

Unifying Learning in Games and Graphical Models

I.Rezek, S.J.Roberts

Department of Engineering Science
University of Oxford
Oxford, OX1 3PJ, U.K.
{irezek,sjrob}@robots.ox.ac.uk

A. Rogers, R.K. Dash, N.Jennings

Electronics & Computer Science
University of Southampton
Southampton, SO17 1BJ, U.K.
{a.rogers,rkd02r,njr}@ecs.soton.ac.uk

Abstract—The ever increasing use of intelligent multi-agent systems poses increasing demands upon them. One of these is the ability to reason consistently under uncertainty. This, in turn, is the dominant characteristic of probabilistic learning in graphical models which, however, lack a natural decentralised formulation. The ideal would, therefore, be a unifying framework which is able to combine the strengths of both multi-agent and probabilistic inference

In this paper we present a unified interpretation of the inference mechanisms in games and graphical models. In particular, we view fictitious play as a method of optimising the Kullback-Leibler distance between current mixed strategies and optimal mixed strategies at Nash equilibrium. In reverse, probabilistic inference in the variational mean-field framework can be viewed as fictitious game play to learn the best strategies which explain a probabilistic graphical model.

I. INTRODUCTION

Multi-agent systems have become an attractive approach to solving complex and distributed activities. They are particularly useful when the task becomes too expensive or impossible to accomplish by a single agent, be it due to physical, geographical or temporal constraints. Application examples in which such problems occur range from control of vehicles in wide geographical terrains, global communication networks, economics and parallel optimisation.

In a multi-agent system, each agent has its own expert knowledge and observation domain. Its action responses are based on its local view (formed by fusing observations from its local environment), observed past actions and its belief of future actions of some of its neighbouring agents. Since there is typically no central controller, the aim is therefore to achieve emergent global behaviour whereby a common goal is solved. However, a common hurdle in real world applications is that information isn't exact and thereby requiring that an agent fuses uncertain information consistently. We may consider two ways of tackling this task: 1) either adapt probabilistic methods to multi-agent domains [1] or 2) incorporate probabilistic reasoning into current multi-agent systems [2], [3]. The former approach, has unfortunately lead to systems that are rigid with regards to the agent's environment. In such systems the topology of all agent interactions is predefined and is often singly connected. In the latter approach, the agent's environment is changing and this will have an affect on

the probabilistic solution. Communication loops, for instance, result in over-confident probability estimates and thus incorrect decision making. In this approach in general, it is far from clear what, if at all, solutions are reached.

Typically, there a few questions one might want to ask about the solutions:

- 1) How good is a solution? By good, we mean robustness to changes in the variables due to noise or other environmental factors.
- 2) Will the system always converge or can it exhibit limit cycle behaviour, and under what conditions can this happen?
- 3) How many equilibria are there and what is the size of the solution clusters?

These are all questions about the global behaviour of locally acting agents that is hard to understand and subject to intensive research. Furthermore, an additional ad hoc embedding of probabilistic inference renders these questions even more intractable.

The most fruitful approach to this problem, in our opinion, is that of finding a unifying cost-function view of the multi-agent learning under uncertainty. The more we know about the cost function, the more can be said about the global behaviour of the system. Here, we assume that there exists a clear functional relationship between the global cost function and the local (agent-specific) cost functions, e.g. team game cost functions in [4].

In this paper we interpret learning in games as that of finding (as a result of iterative play) the strategy closest to the optimal Nash strategy. Closeness is, subjectively, measured by a Kullback-Leibler distance. This measure of distance is typically used for probabilistic inference – the method used to learn optimal models of real-world observations. In particular, a certain simplification of the Kullback-Leibler distance, known as the variational mean-field framework, is shown to be equivalent to independent players competing to find the best explanation of the world.

In section II of this paper we review fictitious play and the cost function view of it. Section IV demonstrates the use of an identical cost function in probabilistic learning and highlights the equivalences between learning games and probabilistic

models. Section V points to extensions that can be made based on the similarities observed in the previous section.

II. (LEARNING IN) GAMES: FICTITIOUS PLAY

We consider non-cooperative strategic-form games with I players, indexed by $i \in \{1, \dots, I\}$ ¹. We split the set of players into two subsets, $I = \{i, \bar{i}\}$, consisting of player i and all other players $\bar{i} = \{1, 2, \dots, i-1, i+1, \dots, I\}$. Each player has a finite set of pure strategies (or actions) S_i - typically assumed to be discrete². The space of all possible combinations of actions is given by the strategy profile $S = \prod_{i=1}^I S_i = S_1 \times S_2 \cdots \times S_I$. Each player's mixed strategy is a distribution over her set of pure strategies $Q_i(S_i) \in \Delta_i(S_i) \forall i = 1, \dots, I$, where $\Delta_i(S_i)$ is the set of all probability distributions over S_i . Analogous to the pure strategy profile, the mixed strategy profile is given by $Q(S) = \prod_{i=1}^I Q_i(S_i) = Q_1(S_1) \times Q_2(S_2) \cdots \times Q_I(S_I)$, and is an element of $\Delta(S) = \prod_{i=1}^I \Delta_i(S_i) \times \Delta_2(S_2) \cdots \times \Delta_I(S_I)$. We will use $S_{\bar{i}}$ and $Q_{\bar{i}}(S_{\bar{i}})$ to mean all other players' pure strategy profile and the mixed strategy profile, respectively, i.e. $Q_{\bar{i}}(S_{\bar{i}}) = \prod_{j \neq i} Q_j(S_j) = Q_1(S_1) \times Q_2(S_2) \cdots \times Q_{i-1}(S_{i-1}) \times Q_{i+1}(S_{i+1}) \cdots \times Q_I(S_I)$ and $S_{\bar{i}} = \prod_{j \neq i} S_j = S_1 \times S_2 \cdots \times S_{i-1} \times S_{i+1} \cdots \times S_I$. The payoff to each player, i , is a function mapping all pure strategy combinations of all players on the real line, i.e. $\ell_i(S) : S \rightarrow \mathbb{R}$ for a given pure strategy profile $S = S_i \times S_{\bar{i}}$.

We define the utility function then as the expected reward obtained by player i , given the mixed strategies $Q_i(S_i)$ of all of i 's opponents

$$R_i(Q(S_i), Q_{\bar{i}}(S_{\bar{i}})) \triangleq E(\ell_i(S)) = \int \cdots \int Q(S) \ell_i(S) dS. \quad (1)$$

When solutions to each player's utility maximisation goal form a fixed point, it is known as an equilibrium. It specifies the joint mixed strategy composed of independent mixed strategies every player adopts for a particular game. In particular, the Nash equilibrium is given as the optimal mixed strategy $Q_i^*(S_i)$ of player i

$$R_i(Q_i^*(S_i), Q_{\bar{i}}^*(S_{\bar{i}})) \geq R_i(Q_i(S_i), Q_{\bar{i}}^*(S_{\bar{i}})) \quad \forall i = 1, \dots, I \quad (2)$$

i.e. each player has no incentive to deviate from the equilibrium strategy assuming none of the other players do.

A model for learning the optimal mixed strategy attained at Nash equilibrium is fictitious play. During play, every player monitors the action of the opponents and continually updates the beliefs about the opponents' strategies. The action each player plays is the best response to the opponents' current mixed strategies. Thus, starting with some prior beliefs about

¹We abuse notation using I to denote both the set of players and the cardinality of the set.

²For reasons of notational clarity any integration over the strategy space will be denoted by an integral, irrespective of whether the space is discrete or continuous.

strategies, at discrete time-step t the strategies are updated according to

$$Q_i^{t+1}(S_i) \in \left(1 - \frac{1}{t+1}\right) Q_i^t(S_i) + \frac{1}{t+1} \beta(Q_{\bar{i}}^t(S_{\bar{i}})) \quad (3)$$

where $\beta(Q_{\bar{i}}^t(S_{\bar{i}}))$, the best response to the other players mixed strategies, is defined as the mapping

$$\beta(Q_{\bar{i}}^t(S_{\bar{i}})) : Q_{\bar{i}}^t(S_{\bar{i}}) \rightarrow Q_i^t(S_i) \quad (4)$$

where

$$\beta(Q_{\bar{i}}^t(S_{\bar{i}})) = \underset{Q_i(S_i) \in \Delta_i(S_i)}{\operatorname{argmax}} R_i(Q_i(S_i), Q_{\bar{i}}^t(S_{\bar{i}})). \quad (5)$$

At Nash equilibrium, the mapping reaches steady state

$$Q_i^*(S_i) = \beta_i(Q_{\bar{i}}^*(S_{\bar{i}})) \quad \forall i = 1, \dots, I \quad (6)$$

and every player adopts the mixed strategy that maximises the expected utility.

One objection to fictitious play has been that players almost never play mixed strategies. Instead they myopically choose a pure strategy which maximises the immediate payoff. As a result, they may constantly switch between the pure strategies with ever increasing cycle durations [5]. To overcome such problems, fictitious play has been generalised to smooth fictitious play, in which players begin play sub-optimally and increasingly play myopic best responses as time, t , passes [5]. In such cases, the utility function is augmented to

$$R_i(Q_i(S_i), Q_{\bar{i}}(S_{\bar{i}})) = \int \cdots \int Q(S) \ell_i(S) dS + \tau H(Q_i(S_i)) \quad (7)$$

so that the *smooth* best response becomes

$$\beta_i(Q_{\bar{i}}(S_{\bar{i}})) = \underset{Q_i(S_i) \in \Delta_i(S_i)}{\operatorname{argmax}} \{R_i(Q_i(S_i), Q_{\bar{i}}(S_{\bar{i}})) + \tau H(Q_i(S_i))\} \quad (8)$$

while the iterative update of the mixed strategies still follows (3). The temperature parameter, τ , controls the degree of sub-optimality played by the players and can be regarded as determining the amount of perturbation of the expected reward $R_i(Q_i(S_i), Q_{\bar{i}}(S_{\bar{i}}))$. The perturbation to player i 's payoffs, $H(Q_i(S_i))$ is required to be a smooth strictly differentiable concave function the slope of which approaches infinity as $Q_i(S_i)$ reaches the boundary of $\Delta_i(S_i)$. One of the functions which satisfies the conditions for the second term in (7) is the entropy function

$$H(Q_i(S_i)) = - \int Q_i(S_i) \log Q_i(S_i) dS_i. \quad (9)$$

Under these assumptions, the best response mixed strategy that maximises (8) can be derived by differentiating (8) with respect to $Q_i(S_i)$ and gives rise to

$$Q_i(S_i) \propto \exp \frac{1}{\tau} \int \cdots \int Q_{\bar{i}}(S_{\bar{i}}) \ell_i(S) dS_{\bar{i}} \quad (10)$$

which, in the game theory literature is often abbreviated to

$$Q_i(S_i) \propto \exp \frac{1}{\tau} R_i(S_i, Q_{\bar{i}}(S_{\bar{i}})) \quad (11)$$

III. INFORMATION THEORETIC VIEW OF FICTITIOUS PLAY

Smooth fictitious play with exponential best response functions, as in (10), can be viewed from an information theoretic perspective. From that, so far over-looked correspondence, the aim of fictitious play is to minimise the Kullback-Leibler divergence, which is defined as

$$D(f(x)||g(x)) = \int f(x) \log \frac{f(x)}{g(x)} dx \quad (12)$$

between any distributions $f(x)$ and $g(x)$ of x .

To see this, consider some distribution, $P_i(V_i, S|\theta_i)$, of payoffs V_i for player i and the set of played strategies of all players, S . The vector θ_i parameterises the mixed strategy distribution of all players, which is unknown but inferred through repeated play. Being a distribution, it can be formulated also in terms of prior payoffs of player i , $P_i(V_i|\theta_i)$, and the posterior distribution over strategies given i 's payoffs, $P_i(S|V_i, \theta_i)$,

$$P_i(V_i|\theta_i) = \frac{P_i(V_i, S|\theta_i)}{P_i(S|V_i, \theta_i)} \quad (13)$$

From an information theoretic perspective, the aim of the game is now to achieve the highest reward probability. The relationship to standard game rewards will be described at the end of this section.

The term on the left of equation (13) is simply (the probability of) the expected reward after all mixed strategy distributions have been integrated out - similar to the integral action in equation (1).

Taking the logarithm on either side and taking the expectation with respect to any arbitrary mixed strategy distribution $Q(S)$, we obtain

$$\begin{aligned} \int \dots \int Q(S) \log(P_i(V_i|\theta_i)) dS = \\ \int \dots \int Q(S) \log(P_i(V_i, S|\theta_i)) dS - \\ \int \dots \int Q(S) \log(P_i(S|V_i, \theta_i)) dS \end{aligned} \quad (14)$$

which, since the term on the right side of the equation is independent of S , can be simplified to

$$\begin{aligned} \log(P_i(V_i|\theta_i)) = \\ \int \dots \int Q(S) \log(P_i(V_i, S|\theta_i)) dS - \\ \int \dots \int Q(S) \log(P_i(S|V_i, \theta_i)) dS \end{aligned} \quad (15)$$

The meaning of the above equation becomes more apparent if we reformulate equation (15) to

$$\begin{aligned} \log(P_i(V_i|\theta_i)) = \\ \int \dots \int Q(S) \log \left(\frac{P_i(V_i, S|\theta_i)}{Q(S)} \right) dS + \\ \int \dots \int Q(S) \log \left(\frac{Q(S)}{P_i(S|V_i, \theta_i)} \right) dS \end{aligned} \quad (16)$$

The last term in (16) is the KL-divergence (equation (12)) between the equilibrium distribution and the current distribution over strategies. It is well-known that the KL-divergence is strictly non-negative and zero if and only if $P_i(S|V_i, \theta_i) = Q(S)$. This fact implies that the left-hand term in (16) is a strict bound on the equilibrium expected reward.

$$\int Q(S) \log \left(\frac{P_i(V_i, S|\theta_i)}{Q(S)} \right) dS \leq \log(P_i(V_i|\theta_i)) \quad (17)$$

As is common when using Energy minimisation concepts for optimisation purposes, we can use Boltzmann distributions to map energy values to probabilities. Using the same idea, we can map rewards to probabilities. In particular, the probabilities are constructed such that the smallest total reward, ℓ_i for player i , leads to the highest probability

$$P(V_i|\theta_i) \propto \exp(\ell_i). \quad (18)$$

Using (18) in (17) leads to the lower bound estimate

$$\int Q(S) \log \left(\frac{P(V_i, S|\theta_i)}{Q(S)} \right) dS \leq \ell_i \quad (19)$$

Equation (17) is identical to (7), bar the fact that we have ignored the temperature parameter τ for ease of presentation reasons. Fictitious play can thus be seen as a two-step iterative procedure. In the

- 1) update the current mixed strategies probabilities given the action observed in the last round,
- 2) the algorithm computes expected rewards based on current belief about the mixed strategies θ_i and, subsequently, the best response.

From the analogy of fictitious play and KL divergence, which is used in graphical model optimisation, we can draw a simple (and for this case rather uninteresting) graphical model of the game, shown in figure 1. The actions are random variables with unknown distributions and referred to as latent random variables (shown as clear nodes in 1). The rewards, akin to measurement observations, are given and considered instantiated random variables (shaded nodes).

IV. VARIATIONAL LEARNING

The Kullback-Leibler minimisation criterion is well known and used in machine learning, for example in the estimation of probabilistic models, known as graphical models. Thus, from the machine learning point of view, fictitious play resembles the well known Expectation Maximisation (EM) algorithm [6], which has a cost function that is, in principle, identical to (17).

Consider again I players, index by $\forall i = 1, \dots, I$, and split into two sets, $I = \{i, \bar{i}\}$. The pure strategy space of each player is denoted by S_i , that of the opponent players by $S_{\bar{i}}$, and the strategy profile by S . The mixed strategy profile is given by $Q(S)$, where make the specific assumption that all players, and subsequently their mixed strategies, are independent

$$Q(S) = \prod_{i=1}^I Q_i(S_i). \quad (20)$$

Fig. 1. Graphical model representation of fictitious play for two players with mixed strategies θ_1/θ_2 , rewards ℓ_1/ℓ_2 and actions S_1/S_2 . Random variables are drawn as circles, observed variables as shaded circles.

This is known as the “mean field” assumption in the variational learning framework.

The payoffs are now generalised from discrete matrix payoffs to payoff functions, $\ell(S, \theta)$ with parameters θ . Typically in machine learning, the payoffs take form of a model postulated to underly the experimental data generating processes (a.k.a. generative model). Furthermore, omnipresent observation noise is captured by (more often than not) additive random perturbations which follow certain probability distributions.

Frequently, the modelling using probability is described as a two-player zero-sum game against nature [7], [8]. In batch learning algorithms, nature is implicitly assumed to play following a stationary (mixed) strategy. The observations the researcher makes are nature’s signals which the research uses to infer nature’s true state (type). This, effectively global payoff function, splits into smaller components so that modelling can be viewed as a cooperative game against nature. The logarithm of the distributions is used for mathematical convenience (the exponential family of distributions plays a very dominant role in machine learning). Thus, the payoffs for modelling observations \mathcal{D} are written as

$$\ell(S, \mathcal{D}|\theta) \triangleq \log P(S, \mathcal{D}|\theta) \quad (21)$$

The analogy between fictitious play and the variational methods can then be established through the cost function. In particular, the variational learning approach finds the maximum entropy equilibrium distribution, $Q(S)$ subject to maximising the expected reward (log-probability of the model)

$$\begin{aligned} D(Q(S)||P(S|\theta)) &= \int \cdots \int Q(S) \ell(S, \mathcal{D}|\theta) dS + \tau H(S) \\ &= \int \cdots \int Q(S) \log \left(\frac{Q(S)^\tau}{p(S, \mathcal{D}|\theta)} \right) dS \end{aligned} \quad (22)$$

The second equality highlights the fact that the variational cost functions (22) is a Kullback-Leibler divergence, as would be

the fictitious play cost function (7) with exponentiated payoffs R .

The optimal distribution $Q_i(S_i)$ of player i is one which minimises the Kullback-Leibler divergence (22), assuming all other players \bar{i} adopt the mixture strategy $Q_{\bar{i}}(S_{\bar{i}})$. It is obtained by partial differentiation of (22). In fact, the definition of Nash equilibrium (2) is a partial differential in disguise. Thus, the variational best response is a mixed strategy distribution which takes the general form [9]

$$Q(S_i) \propto \exp \left\{ \frac{1}{\tau} \int \cdots \int Q_{\bar{i}}(S_{\bar{i}}) \ell(S, \mathcal{D}|\theta) dS_{\bar{i}} \right\} \quad (23)$$

This is identical to smooth fictitious play best response (25), which as shown to converge to equilibrium for fictitious play [10] and for the Expectation Maximisation (EM) algorithm [11], which is a particular variational learning method.

Variational “play” then proceeds in two steps, commonly referred to as the E- and M-step:

- The E-step computes player i ’s best response, $Q(S_i)$, based on the opponents’ mixed strategies $Q_{\bar{i}}(S_{\bar{i}})$.
- The M-step updates the model parameters (mixed strategies) θ .

There are some important differences machine learning and learning in games is the reward.

a) Reward: The major difference is the reward. Rewards are very frequently concave (log) functions. This represents the fact that machine learning tasks deal with modelling real observations. Our lack of complete information about nature, however, makes minimum risk strategy the optimal choice and hence a concave utility [8]. Thus, from a game theoretic view, machine learning is a game against *nature*. The players are the probability distributions imposed on parameters of the experimental data model. They may be independent players, characterised by independent distributions. Thus, they compete against each other to find the best sparse mixed strategy against nature.

By comparison with economic applications of game theory [12], the currency of machine learning games is the “information”, in the sense of log-probability. Just like normal currency, it has no additional attributes to it. Conversion between different monetary currencies is via exchange rates, whilst the information currency is converted through use of conditional probabilities.

b) Temperature: Another difference is the fact, that the greater part of machine learning algorithms keep the temperature parameter constant, bar for example simulated annealing [13], annealed Markov Chain Monte Carlo [14] and deterministic annealing [15].

c) Perturbation function: The parameters of the model are not fixed. They are unknown and, in the Bayesian framework, have associated prior and posterior distributions. These distributions can then be considered as additional players in the game. The value of the game in the Bayesian framework is usually known as the evidence, the marginal probability of the observations. In the EM learning the value becomes the maximum likelihood of the data.

Rewards, in the EM method, can be exactly maximised. That is in the sense that a point estimate for the parameters of the payoff function θ was obtained. However, uncertainty about the rewards may also exist. To deal with this uncertainty in the Bayesian framework, priors are imposed on the payoff function parameters. This leads to an extension of the variational learning scheme, in which the maximum entropy is not taken with respect to a flat measure over the space of possibilities, but with respect to a particular hypothesis - that of the prior. The adjusted cost function (24) takes the form of

$$\begin{aligned} D(Q(S, \theta) || P(S, \theta)) &= \\ &= \iint Q(S)Q(\theta) \log \left(\frac{Q(S)^\tau Q(\theta)^\tau}{p(S, \mathcal{D} | \theta) p(\theta)} \right) dS d\theta \\ &= \iint Q(S)Q(\theta) \log \left(\frac{Q(S)^\tau Q(\theta)^\tau}{p(S, \theta, \mathcal{D})} \right) dS d\theta \end{aligned} \quad (24)$$

This is the Kullback-Leibler divergence between the exact posterior distribution, $p(S, \theta, \mathcal{D})$, and an approximation to it, $Q(S, \theta)$. This KL divergence is typically used when integration to obtain exact marginal distributions, such as $P(S)$ or $P(\theta)$, is intractable [16]. The KL divergence between the approximate and true posterior is minimised leading to coupled update equations for the distributions. In the mean field assumptions they take the general form

$$\begin{aligned} Q(S_i) &\propto \exp \left\{ \frac{1}{\tau} \iint Q_i(S_i) Q(\theta) \ell(S, \theta, \mathcal{D}) dS_i \theta \right\} \\ Q(\theta) &\propto \exp \left\{ \frac{1}{\tau} \int Q(S) \ell(S, \theta, \mathcal{D}) dS \right\} \end{aligned} \quad (25)$$

These are then iterated until convergence. In essence, all variables are now random variables with associated distributions. Thus, the set of players is now augmented to include players whose strategy space are the payoff function parameters.

V. IMPLICATIONS AND EXTENSIONS TO GAME PLAYING

A. Global vs Local Rewards

So far, no specific assumption has been made about the form of the payoffs. In their simplest form, the payoffs may be private to each player

$$\ell(S) = \{\ell_i(S_i, S_{\bar{i}})\} \quad \forall i = 1, \dots, I \quad (26)$$

One the other hand, we may presume that some that some player's payoffs affect only few other players. For example, consider a 3 player extension of the 2-player matching pennies game. In the extension the third player's rewards depend on how the other 2 players play against each other, i.e. they do not consider player 3 in their play. The game designer thus imposes a structural form on the payoff relationships, for example as

$$\ell(S) = \{\ell(S_1, S_2), \ell(S_2, S_1), \ell(S_3, S_1, S_2)\}. \quad (27)$$

Viewed from a probabilistic derivation "payoffs" (22), the structural relationships between their individual components may stem from *a priori* knowledge of causal relationships between parameters of the experimental data model. Thus, they

specifically state knowledge relating local to global reward functions. In general, for computing best response, each player must consider players affecting her, players she influences and other players affecting the players she influences (in graph theory terms the Markov blanket, that is parents, children and children's parents [17]).

B. Independent Mixed Strategies

In addition to assumptions affecting the payoffs, there is the assumption about the independence of players (20). Both, the machine learning as well as game theoretic learning exposition above assumed independent players. This assumption naturally arose from the problem statement, such as *distributed* control. However, the machine learning literature has developed methods for finding marginal equilibrium distributions when there exist some probabilistic dependence between the players' mixed strategies. For certain dependence topologies, e.g. trees, exact methods can be derived [18]. Structure of mixed strategy distributions encode a more direct dependence between the players, one mediated through probabilistic inference rather than payoffs. It is conceivable that such dependence structures find their equivalences in coalition formation in games.

C. Quality of Equilibrium

Quantifying the tightness of the bound (24), and hence the quality of the equilibrium, is subject to current research. Generally speaking it will depend on degree of discrepancy between the true and approximate distribution. One approach uses sampling methods have been used to improve on the equilibrium distribution [19]. Alternatively, the empirical integration in (25) can be performed and has been applied to large scale optimisation problems [20].

ACKNOWLEDGEMENTS

The authors would like to thank S. Reece and D. Leslie for invaluable comments and advice. This research was undertaken as part of the ARGUS II DARP (Defence and Aerospace Research Partnership). This is a collaborative project involving BAE SYSTEMS, QinetiQ, Rolls-Royce, Oxford University and Southampton University, and is funded by the industrial partners together with the EPSRC, MoD and DTI.

REFERENCES

- [1] Y. Xiang, *Probabilistic Reasoning in Multi-agent Systems: a Graphical Models Approach*. Cambridge University Press, 2002.
- [2] J. Dix, M. Nanni, and V. S. Subrahmanian, "Probabilistic agent programs," *ACM Transactions on Computational Logic (TOCL)*, vol. 1, no. 2, pp. 208–246, 2000.
- [3] C. Kreucher, K. Kastella, and A. O. Hero, "Multitarget sensor management using alpha divergence measures," in *First IEEE Conference on Information Processing in Sensor Networks*, Palo Alto, 2003.
- [4] D. Wolpert, "Information Theory - The Bridge Connecting Bounded Rational Game Theory and Statistical Physics," 2004, arXiv.org:cond-mat/0402508.
- [5] D. Fudenberg and D. Levine, *The Theory of Learning in Games*. MIT Press, 1999.
- [6] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society - Series B*, vol. 39, no. 1, pp. 1–38, 1977.

- [7] P. Grünwald and A. Dawid, "Game Theory, Maximum Entropy, Minimum Discrepancy and Robust Bayesian Decision Theory," *Annals of Statistics*, vol. 32, pp. 1367–1433, 2004.
- [8] F. Topsø, "Maximum Entropy versus Minimum Risk and Applications to some Classical Discrete Distributions," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2368–2376, 2002.
- [9] M. Haft, R. Hofmann, and V. Tresp, "Model-Independent Mean Field Theory as a Local Method for Approximate Propagation of Information," *Computation in Neural Systems*, vol. 10, pp. 93–105, 1999.
- [10] J. Shemma and G. Arslan, "Unified Convergence Proofs of Continuous-Time Fictitious Play," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1137–1142, 2004.
- [11] R. Neal and G. Hinton, "A View Of the EM Algorithm That Justifies Incremental, Sparse, and other Variants," in *Learning in Graphical Models*, M. I. Jordan, Ed. Dordrecht: Kluwer Academic Publishers, 1998, pp. 355–368.
- [12] R. Dash, A. Rogers, and N. Jennings, "A mechanism for multiple goods and interdependent valuations," in *Proc. 6th Int. Workshop on Agent-Mediated E-Commerce*, New York, USA, 2004, pp. 748–755.
- [13] E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines*. John Wiley & Sons, 1989.
- [14] R. M. Neal, "Annealed importance sampling," University of Toronto, Tech. Rep., February 1998. [Online]. Available: <ftp://ftp.cs.toronto.edu/pub/radford/ais.ps.Z>
- [15] C. Peterson and J. Anderson, "A Mean Field Theory Learning Algorithm for Neural Networks," *Complex Systems*, vol. 1, pp. 995–1019, 1987.
- [16] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, "An Introduction to Variational Methods for Graphical Models," in *Learning in Graphical Models*, M. Jordan, Ed. Kluwer Academic Press, 1997.
- [17] Z. Ghahramani and M. Beal, *Advanced Mean Field Method—Theory and Practice*. MIT Press, 2000, ch. Graphical models and variational methods.
- [18] J. Yedidia and W. Freeman, "Understanding Belief Propagation and its Generalizations," MERL - Mitsubishi Electric Research Laboratory, TR-2001-22, 2001.
- [19] N. de Freitas, P. A. d. F. R. Højten-Sørensen, and S. J. Russell, "Variational mcmc," in *UAI*, 2001, pp. 120–127.
- [20] T. L. III, M. Epelman, and R. Smith, "A Fictitious Play Approach to Large-Scale Optimization," *Operations Research*, vol. forthcoming, 2005.