Appendix A Gradients for Linear Approximation Cases

As an illustration, we consider the case where a linear approximation is adopted as the physics-based model, i.e., $f_\theta(\mathbf{z}) = \mathbf{A}\mathbf{z} + f_{\mathrm{NN}}(\mathbf{z})$, and the decoder Φ_p is simply an identity matrix of appropriate dimension. Following the derivation of gradients with repect to θ and $\mathbf{z}(t_0)$ given by [46], the gradients under the physics-informed regime can be expressed as the solution of the following differential equation:

$$\begin{aligned} \frac{d\mathbf{a}_{\mathrm{aug}}(t)}{dt} = -\begin{bmatrix} \frac{\partial f}{\partial \mathbf{z}} \\ \frac{\partial f}{\partial \theta} \end{bmatrix} \mathbf{a}(t) & = -\begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix} \mathbf{a}(t) - \begin{bmatrix} \frac{\partial f_{\mathrm{NN}}}{\partial \mathbf{z}} \\ \frac{\partial f_{\mathrm{NN}}}{\partial \theta} \end{bmatrix} \mathbf{a}(t) \\ & = -\mathbf{A}_{\mathrm{aug}} \mathbf{a}_{\mathrm{aug}}(t) - \begin{bmatrix} \frac{\partial f_{\mathrm{NN}}}{\partial \mathbf{z}} & \mathbf{0} \\ \frac{\partial f_{\mathrm{NN}}}{\partial \theta} & \mathbf{0} \end{bmatrix} \mathbf{a}_{\mathrm{aug}}(t), \end{aligned} \tag{A1}$$

where $\mathbf{A}_{\mathrm{aug}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, $\mathbf{a}_{\mathrm{aug}} = \begin{bmatrix} \mathbf{a} \\ \mathbf{a}_\theta \end{bmatrix}$, $\mathbf{a}(t) = \frac{dL}{d\mathbf{z}(t)}$, $\mathbf{a}_\theta(t) = \frac{dL}{d\theta(t)}$. Then the gradients can be approximately given by

$$\mathbf{a}_{\mathrm{aug}}(t_0) = \mathbf{a}_{\mathrm{aug}}(t_1)\exp(\mathbf{A}_{\mathrm{aug}}\Delta_t) + F_{\mathrm{NN}}(t_1), \tag{A2}$$

where the first term is an approximate solution obtained from linear physics-based portion of A1 and the second term F_{NN} accounts for the difference between the linear approximation and the true solution. Suppose the training time steps are $t = t_0, t_1, ..., t_N$, then we can repeat the process and the gradient obtained through back-propagation is:

$$\mathbf{a}_{\mathrm{aug}}(t_0) = \mathbf{a}_{\mathrm{aug}}(t_N)\exp(\mathbf{A}_{\mathrm{aug}} N\Delta_t) + \sum_{i=1}^{N} F_{\mathrm{NN}}(t_i). \tag{A3}$$

The first term of the R.H.S. is brought by the linearized physics-based model and it can be directly back-propagated, while only the discrepancy terms $\sum_{i=1}^{N} F_{\mathrm{NN}}(t_i)$ need to be estimated, which makes the estimated gradients also an approximation to the real ones. As a result from this, the combined gradients are restricted in a regime that is closer to the true function's gradients.