Reply to Reviewer DENr

We thank the referee for the reviewing the manuscript and for emphasizing the importance of developing deep learning methods for multiscale partial differential equations.

Comment The major weakness of the paper is a lack of sufficient novelty. The fact that GD is difficult to learn oscillatory functions with high frequency is known in the literature; see Xu et al. 2020 for regression setting and other references mentioned by the authors. In the PDE setting, Wang et al. 2021 and Wang et. al 2022 also studied the convergence of training process as well as the spectral bias from an NTK perspective. The only difference between the present paper and these earlier works lies in the investigation of the blow-up of the norm of the NTK matrix as shown in Theorem 3.2 and Theorem 3.3. The proofs of both results follow from straightforward calculations of the NTK matrix and do not seem to require new technicalities or ideas.

Reply Thank you for pointing out a deficiency in the original draft of the manuscript, namely that there is no discussion of the similaries and differences between the present work and previous analyses of PINN failure modes.

The present work is certainly inspired by the series of papers by Wang et al. (SIAM J. Sci. Comp. 2021; Comput. Meth. Appl. Mech. Engr. 2021; J. Comput. Phys. 2022) who point out that, fundamentally, training PINNs involves optimizing a multi-objective loss functional; the PDE residual and boundary conditions residual (for stationary problems) need to simulataneously be minimized. Broadly speaking, PINNs can fail to train whenever these two terms in the loss function are imbalanced.

Previous analyses of such imbalances have focused on the case of Poisson-type boundary value problems where the Laplacian was the differential operator (or, in one dimension, the second derivative). Any multiscale features in the problem originated from oscillatory forcing functions, or "right-hand sides". In contrast, in the present work the multiscale nature of the problems considered originates from an oscillatory function *within in the differential operator itself* (which we term "Darcy-type" equations).

The differences between the two cases can be considerable, both in theory and practice, which is a key takeaway of the present work. For example, for the Darcy problem, the spectral radius of the K_{uu} matrix subblock scales like $1/\epsilon^2$, as pointed out in the new Corollary 3.4 and illustrated in Figure 2. For multiscale Poisson problems however, this matrix subblock is independent of ϵ, since the oscillatory forcing function f is independent of the network parameters θ.

We added a new section to the paper that discusses the theoretical differences in more detail; please refer to Section 3.2. We also expanded the numerical experiments to highlight the difference that can arise in practice; please refer to the new Section 4.3.

Comment Unfortunately, the present paper only points out the issue of PINNs and could be substantially improved if the authors were to propose new algorithm/methodology that can solve/alleviate the issue.

Reply We respectfully defer to alternative approaches for solving Darcy-type equations with neural network based approaches found e.g. in Han and Lee (Multiscale Model. Simul., 2023) or Leung et al. (J. Comput. Phys., 2022). As these involve considerable developments beyond the standard PINN methodology (in the latter work) or abandoning it entirely (in the former), we focus in the present work on shedding light on the strengths and weaknesses of existing PINNs techniques, and in particular explaining *why* those weaknesses are present for the class of equations considered here.