We are grateful to the reviewers for the insightful comments on our submission. Below we provide responses to reviewer's major comments. All the minor comments will also be addressed in the revised manuscript.

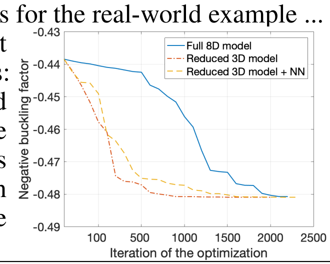
Reviewer 1: "the statement in line 153 ...... in the neighbourhood of $z \iff \langle J_i(z), \nabla f(x)\rangle = 0$."
Response: We appreciate the reviewer's comment and suggestion. We will update line 153 to "$f(x)$ *does not change with a perturbation of* z_i *in the neighborhood of* z*, if and only if* $\langle J_i(z), \nabla f(x)\rangle = 0$." Also, Eqn. (8) will be removed.

Reviewer 1: "in equation 12 and 13 f_1 and f_2 are defined ...... how is the domain of g being determined?"
Response: The loss function does not impose constraints on the domain of $\boldsymbol{z}$, which is why negative values of z_1 appear in Figure 2. As $\boldsymbol{z}$ does not have any physical meanings, it is unnecessary to force $\boldsymbol{z}$ to be in a pre-determined domain during the training. The domain of $\boldsymbol{z}$ can be easily adjusted by translation and dilation after the training process.

Reviewer 1: "emphasize the need for gradient evaluations when you state the observation."
Response: The observation statement (line 118-120) will be updated to "*For a fixed pair* $(\boldsymbol{x}, \boldsymbol{z})$ *satisfying* $\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{x})$*, if* $\boldsymbol{x} = \boldsymbol{g}^{-1}(\boldsymbol{z})$ *moves along a tangent direction,* ***i.e., any direction perpendicular to*** $\nabla f(\boldsymbol{x})$*, of the level set ......*".

Reviewer 1: "ambiguity of the notion of sensitivity (Figure 2 and below)". **Reviewer 4**: "...... for Section 4.1, Figure 2, the authors may want to mention clearly how one should tell from the plots which method is better ....."
Response: We agree with both reviewers that the caption of Figure 2 is not very clear. The caption will be updated to "*.......The first and fourth columns show the relationship between the output and* z_1*,* ***where the performance is better if the curve is thinner (i.e., the thickness of the curves shows the variation of*** $f \circ \boldsymbol{g}^{-1}$ ***w.r.t.*** z_2***).*** *The second and fifth columns show the gradient field (gray arrows) and the vector field of second Jabobian column* $\boldsymbol{J}_2$*,* ***where the performance is better if the gray and black arrows are perpendicular to each other.*** *The third and sixth columns ...*"

Reviewer 2:"Are training times of the NLL low enough and its accuracy high enough to satisfy practitioners?". **Reviewer 4**:"... add an experiment on optimizing the dimensionality reduced functions for the real-world example ... "
Response: To illustrate the significance of NLL to practitioners, we will add a new plot in §4.3 to show the decay of the objective function for the optimal design in 3 cases: (i) using the 8D FEM model, (ii) using the reduced 3D model, (iii) using the reduced 3D model + NN approximation. (i) v.s. (ii) shows the effectiveness of the NLL, i.e., the 3D optimization converges much faster than the 8D optimization; (ii) v.s. (iii) shows that the NN approximation is accurate enough to exploit the dimensionality reduction advantage. In addition, Case (iii) is much faster than Case (ii) because evaluating the NN is very efficient compared to the evaluating the FEM model in Case (ii).



Reviewer 4: "2. The authors may want to add a short paragraph (relatively early in the main text) on ......"
Response: We appreciate the reviewer's suggestion and will add the suggested examples to line 40 as "*... structures of level sets. For example, the existing methods for functions with linear level sets, e.g.,* $f(\boldsymbol{x}) = \sin(x_1 + x_2)$ *(the optimal linear transformation is a* 45° *rotation). When the level sets are nonlinear, e.g.,* $f(\boldsymbol{x}) = \sin(\|\boldsymbol{x}\|^2)$ *(the spherical transformation is optimal), the number of active input dimensions cannot be reduced by linear transformations.*"

Reviewer 4: "3. The mathematical setting in Section 2 is not very clear. ......"
Response: The independence assumption was used to emphasize that our method can deal with functions that have no intrinsic low-D structure in the input space. As independence is not a necessary condition for our method to work, we will remove such assumption, as well as make it clear that $\rho(\boldsymbol{x})$ is a user-specified distribution in the revised manuscript.

Reviewer 4: "4. In Section 2.1, line 87, the function h is not clearly defined ......"
Response: (i) h is an implicitly defined link function mapping from $\boldsymbol{z}$ to y, the composition of h and $\boldsymbol{A}$ is an approximation of f. (ii) The AS and SIR codes used for comparison to the NLL represent the state of the art of both methods. We did not use the SIR code with KDR, because the use of KDR will lead to *irreversible transformations* (i.e., $\boldsymbol{g}^{-1}(\boldsymbol{z})$ may not exist), such that the relationship between $\boldsymbol{z}$ and y may not be a function, i.e., one value of $\boldsymbol{z}$ may be associated with multiple y values. Those explanations will be made clear in the revised manuscript.

Reviewer 4: "5. In Section 3.1, the RevNet seems to require that u_n and v_n have the same dimension, ..."
Response: The used RevNet does require that u_n and v_n have the same dimension. When the dimension of $\boldsymbol{x}$ is odd, we can rewrite/extend $f(\boldsymbol{x})$ to $f(\boldsymbol{x}, x^*)$ by adding one dummy variable x^*. Since y does not really depend on x^*, we will have $\partial f/\partial x^* = 0$, such that the loss function does not impose any constraint on $\partial x^*/\partial z_i$ (the only constraint on x^* is imposed by the regularizer L_2). Even though the $d+1$-dimensional problem is generally harder to solve than the original d-dimensional problem, we do not think it will significantly affect the performance of NLL because the extension $f(\boldsymbol{x}, x^*)$ is not sensitive with respect to x^* from the beginning.

Reviewer 4: "6. ...... the way that the authors generate the training and validation/test sets are not very valid......"
Response: (i) §4.1 is to visualize the nonlinear capability of the NLL approach. To this end, we intended to remove any over-fitting effect by using a dense training set for clear illustration in the first and fourth columns in Figure 2. (ii) §4.2 is to show how the NLL helps alleviate the over-fitting issue, where the validation set with $10,000$ samples were generated uniformly in Ω. (iii) The training set for dimensionality reduction was reused in approximating the transformed function $\boldsymbol{z} \mapsto y$ using neural networks.

Reviewer 4: "7. ...... what are the normalization and derivatives w.r.t.? Is there any activation in the NN approximation?"
Response: The normalization constant is the maximum sensitivity index, such that the biggest sensitivity value is one in Figure 3. The partial derivatives are defined by $\partial y/\partial z_i$ for $i = 1, \dots, d$. Also, $\tanh()$ is used as the activation for NN.