We appreciate the reviewers' attention and thoughtful feedback. We are grateful that there was a consensus that our novel (nonasymptotic) Weyl law and its implications for kernel regression on a manifold represent an important theoretical contribution. Below, we respond to the major themes of the reviews. Several reviewers also provided more technical comments that, while we do not address them here, will be incorporated into our final version of the paper.

Practical relevance. We certainly acknowledge that we may not know the exact heat kernel in many practical situations. While there are means of estimating it from data, our treatment assumes that this is known *a priori*. We consider this paper to be an initial attempt to determine what is possible under ideal circumstances; methods which use alternative fixed or estimated kernels may not do as well, but we provide guidance as to what one might hope to achieve. We believe this provides a firm foundation for future work on manifold function estimation that may be able to address this gap.

Relation to prior work. As some reviewers noted (and as we mention in lines 185–190 and 244–249), our general result on kernel regression (Theorem 1) is similar to previous results such as those in [49] and [50], although our result has the advantage of being applicable to the noiseless case without regularization.

Much of the prior work, such as that in [35] (Caponetto and De Vito) and [40], primarily considers what power-law error rate (i.e., $\|\hat{f} - f^*\|_{L_2}^2 \lesssim n^{-\gamma}$) we can obtain if we are given power-law behavior of the kernel integral operator eigenvalues t_ℓ (or, equivalently, the regularized effective dimension $\operatorname{tr}(\mathcal{T} + \lambda \mathcal{I})^{-1} \mathcal{T})$ as $\lambda \downarrow 0$). Our result is **nonasymptotic** (unlike [35]) and provides **explicit constants** (unlike [35] and [40]). Furthermore, it can be applied to more general kinds of eigenvalue decay (such as the exponential decay that we find for the heat kernel). In the power-law case, when $t_\ell \approx \ell^{-b}$, an optimal choice of the dimension p in the bias and variance terms of our result recovers the standard minimax optimal error rate $\|\hat{f} - f^*\|_{L_2}^2 \lesssim n^{-b/(b+1)}$ that was proved in [35].

A key feature of our work is that the functions we consider are *much smoother* than in the classical models that many of the references consider (e.g., belonging to a Sobolev space, or, similarly, having bounded m^{th} derivative as in [24], which both lead to power-law error rates). Our results show that this smoothness is indeed exploited, as we obtain finite-dimensional error rates (within a logarithmic factor, in the heat kernel case), and the constants in the error and the sample complexity correspond clearly to the finite effective dimension of the function space that we analyze.

Optimality. Our results are certainly optimal (within multiplicative constants) in the case of bandlimited functions. It is well-established that to estimate a function in a p-dimensional subspace G, the (squared) L_2 error due to noise will always be of the order p/n, where n is the number of samples. The sample complexity $n \approx p \log p$ is also optimal, where $p \approx \Omega^m$ and m is the manifold dimension: considering the simple example of the torus T^m, estimating Ω-bandlimited functions requires that every point in the domain is within $O(1/\Omega)$ of a sample point; considering a grid over the domain (which has $O(\Omega^m)$ regions), the coupon collector problem suggests that we will need $O(\Omega^m \log \Omega^m) \approx p \log p$ points sampled uniformly at random to achieve a sufficiently fine sampling.

Additional comments. We assume in our paper that we sample points uniformly over the manifold. As we mention in lines 250–254, it might be an interesting extension to consider other sampling measures. However, the uniform sampling assumption requires the manifold to be compact. We also assume, for simplicity, that the manifold is without boundary (although much of this work could be extended to the case with boundary).

Regarding R3's 3$^{\text{rd}}$ question, our key upper bound on the heat kernel (Lemma 7 in Appendix B) only applies to *positive* upper bounds on sectional curvature. We agree with R3 that it would be very interesting to investigate what happens with a nonpositive upper bound. Regarding the second part of the question, since our nonasymptotic Weyl law (Theorem 2) is only an *upper* bound, its proof in Appendix C only requires an upper bound on the heat kernel, which in turn only requires an upper bound on sectional curvature. A corresponding lower bound in the Weyl law would likely require a corresponding lower bound on the curvature.

Regarding R4's question 1(b), the compactness assumption also ensures that the sectional curvature is uniformly (upper) bounded over the manifold.

Regarding R4's concern regarding the statement of Theorem 3, please see the continuation of the theorem statement onto the next page.

Finally, to clarify the comment on Theorem 4 in lines 291–292, this is intended to be an "alternative interpretation" of the theorem, where we choose a tolerance δ for the bias term. Because the p defined in line 292 only depends logarithmically on δ, we can choose δ to be very small without greatly affecting the "effective dimension" p.