Deep learning based reduced order modeling of seismogram-type acceleration time series model: Part - III

Saumik Dana[1*] and Karthik Reddy Lyathakula[2]

[1*]University of Southern California, Los Angeles, CA, USA.
[2]North Carolina State University, Raleigh, NC, USA.

*Corresponding author(s). E-mail(s): saumik@utexas.edu;
Contributing authors: klyatha@ncsu.edu;

Abstract

The optimal deep learning optimizer can vary depending on the features of the solution of the forward model being trained. The other parameter that can acquire significance in LSTM based encoder decoder frameworks is the number of stacked LSTM layers per encoder and decoder. In this work, we zero-in on the optimal optimizer and number of layers for the reconstruction of the acceleration time series solution of the spring slider damper idealization of a friction model.

Keywords: Time series, LSTM, encoder, decoder, reduced order model

1 Introduction

Deep learning is a practising science form, and the optimal parameters for a problem are obtained with a combination on intuition and experience. We attempt to build reduced order models of complicated forward models using LSTM encoder decoder frameworks, and we observe reconstruction results for different optimizers and batch sizes. The big picture is to train the model, and then used this model as a reduced order model to evaluated inside the Bayesian inference framework to estimate certain parameter(s) of the forward model, as shown in Fig. 1. The forward model is a spring slider damper idealization of the rate and state model for fault friction, and that itself is a piece of a

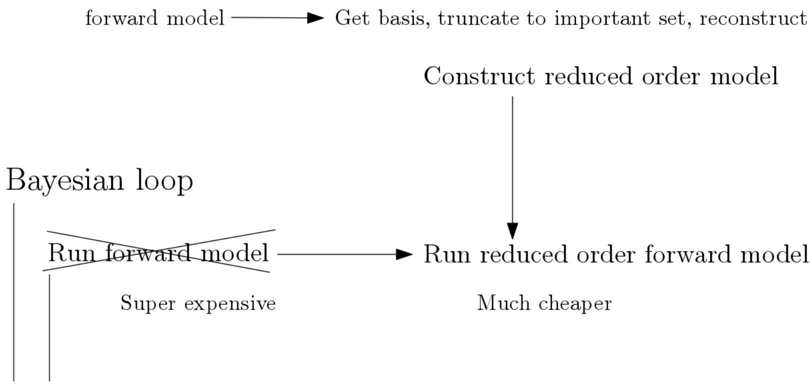
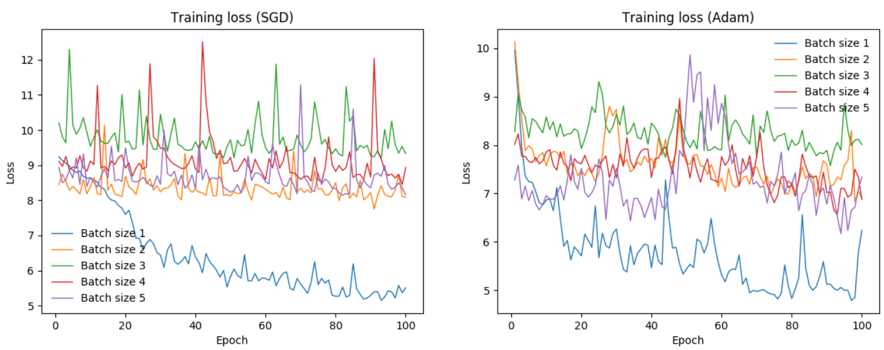


Fig. 1: The big picture

sophisticated coupled flow and geomechanics model [1–9]. The details of the forward model for this work are given in Appendix A. There are three things we are concerned with: the optimizer, the batch size, and the number of LSTM layers which make up both the encoder and decoder.

2 Reconstruction results

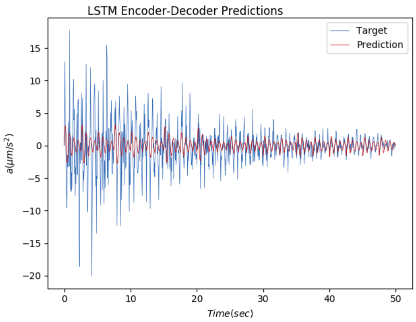


(a) SGD

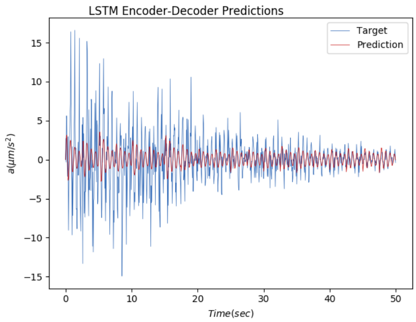
(b) Adam

Fig. 2: Number of LSTM layers is 1

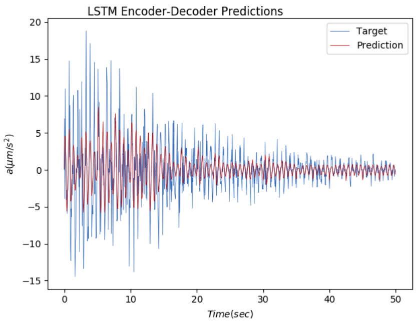
The code was implemented in PyTorch, and run on a basic AMD Ryzen 3 3200U with Radeon Vega Mobile Gfx $\times$ 4 processor. We compare the training losses for the Adam and Stochastic gradient descent optimizers for different batch sizes in Fig. 2. Typically, the loss is least when the batch size is 1, since the weight gradient is updated after every sample. In realistic problems, a batch size of 1 makes the training process incredibly slow because of the size



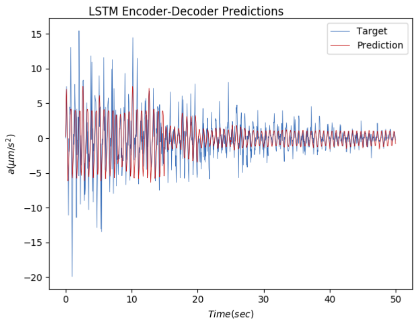
(a) SGD



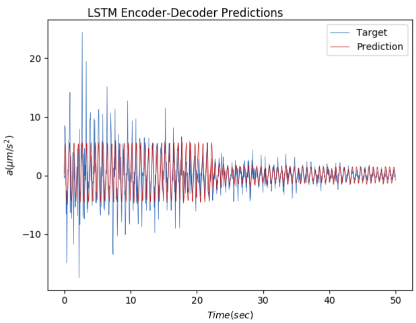
(b) ASGD



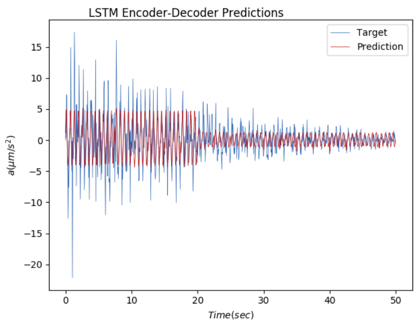
(c) RMSProp



(d) Adam



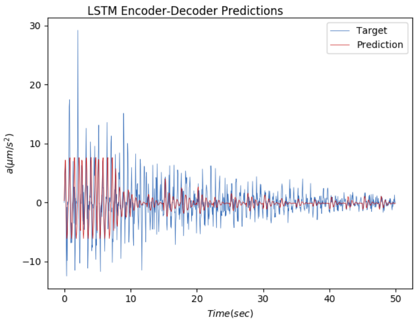
(e) Adamax



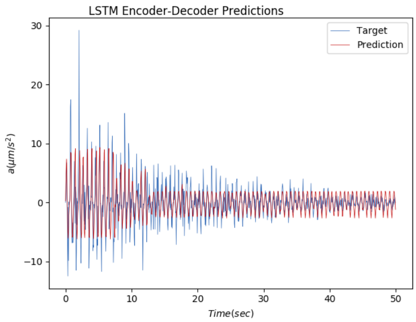
(f) AdamW

Fig. 3: Number of LSTM layers is 1, batch size is 1

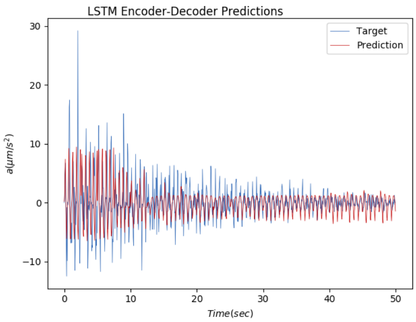
of the data, and making a pass through the network with one sample at a time will make the process infeasible. So, it always makes sense to use a batch size which is large enough to allow fast training, but small enough to factor in enough accuracy in the specified number of epochs. Before getting into that, it makes sense to compare the reconstruction for different optimizers. The data is definitely noisy (which we have ensured after adding noise to the forward model



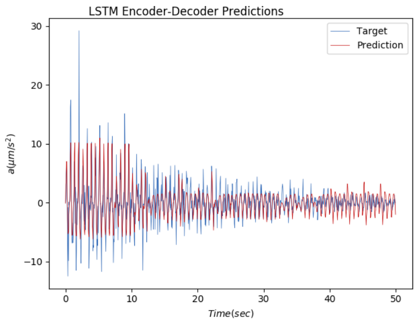
(a) Batch size 1



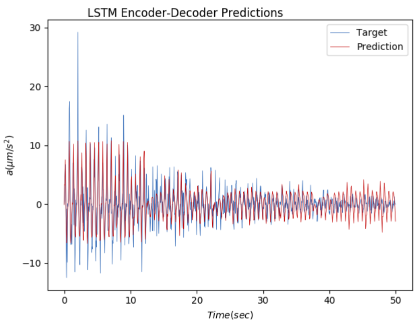
(b) Batch size 2



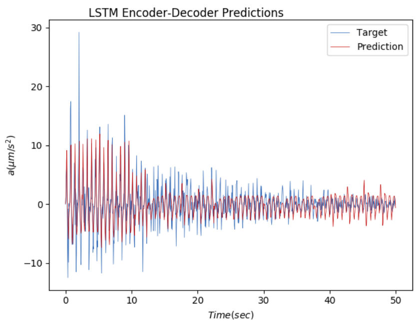
(c) Batch size 3



(d) Batch size 4



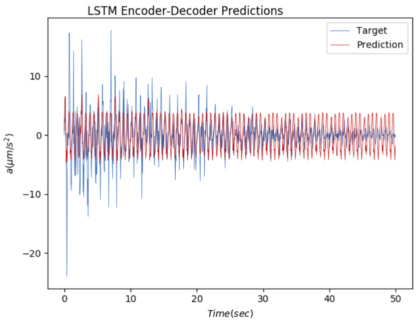
(e) Batch size 5



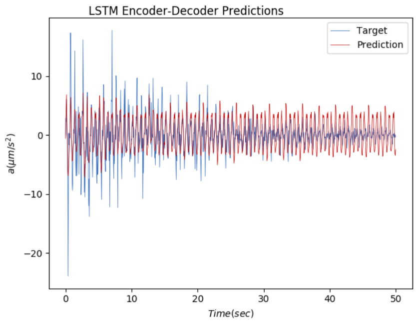
(f) Batch size 6

Fig. 4: Adam optimizer, Number of LSTM layers is 1

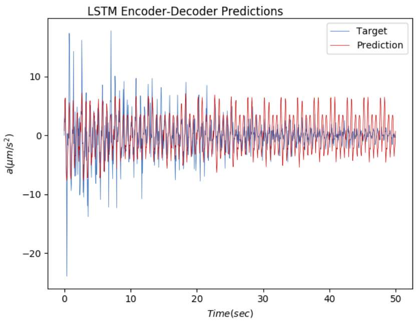
solution), and has the feature that it dies down over a period of time. From Fig. 3, it is a toss-up between RMSProp and Adam as to which is better for the problem, and it is interesting to note that Adamax and AdamW variants of Adam have these boxy patterns in terms of capturing the amplitudes. Now, knowing that we will eventually switch to batch size more than 1 for the realistic models, we compare the reconstruction for different batch sizes with



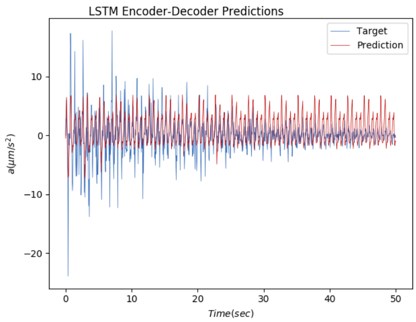
(a) Number of LSTM layers is 1



(b) Number of LSTM layers is 2



(c) Number of LSTM layers is 3



(d) Number of LSTM layers is 4

Fig. 5: Adam optimizer, Batch size is 5

Adam optimizer as shown in Fig. 4. We observe that increasing the batch size does deteriorate the reconstruction when the noisy solution dies down, but does fairly well when the acceleration is substantial. Finally, we compare the reconstruction for different number of LSTM layers for a batch size of 5, and we observe that this particular parameter makes minimal difference to the reconstruction, so we stick to 1 LSTM layer per encoder and decoder in future studies.

Appendix A Forward model details

The quantification of fault slip is achieved using the rate- and state-friction (RSF) model for modeling earthquake cycles on faults [10–15], and given by

$$\begin{aligned} \mu = \mu_0 + A \ln\left(\frac{V}{V_0}\right) + B \ln\left(\frac{V_0 \theta}{d_c}\right), \\ \frac{d\theta}{dt} = 1 - \frac{\theta V}{d_c} \end{aligned} \tag{A1}$$

where $V = |\frac{d\boldsymbol{d}}{dt}|$ is the slip rate magnitude, $a = \frac{dV}{dt}$ which we hypothesize is of the same order as recorded by seismograph, μ_0 is the steady-state friction

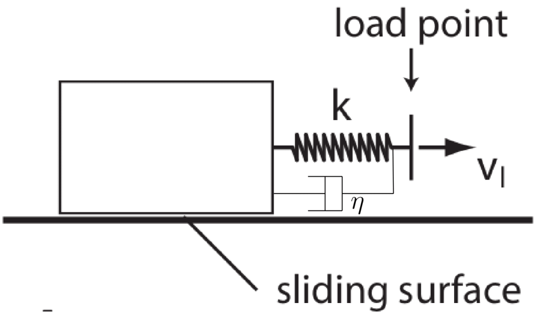


Fig. A1: Spring slider damper idealization of fault behavior

coefficient at the reference slip rate V_0, A and B are empirical dimensionless constants, θ is the macroscopic variable characterizing state of the surface and d_c is a critical slip distance over which a fault loses or regains its frictional strength after a perturbation in the loading conditions [16]. As shown in Fig. A1, we model a fault by a slider spring system [17–19]. The friction coefficient of the block is given by

$$\mu = \frac{\tau}{\sigma} = \frac{\tau_l - k\delta - \eta V}{\sigma} \tag{A2}$$

where σ is the normal stress, τ the shear stress on the interface, τ_l is the remotely applied stress acting on the fault in the absence of slip, -$k\delta$ is the stress relaxation due to fault slip [20] and η is the radiation damping coefficient [21]. We consider the case of a constant stressing rate $\dot{\tau_l} = kV_l$ where V_l is the load point velocity. The stiffness is a function of the fault length l and elastic modulus E as $k \approx \frac{E}{l}$. With $k' = \frac{E}{l\sigma}$, we get

$$\dot{\mu} \approx k'(V_l - V) - k''\dot{V} \tag{A3}$$

where $k'' = \frac{\eta}{\sigma}$. Once the phenomenological form of $\dot{\mu}$ is known, we rewrite Eq. (A1) as

$$\begin{gathered} V = V_0 \exp\left(\tfrac{1}{A}\left(\mu - \mu_0 - B\ln\left(\tfrac{V_0\theta}{d_c}\right)\right)\right), \\ \dot{\theta} = 1 - \tfrac{\theta V}{d_c} \end{gathered} \tag{A4}$$

to get acceleration time series as $a \equiv \dot{V}$ as,

$$\begin{aligned} \dot{V} &= \frac{V}{A}\left(\dot{\mu} - \frac{B}{\theta}\dot{\theta}\right) \\ &= \frac{V_0 \exp\left(\frac{1}{A}\left(\mu - \mu_0 - B\ln\left(\frac{V_0\theta}{d_c}\right)\right)\right)}{A} \end{aligned}$$

$$* \left(k' \left(V_l - V_0 \exp\left(\frac{1}{A} \left(\mu - \mu_0 - B \ln\left(\frac{V_0 \theta}{d_c} \right) \right) \right) \right) - k'' \dot{V} - \frac{B}{\theta} \left(1 - \frac{\theta V}{d_c} \right) \right) \tag{A5}$$

The ballpark values are:
✓ Elastic modulus $E = 5 \times 10^{10}\, Pa$
✓ Critical fault length $l = 3 \times 10^{-2}\, m$
✓ Normal stress $\sigma = 200 \times 10^6 Pa$
✓ Radiation damping coefficient $\eta = 20 \times 10^6 Pa/(m/s)$
✓ $A = 0.011$
✓ $B = 0.014$
✓ $V_0 = 1\mu m/s$
✓ $\theta_0 = 0.6$
✓ $\mu_0 = \mu_0 = 0.6$
✓ $t_{start} = 0,\ t_{end} = 50\, s,\ dt = 0.05\, s$
from which the effective stiffness and damping are obtained as
✓ $k' = \frac{E}{l\sigma} = \frac{5\times10^{10}}{3\times10^{-2}\times2\times10^{8}}[1/m] \approx 10^{-2}[1/\mu m]$
✓ $k'' = \frac{\eta}{\sigma} = \frac{2\times10^{7}}{2\times10^{8}}[s/m] = 10^{-7}[s/\mu m]$

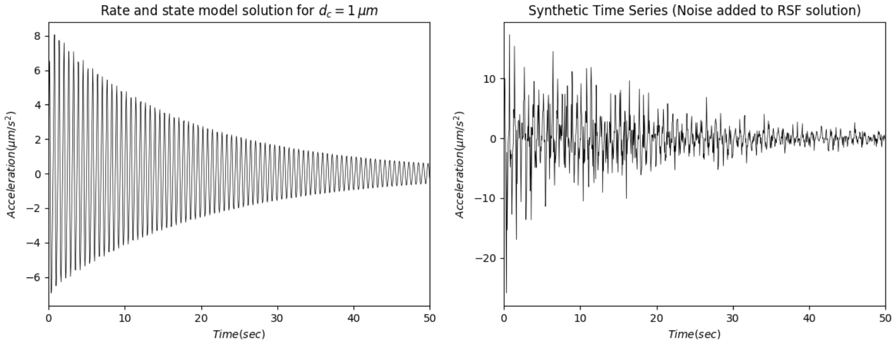


Fig. A2: System response

The influence of critical slip distance on system response to a load point perturbation of the form

$$V_l = V_0(1 + \exp(-t/20)\, sin(10t)) \tag{A6}$$

is shown in Fig. A2.

References

[1] Dana, S., Reddy, K.: Bayesian inference and markov chain monte carlo based estimation of a geoscience model parameter. arXiv preprint arXiv:2201.01868 (2021)

[2] Dana, S., Lyathakula, K.R.: Uncertainty quantification in friction model for earthquakes using bayesian inference. arXiv preprint arXiv:2104.11156 (2021)

[3] Dana, S., Wheeler, M.F.: Convergence analysis of fixed stress split iterative scheme for anisotropic poroelasticity with tensor biot parameter. Computational Geosciences **22**(5), 1219–1230 (2018)

[4] Dana, S., Wheeler, M.F.: Convergence analysis of two-grid fixed stress split iterative scheme for coupled flow and deformation in heterogeneous poroelastic media. Computer Methods in Applied Mechanics and Engineering **341**, 788–806 (2018)

[5] Dana, S., Ganis, B., Wheeler, M.F.: A multiscale fixed stress split iterative scheme for coupled flow and poromechanics in deep subsurface reservoirs. Journal of Computational Physics **352**, 1–22 (2018)

[6] Dana, S., Srinivasan, S., Karra, S., Makedonska, N., Hyman, J.D., O'Malley, D., Viswanathan, H., Srinivasan, G.: Towards real-time forecasting of natural gas production by harnessing graph theory for stochastic discrete fracture networks. Journal of Petroleum Science and Engineering **195**, 107791 (2020)

[7] Dana, S., Jammoul, M., Wheeler, M.F.: Performance studies of the fixed stress split algorithm for immiscible two-phase flow coupled with linear poromechanics. Computational Geosciences, 1–15 (2021)

[8] Dana, S., Ita, J., Wheeler, M.F.: The correspondence between voigt and reuss bounds and the decoupling constraint in a two-grid staggered algorithm for consolidation in heterogeneous porous media. Multiscale Modeling & Simulation **18**(1), 221–239 (2020)

[9] Dana, S.: Addressing challenges in modeling of coupled flow and poromechanics in deep subsurface reservoirs. PhD thesis, The University of Texas at Austin (2018)

[10] Ruina, A.L.: Slip instability and state variable friction laws. Geophys. Res. Lett. **88**, 359–370 (1983)

[11] Scholz, C.H.: Mechanics of faulting. Ann. Rev. Earth Planet. Sci. **17**, 309–334 (1989)

[12] Marone, C.: Laboratory-derived friction laws and their application to seismic faulting. Ann. Rev. Earth Planet. Sci. **26**, 643–696 (1998)

[13] Wei, M., Shi, P.: Synchronization of earthquake cycles of adjacent segments on oceanic transform faults revealed by numerical simulation in

the framework of rate-and-state friction. Journal of Geophysical Research: Solid Earth **126**(1), 2020–020231 (2021)

[14] Jia, Y., Tang, J., Lu, Y., Lu, Z.: The effect of fluid pressure on frictional stability transition from velocity strengthening to velocity weakening and critical slip distance evolution in shale reservoirs. Geomechanics and Geophysics for Geo-Energy and Geo-Resources **7**(1), 1–13 (2021)

[15] Weiqiang, Z., Allison, K., Dunham, E., Yang, Y.: Fault valving and pore pressure evolution in simulations of earthquake sequences and aseismic slip. Nature communications **11**, 4833 (2020)

[16] Palmer, A.C., Rice, J.R.: The growth of slip surfaces in the progressive failure of over-consolidated clay. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **332**(1591), 527–548 (1973)

[17] Rice, J.R., Gu, J.-c.: Earthquake aftereffects and triggered seismic phenomena. Pure and Applied Geophysics **121**(2), 187–219 (1983)

[18] Gu, J.-C., Rice, J.R., Ruina, A.L., Simon, T.T.: Slip motion and stability of a single degree of freedom elastic system with rate and state dependent friction. Journal of the Mechanics and Physics of Solids **32**(3), 167–196 (1984)

[19] Dieterich, J.H.: Earthquake nucleation on faults with rate-and state-dependent strength. Tectonophysics **211**(1-4), 115–134 (1992)

[20] Kanamori, H., Brodsky, E.E.: The physics of earthquakes. Reports on Progress in Physics **67**(8), 1429 (2004)

[21] McClure, M.W., Horne, R.N.: Investigation of injection-induced seismicity using a coupled fluid flow and rate/state friction model. Geophysics **76**(6), 181–198 (2011)