A Data Driven Approach to Control of Large Scale Systems

Suman Chakravorty[1]

[1]Department of Aerospace Engineering

TEXAS A&M UNIVERSITY

Second International Conference on InfoSymbiotics/ DDDAS, Cambridge, August 7-9, 2017

Acknowledgements: Mohammadhussein RafieSakhaei, Dan Yu and P. R. Kumar.
AFOSR DDDAS program, NSF NRI program.



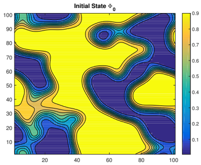
Motivation

- Deep Reinforcement Learning
- AlphaGO, Humanoid motion, Quadruped...

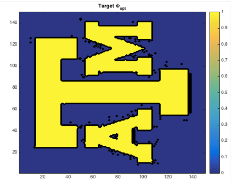
Figure: Deep RL Successes

Motivation

- Extend to Partially Observed Systems?
- Can we extend to very large scale systems such as those governed by PDEs, for instance, Materials Process Design?
- Application of the DDDAS paradigm in RL.



(a) Initial State



(b) Target

Outline

- Preliminaries
- The Curse of Dimensionality (COD)
- Remedies for the COD
- A Separation Principle
- Reinforcement Learning
- Conclusion

Preliminaries

- Control dependent transition density $p(x'/x,u)$ and a cost function $c(x,u)$.
- Stochastic Optimal Control Problem/ Markov Decision Problem (MDP):

$$J_T(x_0) = \min_{u_t(.)} E[\sum_{t=0}^{T} c(x_t, u_t(x_t)) + g(x_T)].$$

- Dynamic Programming Equation:

$$J_N(x) = \min_u \{c(x,u) + E[J_{N-1}(x')]\}, J_0(x) = g(x),$$
$$u_N^*(x) = \arg\min_u \{c(x,u) + E[J_{N-1}(x')]\}.$$

Preliminaries

- Sensing uncertainty given by measurement likelihood $p(z/x)$ $\rightarrow$ Partially Observed/ Belief Space Problem (POMDP):

$$J_T(b_0) = \min_{u_t(.)} E[\sum_{t=0}^{T} c(b_t, u_t(b_t)) + g(b_T)],$$

$$J_N(b) = \min_u \{c(b,u) + E[J_{N-1}(b')]\}, J_0(b) = g(b).$$

- $b(x)$ denotes the "belief state" / pdf of the state governed by the recursive Bayesian Filtering equation.

The Curse of Dimensionality

- Richard Bellman, the discoverer of MDPs and the DP equation, also coined the term "the Curse of Dimensionality".
- Refers to the phenomenon that the complexity of the DP problem increases exponentially in the dimension of the state space of the problem!
- Naively speaking, discretizing the DP equation on a grid with K intervals:

$$J_N(x_i) \approx \min_u \{c(x_i, u) + \sum_j p(x_j/x_i, u) J_{N-1}(x_j)]\},$$

we have to solve a nonlinear recursion with K^d variables.

ADP/ RL

- **Approximate Dynamic Programming (ADP)/ Reinforcement Learning (RL) techniques [1].**
- Policy Evaluation step in policy iteration for discounted DP: we want to evaluate the cost-to-go under a given policy $\mu(.)$, say $J^{\mu}(.)$.
- Assume that the cost-to-go can be linearly parametrized in terms of some "smart" basis functions $\{\phi_1(x), \phi_2(x)..., \phi_K(x)\}$: $J^{\mu}(x) = \sum_{i=1}^{N} \alpha_i \phi_i(x)$.

ADP/ RL

- Policy Evaluation reduces to solving the linear equation for the co-efficients α_i of the cost-to-go function:

$$[I - \beta L]\bar{\alpha} = \bar{c}, \text{ where } \bar{c} = [c_i],$$

$$c_i = \int \underbrace{c(x, \mu(x))}_{c^{\mu}(x)} \phi_i(x) dx, \; i = 1, 2, \cdots N;$$

$$L_{ij} = \int \int p^{\mu}(x'/x)\phi_i(x')\phi_j(x) dx' dx, i, j = 1 \cdots N.$$

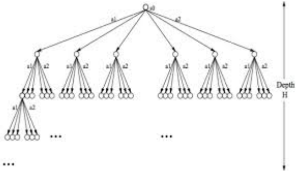
- The integrals above can either be evaluated analytically, for instance, using quadratures, or via Monte Carlo sampling trajectories $\{x_t\}$ as in RL: $L_{ij} \approx \frac{1}{M}\sum_{t=0}^{M-1} \phi_i(x_{t+1})\phi_j(x_t)$.

ADP/ RL

- The issue is that the number of samples required to get a "good" estimate of L_{ij}, and hence the cost-to-go, is still exponential in the dimension of the problem.
- This is due to the fact that a sparse basis Φ is usually never known a priori $\rightarrow$ the number of basis functions is still exponential in the dimension of the problem.
- The set of learning experiments is largely done using heuristics.

MPC

- **Model Predictive Approach**: rather than solve the DP problem backward in time, these approaches explore the reachable space forward in time from a given state [2, 3, 4].
- As shown in the seminal paper [2], these methods are no longer subject to exponential complexity in dimension of the problem.



MPC

- However, the method scales as $(|A||C|)^D$ where D is the depth of the lookahead tree, $|A|$ is the number of actions and $|C|$ is the number of children from every action required for a good estimate of the cost-to-go.
- May be infeasible for continuous state, observation and action space problems.

MPC

- **Model Predictive Control [5]**: rather than solve the DP problem, it solves the deterministic open loop (noise-less) problem at every time step:

$$J_T(x_0) = \min_{u_t} \sum_{t=0}^{T} c(x_t, u_t) + g(x_T).$$

- Can be shown to coincide with DP solution in deterministic systems.
- However, for systems with uncertainty, the MPC approach is heuristic since the optimization above needs to be over control policies $u_t(.)$, and not a control sequence u_t.
- MPC approaches typically fully observed.

A Separation Principle

- Let the transition function be described by the following state space model:

$$x_t = f(x_{t-1}, u_{t-1}, \epsilon w_{t-1}),$$

  where w_t is a white noise sequence, and $\epsilon > 0$ is a "small" parameter.
- Let the feedback law be of the form $u_t(x_t) = \bar{u}_t + K_t \delta x_t$, where $\delta x_t = x_t - \bar{x}_t$, $\bar{x}_t = f(\bar{x}_{t-1}, \bar{u}_{t-1}, 0)$, and K_t is some linear time varying feedback gain.

Basic Idea

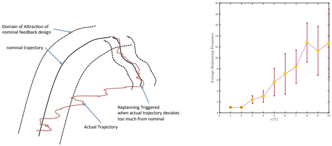
- Let the cost of the nominal trajectory (plan) be given by $\bar{J}_T$ and let the sample stochastic cost be given by $J_T(\omega)$.
- **Main Result:** Given ϵ is sufficiently small, $J_T = \bar{J}_T + \delta J$, and $E[\delta J] = 0$.
- This implies $E(J_T) = \bar{J}_T$, for any nominal control sequence, which in turn implies that this is true also for the optimal sequence.
- Hence, in the small noise case, optimizing the open loop sequence $\bar{u}_t$, and wrapping a (linear) feedback law around it subsequently is near optimal (coincides with DP)!

Basic Idea

- **Separation Principle:** *We may design the open loop optimal law, without considering feedback, since it does not affect the stochastic optimal cost, and hence, the design of the open loop and the closed loop in Stochastic Optimal Control can be separated.*
- Unlike MPC, the design considers the feedback, but shows that it is decoupled from the open loop design.

Basic Idea

- Practically, it means that we do not have to replan at every time step as in MPC.



(d) Replanning (e) Replan freq. vs noise

Figure: Replanning is typically a very rare event ($O(\frac{1}{\epsilon})$ time steps)

Belief Space Generalization (T-LQG)

- **Belief Space Generalization (T-LQG):** Let the observation model be given by, $z_k = h(x_k) + v_k$.
- Assume belief is Gaussian.
- The open loop plan optimizes the nominal, or most likely, evolution of the Gaussian belief, (μ_t, P_t), in particular, it may optimize some measure of the nominal covariance evolution obtained by setting $w_k, v_k = 0$.

Belief Space Generalization (T-LQG)

- The closed loop is designed to track the nominal belief where $u_t(x_t) = \bar{u}_t + K_t(\hat{x}_t - \mu_t)$, K_t is the feedback gain, $\hat{x}_t$ is an estimate of the state from a Kalman filter with gain L_t.
- Ricatti equations for K_t and L_t are decoupled due to the "Separation Principle" of Linear Control theory: reduces complexity of feedback design from $O(d^4)$ to $O(d^2)$.
- Belief space Planning $\rightarrow$ Separation^2!
- Answer to Feldbaum's dual control in the small noise case.

Belief Space Generalization (T-LQG)

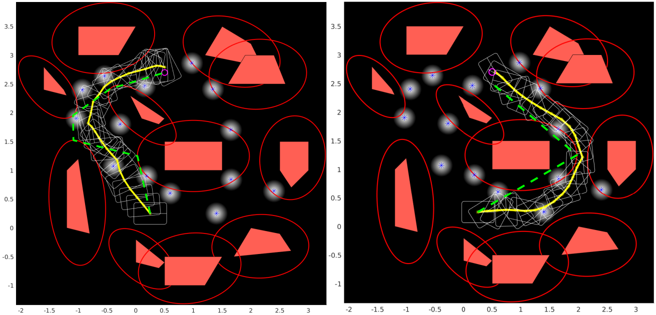


Figure: Youbot base in a complex environment. Solid lines: optimized planned trajectories; dashed lines: optimization initialization trajectories.

Separation based RL

- Reinforcement Learning (RL) "learns" a feedback policy for an unknown nonlinear system from experiments. Access only to a forward generative black-box model.
- The Separation Principle suggests a novel path to accomplish RL.
- The open loop plan $\rightarrow$ optimizing the control sequence $\rightarrow$ a series of *gradient descent* steps $\rightarrow$ a sequence of linear problems.
- The closed loop design $\rightarrow$ identifying a linear time varying (LTV) system around the optimized nominal trajectory.

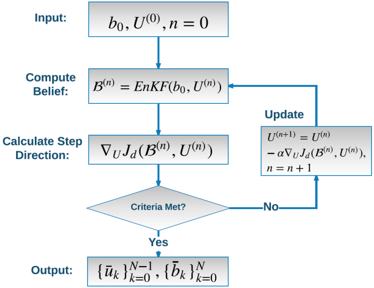
Separation based RL

- Linear Systems are completely determined by their impulse responses.
- This implies we can specify an exact sequence of experiments to perform in order to “learn” the feedback law.
- Allows us to scale to extremely large scale problems: partially observed Partial Differential Equation (PDE) constrained problems.

Separation based RL

Step 1. Open-Loop Trajectory Optimization in Belief Space

Given b_0, solve the deterministic open loop belief state optimization problem (access only to state simulator):



$$\{\bar{u}_k\}_{k=0}^{N-1} = \underset{\{u_k\}}{\text{argmin}} \bar{J}(\{b_k\}, \{u_k\}),$$

$$\text{s.t.} \quad b_{k+1} = \tau(b_k, u_k, \bar{y}_{k+1}),$$

Experiments: $\delta \bar{J}$ given δu_k, for all k.

Separation based RL

Step 2. Linear Time-varying System Identification

Linearize the system around $(\{\bar{\mu}_k\},\{\bar{u}_k\})$ (Only conceptually).

$$\delta x_{k+1} = A_k \delta x_k + B_k(\delta u_k + w_k), \quad \delta y_k = C_k \delta x_k + v_k,$$

Experiments: δy_n given an input δu_k, for all k, n.

Identified deviation system (using time-varying ERA):

$$\delta a_{k+1} = \hat{A}_k \delta a_k + \hat{B}_k(\delta u_k + w_k), \quad \delta y_k = \hat{C}_k \delta a_k + v_k,$$

where $\delta a_k \in \Re^{n_r}, \delta x_k \in \Re^{n_x}$, and $n_r \ll n_x$.

Step 3. Closed Loop Controller Design

Standard LQG controller can be designed for the $A(.), B(.), C(.)$ (or) can learn controller/ estimator directly.

Burgers Equation

Consider the optimal boundary control problem for the Burgers equation (a 1-d analog of the Navier-Stokes equation):

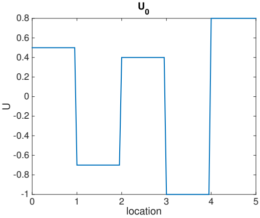
$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} = \mu\frac{\partial^2 U}{\partial x^2},$$

$U(x,t)$: states, μ: viscosity.
Boundary control: $U(0,t) = u_1(t), U(L,t) = u_2(t)$.
Initial condition: $U(.,0) = U_0$,
Control Objective: $U(.,t) = -0.8, t \in [7s, 8s]$.

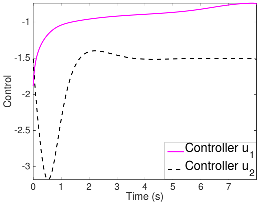
Burgers Equation

System Parameters:

System Dimension	Inputs	Outputs	Identified LTV
100	2	5	10



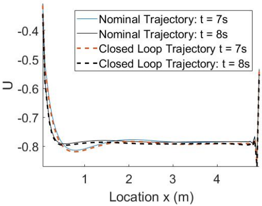
(a) State Evolution



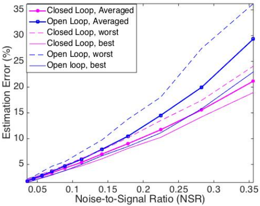
(b) Open Loop Optimal Control

Burgers Equation

Run 100 Monte Carlo Simulations.



(a) Comparison of Closed Loop Belief Trajectory



(b) Comparison of Estimation Error

Online Controller and Estimator Complexity Reduction: $O(10^6)$.

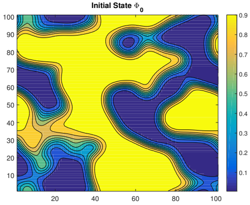
Materials Process Design

Allen-Cahn Phase Field Model:

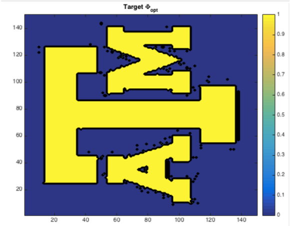
$$\frac{1}{K}\frac{\partial \phi}{\partial t} = \nabla^2 f(\phi, T) - U g'(\phi),$$

$\phi(x,t)$: phase field variable, $T(x,t)$: temperature controller

K, U: constant



(a) Initial State



(b) Target

Conclusion

- The Separation Principle greatly simplifies Stochastic Optimal Control design in a decoupled open loop-closed loop fashion.
- Rigorously generalizes MPC to (partially observed) Stochastic Control problems.
- Allows us to propose a novel RL algorithm that specifies an exact set of experiments to design a feedback plan for a given black-box system.
- Allows us to scale RL to very large scale and partially observed problem: Generalized Motion Planning problems governed by PDEs.
- RL approach needs noise-less simulations.
- Multi-Agent system implications.

References

[1] D. Bertsekas, *Dynamic Programming and Optimal Control: 3rd Ed.* Athena Scientific, 2007.

[2] M. Kearns, Y. Mansour, and A. Ng, "A sparse sampling algorithm for near-optimal planning in large Markov Decision Processes," in *Proceedings of the IJCAI*, 1999.

[3] H. Bai, D. Hsu, W. Lee, and V. Ngo, "Monte carlo value iteration for continuous-state pomdps," *Algorithmic foundations of robotics IX*, pp. 175–191.

[4] H. Kurniawati, D. Hsu, and W. Lee, "SARSOP: Efficient point-based pomdp planning by approximating optimally reachable belief spaces," in *Proceedings of Robotics: Science and Systems*, 2008.

[5] D. Q. Mayne, "Model predictive control: Recent developments and future promise," *Automatica*, vol. 50, pp. 2967–2986, 2014.

[6] J. Van Den Berg, P. Abbeel, and K. Goldberg, "Lqg-mp: Optimized path planning for robots with motion uncertainty and imperfect state information," *The International Journal of Robotics Research*, vol. 30, no. 7, pp. 895–913, 2011.

[7] J. Van Den Berg, S. Patil, and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.

[8] W. Sun, J. van den Berg, and R. Alterovitz, "Stochastic extended lqr for optimization-based motion planning under uncertainty," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 437–447, 2016.

[9] R. Platt, R. Tedrake, L. Kaelbling, and T. Lozano-Perez, "Belief space planning assuming maximum likelihood observatoins," in *Proceedings of Robotics: Science and Systems (RSS)*, June 2010.