Invariants, Monovariants, and Extrema

Matthew Rognlie

Department of Mathematics
Duke University

September 9, 2009

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?
- Find all reachable states.

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?
- Find all reachable states.
- Find all reachable *end* states (states where no more transformations are possible).

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?
- Find all reachable states.
- Find all reachable *end* states (states where no more transformations are possible).
- Will the process inevitably converge to some end state?

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?
- Find all reachable states.
- Find all reachable *end* states (states where no more transformations are possible).
- Will the process inevitably converge to some end state?
- Is there periodicity?

Why Invariants?

Common Questions

Say that we investigating some process, which has a starting state S and a sequence of possible subsequent transformations. We may want to know:

- Is it possible to reach a given state?
- Find all reachable states.
- Find all reachable *end* states (states where no more transformations are possible).
- Will the process inevitably converge to some end state?
- Is there periodicity?

If we want to answer any of these questions, it's likely that we will make use of invariants (or, as we'll discover later, monovariants)!

The Basic Idea

When we are examining a process that involves repeated transformations, we should explore *what does not change*. This is often critical to understanding the properties of the process. Some simple examples:

The Basic Idea

When we are examining a process that involves repeated transformations, we should explore *what does not change*. This is often critical to understanding the properties of the process. Some simple examples:

- In classical physics, the total energy of a closed system is invariant.

The Basic Idea

When we are examining a process that involves repeated transformations, we should explore *what does not change*. This is often critical to understanding the properties of the process. Some simple examples:

- In classical physics, the total energy of a closed system is invariant.
- The parity of a number (whether it is odd or even) is invariant to multiplication by an odd number.

The Basic Idea

When we are examining a process that involves repeated transformations, we should explore *what does not change*. This is often critical to understanding the properties of the process. Some simple examples:

- In classical physics, the total energy of a closed system is invariant.
- The parity of a number (whether it is odd or even) is invariant to multiplication by an odd number.
- Repeatedly sum the digits of a number to get a new number. For instance, we may find:

$$27708 \to 24 \to 6 \to 6 \ldots$$

The value of our number mod 9 is *invariant*.

A Slightly More Involved Example

Suppose n is odd, and the numbers $1, 2, \ldots, 2n$ are written on the blackboard. I repeatedly pick two arbitrary numbers on the blackboard a and b, erase them, and write $|a - b|$ instead. This continues until only one integer is left. Is the number at the end odd or even?

A Slightly More Involved Example

Suppose n is odd, and the numbers $1, 2, \ldots, 2n$ are written on the blackboard. I repeatedly pick two arbitrary numbers on the blackboard a and b, erase them, and write $|a - b|$ instead. This continues until only one integer is left. Is the number at the end odd or even?

Once we realize the sum of numbers on the blackboard mod 2 is invariant, this is **easy**!

A Slightly More Involved Example

Suppose n is odd, and the numbers $1, 2, \ldots, 2n$ are written on the blackboard. I repeatedly pick two arbitrary numbers on the blackboard a and b, erase them, and write $|a - b|$ instead. This continues until only one integer is left. Is the number at the end odd or even?

Once we realize the sum of numbers on the blackboard mod 2 is invariant, this is **easy**!

Indeed, for any a and b, $a + b \equiv a - b \pmod 2$. Thus when I replace two numbers by their difference, the parity of the sum *never changes*. All we need to do now is calculate whether the sum $1 + 2 + \ldots + 2n$ is odd or even:

$$\sum_{k=0}^{2n} k = \frac{2n(2n+1)}{2} = n(2n+1)$$

Since n is odd, the rightmost expression is odd as well, and we can conclude that regardless of the order in which I pick and replace the numbers, the final one will be odd. (Like many examples in this presentation, this is taken from *Problem Solving Strategies* by Arthur Engel, our recommended text for the course.)

Another Contrived Example

Problem

A dragon has 100 heads. A knight can cut off 15, 17, 20, or 5 heads, respectively, with one blow of his sword. In each of these cases, 24, 2, 14, or 17 heads grow on its shoulders. If all heads are blown off, the dragon dies. Can the dragon ever die?

Another Contrived Example

Problem

A dragon has 100 heads. A knight can cut off 15, 17, 20, or 5 heads, respectively, with one blow of his sword. In each of these cases, 24, 2, 14, or 17 heads grow on its shoulders. If all heads are blown off, the dragon dies. Can the dragon ever die?

Proof.

At first glance, this problem is convoluted and intractable. Once we hit upon the idea of using invariants, however, it becomes trivial. We note:

$$(24 - 15) \equiv (2 - 17) \equiv (14 - 20) \equiv (17 - 5) \equiv 0 \pmod 3$$

Another Contrived Example

Problem

A dragon has 100 heads. A knight can cut off 15, 17, 20, or 5 heads, respectively, with one blow of his sword. In each of these cases, 24, 2, 14, or 17 heads grow on its shoulders. If all heads are blown off, the dragon dies. Can the dragon ever die?

Proof.

At first glance, this problem is convoluted and intractable. Once we hit upon the idea of using invariants, however, it becomes trivial. We note:

$$(24 - 15) \equiv (2 - 17) \equiv (14 - 20) \equiv (17 - 5) \equiv 0 \pmod 3$$

The knight's gallantry can never change the number of heads of the dragon mod 3. Since we start at $100 \equiv 1 \pmod 3$, we can never get to 0. The dragon lives! $\square$

One Final Example

Problem

The integers $1, \ldots, n$ *are arranged in any order. In one step, you may switch any two neighboring integers. Prove that you can never reach the initial order after an odd number of steps.*

One Final Example

Problem

The integers $1, \ldots, n$ *are arranged in any order. In one step, you may switch any two neighboring integers. Prove that you can never reach the initial order after an odd number of steps.*

Proof.

Let the sequence of integers be $a_1, \ldots, a_n$. For any indices $i < j$, let b_{ij} equal 0 if $a_i < a_j$ and 1 if $a_j < a_i$. In other words, $b_{ij} = 1$ if a_i and a_j are *out of order*. We assert that switching two neighboring integers will change the sum $\sum_{i<j} b_{ij}$. Indeed, if we switch integers a_k and a_{k+1}, the only value of b_{ij} that changes is $b_{n,n+1}$, which either goes from 0 to 1 or 1 to 0. In any case, its parity changes, and thus the parity of the sum $\sum_{i<j} b_{ij}$ changes. After an odd number of steps, the parity of this sum cannot be the same as the initial parity, and we cannot be at the initial order. □

Some Observations

What We Saw

Some Observations

What We Saw

- These problems were pretty easy! Not all problems using invariants are so simple, but once you realize the right approach you're often close to done.

Some Observations

What We Saw

- These problems were pretty easy! Not all problems using invariants are so simple, but once you realize the right approach you're often close to done.
- Arguments using parity and modular arithmetic in general are *very* commonly associated with invariants.

Some Observations

What We Saw

- These problems were pretty easy! Not all problems using invariants are so simple, but once you realize the right approach you're often close to done.
- Arguments using parity and modular arithmetic in general are *very* commonly associated with invariants.
- Taking sums and exploiting symmetry is a good way to design an invariant.

Some Observations

What We Saw

- These problems were pretty easy! Not all problems using invariants are so simple, but once you realize the right approach you're often close to done.
- Arguments using parity and modular arithmetic in general are *very* commonly associated with invariants.
- Taking sums and exploiting symmetry is a good way to design an invariant.
- We're not limited to looking for values that literally *never change*. Instead, we often want values that change *in a predictable way*. In the last example, for instance, we found a sum that alternated between odd and even parity with each step.

What is a Monovariant?

Many problems require a generalization of the idea of an "invariant." Even if we cannot identify some function of the state of a process that never changes, we may be able to identify a function that always changes *in the same direction*. In fact, this information can be invaluable. Consider the following fact:

What is a Monovariant?

Many problems require a generalization of the idea of an "invariant." Even if we cannot identify some function of the state of a process that never changes, we may be able to identify a function that always changes *in the same direction*. In fact, this information can be invaluable. Consider the following fact:

If there is some positive integral function that decreases at each step of a process, the process must eventually terminate.

What is a Monovariant?

Many problems require a generalization of the idea of an "invariant." Even if we cannot identify some function of the state of a process that never changes, we may be able to identify a function that always changes *in the same direction*. In fact, this information can be invaluable. Consider the following fact:

If there is some positive integral function that decreases at each step of a process, the process must eventually terminate.

This is trivial: if a positive integral function starts at n and decreases with each step, the process certainly cannot continue for more than $n-1$ steps. Yet despite its apparent obviousness, this principle is perhaps the most powerful way for us to draw conclusions about how a process behaves.

First Example

Problem

In the Senate of Kazakhstan, each member has at most three enemies. A member cannot be his own enemy, and enmity is mutual. Prove that the Senate can be divided into two factions such that each Senator has at most one enemy within his faction.

First Example

Problem

In the Senate of Kazakhstan, each member has at most three enemies. A member cannot be his own enemy, and enmity is mutual. Prove that the Senate can be divided into two factions such that each Senator has at most one enemy within his faction.

Proof.

First, we separate the members arbitrarily into two factions. Let H be the sum of all the enemies each member has in his own faction. Suppose one member (let's call him Bob) has at least two enemies in his own faction. Then if Bob switches factions, H will decrease. Let this process continue for all members in the same situation as Bob. Since H is a positive integral function that decreases at each step of the process, the process must terminate. At this point, *no* Senator can have more than one enemy in his own faction, because otherwise the process (by definition) would not have terminated. Thus we have found the desired division of Senators. □

Second Example: from the 2008 Putnam contest!

Problem

Start with a finite sequence $a_1, a_2, \ldots, a_n$ of positive integers. If possible, choose two indices $j < k$ such that a_j does not divide a_k, and replace a_j and a_k by $gcd(a_j, a_k)$ and $lcm(a_j, a_k)$, respectively. Prove that if this process is repeated, it must eventually stop.

Second Example: from the 2008 Putnam contest!

Problem

Start with a finite sequence $a_1, a_2, \ldots, a_n$ of positive integers. If possible, choose two indices $j < k$ such that a_j does not divide a_k, and replace a_j and a_k by $\gcd(a_j, a_k)$ and $\operatorname{lcm}(a_j, a_k)$, respectively. Prove that if this process is repeated, it must eventually stop.

Proof.

First, note that the product $a_1 a_2 \ldots a_n$ is invariant, because $a_j a_k = \gcd(a_j, a_k) \operatorname{lcm}(a_j, a_k)$. Now, each element in the sequence is bounded from above by this product. The last element can never decrease, because it is only replaced by its least common multiple with another integer. Thus eventually it reaches its maximum value and becomes fixed. After this happens, the second-to-last element will never decrease; it also eventually reaches its maximum and becomes constant, and so on, until the entire sequence is fixed. □

Multiple Ways to Solve a Problem

On the actual contest, I solved the previous example using a different monovariant: the *sum* of all elements in the sequence. A little algebra proves:

Multiple Ways to Solve a Problem

On the actual contest, I solved the previous example using a different monovariant: the *sum* of all elements in the sequence. A little algebra proves:

$$a_j \neq a_k \implies \gcd(a_j, a_k) + \operatorname{lcm}(a_j, a_k) < a_j + a_k$$

Multiple Ways to Solve a Problem

On the actual contest, I solved the previous example using a different monovariant: the *sum* of all elements in the sequence. A little algebra proves:

$$a_j \neq a_k \implies \gcd(a_j, a_k) + \operatorname{lcm}(a_j, a_k) < a_j + a_k$$

This is often a nice feature of problems that require monovariants: there isn't a single "correct" monovariant that you need to find before you make any progress. Instead, there may be many candidates, differing in elegance and ease but not in substance. Just remember that most good monovariants will exploit symmetry, and work out some examples by hand to see if any pattern jumps out.

A Scary Monovariant

The following monovariant appears in the solution to a problem from the 1986 International Mathematical Olympiad:

A Scary Monovariant

The following monovariant appears in the solution to a problem from the 1986 International Mathematical Olympiad:

$$\begin{aligned} S(v,w,x,y,z) = &|v| + |w| + |x| + |y| + |z| + |v+w| + |w+x| + |x+y| \\ &+ |y+z| + |z+v| + |v+w+x| + |w+x+y| \\ &+ |x+y+z| + |y+z+v| + |z+v+w| + |v+w+x+y| \\ &+ |w+x+y+z| + |x+y+z+v| + |y+z+v+w| \\ &+ |z+v+w+x| \end{aligned}$$

A Scary Monovariant

The following monovariant appears in the solution to a problem from the 1986 International Mathematical Olympiad:

$$\begin{aligned} S(v,w,x,y,z) = &|v| + |w| + |x| + |y| + |z| + |v+w| + |w+x| + |x+y| \\ &+ |y+z| + |z+v| + |v+w+x| + |w+x+y| \\ &+ |x+y+z| + |y+z+v| + |z+v+w| + |v+w+x+y| \\ &+ |w+x+y+z| + |x+y+z+v| + |y+z+v+w| \\ &+ |z+v+w+x| \end{aligned}$$

Note that this isn't quite as scary as it first appears: if v, w, x, y, z are points on a pentagon, then this is just the sum of the absolute values of sums of individual vertices, pairs, triples, and foursomes. (The original problem involved a procedure applied to integers at the vertices of a pentagon.) In fact, this is a good illustration of how sophisticated symmetries that reflect the underlying structure of a problem can produce impressive results. Some monovariants are simple and elegant, but don't be afraid to try something a little more unwieldy. Experiment!

The Extremal Principle

The extremal principle rests on three important facts (from Engel):

The Extremal Principle

The extremal principle rests on three important facts (from Engel):

1. Every finite nonempty set A of nonnegative integers or real numbers has a minimal element $\min A$ and a maximal element $\max A$, which need not be unique.

The Extremal Principle

The extremal principle rests on three important facts (from Engel):

1. Every finite nonempty set A of nonnegative integers or real numbers has a minimal element $\min A$ and a maximal element $\max A$, which need not be unique.
2. Every nonempty subset of positive integers has a smallest element.

The Extremal Principle

The extremal principle rests on three important facts (from Engel):

1. Every finite nonempty set A of nonnegative integers or real numbers has a minimal element $\min A$ and a maximal element $\max A$, which need not be unique.
2. Every nonempty subset of positive integers has a smallest element.
3. An infinite set A of real numbers need not have a minimal or maximal element. If A is bounded above, however, then it has a smallest upper bound $\sup A$. If A is bounded below, it has a largest lower bound $\inf A$. If $\sup A \in A$, then $\sup A = \max A$, and if $\inf A \in A$, then $\inf A = \min A$.

The Extremal Principle

The extremal principle rests on three important facts (from Engel):

1. Every finite nonempty set A of nonnegative integers or real numbers has a minimal element $\min A$ and a maximal element $\max A$, which need not be unique.
2. Every nonempty subset of positive integers has a smallest element.
3. An infinite set A of real numbers need not have a minimal or maximal element. If A is bounded above, however, then it has a smallest upper bound $\sup A$. If A is bounded below, it has a largest lower bound $\inf A$. If $\sup A \in A$, then $\sup A = \max A$, and if $\inf A \in A$, then $\inf A = \min A$.

Although (indeed, *because*) the extremal principle is useful in a wide range of situations, it is not nearly as easy to recognize as invariance.

A Simple Example: the Senate of Kazakhstan, Again...

Problem

In the Senate of Kazakhstan, each member has at most three enemies. A member cannot be his own enemy, and enmity is mutual. Prove that the Senate can be divided into two factions such that each Senator has at most one enemy within his faction.

A Simple Example: the Senate of Kazakhstan, Again...

Problem

In the Senate of Kazakhstan, each member has at most three enemies. A member cannot be his own enemy, and enmity is mutual. Prove that the Senate can be divided into two factions such that each Senator has at most one enemy within his faction.

Proof.

Consider all partitions of the Senate into factions, and count the total number of enemies E that Senators have in their factions. Now pick a partition of the Senate with minimal E. This partition has the desired property: if some Senator had more than one enemy within his faction, we could move him to the other faction and decrease E, which would be a contradiction. $\square$

A Nice Example

Problem

Given n points in the plane, no three colinear, prove that there is a polygon with all n points as its vertices.

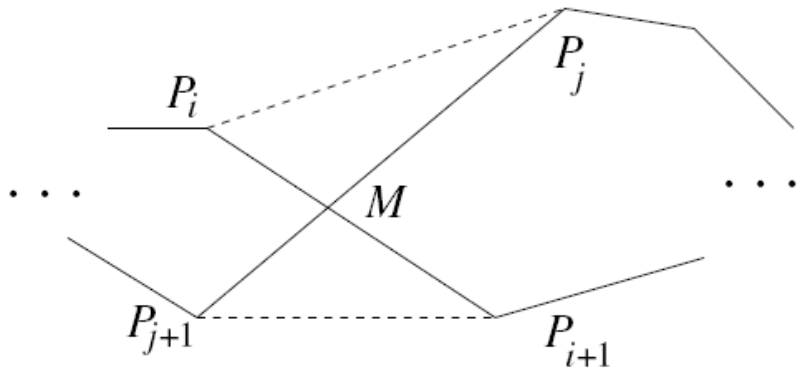
A Nice Example

Problem

Given n points in the plane, no three colinear, prove that there is a polygon with all n points as its vertices.

Proof.

Consider all closed segment paths visiting all n points. Choose a path $P_1P_2 \ldots P_nP_1$ with minimal total length. We assert that this path has no self-intersections, and is thus a polygon. Suppose the contrary: that this minimal path does have at least one self-intersection. Specifically, suppose that the segments P_iP_{i+1} and P_jP_{j+1} in this path intersect at a point M, as in the diagram below. Then the triangle inequality implies that we can reduce total path length by replacing these segments with P_iP_j and $P_{i+1}P_{j+1}$, contradicting the minimality assumption.



A Harder Example

Problem

Let X be a subset of the positive integers with the property that the sum of any two (not necessarily distinct) elements in X is again in X. Suppose that $\{a_1, a_2, \ldots, a_n\}$ is the set of all positive integers not in X. Prove that $a_1 + a_2 + \ldots + a_n \leq n^2$. (from Putnam and Beyond, by Gelca and Andreescu)

A Harder Example

Problem

Let X be a subset of the positive integers with the property that the sum of any two (not necessarily distinct) elements in X is again in X. Suppose that $\{a_1, a_2, \ldots, a_n\}$ is the set of all positive integers not in X. Prove that $a_1 + a_2 + \ldots + a_n \leq n^2$. (from Putnam and Beyond, by Gelca and Andreescu)

Proof.

Start by placing the a_i in increasing order: $a_1 < a_2 < \ldots < a_n$. Since the sum of two elements in X is also in X, if a_i is in the complement of X, for $1 \leq m \leq \frac{a_i}{2}$ either m or $a_i - m$ is not in X. There are $\lceil \frac{a_i}{2} \rceil$ such pairs but only $i-1$ integers less than a_i and not in X. This implies $a_i \leq 2i - 1$, and summing over all $i = 1, \ldots, n$ gives $a_1 + a_2 + \ldots + a_n \leq n^2$. $\square$

Notes about the Extremum Principle

- Often, as in the last problem, it is useful merely to *place the elements in some order*.
- As we saw with the Senate of Kazakhstan, the same basic solution can sometimes be framed using either monovariants or extrema.
- Extrema are *free*: for instance, if positive integral solutions exist to a problem, there is automatically a smallest solution. This provides us with additional information and structure for the problem without any extra effort.
- The Extremum Principle is particularly applicable in discrete problems where extrema are guaranteed to exist. It can be useful in continuous settings as well (we may still be able to put a sequence in order), but certain approaches no longer work. For instance, if we are investigating whether an equation has positive real solutions, we cannot take the "smallest" solution and work from there. It might not exist!