On smooth models for complex domains and distances: list of minor corrections II

Simon Shaw

March 26, 2012

1 Notes on Chapter 1

- p29 Need $\cdots$ between the integrals in (1.3)
- p35 In $B_j^{-1}(x)$ you have $x_j \leq x < x_{j+1}$ which should be $x_j^* \leq x < x_{j+1}^*$
- p36 Your placement of the definition of β_j and δ_j is temporally incorrect and suggests that these are not defined in the statement of $f(x)$ but merely a choice you have made after $f(x)$ is constructed. They need to be defined *before* the statement of $f(x)$ (see, for example p149 of Simon's book where they are stated before equation (4.2)).
- p38 In the definition of $\hat{\mathbf{f}}$ shouldn't we have $\hat{f}(\mathbf{x}_i)$ as its ith component rather than $f(\mathbf{x}_i)$?
- p39 "evaluations of $\hat{f}$ at each of the data" should be "evaluations of $\hat{f}$ at each of the data points".
- p40 $Var(Z_i) = \phi V(\mu^{[k]})$ This should be (as Richard suggested) $Var(Z_i) = \phi V(\mu_i)$. Firstly, you are setting up $Var(Z_i)$ ahead of the algorithm so there's no notion of what $\mu^{[k]}$ is. Also, I know Simon defines it this way (note it is within the PIRLS algorithm!) on p138 of his book but it's a typo and lacking the i as (there) Y_i is a scalar and $\mu^{[k]}$ a vector. A closer match to your approach is p500 of Simon's 2008 JRSS B paper.
- p41 In the pseudodata you've amended $\boldsymbol{\beta}^{[k]}$ to $\hat{\boldsymbol{\beta}}^{[k]}$. This wasn't the issue with the pseudodata I was referring to and I don't think this change is correct either: what is $\hat{\boldsymbol{\beta}}^{[k]}$? The algorithm, at step 2., generates $\boldsymbol{\beta}^{[k]}$. It was the first term of s_i that I was meaning: Simon has $g'(\mu_i^{[k]})(z_i - \mu_i^{[k]})$ you have $g'(\mu_i^{[k]})^2(z_i - \mu_i^{[k]})$
- p41 "Further information on IRLS in the GLM context can be found in chapter 1 of Hardin and Hilbe (2002)." If Hardin and Hilbe (2002) don't cover PIRLS then you should remove this line as it doesn't make sense in the context of the previous work discussing PIRLS. In particular, is IRLS mentioned elsewhere at all in the thesis?
- p41 You've still got an issue with the dimension of the smoothing parameter as it appears as a scalar in step 2. of the PIRLS algorithm. I think you are probably better to write the penalty

- in step 2. of the PIRLS algorithm as $\sum_j \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$. You talk about using $\mathbf{S} = \sum_j \lambda_j \mathbf{S}_j$ on p39 so I'd set up the penalty in terms of $\sum_j \lambda_j \boldsymbol{\beta}^T \mathbf{S}_j \boldsymbol{\beta}$ here i.e. the effective substitution of (1.4) into (1.3) and rewriting $\mathbf{S}$.
- p44 Don't think we mentioned this before but "an measure" should be "a measure".

2 Notes on Chapter 2

- p56 Remove space after "poles".
- p56 In the discussion of the Tweedie distribution you write "prior to using the Tweedie the gamma distribution was used, however since it does not handle exact zeros in the data, the Tweedie (which does allow exact zeros) was used." On p65 however you write "Before employing the class of Tweedie models, simpler gamma distribution-based models were used, diagnostic plots suggested the presence of substantial structure in the residuals." These points should be unified (what was the real reason?) and I think the lines in the quote from p56 could be polished.
- p59 Hadn't noticed this before, but you have numerous $\zeta_\iota(n, e)$ on this page which should be $\zeta_j(n, e)$.
- p62 Delete the sentence "The easiest way ... in (2.3) as." and replace by "Hence, the spatial part of the penalty given in (2.3) is expressed as".
- p62 Delete the sentence "The smoothing ... above clearer". In the thesis version you've omitted the explicit tensor product formulation (equation (2) of the paper) so I'm not sure that that half of the sentence really adds anything in the thesis. Amend the next sentence to start "Rather than estimating λ_{space}, the two ...".
- p65 I think your justification for the need to fix the Tweedie parameter p requires a little more work. Firstly, "the p parameter maybe optimized in a GLM setting however such a method is not available in mgcv for GAMs at this time" is perhaps not an ideal sentence. Whilst it may be true, it gives the impression that all you were interested in doing was running a dataset through existing software and generating the answers which doesn't give a great impression. For example, was there anything stopping you adapting the software to enable the optimisation of p? Secondly, I don't think that something can be classed as "received wisdom" when the only evidence for it is a personal communication. I think the response you want here is given perfectly by p529-530 of this paper.

  `http://onlinelibrary.wiley.com/doi/10.1111/j.1523-1739.2011.01656.x/abstract`
- p66 In Figure 2-3(v) could you colour the points corresponding to exact zeroes in the data in a colour other than grey as you have used grey for the LOESS curve?

3 Notes on Chapter 3

- p75 Four lines from the bottom, add a space before the start of the sentence "The boundary ...".
- p81 In step 2. of the algorithm, "objective function, F, as" should be "objective function, F_k, as".

4 Notes on Chapter 4

- p104 Didn't notice this before but in point 4. of the proposed procedure you have $\mathbf{x_j}$ when you mean $\mathbf{x}_j$.
- p105 My point was really that it's not really clear what the phrase "the values in $\tilde{\mathbf{X}}^*$ have been centred about the origin" means. For example, what are the values? What you mean is that the centroid of the configuration of points is centred about the origin so that the sum of each column of $\tilde{\mathbf{X}}^*$ is zero. Consequently, $\tilde{\mathbf{X}}^{*T}\mathbf{1} = 0$ which is required for $HSH = \mathbb{S}$. Of course, you have to be careful about using $\mathbf{1}$ on p105 as you don't define it until p106. I think you still need to tighten Section 4.2.1 a little. For example, in the opening paragraph, you write "for the moment let us assume that $\mathbf{D}$ is known" and then never come back to a discussion of this (e.g. when is it unknown?). I think the section would really benefit from a early and clear statement of what you are trying to do with MDS. Richard gives you the line in his point 31., that of "trying to find a spatial configuration that reproduces the observed distance matrix". A few pointers.
  - p105 Why do we have n points in n dimensions rather than the conventional n points in p dimensions?
  - p105 "First pretending" I don't like as an opening (I know Richard uses the term "pretending" in his notes but he's being informal). I think it would be better to explicitly say that we have n unknown points $x_1^*, \ldots, x_n^*$ and, from these, we create the matrix $\tilde{\mathbf{X}}^*$ whose ith row is x_i^*.
  - p105 "Note that $\tilde{\mathbf{X}}^*$ may only be found ..." needs to tightened: at this point in the discussion it hasn't been made clear that we are going to try and find a $\tilde{\mathbf{X}}^*$. You should try and relate the centring to the fact that translations and rotations don't alter the Euclidean distance.
  - p106 Clarify the two sentences immediately after equation (4.6), they're rather implicit and I don't know what you mean by factoring $\mathbb{S}$. You need to make explicit the point made by Diaconis et al. (2008, p781). "In summary, given an $n \times n$ matrix of interpoint distances, one can solve for points achieving these distances by the following" and the three steps he lists. This point gets rather lost in your explanation, particularly the third step that, when we don't have $\tilde{\mathbf{X}}^*$ we can extract an $\tilde{\mathbf{X}}^*$ from $\mathbf{D}$.
  - p106 I think Richard's point about "why does the eigen-decomposition have this form" is that he's alluding to your original version giving the impression of a lack of awareness of why this might be the case. It follows from noting that $\mathbb{S}$ is symmetric and applying

the spectral theorem for a symmetric matrix; your initial version mentioned a Choleski decompos but this would additionally require $\mathbb{S}$ to be positive-definite.

- p106 Immediately following (4.7) you note that $\tilde{\mathbf{X}}^*$ is $n \times n$: why?

- p115 "Using theses samples, three methods were fitted to the data" doesn't quite sound right. How about "Three different fitting methods were applied to the data."? The change on p163 is fine.

- p117 Correct the grammar in the phrase "that the smoother estimating the peaks in the peninsula correctly".

- p "MDS+RS must not only to have a sensible physical model" doesn't make sense

5 Notes on Chapter 5

- p155 Stray bracket: "coefficients)"

- p157 In equation (5.2) you need $\eta_{m,d}(x_i)$ to match the form of equation (1.6)

6 Notes on Chapter 6

No issues.

7 Notes on Chapter 7

- p207 Reference should be "Marques and Buckland (2003)" rather than "Marques and Buckland, 2003".

- p208 Typo "it's". (Repeat of Richard's original point 56. "It's" means "It is" or "It has").

- p208 I must confess I'm still not really sure whether your " average detection function" is a statistically reasonable thing to do but I don't think that's a discussion for here.

- p208 Typo "the average performed".

8 Notes on Chapter 8

- p215 Amend "in section 7.3 in chapter 7" to "in section 7.3".

- p215 The phrase "so $\mathbf{Z}$ and z_i are removed" referred to your original version of (8.1) but makes no sense with the revised (8.1).

- p220 Caption to Figure 8-1 should have α_j.

- p229 Typo in the caption to Figure 8-5. "as above but the holding".

- p233 $g(x)$ is a generalisation of (8.1) so the notation should more closely reflect that.