Research Overview

Olalekan Ogunmolu

Department of Electrical Engineering

The University of Texas at Dallas, Richardson, TX

Department of Radiation Oncology
University of Texas Southwestern Medical Center, Dallas, TX

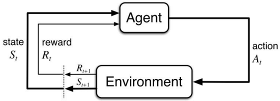
Nov. 09, 2018

Background

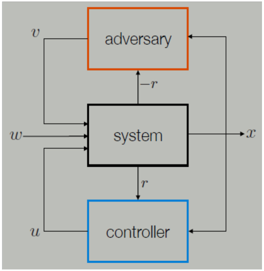

MINIMAX ITERATIVE DYNAMIC GAME: APPLICATION TO NONLINEAR ROBOT CONTROL TASKS.
Authors: Olalekan Ogunmolu, Nick Gans, and Tyler Summers
The Departments of Electrical and Mechanical Engineering,
University of Texas at Dallas, TX, USA

The robustness conundrum

- How to know *a priori* a policy's robustness limits?



- How to inculcate robustness into multistage decision policies?



Minimax iDG Research Overview

- To quantify the brittleness, we optimize the stage cost

$$\max_{\mathbf{v}_t \sim \psi \in \mathbf{\Psi}} \left[\sum_{t=0}^{T} \underbrace{c(\mathbf{y}_t, \mathbf{u}_t)}_{\text{nominal}} - \gamma \underbrace{g(\mathbf{v}_t)}_{\text{adversarial}} \right]$$

- To mitigate lack of robustness, we optimize the *cost-to-go*

$$\mathcal{J}_t(\mathbf{y}_t, \pi, \psi) = \min_{\mathbf{u}_t \sim \boldsymbol{\pi}} \max_{\mathbf{v}_t \sim \psi} \left(\sum_{t=0}^{T-1} \ell_t(\mathbf{y}_t, \mathbf{u}_t, \mathbf{v}_t) + L_T(\mathbf{y}_T) \right),$$

- and seek a saddle point equilibrium policy that satisfies

$$\mathcal{J}_t(\mathbf{y}_t, \pi^*, \psi) \le \mathcal{J}_t(\mathbf{y}_t, \pi^*, \psi^*) \le \mathcal{J}_t(\mathbf{y}_t, \pi, \psi^*),$$

Methods

- Suppose we have two agents interacting in an environment over an horizon T
- Let their dynamics be described as

$$\mathbf{x}_{t+1} = f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t, \mathbf{w}_t), \quad t = 0, \ldots, T-1$$

  - where $\mathbf{x}_t$ and $\mathbf{u}_t$ are state and control variables
  - $\mathbf{v}_t$ and $\mathbf{w}_t$ are the respective disturbance and stochastic random variables.
  - $\mathbf{w} = \{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{T-1}\}$ has distribution, $\mathcal{P}(\mathbf{w}_t | \mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t)$, $t = 0, \ldots, T-1$.
- Furthermore, $\mathbf{u}_t \in \{\pi = \pi_0, \pi_1, \ldots, \pi_T\}$, $\mathbf{v}_t \in \{\psi_0, \psi_1, \ldots, \psi_T\}$, and $\mathbf{w}_t \in \mathcal{P}(\mathbf{w}_t | \cdot)$, $t = 0, \ldots, T-1$

Methods

- Let a policy pair $(\boldsymbol{\pi}, \psi)$ is adopted, cost of a trajectory with initial condition $\mathbf{x}_0$ is

$$\mathcal{J}_0(\mathbf{x}_0, \boldsymbol{\pi}, \psi, \mathbf{w}) = \sum_{t=0}^{T-1} \ell_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) + L_T(\mathbf{x}_T) \qquad (1)$$

  - where $\ell_t, t = 0, \cdots, T-1$ and L_T are nonnegative instantaneous costs

- Average cost is thus

$$\mathcal{J}_0(\mathbf{x}_0, \boldsymbol{\pi}, \psi) = \mathbf{E}_{\mathbf{x}_0(\mathbf{w})=\mathbf{x}_0}[\tilde{\mathcal{J}}_0(\mathbf{x}_0(\mathbf{w}), \boldsymbol{\pi}, \psi, \mathbf{w})]$$

  - where $\mathbf{E}_{\mathbf{x}_0(\mathbf{w})=\mathbf{x}_0}(\cdot)$ is the expectation over random variables $\mathbf{x}_1(\mathbf{w}), \ldots, \mathbf{x}_T(\mathbf{w})$ having value $\mathbf{x}_0$. It will sometimes be abbreviated as $\mathbf{E}_{|\mathbf{x}_t}$ for $\mathbf{x}_t$.

Methods

- Define $\phi(\boldsymbol{\pi}, \boldsymbol{\psi})$ as

$$\phi(\boldsymbol{\pi}, \boldsymbol{\psi}) = \mathbf{E}[\mathcal{J}_0(\mathbf{x}_0, \boldsymbol{\pi}, \boldsymbol{\psi})] \tag{2}$$

  - where $\mathbf{E}$ has been abbreviated as $\mathbf{E}_{\mathbf{x}_0(\mathbf{w})=\mathbf{x}_0}$

so that

$$\phi(\boldsymbol{\pi}, \boldsymbol{\psi}) = \mathbf{E}[\tilde{\mathcal{J}}_0(\mathbf{x}_0(\mathbf{w}), \boldsymbol{\pi}, \boldsymbol{\psi}, \mathbf{w})] \tag{3}$$

Problem Statement

Find an admissible (saddle point equilibrium) policy pair that satisfies, $\mathcal{J}_0(\mathbf{x}_0, \pi^*, \psi) \le \mathcal{J}_0(\mathbf{x}_0, \pi^*, \psi^*) \le \mathcal{J}_0(\mathbf{x}_0, \pi, \psi^*)$, $\forall \pi \in \Pi, \psi \in \Psi$ and $\mathbf{x}_0$.

Problem Setup

- $\mathcal{J}_t(\mathbf{x}_t, \boldsymbol{\pi}, \psi)$ would denote the average process cost with initial condition, $\mathbf{x}_t$ under the policy pair, $(\boldsymbol{\pi}, \psi)$, *i.e.* ,

$$\mathcal{J}_t(\mathbf{x}_t, \boldsymbol{\pi}, \psi) = \min_{\boldsymbol{\pi}} \max_{\psi} \mathbf{E}_{|\mathbf{x}_t} \tilde{\mathcal{J}}_t(\mathbf{x}_t(\mathbf{w}), \boldsymbol{\pi}, \psi, \mathbf{w})$$

- DP transforms the optimization over whole trajectory to a step-wise optimization over $(\mathbf{u}_t, \mathbf{v}_t)$ as

$$\mathcal{J}_t(\cdot) = \min_{\mathbf{u}_t \sim \boldsymbol{\pi}} \max_{\mathbf{v}_t \sim \psi} \mathbf{E}_{|\mathbf{x}_t} \left[\sum_{k=t}^{T-1} \ell_k(\mathbf{x}_k, \boldsymbol{\pi}_k(\mathbf{x}_k), \psi_k(\mathbf{x}_k) + L(\mathbf{x}_k) \right]$$

with boundary condition $\mathcal{J}_T(\mathbf{x}_T, \boldsymbol{\pi}, \psi, \mathbf{w}) = L(\mathbf{x}_T)$

Brittleness Quantification

- Assume an agent's nominal policy, $\boldsymbol{\pi}$, has been found
- Suppose that there is another agent interacting in the nominal agent's environment, so that

$$\begin{aligned} \mathbf{x}_{t+1} &= f_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t), \quad \mathbf{u}_t \sim \pi_t, \mathbf{v}_t \sim \boldsymbol{\psi}_t \\ &:= \tilde{f}_t(\mathbf{x}_t, \mathbf{v}_t), \quad t = 0, \ldots, T-1. \end{aligned} \tag{4}$$

- We pose the stage cost as,

$$\ell_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) = \max_{\mathbf{v}_t \sim \psi} \left[\sum_{t=0}^{T} \underbrace{c(\mathbf{x}_t, \mathbf{u}_t)}_{\text{nominal cost}} - \gamma \underbrace{g(\mathbf{v}_t)}_{\text{opponent's cost}} \right]$$

  - $g_t(\cdot)$ controls the strength of the disturbing agent via the scalar term γ

Transition Slide

This page is left blank intentionally.

Minimax Iterative Dynamic Game

- To mitigate lack of robustness, we optimize

$$\mathcal{J}_t(\mathbf{x}_t, \pi, \psi) = \min_{\mathbf{u}_t \sim \boldsymbol{\pi}} \max_{\mathbf{v}_t \sim \psi} \left[\sum_{t=0}^{T-1} \ell_t(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) + L_T(\mathbf{x}_T) \right]$$

- seeking a saddle point equilibrium policy that satisfies the following inequality,

$$\mathcal{J}_t(\mathbf{x}_t, \pi^*, \psi) \leq \mathcal{J}_t(\mathbf{x}_t, \pi^*, \psi^*) \leq \mathcal{J}_t(\mathbf{x}_t, \pi, \psi^*),$$

Minimax iDG: Case Study

- Continuously solve an online trajectory optimization problem in a minimax fashion
- Essentially a meta-algorithm that is applicable to e.g. iLQG, DDP, GPS, DQN etc.
- Case study is a two-player iDG which proceeds as follows
  - approximate nonlinear system dynamics, $\mathbf{x}_{t+1}$, starting with a schedule of the nominal agent's local controls, $\{\bar{\mathbf{u}}_t\}$, and the opposing agent's local controls $\{\bar{\mathbf{v}}_t\}$ which are assumed to be available,

Minimax iDG: Case Study (Cont'd)

- Case study is a two-player iDG which proceeds as follows
  - run passive dynamics with $\{\bar{\mathbf{u}}\}, \{\bar{\mathbf{v}}\}$ and generate a nominal state trajectory $\{\bar{\mathbf{x}}_t\}$, with neighboring trajectories $\{\mathbf{x}_t\}$
  - we choose a small neighborhood, $\{\delta \mathbf{x}_t\}$ of $\{\mathbf{x}_t\}$, which provides an optimal reduction in cost as the dynamics no longer represent those of $\{\mathbf{x}_t\}$
  - discretizing time, the new state and control sequence pairs become $\delta \mathbf{x}_t = \mathbf{x}_t - \bar{\mathbf{x}}_t$, $\delta \mathbf{u}_t = \mathbf{u}_t - \bar{\mathbf{u}}_t$, $\delta \mathbf{v}_t = \mathbf{v}_t - \bar{\mathbf{v}}_t$.

Minimax iDG: Case Study (Cont'd)

- For

$$\mathcal{J}(\mathbf{x}_t, \boldsymbol{\pi}, \psi) = \min_{\mathbf{u}_t \sim \pi} \max_{\mathbf{v}_t \sim \psi} [\ell(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) + \mathcal{J}(f(\mathbf{x}_{t+1}, \mathbf{u}_{t+1}, \mathbf{v}_{t+1}))]$$

- We consider the Hamiltonian as a perturbation around $\{\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t\}$
- Cost over local neighborhood, $\{\delta \mathbf{x}_t\}$ can be approximated by a 2nd order Taylor expansion,

$$Q(\cdot) \approx \frac{1}{2} \begin{bmatrix} 1 \\ \delta \mathbf{x}_t^T \\ \delta \mathbf{u}_t^T \\ \delta \mathbf{v}_t^T \end{bmatrix}^T \begin{bmatrix} 1 & Q_{\mathbf{x}t}^T & Q_{\mathbf{u}t}^T & Q_{\mathbf{v}t}^T \\ Q_{\mathbf{x}t} & Q_{\mathbf{xx}t} & Q_{\mathbf{xu}t} & Q_{\mathbf{xv}t} \\ Q_{\mathbf{u}t} & Q_{\mathbf{ux}t} & Q_{\mathbf{uu}t} & Q_{\mathbf{uv}t} \\ Q_{\mathbf{v}t} & Q_{\mathbf{vx}t} & Q_{\mathbf{vu}t} & Q_{\mathbf{vv}t} \end{bmatrix} \begin{bmatrix} 1 \\ \delta \mathbf{x}_t \\ \delta \mathbf{u}_t \\ \delta \mathbf{v}_t \end{bmatrix} \tag{5}$$

Minimax iDG: Case Study (Cont'd)

- where

$$
\begin{aligned}
Q_{\mathbf{x}t} &= \ell_{\mathbf{x}t} + f_{\mathbf{x}t}^T V_{\mathbf{x}t+1}, \quad & Q_{\mathbf{u}t} &= \ell_{\mathbf{u}t} + f_{\mathbf{u}t}^T V_{\mathbf{x}t+1} \\
Q_{\mathbf{v}t} &= \ell_{\mathbf{v}t} + f_{\mathbf{v}t}^T V_{\mathbf{x}t+1}, \quad & Q_{\mathbf{xx}t} &= \ell_{\mathbf{xx}t} + f_{\mathbf{x}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{x}t} \\
Q_{\mathbf{ux}t} &= \ell_{\mathbf{ux}t} + f_{\mathbf{u}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{x}t}, & Q_{\mathbf{vx}t} &= \ell_{\mathbf{vx}t} + f_{\mathbf{v}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{x}t} \\
Q_{\mathbf{uu}t} &= \ell_{\mathbf{uu}t} + f_{\mathbf{u}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{u}t}, & Q_{\mathbf{vv}t} &= \ell_{\mathbf{vv}t} + f_{\mathbf{v}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{v}t} \\
Q_{\mathbf{uv}t} &= \ell_{\mathbf{uv}t} + f_{\mathbf{u}t}^T V_{\mathbf{xx}t+1} f_{\mathbf{v}t}.
\end{aligned}
$$

- it is expected that 2nd order terms will dominate the higher-order ones, consistent with linearized methods [Polycarpou and Ioannou (1992)].

Linearization of Dynamics

- LQR approximation to dynamics becomes

$$
\delta \mathbf{x}_{t+1} \approx f_{\mathbf{x}t}\delta\mathbf{x}_t + f_{\mathbf{u}t}\delta\mathbf{u}_t + f_{\mathbf{v}t}\delta\mathbf{v}_t
$$

$$
\ell(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) \approx \frac{1}{2}\begin{bmatrix} 1 \\ \delta\mathbf{x}_t^T \\ \delta\mathbf{u}_t^T \\ \delta\mathbf{v}_t^T \end{bmatrix}^T \begin{bmatrix} \ell_{0t} & \ell_{\mathbf{x}t}^T & \ell_{\mathbf{u}t}^T & \ell_{\mathbf{v}t}^T \\ \ell_{\mathbf{x}t} & \ell_{\mathbf{xx}t} & \ell_{\mathbf{ux}t}^T & \ell_{\mathbf{vx}t}^T \\ \ell_{\mathbf{u}t} & \ell_{\mathbf{ux}t} & \ell_{\mathbf{uu}t} & \ell_{\mathbf{uv}t} \\ \ell_{\mathbf{v}t} & \ell_{\mathbf{vx}t} & \ell_{\mathbf{vu}t} & \ell_{\mathbf{vv}t} \end{bmatrix} \begin{bmatrix} 1 \\ \delta\mathbf{x}_t \\ \delta\mathbf{u}_t \\ \delta\mathbf{v}_t \end{bmatrix} \tag{6}
$$

- It is easy to verify that the feedback controls are

$$
\begin{aligned}
\delta\mathbf{u}_t^\star &= -Q_{\mathbf{uu}t}^{-1}\left[Q_{\mathbf{u}t}^T + Q_{\mathbf{ux}t}\delta\mathbf{x}_t + Q_{\mathbf{uv}t}\delta\mathbf{v}_t\right], \\
\delta\mathbf{v}_t^\star &= -Q_{\mathbf{vv}t}^{-1}\left[Q_{\mathbf{v}t}^T + Q_{\mathbf{vx}t}\delta\mathbf{x}_t + Q_{\mathbf{vu}t}\delta\mathbf{u}_t\right].
\end{aligned} \tag{7}
$$

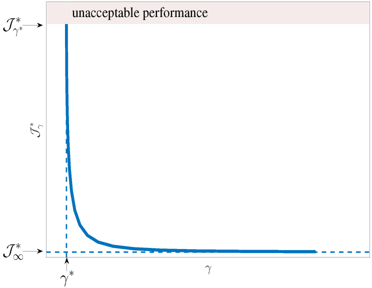
Linearization of Dynamics

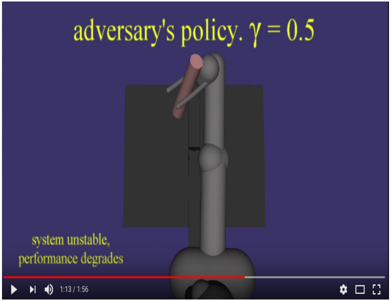
- From which we obtain the recursions,

$$
\begin{aligned}
\Delta V_t &= \mathbf{g}_{\mathbf{u}t} Q_{\mathbf{u}t} + \mathbf{g}_{\mathbf{v}t} Q_{\mathbf{v}t} + \mathbf{g}_{\mathbf{u}t} Q_{\mathbf{uv}t} \mathbf{g}_{\mathbf{v}t} \\
&\qquad + \frac{1}{2}\left(\mathbf{g}_{\mathbf{u}t} Q_{\mathbf{uu}t} \mathbf{g}_{\mathbf{u}t} + \mathbf{g}_{\mathbf{v}t} Q_{\mathbf{vv}t} \mathbf{g}_{\mathbf{v}t}\right) \\
V_{\mathbf{x}t} &= Q_{\mathbf{x}t} + \mathbf{G}_{\mathbf{u}_t}^T Q_{\mathbf{u}t} + \mathbf{G}_{\mathbf{v}_t}^T Q_{\mathbf{v}t} + \mathbf{G}_{\mathbf{u}t}^T Q_{\mathbf{uu}t} \mathbf{g}_{\mathbf{u}t} + \mathbf{g}_{\mathbf{u}t} Q_{\mathbf{ux}t} \\
&+ \mathbf{g}_{\mathbf{v}t} Q_{\mathbf{vx}t} + \mathbf{G}_{\mathbf{v}t}^T Q_{\mathbf{vv}t} \mathbf{g}_{\mathbf{v}t} + \mathbf{G}_{\mathbf{v}t}^T Q_{\mathbf{uv}t}^T \mathbf{g}_{\mathbf{u}t} + \mathbf{G}_{\mathbf{u}t}^T Q_{\mathbf{uv}t} \mathbf{g}_{\mathbf{v}t} \\
V_{\mathbf{xx}t} &= \frac{1}{2}\left(Q_{\mathbf{xx}t} + \mathbf{G}_{\mathbf{u}_t}^T Q_{\mathbf{uu}t} \mathbf{G}_{\mathbf{u}_t} + \mathbf{G}_{\mathbf{v}_t}^T Q_{\mathbf{vv}t} \mathbf{G}_{\mathbf{v}_t}\right) + \mathbf{G}_{\mathbf{u}_t}^T Q_{\mathbf{ux}t} \\
&\qquad + \mathbf{G}_{\mathbf{v}_t}^T Q_{\mathbf{vx}t} + \mathbf{G}_{\mathbf{u}_t}^T Q_{\mathbf{uv}t} \mathbf{G}_{\mathbf{v}_t} \qquad (8)
\end{aligned}
$$

- NB: The gains $\mathbf{g}_{it}$ and $\mathbf{G}_{it}$, i $= \mathbf{u}$ or $\mathbf{v}$ are as defined in ([§II.B]Ogunmolu et al. (2018)).

Results: Brittleness Quantification





Results: Iterative Dynamic Game

Research Overview

Olalekan Ogunmolu

iDG
Approach
Problem Setup
Policy Brittleness Quantification
iDG Problem Setup

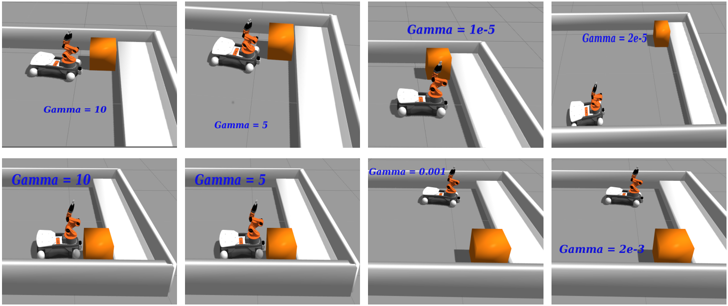
Results
Treatment Types
Frame-based RT
Cyberknife
BOO

Solution
3DOFControl
Adaptive NeuroControl
Neural Network Model

Treatment Plan (TP) Optimization (BOO)
Game Tree

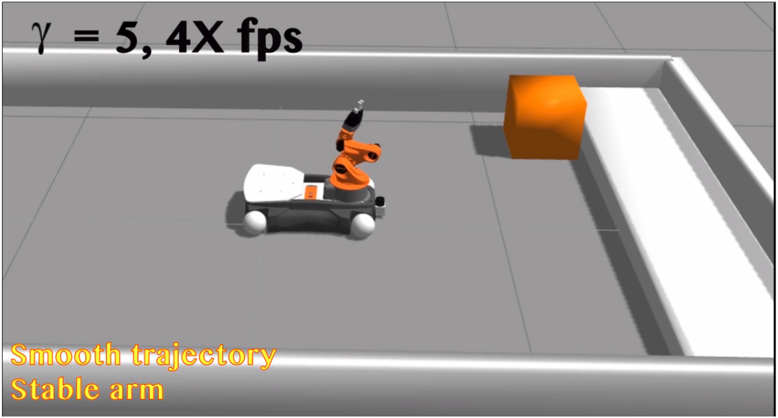
$\mathbf{x}_1^\star$

$\mathbf{x}_2^\star$



End pose of the KUKA platform with our iDG formulation given different goal states and γ-values

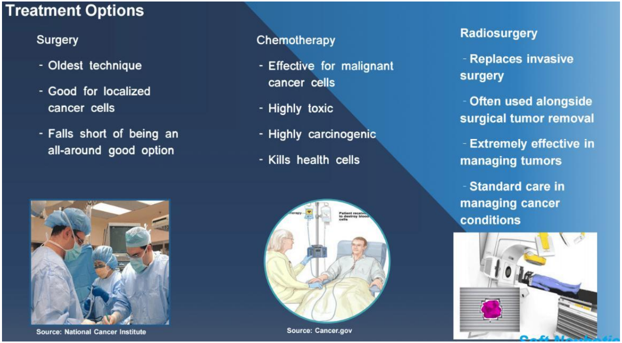
Video of Results



Robotic Radiotherapy

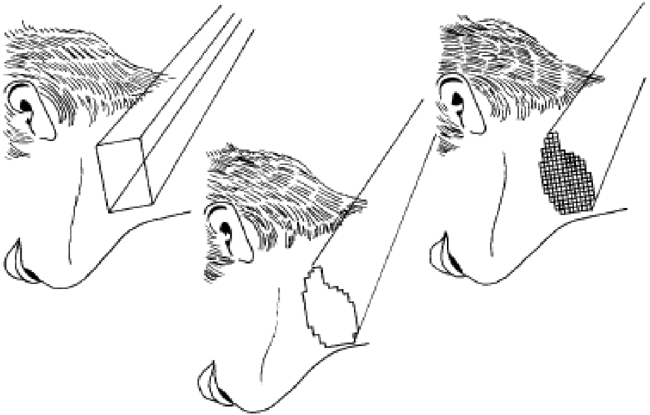
This page is left blank intentionally.

Treatment Options



- Radiation therapy is one of the major cancer therapy modalities
- About 2/3 of U.S. cancer patients receive radiation therapy either alone or in conjunction with surgery, chemotherapy, and immunotherapy e.t.c.

Radiotherapy types



Left: Conventional radiotherapy.
Middle: Conformal radiotherapy (CFRT) without intensity modulation.
Right: CFRT with intensity modulation. Reprinted from Webb (2001).

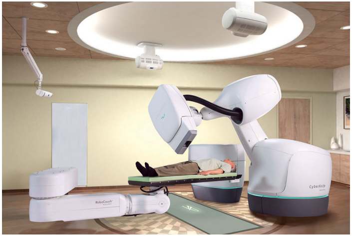
Frame-based Radiotherapy Treatment

Research Overview
Olalekan Ogunmolu
iDG
Approach
Problem Setup
Policy Brittleness Quantification
iDG Problem Setup
Results
Treatment Types
Frame-based RT
Cyberknife
BOO
Solution
3DOFControl
Adaptive NeuroControl
Neural Network Model
Treatment Plan (TP) Optimization (BOO)
Game Tree

- Fractionated radiation dose over many weeks/months

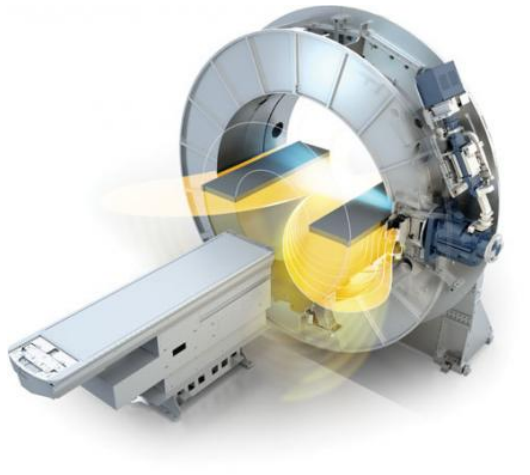
Frameless (Completely non-invasive) RT

High-dose volumes with complex shapes [Adler and Cox (1995)]



©Accuray Inc.

Image Guided Radiotherapy



Reprinted from Imaging Technology News

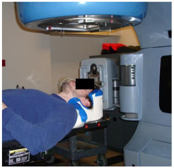
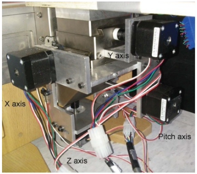
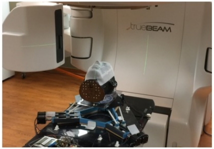
Beam Orientation Optimization

- During treatment planning, a **b**eam **o**rientation **o**ptimization problem (BOO) is separately solved
- Radiation is delivered from $\approx (5-15)$ different beam orientations during IMRT
- BOO determines the best beam angle combinations for delivering radiation.
- Process of determining beamlets' intensities is termed **fluence map optimization** (FMO)

The Immobilization Problem

This page is left blank intentionally.

The Immobilization Problem

Cerviño et al. (2010)	Liu et al. (2015)	Ostyn et al. (2017)
		
Feasibility evaluation	4-D robotic stage couch	6-DoF robotic couch

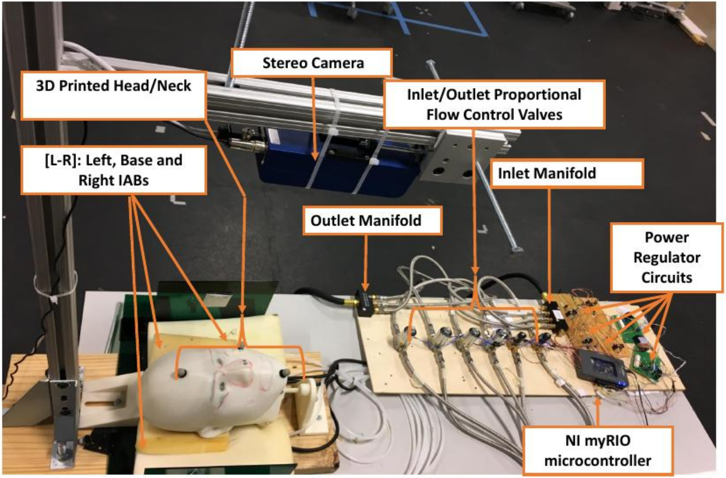
Solution: Soft-Robot Position Correcting Systems

- Eliminate rigid frames and metallic rings
- Eliminate attenuation of X-Ray beams
- Control design
  - Feedback control + optimal regulation + robustness to disturbance ✓

Vision-based 3-DOF Control

Left blank intentionally

Simulation Testbed



Hardware Description

Control Proposals

- Solve a state feedback and feedforward regulation problem
- An adaptation model based on past states and controls:
  - $Z^N = \{u(k), u(k-1), \ldots u(k-n_u), y(k), \ldots, y(k-n_y)\}$
  - Let a persistently exciting input signal $u_{ex} \in L_2 \cap L_\infty$ excite the system's nonlinear modes
- Design Goal:
  - Stabilize states, $\mathbf{y} = [z, \theta, \phi]^T$

Model Reference Adaptive Control

- Model head and bladder dynamics as
  - $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{B}\boldsymbol{\Lambda}\left(\mathbf{u} - f(\mathbf{y}, \mathbf{u})\right) + \mathbf{w}(k)$
    - $\mathbf{A}$, $\boldsymbol{\Lambda}$ unknown, $\mathbf{B}$, sgn$\boldsymbol{\Lambda}$ known
    - $f(\mathbf{y}, \mathbf{u}) \triangleq$ nonlinear function to be adapted for
    - $\mathbf{x} \triangleq$ tuple containing past controls and current outputs
- Approximate $f(\mathbf{y}, \mathbf{u})$ by a neural network with continuous memory states
  - $\hat{f}\left(\mathbf{y}(k), \mathbf{u}(k-d)\right)$ is realized with a *long-short term memory* cell (Horchreiter and Schmidhuber, '91, '97)
  - **purpose**: remember good adaptation gains

Assumptions

- A dynamic RNN with N neurons, $\varphi(\mathbf{x})$, exists
  - maps from a compact input space $\mathbf{u} \subset \mathbb{U}$ to $\mathbf{y} \subset \mathbb{Y}$ on the Lebesgue integrable functions within $[0, T]$ or $[0, \infty)$
- $f(\mathbf{y}, \mathbf{u})$ is exactly $\Theta^T \Phi(\mathbf{y})$
  - f has coefficients $\Theta \in R^{N \times m}$ and a Lipschitz-continuous vector of basis functions $\Phi(\mathbf{y}) \in R^N$
- Inside a ball $\mathbf{Y}_R$ with known, finite radius R,
  - an ideal neural network (NN) approximation $f(\mathbf{y}) : R^n \to R^m$, is realized to a sufficient degree of accuracy, $\varepsilon_f > 0$;
- Outside $\mathbf{Y}_R$,
  - the NN approximation error can be upper-bounded by a known unbounded, scalar function $\varepsilon_{max}(\mathbf{y})$;
  - $\|\varepsilon(\mathbf{y})\| \le \varepsilon_{max}(\mathbf{y}), \quad \forall \quad \mathbf{y} \in \mathbf{Y}_R$;
- There exists an exponentially stable reference model
  - $\dot{\mathbf{y}}_m = \mathbf{A}_m y_m + \mathbf{B}_m \mathbf{r}$

Adaptive Neuro-Control Scheme

- Set control law in terms of parameter estimates from the neural network weights and Lipschitz basis functions
  - $\Phi(\mathbf{y}) = \{\mathbf{y}(k-d), \cdots, \mathbf{y}(k-d-4), \mathbf{u}(k-d) \cdots \mathbf{u}(k-d-5)\}$
  - *i.e.* network looks back in time by 5 time steps at every instant, and then makes a prediction
- Derive adaptive adjustment mechanism from Lyapunov analysis for Adaptive Control (Parks, P., 1966)
- $\mathbf{u} = \underbrace{\hat{\mathbf{K}}_y^{\mathbf{T}} \mathbf{y}}_{\text{state feedback}} + \underbrace{\hat{\mathbf{K}}_r^{\mathbf{T}} \mathbf{r}}_{\text{optimal regulator}} + \underbrace{\hat{f}(\mathbf{y}, \mathbf{u})}_{\text{approximator}}$

Controller formulation

- $\hat{\mathbf{K}}_y$ and $\hat{\mathbf{K}}_r$ are adaptive gains to be designed

Term Contributions

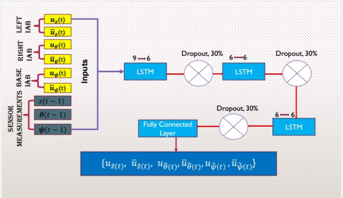
- $\hat{\mathbf{K}}_y^{\mathbf{T}}\mathbf{y}$ term keeps the states of the approximation set $\mathbf{y} \in \mathbf{B}_R$ stable,
- $\hat{\mathbf{K}}_r^{\mathbf{T}}\mathbf{r}$ term causes the states to follow a given reference trajectory
- Function approximator $\hat{f}(\mathbf{y}, \mathbf{u})$ ensures states that start outside the approximation set $\mathbf{y} \in \mathbf{B}_R$ converge to $\mathbf{B}_R$ in finite time

Adaptive Control Formulation

- Assume model matching conditions
  - such that $\hat{\mathbf{K}}_y = \mathbf{K}_y$, and $\hat{\mathbf{K}}_r = \mathbf{K}_r$ (ideally)
- Realize the approximator as $\hat{f}(\mathbf{y}) = \hat{\mathbf{\Theta}}^T \mathbf{\Phi}(\mathbf{y}) + \varepsilon_f(\mathbf{y})$
  - $\hat{\mathbf{\Theta}}^T$ denotes the vectorized weights of the neural network
  - $\mathbf{\Phi}(\mathbf{y})$ denotes the vector of lagged inputs and output,
  - $\varepsilon_f(\mathbf{y})$ is the approximation error.
  - $\mathbf{\Phi}(\mathbf{y}) = \{\mathbf{y}(k-d) \cdots \mathbf{y}(k-d-4), \mathbf{u}(k-d) \cdots \mathbf{u}(k-d-5)\}$

Neural Network Model

Neural Net Architecture



- input: lagged vector of past observations and current control actions
- repeat $\times$ 3
  - pass input through an lstm cell
  - followed by 30% dropout
- output will be control predictions directly fed as valve voltages

Lyapunov Redesign

- **Theorem:**
  - Given correct choice of adaptive gains $\hat{\mathbf{K}}_y$ and $\hat{\mathbf{K}}_r$, the error state vector, $\mathbf{e}(k)$ with closed loop time derivative $\dot{\mathbf{e}}$, is **uniformly ultimately bounded**, and the state $\mathbf{y}$ will converge to a neighborhood of $\mathbf{r}$.
  - Please see proof in Appendix(§6).

Stability Results

We find that

$$\begin{aligned}\dot{\mathbf{V}}(\cdot) &= -\mathbf{e}^T\mathbf{Q}\mathbf{e} - 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\varepsilon_f \\ &\leq -\lambda_{low}\|\mathbf{e}\|^2 + 2\|\mathbf{e}\|\|\mathbf{P}\mathbf{B}\|\lambda_{high}(\boldsymbol{\Lambda})\varepsilon_{max}\end{aligned}$$

- $\lambda_{low}, \lambda_{high} \equiv$ minimum and maximum eigenvalues of Q and $\boldsymbol{\Lambda}$ respectively.
- $\dot{\mathbf{V}}(\cdot)$ is thus negative definite outside the compact set
- $\chi = \left(\mathbf{e} : \|\mathbf{e}\| \leq \frac{2\|\mathbf{P}\mathbf{B}\|\lambda_{high}(\boldsymbol{\Lambda})\varepsilon_{max}(\mathbf{y})}{\lambda_{low}(\mathbf{Q})}\right)$
- thus, we conclude that the error $\mathbf{e}$ is uniformly ultimately bounded.
  - i.e. $\mathbf{y}(t) \to 0$ as $t \to \infty$

Results

- Solving the general form of the lyapunov equation, we have

$$\mathbf{P} = \begin{bmatrix} -\frac{170500}{2668} & 0 & 0 \\ 0 & -\frac{170500}{2668} & 0 \\ 0 & 0 & -\frac{170500}{2668} \end{bmatrix}$$

- Solenoid valves operate in pairs
  - set

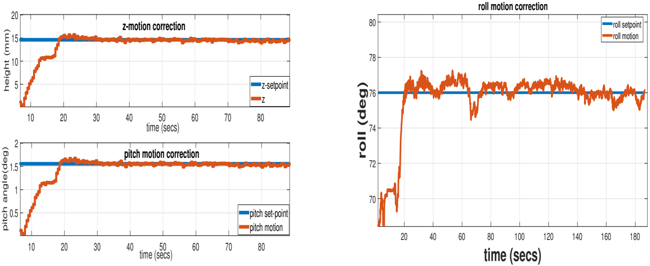
$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- B maps to the 3-axes controllers

$$\begin{bmatrix} u_z & u_\theta & u_\psi \end{bmatrix}^T$$

- non-zero terms are the max. duty-cycle to valves

Results



[Left]: Goal command: $(z, \theta, \phi) = (2.5mm, 0.25^\circ, 35^\circ)$ to $(14mm, 1.6^\circ, 45^\circ)^T$. [Right]: Head roll tracking.

Treatment Plan (TP) Optimization

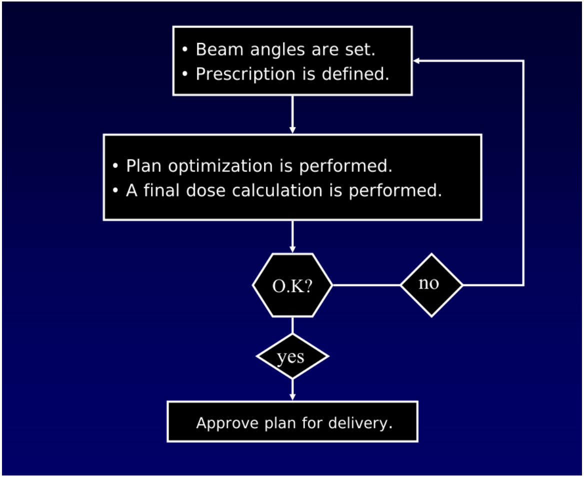
Left blank intentionally.

IMRT TP Overview

- IMRT delivers geometrically-shaped, high-precision photons to tumors in a beam orientation optimization (BOO) process.
- The **BOO problem** aims to find the right beam angle combinations from which to deliver radiation intensities.
    - essentially a combinatorial optimization problem
    - traditional methods fail at real-time feasible results[1].
- Afterwards, the intensity of the fluence is modulated in a **fluence map optimization** (FMO) process.

[1] Current approaches take too long, and are often not optimal.

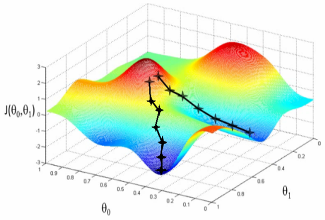
IMRT Treatment Plan Flowchart

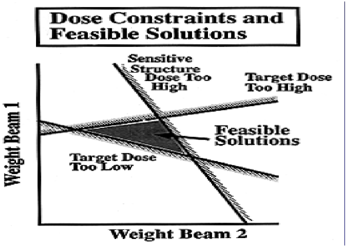


Reprinted from "IMRT Optimization Algorithms. David Shepard. Swedish Cancer Institute. AAPM 2007."

Problem Setup

- Given biological statement of prescriptions
  - find a numerical **objective function**
  - accompanied by **constraints**
- **Challenge**
  - a scalar-valued objective function usually not sufficient





Courtesy of David Shepard

Common Problem Formulation

- Maximize weighted least square dose from for PTVs
- Minimize weighted least square dose for OARs
- DVH-based goal treatments

- Current Approaches and Limitations
  - Stochastic optimization approaches: simulated annealing; genetic algorithms and gradient search, or a combination of genetic and gradient search algorithms
  - Mixed-integer programming, branch and cut/bound algorithms, beam angle elimination algorithms
  - Commercial planners use some highly non-convex objective (actual function is proprietary and unknown to public).
- General weakness: Feasible solution takes too long to find.

Our approach

- Automate the **beam search** problem
- **Ultimate goal** is real-time beam angle prediction given a target volume
- Drawing ideas from
  - **pattern recognition**;
  - **monte carlo evaluations**;
  - **game simulations**; and
  - **approximate dynamic programming**

What we do

- Formulate BOO problem into a large game planning strategy
- Neural fictitious self-play [Heinrich et al. (2015)], to refine policy predictions
  - **purpose**: drive policy weights to a **saddle equilibrium**
- a deep neural network models the nonlinear dynamical system (patient's geometry, robot-linac setup)
  - generating a policy that guides MCTS simulations for two players in a zero-sum Markov game

What we do

- In each episodic markov decision process (MDP) setting, an MCTS lookout strategy guides transition from one beam angle set to another
- Each player in a two-player Markov game finds a best response strategy to their opponent's average strategy
  - driving the policy weights toward an approximate **saddle equilibrium** Heinrich et al. (2015).

Problem Setup

- Let state of the dynamical system be $s \in \mathcal{S}$
- To be controlled by a discrete action $a \in \mathcal{A}$
- States evolve according to an (unknown) dynamics $p(s_{t+1} | s_t, a_t)$ (to be learned)
- Beam angle combination search task defined by a reward function, $R_t = \sum_{t=1}^{N} \gamma^{t-1} r(s_t, a_t)$
  - Can be found by recovering a policy, $p(a_t | s_t; \psi)$
- From now on, we will write $p(a_t | s_t; \psi)$ as $\pi_\psi(a_t | s_t)$.

Preliminaries

- Suppose the first player is p_1, and the second player is p_2
- p_1 chooses its action under a (stochastic) strategy, $\pi^{p_1} = \{\pi_0^{p_1}, \pi_1^{p_1}, \ldots, \pi_T^{p_1}\} \subseteq \Pi^{p_1}$
  - minimizing the game's outcome ζ
- p_2's actions are governed by a policy $\pi^{p_2} = \{\pi_0^{p_2}, \pi_1^{p_2}, \ldots, \pi_T^{p_2}\} \subseteq \Pi^{p_2}$
  - p_2 seeks to maximize ζ in order to guarantee an equilibrium solution for a game without saddle point.
- Π^{p_i} is the set of all possible nonstationary markovian policies

Preliminaries

- Each player bases its decision on a random event's outcome
  - generating a **mixed strategy** determined by **averaging the outcome** of individual plays.
- Both players constitute a two-player **stochastic action selection strategy**, $\pi(s, a) = Pr(a|s) := \{\pi^{p_1}, \pi^{p_2}\}$ that gives the probability of selecting moves in any given state
- Suppose the game simulation starts from an initial condition s_0.

Preliminaries

- One may write the optimal **reward-to-go** value function for state s in stage t, with horizon length T as

$$V_t^*(s) = \inf_{\pi^{p_1} \in \Pi^{p_1}} \sup_{\pi^{p_2} \in \Pi^{p_2}} \mathbb{E}\left[\sum_{i=t}^{T-1} V_t(s_0, f(s_t, \pi^{p_1}, \pi^{p_2}))\right],$$
$$s \in S,\ t = 0, \ldots, H-1$$

- where the terminal value $V_T^*(s) = 0, \ \forall s \in S$;
- $f(\cdot)$ represents the unknown system dynamics
- π^{p_1} and π^{p_2} contain the action/control sequences $\{a_t^{p_1}\}_{0 \le t \le T}$ and $\{a_t^{p_2}\}_{0 \le t \le T}$

Preliminaries

- The **saddle point strategies** for an optimal control sequence pair $\{a_t^{p_1^*}, a_t^{p_2^*}\}$ can be recursively obtained by optimizing a state-action value cost, $\mathcal{J}_t(s, a)$

$$V_t^*(s) = V_t^*(s_t, \pi_t^{p_1}, \pi_t^{p_2}) = \min_{\pi^{p_1} \in \Pi^{p_1}} \max_{\pi^{p_2} \in \Pi^{p_2}} V_t^\star(s_t, \pi^{p_1}, \pi^{p_2})$$
$$\forall s_t \in \mathcal{S}, \pi^{p_1} \in \Pi^{p_1}, \pi^{p_2} \in \Pi^{p_2}.$$

such that

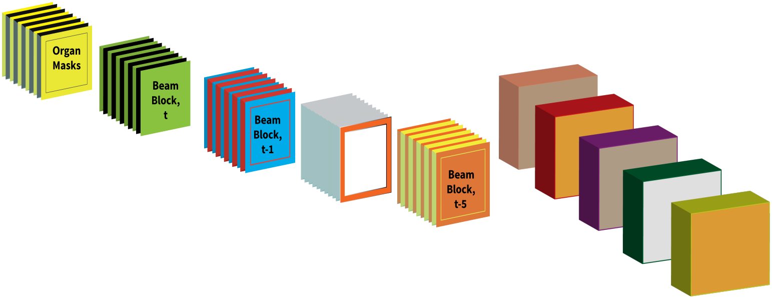
$$V_{p_1}^\star \le V^\star \le V_{p_2}^\star \quad \forall \{\pi_t^{p_1}, \pi_t^{p_2}\}_{0 \le t \le T}.$$

where $V_{p_i}^\star$ are the respective optimal values for each player.

Preliminaries

- Under ideal conditions, we'd like to find the optimal value function under perfect play
- **Caveat:** BOO exhibits Bellman's curse of dimensionality.
- What to do?
  - derive an **approximately optimal** value $V^{\star}_{\psi}(s)$
  - by continually estimating the value function $v^{p}_{\psi}(s)$ using e.g. a policy parameterized by a large function approximator

State Representation



[Left]: Concatenation of the target volume masks and the beam angles before feeding the input planes to the residual tower neural network.
[Right]: Each beam angle in a beam block is represented as shown.

Methods: Search

- Formulated as a bandit search that imposes a regret term on the Q-value
- During the planning process, we estimate a *value*, $v(\mathbf{y}_t)$, that estimates the optimality of a beam block;
- In parallel, we refine the deep neural network policy by optimizing its weight in a separate thread.
  - Network parameters updated by a **mixed strategy** which combines its **pure strategy**,
  - It is a best response to the fictitious opponent's **average pure strategy**.

Methods: Search

- Q-value defined as

$$\bar{Q}(s,a) = Q_j(s,a) + c\sqrt{\frac{2 \ln n(s)}{N(s,a}}, \tag{9}$$

$$a^\star = \arg\max_a \bar{Q}(s,a) \tag{10}$$

Fluence Map Optimization

- Suppose $\mathcal{X}$ is the total discretized of voxels of interest (VOI's) in a target volume
- Let $\mathcal{B}_1 \cup \mathcal{B}_2 \cup \ldots \cup \mathcal{B}_n \subseteq \mathcal{B}$ represents the partition subset of a beam $\mathcal{B}$,

Methods: Fluence Map Optimization

- Suppose further that $\mathcal{D}_{ij}(\theta_k)$ is the matrix that describes each dose influence, d_i
  - delivered to a discretized voxel, i, in a volume of interest, $VOI_h\,(h = 1, \ldots, \mathcal{X})$, from a beam angle, θ_k, $k \in \{1, \ldots, n\}$
- We compute the matrix $\mathcal{D}_{ij}(\theta_k)$ by calculating each d_i for every bixel, j, at every φ°, resolution, where $j \in \mathcal{B}_k$

Methods: FMO problem definition

- The pre-calculated dose term is given by $\mathbf{Ax} = \{\sum_s \frac{w_s}{v_s} \mathcal{D}_{ij}^s \mathbf{x}_s \,|\, \mathcal{D}_{ij} \in \mathbb{R}^{n \times l}, n > l\}$
- Let $w_s = \{\underline{w}_s, \bar{w}_s\}$ be the respective underdosing and overdosing weights for the OARs and PTVs, and v_s represents the total number of voxels in each structure.
- We propose the following cost

$$\frac{1}{v_s} \sum_{s \in \text{OARs}} \|(b_s - \underline{w}_s \mathcal{D}_{ij}^s \mathbf{x}_s)_+\|_2^2 + \frac{1}{v_s} \sum_{s \in \text{PTVs}} \|(\bar{w}_s \mathcal{D}_{ij}^s \mathbf{x}_s - b_s)_+\|_2^2 \tag{11}$$

Methods: FMO

- Rewriting the objective, subject to nonnegative bixel intensity constraints, we have the minimization problem

$$\min \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 \quad \text{subject to } x \geq 0.$$

- The Lagrangian becomes

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \min \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_2^2 - \boldsymbol{\lambda}^T \mathbf{x}.$$

- Since we are solving a large scale problem, we use the ADMM algorithm
- Introducing an auxiliary variable $\mathbf{z}$, we have

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2, \quad \text{subject to } \mathbf{z} = \mathbf{x}, \;\; \mathbf{z} \geq 0,$$

Methods: FMO

- Solving either the $\mathbf{x}$ and $\mathbf{z}$ sub-problems, we have

$$\mathbf{x}^{k+1} = \left(\mathbf{A}^T\mathbf{A} + \rho \boldsymbol{I}\right)^{-1}\left(\mathbf{A}^T\boldsymbol{b} + \rho \mathbf{z}^k - \boldsymbol{\lambda}^k\right). \tag{12}$$

- And using the soft-thresholding operator, $S_{\boldsymbol{\lambda}/\rho}$, we find that

$$\mathbf{z}^{k+1} = S_{\boldsymbol{\lambda}/\rho}\left(\mathbf{x}^{k+1} + \boldsymbol{\lambda}^k\right), \tag{13}$$

where $S_{\boldsymbol{\lambda}/\rho}(\tau) = (x - \boldsymbol{\lambda}/\rho)_+ - (-\tau - \boldsymbol{\lambda}/\rho)_+$. $\boldsymbol{\lambda}$ is updated as

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k - \gamma(\mathbf{z}^{k+1} - \mathbf{x}^{k+1}), \tag{14}$$

where γ is a parameter that controls the step length.

Game Tree Simulation

- For b^d possible move sequences for a robot-patient setup
  - $b =$ beam angles chosen to construct a fluence
  - $d =$ is the total number of discretized angles.
- Suppose $b = 180$ and $d = 5$, we have 180^5 possible search directions
- Exhaustive search becomes real-time infeasible.

Game Tree Simulation; Approach

- Simulate a game of **perfect recall**
  - a sequential simulation of different beam angle combinations g
  - guided by probabilities obtained from a two-player zero-sum game of neural FSP
- the probability distribution is over the possible beam angle subsets, θ^j, in the beam angle space, , $\boldsymbol{\Theta}$
- This strongly discourages classical beam s approaches

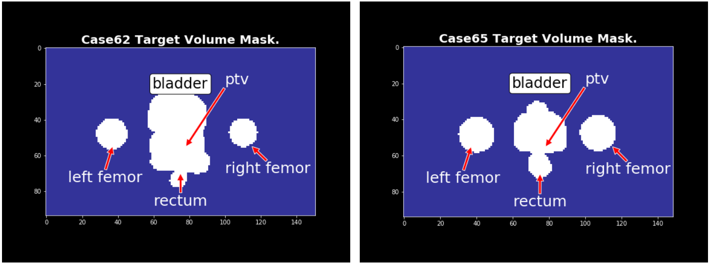
Game Tree Simulation

- Network roll-out policy then efficiently guides the tree's game, Γ toward a *best-first* set of beam angle candidates
- Best-first leaf node encountered is the child node with the highest reward in the tree
- Essentially, a sampling-based lookout algorithm
  - Focuses learning on regions of the state space that are likely to have a good fluence

Transition Slide

This page is left blank intentionally.

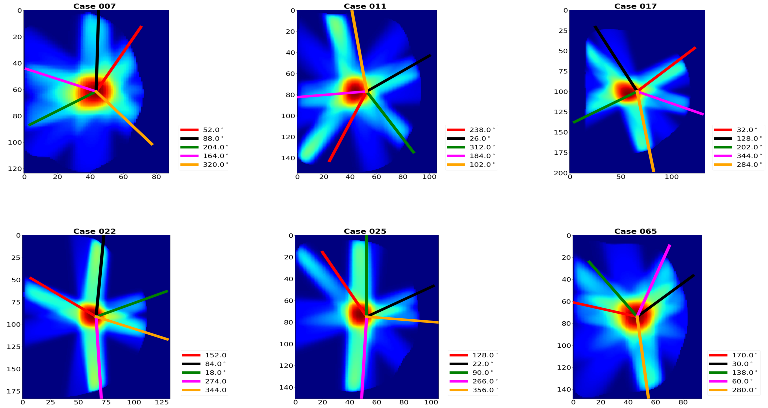
Example Target Volumes



Example Target Volumes. The PTV is engulfed within the bladder in all cases.

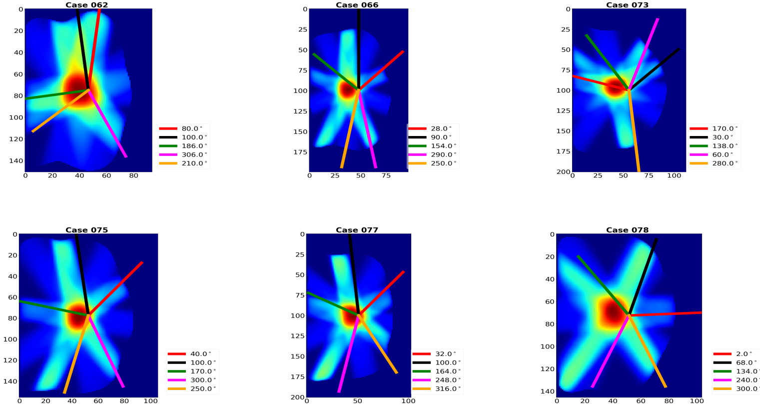
Dose washes for select patients during training of the self-play network

Training Regime



Dose washes for select patients during testing of self-play network

Inference Regime



Transition Slide

This page is left blank intentionally.

Future Work: IMRT Immobilization

- Adopt best practices from one of the SOTA modeling approaches viz
  - *finite element methods* [Coevoet et al. (2017); Bern et al. (2017)],
  - *constant curvature approach* [Godage et al. (2016)],
  - the *continuous Cosserat approach* [Renda et al. (2014)], and
  - the *multi-body hyper-redundant model* [Kang et al. (2012)].

Future Work

- Robust Stabilization via inverse optimality
  - Currently working on imbibing Lyapunov stability into traditional optimal controllers
    - Build on Freeman and Kokotovic's point-wise min-norm robust control lyapunov function to realize a meaningful value function (Int. Journal of Optimal Control, 1996)
    - Simplify algorithm for real-time control
  - Carry out numerical and and experimental verification and validations
  - Important in multistage decision policies, reinforcement learning controllers

Transition Slide

This page is left blank intentionally.

Appendix: 3-DOF Lyapunov Stability Proof

- **Proof**:
  - Choose a Lyapunov function candidate $\mathbf{V}$ in terms of the generalized error state space $\mathbf{e}$, gains, $\tilde{\mathbf{K}}_y^T$, $\tilde{\mathbf{K}}_r^T$, and parameter error $\varepsilon_f(\mathbf{y}(k))$ space

$$\mathbf{V}(\mathbf{e}, \tilde{\mathbf{K}}_y, \tilde{\mathbf{K}}_r^T) = \mathbf{e}^T \mathbf{P} \mathbf{e} + \mathbf{tr}(\tilde{\mathbf{K}}_y^T \boldsymbol{\Gamma}_y^{-1} \tilde{\mathbf{K}}_y^T |\boldsymbol{\Lambda}|) + \mathbf{tr}(\tilde{\mathbf{K}}_r^T \Gamma_r^{-1} \tilde{\mathbf{K}}_r |\boldsymbol{\Lambda}|) \tag{15}$$

Appendix: Stability proof

$$\dot{\mathbf{V}}(\mathbf{e}, \tilde{\mathbf{K}}_y, \tilde{\mathbf{K}}_r) = \dot{\mathbf{e}}^T \mathbf{P}\mathbf{e} + \mathbf{e}^T \mathbf{P}\dot{\mathbf{e}} + 2\mathbf{tr}(\tilde{\mathbf{K}}_y^T \boldsymbol{\Gamma}_{\mathbf{y}}^{-1} \dot{\hat{\mathbf{K}}}_y |\boldsymbol{\Lambda}|) + 2\mathbf{tr}(\tilde{\mathbf{K}}_r^T \boldsymbol{\Gamma}_r^{-1} \dot{\hat{\mathbf{K}}}_r |\boldsymbol{\Lambda}|)$$

$$= \left[\mathbf{A}_m \mathbf{e} + \mathbf{B}\boldsymbol{\Lambda}[\Delta\hat{\mathbf{K}}_r^T \mathbf{r} + \Delta\hat{\mathbf{K}}_x^T \mathbf{x}]\right]^T \mathbf{P}\mathbf{e} + \ldots$$
$$\mathbf{e}^T \mathbf{P}\left[\mathbf{A}_m \mathbf{e} + \mathbf{B}\boldsymbol{\Lambda}[\Delta\hat{\mathbf{K}}_r^T \mathbf{r} + \Delta\hat{\mathbf{K}}_x^T \mathbf{x}]\right] + \ldots$$
$$2\,\mathbf{tr}(\Delta\mathbf{K}_x^T \Gamma_x^{-1} \dot{\hat{K}}_x |\boldsymbol{\Lambda}|) + 2\,\mathbf{tr}(\Delta\mathbf{K}_r^T \Gamma_r^{-1} \dot{\hat{\mathbf{K}}}_r |\boldsymbol{\Lambda}|)$$

Appendix: Stability Analysis

$$= \mathbf{e}^T(\mathbf{P}\,\mathbf{A}_m + \mathbf{A}_m^T\mathbf{P})\mathbf{e} + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\left(\tilde{\mathbf{K}}_y^T\mathbf{y} + \tilde{\mathbf{K}}_r^T\mathbf{r}\right)$$
$$+2\,\mathbf{tr}\left(\tilde{\mathbf{K}}_y^T\boldsymbol{\Gamma}_{\mathbf{y}}^{-1}\dot{\hat{\mathbf{K}}}_y|\boldsymbol{\Lambda}|\right) + 2\,\mathbf{tr}\left(\tilde{\mathbf{K}}_r^T\boldsymbol{\Gamma}_r^{-1}\dot{\hat{\mathbf{K}}}_r|\boldsymbol{\Lambda}|\right)$$

$$= -\mathbf{e}^T\mathbf{Q}\mathbf{e} - 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\varepsilon_f(\mathbf{y}) + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\tilde{\mathbf{K}}_y^T\mathbf{y}$$
$$+ 2\,\mathbf{tr}\left(\tilde{\mathbf{K}}_y^T\boldsymbol{\Gamma}_{\mathbf{y}}^{-1}\dot{\hat{\mathbf{K}}}_y\right) + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\tilde{\mathbf{K}}_r^T\mathbf{r} + 2\,\mathbf{tr}\left(\Delta\mathbf{K}_r^T\boldsymbol{\Gamma}_r^{-1}\dot{\hat{\mathbf{K}}}_r\right)$$

Notice $x^T\,y = \mathbf{tr}\left(y\,x^T\right)$ from trace identity

Appendix: Stability Analysis Cont'd

Therefore,

$$\begin{aligned}\dot{\mathbf{V}}(\cdot) &= -\mathbf{e}^T\mathbf{Q}\mathbf{e} - 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\varepsilon_f \\ &+ 2\,\mathbf{tr}\left(\tilde{\mathbf{K}}_y^T(\boldsymbol{\Gamma}_{\boldsymbol{y}}^{-1}\dot{\hat{\mathbf{K}}}_y + \mathbf{y}\mathbf{e}^T\mathbf{P}\mathbf{B}sgn(\boldsymbol{\Lambda})\right)|\boldsymbol{\Lambda}| \\ &+ 2\,\mathbf{tr}\left(\tilde{\mathbf{K}}_r^T(\boldsymbol{\Gamma}_r^{-1}\dot{\hat{\mathbf{K}}}_r + \mathbf{r}\mathbf{e}^T\mathbf{P}\mathbf{B}sgn(\boldsymbol{\Lambda})\right)|\boldsymbol{\Lambda}|\end{aligned}$$

where for a real-valued x, we have $x = \text{sgn}(x)|x|$.

- first two terms will be negative definite for all $\mathbf{e} \neq 0$
  - since $\mathbf{A}_m$ is Hurwitz
- other terms will be identically null if we choose the adaptation laws

$$\dot{\hat{\mathbf{K}}}_y = -\boldsymbol{\Gamma}_{\boldsymbol{y}}\mathbf{y}\mathbf{e}^T\mathbf{P}\,\mathbf{B}\text{sgn}(\boldsymbol{\Lambda}), \quad \dot{\hat{\mathbf{K}}}_r = -\boldsymbol{\Gamma}_r\mathbf{r}\mathbf{e}^T\mathbf{P}\,\mathbf{B}\,\text{sgn}(\boldsymbol{\Lambda})$$

Appendix: Stability Results

We find that

$$\begin{aligned}\dot{\mathbf{V}}(\cdot) &= -\mathbf{e}^T\mathbf{Q}\mathbf{e} - 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Lambda}\varepsilon_f \\ &\le -\lambda_{low}\|\mathbf{e}\|^2 + 2\|\mathbf{e}\|\|\mathbf{P}\mathbf{B}\|\lambda_{high}(\boldsymbol{\Lambda})\varepsilon_{max}\end{aligned}$$

- $\lambda_{low}, \lambda_{high} \equiv$ minimum and maximum characteristic roots of Q and $\boldsymbol{\Lambda}$ respectively.
- $\dot{\mathbf{V}}(\cdot)$ is thus negative definite outside the compact set
- $\chi = \left(\mathbf{e} : \|\mathbf{e}\| \le \frac{2\|\mathbf{P}\mathbf{B}\|\lambda_{high}(\boldsymbol{\Lambda})\varepsilon_{max}(\mathbf{y})}{\lambda_{low}(\mathbf{Q})}\right)$
- thus, we conclude that the error $\mathbf{e}$ is uniformly ultimately bounded.
  - i.e. $\mathbf{y}(t) \to 0$ as $t \to \infty$

Publications Until Now

- Publications Under Review
  - **Olalekan Ogunmolu**, Xuejun Gu, Steve Jiang, Nicholas Gans. Nonlinear Systems Identification Using Deep Dynamic Neural Networks". Under review at *Physical Review Applied*, American Physical Society, Submitted September 2018.
  - **Olalekan Ogunmolu**, Michael Folkerts, Dan Nguyen, Nicholas Gans, Steve Jiang. Deep BOO 2.0: End-to-End Training of Beam Orientation Selection Policies for Intensity Modulated Radiation Therapy. Under publication invitation review at *International Journal of Robotics Research* (*IJRR*), 2018.
- Working Manuscripts
  - **Olalekan Ogunmolu**, Ayaka Kume, Jethro Tan. A stable Lyapunov approach for designing deep policies for complex robot motion tasks. Robotics and Automation Letters/International Conference on Robotics and Automation (RA-L). 2019.
  - **Olalekan Ogunmolu**, Nick Gans, Xuejun Gu, and Steve Jiang. Simulation and control of a head and neck patient pose correction soft-robot mechanism in intensity modulated radiotherapy. Transactions on Robotics (T-RO) 2018/2019.
- Accepted Publications
  - **Olalekan Ogunmolu**, Michael Folkerts, Dan Nguyen, Nicholas Gans, and Steve Jiang. Deep BOO: Automating Beam Orientation Selection in Intensity Modulated Radiation Therapy. To appear at *The 13th International Workshop on the Algorithmic Foundations of Robotics* (*WAFR*), Mérida, Mexico. December 2018.

Publications Until Now

- Accepted Publications (Cont'd)
  - **Olalekan Ogunmolu**, Nicholas Gans, Tyler Summers. Minimax Iterative Dynamic Game: Application to Nonlinear Robot Control Tasks. *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Madrid, Spain. October 2018.
  - **Olalekan Ogunmolu**, Dan Nguyen, Xun Jia, Weiguo Lu, Nicholas Gans, and Steve Jiang. Automating Beam Orientation Optimization for IMRT Treatment Planning: A Deep Reinforcement Learning Approach. Selected for Oral Presentation at the *John R. Cameron Young Investigators Symposium – 60th Annual Meeting of the American Association of Physicists in Medicine*, Nashville, TN (*AAPM*). July 2018.
  - Yara Almubarak, Joshi Aniket, **Olalekan Ogunmolu**, Xuejun Gu, Steve Jiang, Nicholas Gans, and Yonas Tadesse, Design and Development of Soft Robots for Head and Neck Cancer Radiotherapy. *SPIE: Smart Structures + Nondestructive Evaluation*, (*SPIE*), Denver, CO, U.S.A. March 2018.
  - **Olalekan Ogunmolu**, Adwait Kulkarn, Yonas Tadesse, Xuejun Gu, Steve Jiang, and Nicholas Gans. Soft-NeuroAdapt: A 3-DOF Neuro-Adaptive Pose Correction System For Frameless and Maskless Cancer Radiotherapy. *IEEE/RSJ International Conference on Intelligent Robots and Systems* (*IROS*), Vancouver, BC, Canada. September 2017. DOI: 10.1109/IROS.2017.8206211.

Publications Until Now

- Accepted Publications (Cont'd)
  - **Olalekan Ogunmolu**, Xuejun Gu, Steve Jiang, and Nicholas Gans. Vision-based control of a soft-robot for Maskless Cancer Radiotherapy. *IEEE Conference on Automation Science and Engineering (CASE)*, Fort-Worth, Texas, August 2016. DOI: 10.1109/CoASE.2016.7743378.
  - **Olalekan Ogunmolu**, Xuejun Gu, Steve Jiang, and Nicholas Gans. A Real-Time Soft-Robotic Patient Positioning System for Maskless Head-and-Neck Cancer Radiotherapy. *IEEE Conference on Automation Science and Engineering (CASE)*, Gothenburg, Sweden, August 2015. DOI: 10.1109/CoASE.2015.7294318.

References

References

Marios M Polycarpou and Petros A Ioannou. Modelling, Identification And Stable Adaptive Control Of Continuous-time Nonlinear Dynamical Systems Using Neural Networks. In *American Control Conference, 1992*, pages 36–40. IEEE, 1992.

Olalekan Ogunmolu, Nicholas Gans, and Tyler Summers. Minimax iterative dynamic game: Application to nonlinear robot control tasks. *IEEE International Conference on Robots and Intelligent Systems, (IROS), Madrid, 2018.*, 2018.

Steve Webb. *Intensity-Modulated Radiation Therapy*. Institute of Physics Publishing Ltd, Bristol and Philadelphia, 2001.

Laura I Cerviño, Todd Pawlicki, Joshua D Lawson, and Steve B Jiang. Frame-less and mask-less cranial stereotactic radiosurgery: a feasibility study. *Physics in medicine and biology*, 55(7):1863, 2010.

Xinmin Liu, Andrew H Belcher, Zachary Grelewicz, and Rodney D Wiersma. Robotic Stage for Head Motion Correction in Stereotactic Radiosurgery. In *American Control Conference (ACC), 2015*, pages 5776–5781. IEEE, 2015.

Mark Ostyn, Thomas Dwyer, Matthew Miller, Paden King, Rachel Sacks, Ross Cruikshank, Melvin Rosario, Daniel Martinez, Siyong Kim, and Woon-Hong Yeo. An electromechanical, patient positioning system for head and neck radiotherapy. *Physics in Medicine & Biology*, 62(18): 7520, 2017.

Johannes Heinrich, Marc Lanctot, and David Silver. Fictitious self-play in extensive-form games. In *International Conference on Machine Learning*, pages 805–813, 2015.

Eulalie Coevoet, Adrien Escande, and Christian Duriez. Optimization-based inverse model of soft robots with contact handling. *IEEE Robotics and Automation Letters*, 2 (3):1413–1419, 2017.

James M Bern, Grace Kumagai, and Stelian Coros. Fabrication, modeling, and control of plush robots. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 3739–3746. IEEE, 2017.

Isuru S Godage, Gustavo A Medrano-Cerda, David T Branson, Emanuele Guglielmino, and Darwin G Caldwell. Dynamics for variable length multisection continuum arms. *The International Journal of Robotics Research*, 35(6):695–722, 2016.

Federico Renda, Michele Giorelli, Marcello Calisti, Matteo Cianchetti, and Cecilia Laschi. Dynamic model of a multibending soft robot arm driven by cables. *IEEE Transactions on Robotics*, 30(5):1109–1122, 2014.

Rongjie Kang, David T Branson, Emanuele Guglielmino, and Darwin G Caldwell. Dynamic modeling and control of an octopus inspired multiple continuum arm robot. *Computers & Mathematics with Applications*, 64(5):1004–1016, 2012.