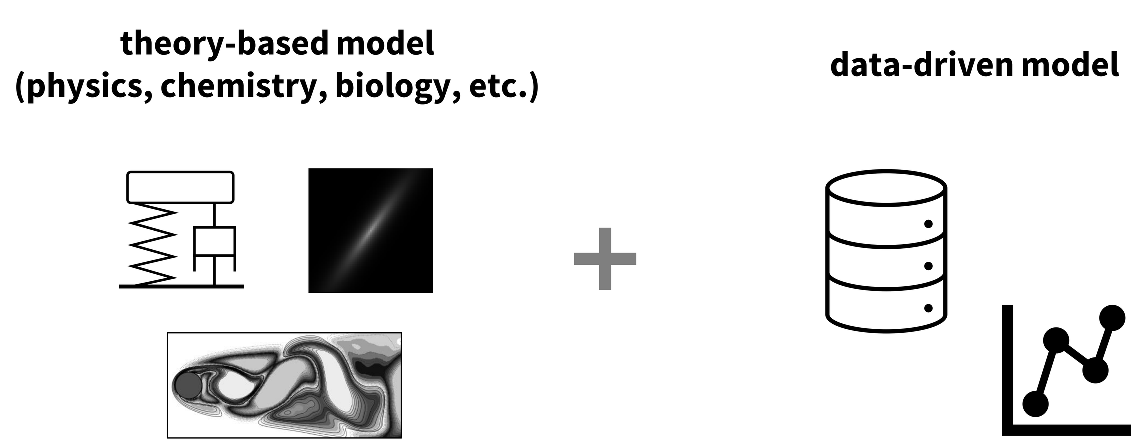
Physics-Integrated Variational Autoencoders for Robust and Interpretable Generative Modeling

Naoya Takeishi, Alexandros Kalousis (University of Applied Sciences and Arts Western Switzerland, HES-SO)

Hes·so GENÈVE
University of Applied Sciences and Arts Western Switzerland

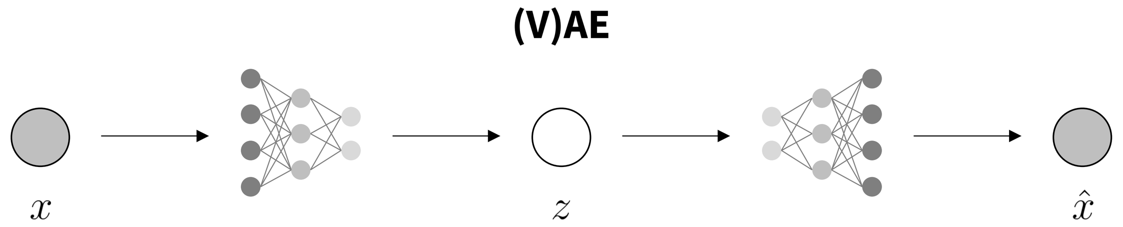
We address grey-box modeling with deep generative models, particularly VAEs. Technical challenge is to strike a balance between theory-based mathematical (physics) models and data-driven models (i.e., neural nets). We propose a set of regularizers to prevent a physics model from being ignored and to ground the semantics of a part of latent variables by an incomplete physics model.

Deep grey-box modeling needs some regularization



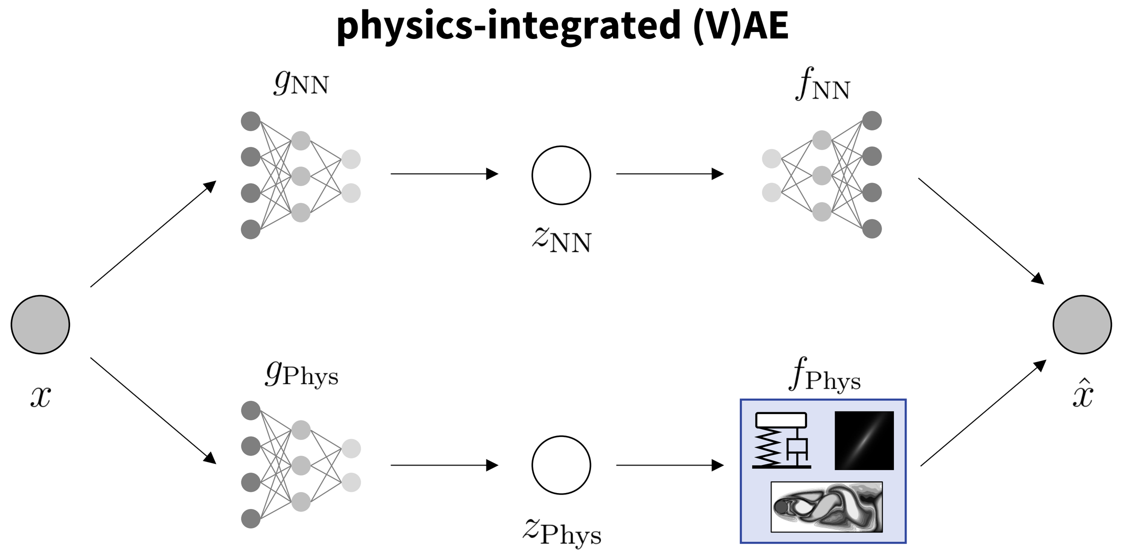
Combining mathematical models based on scientific theories & data-driven models is known as grey-box modeling and can be advantageous in terms of interpretability, extrapolation, and learning efficiency. We consider grey-box modeling with deep neural networks being data-driven part. Particularly, we suppose that VAE's decoder comprises both an incomplete mathematical (physics) model and a neural net (**physics-integrated VAE**).

(V)AE



$\mathbb{E}z = g(x)$ $\mathbb{E}\hat{x} = f(z)$

physics-integrated (V)AE



$\mathbb{E}z_{\mathrm{NN}} = g_{\mathrm{NN}}(x)$

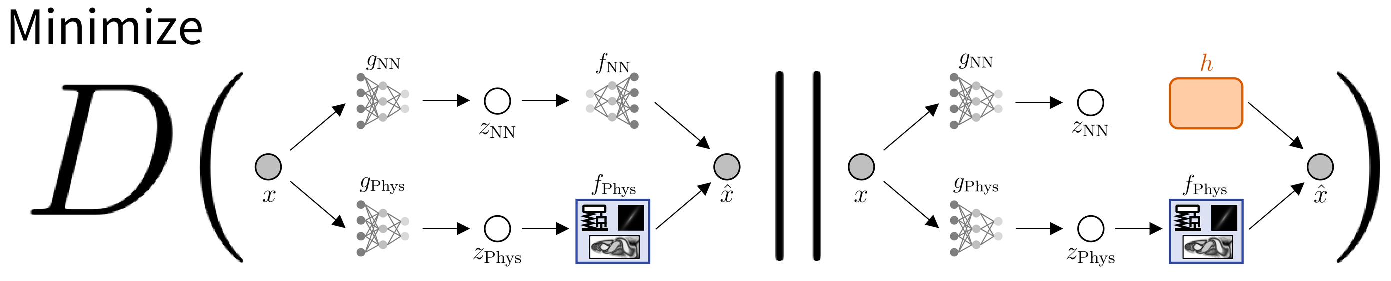
$\mathbb{E}z_{\mathrm{Phys}} = g_{\mathrm{Phys}}(x)$

$\mathbb{E}\hat{x} = \mathcal{F}[f_{\mathrm{NN}}, f_{\mathrm{Phys}}; z_{\mathrm{NN}}, z_{\mathrm{Phys}}]$

Issue: simply by ERM with –ELBO, physics model can be ignored ☹

Regularization 1: Suppressing excess flexibility of NN

Minimize



original model

$\mathbb{E}z_{\mathrm{NN}} = g_{\mathrm{NN}}(x)$

$\mathbb{E}z_{\mathrm{Phys}} = g_{\mathrm{Phys}}(x)$

$\mathbb{E}\hat{x} = \mathcal{F}[f_{\mathrm{NN}}, f_{\mathrm{Phys}}; z_{\mathrm{NN}}, z_{\mathrm{Phys}}]$

$p(\hat{x} \mid X_{\mathrm{data}})$

"physics-only" reduced model

$\cancel{\mathbb{E}z_{\mathrm{NN}} = g_{\mathrm{NN}}(x)}$

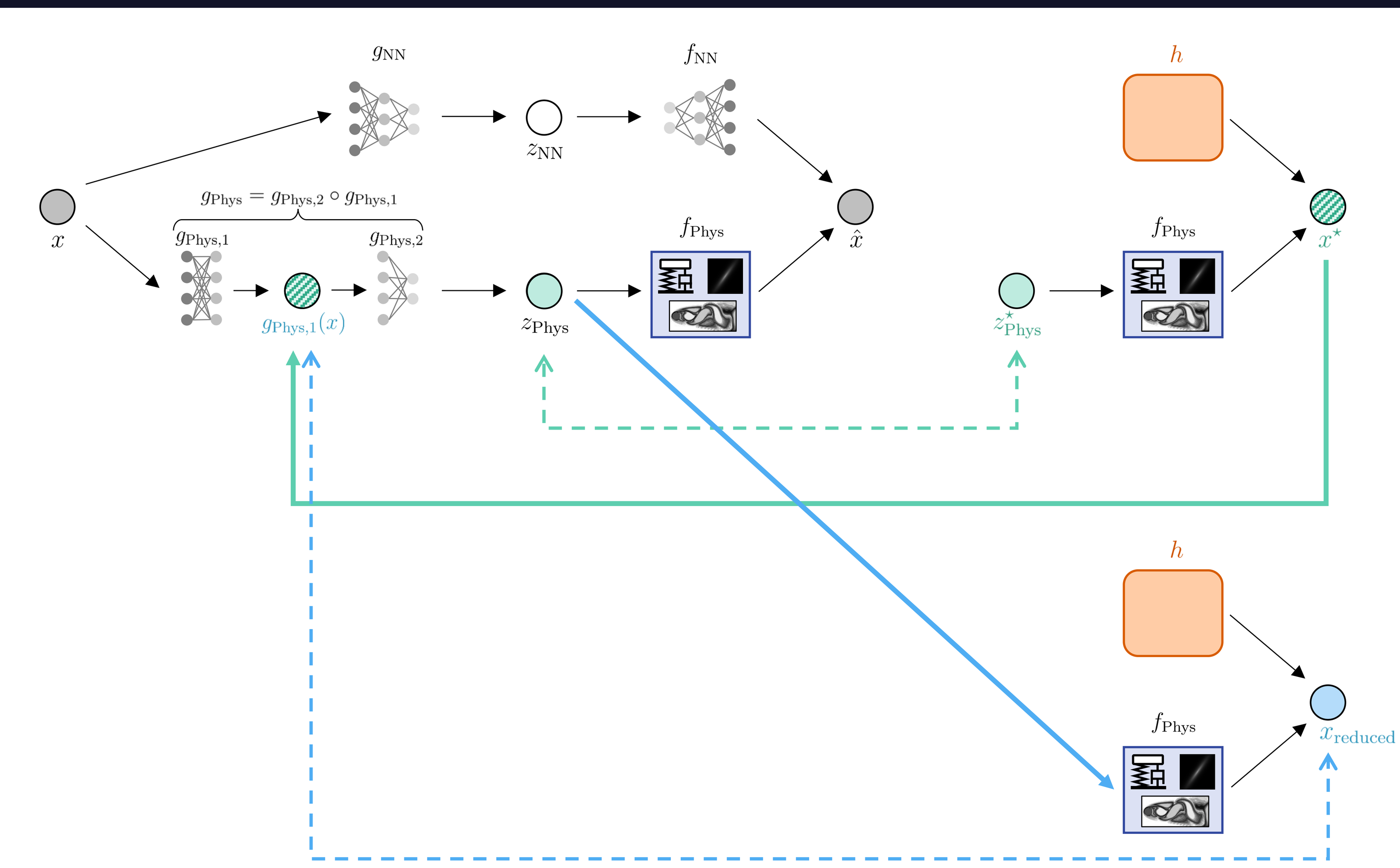
$\mathbb{E}z_{\mathrm{Phys}} = g_{\mathrm{Phys}}(x)$

$\mathbb{E}\hat{x} = \mathcal{F}[\overset{h}{\cancel{f_{\mathrm{NN}},}} f_{\mathrm{Phys}}; \cancel{z_{\mathrm{NN}},} z_{\mathrm{Phys}}]$

$p_{\mathrm{reduced}}(\hat{x} \mid X_{\mathrm{data}})$

Suppress excess flexibility of neural net f_{NN} by minimizing difference between the full model (left) and a "physics-only" reduced model (right), which is created by replacing f_{NN} with some simple replacement function h (e.g., $h(z) = \mathrm{Id}, h(z) = Wz$).

Regularization 2: Data augmentation by physics model



Ground the semantics of z_{Phys} with self-supervision by f_{Phys}. We can generate artificial data x^* from f_{Phys}, but it cannot be put into g_{Phys} directly because f_{Phys} is incomplete, and thus x^* would have different nature from that of real data x.

We try to avoid such difficulty by making the first stage of g_{Phys} output "physics-only" version of x. Let us consider two-stage structure $g_{\mathrm{Phys}} = g_{\mathrm{Phys},2} \circ g_{\mathrm{Phys},1}$. We make $g_{\mathrm{Phys},1}(x)$ and x_{reduced} close each other, where x_{reduced} is created by giving z_{Phys} to the physics-only decoder.

Minimize $\mathbb{E}_{\mathrm{data}} \| g_{\mathrm{Phys},1}(x) - \mathrm{StopGrad}[x_{\mathrm{reduced}}] \|_2^2$ and $\mathbb{E}_{z^\star_{\mathrm{Phys}}} \| g_{\mathrm{Phys},2}(\mathrm{StopGrad}[x^\star]) - z^\star_{\mathrm{Phys}} \|_2^2$

Experiments

Methods compared:

`NN-only`	vanilla VAE, decoder only with NN
`Phys-only`	decoder only with physics model
`NN+solver`	decoder with NN & corresponding solver (e.g., neural ODEs)
`NN+phys`	decoder with NN & physics model, but without proposed regularization
`NN+phys+reg`	proposed method

Pendulum

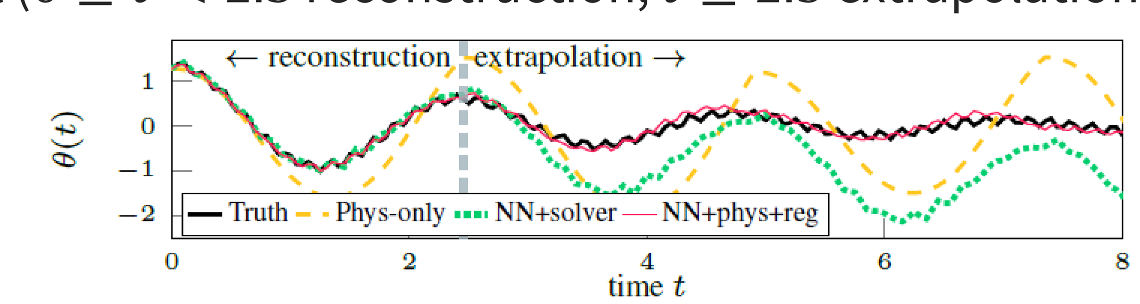
Data generated from: $\ddot{\vartheta} + \omega^2 \sin\vartheta + b\dot{\vartheta} - A\omega^2 \cos(2\pi\phi t) = 0$; 1000 sequences for training.

Model is physics-integrated VAE with decoder:

$$f_{\mathrm{NN},2}\left(\mathrm{ODESolve}_\vartheta\left[f_{\mathrm{Phys}}(\vartheta, z_{\mathrm{Phys}}) + f_{\mathrm{NN},1}(\vartheta, z_{\mathrm{NN}}) = 0 \right], z_{\mathrm{NN}}\right)$$

where $f_{\mathrm{Phys}}(\vartheta, z_{\mathrm{Phys}}) = \ddot{\vartheta} + z_{\mathrm{Phys}}^2 \sin\vartheta$

Result ($0 \le t < 2.5$ reconstruction, $t \ge 2.5$ extrapolation):



Result (reconstruction / inference errors):

		Pendulum				Advection-diffusion			
		MAE of reconst.		MAE of inferred ω		MAE of reconst.		MAE of inferred a	
`NN-only`		0.438	(2.9×10^{-2})	–		0.0396	(2.2×10^{-4})	–	
`Phys-only`		1.55	(7.1×10^{-4})	0.232	(5.9×10^{-3})	0.393	(9.5×10^{-4})	0.0103	(1.5×10^{-3})
`NN+solver`		0.439	(2.3×10^{-2})	–		0.0388	(1.7×10^{-4})	–	
`NN+phys`		0.370	(4.3×10^{-2})	1.04	(2.2×10^{-1})	0.0404	(1.2×10^{-2})	0.258	(3.2×10^{-1})
`NN+phys+reg`		0.363	(4.8×10^{-2})	0.229	(3.8×10^{-2})	0.0437	(1.5×10^{-3})	0.00951	(6.2×10^{-3})
Ablations	$\alpha = 0$	0.396	(4.3×10^{-2})	0.889	(1.9×10^{-1})	0.0461	(1.3×10^{-2})	0.0444	(1.4×10^{-2})
	$\beta = 0$	0.372	(4.1×10^{-2})	0.223	(3.6×10^{-2})	0.0747	(2.4×10^{-2})	0.199	(2.3×10^{-1})
	$\gamma = 0$	0.381	(4.1×10^{-2})	0.276	(4.2×10^{-2})	0.0588	(9.1×10^{-4})	0.0548	(9.4×10^{-7})

Galaxy images

Data comprise a certain class of galaxy; 400 images for training.

Physics-integrated VAE comprises Gaussian light profile & U-net.

Result (random generation):