# Optimizing MRI Pulse Sequence using Reinforcement Learning

*Chongfeng Ling*

A dissertation submitted in partial fulfillment

of the requirements for the degree of

**Master of Science**

of

**University College London**.

Department of Physics and Astronomy

University College London

August 24, 2023

I, Chongfeng Ling, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

# Abstract

Magnetic Resonance Imaging (MRI) is a diagnostic tool in radiology that produces images of the body's internal tissues. The content of MRI images is governed by the pulse sequence, a series of time-dependent magnetic pulses crafted to manipulate the nuclear magnetic resonance phenomenon. Though the principles of MRI are well-defined, the design of a pulse sequence remains challenging due to hardware constraints and the extensive parameter search space. Nowadays, Deep Reinforcement Learning (DRL) has been applied to tackle problems with intricate dynamics. This project focus on the optimization of gradient-echo sequences for 1-D objects using the Deep Deterministic Policy Gradient (DDPG) algorithm under constraints on gradient slew rate. Our DDPG framework consists of an agent generates gradient-echo sequence action that interacts with an environment to reconstructed the object. The error between the reconstructed object and the target object serves as the reward to guide the agent's updates. This model demonstrated its capability to generate pulse sequences that yield satisfactory errors within stipulated constraints. We further talk about the model's limitations and potential improvements to render it more realistic. [1]

[1] Github Repository: `https://github.com/chongfengling/PHAS0077`

# Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor for guiding me in both my research and future direction. His sense of responsibility towards research and his rigorous attitude towards it have inspired me.

Secondly, I want to thank my parents for their supports and trusts.

Lastly, I would like to thank all the partners who have helped me in this project.

# Contents

## Chapter 1

# Introduction

## 1.1 Background

Magnetic Resonance Imaging (MRI) is a medical imaging technique used in radiology to form pictures of tissues of patients. MRI scanner uses strong magnetic fields to interact and manipulates the behavior of the hydrogen nuclei in the body to produce a detectable signal, which is then processed by a computer to form an image of the body. The quality of the image as long as the acquisition time is determined by pulse sequence used.

A pulse sequence in MRI refers to a sequential arrangement of time-dependent magnetic pulse and events designed to manipulate the interaction with nuclei present in body tissues. This sequence defines the order, duration and magnitude of radiofrequency (RF) pulse and gradient pulse. By adjusting these parameters, different contrasts of images and acquisition time can be achieved. Though the principles of MRI are rooted in well-defined physics, the design of pulse sequences is a challenge task, primarily due to hardware limitations and the vast parameter searching space. Figure 1.1 illustrates a gradient echo sequence structure.

In the realm of Deep Reinforcement Learning (DRL), there have been notable advancements in tackling issues with intricate dynamics. Within DRL frameworks, agents are characterized by deep neural networks, which empower them to explore

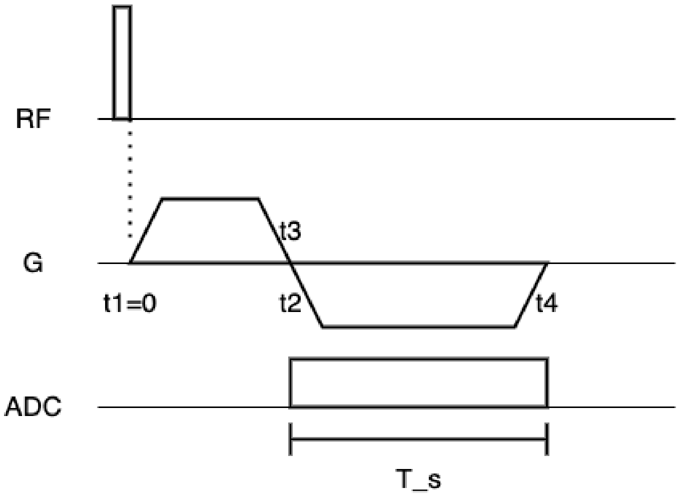


**Figure 1.1:** A 1D imaging protocol for a gradient echo sequence structure.

and learn data representations from environments that structured by certain rules, aiming to maximize their rewards.

## 1.2 Motivation

The design of an optimal pulse sequence is a multifaceted change. Conventional approaches, which export knowledge and iterative experiments, are effective but can be time-consuming and laborious [1]. On the other hand, real-world variables, including differences in the maximum magnetic field strength that MRI scanners can deliver, must be taken into account. These intricacies lead to the lack of a numerical solution for an optimal pulse sequence.

Considering the well-defined physics principles underlying MR signal acquisition, an virtual MR signal simulator can be served as the environment for DRL. Within this framework, the pulse sequence can be conceptualized as an action which is generated by an agent and interacts with the said environment. The rewards can be defined based on the reconstructed image quality, providing a quantitative measure of the performance of the agent and guide the update of the agent.

## 1.3 Objectives

The primary objective of this project is to develop a DRL framework for optimizing gradient echo sequence subject to constraints on gradient slew rate for a 1-D object. To achieve this, an environment will be established that encompasses the entire process from initiating an action to the final reconstruction of the object.

## 1.4 Structure

The structure of this report is arranged as follows. Chapter 2 reviews the current methods of pulse sequence design in MRI and the development of DDPG. Chapter 3 describes the methodology of this project, including the signal simulator, the algorithm with its architecture and evaluation criteria. Chapter 4 offers the details of experiments and chapter 5 will present the results and analysis. Finally, chapter 6 will conclude the project and discuss the future work.

## Chapter 2

# Literature Review

## 2.1 Review of MRI and Gradient-Echo Sequence

The foundational principles of MRI are based on the nuclear magnetic resonance (NMR) phenomenon which was first discovered in the 1930s. Then measurement method of NMR in liquids and solids developed and worked as a groundwork for MRI in the 1950s [2]. Beyond identifying the presence of various nuclei within the sample, gradient-echo sequence establish a relation between the spatial position of nuclei and their precessional frequencies. This relation enables the scanner to determine the origin of the signal and thus conclude the spatial information of nuclei.

Designing an optimal gradient-echo sequence is a challenging task as it faces a mount of constants including gradient hardware limitations and pre-defined MRI protocol parameters [1]. Current method for deriving an optimal sequence including convex optimization [1] and gradient ascent algorithm [3]. However, these methods have their intrinsic limitations. Convex optimization needs a rigorous mathematical model of the system as well as inequality constraint thresholds. On the other hand, gradient ascent algorithm is a local optimization method which is sensitive to the initial guess of the sequence and get stuck in local optimum easily.

## 2.2 Review of Reinforcement Learning

Reinforcement Learning involves an agent interacts with an environment by taking actions based on its current state. The environment then provides a reward and transitions to a new state. The goal of the RL model is to learn to find an optimal reward over time [4]. while the traditional RL algorithms like DQN only work on discrete action space, Deep Deterministic Policy Gradient is designed to address the continuous action space problem [5]. The Actor-Critic Architecture enhances the agent with the capability to assimilate the optimal policy within a high-dimensional state space.

## 2.3 Review of RL in MRI Sequence Design

Recent studies have incorporated the Reinforcement Learning framework to develop an optimal gradient-echo sequence generator for obtaining a consistent reconstructed image compared to the target object.

Zhu [6] employed a model-free Bayesian reinforcement learning to design a gradient-echo pulse sequence generator. Though the work resulted in a satisfactory outcome with a 1-D object, there is a strong assumption on the shape of gradient lobes and lack of constraints on the slew rate of gradient.

Samuel [7] implemented DDPG algorithm to control a scanner to generate signals and update the model based on both the prediction of object and running time.it is noteworthy that the prediction (be correct, incorrect or no prediction) is a component of reward function and an early termination status is return based on it.

## Chapter 3

# Methodology

## 3.1 MR Signal Simulation

### 3.1.1 Bloch Equations

Bloch Equations, named after Felix Bloch who first introduced them in the late 1940s, describe the evolution of nuclear a spin with its surrounding magnetic field over time [2]. Consider the presence of a magnetic field and relaxation terms, the Bloch equations are given by:

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{B}_{ext} + \frac{1}{T_1}(M_0 - M_z)\hat{z} - \frac{1}{T_2}\vec{M}_\perp \tag{3.1}$$

where $\vec{M} = (M_x, M_y, M_z)$ is the position of the spin in a reference frame and $\vec{B}_{ext}$ is the effective magnetic field. $\gamma$ is the gyromagnetic ratio which defines the proportionality between the magnetic moment a nucleus possesses and its angular momentum, it also describes how a nuclear spins responds to an external magnetic field. Relaxation terms $T_1$ and $T_2$ are the longitudinal and transverse relaxation times respectively. $M_0$ is the equilibrium magnetization. The solutions to the Bloch equations are:

$$\begin{aligned}
M_x(t) &= e^{-t/T_2}\left(M_x(0)\cos\omega_0 t + M_y(0)\sin\omega_0 t\right) \\
M_y(t) &= e^{-t/T_2}\left(M_y(0)\cos\omega_0 t - M_x(0)\sin\omega_0 t\right) \\
M_z(t) &= M_z(0)e^{-t/T_1} + M_0\left(1 - e^{-t/T_1}\right)
\end{aligned} \tag{3.2}$$

where $\omega_0$ is the precession frequency of the spin and determined by the external magnetic field $\vec{B}_{ext}$, its position and the gyromagnetic ratio $\gamma$. The detected signal $s(t)$ is a composite of the magnetization $M_x$ and $M_y$.

### 3.1.2 Frequency Encoding

Frequency encoding provides a robust technique for spatially differentiating signals originating from distinct locations within a spin ensemble. After the application of the RF pulse which all spins are bought into the transverse plane, a spatially linearly varying field is applied along some specific direction. As we only consider the imaging of a 1-D object with density profile $\rho(x)$ along the $x$-axis, the direction of the varying field can be chosen to be along the $x$-axis and the new effective magnetic field is updated to

$$B_x(x,t) = B_0 + xG(t) \tag{3.3}$$

where

$$\vec{B}_{ext} = B_0\hat{z} \tag{3.4}$$

and $G(t)$ is the time-dependent gradient field. The precession frequency of spins in Equation 3.2 at different positions are given by:

$$\begin{aligned} \omega(z,t) &\equiv \omega_0 + \omega_G(z,t) \\ &= \omega_0 + \gamma z G(t) \end{aligned} \tag{3.5}$$

Following the signal acquisition, fourier transform is used to recognize the connection between the spin precession frequency and its position. In this context, $s(t)$ and $\rho(x)$ are Fourier transform pairs in the frequency and the spatial domain respectively.

## 3.2 Deep Reinforcement Learning

### 3.2.1 DDPG

Following the environment defined as MR signal simulator, the agent is trained using Deep Deterministic Policy Gradient (DDPG) algorithm to generate pulse sequences. DDPG is an off-policy algorithm integrating the representational power of deep neural networks and the Deterministic Policy Gradient (DPG) algorithm, which aim to address the continuous action problem. Two neural networks are used to approximate the policy $\mu(s|\theta^{\mu})$ and the action-value function $Q(s,a|\theta^{Q})$ respectively. The actor network $\mu(s|\theta^{\mu})$ is used to approximate the optimal action $a$ deterministically for an arbitrary state $s$. The critic network $Q(s,a|\theta^{Q})$ is used to estimate the Q-value of taking an action for the state $s$ and then help the actor improving its policy. Two soft-updated target networks are employed for both actor and critic networks to ensuring training stability. The replay buffer enables off-policy learning of DDPG which allows the agent to learn from past experiences. To balance exploitation and exploration, exploitation noise $\mathcal{N}$ is added to the action during training. Details of DDPG are illustrated in Figure 3.1

**Algorithm 1** DDPG algorithm

```
Randomly initialize critic network $Q(s,a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights $\theta^Q$ and $\theta^\mu$.
Initialize target network $Q'$ and $\mu'$ with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer $R$
for episode = 1, M do
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial observation state $s_1$
  for t = 1, T do
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise
    Execute action $a_t$ and observe reward $r_t$ and observe new state $s_{t+1}$
    Store transition $(s_t, a_t, r_t, s_{t+1})$ in $R$
    Sample a random minibatch of $N$ transitions $(s_i, a_i, r_i, s_{i+1})$ from $R$
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:
        $\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$
    Update the target networks:
        $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
        $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
  end for
end for
```

**Figure 3.1:** Illustration of DDPG algorithm

### 3.2.2 Architecture of Actor-Critic Network

The actor network contains four fully-connected layers with ReLU activation function. The input layer is connected the real and image part of the state. Three output nodes are the percentage of duration of the constant dephasing gradient and rephasing gradient respectively plus a proportion of the current gradient magnitude relative to the maximum gradient strength. These nodes are then followed by a Sigmoid function to ensure the action range is within $[0,1]$.

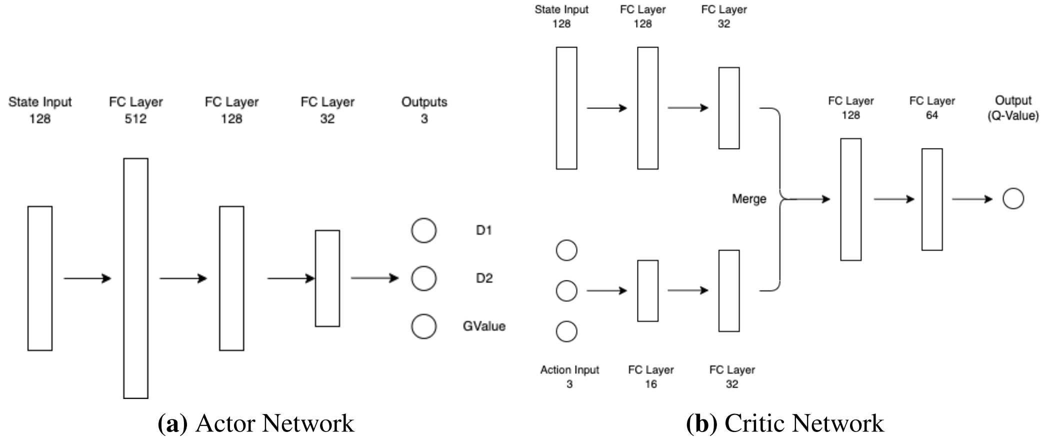
The critic network has a fully-connected state input layer followed by two full-connected layers with ReLU activation function. This is mirrored in the action input which has three nodes. Both inputs are concatenated by a merging layer and followed by two fully-connected layers. Output of the critic network is a single node which is the Q-value of the state-action pair. The architecture of the actor and critic networks are illustrated in Figure 3.2a and Figure 3.2b respectively.

### 3.2.3 Evaluation Criteria

The reward function $r_t$ is defined as the change of Mean Square Error between the current state $s_t$, which evolves within the environment due to the generated action and the target object $\rho$.

$$r_t = -(MSE(s_t,\rho) - MSE(s_{t-1},\rho)) \tag{3.6}$$

If the error is reduced, the reward is positive, and conversely, it becomes negative when the error increases. The agent is trained to maximize the reward, which leads to error minimization. An episode is terminated once the maximum number of steps is attained. Final reward is the cumulative sum of all rewards within this episode.



**(a)** Actor Network

**(b)** Critic Network

**Figure 3.2:** Architecture of two networks in DDPG algorithm

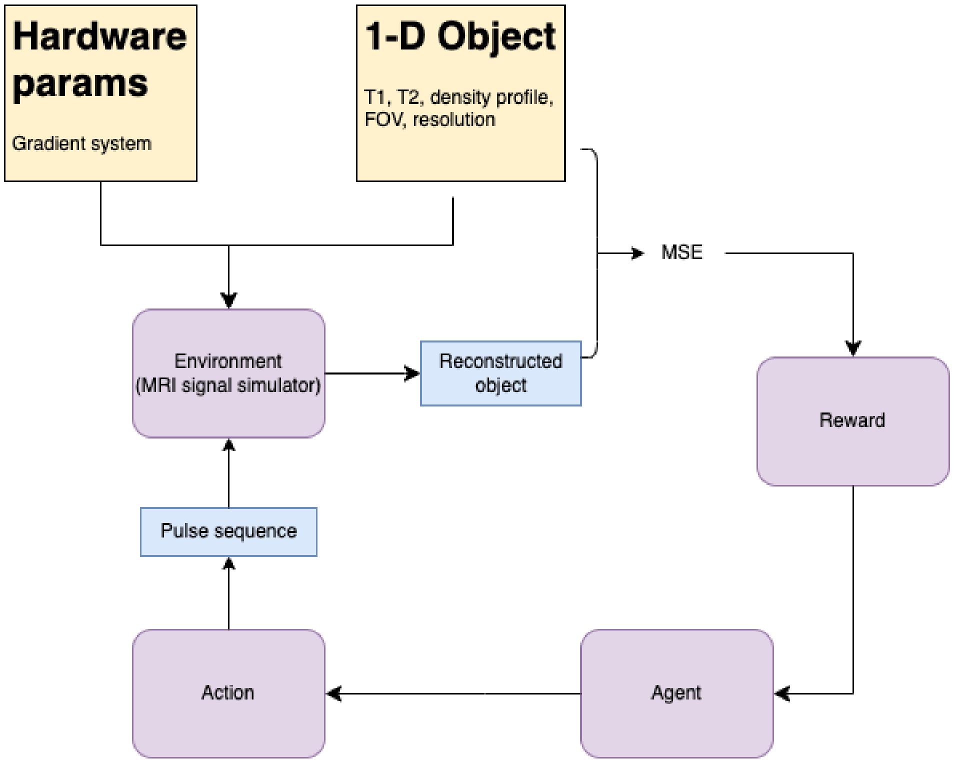


**Figure 3.3:** Schematic of the DDPG framework contains. An agent is designed to that generate gradient echo pulse sequence action. This action interacts with environment to simulate the signal by Bloch equations, complemented frequency encoding. The reward is then defined as the difference between the reconstructed density and the target object, serving as the guiding for the update of the agent.

## Chapter 4

# Experiments

## 4.1 Environment Setup

As we focus on the gradient echo sequence,the simulation begins after RF pulse and all spins are assumed to be on the transverse plane and oriented in the positive $y$-axis direction. The maximum amplitude of gradient in the $x$-axis direction is set to $4 \times 10^{-5}$ T/mm and resolution of the object equals to 1 mm. The dephasing gradient lobe initiates immediately following the conclusion of the RF pulse and is closely followed by a rephasing lobe. The entire sequence culminates when the dephasing lobe reverts to zero. For the sake of simplifying the gradient sequence, we assume that the two lobes are inversely related and that their maximum gradient magnitudes are identical. The duration for rephasing is twice that of dephasing. Analog-to-Digital Converter (ADC) samples the signal during dephasing time and due to the signal is contributed only by spins on the transverse plane, T1 relaxation time is set to infinite, while T2 is defined as 30 ms. A representative simulator is illustrated in Figure 4.1.

## 4.2 Slew Rate Constraint

Slew rate $sl$ is defined by the maximum achievable gradient strength $|G|_{max}$ and rise time $t_r$ as followed:

$$sl = \frac{|G|_{max}}{t_r} \tag{4.1}$$

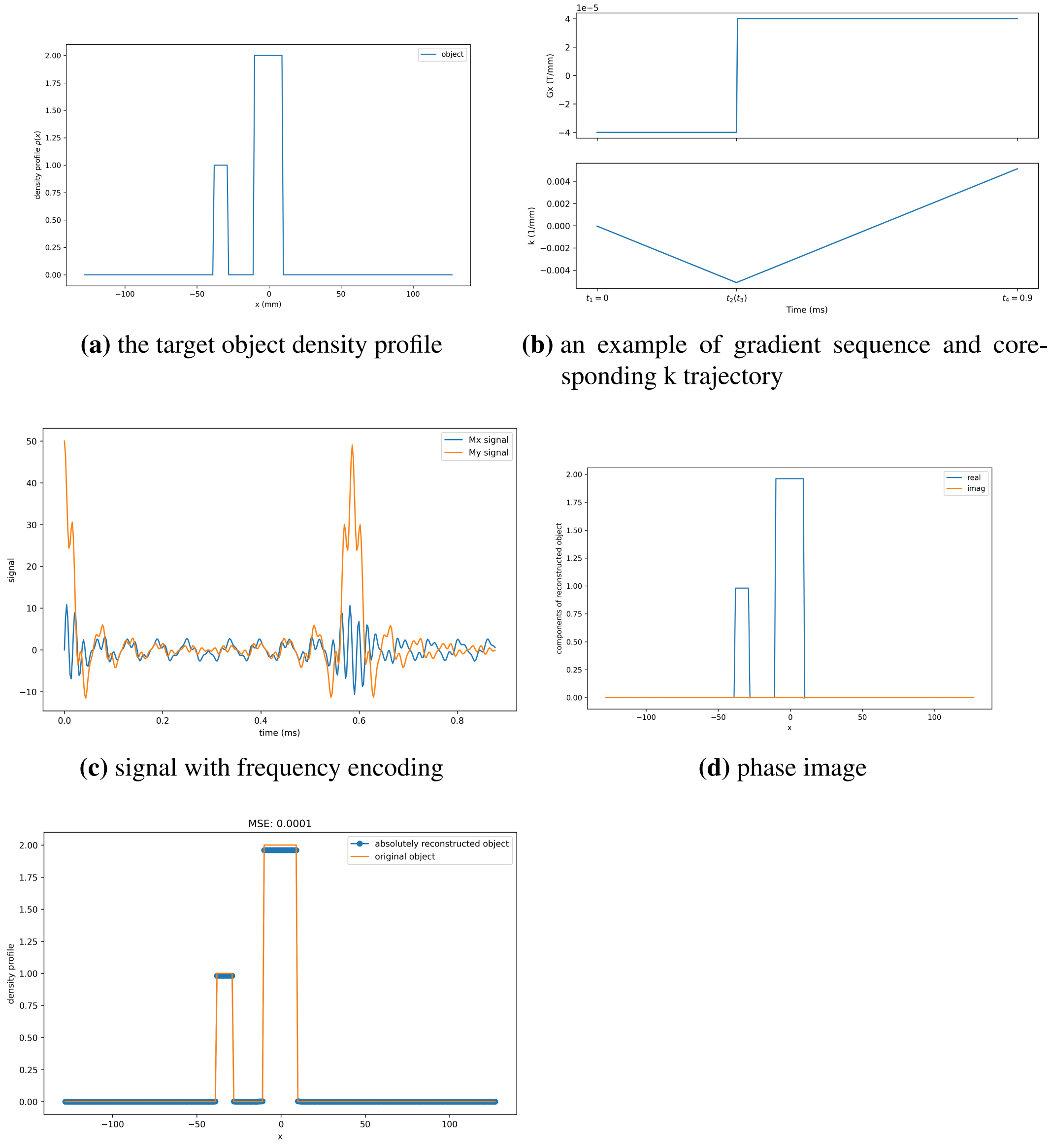
Considering the Nyquist sampling criterion, as the gradient ascends to its peak magnitude, the gradient lobe is a trapezoidal with a consistent slope. The sampling interval on k space is defined as:

$$\Delta k \equiv \frac{\gamma}{2\pi} \int_{t}^{t+\Delta t} dt' G\left(t'\right) = \frac{1}{L} \tag{4.2}$$

where $L$ is the FOV of the object. According to Equation 4.2, sampling interval over time space decreases with the increase of the gradient strength. Consequently, both $\Delta k$ and $\Delta t$ can not remain constant simultaneously. To reduce the complexity of the model, we here assume the $\Delta t$ is only affected by the maximum gradient strength, that is, we assume gradient lobe to be rectangular when defining the sampling interval over both t and k space. Equation 4.2 is written as:

$$\Delta k = \frac{\gamma}{2\pi} G \Delta t \tag{4.3}$$

This assumption enlarge the $\Delta t$ when the gradient strength is growing and $t_r = N\Delta t$ for some integer $N$. Equation 4.1 is then replaced by limiting the duration of the peak gradient to control the slew rate within a certain range.



**(a)** the target object density profile

**(b)** an example of gradient sequence and coresponding k trajectory

**(c)** signal with frequency encoding

**(d)** phase image

**(e)** absolutly reconstruction and object density. Error is cased by T2 relaxation.

**Figure 4.1:** Illustration of a single step in the environment (simulator) from the density profile of the object to the reconstructed density.

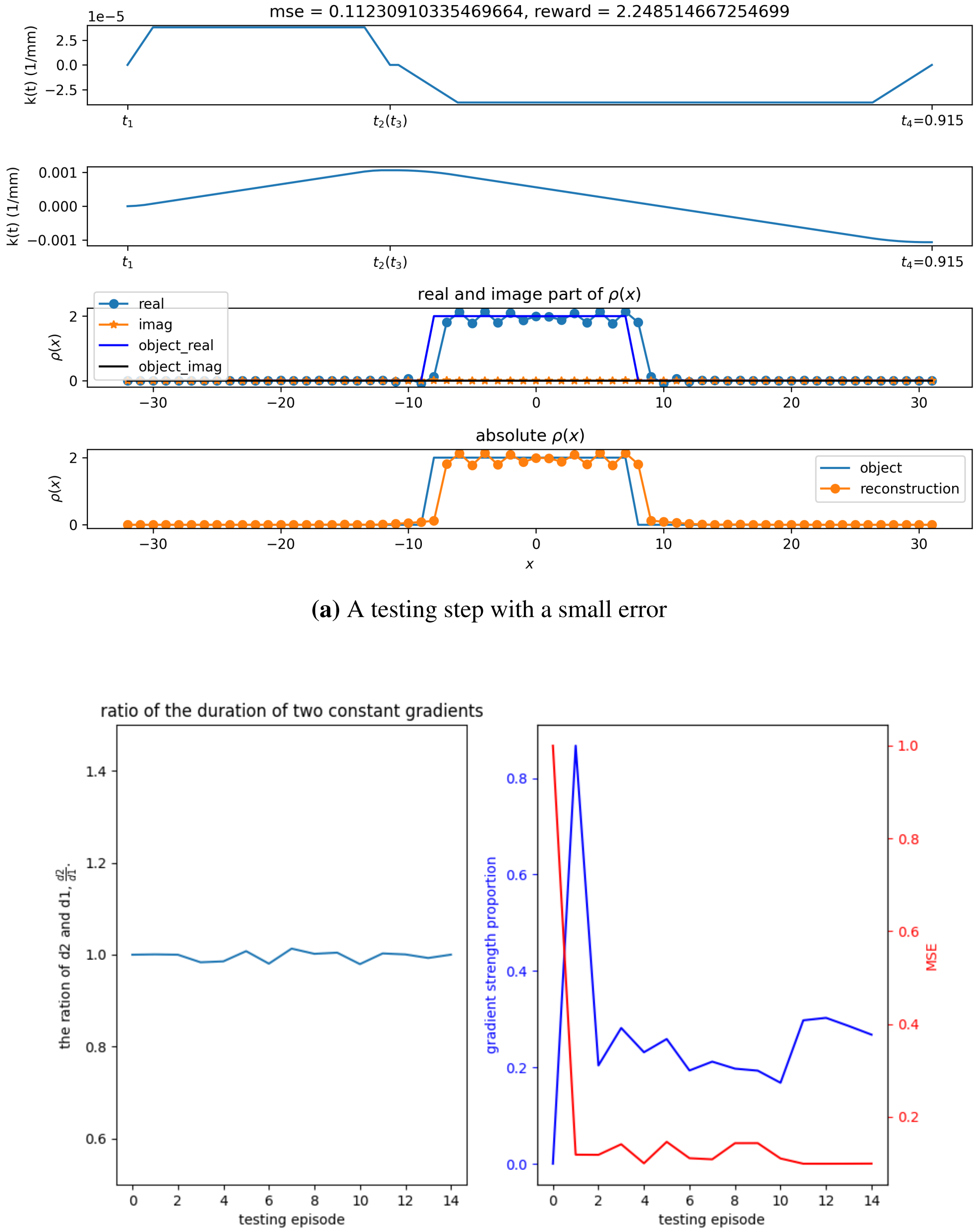
## Chapter 5

# Results and Analysis

We trained the agent to generate pulse sequences for two different objects. Both of two objects are 1-D and have a symmetrical density profile. The first object is a rectangle with a width of 20 mm and high of 2 while the second is composed by two identical rectangles. Tow density profiles are displayed in Figure 5.1a and Figure 5.2a. Each model was trained for 32 episodes, with each episode encompassing 1024 steps. Testing was performed every two training episodes and started from the same state with zero density everywhere.

Two steps in Figure 5.1a and Figure 5.2a have small mean square error related to the target object at around 0.11 and 0.12 respectively. The change of the error (reward) in the first step has a significant jump, while the second step is much smaller, demonstrating the model's balance in exploration and exploitation. Given the constraint on the slew rate, gradients of both steps attain the upper limit, mirroring the slew rate itself. The high peak constant gradients reduce the time of the entire simulation process, consequently minimizing the error introduced by T2 decay. K traverses from $k_{max}$ to $k_{min}$ with the equation $k_{max} = -k_{min}$ holding true. Concurrently, the imaginary part of the reconstructed signal is close to 0, resulting a phase accumulation to 0. An unexpected finding is that the reconstructed signal shifted slightly to the positive $x$-axis direction. Potential explanations for this phenomenon include the presence of a time interval with zero gradient between two gradient lobes and the number of sampling points is insufficient. Overall, these two steps prove that the agent is capable of generating pulse sequences that achieve an acceptable error under the given constraints.

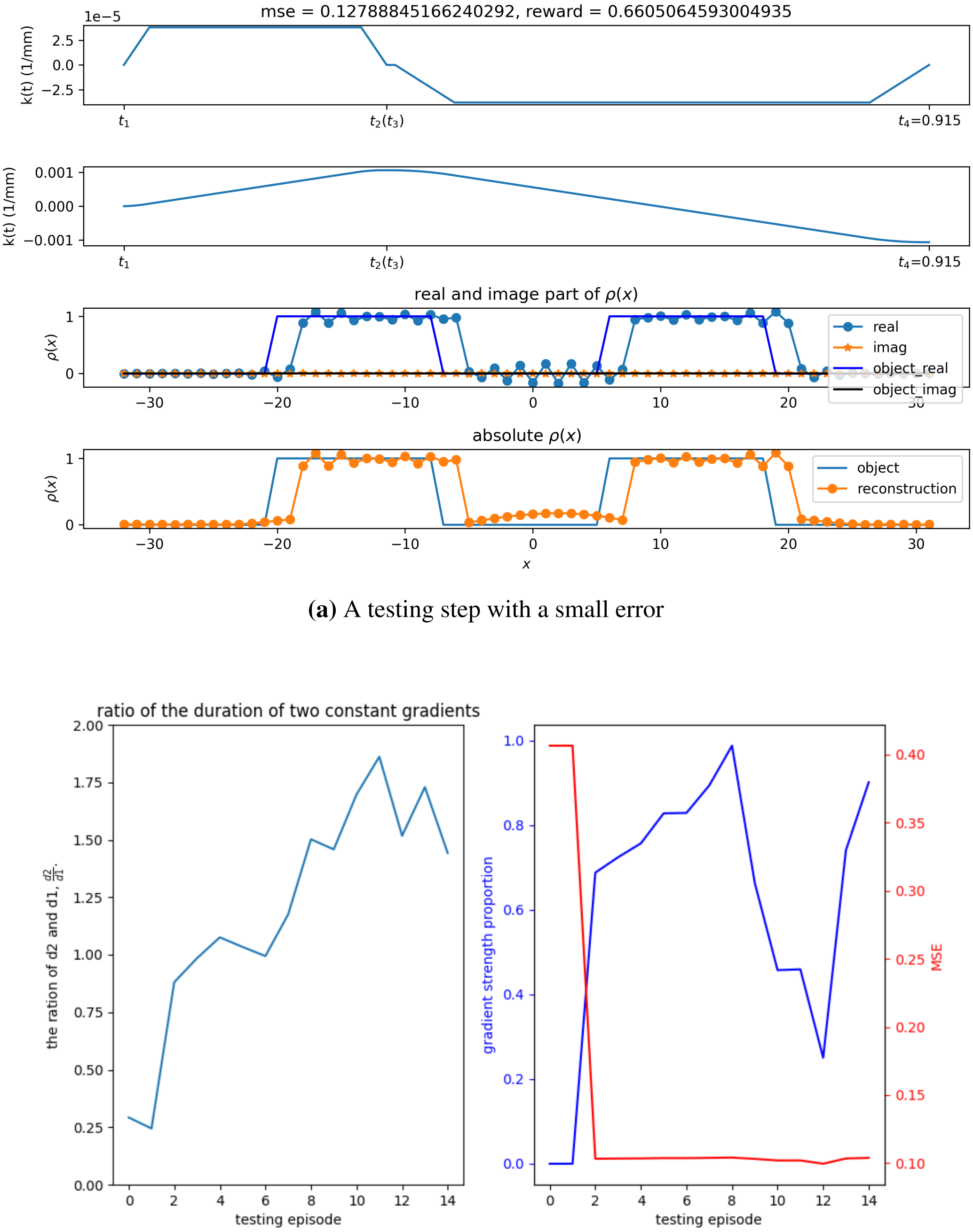
Figure 5.1b and Figure 5.2b display the trend of the optimal action, which corresponds the smallest error within a single testing episode, as the number of training episodes increases. After a small number of training episodes, the MSE of the corresponding optimal action decreases significantly and converge to 0.1 finally. In the first model, the magnitude of the constant gradient remains at a small value and the ratio between the first and second gradient lobe converges to 1. For the second model, the magnitude of the constant gradient is more oscillatory and closed to its upper limit. The ratio of two gradient lobes increases from 0.25 to 1.75. A possible reason is that the duration of two constant gradients is too short so a small change will result in a significant variations in the ratio. In summary, after a amount of training steps, the error between the reconstructed object and the target object converges to a small value.



**(a)** A testing step with a small error

**(b)** Best actions and their error in every testing episodes

**Figure 5.1:** Testing results for Object I



**(a)** A testing step with a small error

**(b)** Best actions and their error in every testing episodes

**Figure 5.2:** Testing results for Object II

# Chapter 6

# Conclusions and Future Work

## 6.1 Conclusions

This project implement a MR signal simulator and a gradient-echo sequence generator using the DDPG algorithm for 1-D object. The generator incorporates constraints on the slew rate of the gradient. The model is trained on two distinct objects and test results of two models not only present the existence of the optimal action, but also demonstrated the stability of the model.

## 6.2 Limitations

The project presents some limitations. Firstly, the decision to maintain a fixed $\Delta t$ despite variations of gradient can introduce inaccuracies in both the signal simulation and the implementation of the slew rate constraint. Secondly, for a more comprehensive evaluation, experiments should be conducted on target objects with more complex density profiles

## 6.3 Future Work

Future work can focus on two parts: improvement to the model and expansion of the tasks. For the model, the first step is to fix the $\Delta t$ problem. Another possible improvement is to define more sophisticated reward function and termination condition. For the tasks, the MR signal simulator can be extended to 2-D object and RF pulse can be added.

# Bibliography

[1] Matthew J. Middione, Michael Loecher, Kévin Moulin, and Daniel B. Ennis. Optimization methods for magnetic resonance imaging gradient waveform design. *NMR in Biomedicine*, 33(12), December 2020.

[2] Robert W. Brown, Yu-Chung N. Cheng, E. Mark Haacke, Michael R. Thompson, and Ramesh Venkatesan, editors. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons Ltd, Chichester, UK, April 2014.

[3] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J. Glaser. Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms. *Journal of Magnetic Resonance*, 172(2):296–305, February 2005.

[4] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[5] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING. 2016.

[6] Bo Zhu, Jeremiah Liu, Neha Koonjoo, Bruce R Rosen, and Matthew S Rosen. AUTOmated pulse SEQuence generation (AUTOSEQ) using Bayesian reinforcement learning in an MRI physics simulation environment. 2018.

[7] Simon Walker-Samuel. Control of a simulated MRI scanner with deep reinforcement learning, May 2023.