Reply to the Reviewer 2:

Thanks for the careful reading and for your helpful comments and suggestions. Please find below point-by-point replies (in black) to your comments and questions (which are reprinted in blue). To give you an overview of all the changes in the paper, we also provide a diff-document that highlights the changes between the initial submission and this re-submission.

I have read the paper with pleasure and have the following suggestions for improvement:

1. Interesting work, generally well-written. Perhaps a small illustrative example with a few figures at the beginning could help understand the procedure intuitively.

   Thank you for this suggestion, which we agree can help the reader better understand the intuition behind the method. We added a 2D blurring kernel example, with figures illustrating the end-to-end PSF method. These are shown in the introduction and the numerical results sections (Sections 1 and 7.3).

2. possibly interesting references for estimating Hessians in linearised seismic imaging include: `https://doi.org/10.1016/j.acha.2007.06.007`, `https://doi.org/10.1190/1.2836323`

   We added text in the paragraph at the end of Section 2 to discuss the (lack of) applicability of our method to wave inverse problems, and added these references.

3. How about comparison with other methods, like those proposed in [7],[52]? and you say a bit more? it is hard to conclude that your method will always outperform these methods.

   We added a numerical comparison of our method with the HODLR method in [7] for a blur kernel in Table 5 of Section 7.3 in the revised manuscript. The results show that the HODLR method requires hundreds to thousands of operator applies to achieve modest accuracies (e.g., 10% error) while our method requires only order 10 operator applies. In Section 2 of the revised manuscript we also added a discussion of the costs of the method proposed in [52]. We present back of the envelope calculations based on complexity estimates given in [52], showing that thousands of operator applies may be required.

4. Perhaps it is illustrative to include some 1D blurring examples to illustrate robustness of method to departure from assumptions (positivity, locality) e.g. on 'indefinite' kernels or kernels with multiple modes.

   We added a numerical study investigating the robustness of the method to violations of the non-negative kernel assumption on two 2D blurring kernel examples (Figure 7). We also added a discussion of the impact of non-locality and negativity on the performance of the method (Section 3.2).

5. is code for applying the developed methods on other inverse problems available?

   Ideally we would like to release the code, however this will take time. In the meantime, we will make the code available upon request.