We thank the reviewers for their kind and thoughtful comments on our work. We especially appreciate the supportive feedback regarding our experimental contributions. Below, we respond to reviewer-specific comments.

Reviewer 1

Our claim that a standard multilayer perceptron "fails to learn high frequencies in theory" is based on the theoretical connection between the MLP's convergence rate and the eigenvalues of the NTK matrix. Standard MLPs have an NTK with such extreme decay in high frequency eigenvalues [3,4] that training them to learn high frequency function components is impractical, and adding an appropriate Fourier feature mapping speeds up convergence by orders of magnitude. It is true that with unbounded training time, a standard MLP could eventually fit to high frequency components; we will revise the text to clarify that we are referring to convergence speed, not representational power in the limit of training time. For example, in the abstract, modifying "a standard MLP fails to learn high frequencies both in theory and in practice" to "a standard MLP has impractically slow convergence to high frequency signal components."

Reviewer 2

We suspect that the benefits of Fourier features may not be as dramatic in high dimensions, since it is harder to find a high-dimensional problem setting with dense observations where stationary kernels would be desirable. However, we agree that generalizations of this approach to higher dimensional data could be an exciting avenue for future work.

Reviewer 4

1. Analysis of accelerated convergence: Fourier features can be used to increase the eigenvalues of the NTK corresponding to higher frequencies, which directly determine convergence speed during training (Eqn. 4, taken from [1,2,15]). We show empirical examples of increasing high-frequency NTK eigenvalues in Figure 2. However, we do not analytically derive the eigenvalues of the NTK as a function of the Fourier features, which would be quite challenging (prior work has only been able to analytically derive NTK eigenvalues for shallow networks).

2. NTK approximation: You are correct that the NTK analysis is a statement about the limiting behavior of MLPs, in the limits of infinite width and infinitesimal SGD learning rate. However, we find that the convergence behavior of our trained MLPs in practice closely matches the predictions provided by the NTK theory for 1D problems. Figures 3b and 3d show that the loss curves predicted by NTK theory match those produced when training the corresponding MLPs.

In higher dimensions, we did not characterize to what degree the precise predictions of NTK theory transfer to MLPs; our focus was instead on using intuition from our 1D experiments to achieve high performance on these "real" higher-dimensional problems. A precise study of the applicability of NTK theory to MLPs in higher dimensions is an exciting avenue for future research.

3. 1D toy examples: Section 5 indeed contains intentionally toy 1D experiments in order to clearly illustrate frequency convergence effects and show that sparse Fourier features are sufficient (which is necessary in higher dimensions). We directly extend the 1D experiment in Figure 4 to a two-dimensional setting in Section 1.4 of the supplement, and demonstrate the same underfitting/overfitting phenomenon.

Clarity of Sections 4 and 5: Thanks for the suggestion. We agree these sections are very important to the flow of the paper and will revise the text to clarify the key takeaways.

More visualizations of results: Thanks for the suggestion. In addition to the 2D and 3D visual results in Section 6 of the appendix, we will also publish our code so that anyone can run experiments with our method on their own data.