Possible Organization for Writing a Thesis including a LaTeXFramework and Examples

by

A Graduate Advisor
B.Sc., University of WhoKnowsWhere, 2053
M.Sc., University of AnotherOne, 2054

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Whichever

Possible Organization for Writing a Thesis including a LaTeXFramework and Examples

by

A Graduate Advisor
B.Sc., University of WhoKnowsWhere, 2053
M.Sc., University of AnotherOne, 2054

Supervisory Committee

Dr. R. Supervisor Main, Supervisor
(Department of Same As Candidate)

Dr. M. Member One, Departmental Member
(Department of Same As Candidate)

Dr. Member Two, Departmental Member
(Department of Same As Candidate)

Dr. Outside Member, Outside Member
(Department of Not Same As Candidate)

Supervisory Committee

Dr. R. Supervisor Main, Supervisor
(Department of Same As Candidate)

Dr. M. Member One, Departmental Member
(Department of Same As Candidate)

Dr. Member Two, Departmental Member
(Department of Same As Candidate)

Dr. Outside Member, Outside Member
(Department of Not Same As Candidate)

ABSTRACT

This document is a possible Latex framework for a thesis or dissertation at UVic. It should work in the Windows, Mac and Unix environments. The content is based on the experience of one supervisor and graduate advisor. It explains the organization that can help write a thesis, especially in a scientific environment where the research contains experimental results as well. There is no claim that this is the *best* or *only* way to structure such a document. Yet in the majority of cases it serves extremely well as a sound basis which can be customized according to the requirements of the members of the supervisory committee and the topic of research. Additionally some examples on using LaTeXare included as a bonus for beginners.

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

my cat, Star Trek, and the weather, for supporting me in the low moments.

Supervisor Main, for mentoring, support, encouragement, and patience.

Grant Organization Name, for funding me with a Scholarship.

I believe I know the only cure, which is to make one's centre of life inside of one's self, not selfishly or excludingly, but with a kind of unassailable serenity-to decorate one's inner house so richly that one is content there, glad to welcome any one who wants to come and stay, but happy all the same in the hours when one is inevitably alone.

Edith Wharton

DEDICATION

Just hoping this is useful!

Chapter 1

Introduction

1.1 Structure of molecules adsorbed to surfaces

Any pictures?

1.2 Experimental probes: IR, Raman, SFG

(Give a brief intro. for people from CS)

1.3 Current approaches to structure elucidation

(Lots reference can go here, previous work)

1.4 Linear programming

(Explain it to Chemistry Department) LP formula: objective function, constraint

complex algorithm

Three kinds of Soluction

Why we select LP as the tool to tackle this problem? The advantage of LP for this problem. Is there any previous work related?

Introduce the simplex method, the toolkit. Linear programming can deal with thousands of variables.

Linear programming is a convex, deterministic process, and it is guaranteed to converge to a single global optimal if there is a solution space. Linear Programming problems are intrinsically easier to solve than general nonlinear problems.

For a linear programming problem, it is by nature that there are three kinds of situations may happen.

First, there is no feasible solution.

Second, there are feasible solutions, and the solutions space is bounded in a polyhedron, and only one globally optimal solution(either a single point or multiple equivalent points along a line) exists. Furthermore, in this unique and convex feasible region, the optimal solution is at one of the vertex of this convex polygon.

Third, the solution space is unbounded, there are multiple solutions, however, no optimal solution for the existing problem.

Because of this property, it is guaranteed that if linear programming solver returns a solution, then this solution is the global optimal one for the problem.

Simplex method: Introduce Linear Programming solver:

Compare linear programming with quadratic programming, why linear programming is a better approach to the problem? (Having problems finding related work or how to prove it myself)

Is there only computational gain? Also consider the model itself and solution space (The problem is defined as "Candidate ratio problem" in Kai's thesis, same here???) to determine the level of similarity between spectra is not an easy task

How should I introduce Linear Programming here? The advantages of linear programming are:

1.5 Aims and scope

What's LP's constraint?

Chapter 2

Methods

2.1 Generating model spectra

In order to procedure our problem, we need to generate candidates data.

2. Define candidates, How do I generate my own candidates?

Draw spectrum graphs for IR, Raman and SFG. What data do I need?

2.1.1 IR absorption spectra

describe the function used to generate these spectroscopy? pictures

2.1.2 Raman scattering spectra

2.1.3 Sum-frequency spectra

2.2 Surface orientation distribution functions

2.3 Linear programming implementation

LP Model

2.4 Scoring methods

Chapter 3

Simplified Molecular Model

3.1 Description

In the last chapter, we have seen how IR, Raman and SFG spectres are produced theoretically. However, before we jump to the real molecule models, we would like to study the properties of our LP model first. We have encountered cases when LP model failed to give us the expected composition, and this is actually the most case. We want to know why and when our LP model would fail?

Therefore, we constructed a toy molecule model to study our LP model first, meanwhile, get a better understanding of the spectroscopy techniques.

For this toy model, we use the following equation to generate the IR spectroscopy,

$$f_\theta(x) = \sum_{q=1}^{4} A_q^2 * cos^2(\theta - \theta_q) \frac{\gamma^2}{(x - w_q)^2 + \gamma^2}$$

$$A : \ amplitude$$

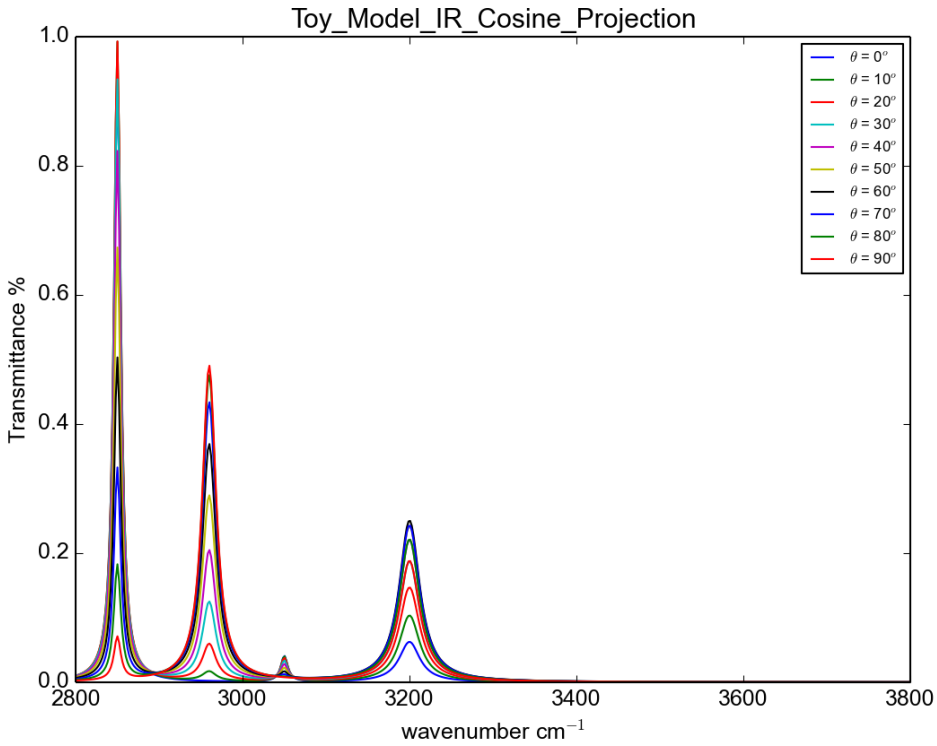
$$\theta_q : angle$$

$$\gamma : width$$

$$w_q : frequency$$

This is the cosine projection of the IR spectra. Furthermore, this molecule only contains 4 vibration modes, with the peaks happen at frequency of 2850, 2960, 3050 and 3200, and the widths for the peaks are 5, 10, 5 and 15, and the heights, which are 1, 0.7, -0.2 and 0.5 respectively. The comparing angle for each peak(What does those angle refer???) is 15, 90, 0 and 60. We generate 10 candidate spectra with 10

different θ values: $\theta = 0°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 90°$



As you can see from the graph that, these 10 candidates have peaks at the same frequency. And the only difference on θ makes them very similar to each other. This also makes it really difficult to obtain exact composition for the targeted spectrum, as various combinations of candidates are possible to achieve the targeted spectrum.

As an example, we compose a target spectrum by combining 15 percent θ of 20 degree's spectrum and 85 percent of θ of 70 degree's spectrum: $0.15 * f_{20}(x) + 0.85 * f_{70}(x)$.

3.2 Linear Programming Model for Spectra Study

The linear programming model that we construct to check if the optimal solution returned actually match the known composition is as following. The model has used in the studies of [?] and [?].

$$minimize \sum_{p=1}^{points} \left| Target - \sum_{c=1}^{candidates} p_c f_\theta(x) \right|$$

$$p_c : percentage\ of\ each\ candidate$$

$$p : number\ of\ points\ selected\ both\ from\ candidates\ and\ Target$$

In this model, the unknown percentage of each candidate is the decision variable, it is denoted by p_c. We select data points along the wavelength frequency,the data point is denoted by p here. Based on each data point, we calculate the absolute residual between target spectrum and the one composed by the decision variables. Then the objective function is to minimize of the sum of the absolute residual over all the data points.

However, in order to actually apply Linear Programming, we need to get rid of the absolute sign in the objective function. We achieve this goal by introducing one more variable X and two more constraints for each data point. Therefore:

For each point:

$$X = \left| Target - \sum_{c=1}^{candidates} p_c f_\theta(x) \right|$$

$$X \geq Target - \sum_{c=1}^{candidates} p_c f_\theta(x)$$

$$X \geq -Target + \sum_{c=1}^{candidates} p_c f_\theta(x)$$

We then convert the previous model into one that can actually be solved by Linear Programming solver:

Experiment index	1	2
Number of Candidates	4	4
Candidates	[0, 10, 20, 30]	[0, 5, 10, 15]
Target Composition	[0.1, 0.5, 0.4, 0]	[0.1, 0.5, 0.4, 0]
Number of Constraints	100(cos)	100(cos)
Return Composition	[0.1, 0.5, 0.4, 0]	[0, 0.796962, 0.103038, 0.1]

$$minimize \sum_{p=1}^{points} X_p$$

$$X_1 - Target_1 + \sum_{c=1}^{candidates} p_c f_\theta(x_1) \geq 0$$

$$X_1 + Target_1 - \sum_{c=1}^{candidates} p_c f_\theta(x_1) \geq 0$$

$$\cdots$$

$$X_n - Target_n + \sum_{c=1}^{candidates} p_c f_\theta(x_n) \geq 0$$

$$X_n + Target_n - \sum_{c=1}^{candidates} p_c f_\theta(x_n) \geq 0$$

$$\sum_{c=1}^{candidates} p_c = 1$$

The last equality restrict the sum of the percentage to 1.

3.3 Experiments

3.3.1 Experiment 1 and 2 Setting

3.3.2 Experiment 2 Result Plotting

The spectroscopy of result composition is almost identical with the target one. Therefore, the LP solver has achieved its optimal.

Experiment index	3
Number of Candidates	10
Candidates	[0, 10, 20, 30, 40, 50, 60, 70, 80, 90]
Target Composition	[0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]
Number of Constraints	100(cos)
Return Composition	[0, 0, 0.730541, 0, 0.212061,0, 0, 0.0573978, 0, 0]

Experiment index	4
Number of Candidates	4
Candidates	[0, 5, 10, 15]
Target Composition	[0.1, 0.5, 0.4, 0]
Number of Constraints	100(cos) + 100(sin)
Return Composition	[0.1, 0.5, 0.4, 0]

Experiment index	5
Number of Candidates	10
Candidates	[0, 10, 20, 30, 40, 50, 60, 70, 80, 90]
Target Composition	[0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]
Number of Constraints	100(cos) + 100(sin)
Return Composition	[0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]

3.3.3 Experiment 3 Setting

3.3.4 Experiment 3 Result Plotting

The spectroscopy of result composition is almost identical with the target one. Therefore, the LP solver has achieved its optimal.

3.3.5 Experiment 4 Setting

3.3.6 Experiment 5 Setting

3.3.7 Constraint Study Based on Experiment 4

3.3.8 Constraint Study Based on Experiment 4

Result:[0, 0.796962, 0.103038, 0.1]
Target: [0.1, 0.5, 0.4, 0]

3.3.9 Constraint Study Based on Experiment 5

3.3.10 Constraint Study Based on Experiment 5

Result:[0, 0, 0.730541, 0, 0.212061,0, 0, 0.0573978, 0, 0]
Target: [0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]

Points	Points Selection	Result
10	[2800, 3300, 50]	[0, 0.796962, 0.103038, 0.1]
20	[2800, 3300, 25]	[0, 0.796962, 0.103038, 0.1]
25	[2800, 3300, 20]	[0, 0.796962, 0.103038, 0.1]
32	[2800, 3300, 15]	[0, 0.796962, 0.103038, 0.1]
50	[2800, 3300, 10]	[0, 0.796962, 0.103038, 0.1]
100	[2800, 3300, 5]	[0, 0.796962, 0.103038, 0.1]
$100 + 1$	[2800, 3300, 5], [2800, 3300, 500]	[0, 0.796962, 0.103038, 0.1]
$100 + 5$	[2800, 3300, 20], [2800, 3300, 100]	[0, 0.796962, 0.103038, 0.1]
$100 + 10$	[2800, 3300, 20], [2800, 3300, 50]	[0, 0.796962, 0.103038, 0.1]
$100 + 50$	[2800, 3300, 20], [2800, 3300, 10]	[0.1, 0.5, 0.4, 0]
$100 + 100$	[2800, 3300, 20], [2800, 3300, 5]	[0.1, 0.5, 0.4, 0]

Points	Point Selection	Result
10	[2800, 3300, 50]	[0.156758, 0, 0, 0.825977, 0, 0, 0, 0, 0, 0.017265]
25	[2800, 3300, 20]	[0, 0, 0.730541, 0, 0.212061, 0, 0, 0.0573978, 0, 0, 0]
50	[2800, 3300, 10]	[0, 0, 0.730541, 0, 0.212061, 0, 0, 0.0573978, 0, 0, 0]
100	[2800, 3300, 5]	[0, 0, 0.730541, 0, 0.212061, 0, 0, 0.0573978, 0, 0, 0]
500	[2800, 3300, 5]	[0, 0, 0.730541, 0, 0.212061, 0, 0, 0.0573978, 0, 0, 0]
$100 + 1$	[2800, 3300, 5], [2800, 3300, 500]	[0, 0, 0.730541, 0, 0.212061, 0, 0, 0.0573978, 0, 0, 0]
$100 + 10$	[2800, 3300, 5], [2800, 3300, 50]	[0.361587, 0, 0.312061, 0.326352, 0, 0, 0, 0, 0]
$100 + 20$	[2800, 3300, 5], [2800, 3300, 25]	[0.174023, 0, 0, 0.791447, 0, 0, 0.0345301, 0, 0, 0]
$100 + 25$	[2800, 3300, 20], [2800, 3300, 20]	[0.174023, 0, 0, 0.791447, 0, 0, 0.0345301, 0, 0, 0]
$100 + 50$	[2800, 3300, 5], [2800, 3300, 10]	[0, 0, 0.753209, 0, 0.146791, 0, 0.1, 0, 0, 0]
$100 + 84$	[2800, 3300, 5], [2800, 3300, 6]	[0.174023, 0, 0, 0.791447, 0, 0, 0.0345301, 0, 0, 0]
$100 + 100$	[2800, 3300, 5], [2800, 3300, 5]	[0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]

Experiment index	6	
Number of Candidates	4	
Candidates	[0, 10, 20, 30]	
Target Composition	[0.1, 0.5, 0.3, 0.1]	
Points	Point Selection	Result
50	[2800, 3300, 10]	[0.2, 0.212061, 0.587939, 0]
100	[2800, 3300, 5]	[0.2, 0.212061, 0.587939, 0]
$100 + 50$	[2800, 3300, 5], [2800, 3300, 10]	[0.787939, 0.0120615, 0.2, 0]
$100 + 100$	[2800, 3300, 5], [2800, 3300, 5]	[0.787939, 0.0120615, 0.2, 0]

Experiment index	7	
Number of Candidates	10	
Candidates	[0, 10, 20, 30, 40, 50, 60, 70, 80, 90]	
Target Composition	[0.1, 0, 0.4, 0, 0.4, 0.1, 0, 0, 0, 0]	
Points	Point Selection	Result
100	[2800, 3300, 5]	[0, 0, 0.731043, 0.149797, 0, 0, 0.119159, 0, 0, 0]
$100 + 100$	[2800, 3300, 5], [2800, 3300, 5]	[0.156072, 0, 0, 0.831607, 0, 0, 0, 0, 0, 0.0123217]

Idea: We can obtain different solutions by have different constraints, as long as the residual is smaller than the threshold we have setted up.

3.3.11 Experiment 6 Setting

Case with 4 candidates when result composition does not match to target one:

3.3.12 Experiment 7 Setting

Case with 10 candidates when result composition does not match to target one:

3.3.13 Experiment 8 Setting

Case with 10 candidates when result composition does not match to target one:

Experiment index	8	
Number of Candidates	10	
Candidates	[0, 5, 10, 15, 20, 25, 30, 35, 40, 45]	
Target Composition	[0.1, 0, 0.5, 0, 0.4, 0, 0, 0, 0, 0]	
Points	Point Selection	Result
10	[2800, 3300, 50]	[0, 0, 0.787939, 0, 0.112061, 0, 0.1, 0, 0, 0]
50	[2800, 3300, 10]	[0, 0, 0.731043, 0.149797, 0, 0, 0.119159, 0, 0, 0]
100	[2800, 3300, 5]	[0, 0, 0.731043, 0.149797, 0, 0, 0.119159, 0, 0, 0]
$100 + 1$	[2800, 3300, 5], [2800, 3300, 500]	[0, 0, 0.731043, 0.149797, 0, 0, 0.119159, 0, 0, 0]
$100 + 100$	[2800, 3300, 5], [2800, 3300, 5]	[0.156072, 0, 0, 0.831607, 0, 0, 0, 0, 0, 0.0123217]

3.3.14 C

oncern(: too easy to find the case that LP solver not returning the target composition.

3.4 Result

3.5 Discussion

During the try-out, we first select the data points at the peaks, which are four points at frequency of 2850, 2960, 3050 and 3200. We then construct the Linear Programming model based on these data points. The result returned by LP-solver matched to the known one. However, if we randomly select four data points, the returned result usually does not match to the know one. At the end, we increase the number of data points, at each 5 wavelength frequency gap, a data point will be selected. And the returned result will eventually match to the known one.

The more data points we select, the more information about the candidates and targeted spectrum we will have in our model. Therefore, the more complicated and complete model we will construct. As we select one data point, a new variable is introduced to our model, meanwhile, two new constraints are brought into our model as well. This means that solution space is further and better restricted. When we select data points only on the peaks, these data points already contains enough information for the solver to distinguish the candidates. However, if we randomly select four data points, they may not contain enough information in order to construct the model to obtain the desired result.

3.5.1 Conclusion

Chapter 4

Realistic Molecular Model

4.1 Description

4.2 Results

After experimenting with the toy problem, we figure out that it is the lack of sufficient information for linear programming model. Because in toy model, we only have four vibrational modes, and the range of XXX the is limited as well.Therefore, we think it is time to introduce the real molecule, which contains more abundant spectroscopy information, and meanwhile, make our study one step closer to real application.

How the scripts are organized to run the experiments?

The goal is to compare different techniques' sensitivity in studying molecule's orientation distribution at surface. (We want to figure out which spectroscopy technique is rich in orientation information.)

4.3 Discussion

4.4 Conclusions

4. How the candidate data is used to organize the experiments

We would like to compare the three spectroscopy techniques to see which tenique is more sensitive to molecular orientation. In order to do this, we set up a group of experiments.

The first experiment is a model based on the data obtained from molecule's IR spectroscopy, and because we are considering theta's Gamma distribution here, we can get the data points from both x and y-polarized spectra.

The second experiment is a model based on the data obtained from molecule's Raman spectroscopy, same reason,

Similar for the third experiment, it is a model based on SFG spectra data, there are three different spectra

However, the fourth to the sixth, is a combination of two of the above techniques, and the seventh one is a combination of all the techniques.

So that we can not only compare which technique is more sensitive, but also we can see that when the information is not only when the candidates get more similar to each other, and when the insufficient situation happens, we can see by adding which technique will improve the case. What's more, we can also know when we will hit the case that by combining all the information we have, it is still not enough to build a model to get desire result.

This group of experiments are listed in below table.

Experiment Index	Spectroscopy Techniques	Polarization
1	IR	x, y
2	Raman	xx, xy, xz, zz
3	SFG	yyz, yzy, zzz
4	IR and Raman	x, y, xx, xy, xz, zz
5	IR and SFG	x, y, yyz, yzy, zzz
6	Raman and SFG	xx, xy, xz, zz, yyz, yzy, zzz
7	IR, Raman and SFG	x, y, xx, xy, xz, zz, yyz, yzy, zzz

5. Experiment set Explaining:

Experiment 1: The first group of candidates we have tried out is that having different amino acids: methionine, leucine, isoleucine, alanine, threonine, and valine. For each amino acid, there are 18 different candidates with theta different from t 0 and 180, expect theta of 90 degree. Because when theta is 90 degree, there is no SFG signal.

We will run the above experiment group 1000 times in order to get a more general conclusion.

In order to conduct experiments based on less pre-assumptions. Every time, we run the experiment group, we will re-shuffle the order of the amino acids in the list. For each amino acid, we then randomly select one of the candidates, then randomly assign percentage to this candidate with the amount that is bigger than 0, but smaller than the reminding percentage $(1 - perassigned\ percentage)$. We

What we find interesting is that, if we only take IR data into our experiment, the returned composition we get from linear programming tool is quite far off the target one. However, if we plot the spectra with the returned composition, it is almost identical as the target spectra, and the residual between these two spectrum is almost negligible, which results at a number as small as 1.84040221222e-05. In order to make a more general conclusion, we run the same experiment 100 times. Form these 100 runs, we observe that the scenario in every single run is the same as the case described before. From this observation, we can see that if we use IR alone to study the molecule's coordination from a group of different candidates, there will not be sufficient information. What's more, it may even be misleading for some cases. Because the spectra generated by plotting the returned composition is so close with the target one. But at the same time, the returned composition is much different from the target one. Furthermore, in reality, the composition is what we need to study, it is the information we do not know ahead. And if we only compare the composed spectra with the target one, they might

Experiment 2: One molecule,

Chapter 5

Evaluation, Analysis and Comparisons

For a Master's research this chapter represents the critical part where **you** are truly evaluated to determine whether you should be given your degree. Even more so for a PhD. Consider carefully what the University calendar states regarding the expectations for a master's thesis, paraphrased here.

1. *A Masters thesis is an original lengthy essay.* The main implication here is that the essay is original, that is, it is completely newly written by you and does not contain any writings from others unless precisely quoted. Any paraphrased items must be cited.

2. *It must demonstrate that:*

   - students understand research methods;
   - students are capable to employ research methods;
   - students demonstrate command of the subject.

3. *The work may be based on:*

   - original data;
   - original exercise from scholarly literature;
   - data by others.

4. *The work must show that:*

- appropriate research methods have been used;
- appropriate methods of critical analysis supplied.

5. *The work must contain:*

   - evidence of some new contribution;
   - evidence of a new perspective on existing knowledge.

Only the last point uses the attribute *new* and it refers almost entirely to giving a new perspective and analysis, even if based on data from others. This truly implies that this current chapter on evaluation and analysis of results is the most important and must be written with care. You are demonstrating here that, even if given data and methods from others, your skills of critical judgment and analysis are now at the level that you can give professional evaluations.

Things are slightly different for a PhD. According to the Graduate Calendar:
a doctoral dissertation must embody original work and constitute a significant contribution to knowledge in the candidate's field of study. It should contain evidence of broad knowledge of the relevant literature, and should demonstrate a critical understanding of the works of scholars closely related to the subject of the dissertation. Material embodied in the dissertation should, in the opinion of scholars in the field, merit publication.

The general form and style of dissertations may differ from department to department, but all dissertations shall be presented in a form which constitutes an integrated submission. The dissertation may include materials already published by the candidate, whether alone or in conjunction with others. Previously published materials must be integrated into the dissertation while at the same time distinguishing the student's own work from the work of other researchers. At the final oral examination, the doctoral candidate is responsible for the entire content of the dissertation. This includes those portions of co-authored papers which comprise part of the dissertation.

The second paragraph makes it clear that one must emphasize what is new and different from others, without arrogance, yet without being too subtle either. The first paragraph implies that for a PhD it is required that one approached an important open problem and gave a new solution altogether, making chapters 3, 4, 5 all part of the body of research being evaluated. In fact at times even the problem may be entirely new, thus including chapter 2 in the examination. This is in contrast to a Master's degree where the minimum requirement is for chapter 5 to be original.

Chapter 6

Possibilities for treating experimental data

6.1 Description

Introduce idea of arbitrary scaling in experimental results (for the case of Raman and SFG) and in the calculation of molecular properties(GAMESS)

6.2 Results

Solution in the terms of slack variables

6.3 Discussion

6.4 Conclusions

Chapter 7

Conclusion and Future Work

7.1 Conclusion

7.1.1 Contributions

7.2 Future Work

Appendix A

Additional Information

This is a good place to put tables, lots of results, perhaps all the data compiled in the experiments. By avoiding putting all the results inside the chapters themselves, the whole thing may become much more readable and the various tables can be linked to appropriately.

The main purpose of an Appendix however should be to take care of the future readers and researchers. This implies listing all the housekeeping facts needed to continue the research. For example: where is the raw data stored? where is the software used? which version of which operating system or library or experimental equipment was used and where can it be accessed again?

Ask yourself: if you were given this thesis to read with the goal that you will be expanding the research presented here, what would you like to have as housekeeping information and what do you need? Be kind to the future graduate students and to your supervisor who will be the one stuck in the middle trying to find where all the stuff was left!

Bibliography

[1] Cliff Atkinson. *Beyond Bullet Points: Using Microsoft Office PowerPoint 2007 to Create Presentations That Inform.* Microsoft Press, 2008.

[2] Wayne Booth. *The Craft of Research.* University of Chicago Press, 2003.