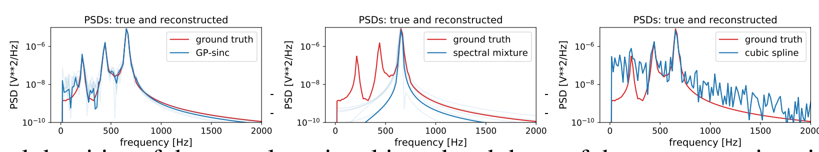
Author response for NeurIPS submission 6935 (Band-Limited Gaussian Processes: The Sinc Kernel)

We are very grateful to all three Reviewers for their valuable suggestions and encouraging comments, these have undoubtedly improved our work. We next address the concerns raised by the Reviewers in 5 parts. The following content (in extended form) and the required code will be part of our final submission if applicable.

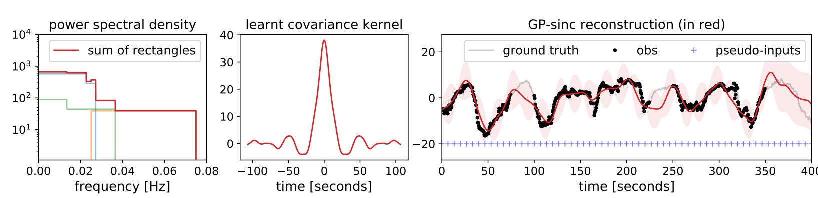
I. Novelty and comparison to other covariance kernels. We agree that, as pointed out by Reviewer 2, the proposed Sinc kernel can be regarded as a Spectral Mixture (SM) kernel, where the RBF is replaced by a Sinc function. However, notice that most kernels are slight modifications of one another when parametrised in the Fourier domain: SM comes from a non-centred RBF, Square Exponential comes from a centred RBF, Cosine comes from a Dirac, ν-Matérn comes from the filter $\frac{1}{(1+\xi^2)^{\nu+1/2}}$, Exponential comes from a Student's t-density and, lastly, Sinc comes from a Rectangle. Therefore, we emphasise our contribution is not the Sinc kernel *per se*, but rather a study about its unique features and how they can enable modern GP-based approaches to Signal Processing. With the ever-growing applications of GPs, we believe that understanding the properties of new kernels is of interest for researchers and practitioners alike.

II. Impact on sparse GP design. Reviewers 1 and 2 recognised the impact of our analysis towards the design of sparse GPs and they recommended us to further explore this connection. To do so, let us first recall that implementing a sparse GP requires (i) the quantity of inducing points M (usually chosen by hand) and their locations (usually jointly optimised with the hyperparameters). If the proposed Sinc kernel is considered, we can rely on the Shannon-Nyquist theorem to parametrise the locations of the inducing inputs as an evenly-spaced grid at (twice) the Nyquist frequency, i.e., the largest frequency in the spectral support. This procedure has two key advantages for sparse GP regression: first, both the number and locations of inducing variables are determined by the Nyquist frequency (the hyperparameter Δ of the kernel) and they need not be chosen or optimised, thus simplifying the optimisation stage. Second, we have theoretical guarantees for perfect, i.e, zero-variance, posterior reconstruction, a property that has been little explored in the GP community. Specifically, if the range of our data is L and the Nyquist frequency is Δ, we can choose $M = 2L\Delta$ inducing points evenly spaced across the range of our data, which makes the training cost $\mathcal{O}(NL^2\Delta^2)$. Notice that this rationale paves the way to quantifying the intuition that the number of inducing points should depend of both the extension of our datapoints (L) and the data's frequency content (Δ).

III. Additional experiments. The reviewers highlighted key aspects in our manuscript that were discussed but lacked experimental validation. Next, we present two sets of experiments addressing the comparison of GP-sinc against spectral mixture and classic interpolation relationship (see Part III.A), and implementation of the generalised Sinc kernel, i.e., a Sinc mixture, using the previous concept of sparse approximation (explained in Part III.B).



III.A Comparison against spectral mixture and classic interpolation. We considered the reconstruction of a band-limited audio signal (form TIMIT) using 20% unevenly-sampled observations. The plots show the spectral densities of the complete signal in red and those of the reconstructions in blue for the proposed GP-sinc (left), GP spectral mixture (centre) and a cubic spline (right). The proposed GP-sinc outperformed the benchmarks due to the rectangular PSD function, which allows frequencies with high and zero energies to be arbitrarilty close.



III.B Mixture of Sincs and sparse implementation. We trained a GP with the generalised Sinc kernel (GSK), i.e., a Sinc mixture, using a heart rate signal with unobserved regions. Recall that the GSK provides a compact-support and frequency-varying spectrum, therefore, we trained a sparse version of GSK using the rationale described in Part II. The plots show the PSD at the left (components in colours and mixture in red), the kernel at the centre, and the time series (ground truth, observations, and reconstruction) at the right. Notice from the right plot that though $N = 600$ observations were considered (black dots), only $M = 54$ inducing locations (blue crosses) were needed following the discussion in Part II.

IV. Uncorrelated process and noiseless observations (Sec. 5). Reviewer 2 asked for the necessity of these assumptions. We clarify that we were only interpreting classical Signal Processing approaches into our setting, where no time correlation or observation model is assumed. In such case, we showed that our model collapses to the Whittaker–Shannon interpolation. However, without this assumption our model still holds through eq. (23).

V. The Nyquist frequency (NF). We agree with Reviewer 1 in their recommendation to include a brief introduction to NF and the concept of perfect reconstruction. Many thanks for this suggestion.