THE TEACHING OF MATHEMATICS
2013, Vol. XVI, 2, pp. 67–78

A NOTE ON INFINITE DESCENT PRINCIPLE

Miljan Knežević and Đorđe Krtinić

Abstract. We provide some theoretical background for infinite descent principle and its relation to the principle of mathematical induction and the well ordering property. Also, we provide some interesting examples by applying the infinite descent principle as an extremal principle in several situations. At the end we prove several assertions which confirm the understanding that principle is very important for the students when solving various complex problems they are faced with.

MathEduc Subject Classification: E65, F65, K25

MSC Subject Classification: 97E60, 97F60, 97K20

Key words and phrases: Infinite descent principle; principle of mathematical induction; well ordering property.

1. Introduction

The first encounter with basic mathematical concepts for every man is closely related to the creation of intuition about natural numbers. At first, we work with some concrete natural numbers, we add them, multiply them and we are confident we will always get a natural number as a result. Even as beginners, we also create a similar intuition about the other sets of numbers.

Natural numbers arose from the counting needs and the set of natural numbers $\mathbb{N} = \{1, 2, \dots\}$ was the first nontrivial abstract notion for humans. For teachers, one of the main tasks in teaching mathematics is to develop, in the right way, the concepts of numbers we are dealing with, especially the concept of natural numbers. Most of our current students will also be teachers tomorrow, so we must constantly strive to give our students a better idea about the notion of natural numbers and have to extend it and deepen it. Thus, we have to introduce our students to Peano's axioms and in order to establish the set of the naturals formally, we need to show them their basic properties and their consequences and to prove to students some other logically equivalent statements. The most important part of Peano's axioms is the postulate of the mathematical induction (axiom $A5$), which, in its original form, is not the same as the well know method of proving by using mathematical induction principle (WPMI – weak mathematical induction principle).

Historically, mathematical induction principle was first used by mathematician Francesco Maurolico (1494–1575) in the 16th century. In his work he stated a lot of properties of natural numbers and he proved some of them by using WPMI. Otherwise, the first formal explanation of mathematical induction principle was presented by the mathematician De Morgan, in 1838, and this was also the first time that the term induction was used. On the other hand, in contrast to geometry, axiomatized since the ancient times, mathematician Peano was the first who gave an axiomatic foundation of the set of natural numbers $\mathbb{N}$, that is established on three basic concepts: natural number, immediate successor and constant 1. He arranged them according to the following axioms:

(A1) 1 is a natural number.

(A2) The immediate successor x' of any natural number x is a natural number.

(A3) Different natural numbers never have the same immediate successor.

(A4) 1 is not the immediate successor of any natural number.

(A5) If $M \subset \mathbb{N}$ is any set for which:

(1) $1 \in M$ and

(2) for all $x \in \mathbb{N}$, if $x \in M$ then $x' \in M$.

Then $M = \mathbb{N}$.

The last axiom (A5) is known as the induction axiom.

Usually, high school students study the method of proving some properties of natural numbers by mathematical induction principle (see Proposition 1), but they do not know that this important principle can be easily deduced from the induction axiom (A5). Also, they do not understand the principle completely and they use it for granted in many situations (number theory, geometry, combinatorics, etc).

In order to lead students closer to the point, we first need to formulate and then prove the following proposition known as the weak mathematical induction principle (WPMI) for them. In literature, this form is used instead of the induction axiom (A5). Of course, we assume here that the basic properties of the set of natural numbers $\mathbb{N}$ are introduced, including notation, basic operations and ordering. For example, instead of $\mathbb{N} = \{1, 1', (1')', \dots\}$, we use decimal representation of natural numbers and instead of the immediate successor x', we write $x + 1$, for all $x \in \mathbb{N}$.

Proposition 1. (WPMI) *Let $p(n)$, $n \in \mathbb{N}$, be an arbitrary statement that depends on the natural number n. Suppose that:*

(a) *$p(1)$ is a true statement and*

(b) *for all $n \in \mathbb{N}$, if the statement $p(n)$ is true then the statement $p(n+1)$ is also true.*

Then, the statement $p(n)$ is true for all natural numbers $n \in \mathbb{N}$.

Proof. Let $M \subset \mathbb{N}$ be the set of all natural numbers n for which the statement $p(n)$ is a true statement. Obviously, $1 \in M$, since the condition (a) holds. On the other hand, if $n \in N$ is an arbitrary number for which $n \in M$, i.e. for which the statement $p(n)$ is true, then, by the condition (b), the statement $p(n+1)$ is also true. That is $n' = n + 1 \in M$. Finally, according to the axiom $(A5)$, $M = \mathbb{N}$. ∎

2. Infinite descent principle

Roughly speaking, the infinite descent principle (PID) is known as a method, or a proving technique, for solving various problems in number theory and geometry and it is used in very simple situations, as well as in proving numerous important theorems in modern mathematics. For the first time it was mentioned by that name by the great amateur mathematician Pierre de Fermat (1601–1665) while he was working in Diophantine equation analysis. Now, our students find this important principle useful even as a special method of proving by contradiction, even though it is logically equivalent to the mathematical induction principle, weak and strong, and to the well ordering property of nonempty subsets of the naturals. Frequently, one can use the infinite descent principle to show the nonexistence of a subset of the set of natural numbers with certain property also. Note that there are many variants of that principle and we will mention some of them, especially those that will be important to our students for improving mathematical reasoning.

In its original form, the infinite descent principle appears as follows:

PID. *Let $p(n)$, $n \in \mathbb{N}$, be an arbitrary property of natural number n. Assume that*

(e) *$p(1)$ is a true statement and*

(f) *for all $k \in \mathbb{N}$, if the statement $p(k)$ is false then there exists a natural number m, $m < k$, for which the statement $p(m)$ is also false.*

Then, the statement $p(n)$ is true for all natural numbers $n \in \mathbb{N}$.

In our work with young gifted students the following variant of the infinite descent principle is useful. Thus we state it as a proposition below.

Proposition 2. *There is no infinite strictly decreasing sequence of natural numbers.*

Proof. Let us presume for a moment that the statement of the infinite descent principle holds. Next, we assume that there exists a sequence (a_n), $a_n \in \mathbb{N}$, $n \in \mathbb{N}$, with the property $a_1 > a_2 > a_3 > \dots$. Therefore, if we define a set $A = \{a_1, a_2, \dots\}$, then we can construct a property $p(n)$ of a natural number n as $n \notin A$. Obviously, the statement $p(1)$ is true, since $1 \notin A$. Thus, the condition (e) is satisfied. Also, by the construction of the property $p(n)$ and the sequence (a_n), the condition (f) is satisfied too, since if there is a natural number k, for which the statement $p(k)$ is not true, one can always find the smaller number m with the same property. This leads us to a contradiction related to the validity of the infinite descent principle, since $p(a_n)$ is not a true statement, for all $n \in \mathbb{N}$.

Otherwise, if we are dealing with an arbitrary property $p(n)$ of natural numbers $n \in \mathbb{N}$, by assuming that there is no infinite strictly decreasing sequence of the naturals, we could easily obtain the infinite descent principle. Indeed, if the conditions (e) and (f) for the given property hold, then if there is any $k > 1$, for which the statement $p(k)$ is false, we can find a smaller number $1 < m < k$ with the same property and so on. Hence, we will construct an infinite decreasing sequence of naturals that contradicts the assumption about the number k. Thus, the statement $p(n)$ is true, for all $n \in \mathbb{N}$, and the infinite descent principle holds. ∎

Remark. As an important fact here, we note that our primary school students use the previous variant of the infinite descent principle to show that the number $\sqrt{2}$ is irrational. Also, it would be very interesting to investigate whether the infinite descent principle is more practical and much more appropriate at this age in relation to the weak induction principle. So, the major aim of this paper is to show the beauty of the infinite descent principle and its applications.

3. Strong induction principle and well ordering property

In working with young students, we are often faced with different variants of the induction. Usually, it is important for them to understand the concepts of the strong mathematical induction principle (SPMI) and well ordering property (WOP). These principles can be applied in many situations.

SPMI. *Assume that:*

(c) $p(1)$ *is a true statement.*

(d) *For all* $k \in \mathbb{N}$, *if the statements* $p(1), p(2), \dots, p(k)$ *are true then the statement* $p(k+1)$ *is also true.*

Then, the statement $p(n)$ *is true for all natural numbers* $n \in \mathbb{N}$.

WOP. *Every nonempty subset of the set of natural numbers* $\mathbb{N}$ *has the smallest element.*

Here we show that these principles have interesting properties.

Proposition 3. WPMI $\Rightarrow$ SPMI.

Proof. Let $p(n)$, $n \in \mathbb{N}$, be an arbitrary statement for which the conditions (c) and (d) are satisfied. We define a new statement $q(n) = p(1) \wedge p(2) \wedge \ldots \wedge p(n)$, $n \in \mathbb{N}$. Obviously, by (c), $q(1) = p(1)$ is a true statement. Assume now that for some $k \in \mathbb{N}$ the statement $q(k)$ is true. Then, since $q(k) = p(1) \wedge p(2) \wedge \ldots \wedge p(k)$, the statements $p(1), p(2), \dots, p(k)$ are all true. Therefore, from the condition (d), we conclude that the statement $p(k+1)$ is also true. Thus, $q(k+1) = p(1) \wedge p(2) \wedge \ldots \wedge p(k) \wedge p(k+1)$ is a true statement. By applying the WPMI to the statement $q(n)$, $n \in \mathbb{N}$, we obtain that the statement $q(n)$ is true for all $n \in \mathbb{N}$. The same holds for the statement $p(n)$, since $q(n) = p(1) \wedge p(2) \wedge \ldots \wedge p(n)$, $n \in \mathbb{N}$. ∎

Proposition 4. SPMI $\Rightarrow$ WOP.

Proof. Let $A \subset \mathbb{N}$, $A \neq \emptyset$, be an arbitrary set. We will show that the set A has the smallest element. First, we define a property $p(n)$, of natural number n, by the following statement: if $A \subset \mathbb{N}$ and $n \in A$, then A has a least element. Then, if we prove that the statement $p(n)$ is true for all $n \in \mathbb{N}$, since the set A is nonempty, it must contain some $k \in \mathbb{N}$ and, therefore, has a least element, because the statement $p(k)$ is true.

Let us now prove that the property $p(n)$ is true for all natural numbers. Clearly, if $1 \in A$, then 1 is the least element of the set A indeed. Thus, the statement $p(1)$ is true. On the other hand, assume that the statements $p(m)$ are true for all $1 \leqslant m \leqslant k$, where k is an arbitrary natural number. If the number $k+1 \in A$, we have the following two cases. In the first case if the number $k+1$ is the least element of the set A, the statement $p(k+1)$ is true. In the second case, if the number $k+1$ is not the least element of the set A, then there exists a natural number m, $1 \leqslant m \leqslant k$, with the property that $m \in A$, since $k+1 \in A$. According to the assumption, since $p(m)$ is a true statement, we conclude the set A has the smallest element and, hence, the statement $p(k+1)$ is obviously true. According to the strong mathematical induction principle, we obtain that the statement $p(n)$ is a true statement for all $n \in \mathbb{N}$. This completes the proof by the argument above. ■

Proposition 5. WOP $\Rightarrow$ PID.

Proof. Assume that, for the given property $p(n)$, the conditions (e) and (f) are satisfied. We define a set $A \subset \mathbb{N}$ by $A = \{n \in \mathbb{N} : p(n)\text{is not a true statement}\}$. If the set A is empty, there is nothing to prove, but if not, by applying the well ordering principle to the set A, there is some $k \in \mathbb{N}$, $2 \leqslant k$, since $1 \notin A$ by the condition (e), which is the smallest element of the set A. However, by using the condition (f), since the statement $p(k)$ is false, there exists a natural number m, $m < k$, for which the statement $p(m)$ is also false. Thus, $m \in A$, $m < k$, and this contradicts the assumption that k is the smallest element of the set A. Hence, the set A is empty, i.e. the statement $p(n)$ is true for all $n \in \mathbb{N}$. ■

Proposition 6. PID $\Rightarrow$ WPMI.

Proof. Let $p(n)$, $n \in \mathbb{N}$, be an arbitrary statement that depends on the natural number n for which the conditions (a) and (b) are satisfied. We have to prove that the condition (f) is satisfied too and, since the condition (e) is trivially satisfied, we could apply the infinite descent principle to conclude that the property $p(n)$ is true for all $n \in \mathbb{N}$.

To do so, we assume that the condition (f) is not valid. Then, there exists $k \in \mathbb{N}$ for which the statement $p(k)$ is false, but for every $m \in \mathbb{N}$, $m < k$, the statement $p(m)$ is true. Hence, since $k \geqslant 2$, we find the number $k_1 = k - 1 \in \mathbb{N}$, $k_1 < k$, for which the statement $p(k_1)$ is true, but the statement $p(k_1 + 1) = p(k)$ is not. Thus, the condition (b) does not hold and we get a contradiction. Finally, by applying the infinite descent principle, the property $p(n)$ is true for all $n \in \mathbb{N}$. ■

4. Interesting examples

In the previous section we have seen that the infinite descent principle is equivalent to the mathematical induction principle and in some examples is very often a more natural way of thinking. Students cannot be expected to think in that direction if they have never encountered the notion of infinite descent and similar ideas. Thus, it is very important to introduce them to such a way of thinking. Therefore, in this section we are going to present a few examples which could help students understand the concept of the infinite descent principle which was based on the experience we have gained in working with students candidates for the IMO (International Mathematical Olympiad) team of Serbia.

It would be beneficial for students to encounter the infinite descent principle for the first time through different kinds of games. This approach is also both suitable and easy for teachers. Frequently, those examples are much easier then others and are often based in some way of symmetry preserving. We illustrate such a view by using the following examples.

EXAMPLE 1. *Two players are playing two-moves chess, i.e. chess in which each player can make two moves. Prove that the player that plays with the black figures does not have the winning strategy.*

Solution. Suppose, for the purpose of contradiction, that the black player has a winning strategy. Hence, for every initial two moves that the white player makes, the black player has a winning strategy. Consider the following two initial moves made by the white player: $G1 - F3$ and $F3 - G1$. Note that these two moves do not alter the board, so the black player has the same position as the white player had before making his moves. Since this new position is essentially the same as the original one, except that the board is flipped, then the white player has the winning strategy now. This is a contradiction. ∎

EXAMPLE 2. *Two players, white and black, play the following game on an ordinary chess board. At the beginning in every square of the first row there is a white token, while in every square of the eighth row there is a black token. The players alternately make the moves and each player can move the token of the corresponding color. A move consists of displacing the token by the player in some square in the same column in which it is located, so the tokens should approach each other. The players cannot move the tokens backward and the tokens cannot be skipped. The game ends when one of the players cannot make the move. Prove that the black player has a forced victory.*

Solution. Since the number of columns on the board are even, one could divide them into pairs. After the first move made by the white player, which consists of moving a token for i places in the column that is chosen, the black player can respond by moving the corresponding token for i places in the column that is paired with a column in which the white player made a move. Thus, since the total number of moves the player could make is a finite number (it could be estimated by considering the number of empty squares), the game must be completed. So, the black player has the winning strategy. ∎

At first glance, the previous examples are very similar, but they show two totaly different situations. The first example points to maintaining the same position, while in the other students are faced with the notion of *invariant*. Examples of the second type are easily accepted by students even in the fifth grade of elementary school (there are lot of examples of games in which the winning strategy is reduced in order to maintain some form of symmetry). Observe that in the second example we consider that a finite subset of the set of all natural numbers is bounded above. Otherwise, in many situations it is more natural to refer shortly to the infinite descent principle as *extremal principle*. The next example is typical for introducing students to such a way of thinking.

EXAMPLE 3. *At a tournament, in which there were $n \geqslant 2$ players, each player played exactly one game with each of the remaining players. After the tournament each player made a list of players that he had defeated. Prove that there is such a player on whose list, as well as on the lists of the other players he defeated, one can find the names of all the remaining players.*

This example shows that we require of the students to think less about some form of symmetry, but about the existence of an "extremal" player. In this specific case, it would be the player with the highest number of wins, i.e. one of such players if there are players with the same extremal number of wins. Such a player obviously exists (since any nonempty finite subset of $\mathbb{N}$ has the greatest element) and we denote him by A. Suppose then that there exists a player B that is not on the list of player A, neither on the lists of the players which that defeated by player A. Thus, player B defeated player A and, also, all the players that were defeated by player A. Hence, player B has more wins then player A, which contradicts the assumption.

Also, one of the typical ways of comparison is a comparison by using the lexicographic order (in the case of numbers it is the concept of number system). These concepts can be naturally associated with *infinite descent principle*. Also, the main emphasis in the implementation of the principle usually lies in finding the "right order".

EXAMPLE 4. *Several positive integers are written in a row. Iteratively, one after another (if it is possible) a player performs the following operations. He chooses two adjacent numbers a and b, such that $a > b$ and a is on the right of b. Then, he replaces the pair (b, a) by either $(a, b+1)$ or $(a, a-1)$. Prove that the player can perform only a finite number of such operations.*

Solution. First, observe that those operations do not change the greatest element of the sequence that is written, i.e. it appears in the sequence by applying the operation and it could not be increased. Suppose that at the beginning it was written $(x_1, x_2, \ldots, x_n)$, $n \in \mathbb{N}$. We say that $(x_1, x_2, \ldots, x_n) \prec (y_1, y_2 \ldots, y_n)$ if and only if $x_j < y_j$, for some j, and for all $i < j$ holds $x_i = y_i$. The operations the player could perform strictly increase $n-$tuple in the previous sense, i.e. with respect to the relation we defined. Since the greatest element is bounded we conclude that the total number of such $n-$tuples the player can obtain are bounded also, so he could perform only the finite number of those operations. ■

Note that there are different approaches for solving the previous problem. Sometimes it is very difficult to lead the students to find an appropriate invariant that could help them find a solution for many problems they are faced with. The next example we give is quite difficult even for the high school students.

Example 5. *Suppose that on the table there are $\frac{n(n+1)}{2}$ balls arranged in several piles. In one move the player takes one ball from each pile and forms a new pile from these balls. If only one ball was in a pile, after the move that pile disappears. Prove that after a finite number of moves exactly n piles will remain, so that in i-th pile there will be exactly i balls, $1 \leqslant i \leqslant n$.*

Solution. Let us suppose the piles are sorted by size in non increasing order. Let us assign to the ball j, lying in the i-th pile, the square of integer grid whose upper right vertex has coordinates (i, j). We will call that square the square (i, j). Let the "characteristics" of that square be $i + j$. In this way the game is quite easy for graphical presentation. The number of balls in the new pile is equal to the number of piles from the previous step, that is the number of assigned squares in the first row. It follows that in this model a move can be seen as "taking" squares from the first row, pushing other squares one position right and one position down, "setting" the previous first row to a place of the first column and, if necessary, rearranging columns by size (in the case when new first column is not the largest).

Let S be the sum of "characteristics" of all associated squares. In the described procedure, if the new pile is the largest, the ball that was assigned to the square $(1, j)$ after operation will be assigned to the square $(j, 1)$. Also, the ball that was assigned to the square (i, j) for $i \geqslant 2$ will be assigned to the square $(i - 1, j + 1)$. It follows that in both situations the square with the same "characteristics" will be assigned to the same ball, so these operations will not change S. If the new pile is not the largest, columns will be exchanged, and in that case S will decrease.

According to PID, at one moment, S will reach a value that will not be decreased and no column exchange will happen. This means that the "path" of the ball assigned to the square (i, j) will be $(i, j) \to (i+1, j-1) \to \ldots \to (1, i+j-1) \to (i+j-1, 1) \to \ldots$, i.e. the square will move periodically on "diagonal" of the coordinate grid, with a period $i + j$. In addition, under the conditions of the problem, at the moment the ball is assigned square (i, j), some balls will be assigned to squares (i, k) for every $k < j$ (there must not be a "hole"). Since the lengths of two consecutive diagonals are relatively prime integers, it follows that if the diagonal $i + j$ contains a square assigned to a ball, then all the squares on diagonal $i + j - 1$ are assigned to balls (and therefore it applies for all k diagonals, for $k < i + j$).

Thus the diagonals must be filled in order, so it follows that first n diagonals are filled, i.e. after the finite number of moves balls will be distributed at n piles, so that i-th pile contains i balls for $1 \leqslant i \leqslant n$ (for arbitrary initial position). ∎

For most of the high school students the first acquaintance with PID is related to the basic concepts in number theory. Usually, the appearance of the PID in that area comes later than in combinatorics, since it requires a higher level of knowledge. However all of these ideas may be developed in primary school. Historically, the following two examples are well known. A student could encounter the first one even in the primary school.

Example 6. *Prove that $\sqrt{2}$ is not a rational number.*

Solution. Suppose that $\sqrt{2} = \frac{p}{q}$, $p, q \in \mathbb{N}$, such that q is as small as possible.

Of course, we have $p > q$. It follows that $2q^2 = p^2$, so we have $2|p$. Thus, $p = 2p_1$, for some $p_1 \in \mathbb{N}$, and $q^2 = 2p_1^2$. Hence, $\sqrt{2} = \frac{q}{p_1}$ and we have $q > p_1$, which contradicts the choice of q. ■

EXERCISE 1. Let p be a prime number of the form $4k - 1$, for some $k \in \mathbb{N}$. Prove that the equation $x^2 + y^2 = p(u^2 + v^2)$ has only the trivial solution in $\mathbb{Z}^4$.

The following example is suitable for high school students.

EXAMPLE 7. *Prove that the equation $x^4 + y^4 = z^2$ has no solutions in positive integers.*

Solution. If there is a solution (x, y, z) with $p \mid x$ and $p \mid y$, then $p^4 \mid z^2$, i.e. $p^2 \mid z$, then the triple $(\frac{x}{p}, \frac{y}{p}, \frac{z}{p^2})$ is also a solution. Thus, we suppose that there is a solution (x, y, z) with $(x, y) = 1$. We will chose such a solution for which z is as small as possible.

The numbers x and y can not both be odd. If that is true, then $x^4+y^4 \equiv 2 \equiv z^2 \pmod 4$, but the squares modulo 4 are 0 and 1. It follows that (x^2, y^2, z) is a primitive Pithagorean triple, so we have $x^2 = k^2 - l^2$, $y^2 = 2kl$, $z = k^2 + l^2$, for some mutually prime k, l (here we assumed that $2 \nmid x$ and $2 \mid y$). From $y^2 = 2kl$ and $(k, l) = 1$ we have $k = k_1^2$, $l = 2l_1^2$ (if $k = 2k_1^2$, $l = l_1^2$, then $1 \equiv x^2 = k^2 - l^2 \equiv 0 - 1 \equiv -1 \pmod 4$). Thus, $k^2 = l^2 + x^2$ and since $(k, l) = 1$ and $2 \nmid x$, we have $k_1^2 = k = s^2 + t^2$ and $2l_1^2 = l = 2st$, $(s, t) = 1$. Therefore, $s = u^2$, $t = v^2$ and $k_1^2 = k = u^4 + v^4$, $(u, v) = 1$. Thus, the triple (u, v, k_1) is also a solution with $k_1 \leqslant k < k^2 + l^2 = z$, which contradicts the choice of z. ■

We will present below a few examples of the PID in the field of number theory which prove that it is hard to expect of students to use this principle for solving the tasks if they are not already familiar with such ideas. The first of these was the most difficult problem that appeared at the IMO in 1988. Namely, that problem was done worst in the competition and only 11 of the 268 competitors completely solved the problem. The members of the SFRY team (at that time the students from Serbia were a part of the SFRY team) had zero points on that assignment. Average points of all competitors on that problem was 0.634. Here we note that each task at every IMO carries 7 points. Also, that problem, among all other things, contributed to the PID to become an integral part of the preparation of almost all teams, especially for those that are highly ranked.

Such an approach in working with students brought results. Indeed, almost the same type of assignment that appeared at the IMO in 2007 was solved better by the contestants, although at the same competition there were two of the problems solved worse than that. The members of the Serbian national team won 18 points on that problem. However, 94 of the 520 competitors who competed at that IMO solved the problem entirely. At that time the average of all the points on that problem which were won by the contestants was 1.898. Note that the improvement was greater since large number of countries, that were for the first time involved at such a kind of competition, were typically below the average of the points of all contestants who took part at the IMO in 1988.

EXAMPLE 8. *Let* a, b *be positive integers, such that* $ab+1 \mid a^2+b^2$. *Prove that* $\dfrac{a^2+b^2}{ab+1}$ *is the square of an integer.*

Solution. Let $\frac{a^2+b^2}{ab+1} = k \in \mathbb{N}$. Obviously, $\min\{a^2, b^2\} \geqslant k$. If not, say $a^2 < k$, we have $b^2 - kab = b(b-ka) = k - a^2 > 0$, i.e. $b > ka$. Hence, $b = ka + x$, for some $x \in \mathbb{N}$. Therefore, from $a^2+b^2 = k(ab+1)$, we obtain $a^2+(ka+x)^2 = ka(ka+x)+k$, i.e. $a^2 + kax + x^2 = k$. But, $k \leqslant kax < a^2 + kax + x^2 = k$, which is impossible.

Let $S = \{(m,n) \in \mathbb{N}^2 : \frac{m^2+n^2}{mn+1} = k,\ m \geqslant n\}$. Set S is nonempty, since $(a,b) \in S$ or $(b,a) \in S$. According to the PID, there is a pair $(c,d) \in S$ for which the second coordinate, i.e. number d, is as small as possible. Thus, $c^2 - kdc + d^2 - k = 0$, i.e. number c is a solution to that quadratic equation. Let c' be the second solution. Due to the Viet's formulae we have $c + c' = kd$ and $cc' = d^2 - k$. Therefore, $c' = \frac{d^2-k}{c} \leqslant \frac{d^2-k}{d} < d$. Recall that $d^2 \geqslant k$, so $c' \geqslant 0$. If $c' > 0$, then $(d, c') \in S$ also. But that pair has the second coordinate smaller than the pair (c,d), which is opposite to the choice of d. Hence, $c' = 0$ and thus $k = d^2$. ∎

EXERCISE 2. Let a, b be positive integers. If $4ab-1$ divides $(4a^2-1)^2$, prove that $a = b$.

We will prove a more general statement.

EXAMPLE 9. *Let* a, b, k *be positive integers and* $k > 1$. *If* $kab-1$ *divides* $(ka^2-1)^2$, *then* $a = b$.

Solution. Obviously $0 \equiv (ka^2-1)^2 \equiv \big(ka^2 - kab\big)^2 = k^2a^2(a-b)^2 \pmod{kab-1}$, since $kab \equiv 1 \pmod{kab-1}$. Thus, since $(ka, kab-1) = 1$, we have $kab-1 \mid (a-b)^2$.

Assume now that there are some positive integers $a \neq b$ such that $kab-1 \mid (a-b)^2$. Let $\dfrac{(a-b)^2}{kab-1} = q \in \mathbb{N}$ and $S = \{(m,n) \in \mathbb{N}^2 : \dfrac{(m-n)^2}{kmn-1} = q\}$. We now proceed as in the previous example.

Set S is nonempty. According to the PID, there is a pair $(A,B) \in S$ which minimizes $m+n$, over all $(m,n) \in S$. Without the loss of generality assume $A > B$. Consider now the quadratic equation

$$\frac{(x-B)^2}{kxB-1} = q \quad \Leftrightarrow x^2 - (2B+kqB)x + B^2 + q = 0,$$

which has a solution $x_1 = A$. If x_2 is the second solution to that equation, from Vieta's formulae we have $x_2 = 2B + kqB - A = \dfrac{B^2+q}{A}$. Obviously, $x_2 \in \mathbb{N}$ and the pair $(x_2, B) \in S$. Due to the minimality of the pair (A,B), we get $x_2 + B \geqslant A + B$, i.e. $x_2 \geqslant A$. Thus, $\dfrac{B^2+q}{A} \geqslant A$, i.e. $q \geqslant A^2 - B^2$. Hence, $q = \dfrac{(A-B)^2}{kAB-1} \geqslant A^2 - B^2$, or equivalently $A - B \geqslant (A+B)(kAB-1)$, which is impossible, since $A, B, k \in \mathbb{N}$ and $k > 1$. Therefore, $a = b$. ∎

After being introduced to these ideas students accept some nontrivial methods very easily, such as solving Pell's equation and applying the PID in solving functional equations or in analysis. These examples are widely known, and we have singled out only one that was at the IMO few years ago. Finally, we show that the PID, as an extremal principle, can easily be found at "less expected" places, such as in geometry. We will illustrate the fact with two examples, historically very important, known as Sylvester's problem and Littlewood's problem. The proof of the second one is left to an eager reader as an exercise.

EXAMPLE 10. *Let* $f : \mathbb{N} \to \mathbb{N}$ *be such that* $f(n+1) > f\big(f(n)\big)$*, for all* $n \in \mathbb{N}$*. Prove that* $f(n) = n$*, for all* $n \in \mathbb{N}$*.*

Solution. Due to the PID, there is $d = \min\{f(n) : n \in \mathbb{N}\}$. We chose $m \in \mathbb{N}$ such that $d = f(m)$. If $m > 1$, then $d = f(m) > f\big(f(m-1)\big)$, so $f(m-1) \in \mathbb{N}$ is an element to which the smaller value is attained. From the choice of d, it follows that $m = 1$.

Similarly, the smallest element of the set $\{f(n) : n \in \mathbb{N}, n \geqslant 2\}$ is $f(2)$. If $f(1) = f(2)$, then $f(1) = f(2) > f\big(f(1)\big)$, which is opposite to the choice of $f(1)$. If we proceed with this approach, we get $f(1) < f(2) < \ldots < f(n) < \ldots$, and since $f(1) \geqslant 1$, we obtain $f(k) \geqslant k$.

Finally, if $f(k) > k$, for some $k \in \mathbb{N}$, then $f(k) \geqslant k+1$, so $f\big(f(k)\big) \geqslant f(k+1)$. On the other hand, according to the formulation of the problem, we have $f\big(f(k)\big) < f(k+1)$, which is impossible. Thus, $f(k) = k$, for all $k \in \mathbb{N}$. ∎

EXAMPLE 11. [Sylvester's problem] *A finite set of* $n \geqslant 3$ *points in the plane has the property that any line thought two of them passes through a third. Then all the points lie on a line.*

Solution. Let $\mathcal{P}$ be the set of points given. Suppose that these points are not all collinear. Consider the set $\mathcal{L}$ of all lines that are determined by the pairs of points in $\mathcal{P}$. Now, for each point $P \in \mathcal{P}$ and each line $\ell \in \mathcal{L}$, with $P \notin \ell$, consider the pair (P, ℓ). Obviously, the number of such pairs is finite. Thus, there exists a point P_0 and a line ℓ_0, i.e. the pair (P_0, ℓ_0), for which the distance from point P_0 to line ℓ_0 is as small as possible, among all pairs. Let Q be the projection of point P_0 on line ℓ_0. Since line ℓ_0 contains at least tree different points of set $\mathcal{P}$, at least two of them will lie on the same side of Q. Denote those points with R and S, with $Q - R - S$ (we could have the situation in which points Q and R are the same). Hence, point R has the distance to the line determined by the points P_0 and S which is smaller than the distance from point P_0 to line ℓ_0. This fact leads us to a contradiction. ∎

EXERCISE 3. [Littlewood's problem] A cube given can not be divided into a finite number of pairwise distinct cubes.

ACKNOWLEDGMENT. We are taking this opportunity to thank to professors Zoran Kadelburg, Vladimir Janković and Vladimir Božin who read this article and gave us useful comments and suggestions.

REFERENCES

[1] C. B. Boyer, *A History of Mathematics*, Wiley, New York, 1968.

[2] A. Engel, *Problem Solving Strategies*, Springer, 1998.

[3] D. Djukić, V. Janković, I. Matić, N. Petrović, *The IMO Compendium: A Collection of Problems Suggested for the International Mathematical Olympiads*: 1959-2004, Springer, 2006.

[4] D. S. Gunderson, *Handbook of Mathematical Induction: Theory and Applications*, Discrete Mathematics and Its Applications, 2010 by Chapman and Hall/CRC.

[5] A. G. Kurosh, *Lectures on General Algebra*, Translated from the Russian edition (Moscow, 1960) by K. A. Hirsch. Chelsea, New York, 1963.

Faculty of Mathematics, University of Belgrade, Studentski trg 16, 11000 Beograd, Serbia

E-mail: kmiljan@math.rs, georg@matf.bg.ac.rs