Accelerating Wound Healing using Deep Reinforcement Learning

Fan Lu*[1], Ksenia Zlobina[1], Nicholas A. Rondoni[1], Sam Teymoori[1], and Marcella Gomez[1]

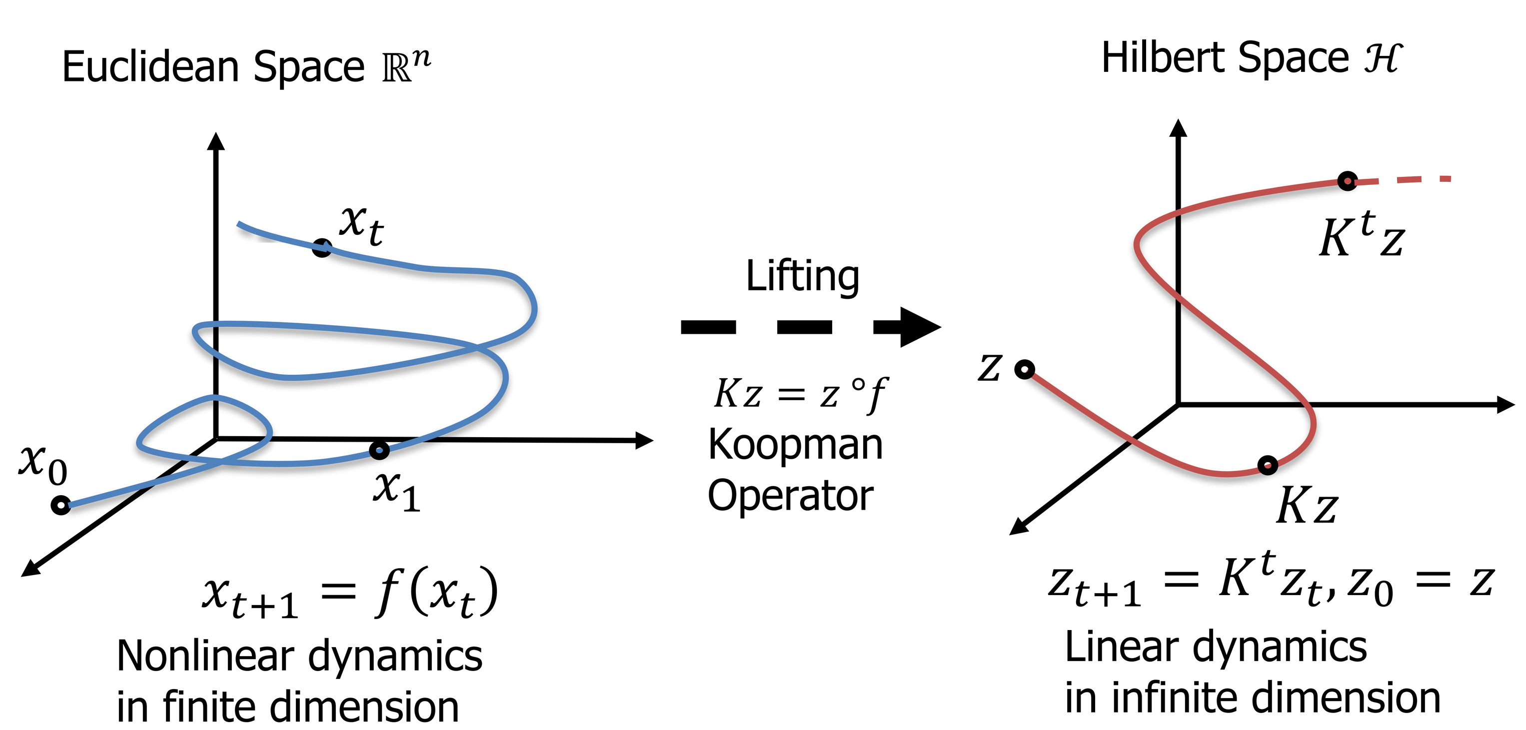
[1]The Department of Applied Mathematics, University of California Santa Cruz

Motivations

- Finding optimal treatments for wound healing is important
- Challenges in applying Deep Reinforcement Learning (DRL) to biology:
  - Exploration should be constrained: ethical and safety concerns
  - We need guidance that help incrementally increase the performance of target policy
- Wound dynamics are intricate and nonlinear, making mathematical modeling hard
- Getting optimal controls for nonlinear dynamics is difficult
- Theory in linear models is well-developed

Objective: Design a DRL-based algorithm for accelerating wound healing while balancing the trade-off between exploration and exploitation

Learning Linear Representation Example: Koopman Operator Theory



Challenges:

- The optimization lies in function space, rendering it intractable in practical applications
- It is often hard to account for the effect of inputs and control in nonlinear systems
- Generalizations consider the control effect using deep learning, but still heavily rely on overfitting the models without incrementally and adaptively online learning

DeepMapper

Chain Rule: $\frac{dz}{dt} = \frac{dz}{dx}\frac{dx}{dt} = A^\omega z + J_\theta(x)Bu$

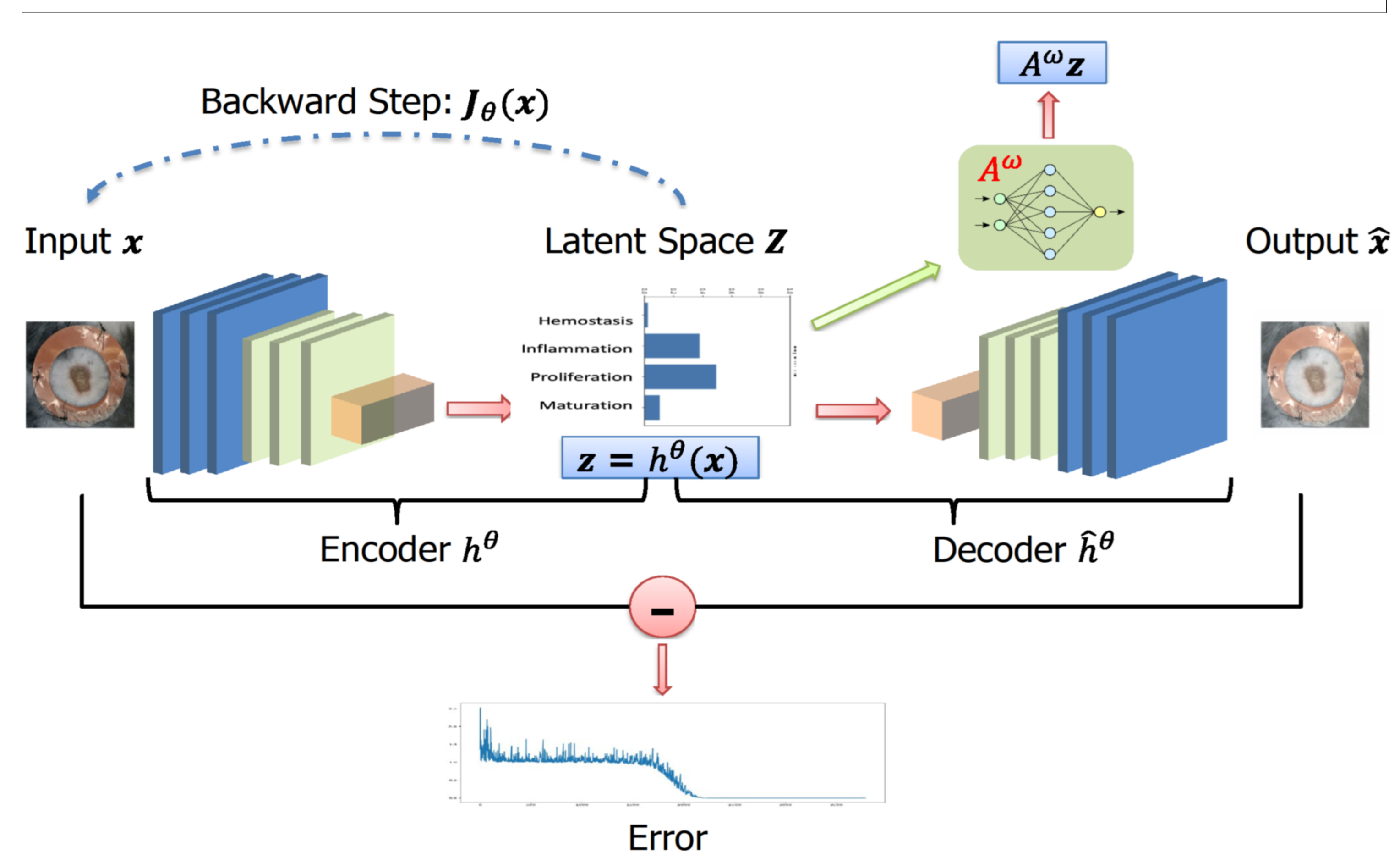
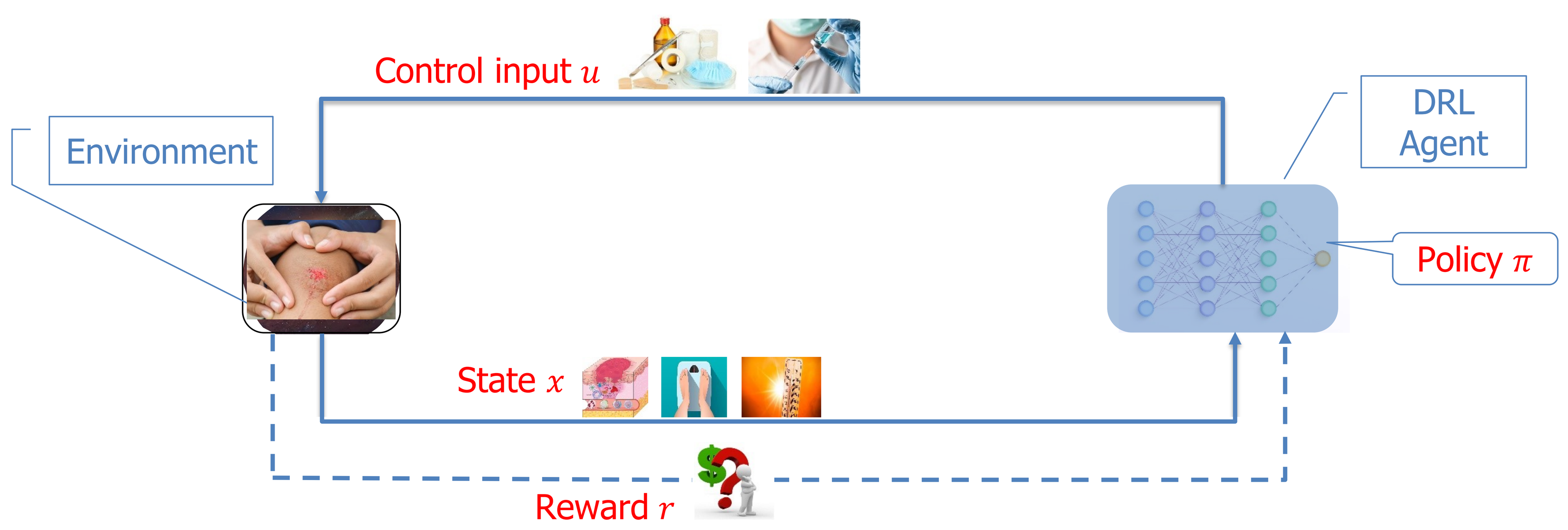


Figure: A self-adaptive linearity abstraction for learning the linear representation of latent space in Autoencoders for nonlinear wound healing dynamics

Deep Q-Learning

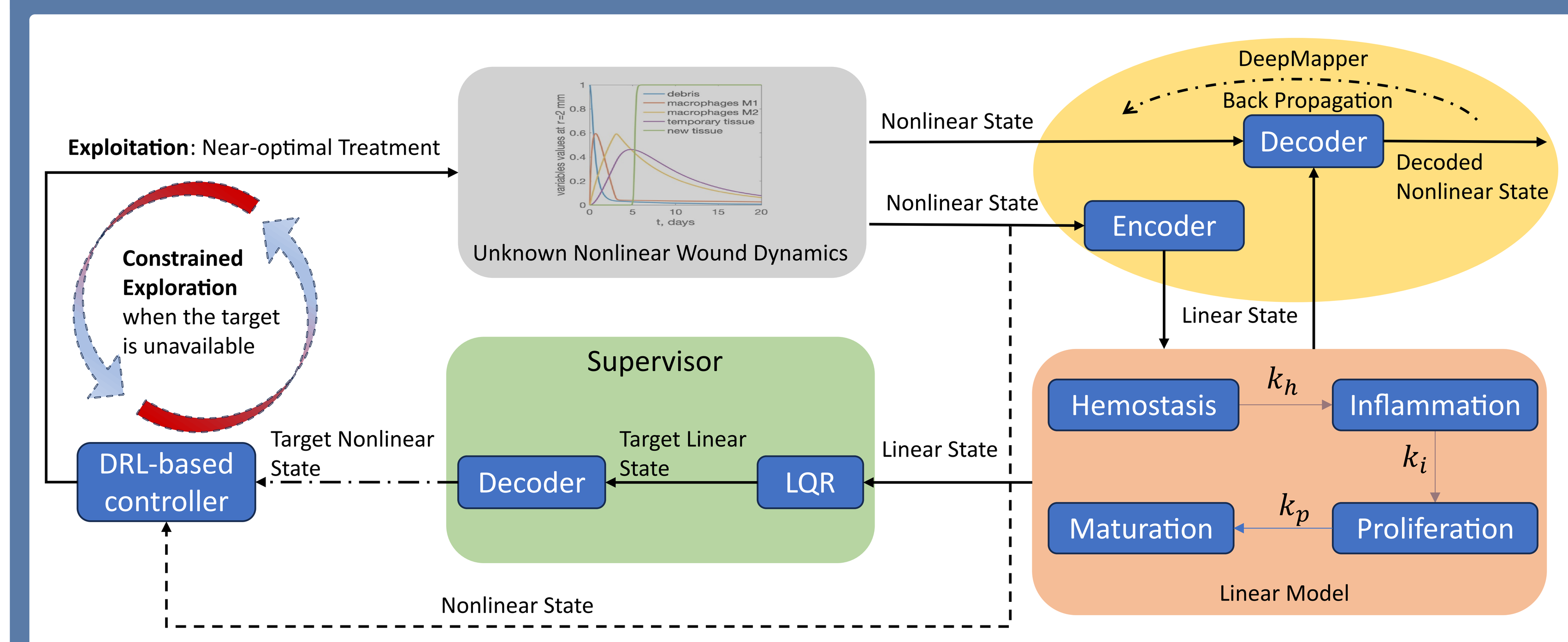


Goal: Find a policy π^θ such that $\sum_{t=0}^{\infty}\gamma^t r(x_t, u_t)$ is maximized with θ the weight of DNNs.

$$\theta_{t+1} = \theta_t + \alpha_{t+1}\left[r(x_t,u_t) + \underline{Q}^{\theta_{t-1}}(x_{t+1}) - \gamma Q^{\theta_t}(x_t,u_t)\right]\nabla_\theta Q^\theta(x_t,u_t)|_{\theta=\theta_t}$$

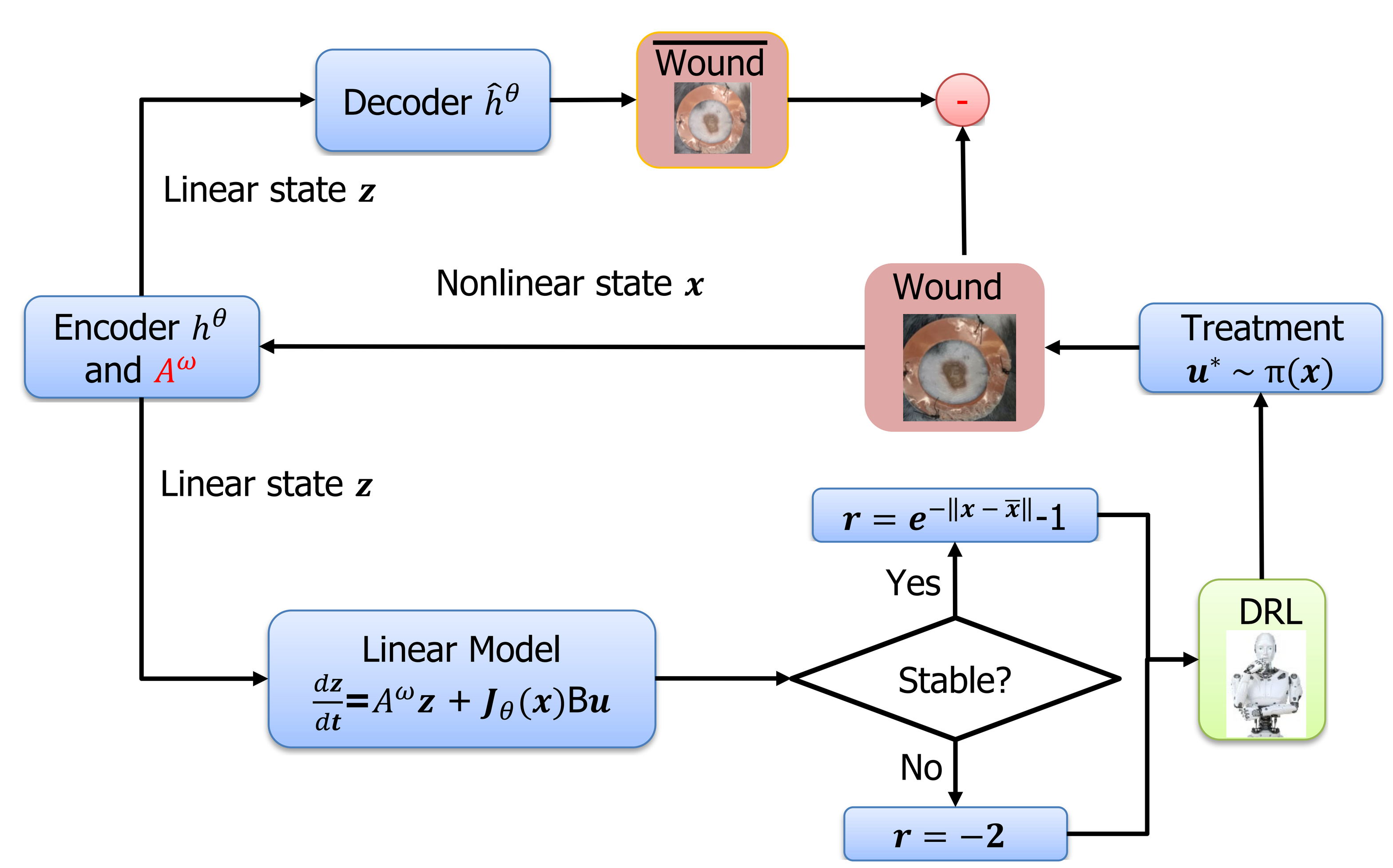
$$\text{where } \pi^\theta(x) \in \arg\max_{u\in\mathsf{U}} Q^\theta(x,u), \quad \underline{Q}^\theta(x) := \max_{u\in\mathsf{U}} Q^\theta(x,u)$$

Framework



DeepMapper: Where Does the Control u Come From

- DRL agent wants to find a policy π that tracks the decoded nonlinear signal of the linear signal that is optimal if it is stable
- DRL agent will get punished for generating control that makes the linear representation unstable
- We have formulated a closed-loop control framework that learns optimal control policy adaptively and incrementally with a safety guide



Numerical Study: Problem Setup

Nonlinear Wound dynamics:

Debris: $\dot a = -a m_1$

M1: $\dot m_1 = \beta a - \dot a - \rho\frac{m_1^q}{k^q+m_1^q} - \gamma_1 m_1 + \tilde D\left[\frac{1}{\tilde r}\frac{\partial m_1}{\partial \tilde r} + \frac{\partial^2 m_1}{\partial \tilde r^2}\right] - u m_1$

M2: $\dot m_2 = \rho\frac{m_1^q}{k^q+m_1^q} - \gamma_2 m_2 + \tilde D\left[\frac{1}{\tilde r}\frac{\partial m_2}{\partial \tilde r} + \frac{\partial^2 m_2}{\partial \tilde r^2}\right] + u m_1$

Temporary Tissue: $\dot c = m_2 - \mu c$

New Tissue: $\dot n = c\left[\tilde\alpha n(1-n) + \tilde D_n\left[\frac{1}{\tilde r}\frac{\partial n}{\partial \tilde r} + \frac{\partial^2 n}{\partial \tilde r^2}\right]\right]$

Linear Representation:

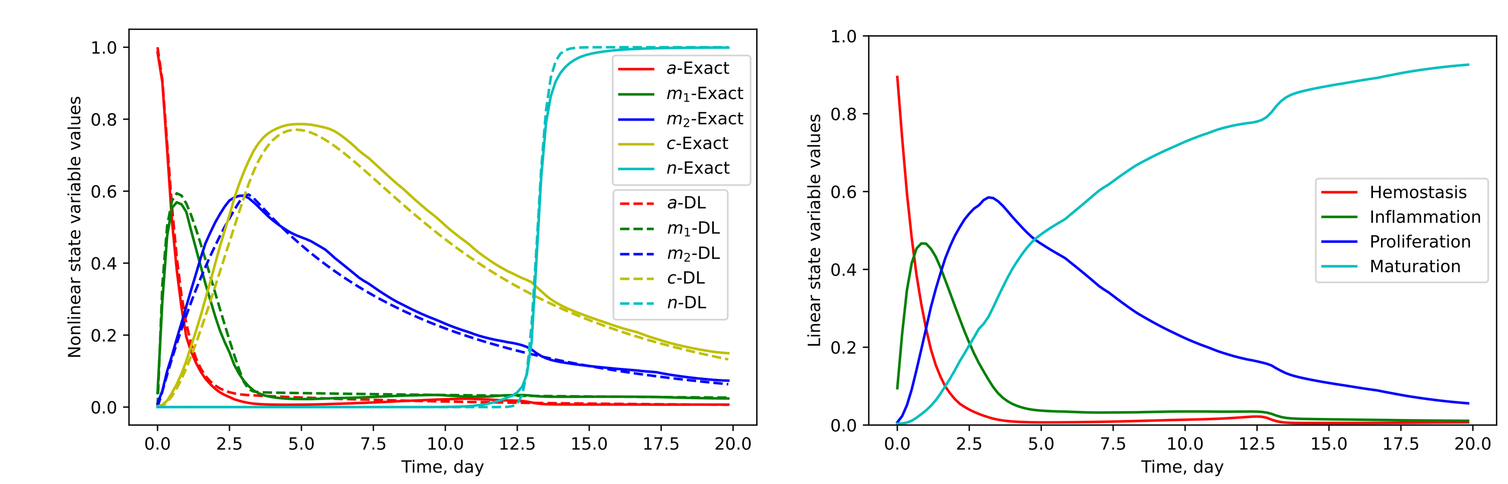
Hemostasis: $\dot H = -k_h H$

Inflammation: $\dot I = k_h H - k_i I$

Proliferation: $\dot P = k_i I - k_p P$

Maturation: $\dot M = k_p P$

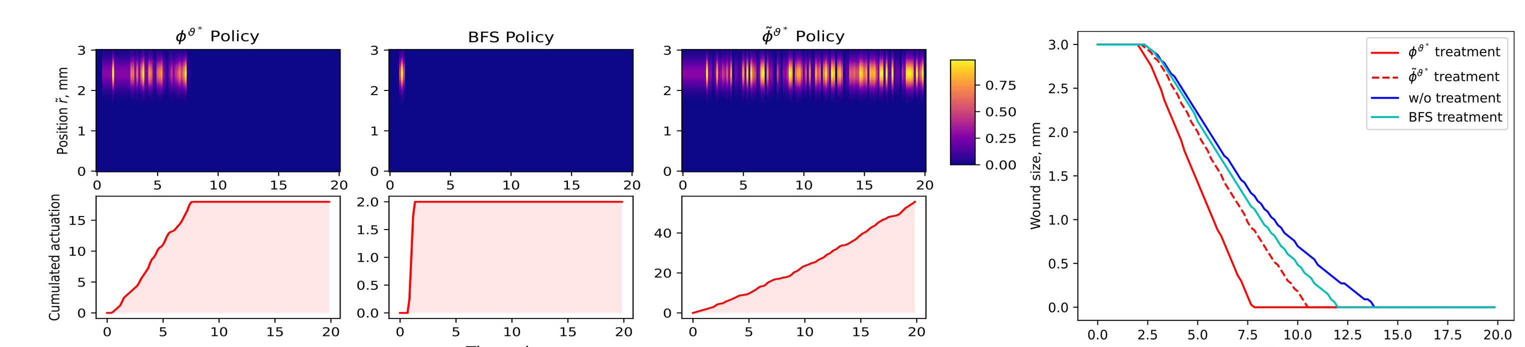
Numerical Study: Performance of DeepMapper



(a) Results of time dependent variables in the nonlinear model and decoded states

(b) Results of time dependence of all variables in the linear representation.

Numerical Study: Policy Comparison



(a) Different policies of wound treatment vs. time & pos

(b) Wound size vs. time

Conclusion & Future Work

- A general closed-loop control framework that incorporates deep learning, optimal adaptive control, and reinforcement learning for accelerating wound healing
- The proposed method has successfully reduced the wound healing time by a significant quantity compared to those without any treatment and treatment found by [4], as well as capturing the concerns of safety and less drug usage
- Optimality guarantee from the DRL algorithm [1, 2, 3]?

References

[1] Fan Lu, Prashant G. Mehta, Sean P. Meyn and Gergel Neu *Convex Q-learning* American Control Conference (ACC). Dec. 2021.

[2] Fan Lu, Prashant G. Mehta, Sean P. Meyn and Gergel Neu *Convex Analytic Theory for Convex Q-Learning* Control Conference on Decision and Control (CDC). Dec. 2022.

[3] Fan Lu, and Sean P. Meyn *Convex Q Learning in a Stochastic Environment* Control Conference on Decision and Control (CDC). Dec. 2023 (To Appear).

[4] Ksenia Zlobina1, Jiahao Xue, and Marcella Gomez *Effective Spatio-Temporal Regimes for Wound Treatment by Way of Macrophage Polarization: A Mathematical Model* Frontiers. Dec. 2022.