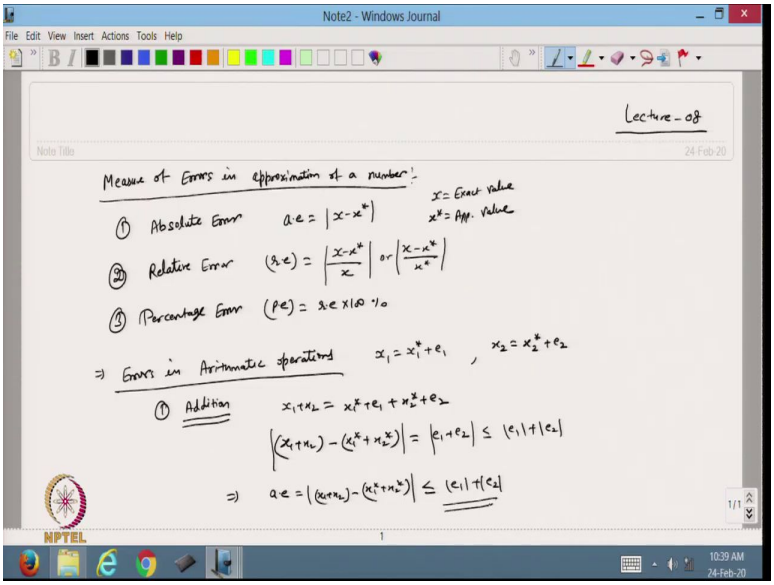
Scientific Computing Using Matlab
Professor Vivek Aggarwal
Professor Mani Mehra
Department of Mathematics
Indian Institute of Technology Delhi
Delhi Technological University
Lecture 8
Errors Arithmetic

Welcome back to this course. So today we will discuss lecture 8. So in the previous lecture, we have discussed about the rounding off error and the chopping off error.
(Refer Slide Time: 0:37)



So to today we will go further and we will discuss the measure of errors in approximation of a number. So they are, basically we generally calculate three types of errors, so the first one is the absolute error. So absolute error is we write is ae, so this is equal to x that is a number and x star we have taken the approximation of that number. So that is the called the absolute error. The second one we will go for relative error. So the relative error means, it gives you the error corresponding to the given number. So this can be written as x minus x star and I am dividing it by either x if the x is there or we can write as or $(x - x^*)/ x^*$.

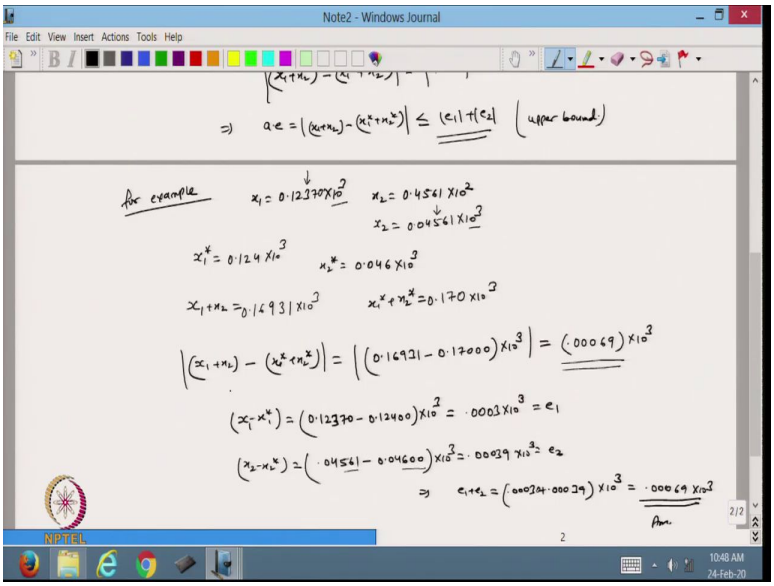
So this is the relative error corresponding to the exact number and this is the corresponding to the approximate number because my x is the exact x, exact value and my x^* is the approximate

value. So this is an approximate value, and the third one is the percentage error. So percentage error we call it P, so this is a relative error multiplied 100. So it gives you how much percentage error is there in the given computation.

So now we will start further and let us do the errors in the arithmetic operations like we have two numbers, like I have a number x_1 and this is approximated by the number x_1^*. And I know that some error is there. So let e_1 is the error involved in the x_1. Similarly, I have another number x_2 that is also we have approximated and that has some error that is called e_2.

So what I do, what will happen so let me take the first one. So when we add these two numbers, so add. Now what will happen if I add this number x_1 and x_2? So let us see this one, so if I add this number $x_1 + x_2$, so this number I can write as $x_1^* + e_1 + x_2^* + e_2$. So this is my number, now $x_1 + x$ minus so I can take this number $x_1^* + x_2^*$ on the left hand side.

On the right hand side, I have only $e_1 + e_2$ left. Now I take the modulus value.
$|e_1+e_2| \leq |e_1| + |e_2|$. So from here I can say that the absolute error in this, in addition, so absolute error that is equal to $x_1 + x_2 \leq |x_1| + |x_2|$.
(Refer Slide Time: 5:06)



So this gives me the, not the exact value but from here you can see that this gives me the upper bound. So that is the upper bound. Upper bound means in this case, if I add two numbers and there are some errors in the numbers already, errors in the number, then the maximum errors can happen in the addition of the number is always less than equal to the modulus of the error in the corresponding numbers. So that gives me the upper bound.

So, for example I take a number $x_1 = 0.1237 \times 10^3$. Another number, I take $x_2 = 0.4561 \times 10^2$. So I have two numbers and let so I will write x_2 as because in this case I should have the same power of x, so I will write 0., so this one can be written as 0.04561×10^3 .

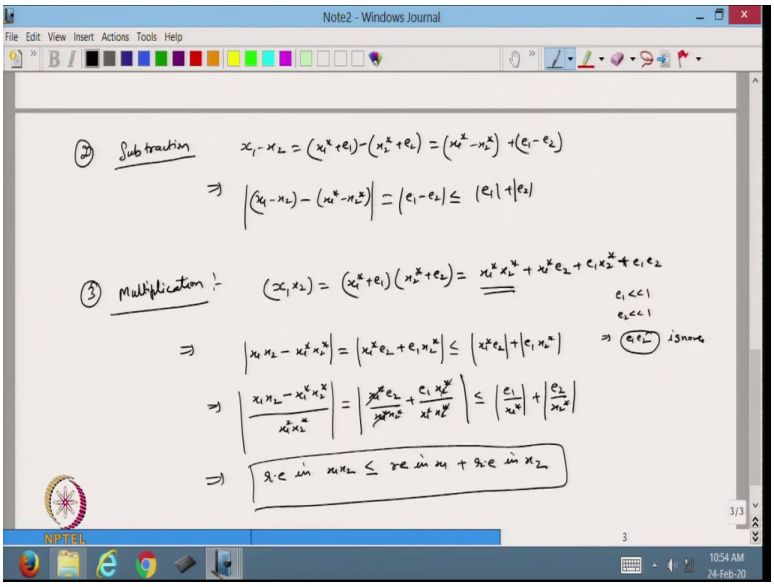
So this one is, I can write this number because I have added one 0 there. Now both the numbers have the same power. Now suppose I take an approximate value of this x_1^* and I suppose I round off this number at third place and similarly, I round off the number at third place. So if I round off the number at third place, I will get 0.124×10^3 .

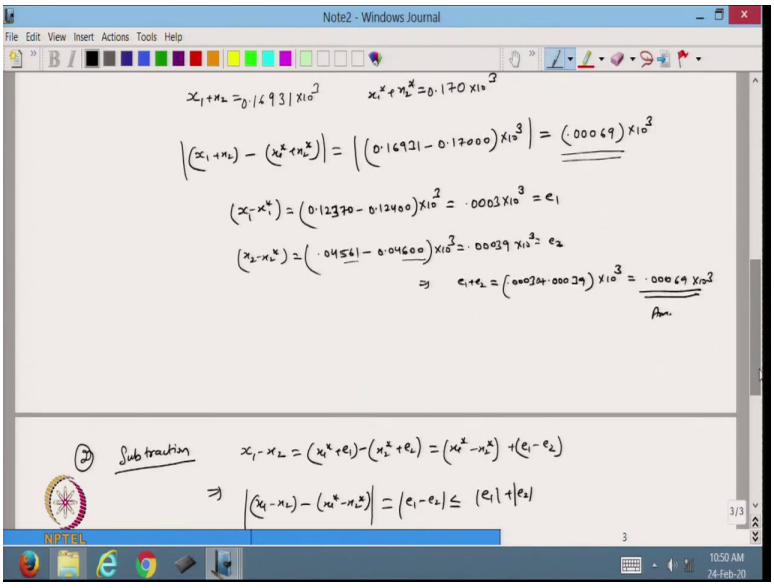
So this is the number we got and my $x_2^* = 0.046 \times 10^3$. So that is the number we got. So in this case suppose I want to find my $x_1 + x_2$. So this will be, so this is the number we got if I add $x_1 + x_2$. Now if I add $x_1^* + x_2^*$, so $x_1^* + x_2^*$ = 0.467 and 1, 10^3 . So this is a number we got from x_1 plus x_2^* .

Now, suppose I want to find $x_1 + x_2 - x_1^* + x_2^*$, so this is the number I want to find. So this will be equal to minus, so this is equal to so it will be so this is a bigger number. So from here I can calculate this value. So this will 0. so 931, so 170 is 931. So this one can be written as 0.00069 into 10^3 . So that is the value we got from the errors when we add these two numbers.

Now I want to see what will happen if I calculate the error in x minus $x_1 - x_1^*$. So $x_1 - x_1^*$, this is the number 12370 - 12400 x 10^3 . So it will be 400 so 370, so it is 0.0003×10^3 . So that is the error we got in the, so this is my e_1 and $x_2 - x_2^*$ so this is my x_2 so 0.4561 - 0.04600 into 10^3. So in this case, it is 600 and 561. So it is 0., I can call it 0.00039 because this 600 minus 561 is 39 and before that is two 0.

So this is the, into 10^3 so that will be my e_2. So from here I can say what is my $e_1 + e_2$. So e_1 is + 0.00039×10^3 . So it is 0.000 three 0, so it is 69×10^3 . So that is the maximum error it can happen when we do the addition. So that is the answer we have, so that is the error that happened in the term of $x_1 + x_2$. So I think this calculation is okay and we can move further. So then this is the addition we have done and we got this error in the addition.
(Refer Slide Time: 11:26)





Now we, next one we go for the subtraction. So suppose we have two numbers, the same number and I want to subtract $x_1 - x_2$. So this one can be written as $x_1^* + e_1 - x_2^* - e_2$, so from here I can write $x_1^* - x_2^* + e_1 - e_2$. So from here, I can write that $x_1 - x_2 - x_1^* - x_2^*$. So in this case I am taking the subtraction of two numbers. So this will be equal to $e_1 - e_2$.

So by the triangle law it can be written as $e_1 - e_2$. So it has the upper bound for the subtraction. Similarly, in the case of additions we have seen. So that is the absolute error in this case. The third one is what will happen if we do the multiplication because if you see in the last in the previous also the maximum error even in two was 0.0069 and if I took the error in this case so that will also come 0.0069.

So it is by chance that these two numbers are coming. But definitely this error cannot be bigger than this one because this is the upper bound we have, we can have. So that is the upper bound we can have in this case. So in the multiplication what you will do? Suppose I want to calculate, what is the $x_1 \ x_2$, so this one I want to calculate and we want to see how the error will propagate if we multiply the two numbers.

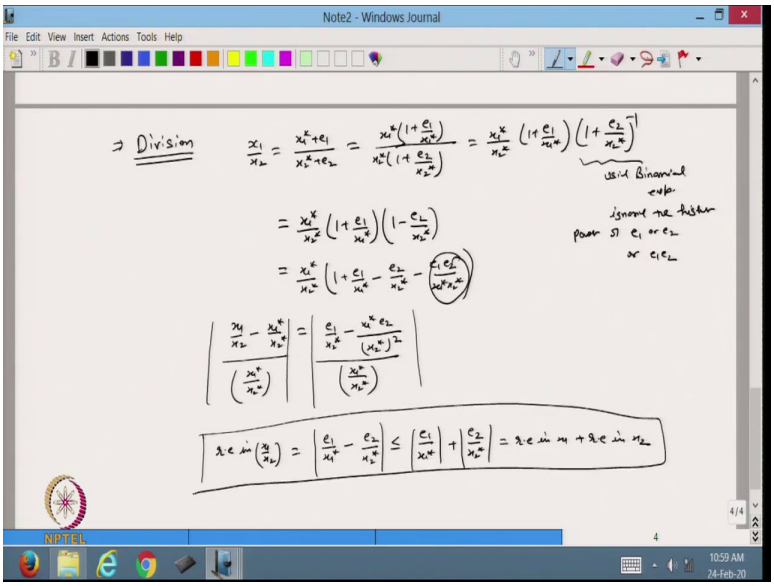
So this is x_1 can be written as $x_1^* + e_1$ $x_2^* + e_2$, so this is the number we have. If I multiply this one so it can be written as $x_1^* \ x_2^* + x_1^* \ e_2 + e_1 \ x_2^* + e_1 + e_2$. So now in this case, we know that the error is always very small, so e_1 is very, very small. e_2 is also very, very small. Then from here I can say that e_1^2 and e_2^2 is very much smaller as compared to e1 and e2.

So this two values are very small and we can ignore this value, so we can ignore this one. So from here I can write that x1 into x2 minus, so this factor I will take on the left hand side. On the right hand side, I will get $x_1^* \ e_2 + e_1 \ x_2^*$, so this we got. Now suppose I take the modulus of this, then it is not going to give any information because the right hand side vector is also dependent on, so I can write like this one.

But in the right hand side vector it is also depending on x_1^* and x_2^* . So we cannot make any conclusion in the terms of the absolute error. So in this case, what you do, I will try to find the relative error. So what I do is I will divide it by the number $x_2 \ x_1^* \ x_2^*$ and then I take the modulus. So this one I can write as $x_1^* \ e_2 \ / \ (x_1^* \ x_2^* + e_1 \ x_2^* \ x_1^* + x_2^*)$, not plus it is multiplication.

So that is equal to $x_1^* \ x_2^*$. So this will cancel out. So from here I can write that this should be equal to $e_1 \ / \ x^*_1 + e_2 \ / \ x^*_2$ and that is what is this? It is the relative error in x1 and this is the relative error in x2. So from here I can say that the relative error in x1 x2 is less than equal to relative error in x1 plus relative error in x2. So that is we can have the upper bound in the relative error when we multiply the two numbers.

(Refer Slide Time: 16:38)



So this is what we have done for the multiplication and the last one we can do the division. So what will happen in the division? Now let me divide the number x1 by x2. So in this case what we will do? This can be written as plus e1 divided by x2 star plus e2. Now I can take the number x1 common from the numerator and x2 star from the denominator, so this will become one plus e1 by x1 star and this one can be written as one plus e2 by x2 star.

So this can be written as, now I can write as a x1 star x2 star. And this one, I can write as x1 star 1 plus e2 x2 star. So this one number I can take and now this number is this is a binomial so I can expand this one using binomial expansion. But I will ignore the higher power of, so ignoring the higher power of e1 or e2 or e1 into e2.

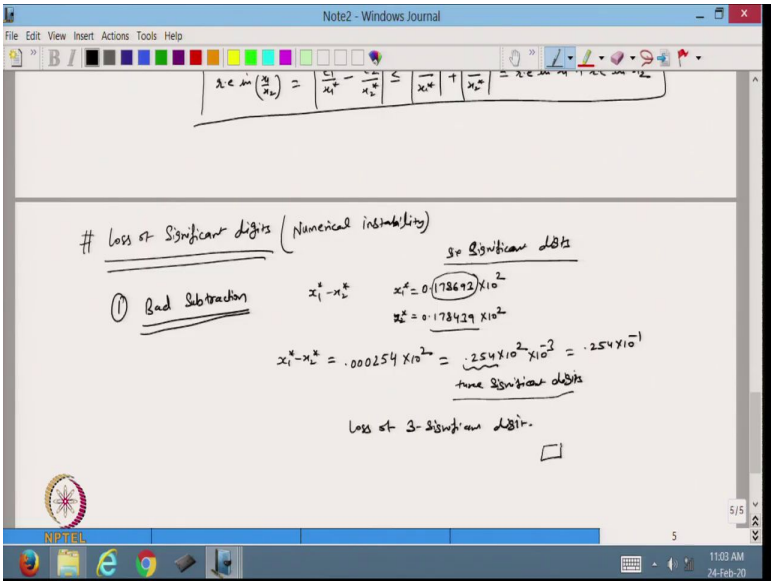
So like this one I can ignore. So from here I can write x1 star x2 star, so 1 plus e1 by x1 star 1 minus e2 by x2 star and all other term will carry the higher power so that we can ignore and then we can do the further calculation. So this will, it will be 1 plus e1 by x1 star minus e2 by x2 star and then minus of e1 e2 x1 star x2 star. So this one I can ignore because it gives the value of e1 and e2, which become a very small quantity.

So from here I can write my x1 over x2, so I will take this quantity on the left hand side. So this will become x1 star over x2 star. On the right hand side, I will get this x1 star will cancel out, so I will get e1 over x2 star minus and here I will get x1 star e2 divided by x2 star whole square. So in this case also if I want to find the absolute error it is not going to give any upper bound because it is dependent on this one.

So in this case what I will do, I want to find the relative error. So I will divide this by x1 star

over x2 star, so I will divide this number like this, x1 star over x2 star. So this number will give you, so that is and then I will take the absolute value. So this on the left hand side gives the relative error in x1 by x2. On the right hand side, if you see this we will divide by this, so x2 will cancel out. It will give you e1 over x1 star from here minus, so this one x_1 will cancel out, so it will give you only e_2 / x_2^*.

So that is the number we got in this case. So this is 1 plus, so this is the 1 I have taken on the left side and this. So that can be written as e_1 over x_1^* + e_2 over x_2^*. So this can be written as relative error in x_1 plus relative error in x_2. So by using this one we can define the relative error in x_1 / x_2 is always less than equal to the relative error in x_1 plus relative error in x_2. So in this way we can find the upper bounds on the relative error when we divide the number x_1 by x_2. So these numbers we are always given the upper bound. The next thing we can give is how the calculation is when we do the calculation how the number of significant digits can be lost.
(Refer Slide Time: 21:47)



So this is the, I will give you the one example of loss of a significant digit. So that also comes as a numerical instability. So what is the numerical instability, because sometimes we do the calculation, we know that at each time we start doing the computation, we are multiplying the two numbers, subtracting the two numbers, dividing the two numbers. So in this case, whatever the, at the starting we introduce some rounding off error or the chopping off error so that errors will grow as we do the computation.

So it depends that sometimes it will happen that then my results will become unbounded or it tends to diverge. So in that case, we say that our computation is numerically unstable. Otherwise, if it is giving the correct answer, then it will go, we call it the numerical stable. So how will this

happen? So this is the one form I can say that first one is what we call it bad subtraction.

So what is the bad subtraction? Suppose I have a number one number is x_1 approximate number and another number is x_2, only thing is that x_1 and x_2 is of the same magnitude, then sometime it will happen if I subtract one number from the another number and both the number are of the same magnitude, then what will happen? We lose some significant digits. So in this case, it is called the bad subtraction.

For example, suppose I take a number x_1 star as some number I am taking. It is 178693 x 10 raise to power. Maybe 2 suppose I write and another is y_2^* or x_2^* , that is another number, and that is 0 equal to 178439 x 10^2 . So suppose these are the two numbers and if you see that this number has the same magnitude, it is 1 0.1786, it is 0.1784 and the same power 10 raises to power this.

So these two numbers are of the same magnitude and suppose I want to do subtraction. So if I do the subtraction in this case, what I will get, I will subtract from here. So it will be 452 and then 0000 into 10 raise to power 2. So in this case, if you see from here that my number in these numbers, I have a (sig), six significant digits. In this also we have six significant digits.

But in this case, I can write this number as 0.254 x 10^2 was already there. And I will write it 10^{-3}. So in this case, what will happen, I will lose my significant digit. So now this number has three significant digits. So earlier I started with the six significant digits, now I deal with the number that has three significant digits.

So this number is 0.254 x 10^{-1} . So from here, I can say that I, this is the loss of 3 significant digits. So this is all about that, how the error grows or how we can find out the upper bounds of the error when we do the subtraction, addition, multiplication and division of the two numbers. So in this we have discussed how we can also lose, in the computation how a significant digit can be lost. So this is all about in this lecture. Thanks very much. In the next lecture we will go further. Thanks very much.