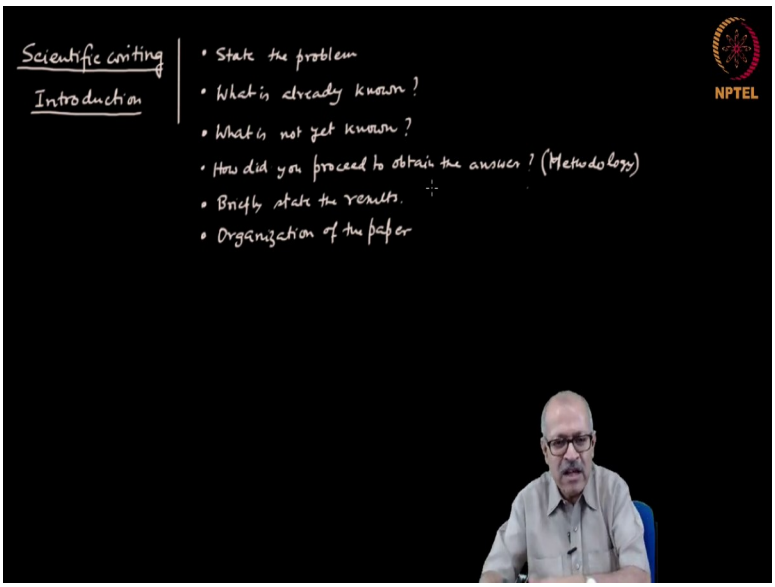
Research Methodology
Prof. Soumitro Banerjee
Department of Physical Sciences
Indian Institute of Science Education and Research, Kolkata

Lecture - 59
Scientific Writing: Journal Papers Part 3

(Refer Slide Time: 00:20)



We are dealing with scientific writing. The purpose clearly is to make the ideas that we have generated reach the reader. So, that is the basic idea: the results that you have got, the experiments that you have performed, should very clearly reach the readers, so that they can use it. In the writing of a paper, as I said, the title is the most important thing. The abstract is the next most important thing, and then comes the importance of the Introduction.

Now we come to the actual body of the paper, in which the first section is the Introduction. In the Introduction, the reader would expect to find answers to the following things, and therefore, the writer has to put these in the introduction in an appropriate way.

In the introduction you have to state the problem that is being tackled in that paper. State the problem right up front. That means, the first paragraph. In the first paragraph, the reader will expect to see what the problem is all about. Therefore, the problem has to be stated at least in a preliminary form in the first paragraph. In more refined form

somewhere down the introduction section, but at least in the first paragraph the reader should find what it is all about, the problem that is being attacked.

After you have stated ‘this is what I am trying to find out’, you have to tell the reader what is already known. What is already known on this issue, or the background material. That is contained in what is known as literature survey. Now, the writing the literature survey again is a bit difficult because you have to decide up front which audience you are addressing, and depending on that you will have to go to different extents of detailing.

If you are writing for a knowledgeable audience (knowledgeable means knowledgeable in that particular area of research) then you need not talk about the basics of that area. But if it is meant for a general audience, you have to talk about the basics of that area. So, where you start the literature survey depends on the what kind of audience you are expecting.

If you are talking, for example, on the phenomenon of superconductivity, you might start with the first observation of superconductivity early in the last century, or you can talk about the relatively later developments on superconductivity, especially the theory of superconductivity which happened in the later part of the last century. So, depending on the audience you might start.

But the point is that, when you end the literature survey, the reader should have a very clear idea about what is already known. Then you have to state what is not yet known. I will come to that. But when you are talking about what is known, you have to view it as a narrative, as a story line, so that it is comfortable to read. The narration proceeds in a way that follows the development of ideas.

Somebody proposed a hypothesis, somebody else tested the hypothesis and found it to be incorrect. Another hypothesis was proposed by somebody, somebody tested it and found that he or she cannot eliminate the null hypothesis. The hypothesis was a straight line relationship, so on and so forth. You view it as a narrative and at each point you have to cite who has done which work.

How to cite, the stylistics of the citation, I will come to that later. But the literature survey is a prime place where you refer to the earlier work, you cite the earlier work.

Now, citing the earlier work and talking about what is already known is not optional. It is mandatory. In any paper, even in the briefest possible papers, you have to talk about what is already known, and then prepare the ground for stating what is not known.

Therefore, on the basis of the literature survey, you have to establish what is not yet known. So, after having told what is known, then you have to tell what is not yet known, and that is my subject of investigation in this paper. Thus, you establish the motivation: this is known, on that basis I can see that this is unknown, and therefore, this is my subject matter.

Now, when you state what is unknown, as you know all the research is essentially trying to answer some question. It is a question, and therefore, when you state what is yet unknown, you have to state it in the form of a question whose answer is yet unknown. A question means something that starts with which, when, where, how, and ends with a question mark.

So, it has to be a question. The most common error is to make statements rather than questions, like, ‘in this paper we study the property of ...’. This is a statement. You are not stating which question’s answer is yet unknown. When you are saying that ‘in this paper we studied ...’ (something) then some part of that is known, and on that basis, some part is unknown. You have to state very clearly what is yet unknown.

That needs to be stated in the form of questions: ‘this question’s answer is not yet known’. And then you have to state how did you proceed to obtain the answer, your methodology. You may have performed an experiment, you may have performed some simulation studies, you may have obtained a derivation, you may have proved a theorem. So, what did you do in order to obtain the answer to that question? What method did you follow? That needs to be very clearly stated. This is actually a question of methodology and since there are various techniques and methods available in science: experiment, theory building and all that, you have to state what you did.

Then briefly state the results. Do not keep stating the results for the Conclusion section. There will be a section, the conclusions out of the paper. But the reason we include a brief statement about the principal outcome of the paper—the result—is that, by reading the Introduction, the reader clearly knows what to expect in the paper.

‘In this paper we show that ...’ kind of language we use. That means, the reader immediately expects that somewhere down the line he would be convinced that this result is true and therefore, he would check whether the method followed actually derive that result or not, and so on so forth. So, it is necessary to state briefly the results, not in detail, but briefly the principal results.

And then you have to state the organization of the paper. The organization of the paper means how you have broken it up into sections, each section dealing with a specific aspect: some section on deriving the model, the simulation, the results. So, each would be a different section. How you have sectionalized the paper -- that has to be stated. So, its language would be that ‘in section 2 we introduce the system model, in section 3 we simulate the system model using certain parameter values, in section 4 we present the results’, so on and so forth, which means that after reading that, the reader knows where to expect what. This is necessary.

If you notice your own reading style, you should expect the readers’ reading style to be the same. What is your reading style? What is everybody’s reading style? We jump around. After having read the Introduction, sometimes we simply go to the conclusion and find out what is the conclusion out of that and then maybe we go to the relevant sections that are of our interest. Rarely we read from the beginning to the end in one go. We jump around. We will look at some pictures: this figure is of interest to me, so I will come back to that. If you want to jump around, you should know how to do that. The organization of the paper: that paragraph helps in doing that.

So, these are the things to be included in the introduction. Many people write the Introduction at the end, after having composed the rest of the paper, they start to write the Introduction. I normally do not write that way, because, for me, it is sort of a wholesome story that starts with the Introduction. Therefore, the development of the thought, development of the idea that starts from the Introduction and goes forward, there is a continuity. That might be broken if the Introduction is written at the end. So, I personally prefer to write in one go after having formed a complete idea about what is to be presented in the paper.

But different people have different techniques of doing that. I am not really disregarding the other possibility, but it is just my choice.

After you have put down the Introduction, comes the middle part of the paper. Middle part of the paper means the other sections that are there in the paper. In most papers, we put the literature survey inside the introduction. Right after introducing the problem to be attacked, we state what is already known and therefore, it comes there.

But in some rare situations this literature survey part may be a bit longish. For practical reasons you might need to tell the reader what is known and that might be a bit longish. In that case some people separate the literature survey as a separate section called 'Background' and they present that way, but that is relatively rarer. Normally the Introduction contains the literature survey and after that we need to present the hypothesis and the methodology of testing the hypothesis.

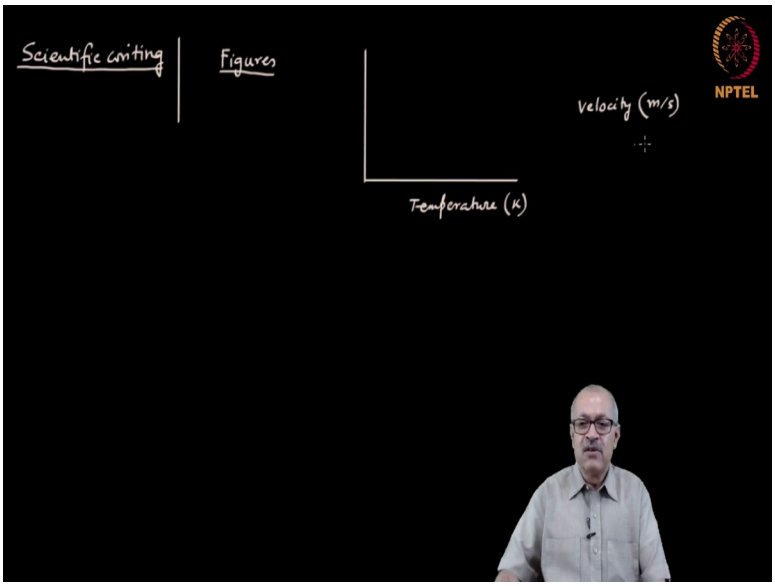
Parameters used and results obtained -- all these would come one after the other in the middle part of the paper. Now, while writing the middle part of the paper, you have to keep in mind the reading habit of the reader: as I said, readers normally jump around. After reading the Introduction or some people go directly to the conclusion and find out what the conclusions are and then come back to the relevant portions. And in the relevant portion, the important things are the figures, tables, charts, equations.

So, these are the landing grounds. That means, after having jumped around, the reader normally looks at what is interesting. And then if he or she is convinced that it is worth reading the whole paper, then he or she will read the whole paper.

The writer has to keep this in mind that the reader might jump around. So, there has to be enough landing grounds that the writer directs the reader to jump around in the places. Where the writer wants them to jump around, where the writer wants the reader to land, say, a particular figure is central to this paper; that means, it conveys the basic idea. So, the organization should be such that while the reader jumps around, he or she will inevitably land on that figure.

The writing should be such: in this part you have to state what the mathematical model is, you have to state the assumptions, state the parameter values. All that has to go into this middle part of the paper. But in the middle part of the paper, the most important thing are the figures because a figure can convey in very precise and accurate way what 10 paragraphs cannot. Much more is conveyed in just one figure if the figure is properly thought of and composed.

(Refer Slide Time: 17:55)



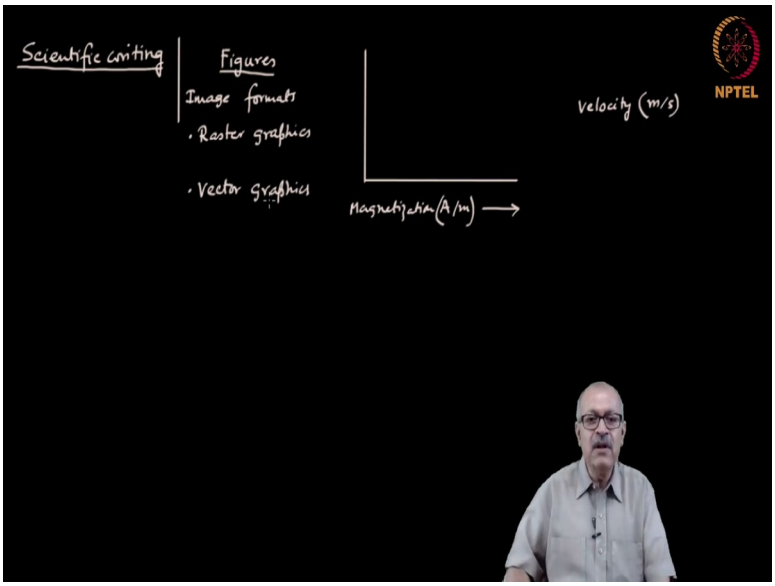
So, figures are most important things in a paper. I am thinking of scientific papers in which there would normally be figures. Therefore, I will spend some time in discussing figures because many people make mistakes in producing figures.

Firstly, if any data, primary data, or simulation results of a theoretical development can be presented as a graph, always do that, because that conveys the idea to the reader in a fraction of a second what 5 paragraphs would not. So, it is important to make pictures whenever possible. A paper should be dominated by figures rather than by text, unless the content of the paper is such that there cannot be any figures, there can be only text.

Now some things: possibly known, but people do mistakes. That is why I have to say: whenever there is a graph there are axes and the axes have to be marked. I have seen that many times people don't do it right.

For example, here temperature, you have to also state the unit in which it is given. Like this: within bracket the unit should be there. For example, velocity, within bracket meters per second. So, this is the right way of labelling the axes.

(Refer Slide Time: 20:26)



If you want to write, say, magnetization, you should not simply say amperes per meter. No, it is ‘Magnetization’ within bracket amperes per meter. That is the right way of labelling the graph.

Another point is that these are written in the letters, and after you have produced the figure it will be resized when it is put into the paper. And when it is resized, mostly it is shortened, the dimensions reduce, and as a result of which, these letters become so small that they cannot be read. So, the rule of thumb is that these letters should be of such a font such size, so that after reduction, their size is one size smaller than the size in the text. Normally we use, in the caption, one size smaller font and this should be of the same size after reproduction, that means, after reduction.

So, these should be produced accordingly. Remember, normally we produce the graphs using some graph plotting program like gnuplot, like Origin, like MATLAB, Mathematica or something like that. They normally would not produce these in sufficiently large sizes. So, you have to do something to tell the graph plotting program to write the axis markings in sufficiently large font size. Every program has that facility and you have to use that.

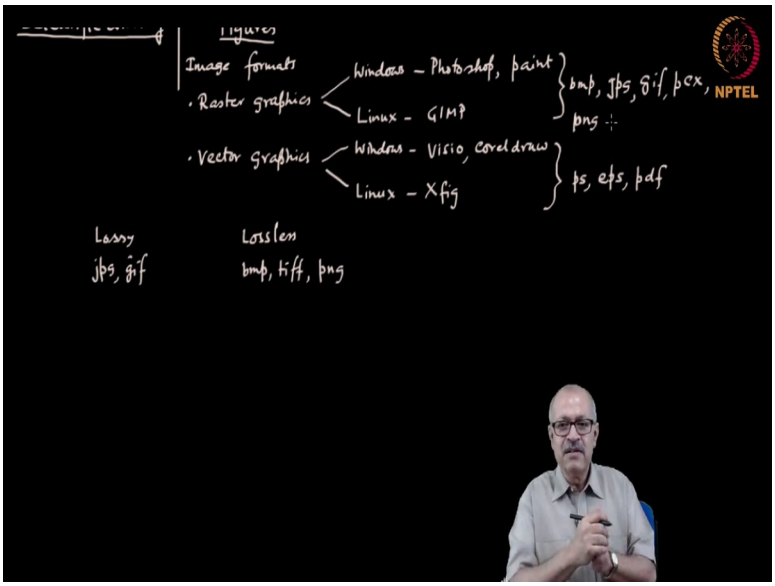
Just plotting the graph, whatever you get you put in the paper -- that is not the right way. So, these have to be sufficiently big size. Then, in the figures you are putting some

images, and images come in different formats. If you talk about image formats, there are basically two; one is called the raster graphics and the other is called vector graphics.

A raster graphics image is essentially a collection of dots. So, when you enlarge it, it becomes blocky while a vector graphic image is essentially where the locations are joined by certain curves, as a result of which, if you enlarge, it does not break, it does not become blocky. That is why publishers normally prefer vector graphics images. And programs like MATLAB, like mathematica, and things like that, they can produce vector graphics images.

So, wherever possible you should try to produce vector graphics images. Now what are the raster graphics formats? In windows as well as in LINUX and MAC, you have both these formats available in different programs.

(Refer Slide Time: 24:51)



For example, in Windows, you have Photoshop, you have the smaller program Paint -- these are raster graphics programs. In Linux, you have GIMP, Gnu image Manipulation Program, and there are various other programs like that. And in vector graphics , in Windows you have Visio, Coreldraw, etc. In Linux, Xfig, gnuplot, etc. What I am writing are definitely not exhaustive and I am just giving examples. The usual graph plotting programs can export in both these formats: both raster graphics as well as vector graphics.

The raster graphics formats are bmp, jpg, gif, pcx, png and so on and so forth. These are all raster graphics formats. Vector graphics formats are ps (postscript), eps (encapsulated postscript) and pdf. These are normal vector graphics formats. But these can also take raster graphics images. So, these are actually usable by both techniques.

Now, there are some raster graphics formats which are lossy, and there are some which are lossless. For example, lossless includes bmp, tiff, png and so on and so forth. Lossy includes say jpg, gif, etcetera.

Lossiness depends on the fidelity that you set while saving an image. You can set the fidelity in, say, a jpg image. Accordingly it will incur some losses. If you set it 100 percent, then there is no loss. Otherwise there will be some losses. In general the losses are such that on screen they are not very apparent, but if you want to print in a very high quality mode, then the losses would be visible.

So, if you are using only raster graphics images, then normally it might be a bit blocky. The way to avoid that is to create the figure in a relatively large size, so that, when reduced in size, the blockiness goes away. So, create figures in relatively larger size.

But if you do it in a relatively large size and plan to reduce in the final reproduction, then lines will become thin. Thin lines do not show well and therefore, you have to create the picture with relatively thicker lines, so that even when they are shrunk, they are compressed, the line widths are comfortable to the eye.