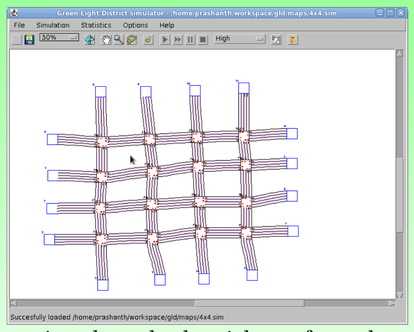
Stochastic Systems Lab

Faculty: Prof.Shalabh Bhatnagar **Students:** V.K. Mishra, Ajin G. J, Prashanth. L.A, Sindhu. P. R, Prasad. H.L, Abhranil. C, Lakshmanan. K, Saswata. C, Chandrashekar. L, Nikhil. M, Ravindra. V, Prabuchandran. K. J., Sunil. K. M, Brijendra Kumar, Debarghya. G

http://stochastic.csa.iisc.ernet.in

Road Traffic Control



Reinforcement learning based algorithms for adaptive traffic signal control.

Ref.: Prashanth L. A. and Shalabh Bhatnagar, "Reinforcement Learning with Function Approximation for Traffic signal control", IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 2, pp.412-421, 2011.

Online General-Sum Stochastic Games

For agent $i \in \{1, 2, \ldots, N\}$, let $v^i : S \to \Re$ be value function, $\pi^i : S \times A^i \to [0, 1]$ be policy and β be the discounted factor. Also, let $r^i \in \Re$ be the reward to agent i for taking action $a^i \in A^i$ in current state $x \in S$. Let $y \in S$ be the next state. Then, an update procedure,

$$\pi^i_{n+1}(x, a^i) = \Gamma(\pi^i_n(x, a^i) + b(n)\sqrt{\pi^i_n(x, a^i)}g^i_n),$$
$$v^i_{n+1}(x) = v^i_n(x) + c(n)g^i_n,$$

where $g^i_n = r^i + \beta v^i(y) - v^i(x)$, will converge to (v^*, π^*) with π^* being a Nash strategy-tuple of the underlying N-agent general-sum discounted stochastic game.

Ref.: Prasad H. L., Shalabh Bhatnagar, "Algorithms for Nash Equilibria in General-Sum Stochastic Games", Submitted to JMLR, 2011.

Actor-Critic algorithm for Constrained MDP

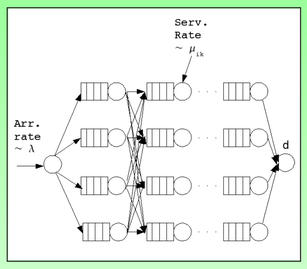
We consider the problem of control subject to inequality constraints for both discounted cost and long-run average cost criteria. We incorporate function approximation in both the objective and constraint functions resulting in a corresponding constrained parameter optimization problem. The general problem has the following form:

$$\min_{\theta} \quad J(\theta)$$
$$\text{s.t} \quad G_i(\theta) \leq \alpha_i, i = 1, \ldots, N.$$

We develop the first actor-critic algorithms in this setting of function approximation.

Ref.: S.Bhatnagar, An actor-critic algorithm with function approximation for discounted cost constrained Markov decision processes, Systems and Control Letters, Vol. 59, pp.760-766, 2010

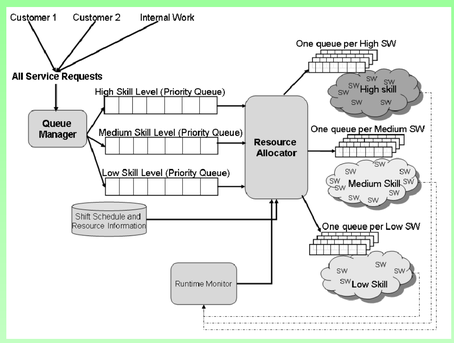
queueing networks



- Stochastic gradient algorithms for routing in communication networks using the total end-to-end delay.
- Actor-Critic and Q-learning based algorithms in networks with arbitrary topology and non-identical servers.

Ref.: K. Lakshmanan and Shalabh Bhatnagar, "Smoothed functional and Quasi-Newton Algorithms for Routing in Multi-stage Queueing Network with Constraints", ICDCIT 2011.

Service Systems



Problem:
Find the optimal staffing levels in a service system (SS) for a given dispatching policy (mapping from service requests to service workers) while maintaining system steady-state and compliance to aggregate SLA constraints.

Our Work:
We developed several stochastic optimization algorithms that performed significantly better than the state-of-the-art OptQuest optimization toolkit, which is being used in IBM's pool simulation project. We incorporated SPSA and smoothed functional based gradient estimation for 'primal descent' in these algorithms.

Ref: Prashanth L.A., H.L.Prasad, N.Desai, S.Bhatnagar and G.Dasgupta, Stochastic optimization for adaptive labor staffing in service systems, Proceedings of 9th International Conference on Service Oriented Computing (ICSOC) (accepted), 2011.

Actor-critic algorithms based on ALP

Details: We solve infinite horizon discounted reward markov decision process (MDP). Motivated by the approximation properties of ALP, we explore the use of an ALP based critic in the actor-critic framework.

ALP-critic suffers from two important limitations

1. Oscillatory Convergence behavior of primal-descent-dual-ascent ($PDDA$) scheme.
2. Large number of constraint in case of problems with large number of states.

The limitations are overcome by

1. Restricting the approximation to state aggregation.
2. Solving a Reduced Approximate Linear Program, with fewer number of constraints.
3. Adapting the constraints of the $RALP$ to include the active constraints of the ALP-critic by means of smoothed function gradient scheme.

Stability Conditions for Asynchronous SA

1. Consider a system comprising n processors each of which updates a parameter component using a stochastic approximation (SA) scheme.
2. At the end of this computation, each processor passes this information to all other processors which reaches other processor with a random delay.
3. Each processor performs updates using its own local clock and with the most recent information on updates it has about the other processors.
4. Our work develops such conditions for the case of asynchronous stochastic updates with delays as described above.

Ref.: S.Bhatnagar, The Borkar-Meyn Theorem for Asynchronous Stochastic Approximations, Systems and Control Letters, Vol. 60, pp. 472-478, 2011.

Dynamic Mechanism Design

1. We design a dynamic mechanism with dynamic types, where the agents have limited capacity.
2. We show that our mechanism possesses the desirable properties of allocative efficiency and incentive compatibility. Further, we also show that our mechanism is 'marginally compensating' the loss caused by an agent's misreport.
3. We are investigating application of this mechanism to the problem of work dispatch in service systems.

Ref.: Prashanth L.A., H.L.Prasad, N.Desai and S.Bhatnagar, Dynamic Mechanism Design with Capacity Constraints, AAMAS, Submitted 2011.

Stochastic Systems Lab

Faculty: Prof.Shalabh Bhatnagar Students: V.K. Mishra, Ajin G. J, Prashanth. L.A, Sindhu. P. R, Prasad. H.L, Abhranil. C, Lakshmanan. K, Saswata. C, Chandrashekar. L, Nikhil. M, Ravindra. V, Prabuchandran. K. J., Sunil. K. M, Brijendra Kumar, Debarghya. G

http://stochastic.csa.iisc.ernet.in

Road Traffic Control

Reinforcement learning based algorithms for adaptive traffic signal control.

Ref.: Prashanth L. A. and Shalabh Bhatnagar, "Reinforcement Learning with Function Approximation for Traffic signal control", IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 2, pp.412-421, 2011.

Online General-Sum Stochastic Games

For agent $i \in \{1, 2, \ldots, N\}$, let $v^i : S \to \Re$ be value function, $\pi^i : S \times A^i \to [0, 1]$ be policy and β be the discounted factor. Also, let $r^i \in \Re$ be the reward to agent i for taking action $a^i \in A^i$ in current state $x \in S$. Let $y \in S$ be the next state. Then, an update procedure,

$$\pi^i_{n+1}(x, a^i) = \Gamma(\pi^i_n(x, a^i) + b(n)\sqrt{\pi^i_n(x, a^i)}g^i_n),$$
$$v^i_{n+1}(x) = v^i_n(x) + c(n)g^i_n,$$

where $g^i_n = r^i + \beta v^i(y) - v^i(x)$, will converge to (v^*, π^*) with π^* being a Nash strategy-tuple of the underlying N-agent general-sum discounted stochastic game.

Ref.: Prasad H. L., Shalabh Bhatnagar, "Algorithms for Nash Equilibria in General-Sum Stochastic Games", Submitted to JMLR, 2011.

Actor-Critic algorithm for Constrained MDP

We consider the problem of control subject to inequality constraints for both discounted cost and long-run average cost criteria. We incorporate function approximation in both the objective and constraint functions resulting in a corresponding constrained parameter optimization problem. The general problem has the following form:

$$\min_{\theta} \quad J(\theta)$$
$$\text{s.t} \quad G_i(\theta) \leq \alpha_i, i = 1, \ldots, N.$$

We develop the first actor-critic algorithms in this setting of function approximation.

Ref.: S.Bhatnagar, An actor-critic algorithm with function approximation for discounted cost constrained Markov decision processes, Systems and Control Letters, Vol. 59, pp.760-766, 2010

queueing networks

- Stochastic gradient algorithms for routing in communication networks using the total end-to-end delay.
- Actor-Critic and Q-learning based algorithms in networks with arbitrary topology and non-identical servers.

Ref.: K. Lakshmanan and Shalabh Bhatnagar, "Smoothed functional and Quasi-Newton Algorithms for Routing in Multi-stage Queueing Network with Constraints", ICDCIT 2011.

Service Systems

Problem:
Find the optimal staffing levels in a service system (SS) for a given dispatching policy (mapping from service requests to service workers) while maintaining system steady-state and compliance to aggregate SLA constraints.

Our Work:
We developed several stochastic optimization algorithms that performed significantly better than the state-of-the-art OptQuest optimization toolkit, which is being used in IBM's pool simulation project. We incorporated SPSA and smoothed functional based gradient estimation for 'primal descent' in these algorithms.

Ref: Prashanth L.A., H.L.Prasad, N.Desai, S.Bhatnagar and G.Dasgupta, Stochastic optimization for adaptive labor staffing in service systems, Proceedings of 9th International Conference on Service Oriented Computing (ICSOC) (accepted), 2011.

Actor-critic algorithms based on ALP

Details: We solve infinite horizon discounted reward markov decision process (MDP). Motivated by the approximation properties of ALP, we explore the use of an ALP based critic in the actor-critic framework.

ALP-critic suffers from two important limitations

1. Oscillatory Convergence behavior of primal-descent-dual-ascent ($PDDA$) scheme.
2. Large number of constraint in case of problems with large number of states.

The limitations are overcome by

1. Restricting the approximation to state aggregation.
2. Solving a Reduced Approximate Linear Program, with fewer number of constraints.
3. Adapting the constraints of the $RALP$ to include the active constraints of the ALP-critic by means of smoothed function gradient scheme.

Stability Conditions for Asynchronous SA

1. Consider a system comprising n processors each of which updates a parameter component using a stochastic approximation (SA) scheme.
2. At the end of this computation, each processor passes this information to all other processors which reaches other processor with a random delay.
3. Each processor performs updates using its own local clock and with the most recent information on updates it has about the other processors.
4. Our work develops such conditions for the case of asynchronous stochastic updates with delays as described above.

Ref.: S.Bhatnagar, The Borkar-Meyn Theorem for Asynchronous Stochastic Approximations, Systems and Control Letters, Vol. 60, pp. 472-478, 2011.

Dynamic Mechanism Design

1. We design a dynamic mechanism with dynamic types, where the agents have limited capacity.
2. We show that our mechanism possesses the desirable properties of allocative efficiency and incentive compatibility. Further, we also show that our mechanism is 'marginally compensating' the loss caused by an agent's misreport.
3. We are investigating application of this mechanism to the problem of work dispatch in service systems.

Ref.: Prashanth L.A., H.L.Prasad, N.Desai and S.Bhatnagar, Dynamic Mechanism Design with Capacity Constraints, AAMAS, Submitted 2011.