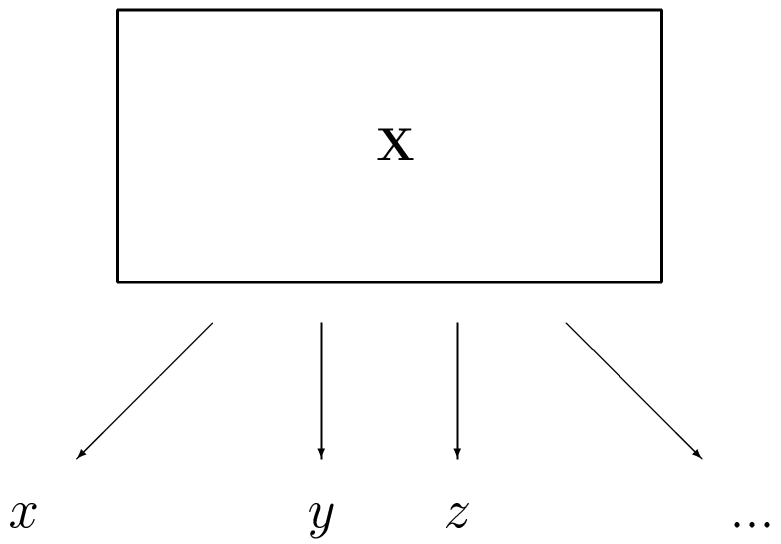
Black Box Groups

Lyon, 10 April 2003

Alexandre Borovik

Black Box



An example when we are not given actual elements, only partial information:

Galois group of a polynomial

$$\begin{tikzcd}[column sep=small]
 & & \boxed{\mathbf{f}(\mathbf{x}) \in \mathbb{Z}[\mathbf{x}]} \arrow[dll] \arrow[dl] \arrow[d] \arrow[dr] & & \\
\mathrm{mod}\ p & \mathrm{mod}\ q & \mathrm{mod}\ r & \ldots &
\end{tikzcd}$$

Reduction modulo random prime number p:

$$\bar{f}(x) = \bar{f}_1 \cdots \bar{f}_k, \quad \deg \bar{f}_i = d_i,$$

f_i irreducible over $\mathbb{F}_p$

$(d_1, \ldots, d_k)$ cycle structure of random $g \in \mathrm{Gal}(f)$

$f(x) \in \mathbb{Z}[x]$ irreducible polynomial, $\deg f = n$

Reduction modulo random prime number p:

$$\bar{f}(x) = \bar{f}_1 \cdots \bar{f}_k, \quad \deg \bar{f}_i = d_i,$$

$$f_i \text{ irreducible over } \mathbb{F}_p$$

Čebotarev Density Theorem: $(d_1, \ldots, d_k)$ are lengths of cycles of a *random* element from $G = \mathrm{Gal}(f)$ in its action on the roots of f.

This is a black box!

$\mathrm{Sym_n}$ and $\mathrm{Alt_n}$ as Galois groups

$f(x) \in \mathbb{Z}[x]$ irreducible polynomial, $\deg f = n$

Generically: $\mathbf{Gal}(f) = \mathrm{Sym}_n$ or Alt_n.

How can one check that $f(x)$ is generic?

$\mathbf{Gal(f) = Sym_n}$ or $\mathbf{Alt_n}$?

Reduction modulo random prime number p:

$$\bar{f}(x) = \bar{f}_1 \cdots \bar{f}_k, \ \deg \bar{f}_i = d_i, \ \bar{f}_i \text{ irreducible}$$

$(d_1, \ldots, d_k)$ cycle structure of random $g \in \mathrm{Gal}(f)$

Davenport & Smith: If $G = \mathrm{Sym}_n$ or Alt_n then, with probability $\frac{\log 2}{\log n}$, we get a cycle of prime length $n/2 < q < n$,

which confirms that

G is at least 4-transitive (**Jordan**)

C. Jordan: If a transitive permutation group on n letters contains a cycle of prime length

$$n/2 < q < n$$

then G is at least 4-transitive

Classification of Finite Simple Groups (CFSG):

4-transitive groups are

Sym_n, Alt_n or a Mathieu group: M_{11}, M_{12}, M_{22}, M_{23}, M_{24}.

Elementary observations:

If $\mathrm{Gal}(F) = \mathrm{Sym}_n$,

- an n-cycle occurs with probability $1/n$.
- A permutation with 2 cycles

  $$(p_1, p_2), \quad p_1 + p_2 = n$$

  occurs with probability

  $$\frac{1}{p_1 p_2}$$

In general: every finite simple group G has "frequent" elements of orders specific for G.

Typical Example: Matrix Groups

$$\mathbf{X} = \langle x_1, \ldots, x_k \rangle \leqslant GL_n(\mathbb{F}_q)$$

is the matrix group of big dimension, so that $|\mathbf{X}|$ is astronomical.

- Statistical study of 'random' products of $x_1, \ldots, x_k$ is the only known approach to identification of $\mathbf{X}$.
- Determination of orders involves either
  - factorisation of integers into primes, or
  - the discrete logarithm problem over finite fields.

Orders of elements in $GL_n(q)$:

- $|x|$ is the minimal **divisor** d of $|GL_n(q)|$ such that
$$x^d = 1$$
- Factorisation is hard
- Exponentiation is easy:
$$\begin{aligned} x^{100} &= x^{2^6+2^5+2^2} \\ &= x^{2^{2^{2^{2^{2^2}}}}} \cdot x^{2^{2^{2^{2^2}}}} \cdot x^{2^2} \end{aligned}$$

x^d requires $< 2\log_2 d$ multiplications

Construction of black boxes for:

$$\mathbf{X} \leqslant GL_N(\mathbb{F}_q),$$

$$\mathbf{X} = \langle x_1, \ldots, x_k \rangle$$

Product Replacement Algorithm: Celler / Leedham-Green / Murray / Niemeyer / O'Brian

Product replacement algorithm

- Start with $(x_1, \ldots, x_k)$ generating $\mathbf{X}$
- Randomly replace

$$(x_1, \ldots, x_i, \ldots, x_k) \longleftarrow (x_1, \ldots, x_j^{\pm 1} x_i, \ldots, x_k)$$

$$(x_1, \ldots, x_i, \ldots, x_k) \longleftarrow (x_1, \ldots, x_i x_j^{\pm 1}, \ldots, x_k)$$

- select random x_i.

Product replacement algorithm:

random walk on the graph:

$\Gamma_k(G)$ set of generating k-tuples

graph with edges:

$$(g_1, \dots, g_i, \dots, g_k) \longrightarrow (g_1, \dots, g_j^{\pm 1} g_i, \dots, g_k)$$

$$(g_1, \dots, g_i, \dots, g_k) \longrightarrow (g_1, \dots, g_i g_j^{\pm 1}, \dots, g_k)$$

Product Replacement Algorithm:

- Very good practical performance
- Biased in artificial examples, say, in
$$X = (Alt_n)^{n!/8}$$
(Babai-Pak).
- In general, $\Gamma_k(G)$ is not connected
- $\Gamma_k(G)$ is connected for large $k > 2\log_2|G|$ (Babai). In that case
$$\text{the diameter} < C \cdot \log^2|G|$$
- **Conjecture.** If G is simple, $\Gamma_k(G)$ is connected for $k \geqslant 3$.

Product Replacement Algorithm:

- **Pak:** If G_n are simple, $|G_n| \longrightarrow \infty$, then $\Gamma_k(G)$ has a connected component $\Gamma'_k(G)$,

$$\frac{|\Gamma'_k(G)|}{|\Gamma_k(G)|} \longrightarrow 1$$

- **Pak:** If k is sufficiently big, mixing time of a random walk on $\Gamma_k(G)$ is bounded by a polynomial in k and $\log |G|$.

Lubotzky–Pak:

If $Aut\,F_k$ satisfies Kazhdan property (T), then mixing time of a random walk on a component of $\Gamma_k(G)$

$$\text{mix} \leqslant C(k) \cdot \log_2 |G|.$$

Conjecture. For $k \geqslant 4$, $Aut\,F_k$ has (T).

Questions about a black box group X:

Identification problem:

Determine the isomorphism type of $\mathbf{X}$ with given degree of certainty.

Verification problem:

Is $\mathbf{X}$ isomorphic to the given group $\mathbb{G}$?

Constructive recognition:

Find an explicit isomorphism

$$\mathbf{X} \longrightarrow \mathbb{G}$$

Bratus & Pak, Constructive recognition of $\mathbf{Sym_n}$, assuming that we can determine the orders of the elements.

We suspect that $\mathbf{X} \simeq \mathrm{Sym}_n$, $n = 2k$.

How can we **prove** this?

We suspect that $\mathbf{X} \simeq \mathrm{Sym}_n$, $n = 2k$. How can we **prove** this?

- Goldbach's Conjecture: $n = p_1 + p_2$, p_1, p_2 distinct primes (checked for $n < 10^{20}$).
- Find x, $|x| = p_1 p_2$.

$$\begin{aligned} y_1 &= x^{p_2} = (1, 2, \ldots, p_1) \\ y_2 &= x^{p_1} = (p_1 + 1, \ldots, n) \end{aligned}$$

$$
\begin{aligned}
y_1 &= x^{p_2} = (1, 2, \ldots, p_2) \\
y_2 &= x^{p_1} = (p_1 + 1, \ldots, n)
\end{aligned}
$$

- Find y with $|y| = 2q_1q_2$, q_1, q_2 odd primes. Compute a transposition

$$a = x^{q_1 q_2}.$$

- Check if a interchanges a point from y_1 with a point from y_2, that is,

$$ay_1 \neq y_1 a \text{ and } ay_2 \neq y_2 a.$$

If not, repeat the search for a transposition.

$$
\begin{aligned}
y_1 &= x^{p_2} = (1, 2, \ldots, p_1) \\
y_2 &= x^{p_1} = (p_1 + 1, \ldots, n) \\
a & \quad \text{transposition mixing } y_1 \text{ and } y_2
\end{aligned}
$$

- Compute $b = y_1 a y_2$. This is a cycle of length $p_1 + p_2 = n$.
- Compute $a_1 = a^{b^{p_1 - 1}}$. Without loss of generality:

  $b = (1, 2, \ldots, n),\ a = (p_1, p_1{+}1),\ a_1 = (1, 2)$

  b and a_1 generate Sym_n.

$$b = (1, 2, \ldots, n), \quad a_1 = (1, 2)$$

- Compute, by conjugating by b,

  $a_2 = (2, 3), \ldots, a_{n-1} = (n-1, n), a_n = (n, 1)$.

So we have the standard generators

$a_1 = (1, 2), \ldots, a_{n-1} = (n - 1, n), a_n = (n, 1)$.

So we have the standard generators

$$a_1 = (1,2), a_2 = (2,3), \ldots, a_n = (n,1).$$

- For an arbitrary $x \in \mathbf{X}$, compute

  $$x^{-1} a_1 x = (x(1), x(2)).$$

  By checking which of a_i this element commutes with, determine $\{x(1), x(2)\}$.

- Analogously, find

  $$\{x(2), x(3)\}, \ldots, \{x(n), x(1)\}.$$

- This will give

  $$x(1), \ldots x(n)$$

- So we can efficiently compute

$$x(1), \ldots x(n)$$

- which establishes establishes an isomorphism

$$\mathbf{X} \longrightarrow \mathrm{Sym}_n.$$

The effort so far was concentrated on

$\mathbf{X} \leqslant GL_N(\mathbb{F}_q)$ *simple* matrix group

$\mathbf{X} = \langle x_1, \ldots, x_k \rangle$

Distribution of orders of elements is specific for every finite simple group and allows to determine $\mathbf{X}$

Routines for GAP and MAGMA

People working in the area:

Babai, Beals, Bratus, Brookshank, Celler, Cooperman, Conder, Finkelstein, Holt, Leedham-Green, Linton, Lübeck, Kantor, Lubotzky, Murray, Neumann, Niemeyer, O'Brien, Pak, Palfy, Praeger, Rees, Seres, . . .

Comments

1. Substitute for prime factorisation

$$
\begin{aligned}
\mathbf{X} &\leqslant GL_N(\mathbb{F}_q) \\
|GL_N(\mathbb{F}_q)| &= q^{N(N-1)/2}(q-1)\cdots(q^N-1) \\
&= r_1^{a_1}\cdots r_m^{a_m} \quad \text{the finest one can get} \\
|x|_{\approx} &= \text{the least number} \\
& \quad l = r_1^{b_1}\cdots r_m^{b_m}, \\
& \quad x^l = 1.
\end{aligned}
$$

2. Nature of algorithms

Verification algorithms are **one-sided**:

Is $\mathbf{X} \simeq \mathbb{G}$ (the target group)?

If $\mathbf{X}$ contains an element of order not present in $\mathbb{G}$, then

$\mathbf{X} \not\simeq \mathbb{G}$ **absolutely**

3. The landscape is very different from the traditional group theory:

$$\mathbf{X} = SL_2(2^n)$$

- elements of even order are involutions
- involutions are conjugate

- The frequency involutions in $SL_2(2^n)$:

$$
\begin{aligned}
C &= C_{\mathbf{X}}\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right) \\
&= \left\{\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}\right\} \\
|C| &= 2^n \\
\#(\text{involutions }) &= \frac{|\mathbf{X}|}{2^n} \\
P(\text{even order}) &= \frac{|\mathbf{X}|}{2^n} \cdot \frac{1}{|\mathbf{X}|} = \frac{1}{2^n}
\end{aligned}
$$

If $n \gg 1$, $\mathbf{X}$, for all practical purposes, contains no involution and no elements of even order.

4. Statistical approach fails for

$$
\begin{aligned}
B_n(q) &= \Omega_{2n+1}(q), \\
C_n(q) &= PSp_{2n}(q),
\end{aligned}
$$

q odd:

they have virtually the same statistics of orders of elements.

Here,

$\Omega_{2n+1}(q)$ is the subgroup of index 2 in the orthogonal group $SO_{2n+1}(q)$,

$PSp_{2n}(q)$ is the projective symplectic group.

Why?

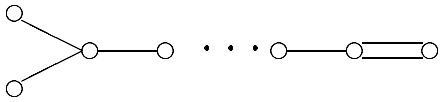
Let $G = G(\overline{\mathbb{F}}_q)$ be a simple algebraic group

- regular semisimple elements form an open subset of G
- statistics of orders of regular semisimple elements is determined by the **Dynkin diagram** of G, which are the same in the case of groups B_n and C_n, $n \geqslant 3$:

$BC_n,\ n \geq 2$

- But the conjugacy classes and the structure of centralisers of *involutions* (elements of order 2) are determined by the **extended Dynkin diagrams** which are different:

$\widetilde{B}_n, \quad n \geq 3$



$\widetilde{C}_n, \ \ n \geq 3$

- Luckily, the frequency of elements of even order in finite algebraic groups over fields of odd characteristic is at least $\frac{1}{4}$.

Idea: use involutions and their centralisers — if we can construct them.

Involutions in Black Box Groups

Miller-Rabin Primality Test:

An odd integer n is prime

$$\Updownarrow$$

$(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic of order $n-1$ and contains exactly one involution -1.

This is a verification problem: is $\mathbf{X} \simeq \mathbb{Z}_{\mathbf{n-1}}$?

Explanation:

If $n = n_1 \cdots n_k$ with $n_1, \ldots, n_k$ pairwise coprime, then

$$(\mathbb{Z}/n\mathbb{Z})^* = (\mathbb{Z}/n_1\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})^*$$

$$-1 \bmod n = -1 \bmod n_1 \times \cdots \times -1 \bmod n_k$$

(Chinese Remainder Theorem)

Miller-Rabin Primality Test:

- Factorise $n - 1 = 2^l \cdot m$, m odd.
- Take random $x \in (\mathbb{Z}/n\mathbb{Z})^*$.
- Compute

$$x \mapsto x^m, (x^m)^2, \ldots, (x^m)^{2^r} \neq 1, 1$$

$$\mathrm{i}(x) := (x^m)^{2^r} \quad \text{is an involution}$$

- If $\mathrm{i}(x) \neq -1$, **return** "n IS NOT PRIME,,
- If $\mathrm{i}(x) = -1$ for k random x, **return**

  n IS PRIME WITH PROBABILITY OF ERROR

$$\leqslant \frac{1}{4^k}$$

Centralisers of involutions

Assume that we know a computationally feasible n such that $x^n = 1$ for all $x \in \mathbf{X}$ (for example, when $\mathbf{X} \leqslant GL_d(\mathbb{F}_q)$ for reasonably sized d and q.) Factorise

$$n = 2^l \cdot m \quad , m \text{ odd.}$$

Construction of involutions:

$$\begin{aligned} x &\mapsto x^m, (x^m)^2, \ldots, (x^m)^{2^r} \neq 1, 1 \\ \mathrm{i}(x) &:= (x^m)^{2^r} \quad \text{is an involution} \end{aligned}$$

Centralisers of involutions: If t is an involution, we have a partial map

$$\begin{aligned} \zeta : \mathbf{X} &\longrightarrow C_{\mathbf{X}}(t) \\ x &\mapsto (t \cdot t^x)^{(m+1)/2} \cdot x^{-1}, \end{aligned}$$

when $|t \cdot t^x|$ odd

Why? If $|t \cdot t^x| = s$ is odd, the group $\langle t, t^x \rangle$ is the dihedral group of order $2s$, and, by Sylow's Theorem, $t^y = t^x$ for some $y \in \langle t \cdot t^x \rangle$. Then $yx^{-1} \in C_{\mathbf{X}}(t)$. It can be seen that

$$y = (t \cdot t^x)^{(m+1)/2}.$$

Construction of a black box for $C_{\mathbf{X}}(t)$ from a black box for $\mathbf{X}$

If $c \in C_{\mathbf{X}}(t)$,

$$\begin{aligned}
\zeta(cx) &= (tt^{cx})^{(m+1)/2} \cdot x^{-1} \cdot c^{-1} \\
&= (tt^{x})^{(m+1)/2} \cdot x^{-1} \cdot c^{-1} \\
&= \zeta(x) \cdot c^{-1}
\end{aligned}$$

If $x \in \mathbf{X}$ are uniformly distributed and independent,

then $\zeta_1(x)$ are uniformly distributed and independent in $C_{\mathbf{X}}(t)$

We have a remarkable analogy with the Brauer's programme of classification of finite simple groups in terms of the centralisers of involutions.

Altseimer-Borovik:

- Used centralisers of involution for telling $B_n(q)$ from $C_n(q)$.
- Systematic development of involution-based algorithms of black box recognition.