Decentralized Stochastic Planning via Approximate Dynamic Programming

Silvia Ferrari
Paul Ruffin Scarborough Associate Professor of Engineering
Department of Mechanical Engineering and Materials Science
Duke University

2011 ONR MURI: Nonparametric Bayesian Models to Represent Knowledge and Uncertainty for Decentralized Planning

Year 1 Review Meeting, MIT, Boston (MA)
October 19, 2012

Technical Accomplishments – Year 1

- Analysis of **Distributed Optimal Control** (DOC) Method and Algorithms.
  - Important for performing decentralized stochastic planning and control over large spatial and temporal scales

- Developed new **information value functions**: (1) representing the probability of multiple detections for maneuvering targets represented by Markov motion models, and (2) representing the value of information in NPBM.
  - Important for obtaining DOC performance functions that are integral function of **X**, and of nonparametric Bayesian models of sensed environment and target behaviors

- Developed **approximate dynamic programming** (ADP) approach for **hybrid systems**.
  - Important for performing distributed learning through ADP, for teams of heterogeneous autonomous static and mobile agents, which typically involve both discrete and continuous state and control variables

- Developed a **decentralized KDE-consensus algorithm** for computing DOC control laws for individual agents, through the diffusion of local inferences and optimality conditions.
  - Important for implementing decentralized planning for large-scale autonomous agents with limited communications

Optimal Control Framework

- Approximate dynamic programming and control based on optimal control problem:

An integral objective function of state and control,

$$J = \phi\left[\mathbf{x}(t_f), t_f\right] + \int_{t_0}^{t_f} L\left[\mathbf{x}(t), \mathbf{u}(t), t\right] dt, \quad \text{with I.C. } \mathbf{x}(t_0)$$

is to be optimized w.r.t. $\mathbf{u}(t)$ and $\mathbf{x}(t)$, subject to the **agent** dynamics,

$$\dot{\mathbf{x}}(t) = \mathbf{f}\left[\mathbf{x}(t), \mathbf{p}(t), \mathbf{u}(t), t\right]$$

and subject to equality and inequality mission constraints

$$\mathbf{c}\left[\mathbf{x}(t), \mathbf{u}(t)\right] \geq 0$$

- Approximated dynamic programming (ADP) can be applied to the above OC problem to learn to improve performance continuously over time, subject to modeling errors, parameter variations, and partial state information.

➢ Technical Challenge: The computations required to solve the above optimal control problem for many agents with decoupled dynamics, but couplings in rewards and constraints (e.g. TI-MDPs), are prohibitive.

Distributed Optimal Control Framework

Objective: Develop decentralized learning and planning theory and algorithms for stochastic multiscale dynamical systems.

- The system is comprised of many agents or processes that, on small spatial and time scales, can each be described by a detailed microscopic model,

ODE: $\dot{\mathbf{x}}_i(t) = \mathbf{f}[\mathbf{x}_i(t), \mathbf{u}_i(t), \mathbf{w}_i(t)], \quad i = 1, \ldots, N$

or

MDP: $\theta_i = \{S_i, A_i, P_i(s,s'), R_i(s,s')\}, \quad i = 1, \ldots, N$

→ "Solution Operator"

$\mathbf{x}_i(k + \Delta t) = \mathcal{T}_{d_i}^{t} \mathbf{x}_i(k), \quad \mathbf{x}_i \in \Re^n,$

$N >> 1$

- On larger spatial and time scales, the interactions of microscopic agents give rise to macroscopic **coherent** behavior or coarse dynamics, and performance.

- The macroscopic description $\mathbf{X} \in \Re^M, M << N,$ is based on the statistics of interest, and determines the *restriction operator* $\mathcal{M}$, and an appropriate lifting operator μ, s.t.,

$\mathbf{X} = \mathcal{M}\mathbf{x}_i \rightarrow \mathcal{T}_c^{\tau} = \mathcal{M}\mathcal{T}_{d_i}^{\tau}\mu$ "Coarse Time Stepper" with τ = coarse time.

$\mathcal{M}$ may involve some averaging, and could consist of a probability density function (PDF), its moments, or a maximum likelihood (ML) inference based field estimator.

- Define $\mathcal{M}$ based on the decentralized nonparametric models (DP and BP) and covariates.

Distributed Optimal Control (DOC)

Assume the macroscopic state of the agents can be represented by a restriction operator, such as a probability density function (PDF): $p[\mathbf{x}(t), t]$.

The distribution of agents $p[\mathbf{x}(t),t]$ is to be optimized such that its macroscopic performance,

$$J = \phi\{p[\mathbf{x}_j(t_f), t_f]\} + \int_{t_0}^{t_f} L\{p[\mathbf{x}_j(t), t], \mathbf{u}_j(t), t\} dt,$$

is maximized, subject to the microscopic agent's dynamic equation,

$$\dot{\mathbf{x}}_j(t) = \mathbf{f}[\mathbf{x}_j(t), \mathbf{u}_j(t), t], \quad j = 1, \ldots, N$$

and subject to equality and inequality constraints on the agent's microscopic state and control

$$\mathbf{c}_j[\mathbf{x}_j(t), \mathbf{u}_j(t)] \geq 0$$

Theoretical Results:

- ✓ Necessary conditions for optimality
- ✓ Conservation law analysis
- ✓ Numerical method of solution based on finite volume (FV) approach
- ✓ Computational complexity analysis

DOC Problem and Solution

Assuming agents are neither created nor destroyed inside the pre-defined region of interested (ROI), A, the macroscopic dynamic equation consists of the partial differential equation (PDE) known as **advection equation**:

$$\frac{\partial p[\mathbf{x}(t),t]}{\partial \mathbf{x}} = -\nabla \cdot \{p[\mathbf{x}(t),t] \;\; \mathbf{f}[\mathbf{x}(t),\mathbf{u}(t),t]\}$$

$$\begin{cases} \text{IC: } p[\mathbf{x}(t_0),\, t_0] = p_0. \\ \text{BC: } p[\mathbf{x} \in \partial A, t] = 0 \end{cases}$$

Furthermore, the distribution must obey the normalization condition,

$$\int_A p[\mathbf{x}(t),t]d\mathbf{x} = 1$$

and the constraints $p[\mathbf{x} \notin A, t] = 0$

which indicate the support of the distribution is the interior of A.

Comparing this problem formulation with the classical optimal control problem, it can be seen that classical optimality conditions do not apply.

DOC constitutes a new class of optimal control problem, where a time-varying probability density function, $p(\cdot)$, is to be determined by optimizing its performance over time, subject to a PDE.

DOC Optimality Conditions

- Introduce the following Hamiltonian: $H[p,\mathbf{u},t] \equiv L[p,\mathbf{u},t] + \lambda(t)\, p(\nabla \cdot \mathbf{f})$

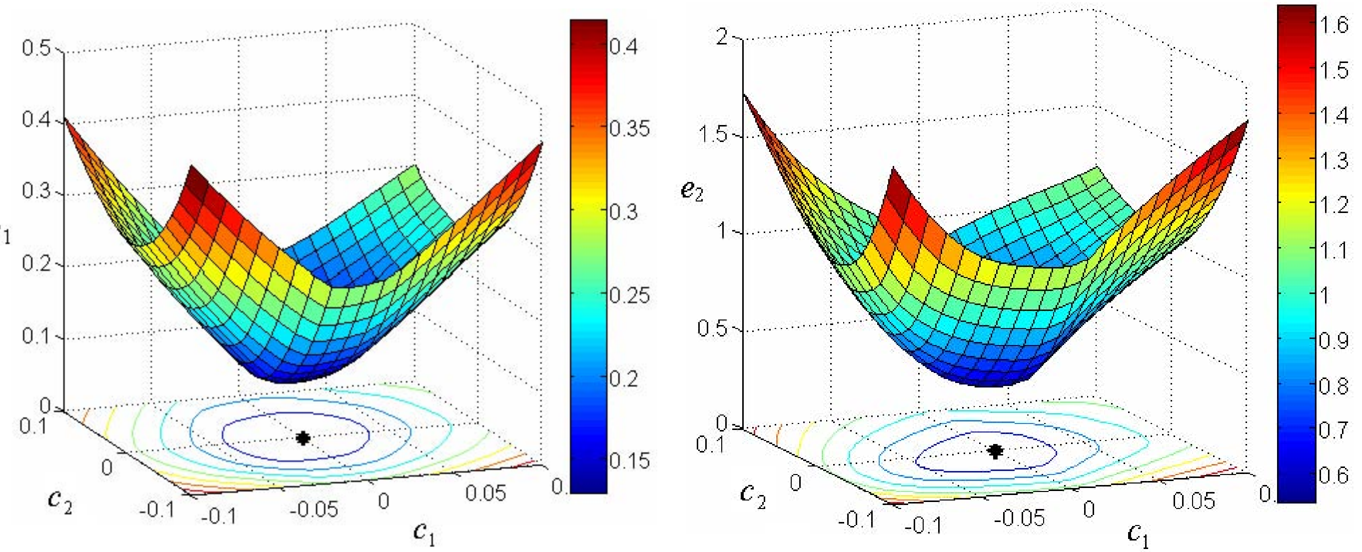
- Using Calculus of Variations, the following optimality conditions are obtained in the form of PDEs, and must be satisfied for $t_0 \le t \le t_f$,

$$\dot{\lambda}\nabla p = \frac{\partial L[\cdot]}{\partial \mathbf{x}} + \lambda\left\{\nabla p(\nabla \cdot \mathbf{f}) + p[\nabla \mathbf{F}]^T - \nabla^2 p \cdot \mathbf{f} - \frac{\partial}{\partial t}(\nabla p)\right\}$$

$$\frac{\partial L[p,\mathbf{u},t]}{\partial \mathbf{u}} + \lambda p[\nabla \mathbf{G}]^T = \mathbf{0},$$ where, $\mathbf{F} \equiv \partial \mathbf{f} / \partial \mathbf{x}$ and $\mathbf{G} \equiv \partial \mathbf{f} / \partial \mathbf{u}$

and subject to the boundary conditions (BCs) provided by the normalization condition.

Parametric study:
(numerical validation)



Agent's Feedback Control Law

- Each agent i moves according to a potential navigation function defined as a linear combination of an attractive potential, which depends on the optimal density of agents $p^*(\mathbf{x}, t)$, and a repulsive potential for local objectives (e.g. collision avoidance).

$$U(\mathbf{x}_i, t) = w_1 \cdot U_{att}(\mathbf{x}_i, t) + w_2 \cdot \sum_{l=1, l \neq i} U_{l_{rep}}(\mathbf{x}_i, t)$$

Where, $U_{att}(\mathbf{x}_i, t) = \hat{p}(\mathbf{x}_i, t) - p^*(\mathbf{x}_i, t + t_d)$ $\quad t_d$ = time-shifting parameter,

and $$U_{l_{rep}}(\mathbf{x}_i, t) = \begin{cases} \frac{1}{2}\left(\frac{1}{\|\mathbf{x}_i(t) - \mathbf{x}_l(t)\|} - \frac{1}{\rho_0}\right)^2 & if \ \|\mathbf{x}_i(t) - \mathbf{x}_l(t)\| \leq \rho_0 \\ 0 & if \ \|\mathbf{x}_i(t) - \mathbf{x}_l(t)\| > \rho_0 \end{cases}$$

where ρ_0 is a distance-of-influence parameter of the repulsive potential.

- The feedback control input for the i^{th} agent is found by following the direction of the negative gradient of the navigation function, i.e.:

$$-\nabla U(\mathbf{x}_i, t) = -w_1 \cdot \nabla U_{att}(\mathbf{x}_i, t) - w_2 \cdot \sum_{i=1, i \neq j} \nabla U_{l_{rep}}(\mathbf{x}_i, t)$$

Computational Complexity Analysis

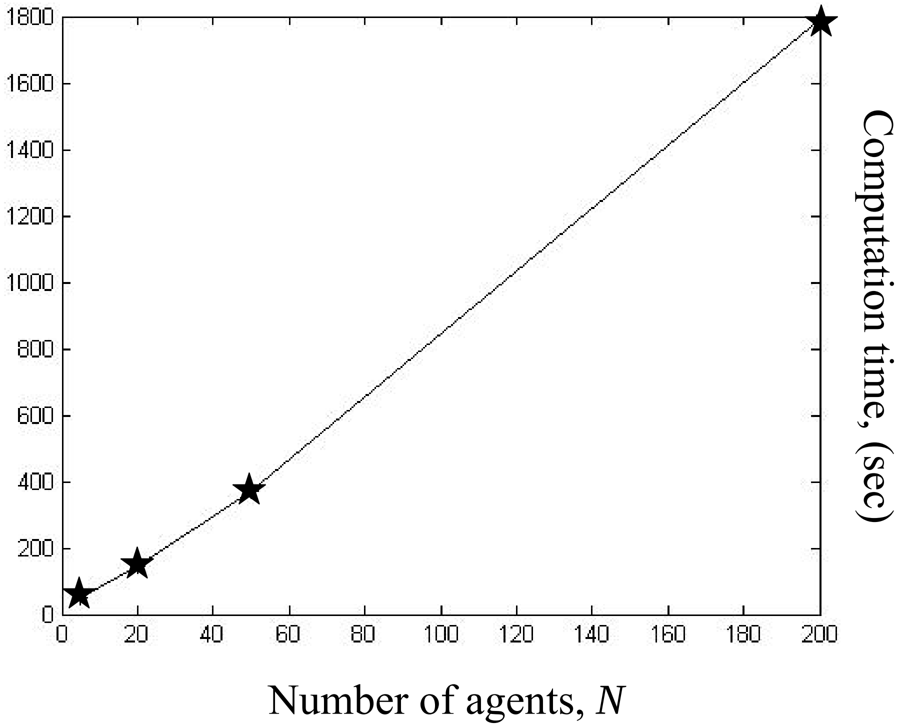
- Since the optimization is performed on the macroscopic agent distribution, the number of agents does not influence the computation time of the optimal PDF.
- The computation time required by the *centralized* agents' microscopic control laws varies linearly with the number of agents, N.

Optimal Control (OC) Solution:

Subproblem	DOC	Classical OC
Hessian update	$O(zXK^2)$	$O(nmN^2K^2)$
QP	$O(z^2XK^3)$	$O(nm^2N^2K^3)$
Line search	$O(XK)$	$O(nNK)$

Dimensions:
z = components; n = agent state; m = agent controls; X = state collocation points; K = time collocation points; N = number of agents.

Agent Control Law Computation:

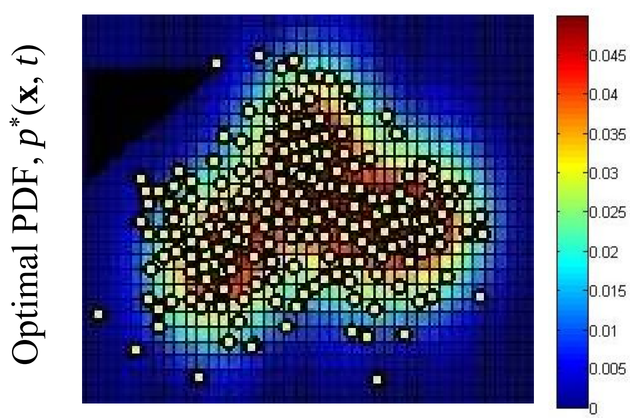


Decentralized DOC Method

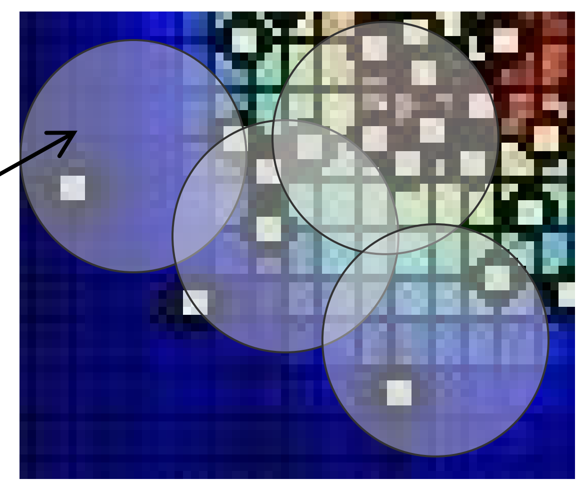
- Couplings between agents arise primarily through common mission objectives

$$J = \phi\{p[\mathbf{x}_j(t_f), t_f]\} + \int_{t_0}^{t_f} L\{p[\mathbf{x}_j(t), t], \mathbf{u}_j(t), t\} dt,$$

- The optimal PDF, $p^*(\mathbf{x}, t)$, represents ideal macroscopic state (incl. couplings), optimized subject to microscopic dynamics and controls (reachability).
- Decentralized DOC: integrate DOC control law with gossip-like paradigm for the diffusion of local inferences, control laws, and connectivity constraints, and vice versa for proving reachability under limited and local communication assumptions.



Local (microscopic) communication



Decentralized KDE for Control

- The DOC feedback control law can be calculated in a decentralized manner by estimating the actual agent PDF $\hat{p}$, using **decentralized kernel density estimation**.

$$U_{att}(\mathbf{x}_i, t_k) = \boxed{\hat{p}(\mathbf{x}_i, t_k)} - \boxed{p^*(\mathbf{x}_i, t_k + t_d)}$$

Macroscopic performance

- Instead of using centralized estimation of the actual agents' PDF, each agent maintains a local estimation, governed by a stored kernel set,

$$S_i = \left\{ \left\langle w_{i,k}, \mathbf{x}_{i,k}, \mathbf{H}_{i,k} \right\rangle, k = 1, \ldots, N_i \right\}$$

- Initially, each agent has only one kernel stored (centered at its own position) and shares kernel information with other sensors through an **information spreading** protocol.

Kernel parameters for the *i*th sensor:

$K_{\mathbf{H}_{i,k}}$ = kernel	$\mathbf{H}_{i,k}$ = bandwidth matrix
$w_{i,k}$ = weighting coefficients	N_i = number of kernels stored by sensor *i*

Decentralized KDE Estimate of Agents' PDF

❖ The k^{th} kernel stored by agent i, and centered at $\mathbf{x}_j$, is defined as,

$$K_{\mathbf{H}_{i,k}}(\mathbf{x}-\mathbf{x}_i) \equiv \left|\mathbf{H}_{i,k}\right|^{-1/2} K(\mathbf{H}_{i,k}{}^{-1/2}(\mathbf{x}-\mathbf{x}_j))$$

where the kernel function is chosen as the standard two-dimensional Gaussian kernel,

$$K(\mathbf{x}) = \frac{1}{2\pi} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}$$

❖ Then the local sensor PDF estimation by sensor i is calculated as a weighted sum of kernels,

$$\hat{p}_i(\mathbf{x}_i, t_k) = \sum_{k=1}^{N_i} w_{i,k} K_{\mathbf{H}_{i,k}}(\mathbf{x}-\mathbf{x}_{i,k})$$

Kernel parameters for the i^{th} sensor:

$K_{\mathbf{H}_{i,k}}$ = kernel	$\mathbf{H}_{i,k}$ = bandwidth matrix
$w_{i,k}$ = weighting coefficients	N_i = number of kernels stored by sensor i

Benchmark Problem: Multi-agent Path Planning

❖ The agent microscopic dynamics are given by the unicycle model:

Agent i:

$$\begin{cases} \dot{x}_i = v\cos(\theta_i) & \dot{y}_i = v\sin(\theta_i) \\ \dot{\theta}_i = u_{\omega i} & \dot{v}_i = u_{ai} \end{cases}$$

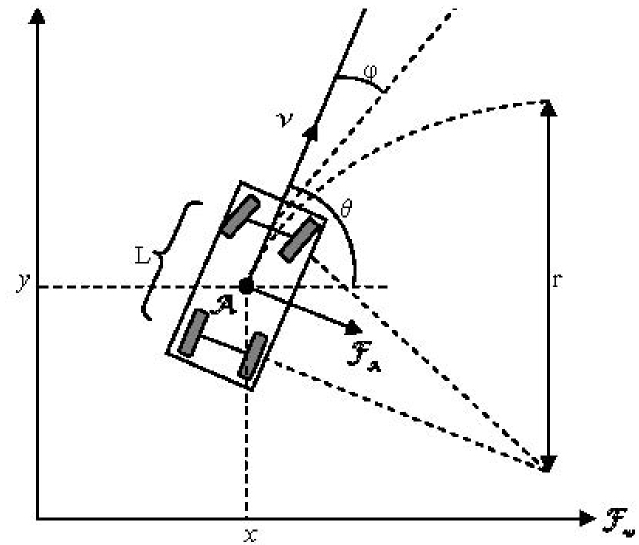
Where:

x: $x-coordinate$ y: $y-coordinate$

θ: $heading\ angle$ v: $linear\ velocity$

u_ω: $angular\ velocity\ control$

u_a: $linear\ acceleration\ control$



❖ The DOC Lagrangian is formulated from macroscopic path-planning objectives:

$$L[\cdot] = w_d \boxed{D_\alpha\left(p(\mathbf{x}_i,t)\,\|\,h(\mathbf{x}_i,t_f)\right)} + \int_A \left[w_p\, p(\mathbf{x}_i,t)U_{rep}(\mathbf{x}_i) + w_e \mathbf{u}_i^T(t)\mathbf{R}\mathbf{u}_i(t)\right] d\mathbf{x}_i$$

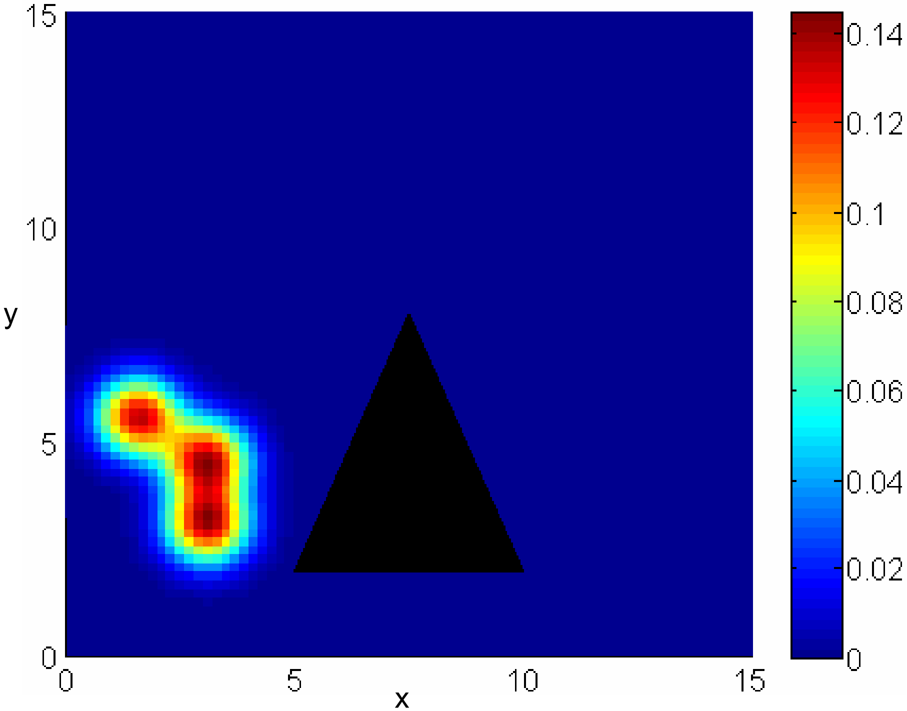
Planning and sensing objectives in the DOC Lagrangian:

Information theoretic functions (α-divergence or KL-divergence) for NPBMs.

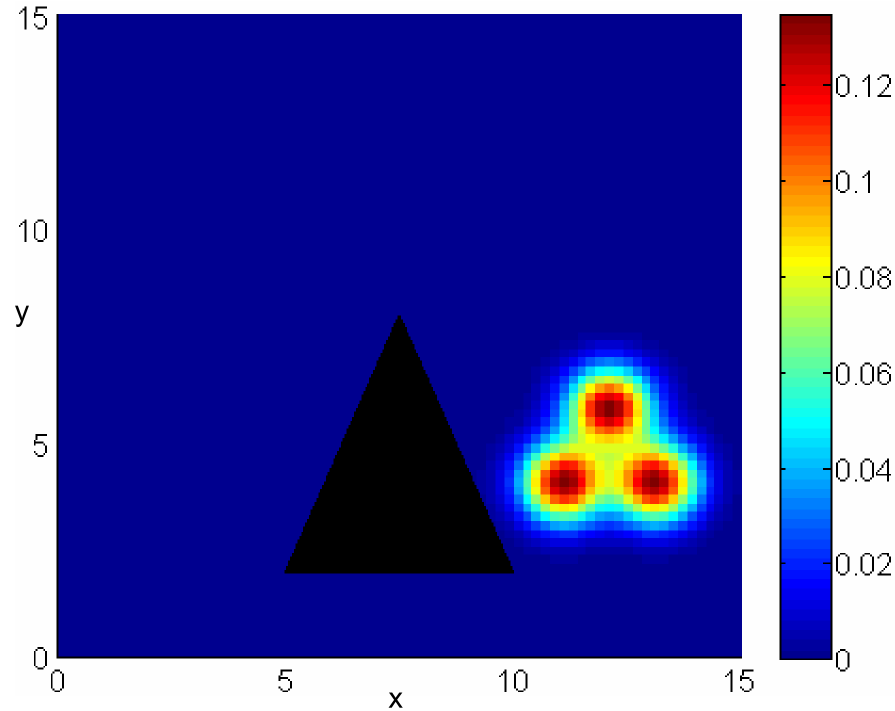
Example: Path Planning/Formation

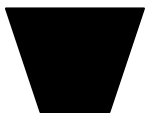
- Agents must maintain a constant distance between the centers of (z = 3)-mixture components (e.g. for communication) while traveling from initial to goal PDF.

Initial PDF, $p(x_i, t_0)$ Pr($\mathbf{x}_i$)

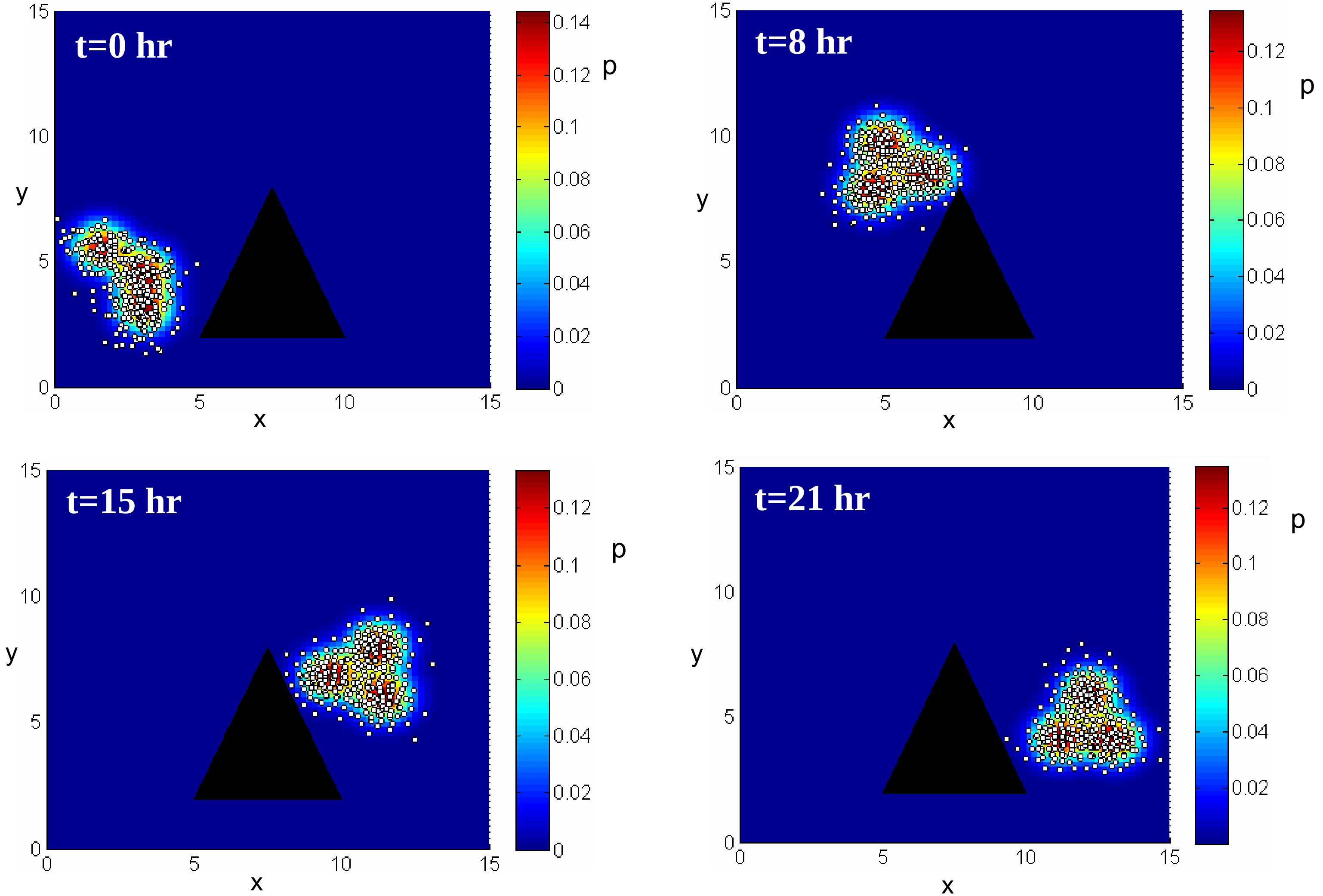


Goal PDF, $h(x_i, t_f)$ Pr($\mathbf{x}_i$)



 : Fixed obstacle

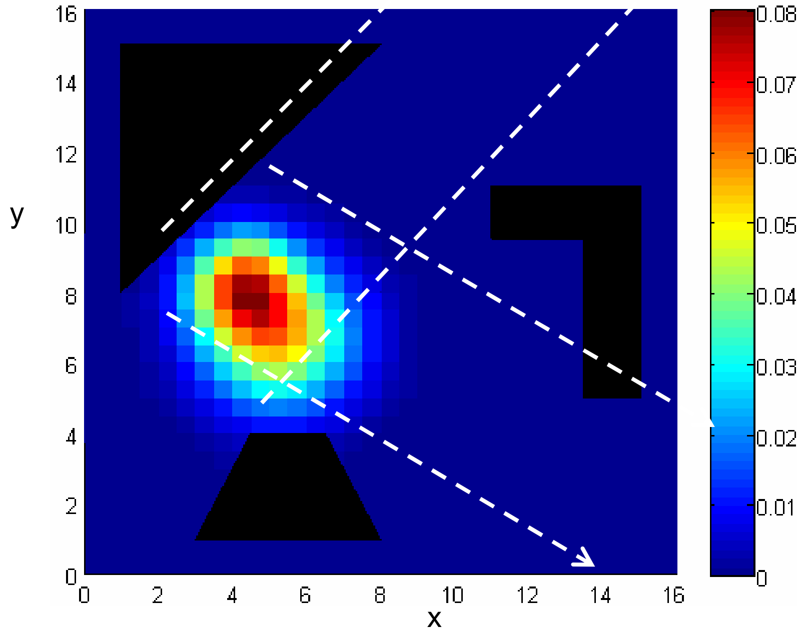
Results: Path Planning/Formation



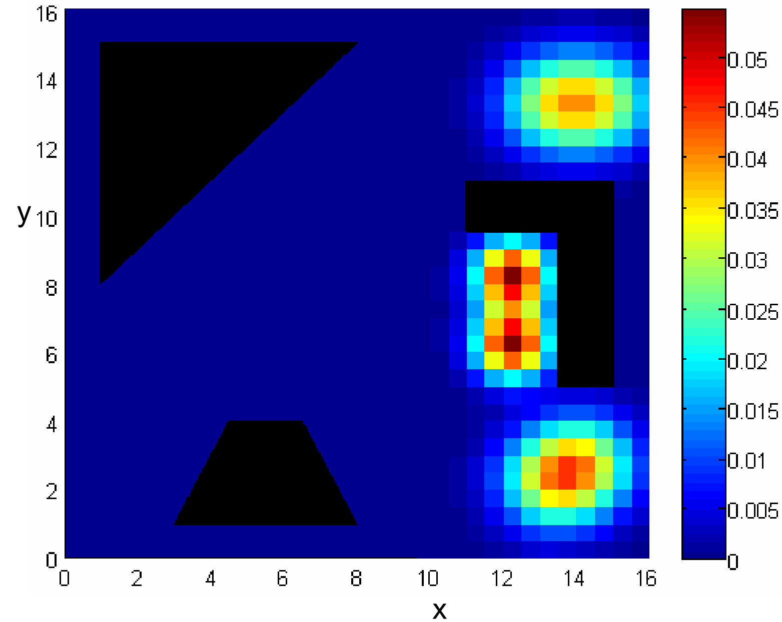
Example: Target Detection

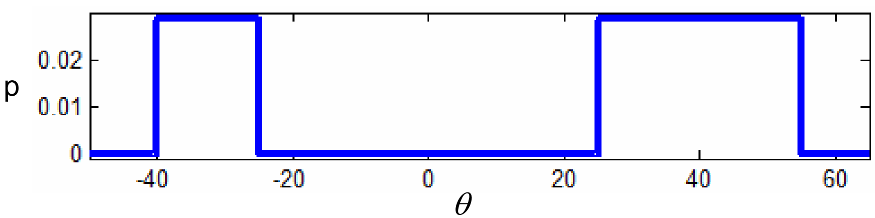
- Agents must obtain at least *k* detections from a set of moving targets (Markov model).

Targets' PDF, $f_T(x_i, t_0)$



Initial sensor PDF, $p(x_i, t_0)$





Target heading PDF, $f_\theta(\theta)$

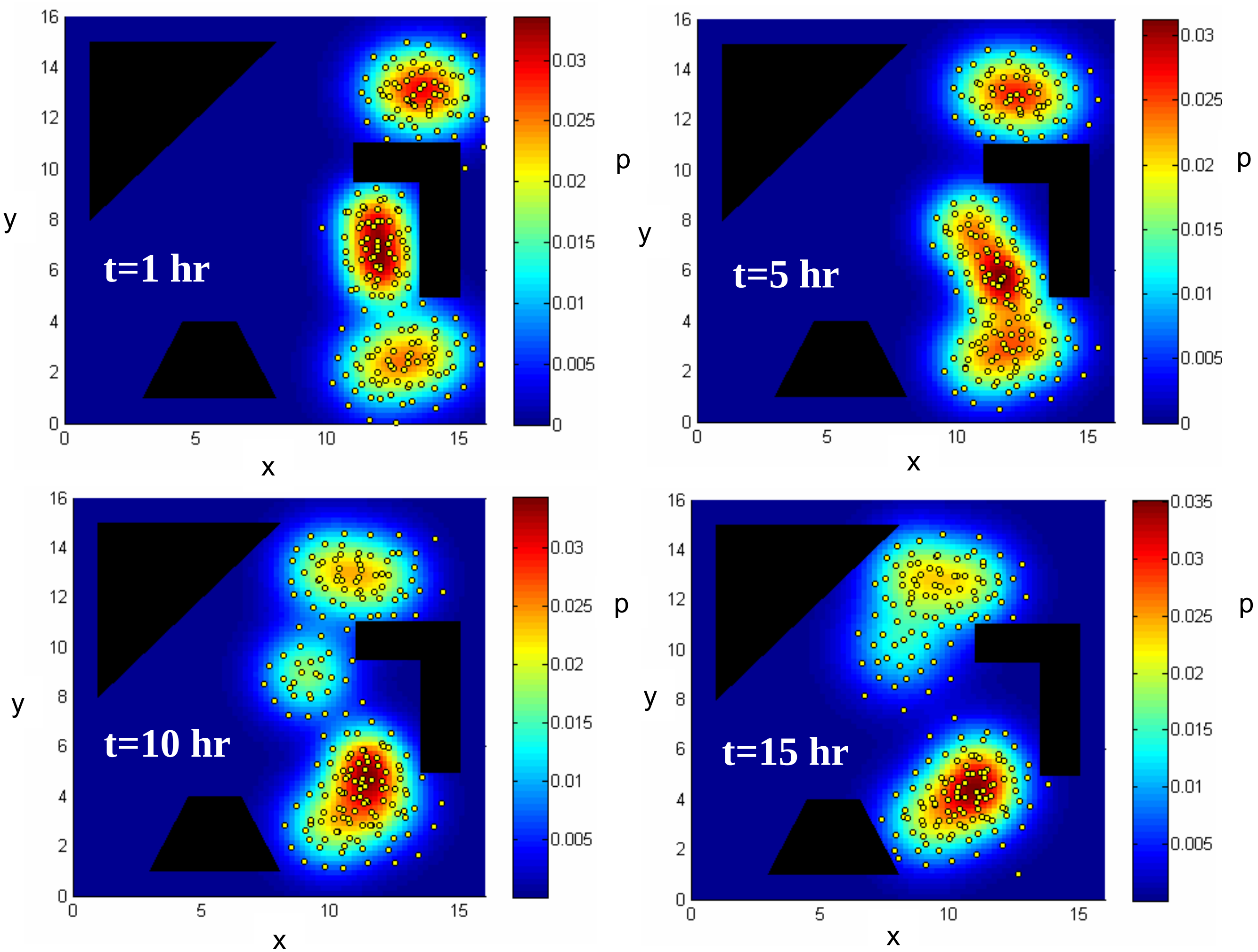
$k = 2$ detections required per target

$N = 200$ sensors

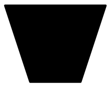
$r = 0.3$ km (sensor detection range)

$z = 6$ mixture components

Results: Target Detection



○ : Agent position



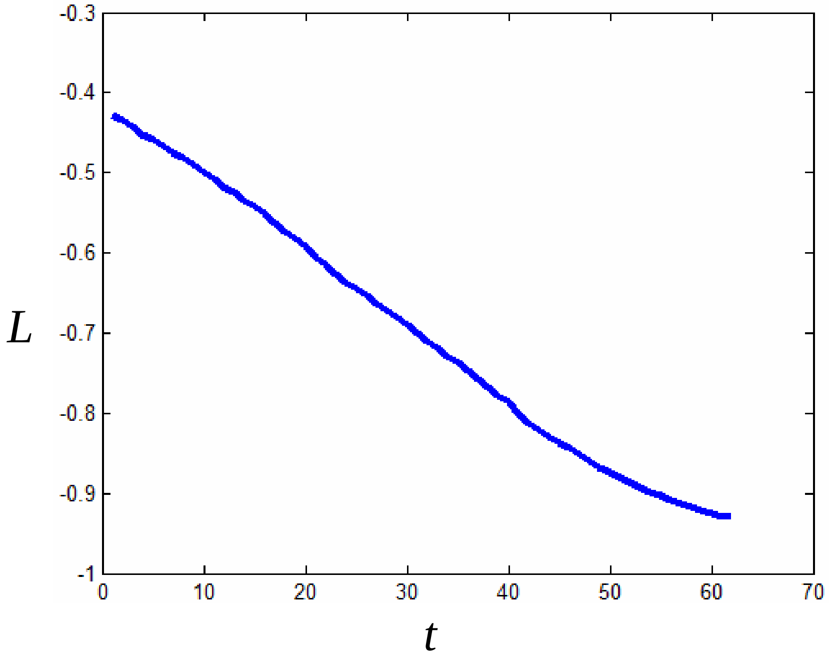
: Fixed obstacle

Performance Results

❖ The solution computed using the DOC approach optimizes the multiscale dynamical system performance throughout the time interval $(t_0, t_f]$.

❖ The DOC planning and control laws outperforms other (scalable) strategies, such as those shown in the table below.

Target-detection Lagrangian for DOC solution



Comparison of Cost Evaluations between Alternate Strategies

Method	Cost, J
DOC	-0.708
Uniform PDF (static)	-0.410
Grid (static)	-0.566
Random (static, avg. over 20 runs)	-0.534

Summary and Future Work

Technical Accomplishments – Year 1:

➢ DOC approach for information-driven mobile sensor agents (MSAs) planning and control over large spatial and temporal scales.

➢ New **information value functions** for Markov motion models, and NPBMs. [Jordan, Carin]

➢ New **approximate dynamic programming** (ADP) approach for **hybrid systems**. [Darrell, How]

➢ New **decentralized KDE-consensus algorithm** for distributed optimal control (DOC). [How, Willsky]

Future Work – Year 2,3:

- Distributed learning: develop ADP relations for DOC to adapt $\mathbf{X}^*$ over time [Darrell]
- Policy iteration for local agent adaptation subject to local constraints [Leonard]
- Value iteration for learning local rewards subject to nonparametric inference [Carin]
- Networks scaling and convergence of multiscale ADP algorithms [Fisher, Wainright]
- Adaptive CBBA/DOP formalism for heterogeneous systems [Roy, How]

LISC

LABORATORY FOR INTELLIGENT SYSTEMS AND CONTROLS

Publications – Year 1

S. Ferrari, R. Fierro, and T. A. Wettergren, *Modeling and Control of Dynamic Sensor Networks*, CRC Press, Boca Raton, FL, ISBN 1439866791, under contract, to appear in 2013.

G. Foderaro, S. Ferrari, and M. Zavlanos, "A Decentralized Kernel Density Estimation Approach for Planning Paths of Distributed Sensor Networks," *NIPS 2012 Workshop on Bayesian Nonparametric Models (BNPM) for Reliable Planning and Decision-Making Under Uncertainty*, submitted.

H. Wei, W. Lu, and S. Ferrari, "An Information Value Function for Nonparametric Gaussian Processes," *NIPS 2012 Workshop on Bayesian Nonparametric Models (BNPM) for Reliable Planning and Decision-Making Under Uncertainty*, submitted.

H. Wei and S. Ferrari, "A Geometric Transversals Approach to Analyzing the Probability of Track Detection for Maneuvering Targets," *IEEE Transactions on Computers*, submitted.

G. Foderaro, S. Ferrari, T. A.Wettergren, "Distributed Optimal Control for Multi-agent Trajectory Optimization," *Automatica*, in revision.

G. Zhang, W. Lu, and S. Ferrari, "An Information Potential Approach to Integrated Sensor Path Planning and Control," *IEEE Transactions on Robotics*, in revision.

W. Lu, G. Zhang, S. Ferrari, M. Anderson, and R. Fierro, "An Information Potential Approach for Tracking and Surveilling Multiple Moving Targets using Mobile Sensor Agents, " *Journal of Defense Modeling and Simulation*, accepted.

G. Zhang, S. Ferrari, and W. Lu, "A Comparison of Information Theoretic Functions for Tracking Maneuvering Targets," *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Ann Arbor, MI, August 2012, in press.

Publications – Year 1

D. Tolic, R. Fierro and S. Ferrari, "Optimal Self-Triggering for Nonlinear Systems via Approximate Dynamic Programming," *Proc. IEEE Multi-Conference on Systems and Control (MSC)*, Dubrovnik, Croatia, October 2012, in press.

W. Lu, S. Ferrari, R. Fierro, and T. Wettergren, "Approximate Dynamic Programming (ADP) Recurrence Relationships for a Hybrid Optimal Control Problem," invited paper, *Proc. SPIE Conference, Unmanned Systems Technology XIII, Session on Intelligent Behaviors*, Baltimore, MD, April 2012, in press.

W. Lu, H. Wei, and S. Ferrari, "A Kalman-Particle Filter for Estimating the Number and State of Multiple Targets," *Proc. International Conference on Management Sciences and Information Technology*, Changsha, China, July 2012, in press.

G. Foderaro, A. Swingler, and S. Ferrari, “A Model-based Cell Decomposition Approach to Online Pursuit-Evasion Path Planning and the Video Game Ms. Pac-Man," *Proc. IEEE Conference on Computational Intelligence and Games*, Granada, Spain, September 2012, in press.

S. Ferrari, M. Anderson, R. Fierro, and W. Lu, "Cooperative Navigation for Heterogeneous Autonomous Vehicles via Approximate Dynamic Programming," invited paper, *Proc. of the IEEE Conference on Decision and Control*, Orlando, FL, December 2011, pp. 121-127.

S. Ferrari, G. Zhang, and C. Cai, “A Comparison of Information Functions and Search Strategies for Sensor Planning," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, Vol. 42, No. 1, 2012.

Acknowledgements:
This research was funded by the ONR MURI Grant N00014-11-1-0688.

Questions?