Thesis Proposal Defense

Brett Bethke
Aerospace Controls Lab, MIT

December 5, 2008

Outline

- Introduction
  - Thesis goals & contributions
  - Literature review
- Work to date
- Proposed work

Introduction

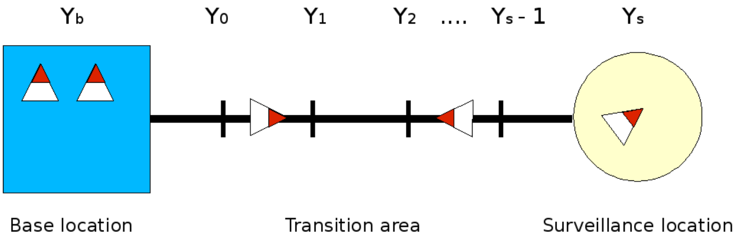
- Overall thesis objective: development of new strategies for addressing multi-agent planning problems under uncertainty
  - In particular, focus on modeling and solving "health management" problems
- Main areas of thesis contributions:
  1. "Health-aware" multi-agent planning problems as MDPs (formulating a meaningful problem of interest)
  2. Kernel-based approximate dynamic programming algorithms (development of general methods to solve the problems)
  3. Online adaptation to changing / poorly known models (solving the problem in the face of model uncertainty)
  4. Flight demonstrations (experimental verification of the usefulness of the proposed problem models and solution techniques)

Literature Review

- Recent advances from machine learning community starting to be applied to Approximate Dynamic Programming (ADP)
  - Kernelized approximate linear programming formulation [11]
  - Kernelized approximate value iteration [17, 10]
  - TD learning using Gaussian processes [14, 13, 12, 19]
  - LSTD using support vector machines [23]
  - Manifold-based kernels as cost approximation architectures [21, 22, 15, 16, 3, 2, 20]
- But kernel-based ADP is a young area of research...

Area 1: "Health-aware" MDP Formulations

- Persistent surveillance under stochastic fuel usage
  - Goal: maintain a specified number of UAVs over a surveillance area at all times
  - UAVs have finite fuel capacity. Amount of fuel used at each time step is a random variable
  - UAVs can refuel at base, but crash if they run out of fuel while flying
  - Surveillance area far from base location → takes finite time to fly between the two, replacement UAVs must be dispatched early
  - Publication: [9] (ACC '08)



Area 2: Kernel-based ADP

- Would like to be able to solve large problems $\rightarrow$ approximation methods needed
- General observations / motivation
  - Nonparametric, kernel-based techniques (support vector regression, Gaussian process regression, etc) provide powerful and flexible cost approximation architectures
  - Bellman residual approaches: evaluate the policy μ by solving

$$\min_{\widetilde{J}_\mu \in \mathcal{J}} \sum_{i \in \widetilde{\mathcal{S}}} \left(\widetilde{J}_\mu(i) - T_\mu \widetilde{J}_\mu(i) \right)^2,$$

    then perform policy improvement
  - Objective function bounded below by zero. Goal: find a cost function $\widetilde{J}_\mu$ that achieves this bound (Bellman Residual Elimination)
  - Publications: [1, 8, 7] (CDC '08, JMLR '08, ACC '09)

Area 2: Kernel-based ADP - Basic Idea I

1. Select a kernel function:

$$k(i, i') = \langle \underline{\Phi}(i), \underline{\Phi}(i') \rangle$$

2. Functional form of the cost function:

$$\widetilde{J}_\mu(i) = \langle \underline{\Theta}, \underline{\Phi}(i) \rangle$$

where $\underline{\Theta}, \underline{\Phi}(i) \in \mathcal{H}_k$

- Identical to the standard "linear combination of basis functions" approach, except that the dimensionality of $\underline{\Theta}$ and $\underline{\Phi}(i)$ may be very large...

Area 2: Kernel-based ADP - Basic Idea II

3. Rewrite Bellman residual:

$$
\begin{aligned}
BR(i) &= \widetilde{J}_\mu(i) - T_\mu \widetilde{J}_\mu(i) \\
&= \widetilde{J}_\mu(i) - \left(g_i^\mu + \alpha \sum_{j \in \mathcal{S}} P_{ij}^\mu \widetilde{J}_\mu(j) \right) \\
&= \langle \underline{\Theta}, \underline{\Phi}(i) \rangle - \left(g_i^\mu + \alpha \sum_{j \in \mathcal{S}} P_{ij}^\mu \langle \underline{\Theta}, \underline{\Phi}(j) \rangle \right) \\
&= -g_i^\mu + \langle \underline{\Theta}, \left(\underline{\Phi}(i) - \alpha \sum_{j \in \mathcal{S}} P_{ij}^\mu \underline{\Phi}(j) \right) \rangle \\
&= -g_i^\mu + \langle \underline{\Theta}, \underline{\Psi}(i) \rangle
\end{aligned}
$$

Area 2: Kernel-based ADP - Basic Idea III

4. Using the new feature mapping $\underline{\Psi}(i)$, define the associated Bellman kernel

$$\mathcal{K}(i, i') = \langle \underline{\Psi}(i), \underline{\Psi}(i') \rangle$$

and the associated residual function $\widetilde{W}_\mu(i) \in \mathcal{H}_\mathcal{K}$

$$\widetilde{W}_\mu(i) \equiv \langle \underline{\Theta}, \underline{\Psi}(i) \rangle$$

5. The desired property

$$BR(i) = 0 \quad \forall i \in \widetilde{\mathcal{S}}$$

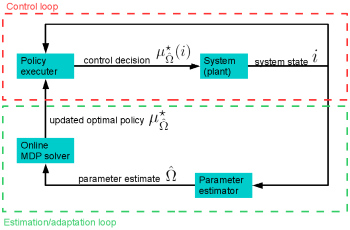
is now equivalent to the regression problem

$$\widetilde{W}_\mu(i) = g_i^\mu \quad \forall i \in \widetilde{\mathcal{S}}$$

We can solve this regression problem using any kernel-based regression technique (support vector regression, Gaussian process regression, etc)

Area 3: Online Adaptation

- Our BRE algorithms are model-based
  - In many applications, general form of the dynamic equations may be known (but may exhibit parametric uncertainty)
- What if the model is poorly known and/or changing with time?
- Have developed an online adaptation mechanism to simultaneously estimate system model and re-solve the MDP
- Advantage: separates MDP solution from model estimation



- Publications: [6, 5, 4, 18] (GNC '08, RAM '08, ACC '09, Infotech '09)

Area 4: Large-scale implementation / Flight Experiments

- Bellman Residual Elimination algorithm calculations are amenable to distributed computation
- Have designed and implemented large-scale, distributed software architecture for testing BRE on large problems
  - Uses Message Passing Interface (MPI), a parallel computing framework originally developed for supercomputers
  - Currently running experiments on a 24-processor cluster
  - Implementation scalable to 1000s of processors

Proposed Work

- For completion of the thesis, the following areas of work are proposed:
  - Further BRE Algorithm Development/Extension
  - Large-Scale Health Management Flight Demonstrations

Further BRE Algorithm Development/Extension

- "*n*-stage" Bellman Residual Elimination: solving

$$T_{\mu}^{n} J_{\mu} = J_{\mu}$$

- Investigation of manifold-based kernels, and their relationship to *n*-stage BRE
- Further decentralization of BRE (i.e. when computational nodes have limited communication bandwidth)
- Extension of BRE to model-free learning: stochastic approximations of the associated Bellman kernel

Large-Scale Health Management Flight Demonstrations

- Continue using high-performance BRE implementation to experiment with solving large-scale problems
- Use computed policy as planning element for persistent surveillance flight demonstrations in RAVEN
  - Goal: demonstrate applicability and usefulness of both the persistent surveillance problem formulation as well as the BRE solution technique
- Use BRE algorithms for adaptive planning with online model estimation

References I

B. Bethke, J. How, A. Ozdaglar.
Approximate Dynamic Programming Using Support Vector Regression.

In *Proceedings of the 2008 IEEE Conference on Decision and Control*, Cancun, Mexico, 2008.

M. Belkin and P. Niyogi.
Semi-supervised learning on riemannian manifolds.
Machine Learning, 56(1-3):209–239, 2004.

M. Belkin and P. Niyogi.
Towards a theoretical foundation for laplacian-based manifold methods.
In Peter Auer and Ron Meir, editors, *COLT*, volume 3559 of *Lecture Notes in Computer Science*, pages 486–500. Springer, 2005.

References II

L. Bertuccelli, B. Bethke, and J. How.
Robust adaptive markov decision processes in multi-vehicle applications.
In *Proceedings of the American Control Conference (to appear)*, 2009.

B. Bethke, L. Bertuccelli, and J. How.
Real-time adaptive mdp-based planning.
IEEE Robotics and Automation Magazine (to appear), 2008.

B. Bethke, L. Bertuccelli, and J. P. How.
Experimental Demonstration of MDP- Based Planning with Model Uncertainty.
In *AIAA Guidance Navigation and Control Conference*, Aug 2008. AIAA-2008-6322.

References III

B. Bethke and J. How.
Approximate dynamic programming using bellman residual elimination and gaussian process regression.
In *Proceedings of the American Control Conference (to appear)*, 2009.

B. Bethke, J. How, and A. Ozdaglar.
Kernel-based reinforcement learning using bellman residual elimination.
Journal of Machine Learning Research (to appear), 2008.

B. Bethke, J. How, and J. Vian.
Group health management of UAV teams with applications to persistent surveillance.
In *Proceedings of the American Control Conference*, 2008.

References IV

M. Deisenroth, J. Peters, and C. Rasmussen.
Approximate dynamic programming with gaussian processes.
In *Proceedings of the American Control Conference*, 2008.

T. Dietterich and X. Wang.
Batch value function approximation via support vectors.
In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani, editors, *NIPS*, pages 1491–1498. MIT Press, 2001.

Y. Engel.
Algorithms and Representations for Reinforcement Learning.
PhD thesis, Hebrew University, 2005.

References V

Y. Engel, S. Mannor, and R. Meir.
Reinforcement learning with gaussian processes.
In Luc De Raedt and Stefan Wrobel, editors, *ICML*, volume 119 of *ACM International Conference Proceeding Series*, pages 201–208. ACM, 2005.

Y. Engel, S. Mannor, and Ron Meir.
Bayes meets bellman: The gaussian process approach to temporal difference learning.
In Tom Fawcett and Nina Mishra, editors, *ICML*, pages 154–161. AAAI Press, 2003.

S. Mahadevan.
Proto-value functions: Developmental reinforcement learning.
In *International Conference on Machine Learning*, 2005.

References VI

S. Mahadevan and M. Maggioni.
Value function approximation with diffusion wavelets and laplacian eigenfunctions.
In *NIPS*, 2005.

D. Ormoneit and Ś. Sen.
Kernel-Based Reinforcement Learning.
Machine Learning, 49(2):161–178, 2002.

J. Redding, B. Bethke, L. Bertuccelli, and J. How.
Experimental demonstration of exploration toward model learning under an adaptive mdp-based planner.
In *Proceedings of the AIAA Infotech Conference (to appear)*, 2009.

References VII

J. Reisinger, P. Stone, and R. Miikkulainen.
Online kernel selection for bayesian reinforcement learning.
In *Proceedings of the 25th International Conference on Machine Learning*, 2008.

W. Smart.
Explicit manifold representations for value-function approximation in reinforcement learning.
In *AMAI*, 2004.

M. Sugiyama, H. Hachiya, C. Towell, and S. Vijayakumar.
Geodesic gaussian kernels for value function approximation.
In *Workshop on Information-Based Induction Sciences*, 2006.

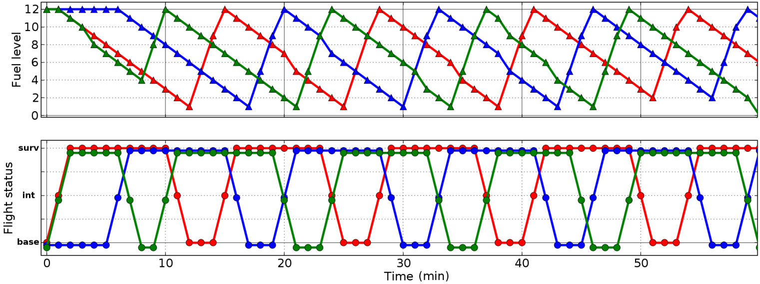
References VIII

M. Sugiyama, H. Hachiya, C. Towell, and S. Vijayakumar.
Value function approximation on non-linear manifolds for robot motor control.
In *Proc. of the IEEE International Conference on Robotics and Automation*, 2007.

J. Tobias and P. Daniel.
Least squares svm for least squares td learning.
In Gerhard Brewka, Silvia Coradeschi, Anna Perini, and Paolo Traverso, editors, *ECAI*, pages 499–503. IOS Press, 2006.

Area 1: Persistent Surveillance Results

- For small problems, exact solution using value iteration is possible
- Example: 3 UAVs, 2 requested at surveillance area



- Health-aware properties of the optimal solution
  - UAVs return to base with extra fuel (hedge against fuel usage uncertainty)
  - Green UAV returns well before out of fuel (allows a regular switching pattern to be established)

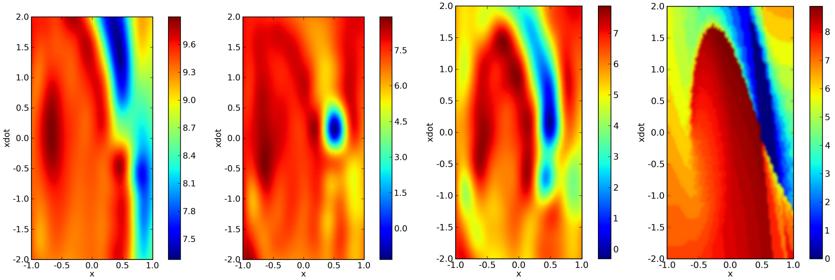
Area 2: Kernel-based Approximate Dynamic Programming

- Advantages of our approach:
  - Bellman residuals provably zero at the sample states
  - Provably exact ($\widetilde{J}_\mu = J_\mu$) in the limit $\widetilde{\mathcal{S}} \to \mathcal{S}$
  - No trajectory simulations required $\to$ no simulation noise effects
  - Algorithm based on Gaussian process regression provides natural error bounds on the solution and allows for automatic adjustment of kernel hyperparameters
  - Computational requirements scale with the number of sample states chosen (under designer's control)
    - Entire algorithm distributable over many computational resources

BRE(SV) Results

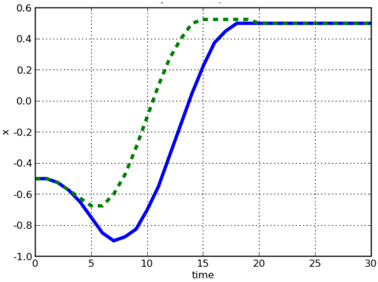
- Mountain car problem, with 9x9 grid of sample states
- Using BRE(SV) (support vector regression variant)
- Kernel function:

$$k((x_1, \dot{x}_1), (x_2, \dot{x}_2)) = \exp\left(-(x_1 - x_2)^2/(0.25)^2 - (\dot{x}_1 - \dot{x}_2)^2/(0.40)^2\right).$$



BRE(SV) Results

- System response:

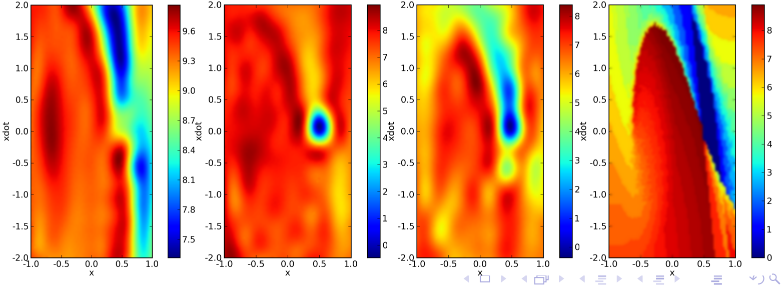


- Questions:
  - How to choose kernel parameters?
  - Error bounds?

BRE(GP) Results

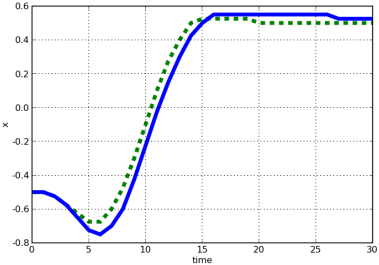
- BRE(GP) (Gaussian process regression variant) can address these questions
- Automatically learns kernel parameters using marginal likelihood maximization
- Provides error bounds using posterior covariance
- Mountain car kernel function (poorly known initial parameters):

$$k((x_1, \dot{x}_1), (x_2, \dot{x}_2); \underline{\Omega}) = \exp\left(-(x_1 - x_2)^2/(\Omega_1)^2 - (\dot{x}_1 - \dot{x}_2)^2/(\Omega_2)^2\right).$$



BRE(GP) Results

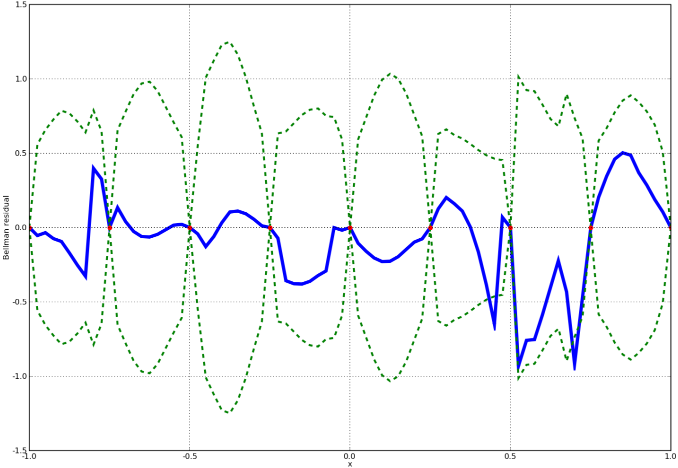
- System response:



- BRE(GP) successfully and automatically identifies a better set of kernel parameters than were chosen by hand for BRE(SV)

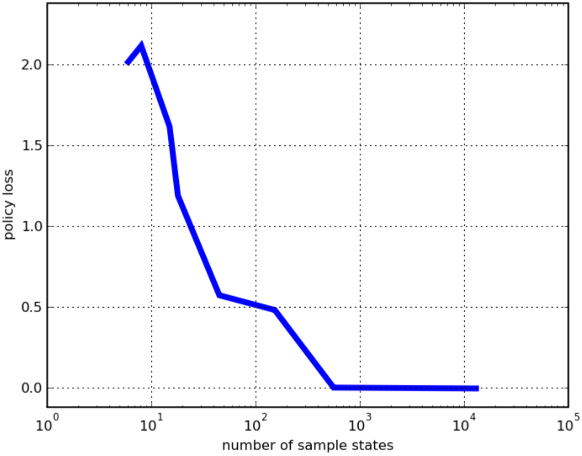
BRE(GP) Results

- Verify that Bellman residual are zero at sample states
- Examine 2σ error bounds



BRE(GP) Results

- Verify that BRE(GP) yields optimal policy in limit of sampling entire space



Area 3: Results

- Can utilize "bootstrapping" to reduce time needed to solve the MDP online, given a previous solution
- Flight results, using value iteration as the MDP solution mechanism: