**Table for reporting responses to reviewers’ comments**

|  |  |  |  |
| --- | --- | --- | --- |
| **Examiners comment** | **Student’s response** | **Change made to thesis (if any)** | **Page reference** |
| Page 6, para 1, line 9. Change ” at the end of the line to “ and make this change throughout as  needed | Fixed. This error was not present anywhere else in the thesis. | Changed as requested. | 6 |
| Page 6, para 3, line 3. Change graph to graphs. | Fixed. | Changed as requested. | 6 |
| Page 7, line 7. You may want to point to the definition of coupling time since it has not yet been defined | Added reference forward to coupling time section. | Added reference forward to coupling time section. | 7 |
| Page 9, line 2. Change n to *n* | Fixed. | Changed as requested. | 9 |
| Page 18, section 2.3.2.3, para1, last line. There are extra commas around the .... Also, it seems like t\_2 is missing from the condition. | t\_2 was indeed missing from the condition and has been added back in. The commas around the ellipses are a deliberate stylistic choice and **have not been changed.** | Added t\_2 to the condition | 18 |
| Page 48, last line, sentence starting with “Since our proof...”. It was unclear how you meant a similar approach could extend your results: To the criticial temperature? You do address some of this in the  conclusions to Part 1, but it might be good to clarify here what you mean. | Changed this sentence to clarify our meaning and explicitly mention how a similar approach could extend our results. | Changed sentence starting with ``Since our proof...’’ to  “Since our proof is also based on information percolation there is good reason to think that a similar approach could also work to extend our result all the way to the critical temperature.” | 48-49 |
| For the second part of the thesis, it might be nice to have a brief summary of your results right before/after explaining Lindsay’s approach in section 6.2. I was left a bit unsure as to what the contribution of the thesis was until I had read through a good deal of Chapter 6. It would be nice to set your results in the context of the introduction is possible.  It may also be nice to give a few  specific directions for future work in the area in section 6.5. | Added a summary of our contribution directly before Section 6.2.  Added | Added the following paragraph directly before Section 6.2.  “In this Chapter, we will explore current results concerning $K\_\vect{x}$, and expand on these to produce new bounds for $K\_\vect{x}$ in a variety of settings. We look at current results in Section \ref{sec:summary of Lindsay}, and present new results in Section \ref{sec:mixture results}. Our main results are Theorems \ref{thm:n=2 inflection result}, \ref{thm:general n C1 bound}, \ref{thm:bounding Cm}, and \ref{thm:probability in C1}; as well as a simple corollary that results from this last theorem. We also make use of an empirical analysis of $K\_\vect{x}$ in Section \ref{sec:empirical results} to frame the problem from a different perspective, and to give us a point of comparison for our bounds. The discussion of current results in Section \ref{sec:summary of Lindsay} also contains some original work placed throughout where it is directly applicable. In particular, Theorem \ref{thm:all points separated by alpha} and Section \ref{sec:KKT conditions} are both new.”  Added the following paragraph to the conclusion of Chapter 6.  “One might hope to extend the results above to a wider class of densities, or to generalize the results for $n=2$ to $n > 2$. This is indeed an area of interest for future research, but there are some immediate apparent difficulties that arise if we try to use the same approaches as in this chapter. In particular, when $n=2$ we are able to simplify our problem to one concerning the curvature of the likelihood curve. For $n>2$, the problem cannot be answered by simply considering the curvature. For normal densities, we were able to obtain results for $n>2$ by taking advantage of the simple form that the derivatives of the normal density take. In general, we are not guaranteed that derivatives for other densities will simplify as well. However, we may be able to generalise our results to other densities in the one-parameter exponential family, in a similar fashion to Lindsay in \cite{Lindsay1983a-he}.” | 71-72, 102 |
| There is a problem with Figure 6.6 on page 80. The plot claims to show the ‘likelihood curve’ (f\_θ(x\_1), f\_θ (x\_2)) as θ traverses the parameter space, with  (x\_1, x\_2) = (0,0.4) and f θ (x) = f(x−θ) where f(·) is the triangular density with width 1/2. However, the form of this density is  f(x) = {  4 − 16|x| for |x| ≤ 0.25,  0 otherwise.  }  In particular the maximal value is 4, not 1 as suggested by the plot. The problem can be remedied by  - mutliplying the markings on the axes by 4 and  - replacing x\_2 with 0.2  without changing the qualitative features of the plot. | This problem can also be rectified by taking f to be the triangular density with width 2 and changing x\_2 to 0.8. This was the original intention and so we have made this change instead of changing the figure. | Changed from x = (0, 0.4) to x = (0, 0.8) and changed ‘f is a symmetric triangular density with width 1/2’ to ‘f is a symmetric triangular density with width 2.’ | 80 |
| The estimator (7.13) is not necessarily consistent without extra assumptions on the distribution of the errors U\_jk . In particular the Fourier transform φ\_U(t) must be real-valued, so the U\_jk 's need have a symmetric distribution about zero; see condition (2.2) in Delaigle et al. (2008). This should be mentioned at some point in the lead-up to (7.13); this is an opportune moment  to foreshadow the assumption A7 that is made in the following section. | Added the extra assumptions required for (7.13) to be a consistent estimator. | Extended sentence leading up to (7.13) to mention the extra assumptions required to make it a consistent estimator. | 105-106 |
| The comparison of the MATLAB and R implementations of the methods of Chapter 7 is inadequate. Since computational implementation is the focus of Chapter 7, more work is needed to explain the observed differences in  performance between the MATLAB and R versions, other than to say ‘We are unsure as to why out implementation in R tends to produce worse objective values than out implementation in MATLAB....we do not know exactly what is going on.’ This last phrase should preferably not appear in a PhD  thesis, it would be better to identify a few possible causes and investigate them; even if they do not reveal anything it is important to show that such  issues are not to blame to assist other future researchers (perhaps this was done but not mentioned?).  There are various aspects one could explore, I list a few below (there are many others):  - Are there any tuning parameters used in either the MATLAB or R  implementations? If so, are they set to the same values?  - The R package NlcOptim is a third-party package written by an ‘ordinary R user’ (as opposed to a built-in function developed by the R Core Team). It is written entirely in R code, and does not appear to call any compiled C, Fortran or C++ code, unlike the general-purpose optimisation functions ‘built-in’ to R. Can the same be said of the MATLAB implementation? Does a Fortran, C or C++ version of this  algorithm exist which could be used in R (see the R extensions manual:  https://cran.r-project.org/doc/manuals/R-exts.pdf for how to  do this).  At least two possible causes should be investigated and reported on. | The suggested aspects to explore were things that we had already investigated but failed to report on. We have added discussion on the steps we took to try to resolve the differences and added some speculation as to why MATLAB is performing better. | Added following sentence to the section introducing the method used in our R implementation: ‘We set all available tuning parameters to be equivalent to the ones used in our MATLAB implementation.’  Removed the paragraph that was commented on and added the following paragraphs in its place:  We are unsure as to exactly why our implementation in R tends to produce worse objective values than our implementation in MATLAB. One possibility is that the optimization routines used in MATLAB's \emph{fmincon} function are more sophisticated than those used in the Nlcoptim package for R. We note that MATLAB's algorithms have been refined over many years by a large number of contributors, whereas Nlcoptim is written by just two authors. It should be noted that we tried several other R packages to perform the optimization. These included alabama \cite{Varadhan2015-ra}, nloptr \cite{Ypma2018-og}, and the built in base R function, optim. This last function does not natively allow for non-linear constraints and so we instead encoded our constraints using an infinite penalty in the objective. We were unable to obtain satisfactory results with any of these packages.  However, we are not particularly concerned about the difference in objective values between our two implementations since the estimated densities we obtain using R appear to be just as good as those we get in MATLAB. This is possibly related to the discussion earlier in this section about Figure \ref{fig:moving masses m40 example} in which we pointed out that worse objective values in our first optimization can sometimes produce better looking densities. | 120, 121 |
| The reference numbered [56] seems to be incomplete and possibly incorrect; I could not find the work it seems to refer to | Citation was incomplete and had wrong year and page numbers. | Changed year from 2018 to 2019 and added volume and issue number. Also fixed page numbers. | 130 |
| Page 75 I believe that σ\_2 = 0.4^2 should in fact be σ^2 = 0.4^2 ; please check. | Fixed. | Changed σ\_2 to σ^2 | 75 |
| Page 78 It should be explicitly pointed out that H and H\_Q are the same, or the definition at (6.32) could be written as  H = H Q = ···. | We use H (without subscript) twice and in both cases we specifically mean H\_\hat{Q}. So we have changed these two occurrences from H to H\_\hat{Q}. | Changed H to H\_\hat{Q} in two places | 78 |
| Page 83 I believe that in inequality (6.65), the θ\_j should be θ^∗ ; please check. | Fixed. | Changed as requested | 83 |
| Pages 92-3 In the proof of Theorem 6.9, reference is made to results in Lemmas which have not yet been presented e.g. Lemma 6.10. In such cases  it would improve readability slightly to insert the word ‘below’ to make clear  to someone reading the work linearly that it has not yet been read (this is done at other points in the thesis but not here). | Fixed. | Added word ‘below’ after the first references of Lemmas 6.10 and 6.11 in the proof of Theorem 6.9 | 93 |
| Page 112 At the end of the second paragraph of 7.2.4, it says ‘...we will  explore this further in Section 6.3’. It is perhaps the case that the intention was 7.3. Please confirm. | Fixed. | Fixed reference from 6.3 to 7.3 | 112 |
| Page 116 It is better to use the term ‘mass point(s)’ when referring to the points of support of a discrete distribution. The term ‘mass(es)’ is ambiguous, is sometimes used for the actual probability/weight supported on the point, as well as for the point itself. | Changed ‘masses’ to ‘mass points’ where appropriate on page 116. There were also other uses of ‘masses’ later on which we could make less ambiguous by using `mass points`. | Changed ‘masses’ to ‘mass points’ throughout discussion on pages 116-118 where we mean the mass points and not just the probability weights. | 116,118 |