How badly does unique factorization fail, usually?

Paul Pollack (UGA)

joint with Enrique Treviño (Lake Forest)

Unique factorization?

Let R be an integral domain. A nonzero, nonunit element $\pi \in R$ is **irreducible** if π cannot be written as a product of two nonunits.

A domain R is a **unique factorization domain (UFD)** if every nonzero nonunit is a product of irreducibles and this expression is unique up to order and up to unit factors.

Unique factorization?

Let R be an integral domain. A nonzero, nonunit element $\pi \in R$ is **irreducible** if π cannot be written as a product of two nonunits.

A domain R is a **unique factorization domain (UFD)** if every nonzero nonunit is a product of irreducibles and this expression is unique up to order and up to unit factors.

More precisely, we require that if $\pi_1 \cdots \pi_k = \rho_1 \ldots \rho_\ell$, with all the π_i and ρ_j irreducible, then

(a) $k = \ell$,

(b) after rearranging, π_i is a R-unit multiple of ρ_i for all $i = 1, 2, \ldots, k$.

Hits and misses

In a first algebra course, one sees many examples of UFDs. These can lull one into a false sense of security!

The following near-canonical example of non-unique factorization is helpful to keep students on their toes: In the ring $\mathbb{Z}[\sqrt{-5}]$,

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

This is a *genuine* example of non-unique factorization: all of $2, 3$, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible in $\mathbb{Z}[\sqrt{-5}]$. Furthermore, the only units in $\mathbb{Z}[\sqrt{-5}]$ are ± 1, so there is no chance that the irreducibles on the left are unit multiples of those on the right.

Hits and misses

In a first algebra course, one sees many examples of UFDs. These can lull one into a false sense of security!

The following near-canonical example of non-unique factorization is helpful to keep students on their toes: In the ring $\mathbb{Z}[\sqrt{-5}]$,

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

This is a *genuine* example of non-unique factorization: all of $2, 3$, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible in $\mathbb{Z}[\sqrt{-5}]$. Furthermore, the only units in $\mathbb{Z}[\sqrt{-5}]$ are ± 1, so there is no chance that the irreducibles on the left are unit multiples of those on the right.

Thus, $\mathbb{Z}[\sqrt{-5}]$ is not a UFD!

Hits and misses

In a first algebra course, one sees many examples of UFDs. These can lull one into a false sense of security!

The following near-canonical example of non-unique factorization is helpful to keep students on their toes: In the ring $\mathbb{Z}[\sqrt{-5}]$,

$$2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

This is a *genuine* example of non-unique factorization: all of $2, 3$, $1 + \sqrt{-5}$ and $1 - \sqrt{-5}$ are irreducible in $\mathbb{Z}[\sqrt{-5}]$. Furthermore, the only units in $\mathbb{Z}[\sqrt{-5}]$ are ± 1, so there is no chance that the irreducibles on the left are unit multiples of those on the right.

Thus, $\mathbb{Z}[\sqrt{-5}]$ is not a UFD! However, it's not that far from being one. There are two irreducibles involved in both of our factorizations of 6. And this is the case for *all* counterexamples to uniqueness. We call $\mathbb{Z}[\sqrt{-5}]$ a **half-factorial domain** (HFD).

Stretching, the truth about unique factorization

Let R be a domain where every nonzero nonunit factors into irreducibles in at least one way, but only finitely many different ways (up to order and units). For each nonzero nonunit $\alpha \in R$, we define the **length spectrum** of α by

$$\mathcal{L}(\alpha) = \{\text{all lengths } k \text{ of irreducible factorizations } \alpha = \pi_1 \cdots \pi_k\}.$$

We define the **elasticity** of α by

$$\rho(\alpha) = \frac{\max \mathcal{L}(\alpha)}{\min \mathcal{L}(\alpha)}.$$

Finally, we define the elasticity $\rho(R)$ of R by

$$\rho(R) = \sup_{\alpha} \rho(\alpha).$$

So $\rho(R) = 1$ if and only if R is an HFD.

Fun. Theorem of Stretchiness

When K is a number field (finite extension of $\mathbb{Q}$), it is known that the elasticity of the ring of integers $\mathcal{O}_K$ is a finite number expresible in terms of a certain combinatorial constant associated to the class group.

Definition

Let G be a finite abelian group. The **Davenport constant** of G is the smallest positive integer $D = D(G)$ with the following property:

> *Every sequence $g_1, g_2, \ldots, g_D$ of elements of G contains a nonempty subsequence multiplying to the identity.*

Fun. Theorem of Stretchiness, ctd.

Fun Theorem of Stretchiness (Narkiewicz, Steffan, Valenza)

Whenever $\mathcal{O}_K$ is not a unique factorization domain,

$$\rho(\mathcal{O}_K) = \frac{1}{2}D,$$

where $D = \operatorname{Dav} \operatorname{Cl}(\mathcal{O}_K)$.

Fun. Theorem of Stretchiness, ctd.

Fun Theorem of Stretchiness (Narkiewicz, Steffan, Valenza)

Whenever $\mathcal{O}_K$ is not a unique factorization domain,

$$\rho(\mathcal{O}_K) = \frac{1}{2}D,$$

where $D = \operatorname{Dav} \operatorname{Cl}(\mathcal{O}_K)$.

This has some nice corollaries. For example, $\mathcal{O}_K$ is a HFD precisely when $\operatorname{Cl}(\mathcal{O}_K)$ has size 1 or 2 (Carlitz, 1960). In particular, $\mathbb{Z}[\sqrt{-5}]$ (of class number 2) is an HFD.

Fun. Theorem of Stretchiness, ctd.

Fun Theorem of Stretchiness (Narkiewicz, Steffan, Valenza)

Whenever $\mathcal{O}_K$ is not a unique factorization domain,

$$\rho(\mathcal{O}_K) = \frac{1}{2}D,$$

where $D = \mathrm{Dav}\,\mathrm{Cl}(\mathcal{O}_K)$.

This has some nice corollaries. For example, $\mathcal{O}_K$ is a HFD precisely when $\mathrm{Cl}(\mathcal{O}_K)$ has size 1 or 2 (Carlitz, 1960). In particular, $\mathbb{Z}[\sqrt{-5}]$ (of class number 2) is an HFD. To take another example, let $K = \mathbb{Q}(\sqrt{-23})$, for which $\mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{-23}}{2}]$. Then $\#\mathrm{Cl}(\mathcal{O}_K) = 3$, and so $\rho(\mathcal{O}_K) = \frac{3}{2}$. That $\rho(\mathcal{O}_K) \geq \frac{3}{2}$ is immediate from the example

$$(2 + \sqrt{-23})(2 - \sqrt{-23}) = 3 \cdot 3 \cdot 3.$$

The anatomy of an irreducible

The Davenport constant enters the picture through an observation made by Harold Davenport in 1966, at the Midwestern Conference on Group Theory and Number Theory at Ohio State.

The anatomy of an irreducible, ctd.

Let K be a number field, and let π be an irreducible element of $\mathcal{O}_K$. Factor $\pi\mathcal{O}_K$ (uniquely) as a product of prime ideals,

$$\pi\mathcal{O}_K = P_1 \cdots P_g.$$

No nonempty, proper subproduct of $P_1, \dots, P_g$ can multiply to a principal ideal: Suppose $P_1 \cdots P_r = \alpha\mathcal{O}_K$ for some nonunit $\alpha \in \mathcal{O}_K$, where $r < g$. Then $P_{r+1} \cdots P_g = \beta\mathcal{O}_K$ for some nonunit $\beta \in \mathcal{O}_K$. Hence,

$$\pi\mathcal{O}_K = \alpha\beta\mathcal{O}_K$$

and after adjusting α by a unit, $\pi = \alpha\beta$, a nontrivial factorization.

The anatomy of an irreducible, ctd.

Ergo: If g is the number of prime ideal factors of π, counted with multiplicity, then

$$g \leq \operatorname{Dav} \operatorname{Cl}(\mathcal{O}_K).$$

Conversely, we construct an irreducible composed of $D := \operatorname{Dav} \operatorname{Cl}(\mathcal{O}_K)$ prime ideals. Every ideal class is represented by a prime ideal (Landau). Thus, we can pick $D-1$ prime ideals $P_1, \dots, P_{D-1}$ where no nonempty subproduct multiples to the identity. Let P_D be a prime ideal in the class inverse to $P_1 \cdots P_{D-1}$. Then

$$P_1 \cdots P_D = \pi \mathcal{O}_K$$

for some nonzero, nonunit $\pi \in \mathcal{O}_K$. Exercise: π is irreducible.

Davenport's observation: $\operatorname{Dav} \operatorname{Cl}(\mathcal{O}_K) = \max \#$ prime factors of π.

Half the Fun (Theorem)

Suppose π is an irreducible composed of D distinct prime ideals, $\pi\mathcal{O}_K = P_1 \cdots P_D$. Choose prime ideals $P_1', \ldots, P_D'$ in classes inverse to $P_1, \ldots, P_D$, respectively.

Then $P_1' \cdots P_D' = \pi'\mathcal{O}_K$, where π' is also irreducible. Furthermore, $P_i P_i' = \gamma_i \mathcal{O}_K$, for a nonunit $\gamma_i \in \mathcal{O}_K$. The equation $P_1 \cdots P_D P_1' \cdots P_D' = (P_1 P_1') \cdots (P_D P_D')$ gives

$$(\pi\pi')\mathcal{O}_K = (\gamma_1 \cdots \gamma_D)\mathcal{O}_K.$$

Adjusting γ_1 by a unit, we get an equality of elements. Since $\gamma_1 \cdots \gamma_D$ can be factored into at least D irreducibles,

$$\rho(\mathcal{O}_K) \geq \rho(\pi\pi') \geq \frac{D}{2}.$$

This is half the Fundamental Theorem.

Stirring the sup

One can view $\rho(\alpha)$ as an arithmetic function on (nonzero, nonunits of) $\mathcal{O}_K$, taking values in $\mathbb{Q}^{\geq 1}$.

The Fundamental Theorem of Stretchiness is a theorem about large values of this function:

$$\sup_{\alpha} \rho(\alpha) = \frac{1}{2}D.$$

(The construction we gave before shows that the supremum is actually a maximum.)

Taking the supremum is a coping mechanism of sorts. Rather than confronting the entire sea of numbers $\{\rho(\alpha)\}$, we get to focus on a single "worst-case" quantity.

Stirring the sup

One can view $\rho(\alpha)$ as an arithmetic function on (nonzero, nonunits of) $\mathcal{O}_K$, taking values in $\mathbb{Q}^{\geq 1}$.

The Fundamental Theorem of Stretchiness is a theorem about large values of this function:

$$\sup_{\alpha} \rho(\alpha) = \frac{1}{2}D.$$

(The construction we gave before shows that the supremum is actually a maximum.)

Taking the supremum is a coping mechanism of sorts. Rather than confronting the entire sea of numbers $\{\rho(\alpha)\}$, we get to focus on a single "worst-case" quantity. But we are number theorists! We want the sea of numbers! We are fearless!

Stirring the sup, ctd.

How are the numbers $\rho(\alpha)$ distributed? We know the max/supremum.

We also know the min: By definition, $\rho(\alpha) \geq 1$ always. And $\rho(\alpha) = 1$ infinitely often, e.g., at each irreducible α of $\mathcal{O}_K$.

What else might we want to know? Is every value in $[1, \rho(\mathcal{O}_K)]$ attained by some $\rho(\alpha)$?

Stirring the sup, ctd.

How are the numbers $\rho(\alpha)$ distributed? We know the max/supremum.

We also know the min: By definition, $\rho(\alpha) \geq 1$ always. And $\rho(\alpha) = 1$ infinitely often, e.g., at each irreducible α of $\mathcal{O}_K$.

What else might we want to know? Is every value in $[1, \rho(\mathcal{O}_K)]$ attained by some $\rho(\alpha)$? Yes! Baginski, Chapman, Crutchfield, Grace Kennedy and Wright, 2006.

What about more "statistical" properties? Is there a mean value?

Stirring the sup, ctd.

How are the numbers $\rho(\alpha)$ distributed? We know the max/supremum.

We also know the min: By definition, $\rho(\alpha) \geq 1$ always. And $\rho(\alpha) = 1$ infinitely often, e.g., at each irreducible α of $\mathcal{O}_K$.

What else might we want to know? Is every value in $[1, \rho(\mathcal{O}_K)]$ attained by some $\rho(\alpha)$? Yes! Baginski, Chapman, Crutchfield, Grace Kennedy and Wright, 2006.

What about more "statistical" properties? Is there a mean value?

Is there a "typical" value ρ_{typ}? "Typical" means 100% of the numbers $\rho(\alpha)$ lie within ϵ of ρ_{typ}.

Stirring the sup, ctd.

How are the numbers $\rho(\alpha)$ distributed? We know the max/supremum.

We also know the min: By definition, $\rho(\alpha) \geq 1$ always. And $\rho(\alpha) = 1$ infinitely often, e.g., at each irreducible α of $\mathcal{O}_K$.

What else might we want to know? Is every value in $[1, \rho(\mathcal{O}_K)]$ attained by some $\rho(\alpha)$? Yes! Baginski, Chapman, Crutchfield, Grace Kennedy and Wright, 2006.

What about more "statistical" properties? Is there a mean value?

Is there a "typical" value ρ_{typ}? "Typical" means 100% of the numbers $\rho(\alpha)$ lie within ϵ of ρ_{typ}.

Since ρ is nonnegative and bounded, if ρ_{typ} exists, it is also the average value.

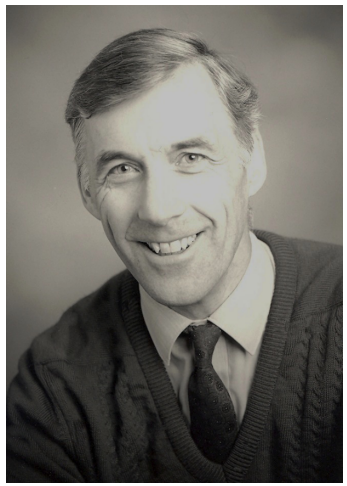
Stirring the sup, ctd.

Theorem (Narkiewicz and Sliwa, 1977)

Yes, $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ exists. Furthermore, $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ depends entirely on the (isomorphism type of) $\mathrm{Cl}(\mathcal{O}_K)$.

Similar results were obtained, independently, by Allen and Pleasants in 1980.

W. Narkiewicz



P. A. B. Pleasants

Playing solitaire, in a group!

To determine $\rho_{\mathrm{typ}}(\mathcal{O}_K)$ requires determining the optimal way to play a certain game on the class group of $\mathcal{O}_K$. Enrique and I call this game "group solitaire."

Let G be a finite abelian group. (For the application to $\rho_{\mathrm{typ}}(\mathcal{O}_K)$, we will take G as the class group of $\mathcal{O}_K$, but the setup of the game doesn't require that.) Observe that

$$\left(\prod_{g\in G} g\right)^2 = \prod_{g\in G} g \prod_{g\in G} g^{-1} = e.$$

This will be useful momentarily.

Rules are rules

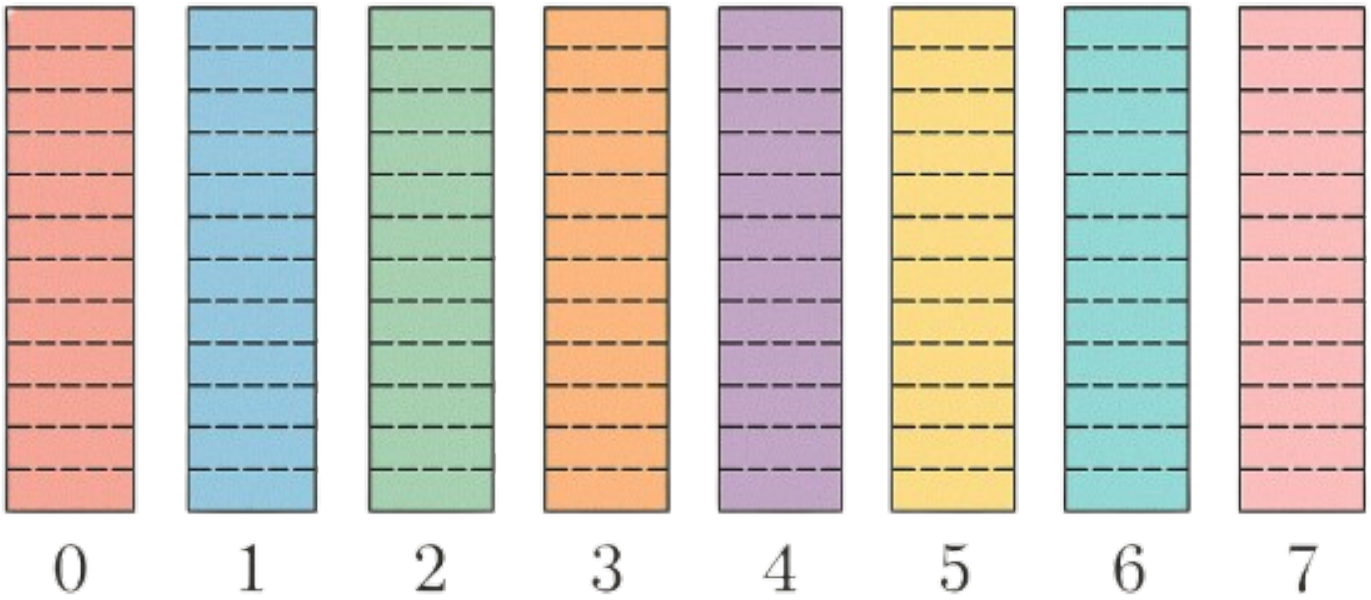
Let's suppose G has n elements. Take n stacks of poker chips, each of the same height X (height = number of chips). We view each chip as representing an element of the group, with the different stacks corresponding to the different elements.

Let's assume X is even. Then the product of all the poker chips is the identity e. (It is $(\prod_{g \in G} g)^X$, and $(\prod_{g \in G} g)^2$ is the identity.)

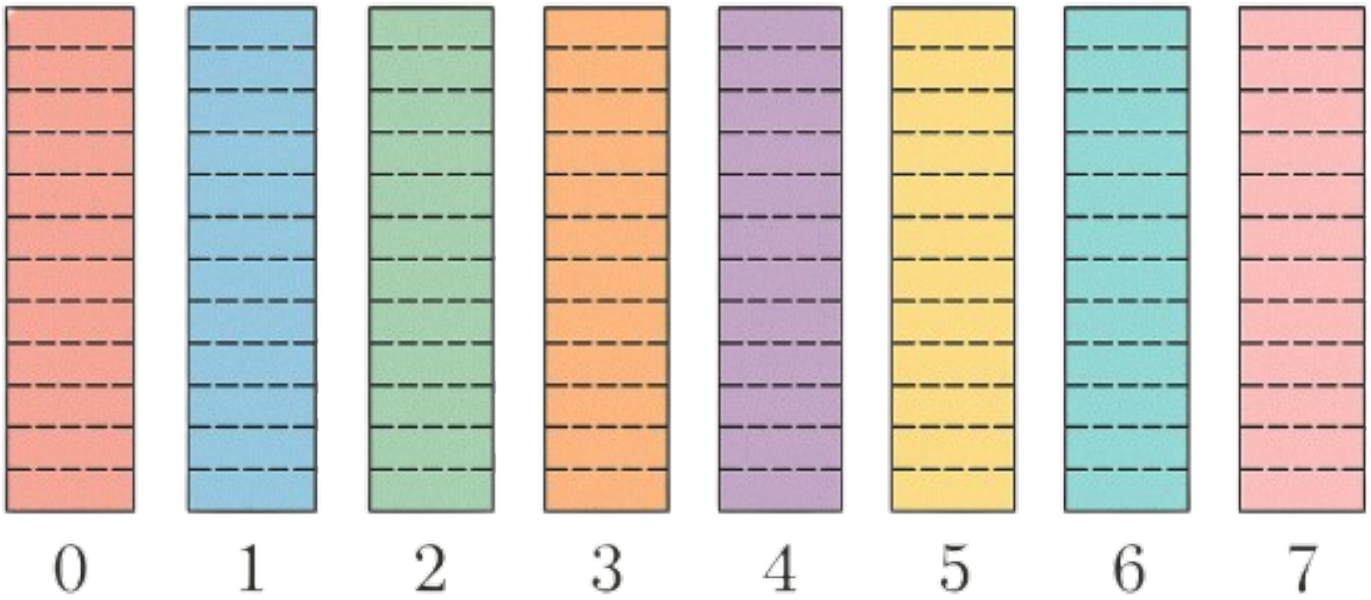
An allowable move consists of removing a collection of chips which multiply to the identity, where no proper subcollection also multiplies to the identity.

Objective: Clear all the stacks in as few moves as possible!

An example: $\mathbb{Z}/8\mathbb{Z}$ solitaire



An example: $\mathbb{Z}/8\mathbb{Z}$ solitaire



Exercise

Take $G = \mathbb{Z}/8\mathbb{Z}$. Show that if each stack has height X, then any way of clearing the stacks requires at least $\sim \frac{5}{2}X$ moves.

Am I just a game to you?

Let $\Sigma(G;X)$ be the minimal number of moves needed to clear the board in G-solitaire, starting with stacks of equal (even) height X. The ratio $\frac{\Sigma(G;X)}{X}$ tends to a limit; we define

$$\operatorname{Clr} G = \lim_{X\to\infty} \frac{1}{X}\Sigma(G;X)$$

and call it the **clearing constant** associated to G. For example, $\operatorname{Clr} \mathbb{Z}/8\mathbb{Z} = \frac{5}{2}$.

The existence of $\operatorname{Clr} G$ follows from linear programming. Those same methods show that $\operatorname{Clr} G \in \mathbb{Q}^{\geq 1}$.

Theorem

For every number field K,

$$\rho_{\text{typ}}(\mathcal{O}_K) = \frac{1+\#\operatorname{Cl}(\mathcal{O}_K)}{2\operatorname{Clr}\operatorname{Cl}(\mathcal{O}_K)}.$$

Am I just a game to you?

I just give some ideas/impressions of the proof.

A legal move corresponds to constructing an irreducible, where the prime ideals are from the classes of the removed elements.

We make all stacks the same size to capture typical behavior.

Generalization of Hardy–Ramanujan: For almost all (asymptotically 100%) of $\alpha \in \mathcal{O}_K$, the number of prime ideals from a given class involved in the prime factorization of α is

$$\approx \frac{1}{\#\mathrm{Cl}(\mathcal{O}_K)} \log\log|\mathrm{Norm}(\alpha)|.$$

In particular, the prime ideal divisors of α are approximately uniformly distributed among the ideal classes.

What have you done ?!?!

This is all well and good (I hope you agree)…but what did Enrique and I do?

We determined $\mathrm{Clr}\, G$ for two families of abelian groups G.

Allen & Pleasants (1980) had determined $\mathrm{Clr}\, G$ for all groups $G = (\mathbb{Z}/p^r\mathbb{Z})^s$ (homocyclic p-groups). In particular, their results handle any elementary p-group.

These same authors also determined, by ad hoc methods, that

$$\mathrm{Clr}(\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}) = \frac{8}{3} \quad \text{and} \quad \mathrm{Clr}\, \mathbb{Z}/6\mathbb{Z} = \frac{13}{6}.$$

We locate these last two results as special cases of more general formulas.

What have you done ?!?! ctd.

Theorem (E.T. and P.)

Let p and q be distinct primes.

(a) $\mathrm{Clr}(\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z}) = 1 + \frac{2p^2-p-1}{2p-1}$,

(b) $\mathrm{Clr}(\mathbb{Z}/pq\mathbb{Z}) = 3 - \frac{1}{p} - \frac{1}{q}$.

Application. Suppose $\mathrm{Cl}(\mathcal{O}_K) \cong \mathbb{Z}/10\mathbb{Z}$. Using (b) with $p = 2, q = 5$ ...

Scenario	**Elasticity**
Ideal Case (UFD or HFD)	$\rho = 1$
Typical Case (ρ_{typ})	$\rho = \frac{55}{23} \approx \mathbf{2.39}$
Worst Case (ρ)	$\rho = 5$

Charging ahead

In the interests of time, I describe only some ideas associated with (a). (The proof of (b) is a different, explicit construction.)

The proof revolves the notion of **charge**.

Let G be a finite abelian group. We work with vectors $\mathbf{v} \in \mathbb{R}^G$, meaning real vectors indexed by elements of G. The gth component of $\mathbf{v}$ will be denoted $\mathbf{v}[g]$. Every multiset of elements of G naturally corresponds to a vector of this kind — the gth component counting the number of occurrences of g. In particular, we can associate a vector to each move, and to the initial configuration of chips.

By the **charge** of $\mathbf{v}$, we mean the quantity

$$\sum_{g \in G} \frac{\mathbf{v}[g]}{\mathrm{ord}(g)},$$

where $\mathrm{ord}(g)$ is the order in G.

Charging ahead

The maximum charge of a legal move in G-solitaire has been extensively studied, beginning with work of Ulrich Krause in 1984. He called this **cross number** of G. We will denote it by $k(G)$.

In any game of group solitaire, the X chips corresponding to the identity have to be removed one-at-a-time. Let $k_{\text{initial}}(X, G) = X \sum_{g \in G, g \neq 0} 1/\operatorname{ord}(g)$ denote the charge of the non-identity elements in the initial configuration.

Each move decreases the charge of the configuration by at most $k(G)$, and so

$$\Sigma(G; X) \geq X + \frac{k_{\text{initial}}(X, G)}{k(G)}.$$

Dividing by X and sending X to infinity ...

Charging ahead, ctd.

Each move decreases the charge of the configuration by at most $k(G)$, and so

$$\Sigma(G;X) \geq X + \frac{k_{\text{initial}}(X,G)}{k(G)}.$$

Dividing by X and sending X to infinity,

$$\mathrm{Clr}(G) \geq 1 + \frac{1}{k(G)} \sum_{g \in G, g \neq 0} \frac{1}{\mathrm{ord}(g)}.$$

We call finite abelian groups G for which equality holds here **economical**. One can reinterpret Allen and Pleasants' calculation of $\mathrm{Clr}\,(\mathbb{Z}/p^r\mathbb{Z})^s$ as proving that homocyclic p-groups are economical.

Charging ahead, ctd.

Our Theorem(a) is no more and no less than the claim that the groups $G = \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z}$ are all economical.

One can interpret economical-ity as the feasibility of linear programming problem in $\mathbb{R}^G$. This problem becomes intimidating as p grows, because $\#G$ grows as well.

But in fact, one can reformulate the linear programming problem replacing G by $G/\sim$, where $\sim$ is automorphism-equivalence. And $G/\sim$ is always a 4-element space, regardless of p. In this small-dimensional space, it is possible to "eyeball" a solution.

Charging ahead, ctd.

Our Theorem(a) is no more and no less than the claim that the groups $G = \mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^2\mathbb{Z}$ are all economical.

One can interpret economical-ity as the feasibility of linear programming problem in $\mathbb{R}^G$. This problem becomes intimidating as p grows, because $\#G$ grows as well.

But in fact, one can reformulate the linear programming problem replacing G by $G/\sim$, where $\sim$ is automorphism-equivalence. And $G/\sim$ is always a 4-element space, regardless of p. In this small-dimensional space, it is possible to "eyeball" a solution.

Challenge: Determine the clearing constant for $\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/p^3\mathbb{Z}$!

THANK YOU FOR LISTENING

and

HAPPY BIRTHDAY, KRISHNA!