Recap: What is the paper's goal, and why? Bound the statistical complexity of tuning SM Kernel hyperparameters, when **noise is fully adversarial** in the **active regression** setting. While extremely useful, the SM Kernel is notoriously hard to tune (L147-148). By proving that the statistical cost of tuning the SM Kernel is low, even in this hard regime, we show that we need better algorithms for hyperparameter tuning, and not to just collect more data. Also, we provide a new mathematical framework to analyze hyperparameter tuning, which hopefully opens the door to new algorithms.

We thank the reviewers for their constructive and broadly positive feedback. We are pleased to see the reviewers find our problem statement interesting and relevant, find our technical results to be sound and relevant to practitioners, and find our writing to be clear. We also appreciate the reviewers asking about potential future directions of our work, either for new algorithms or for a different statistical setup.

One important open question identified by (R1, R3, R4) is about **finding a polynomial time algorithm**. While our paper does not have a polynomial time algorithm for SM Kernel hyperparameter tuning, **it opens the door for more research towards that goal**. Previously, it was unknown if exponentially many samples were needed to even *information-theoretically* tune a SM Kernel. If this were the case, then there would be no hope for finding a polynomial time algorithm. We show that this barrier does not exist, which gives more hope for a fast algorithm.

Along the same lines, R1 asks us to **"discuss the possible direction[s] of developing the corresponding algorithm"**. This highlights another potential benefit of our paper: prior work does not frame hyperparameter tuning as a Fourier fitting problem. By rigorously and directly connecting the two settings, our paper makes it easier for a polynomial time algorithm to originate from signal processing and benefit kernel hyperparameter tuning. **The Sparse Fourier Transform and Compressed Sensing literatures seem promising in this regard** – notably, they both feature fast algorithms with statistical guarantees for many Fourier fitting problems. We agree that the paper should discuss these directions explicitly, and we have added such a discussion to our conclusion.

R2 is concerned about the **applications and broader impact of our paper**. We study kernel ridge regression with adversarial noise. There are many applications of this setting. For instance, we cite two papers that use the SM Kernel in distinct ways: [HSSM15] uses the SM Kernel in an analysis of the lifespan of lithium-ion batteries and [WDLX15] uses the SM Kernel to model human decision-making processes. Many more applications can by found in the papers that cite [WA13], which proposed the SM Kernel. So, **the SM Kernel is relevant to many applications** despite being hard to tune. Our **broader impact depends on how the SM Kernel is used**. We agree the introduction and broader impact sections would benefit from a clearer discussion of applications of our framework, and have added that to the paper.

R1 also brings up the **novelty and impact of our paper, and its relationship to [AKM+19]**. [AKM+19] is a recent STOC paper about Fourier function fitting in the same adversarial noise setting as our paper. They show that kernel ridge regression provides a good interpolant when the kernel is fixed (i.e. hyperparameters are *known*). The novel technical contribution of our work is the extension of [AKM+19] to the setting with an *unknown* kernel. This allows us to frame kernel hyperparameter optimization as a Fourier fitting problem, so **[AKM+19] lets us prove the first bounds on the statistical complexity of learning kernels with totally adversarial noise**. Mathematically, this comparison is made clear in our introduction (Problem 1 versus Problem 2).

R4 asks **if our results can apply to the fixed data setting**. This is an interesting question, but answering it requires a different noise model than our totally adversarial noise model. If the data was fixed, then an adversary could arbitrarily perturb the signal $y(t)$ only at the observed times. So, a small-norm noise signal would make approximate recovery of $y(t)$ totally impossible. Instead, perhaps the adversary could only perturb an ε fraction of the observations? We think this is an interesting direction, and it is plausible that our techniques could generalize such a fixed-kernel result into a hyperparameter optimization result, but this speculation and out of scope for our paper.

R2 asks for a **"intuition behind the statistical dimension" and how it varies between kernels**. Kernels are nonparametric functions that can involve infinitely many features (in our paper, a feature of a kernel is like the Fourier transform of a kernel at a given frequency). Nevertheless, in Kernel Ridge Regression, thanks in part to regularization, kernels can typically be well approximated with a finite number of features (this is the idea behind e.g. Random Fourier Features). Statistical dimension captures exactly how many features are needed. Since regression in d features requires roughly d samples, this means kernel ridge regression needs roughly statistical dimension many observations. For the RBF kernel, statistical dimension is linear in the lengthscale parameter and the duration of time we interpolate over. The intuition is that if either the lengthscale or the duration of time increase, then we can represent more complex functions with the same Fourier-norm, so this necessitates more samples (i.e., larger statistical dimension).

R2 asks for **Clarification about $O(1)$ and C in our bounds**. These all refer to a universal constant independent of any problem parameters such that the left hand side is bounded by the constant times the right hand side. We thank R2, have updated our paper to always use the C notation, and always describe C as a universal constant. R2 asks us to include a **better intuition for the Energy term in the introduction**, beyond what is on lines 33-35. We thank R2 and agree; we have now included a stronger mathematical intuition for the Energy term in the introduction.