# INTRODUCTION TO INEQUALITIES

VIPUL NAIK

Abstract. This is a somewhat modified version of the notes I had prepared for a lecture on inequalities that formed part of a training camp organized by the Association of Mathematics Teachers of India for preparation for the Indian National Mathematical Olympiad (INMO) for students from Tamil Nadu.

## 1. Basic idea of inequalities

1.1. **What we need to prove.** An "inequation" is an expression of the form:

$$F \geq 0$$

where $F$ is an expression in terms of certain variables. An "inequality"' is an inequation that is satisfied for all values of the variables (within a certain range).

For instance:

$$x^2 - x + 1 \geq 0$$

and

$$x^2 - x - 1 \geq 0$$

are both inequations. Among these, the first inequation is true for *all* real $x$, while the second inequation is true for all values of $x$ within a certain range.

Thus, when we talk of an inequality, we have the following in mind:

- The underlying *inequation*
- The *range of values* over which the inequality is true

A *strict* inequation is an inequation of the form:

$$F > 0$$

where $F$ is an expression in terms of the variables.

Given any inequation $F \geq 0$ we can consider the corresponding strict inequation $F > 0$.

Thus, when studying an inequality, we are interested in:

- The underlying *inequation*
- The *range of values* over which the inequality is true
- The values for which exact equality holds

Some other points to note:

- Any inequation of the form $F \geq G$ where $F$ and $G$ are both expressions can be written in the standard form as $F - G \geq 0$. The original inequation is true for precisely those values for which the standard form is true. The equality conditions are also the same.
- An inequation of the form $F \leq G$ can be expressed as $G - F \geq 0$. Again, the original inequation is true for precisely those values for which the standard form is true. The equality conditions are also the same.

1.2. **No square is negative.** This basic inequality states:

$$x^2 \geq 0$$

The range is all $x \in \mathbb{R}$ and equality holds iff $x = 0$.

This can be generalized to something of the form:

$$(f(x_1, x_2, \ldots, x_n))^2 + (g(x_1, x_2, \ldots, x_n))^2 \geq 0$$

The range is all $x \in \mathbb{R}$ and equality holds iff $f(x_1, x_2, \ldots, x_n) = g(x_1, x_2, \ldots, x_n) = 0$.

**Problem 1.** Prove that $x^4 - x^2y^2 + y^4 \geq 0$ for all real $x$ and $y$, equality holding iff $x = y = 0$.

*Proof.* We use:

$$x^4 - x^2y^2 + y^4 = (x^2 - y^2)^2 + (xy)^2$$

Thus, $(x^2 - y^2)$ plays th role of $f$ above and $xy$ plays the role of $g$

Clearly then, the left-hand-side is nonnegative, and is 0 if and only if $x^2 = y^2$ and $xy = 0$, thus forcing $x = y = 0$. □

We can extend the idea to sums of more than two squares:

**Problem 2.** Prove that $a^2 + b^2 + c^2 + ab + bc + ca \geq 0$ with equality holding only if $a = b = c = 0$.

*Proof.* The left-hand-side can be expressed as $1/2(a^2 + b^2 + c^2 + (a + b + c)^2)$. So it is nonnegative and can be zero only if $a = b = c = 0$.

Alternatively, the left hand side can also be written as $1/2((a + b)^2 + (b + c)^2 + (c + a)^2)$ and is hence nonnegative, taking the value 0 if and only if $a = b = c = 0$ □

Another problem (for which I'm not writing the solution here):

**Problem 3.** Prove that $a^2 + b^2 + c^2 - (ab + bc + ca) \geq 0$ with equality holding only if $a = b = c$.

It turns out that one of the solution techniques for the previous problem can be applied to this one.

1.3. **Manipulating about the inequality symbol.** The following results are typically used for manipulating inequalities:

- We can *add* two inequalities. The greater side gets added to the greater side, the smaller side to the smaller side. If either inequality is strict, the resultant inequality is again strict. More generally, the set of values for which the resultant inequality becomes equality is the intersection of the corresponding sets for each inequality.
- We can multiply both sides of an inequality by a positive number. In general, however, we cannot multiply two inequalities.

## 2. Mean inequalities

2.1. **Definition of means.** A *mean* is a good notion of *average* for a collection of numbers. A mean of $n$ numbers is thus typically a function from $n$-tuples of reals to reals, such that:

- If all the members of the tuple are equal, the mean should be equal to all of them. That is, if $a = a_1 = a_2 = \ldots a_n$ then the mean of $a_1, a_2, \ldots, a_n$ is $a$.
- The mean is a symmetric function of all the elements of the tuple, that is, if the elements are permuted, the value of the mean remains unchanged. That is, the mean of $a_1, a_2, \ldots, a_n$ is the same as the mean of $a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(n)}$.

- The mean of a collection of positive numbers should be between the smallest number and the largest number. That is, if $a_1 \leq a_2 \leq \ldots \leq a_n$, the mean lies between $a_1$ and $a_n$.
- The mean is an increasing function in each of the arguments. That is, if $a_i \leq a_i'$, then the mean of $a_1, a_2, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n$ is less than or equal to the mean of $a_1, a_2, \ldots, a_{i-1}, a_i', a_{i+1}, \ldots, a_n$.

We now define some typical notions of mean:

**Definition.** (1) The **arithmetic mean**(defined) of $n$ real numbers $a_1$, $a_2$, $a_3, \ldots, a_n$ is defined as:

$$\frac{a_1 + a_2 + \ldots a_n}{n}$$

The arithmetic mean is a well-defined notion for *any* collection of real numbers (positive, negative or zero).

(2) The **geometric mean**(defined) of $n$ positive real numbers $a_1$, $a_2$, $a_3, \ldots, a_n$ is defined as

$$(a_1 a_2 \ldots a_n)^{1/n}$$

The geometric mean is defined only for *positive* numbers.

(3) The **quadratic mean**(defined) or the root-mean-square of $n$ real numbers $a_1$, $a_2$, $a_3, \ldots, a_n$ is defined as:

$$\sqrt{\frac{a_1^2 + a_2^2 + \ldots + a_n^2}{n}}$$

(4) The **harmonic mean**(defined) of $n$ nonzero real numbers $a_1$, $a_2$, $a_3, \ldots, a_n$ is defined as:

$$\frac{a_1^{-1} + a_2^{-1} + \ldots + a_n^{-1}}{n}^{-1}$$

For two positive reals $a$ and $b$, these boil down to the formulas:

| Name of the mean | Value |
|---|---|
| Arithmetic mean | $\frac{(a+b)}{2}$ |
| Geometric mean | $\sqrt{ab}$ |
| Quadratic mean | $\sqrt{\frac{a^2+b^2}{2}}$ |
| Harmonic mean | $\frac{2ab}{a+b}$ |

### 2.2. **Inequalities for two variables.**

**Claim.** For positive reals $a$ and $b$, Q.M. $\geq$ A.M. $\geq$ G.M. $\geq$ H.M.

*Proof.* We prove Q.M. $\geq$ A.M. The remaining proofs follow along similar lines:

What we would like to show is that, for all reals $a$ and $b$:

$$\sqrt{\frac{a^2 + b^2}{2}} \geq \frac{a + b}{2}$$

Since the left side is nonnegative, it suffices to show that the *square* of the left side is greater than or equal to the square of the right side. That is, we need to show that:

$$\frac{a^2 + b^2}{2} \geq \frac{(a + b)^2}{4}$$

But the latter rearranges to $(a - b)^2 \geq 0$. This tells us that the inequality is valid for all real $a$ and $b$ with equality holding iff $a = b$. ☐

Let's look at the pattern. The Q.M. is essentially obtained by taking the arithmetic mean of squares and then taking squareroot. The A.M. is obtained by taking the arithmetic mean of first powers and then taking the first root. The H.M. is obtained by taking the arithmetic mean of inverses and then taking the inverse. This suggests a general definition:

$$M_r(a,b) = \left(\frac{a^r + b^r}{2}\right)^{1/r}$$

Then the quadratic mean is $M_2$, the arithmetic mean is $M_1$, and the harmonic mean is $M_{-1}$.

By this definition, $M_0$ does not make sense. But it turns out that, through a suitable limit argument, we can take $M_0$ as the geometric mean. In that case, we have:

$$M_2 \geq M_1 \geq M_0 \geq M_{-1}$$

We also know that:

$$2 \geq 1 \geq 0 \geq -1$$

Does this suggest something?

2.3. **The mean inequalities: an explanation.** Let $a$ and $b$ be positive reals. What can we say about the behaviour of $M_r(a,b)$ as $r$ varies from $-\infty$ to $\infty$. It turns out that as $r \to -\infty$, $M_r$ approaches $\min\{a,b\}$, and as $r \to \infty$, $M_r \to \max\{a,b\}$. Thus, as $r$ steadily increases, $M_r(a,b)$ steadily goes from the minimum to the maximum.

The explanation for this can be sought by viewing the $r$ as a kind of *weighting* of $a$ and $b$. The greater the value of $r$, the greater the *dominance* of the bigger term, and hence, the greater the mean is to the bigger term.

2.4. **The mean inequalities for many variables.** The same phenomena which we observe for two variables also generalize to more than two variables. We define:

$$M_r(a_1, a_2, \ldots, a_n) = \left(\frac{a_1^r + a_2^r + \ldots a_n^r}{n}\right)^{1/r}$$

Again, as $r \to -\infty$, $M_r$ approaches the minimum of the $a_n$s, and as $r \to \infty$, $M_r$ approaches the maximum of the $a_n$s.

## 3. Cauchy-Schwarz inequality

3.1. **Statement.** Let $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be two $n$-tuples of real numbers. Then:

$$(\sum_i a_i^2)(\sum_i b_i^2) \geq (\sum_i (a_i b_i))^2$$

With equality holding if and only if one of the tuples is zero or if $b_i = \lambda a_i$ for some fixed $\lambda$ independent of $i$ (that is, the tuple of $b_i$s is a scalar multiple of the tuple of $a_i$s).

3.2. **Vector interpretation.** The vector interpretation of Cauchy Schwarz inequality looks at both $a = (a_1, ma_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ as vectors in $\mathbb{R}^n$. Then, the left-hand-side is:

$$|a|^2 \, |b|^2$$

where $|a|$ denotes the magnitude or length of the vector $a$

The right-hand-side is the square of the dot product of the vectors, which is the same as:

$$(a.b)^2 = |a|^2 \, |b|^2 \cos^2\theta$$

where $\theta$ is the angle between the vectors. Since $\cos^2\theta \leq 1$ and quality holds if and only if $a$ and $b$ are collinear, we get a *geometric* proof of Cauchy-Schwarz inequality.

3.3. **A trigonometric problem.** Consider the following problem:

**Problem 4.** Maximize

$$a\cos\theta + b\sin\theta$$

as a function of $\theta$ where $a$ and $b$ are fixed reals (and not both zero).

The idea is to view this as a *dot product* of vectors $(a, b)$ and $(\cos\theta, \sin\theta)$. We have:

$$(a^2 + b^2)(\cos^2\theta + \sin^2\theta) \geq (a\cos\theta + b\sin\theta)^2$$

Since $\cos^2\theta + \sin^2\theta = 1$, we obtain:

$$(a\cos\theta + b\sin\theta) \leq \sqrt{a^2 + b^2}$$

A necessary and sufficient condition for the magnitude of the left-hand side to be $\sqrt{a^2+b^2}$ is that $a/\cos\theta = b/\sin\theta$, giving $\tan\theta = b/a$. Among the two possible values for the pair $(\cos\theta, \sin\theta)$ we must pick the one making $a\cos\theta + b\sin\theta$ positive.

3.4. **A geometric problem.** Consider the following problem:

**Problem 5.** Let $A$ and $B$ be two points in a plane at distance 1. Find the maximum length of a path from $A$ to $B$, comprising at most $n$ line segments, with the property that at every stage, the distance from $B$ is reducing.

The answer is $\sqrt{n}$.

*Proof.* The idea of the proof is to use induction on $n$. Let $f(n)$ denote the maximum value for a given $n$.

We observe that any such optimal path is *memoryless* in the following sense:

Suppose $\gamma$ is a path from $A$ to $B$ comprising at most $n$ line segments, and suppose that the first line segment of $\gamma$ ends at a point $P$. Now, the part from $P$ to $B$ must be composed of $(n-1)$ line segments with the property that at every stage, the distance from $B$ is reducing.

Now, whatever path we choose, we could replace it by a path of maximum length from $P$ to $B$ comprising $(n-1)$ line segments and with the property that distance from $B$ is reducing. Since the original thing was *longest*, we conclude that the part from $P$ to $B$ must also be the *longest* one.

Now what is the longest possible path of $(n-1)$ line segments from $P$ to $B$? Since lengths scale, it is the length $PB$ times the value $f(n-1)$. We thus get:

$$\text{length of} \gamma = AP + PBf(n-1)$$

Thus the maximum of the possible lengths of $\gamma$ is the maximum over all $P$ of the above expression.

Now, from the fact that along the path $AP$, the distance from $P$ is steadily reducing, we obtain that the angle $\angle APB$ is either obtuse or right. Thus, in particular, for any given length $AP$, we have:

$$PB \leq \sqrt{1 - AP^2}$$

If equality does not hold, we could replace $P$ by another point $Q$ such that $AQ = AP$ and such that $\angle AQB = \pi/2$. Then, $QB$ would be greater than $PB$, and hence, the length of the longest path would increase. Hence, we conclude that equality does indeed hold for the longest path, viz $\angle APB = \pi/2$.

Let $\theta$ be $\angle BAP$. Then $AP = \cos\theta$ and $PB = \sin\theta$. We thus get:

$$\text{length of } \gamma = \max_\theta \cos\theta + f(n-1)\sin\theta$$

Thus, applying the result of the previous problem:

$$f(n) = \sqrt{1 + (f(n-1))^2}$$

Since $f(1) = 1$ (clearly) we get $f(n) = \sqrt{n}$. $\square$

## 4. Rearrangement and Chebyshev inequality

4.1. **Rearrangement inequality: statement.** Let $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ be two $n$-tuples of real numbers such that $a_1 \geq a_2 \geq \ldots \geq a_n$ and $b_1 \geq b_2 \geq \ldots b_n$. Let $\sigma$ be a permutation of the numbers $1, 2, \ldots, n$. Then:

$$\sum_i a_i b_i \geq \sum_i a_i b_{\sigma(i)}$$

In other words, the sum of pairwise products is maximum if we pair the *largest* with the largest, the second largest with the second largest, and so on.

Equality holds if and only if, for each $i$, $a_i = a_{\sigma(i)}$ or $b_i = b_{\sigma(i)}$.

Further:

$$\sum_i a_i b_{\sigma(i)} \geq \sum_i a_i b_{n+1-i}$$

In other words, the sum of pairwise products is minimum if we pair the largest with the smallest, the second largest with the second smallest, and so on.

4.2. **Idea behind the inequality.** Think of it as a resource allocation problem. For instance, suppose a thief has 3 bags and 3 kinds of coins (gold, silver, copper) to pack in them, and she must pack a different kind of coin in each bag. Assume further that the coins are available in unlimited quantities. Then, in order to maximize her loot, she will put the gold coins in the biggest bag, the silver coins in the second biggest bag, and the copper coins in the third biggest bag.

The idea is: send the most to the best. Such an allocation principle is often called a *greedy* allocation principle.

The Rearrangement inequality is best proved for two elements, and then extended by induction. Let $a_1 \geq a_2$ and $b_1 \geq b_2$. Then we have:

$$(a_1 - a_2)(b_1 - b_2) \geq 0$$

Manipulating this gives us:

$$a_1 b_1 + a_2 b_2 \geq a_1 b_2 + a_2 b_1$$

The rearrangement inequality thus illustrates the general statement the principles of optimization and equality are often at crossroads.

To use this to prove the result globally, we start with the expression $\sum_i a_i b_{\sigma(i)}$ and locate indices $i, j$ for which $i < j$ but $\sigma(i) > \sigma(j)$. We then change the permutation to one sending $i$ to $\sigma(j)$ and $j$ to $\sigma(i)$ (and having the same effect as $\sigma$ on the others). This local change increases the value of the expression and hence it is clearly not the optimum value.

Note here that equality holds only if $a_i = a_j$ or $b_{\sigma(i)} = b_{\sigma(j)}$.

4.3. **An application of rearrangement.** Consider the following problem I had mentioned earlier:

Prove that $a^2 + b^2 + c^2 - (ab + bc + ca) \geq 0$ with equality holding only if $a = b = c$.

This problem can also be solved using the rearrangement inequality. First observe that since the expression is symmetric in $a$, $b$ and $c$, we can assume without loss of generality that $a \geq b \geq c$.

Consider the triple $(a, b, c)$. This is an ordered triple with the property that the elements are in non-increasing order. Then $(b, c, a)$ is a permutation of this expression. Thus, by rearrangement inequality:

$$aa + bb + cc \geq ab + bc + ca$$

Which gives us what we want.

Also note that in this case, equality holds if and only if $a = b = c$.

4.4. **Chebyshev inequality.** Chebyshev inequality says that sending the most to the best is better than giving the average to the average. More formally, if $(a_1, a_2, \ldots, a_n)$ and $(b_1, b_2, \ldots, b_n)$ are two $n$-tuples of decreasing reals:

$$\sum_i a_i b_i \geq \frac{\sum_i a_i \sum_i b_i}{n}$$

Where equality holds iff either all the $a_i$s are equal or all the $b_i$s are equal.

4.5. **Fundamental difference between Chebyshev and Cauchy-Schwarz.** Both the Chebyshev and the Cauchy-Schwarz inequalities are similar in the following sense:

- They are both true for all reals
- They both provide bounds of $\sum_i a_i b_i$

But they are different in the following ways:

- In Chebyshev, it is important to order the $a_i$s and $b_i$s in descending order, whereas Cauchy-Schwarz is applicable for any ordering
- Chebyshev gives a bound in terms of $\sum_i a_i$ and $\sum_i b_i$ while Cauchy-Schwarz gives a bound in terms of the sums of their squares.
- Chebyshev provides a *lower bound* on $\sum_i a_i b_i$ while Cauchy-Schwarz provides an upper bound
- The equality case is different in both. In Chebyshev, equality holds if all the elements in one of the tuples are equal. In Cauchy-Schwarz, equality holds if the two tuples are scalar multiples of one another.

A word of caution, though, when deciding whether to apply Chebyshev or Cauchy-Schwarz. Just because the inequality seems to require a lower bound on $\sum_i F_i G_i$, does not mean that Chebyshev is the one to be used. In fact, we could still use Cauchy-Schwarz by taking $a_i = F_i G_i$ and $b_i$ to be $1/F_i$.

## 5. Nesbitt's inequality

5.1. **Statement of the inequality.**

**Problem 6** (Nesbitt's inequality)**.** For positive $a$, $b$ and $c$, prove that:

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} \geq \frac{3}{2}$$

with equality holding if and only if $a = b = c$.

5.2. **Applying Cauchy-Schwarz (direct application fails).** To apply Cauchy-Schwarz we need to put the terms $\frac{a}{b+c}$ and its analogues on the *left* side, which means we should view each of them as a square. Their squareroots are $\sqrt{\frac{a}{b+c}}$ and its analogues. Thus, one tuple is:

$$\left(\sqrt{\frac{a}{b+c}}, \sqrt{\frac{b}{c+a}}, \sqrt{\frac{c}{a+b}}\right)$$

We would like the other tuple to be something that cancels the denominator. A natural choice is $\left(\sqrt{b+c}, \sqrt{c+a}, \sqrt{a+b}\right)$. Unfortunately, this fails to yield the answer, because the expression that we get is upper-bounded, rather than lower-bounded, in the case of equality.

5.3. **Applying Chebyshev.** Consider the tuples $(a, b, c)$ and $((b+c)^{-1}, (c+a)^{-1}, (a+b)^{-1})$. We first need to determine whether they are arranged in the same order. Assume without loss of generality that $a \geq b \geq c$. Then $b + c \leq c + a \leq a + b$, and taking inverses, we obtain that the second tuple also has its coordinates in descending order.

We are thus in a position to apply Chebyshev's and obtain that the give expression is at least:

$$\frac{(a+b+c)((b+c)^{-1} + (c+a)^{-1} + (a+b)^{-1})}{3}$$

Now using A.M.-H.M. inequality for the quantities $(b+c)$, $(c+a)$ and $(a+b)$, we get the required result.

5.4. **A short proof.** Another way of proving the result is to add and subtract 3, thus writing it as:

$$(a+b+c)\left(\frac{1}{b+c} + \frac{1}{c+a} + \frac{1}{a+b}\right)$$

And now apply the A.M.-H.M. inequality.

## 6. A past IMO problem

### 6.1. The problem statement.

**Problem 7** (IMO 1995)**.** Prove that if $a$, $b$ and $c$ are positive reals such that $abc = 1$, then:

$$\frac{1}{a^3(b+c)} + \frac{1}{b^3(c+a)} + \frac{1}{c^3(a+b)} \geq \frac{3}{2}$$

The first trick is to put $x = 1/a$, $y = 1/b$ and $z = 1/c$. The left-hand side becomes:

$$\frac{x^2}{y+z} + \frac{y^2}{z+x} + \frac{z^2}{x+y}$$

**6.2. Cauchy-Schwarz.** After this point, the first possibility to consider is Cauchy-Schwarz. Since we want to lower-bound the sum here, we must view $\frac{x^2}{y+z}$ and its analogues as squares of a tuple. The other tuple is obtained by cancelling denominators from third tuple. We thus have tuples:

$$\left(\frac{x}{\sqrt{y+z}}, \frac{y}{\sqrt{z+x}}, \frac{z}{\sqrt{x+y}}\right)$$

and

$$\left(\sqrt{y+z}, \sqrt{z+x}, \sqrt{x+y}\right)$$

We apply Cauchy-Schwarz to these tuples, and then use A.M.-G.M. inequality and the fact that $xyz = 1$.

If we keep track of the inequality constraints at each step, we obtain that equality holds if and only if $x = y = z = 1$, and hence $a = b = c = 1$.

# Index