Project Topics:

Here are some suggestions and references. The references listed here contain more references that may be suitable for projects:

- **Methodological aspects of approximate DP with cost function approximation** (policy evaluation, approximate and optimistic policy iteration, Q-learning, etc)

  The web pages of Dimitri Bertsekas, John Tsitsiklis, Ben Van Roy, Huizhen (Janey) Yu, and Shie Mannor

  http://www.mit.edu/~dimitrib/home.html
  http://www.mit.edu/~jnt/home.html
  http://www.stanford.edu/~bvr/
  http://www.mit.edu/~janey_yu/note_asaproofs.pdf
  http://www.ece.mcgill.ca/~smanno1/pubs.htm

- **Issues of Exploration in Approximate Policy Iteration**

  The material on geometric and free-form sampling in 6.4 of Vol. II of your text, and the associated relevant references.

  Bertsekas, D. P., "Lambda-Policy Iteration: A Review and a New Implementation", Lab. for Information and Decision Systems Report LIDS-P-2874, MIT, October 2011. (This is new and your instructor would love to see some computational evaluation/comparisons.)

  Bertsekas, D. P., and Yu, H., "Q-Learning and Enhanced Policy Iteration in Discounted Dynamic Programming," Lab. for Information and Decision Systems Report LIDS-P-2831, MIT, April, 2010 (revised October 2010); Math. of OR (to appear).

  Sutton, R. S., and Barto, A. G., 1998. Reinforcement Learning, MIT Press, Cambridge, MA. (The discussion of on-policy and off-policy training methods)

- **Q-Learning and Policy Iteration**

  Tsitsiklis, J. N., 1994. "Asynchronous Stochastic Approximation and Q-Learning," Machine Learning, Vol. 16, pp. 185-202.

  Yu, H., and Bertsekas, D. P., 2007. "A Least Squares Q-Learning Algorithm for Optimal Stopping Problems," Lab. for Information and Decision Systems Report 2731, MIT; also in Proc. European Control Conference 2007, Kos, Greece.

  Bertsekas, D. P., and Yu, H., "Q-Learning and Enhanced Policy Iteration in Discounted Dynamic Programming," Lab. for Information and Decision Systems Report LIDS-P-2831, MIT, April, 2010 (revised October 2010).

  Yu, H., and Bertsekas, D. P., "Q-Learning and Policy Iteration Algorithms for Stochastic Shortest Path Problems," Lab. for Information and Decision Systems Report LIDS-P-2871, MIT, September 2011. (This is new and your instructor would love to see some computational evaluation/comparisons.)

  Yu, H., and Bertsekas, D. P., "On Boundedness of Q-Learning Iterates for Stochastic Shortest Path Problems," Lab. for Information and Decision Systems Report LIDS-P-2859, MIT, March 2011; revised Sept. 2011.

- **Monte Carlo Linear Algebra**

  The discussion and references of Section 7.3, Vol. II of your text.

- **Rollout Methodology for Deterministic and Stochastic Optimization**

  The discussion and references of Chapter 6, Vol. I is a good starting point (see also the applications papers below, which use rollout.)

- **Approximate DP via Linear Programming**

  de Farias, D. P., and Van Roy, B., 2003. "The Linear Programming Approach to Approximate Dynamic Programming," Operations Research, Vol. 51, pp. 850-865.

  de Farias, D. P., and Van Roy, B., 2004. "On Constraint Sampling in the Linear Programming Approach to Approximate Dynamic Programming," Mathematics of Operations Research, Vol. 29, pp. 462-478.

  Desai, V. V., Farias, V. F., Moallemi, C. C., 2009. "The Smoothed Approximate Linear Program," Research Report, MIT and Columbia University.

  The web pages of Profs. John Tsitsiklis, Vivek Farias, Ben Van Roy

  http://www.mit.edu/~jnt/home.html
  http://web.mit.edu/vivekf/www/mypapers2.html
  http://www.stanford.edu/~bvr/

- **Exact and Approximate Algorithms for Partially Observable MDP** There is a large number of survey articles on this topic. In Google Scholar type "POMDP Survey" to see several, ranging from the 1991 survey by Lovejoy to the present. Some other papers are listed below.

  Hauskrecht, M., 2000. "Value-Function Approximations for Partially Observable Markov Decision Processes, Journal of Artificial Intelligence Research, Vol. 13, pp. 33-95.

  Poupart, P., and Boutilier, C., 2004. "Bounded Finite State Controllers," Advances in Neural Information Processing Systems.

  Yu, H., and Bertsekas, D. P., 2004. "Discretized Approximations for POMDP with Average Cost," Proc. of the 20th Conference on Uncertainty in Artificial Intelligence, Banff, Canada.

  Yu, H., and Bertsekas, D. P., 2008. "On Near-Optimality of the Set of Finite-State Controllers for Average Cost POMDP," Mathematics of Operations Research, Vol. 33, pp. 1-11.

  Chong, E. K. P., Kreucher, C., and Hero, A. O., 2009. "Partially Observable Markov Decision Process Approximations for Adaptive Sensing," Discrete Event Dynamic Systems J. (check the references on particle filtering)

  Patek, S. D., Partially Observed Stochastic Shortest Path Problems With Approximate Solution by Neurodynamic Programming, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Sept. 2007.

- **Model Predictive Control;** related methodology and applications (a vast literature here in within the control systems field). The following two are based on a DP viewpoint and an optimization viewpoint, respectively.

Bertsekas, D. P., 2005. "Dynamic Programming and Suboptimal Control: A Survey from ADP to MPC," in Fundamental Issues in Control, European J. of Control, Vol. 11. (A survey with many other references.)

Y. Wang, and S. Boyd, 2009. "Performance bounds for linear stochastic control," Systems and Control Letters, Vol. 58, pp. 178-182.

The web page of Manfred Morari, ETH, Zurich

http://control.ee.ethz.ch/~morari/

- **Simulation-Based Scientific Computation and Least Squares Inference Problems**

Bertsekas, D. P., and Yu, H., 2009. "Projected Equation Methods for Approximate Solution of Large Linear Systems," Journal of Computational and Applied Mathematics, Vol. 227, pp. 27-50.

Bertsekas, D. P., 2011. "Temporal Difference Methods for General Projected Equations," IEEE Trans. on Automatic Control, Vol. 56, pp. 2128 - 2139.

Yu, H., 2010. "Least Squares Temporal Difference Methods: An Analysis Under General Conditions," Technical report C-2010-39, Dept. Computer Science, Univ. of Helsinki.

Yu, H., 2010. "Convergence of Least Squares Temporal Difference Methods Under General Conditions," Proc. of the 27th ICML, Haifa, Israel.

- **Distributed Synchronous and Asynchronous Dynamic Programming**

Bertsekas, D. P., and Yu, H., "Distributed Asynchronous Policy Iteration in Dynamic Programming," Proc. of 2010 Allerton Conference on Communication, Control, and Computing, Allerton Park, ILL, Sept. 2010.

Bertsekas, D. P., 1982. "Distributed Dynamic Programming," IEEE Trans. Automatic Control, Vol. AC-27, pp. 610-616.

Bertsekas, D. P., and Tsitsiklis, J. N., 1989. Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, N. J.; republished by Athena Scientific, Belmont, MA, 1997 (available on line).

Bertsekas, D. P., 2007. "Separable Dynamic Programming and Approximate Decomposition Methods," IEEE Trans. on Aut. Control, Vol. 52, pp. 911-916.

Busoniu, Babuska, and De Schutter, 2008. "A comprehensive survey of multiagent reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics.

Athena is MIT's UNIX-based computing environment. OCW does not provide access to it.

Venkat, A. N., Rawlings, J. B., Wright, S. J., 2007. Distributed model predictive control of large-scale systems, Lect. Notes in Control and Info., Springer (the 2006 thesis by Venkat, Distributed model predictive control: Theory and applications is also available on line)

Mannor, S., and Shamma, J. 2007. "Multi-agent Learning for Engineers," Artificial Intelligence, Vol. 171, pp. 417-422.

J. R. Kok, and N. Vlassis, 2006. "Collaborative Multiagent Reinforcement Learning by Payoff Propagation," Journal of Machine Learning Research, Vol. 7, pp. 17891828.

Barto, A. G., Mahadevan, S, 2003. "Recent advances in hierarchical reinforcement learning, Discrete Event Dynamic Systems J.

- **Approximation in Policy Space and Policy Gradient Methods**

Konda, V. R., and Tsitsiklis, J. N., 2003. "Actor-Critic Algorithms," SIAM J. on Control and Optimization, Vol. 42, pp. 1143-1166.

Konda, V. R., 2002. Actor-Critic Algorithms, Ph.D. Thesis, Dept. of EECS, M.I.T., Cambridge, MA.

Marbach, P., and Tsitsiklis, J. N., 2001. "Simulation-Based Optimization of Markov Reward Processes," IEEE Transactions on Automatic Control, Vol. 46, pp. 191-209.

Marbach, P., and Tsitsiklis, J. N., 2003. "Approximate Gradient Methods in Policy-Space Optimization of Markov Reward Processes," J. Discrete Event Dynamic Systems, Vol. 13, pp. 111-148.

Munos, R., 2005. "Policy gradient in continuous time," INRIA Report.

Many references on the subject at the end of the chapter 7 of Vol. II of your text.

- **Error Bounds for Approximate Dynamic Programming Methods**

Yu, H., and Bertsekas, D. P., 2010. "New Error Bounds for Approximations from Projected Linear Equations," Mathematics of Operations Research, Vol. 35, pp. 306-329.

Antos, Szepesvari, C., Munos, R., 2006. "Learning near-optimal policies with Bellman-residual minimization based fitted policy iteration and a single sample path," Conference On Learning Theory, Pittsburgh, USA; Extended version in Machine Learning, 2008.

Munos, R., Szepesvari, C., 2008. "Finite Time Bounds for Sampling-Based Fitted Value Iteration," Journal of Machine Learning Research.

Munos, R., 2005. "Error Bounds for Approximate Value Iteration," American Conference on Artificial Intelligence.

Other papers on bounds by R. Munos

- **Basis Function Adaptation** (the optimization of basis function selection in approximate DP, as well as the automatic basis function generation)

  Keller, P. W., Mannor, S., and Precup, D., 2006. "Automatic Basis Function Construction for Approximate Dynamic Programming and Reinforcement Laerning," Proc. of the 23rd ICML, Pittsburgh, Penn.

  Jung, T., and Polani, D., 2007. "Kernelizing LSPE(λ)," in Proc. 2007 IEEE Symposium on Approximate Dynamic Programming and Reinforcement Learning, Honolulu, Hawaii. pp. 338-345.

  I. Menache, S. Mannor, and N. Shimkin, 2005. "Basis function adaptation in temporal difference reinforcement learning," Annals of Operations Research, Vol. 134, pp. 215-238.

  Yu, H., and Bertsekas, D. P., "Basis Function Adaptation Methods for Cost Approximation in MDP," Proc. of IEEE International Symposium on Adaptive Dynamic Programming and Reinforcement Learning 2009, Nashville, TN.

- **Applications of Approximate DP in Specific Problem Areas,** e.g., scheduling, routing, trajectory planning, resource allocation, revenue management, dynamic investment, economic growth, queuing, adaptive sensing, etc. (There is a vast literature here, the following is just a small indicative sample.)

  Sun, Zhao, Luh, and Tomastik, 2008. "Optimization of joint replacement policies for multipart systems by a rollout framework," IEEE Transactions on Automation Science and Engineering.

  Bertsekas, D. P., and Castanon, D. A., 1999. "Rollout Algorithms for Stochastic Scheduling Problems," Heuristics, Vol. 5, pp. 89-108.

  Chong, E. K. P., Kreucher, C., and Hero, A. O., 2009. "Partially Observable Markov Decision Process Approximations for Adaptive Sensing," Discrete Event Dynamic Systems J. (has many additional refs by the same authors)

  Secomandi, N., 2001. "A Rollout Policy for the Vehicle Routing Problem with Stochastic Demands," Operations Research, Vol. 49, pp. 796-802.

  The web page of Vivek Farias has references on yield management
  http://web.mit.edu/vivekf/www/mypapers2.html

- **Approximate DP Applications in Finance**

  Longstaff, F. A., and Schwartz, E. S., 2001. "Valuing American Options by Simulation: A Simple Least-Squares Approach," Review of Financial Studies, Vol. 14, pp. 113-147.

  Li, Y., Szepesvari, C., and Schuurmans, D., 2009. "Learning Exercise Policies for American Options," Proc. of the Twelfth International Conference on Artificial Intelligence and Statistics, Clearwater Beach, Fla.

  Tsitsiklis, J. N., and Van Roy, B., 2001. "Regression Methods for Pricing Complex American-Style Options," IEEE Trans. on Neural Networks, Vol. 12, pp. 694-703.

  Yu, H., and Bertsekas, D. P., 2007. "A Least Squares Q-Learning Algorithm for Optimal Stopping Problems," Lab. for Information and Decision Systems Report 2731, MIT; also in Proc. European Control Conference 2007, Kos, Greece.

- **Adaptive Dynamic Programming**

  Bertsekas, D. P., "Value and Policy Iteration in Optimal Control and Adaptive Dynamic Programming", Lab. for Information and Decision Systems Report LIDS-P-3174, MIT, May 2015 (revised Sept. 2015), and the references given there.

MIT OpenCourseWare
http://ocw.mit.edu

6.231 Dynamic Programming and Stochastic Control

Fall 2015

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.