Approximate State Estimation in Multiagent Settings with Continuous or Large Discrete State Spaces

Prashant Doshi
Department of Computer Science
University of Georgia
Athens, GA 30602
pdoshi@cs.uga.edu

ABSTRACT

We present a new method for carrying out state estimation in multiagent settings that are characterized by continuous or large discrete state spaces. State estimation in multiagent settings involves updating an agent's belief over the physical states and the space of other agents' models. We factor out the models of the other agents and update the agent's belief over these models, as exactly as possible. Simultaneously, we sample particles from the distribution over the large physical state space and project the particles in time.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems

General Terms

Algorithms, Performance

Keywords

Multiagent state estimation, continuous state spaces, particle filters

1. INTRODUCTION

In order to act rationally, an agent must keep track of the state of the environment over time based on its actions and observations. When the agent is acting alone in the environment, it must track the evolution of the *physical* state; this is usually accomplished using the Bayes filter. The filter, in practice, manifests as the Kalman filter when the dynamics are linear Gaussian and the agent's prior belief is Gaussian, or as the particle filter (PF) [3] when no assumptions about the dynamics or prior beliefs are made. In the presence of other agents who themselves act, observe, and update their beliefs the agent must track not only the physical state but also the possible states of others. This is because other agents' actions may affect the evolution of the physical state and the agent's payoffs. A naive way to do this is to consider the other agents as automatons, whose actions follow a fixed and known probability distribution, allowing the use of the original Bayes filter for the state estimation. A more sophisticated approach is to generalize the Bayes filter to multiagent settings as shown in [1], in which an agent tracks the evolution of the *interactive* state over time. In practice, the estimation may be carried out using the interactive PF (I-PF) that is an analogous generalization of the PF to the multiagent setting [1, 2].

Previous applications of the I-PF have been confined to simple problems with a very small number of discrete physical states. This is because a large number of particles must be sampled, at the expense of computational efficiency, to achieve good approximation quality. While this limitation also affects the traditional PF, it is especially acute for the I-PF. This is because the interactive state space from which the particles are sampled tends to get large as it includes the nested beliefs of the other agents.

The above mentioned limitation of the I-PF becomes more potent in the context of a continuous or very large physical state space, as exhibited by many real-world application settings. In this paper, we present a new method for approximately carrying out the state estimation in multiagent settings characterized by continuous or large discrete physical state spaces. Our approach involves factoring out some dimensions of the interactive state space and updating the belief over these dimensions as exactly as possible, while sampling and propagating the remaining ones. In particular, we factor out the models of the other agents and update the agent's belief over these models. In performing this update, all distributions that can be handled exactly are handled exactly, while tight approximations are used for the remaining ones. Simultaneously, we sample particles from the distribution over the large physical state space and project the particles in time. When compared with the performance of the I-PF on continuous settings, we show that our approach achieves better approximation quality while consuming less computational resources as measured by the number of particles and run time.

2. FACTORING STATE ESTIMATION

We decompose the state estimation process into two factors, one of which represents the update of the belief over the physical state space, and the other is the update of the belief over the other agent's models conditioned on a physical state.

We show the decomposition below:

$$b^t_{i,1}(is^t) = \int_{is^{t-1}} \sum_{a_j^{t-1}} Pr(s^t|a_i^{t-1}, a_j^{t-1}, o_i^t, s^{t-1})\ Pr(\theta^t_{j,0}|s^t, a_i^{t-1}, a_j^{t-1}, o_i^t, \theta^{t-1}_{j,0})\ Pr(a_j^{t-1}|\theta^{t-1}_{j,0})\ b^{t-1}_{i,1}(is^{t-1})\ d\ is^{t-1} \quad (1)$$

Here, $Pr(a_j^{t-1}|\theta^{t-1}_{j,0})$ is the probability that a_j^{t-1} is Bayes rational for the agent described by $\theta^{t-1}_{j,0}$.

We may expand the first term of the above equation as,

$$Pr(s^t|a_i^{t-1}, a_j^{t-1}, o_i^t, s^{t-1}) = \alpha\ O_i(s^t, a_i^{t-1}, a_j^{t-1}, o_i^t) \times T_i(s^{t-1}, a_i^{t-1}, a_j^{t-1}, s^t)$$

where α is the normalization constant, T_i and O_i are the transi-

AAMAS'07, May 14–18, 2007, Honolulu, Hawai'i, USA.

978-81-904262-7-5 (RPS)

tion and observation functions, respectively. The other term in the equation, $Pr(\theta^t_{j,0} \mid s^t, a_i^{t-1}, a_j^{t-1}, o_i^t, \theta^{t-1}_{j,0})$, may be rewritten as,

$$Pr(\theta^t_{j,0}|s^t, a_i^{t-1}, a_j^{t-1}, o_i^t, \theta^{t-1}_{j,0}) = \sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t) \times \delta_D(SE_{\widehat{\theta}^t_j}(b^{t-1}_{j,0}, a_j^{t-1}, o_j^t) - b^t_{j,0})$$

$SE_{\widehat{\theta}^t_j}$ stands for the update of the complete belief using the transition and observation functions in the frame $\widehat{\theta}^t_j$.

3. REPRESENTING PRIOR BELIEFS

In order to sample from agent i's beliefs, we first need to represent them. Let $\mathbf{X} = \{X_1, X_2, \ldots, X_k\}$ be the set of k *continuous*-valued variables, where $k \in \mathbb{N}$, and $\mathbf{Y}$ be the set of *discrete*-valued elements of the (hybrid) physical state space, S. Let $\mathbf{x}$ be an instantiation of $\mathbf{X}$ and analogously for $\mathbf{y}$. Together, $\mathbf{X}$ and $\mathbf{Y}$ completely describe the physical state space.

We represent agent j's level 0 belief, $b^{t-1}_{j,0} \in \Delta(S)$, using a factorization of the physical state space. In other words, $b^{t-1}_{j,0}(\mathbf{XY}) = p_{j,0}(\mathbf{Y})p_{j,0}(\mathbf{X}|\mathbf{Y})$. While $p_{j,0}(\mathbf{Y})$ is a discrete probability distribution over $\mathbf{Y}$, $p_{j,0}(\mathbf{X}|\mathbf{Y})$ is a collection of multivariate Gaussian densities, each of which is defined over the variables in $\mathbf{X}$. Each Gaussian within the collection may have a different set of parameters specific to the instantiation of $\mathbf{Y}$.

Agent i's level 1 belief, $b^{t-1}_{i,1} \in \Delta(S \times \Theta^{t-1}_{j,0})$, is a distribution over the level 0 beliefs of agent j for each physical state and frame of the other agent. In order to represent this belief, we factor it as, $b^{t-1}_{i,1}(s, \theta^{t-1}_{j,0}) = b^{t-1}_{i,1}(s^{t-1})\, b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{t-1})$. The term $b^{t-1}_{i,1}(s^{t-1})$, is a distribution over the physical state space analogous to j's level 0 belief, and may be represented similarly. The second factor, $b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{t-1})$ is a distribution over the level 0 beliefs of j conditioned on the physical state (assuming j's frame is known). Because j's belief is represented using a collection of Gaussians, as mentioned previously, that are described by their means and covariance matrices, i's level 1 beliefs are densities over these parameters. We represent i's level 1 belief conditioned on the physical state using a conditional linear Gaussian (CLG) density function.

4. RAO-BLACKWELLISED I-PF (RB-IPF)

We sample N particles from agent i's belief over the physical state space, $b^{t-1}_{i,1}(s^{t-1})$, represented as mentioned in Section 3, resulting in a set of particles $\{s^{(1)}, s^{(2)}, \ldots, s^{(N)}\}$ that together approximate the belief over the large state space. The belief over the complete interactive state space is then given by the following set of particles, $\{(s^{(n)}, b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{(n)})\}^N_{i=1}$:

$$b^{t-1}_{i,1}(s^{t-1}, \theta^{t-1}_{j,0}) \approx \frac{1}{N}\sum_{n=1}^{N} \delta_D(s^{t-1} - s^{(n)}) b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{(n)}) \quad (2)$$

where δ_D is the Dirac delta function.

Substituting Eq. 2 into Eq. 1 and expanding $is^{t-1} = \langle s^{t-1}, \theta^{t-1}_{j,0}\rangle$, we have the following:

$$b^t_{i,1}(is^t) \approx \tfrac{1}{N} \sum_{a_j^{t-1}} \sum_n \int_{s^{t-1}} \delta_D(s^{t-1} - s^{(n)})\, Pr(s^t|a_i^{t-1}, a_j^{t-1}, o_i^t, s^{t-1})\, d\, s^{t-1} \times \int_{\theta^{t-1}_{j,0}} Pr(\theta^t_{j,0}|s^t, a_i^{t-1}, a_j^{t-1}, o_i^t, \theta^{t-1}_{j,0})\, Pr(a_j^{t-1}|\theta^{t-1}_{j,0})\, b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{(n)})\, d\, \theta^{t-1}_{j,0}$$

Thus, agent i's state estimation in the multiagent setting may take the following approximate form:

$$b^t_{i,1}(is^t) \approx \tfrac{\alpha}{N} \sum_{a_j^{t-1}} \sum_{n=1}^{N} \rho^{(n)}_{a_j^{t-1}}(s^t)\, \kappa^{(n)}_{a_j^{t-1}}(\theta^t_{j,0}|s^t), \text{ where}$$

$$\rho^{(n)}_{a_j^{t-1}}(s^t) \stackrel{def}{=} O_i(s^t, a_i^{t-1}, a_j^{t-1}, o_i^t)\, T_i(s^{(n)}, a_i^{t-1}, a_j^{t-1}, s^t) \quad (3)$$

and

$$\kappa^{(n)}_{a_j^{t-1}}(\theta^t_{j,0}|s^t) \stackrel{def}{=} \int_{\theta^{t-1}_{j,0}} \sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t) \delta_D(SE_{\widehat{\theta}^t_j}(b^{t-1}_{j,0}, a_j^{t-1}, o_j^t) - b^t_{j,0})\, Pr(a_j^{t-1}|\theta^{t-1}_{j,0})\, b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{(n)})\, d\theta^{t-1}_{j,0} \quad (4)$$

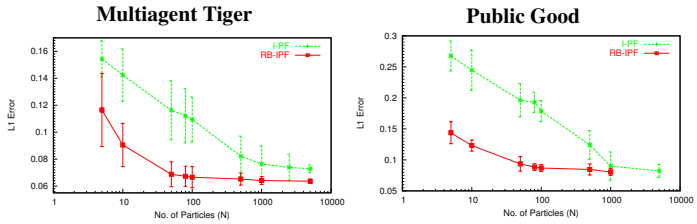
We estimate the physical state, denoted by $\rho^{(n)}_{a_j^{t-1}}$, by propagating the particles as in the PF, while the estimation of the other agent's models given by $\kappa^{(n)}_{a_j^{t-1}}$ is updated as we show next.

5. BELIEF UPDATE OVER MODELS

For settings where the physical state space is continuous, both agents' observation functions may be represented using Gaussian or *softmax* (also known as a *logistic*) density functions. We represent the transition functions as CLGs. For clarity, we subdivide the process of updating i's conditional beliefs over j's models (Eq. 4) into three steps:

Step 1: Given that agent j's prior level 0 beliefs are Gaussian (see Section 3), and we may closely approximate the product of a Gaussian and softmax function by a Gaussian [6], j's level 0 belief update, $SE_{\widehat{\theta}^t_j}(b^{t-1}_{j,0}, a_j^{t-1}, o_j^t)$ is carried out analytically. Specifically, we derive a Gaussian that forms a lower bound to the softmax density using *variational methods* (see [6] for the derivation; [5] for an introduction to variational methods). We note that while the variational Gaussian may not be a tight approximation to the softmax, the product Gaussian is a tight approximation to the product of the softmax and the Gaussian (for example, see [6]). Therefore, analogous to the Kalman filter, j's posterior belief is also a Gaussian whose mean and covariance are functions of the mean and covariance of the prior belief, $\mu^t_{j,0} = f^{\mu}_{a_j,o_j}(\mu^{t-1}_{j,0}, \Sigma^{t-1}_{j,0})$ and $\Sigma^t_{j,0} = f^{\Sigma}_{a_j,o_j}(\Sigma^{t-1}_{j,0})$. We may rewrite the above as, $\mu^{t-1}_{j,0} = g^{\mu}_{a_j,o_j}(\mu^t_{j,0}, \Sigma^t_{j,0})$ and $\Sigma^{t-1}_{j,0} = g^{\Sigma}_{a_j,o_j}(\Sigma^t_{j,0})$, where $g^{\mu}_{a_j,o_j}$ and $g^{\Sigma}_{a_j,o_j}$ may be seen as inverses under the assumption that $f^{\mu}_{a_j,o_j}$ and $f^{\Sigma}_{a_j,o_j}$ are 1:1 maps.

Step 2: We now turn our attention to the integral in Eq. 4. As we mentioned previously, the term $Pr(a_j^{t-1}|\theta^{t-1}_{j,0})$ is the probability that action, a_j^{t-1}, is Bayes rational for the model $\theta^{t-1}_{j,0}$, ie. $a_j^{t-1} \in OPT(\theta^{t-1}_{j,0})$ where OPT is the set of actions that are Bayes rational given the model. Let $\mathcal{R}_{a_j^{t-1}} \subseteq \Theta^{t-1}_{j,0}$ be the contiguous region of j's models for which the action, a_j^{t-1} is Bayes rational [7] ie. define $\mathcal{R}_{a_j^{t-1}}$ such that, $\forall\ \theta^{t-1}_{j,0} \in \mathcal{R}_{a_j^{t-1}}, a_j^{t-1} \in OPT(\theta^{t-1}_{j,0})$. Because, $Pr(a_j^{t-1}|\theta^{t-1}_{j,0})$ may be rewritten as, $\frac{1}{|OPT(\theta^{t-1}_{j,0})|} \times \delta_D(OPT(\theta^{t-1}_{j,0}) - a_j^{t-1})$, the integral becomes, $\int_{\mathcal{R}_{a_j^{t-1}}} \frac{1}{|OPT(\theta^{t-1}_{j,0})|} \sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t)\, \delta_D(SE_{\widehat{\theta}^t_j}(b^{t-1}_{j,0}, a_j^{t-1}, o_j^t) - b^t_{j,0})\, b^{t-1}_{i,1}(\theta^{t-1}_{j,0}|s^{(n)})\, d\theta^{t-1}_{j,0}$. Substituting i's level 1 Gaussian belief into the above we get, $\int_{\mathcal{R}_{a_j^{t-1}}} \frac{1}{|OPT(\theta^{t-1}_{j,0})|} \sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t)\, \delta_D(SE_{\widehat{\theta}^t_j}(b^{t-1}_{j,0}, a_j^{t-1}, o_j^t) - b^t_{j,0})\, \mathcal{N}([\mathbf{w}^{\mathbf{y}^{(n)}} \cdot \mathbf{x}^{(n)}]; \Sigma^{\mathbf{y}^{(n)}})(\mu^{t-1}_{j,0}, \mathbf{\Sigma}^{t-1}_{j,0})\, d\theta^{t-1}_{j,0}$
$= \sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t)\, \mathcal{N}([\mathbf{w}^{\mathbf{y}^{(n)}} \cdot \mathbf{x}^{(n)}]; \Sigma^{\mathbf{y}^{(n)}})$
$(g^{\mu}_{a_j,o_j}(\mu^t_{j,0}, \Sigma^t_{j,0}), g^{\Sigma}_{a_j,o_j}(\Sigma^t_{j,0}))$, if $\langle g^{\mu}_{a_j,o_j}(\mu^t_{j,0}, \Sigma^t_{j,0}), g^{\Sigma}_{a_j,o_j}(\Sigma^t_{j,0})$



(a) (b)

	Run times in secs	
Problem	**RB-IPF**	**I-PF**
Tiger	50.93 ± 0.96	151.05 ± 0.4
PG	32.06 ± 0.3	101.07 ± 0.45

(c)

Figure 1: Performance profiles for the (a) multiagent tiger and (b) PG problems. The estimation accuracy of RB-IPF is significantly better than the I-PF's given the same number of particles. (c) For an identical L1 error, the RB-IPF takes less run time than the I-PF (Linux, Xeon 3.4GHz, 4GB).

$\rangle$ belongs in $\mathcal{R}_{a_j^{t-1}}$, 0 otherwise. Note that $\langle \mu_{j,0}^t, \Sigma_{j,0}^t \rangle$ parameterize $b_{j,0}^t$ and $\frac{1}{|OPT(\cdot)|}$ is absorbed into the density. We focus on the second term of the previous expression next. Intuitively, this density is i's updated belief that results when the transformations, f_{a_j,o_j}^{μ} and f_{a_j,o_j}^{Σ} are applied to the variate, $\langle \mu_{j,0}^{t-1}, \Sigma_{j,0}^{t-1} \rangle$ at which a_j^{t-1} is Bayes rational. If the transformations are not linear, the resulting density over $\langle \mu_{j,0}^t, \Sigma_{j,0}^t \rangle$ may not be Gaussian. In this case, we numerically estimate the Gaussian, $\mathcal{N}\ (\mu_{a_j,o_j}^{(n)}; \Sigma_{a_j,o_j}^{(n)})$ $(\mu_{j,0}^t, \Sigma_{j,0}^t)$, that best fits the previous density.

Step 3: The final step in calculating $\kappa_{a_j^{t-1}}^{(n)}(\theta_{j,0}^t | s^t)$ involves the product, $\sum_{o_j^t} O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t) \mathcal{N}(\mu_{a_j,o_j}^{(n)}, \Sigma_{a_j,o_j}^{(n)})(\mu_{j,0}^t, \Sigma_{j,0}^t)$ where $\mathcal{N}(\mu_{a_j,o_j}^{(n)}, \Sigma_{a_j,o_j}^{(n)})$ is the fitted Gaussian density. Notice that $\kappa_{a_j^{t-1}}^{(n)}$ is a *mixture* of Gaussians where $O_j(s^t, a_j^{t-1}, a_i^{t-1}, o_j^t)$ is the weight assigned to each participating Gaussian in the mixture. This weight is the probability with which j received its observation, o_j^t, on performing action, a_j^{t-1}.

6. EXPERIMENTS

Our first problem domain is the *continuous multi-agent tiger problem*, a modified version of the persistent multi-agent tiger problem discussed in [7]. While in the classical, discrete, version the tiger could be at one of two locations (doors) unknown to the agents, in this version, the tiger is located on a continuous axis, $-1 \leq x \leq 1$. The gold is always located symmetrically (about $x = 0$) from the tiger's location. Hence, knowing the tiger's location allows one to exactly infer the location of the gold. We assume a discrete action space that involves each agent calling out the location of the gold. Here, to keep matters simple, this could be *left* (OL) or *right* (OR). Left, for example, could signify that the gold is located at some point, $x \leq 0$. Each agent may also listen (L), in which case, the agent hears a growl from the left (GL) or right (GR) that informs the agent, noisily, the location of the tiger. The agent also overhears, noisily, the location, if called out by the other agent. Once a location has been called out, the tiger (and the gold) likely persist at their original location in the next time step.

The second problem domain is a *sequential* version of the public good (PG) problem with punishment [4]. Let $x_u \in X_u$ represent the quantity of resources in the public pot. We assume in our formulation that X_u is hidden from the agents. However, each agent on performing an action receives an observation of *plenty* (PY) or *meager* (MR) symbolizing the state of the public pot. The resources in agent i's private pot, $x_{r,i} \in X_{r,i}$, are observable to i.

• The state space is, $X_i = X_u \times X_{r,i}$, which represents the amount of resources in the public pot and private pot of agent i • The set of actions for agent i is, A_i = {C, D}. The agent contributes some fixed amount, $x_c < X_T$, during the contribute action. Let $A = A_i \times A_j$, where $A_j = A_i$ • The observations of agent i are, Ω_i = {PY, MR} • The transition function, $T_i : X_i \times A \times X_i \rightarrow [0, 1]$. Because the amount of contributions are fixed and known, the transitions are deterministic. Note that both agents' actions affect X_u while $X_{r,i}$ is affected only by i's action. • The observation function is, $O_i : X_u \times \Omega_i \rightarrow [0, 1]$ • The reward function is, $R_i : X_i \times A \rightarrow \mathbb{R}$. The reward is determined as follows: $R_i(x_i, a_i, a_j) = x_{r,i} + c_i x_u - 1_D(a_i) 1_C(a_j) P - 1_C(a_i) 1_D(a_j) c_p$, where c_i is the marginal private return. We assume that $c_i < 1$ so that the agent does not benefit enough that it contributes to the public pot for private gain. Simultaneously, $c_i M > 1$, making collective contribution pareto optimal. P is the punishment meted out to the defecting agent and c_p is the non-zero cost of punishing for the contributing agent. For simplicity, let the cost of punishing be same for both the agents. $1_D(\cdot)$ is an indicator function which is 1 if its argument is D, 0 otherwise.

We demonstrate that the RB-IPF is statistically more efficient as compared to the I-PF. In other words, the RB-IPF estimates the hidden interactive state more accurately as compared to the I-PF, while consuming less particles and hence less computational resources. In Figs. $1(a)$ and (b), we show the line plots of the L1 error with respect to the number of particles (N) allocated to the two filters in the two problem domains. Because it is not possible to carry out the belief update exactly, we used the I-PF with half a million particles to compute the 'exact' beliefs. Each data point in both the plots is the average of 10 runs of the respective filter. For the multi-agent tiger problem, the estimation was performed for the case where agent i listens (L) and hears a growl from the left but does not overhear anything ($\langle GL, S \rangle$). In the PG problem, the beliefs resulting from agent i contributing (C) and perceiving plenty of resources (PY) in the public pot were used for comparison. We observe that for both the problems, the posterior belief generated by the RB-IPF is much closer to the truth than the one generated by the I-PF for the same number of particles. Notice that, on average, exponentially many more particles are required for the I-PF to reach an identical L1 error as the RB-IPF. This is somewhat exemplified by the difference in the run times of the two methods (see Fig. $1(c)$) for an identical estimation accuracy.

7. REFERENCES

[1] P. Doshi and P. J. Gmytrasiewicz. Approximating state estimation in multiagent settings using particle filters. In *AAMAS*, 2005.

[2] P. Doshi and P. J. Gmytrasiewicz. A particle filtering based approach to approximating interactive pomdps. In *AAAI*, 2005.

[3] A. Doucet, N. D. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.

[4] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[5] M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

[6] K. Murphy. A variational approximation for bayesian networks with discrete and continuous variables. In *UAI*, 1999.

[7] B. Rathnas, P. Doshi, and P. Gmytrasiewicz. Exact solutions to interactive pomdps using behavioral equivalence. In *AAMAS*, 2006.