Learning of cooperative behaviour in robot populations

Michalis Smyrnakis[1] and Dario Bauso[2] and Paul A. Trodden[3] and Sandor M. Veres[4]

***Abstract*— This paper addresses convergence and equilibrium properties of game theoretic learning algorithms in robot populations using simple and broadly applicable reward/cost models of cooperation between robotic agents. Regret based methods of robot learning to cooperate are joined with relevant results of midfield game theory to achieve results on asymptotic second moment boundedness in the variation of cooperative behaviour. A second result proves worst case stability of behaviours holds asymptotically and the results are illustrated in simulation.**

I. INTRODUCTION

Learning of cooperative behaviour in multi-agent systems and robots is an actively researched area where there are few unifying themes today. This paper is looking convergence of population behaviour and learning in a relatively simple game theoretic framework where robotic agents are awarded for cooperating with others and also pay costs for cooperation. Related research is wide and diverse. [1] investigated probabilistic convergence of joint actions in discrete and continuous strategy potential games under the log-linear learning algorithm. [2] present a cooperative learning algorithm to accommodate for the individualistic Q-Learning as well as the collaborative advice sharing, resulting in an approach where a multi-robot team can improve cooperation and also enhance individual performance concurrently.

[3] discusses how multiple robots can emerge cooperative behaviours through co-evolutionary processes in an actor-critic architecture as an integration of local predictive models and vector-valued reward functions. [4] propose a framework for addressing sensor and actuator failure tolerances by incorporating a learning-based supervisor synthesis approach and control reconfiguration mechanism in multi-agent systems. [5] presents a method for two wheeled mobile robots to navigate in unknown environments while learning to carry an object together. In their navigation leader robot and a follower robot cooperatively perform obstacle boundary following or target seeking (TS) to reach a target location. In [6] a dynamic neural network model was used in an experiment of robots learning to make them cooperate with others under potential unpredictability of the others' behaviours. [7] addresses the problem of robot learning cooperative behaviour in an interactive task with a human - they use probabilistic learning of dynamical system models to encode the robot's motion along the task.

[8] addresses the complexity of decentralised learning of behaviours by proposing time-sharing tracking framework in which a joint-state best-response Q-learning method serves as the primary algorithm to adapt to the cooperating robots' policies where agents learn the optimal cooperative policy eventually - as illustrated in simulation examples. In [9] a top down and iterative approach is proposed following a divide-and-conquer where a multi-agent system is modelled as a concurrent discrete-event system defined by a collection of finite automata that interact with each other.

This paper considers a simple cost-benefits model of cooperation between robotic agents where results from midfield game theory can be utilised. We prove that there are regret based [10], [11] strategies of learning to cooperate with other agents which leads to convergence to an equilibrium of cooperative behaviour.

The next section outlines some scenarios to motivate our problem through examples where robot cooperation is evidently useful and also defines the basic cost-reward model. Section 3 introduces the agent cooperation problem in its most general form where probabilities of creating and dropping cooperation and evolution of agents' cooperation levels are discussed. Section 4 contains the main results in Theorem 1 on second moment boundedness and Theorem 2 on worst-case stability cooperative behaviours. Sections 5 and 6 presents simulations and conclude the paper, respectively.

II. SOME PROBLEMS IN COOPERATIVE ROBOTICS

This section introduces some novel models of robot collaboration for which the cooperation paradigm of the theory presented in this paper can be applied.

A. Assistive observation relay

A large group of robots is collecting a set of small moving objects (animals or useful materials) in an environment. The robots have limited sensing capability of vision with 55^o angle of view with 3 cameras which are best used to observe a scene at a distance of about $d > 5$ meters. Area coverage of the cameras is very poor to enable a robot to its own vicinity at once, while another robot can observe a robots whole environment from a distance of about d using a mono camera (the robot can only see a small section of its own immediate environment around it). Also the robot

*This work was not supported by any organization

[1]Michalis Smyrnakis is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mapping Street, S1 3JD, UK m.smyrnakis@sheffield.ac.uk

[2]Dario Bauso is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mapping Street, S1 3JD, UK and with Dipartimento di Ingegneria Chimica, Informatica, Gestionale, e Meccanica, Università di Palermo, Italy d.bauso@sheffield.ac.uk

[3]Paul A. Trodden is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mapping Street, S1 3JD, UK p.trodden@sheffield.ac.uk

[4]Sandor M. Veres is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mapping Street, S1 3JD, UK s.veres@sheffield.ac.uk

has difficulty to observe its own body if it is covered by the moving objects (for instance pollution). Robots can help each other in such a situation to deal with the moving objects by collecting them (killing them or wiping them clean if they are pollution phenomena).

This is not unlike when friends observe each other's clothing and adjust a collar or a misaligned tie. This is cooperation among humans to help each other to look good. Lack of cooperation can leave them sometime not looking the best, hence friendship pays off in many ways. When our robots deal with a persistent set of reoccurring pollutants to be collected the benefits of cooperation in sharing observations is immediate.

We can assume that robot's arms are independently operated from its observation cameras. A robot i observing the body and environment of another robot j costs cx_i in terms of observation made and imaging relayed to j at modelling quality level x_i. Vice versa, robot j can observe robot i and relay the information to it at modelling quality level x_i. Assuming that the benefits for the control system of each robot i (for catching the moving objects) are proportional by a factor b to the quality of model x_j relayed by another robot, it follows that robot i benefits by bx_j.

B. Clearing an area by robots

Consider that a set of robots in a large industrial site need to clear an area from various objects which can weigh multiple of the carrying capability of a single robot. Whenever they cooperated in carrying an object in pairs, three working together or four of them, their cost of contribution will proportionally depend on their weight carrying capability x_j which is proportional to their energy consumption by a factor of c . On the other hand, each time a heavy object is carried (which they were not able to carry themselves) they get a contribution measured by bx_i by each collaborating partner. In this model the benefit for each robot in cooperation is then approximately proportional to the total weight of the object carried, with a proportionality factor b.

C. Cooperation in model predictive control

In some environmental control problems, robots may need to coordinate precise movements to handle objects. An obvious choice for real time control of their movements is distributed model predictive control (DMPC). In DMPC, the overall system is controlled by a number of independent but interacting controllers or control agents, which share information (predictions) in order to achieve coordinated control [12]. For example, consider the simple scenario where a linear time-invariant system

$$\begin{bmatrix} z_1^+ \\ z_2^+ \end{bmatrix} = A \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + B \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

is to be controlled by two distributed control agents, such that constraints $z \in \mathbb{Z}$, $u \in \mathbb{U}$ are met. The non-cooperative optimal control problem for agent $i \in \{1, 2\}$ is

$$V_i^0(z_i; \mathbf{u}_j) = \min_{\mathbf{u}_i} \{ V_i(z_i, \mathbf{u}_i) : \mathbf{u}_i \in \mathcal{U}_i(z_i; \mathbf{u}_j) \}$$

where V_i is a finite-horizon cost function, $\mathbf{u}_i$ is the predicted sequence of controls, and $\mathcal{U}_i$ is a feasible set defined by the dynamical model and the constraints. The problem is coupled because the subsystems $i = 1, 2$ may be dynamically coupled (via off-diagonal terms in A, B) or share constraints. Hence, the feasible set $\mathcal{U}_i$ – and the optimal cost V_i^0 – depends on the current plan of the other agent, $\mathbf{u}_j$.

A non-cooperative approach here is known to lead to optimized controls that are not Pareto efficient [13] with respect to a system-wide cost

$$V_1(z_1, \mathbf{u}_1) + V_2(z_2, \mathbf{u}_2),$$

Therefore, in cooperative approaches [14]–[16], each agent attempts to optimize the system-wide cost rather than just its own cost, by minimizing

$$(1 - x_i)V_i(z_i, \mathbf{u}_i) + x_i V_j(z_j, \mathbf{u}_j).$$

The scalar x_i denotes the level of cooperation: with $x_i = 0$ the agent is non-cooperative, while $x_i = 1$ results in entirely altruistic behaviour. Situations exist for both dynamically coupled systems [13] and dynamically decoupled systems with coupled constraints [16] such that

$$\begin{aligned} V_1(\mathbf{u}_1^*) &> V_1(\mathbf{u}_1^0) \\ \text{and } V_2(\mathbf{u}_2^*) &< V_2(\mathbf{u}_2^0) \\ \text{but } V_1(\mathbf{u}_1^*) + V_2(\mathbf{u}_1^*) &< V_1(\mathbf{u}_1^0) + V_2(\mathbf{u}_1^0), \end{aligned}$$

where $(\mathbf{u}_1^0, \mathbf{u}_2^0)$ are the solutions obtained from the non-cooperative MPC optimizations, and $(\mathbf{u}_1^*, \mathbf{u}_2^*)$ are those from cooperative MPC. Thus, the cost to agent 1 here is

$$V_1(\mathbf{u}_1^*) - V_1(\mathbf{u}_1^0)$$

while the benefit to agent 2 is

$$V_2(\mathbf{u}_2^0) - V_2(\mathbf{u}_2^*).$$

Therefore, in adopting a cooperative objective, a control agent may sacrifice his own minimal cost for the benefit of the overall system.

D. Cooperation of autonomous cars

Consider the case of autonomous cars move along a road. The behaviour of the cars, the safe distance they keep from the car ahead and behind, can vary. This can be seen as the case where x_i represents the behaviour of each car. A cooperative car with $x_i = 1$, will show "polite" behaviour with respect to the vehicle in front as it will keep a safe distance. On the other hand a non-cooperative car with $x_i = 0$ will leave a minimum distance to the car in front. In addition, each individual car would like to adapt its behaviour according to the behaviour of other cars as drivers do.

In [17] and [18] a model predictive control approach was used to update vehicle velocities by optimising a cost function, using the optimal velocity model. The cost function to update the velocity of vehicle i has been defined as

follows, given that vehicles j and k are ahead and behind, respectively.

$$L(y,v,u) = \begin{array}{l} w_v(v_i(t)-V_d)^2 + w_u u_i^2 \\ +w_f g^2(y_j(t),y_i(t),v_j(t)) + w_s(t)S_{err}^2(t) \end{array} \quad (1)$$

where $y_i(t)$ and $y_j(t)$ denote the position of vehicles i and j, $v_i(t)$ and $v_j(t)$ denote their velocities, u_i is the control cost of vehicle i's velocity update. The cost term $g(y_j(t),y_i(t),v_j(t))$ is defined as: $g(y_j(t),y_i(t),v_j(t)) = \kappa(V_1 + V_2\tanh[C_1(y_i - y_j - l_m) - C_2] - v_j)$, with κ, V_1, V_2, C_1 and C_2 being constants and l_m is the minimum allowable distance between vehicles i and j. w_v, w_u and w_f are constant weights, $w_s(t)$ is a time varying weight and $S_{err}(t) = S_0 + t_h v_i(t) - (y_k - y_i)$, is the penalty cost that takes into account the distance with vehicle k, t_h and S_0 are constants. The first term of the cost function represents the cost of not moving with the desired velocity V_d, the second term is the cost of updating the velocity, the third and forth terms represent the costs of changing the distance with vehicle j and k, respectively.

Based on the cooperation level of x_i of vehicle i, the weights of (1) can change in order to take into account the behaviour of i. Thus (1) can be written as:

$$L(y,v,u) = \begin{array}{l} (1-x_i)(v_i(t)-V_d)^2 + w_u u_i^2 \\ +x_i g^2(y_j(t),y_i(t),v_j(t)) + x_i S_{err}^2(t) \end{array} \quad (2)$$

A cooperative vehicle i will have small cost for not following its desired velocity, V_d, when the cost for having smaller distance than the desired one with the other vehicles, will be the most important factors in the optimisation process of (2). On the other hand the first term will be the most important term in the optimisation process of a non-cooperative vehicle.

Another approach that is widely used to describe traffic is based on cellular automata. A popular model that describes traffic in motorways using cellular automata is the Nagel-Schrekenberg model [19]. In [19] the area of interest was divided in cells of 7.5 meters width, and each vehicle's velocity, v^i, which denotes the number of cells that a vehicle will move, is an integer between zero and v_{max}. Then the following rules are applying in order to update a vehicle's velocity at time t, v_i^t:

- $v_i^t = min(v_i^{t-1}+1, v_{max})$ if there are $v_i^{t-1}+1$ or v_{max} unoccupied cells in front of vehicle i
- $v_i^t = v_p$ if there are only v_p, $v_p \leq v_i^{t-1}+1$ unoccupied cells in front of vehicle i
- $v_i^t = max(v_i^{t-1}-1, 0)$ with probability p

where p is a small probability under which some unexpected slowdown of the vehicles can occur.

The Nagel-Schrekenberg model can be extended in order to take into account the coordination levels of the vehicles as follows:

1) Each vehicle i at time t computes the distance it wants to keep from the next vehicle j as $d_{i,j} = \lfloor x_i d_{min} \rfloor$
2) Update vehicle i's velocity as follows:
   - $v_i^t = min(v_i^{t-1}+1, v_{max})$, if vehicle j, is either $v_{max}+d_{i,j}$ or $v_i^{t-1}+1+d_{i,j}$ cells away, otherwise
   - $v_i^t = v_d$ where $v_d = n_{tot}^t - d_{i,j}$, and n_{tot}^t are the total number of empty cells in front of vehicle i, otherwise
   - $v_i^t = max(v_i^{t-1}-1, 0)$ with probability p.

where d_min is the minimum safe distance, minimum number of cells, a cooperative vehicle wants to have from the other vehicles

E. Cost and reward model

There is a common model for all the above examples. For simplicity and to be able to treat the problem analytically, the $success$ (or $reward$) of a node i at time t is measured by

$$-deg[v_i]^t c x_i^t + b \sum_{j \in \mathcal{N}_i^t} x_j^t \quad (3)$$

where $\mathcal{N}_i^t = \{j \in \mathcal{I} : w_{ij} > 0\}$, is the set of robotic agents cooperating with agent i or neighbours of agent i hereafter, $deg[v_i]^t = |\mathcal{N}_i^t|$ is the cardinality of $\mathcal{N}_i^t$, and b, c are constants with $b > c$ the benefit (reward) and cost coefficients.

III. PROBLEM STATEMENT

Cooperative robotics can find inspiration from the results in the evolution theory of cooperative networks. In [20] the evolution of networks in terms of cooperative behaviour and connections among the nodes has been studied. In particular the evolution of networks has been addressed, which start from a single node and expand up to N_{tot} nodes under assumptions that the nodes behaviour and tendency to create new links is based on the Simultaneous Emergence and Evolution (SEE) model.

Robotic swarms dealing with the environment can benefit from cooperative behaviour: they can create new links of cooperation with other robots to improve what they can achieve. They can benefit from cooperation and overall the price they pay for cooperation can be outweighed by the benefits they receive. The advantage of SEE against other alternatives, such as the one in [21], is that SEE takes into account also the behavioural properties of other robotic nodes.

In this article a new model is proposed, equally applicable to abstract nodes, not only robots but any kind of agents, physical or non-physical on networks or in physical environments, which extend SEE in order to take into account past interactions and the probability to drop an existing connection or create a new link. In contrast to SEE, the focus is placed on the evolution of connections and behaviour of networks with constant upper limit of the number of nodes N_{tot}.

Similarly to prior work on SEE [20], we consider networks which can be represented as graphs, $G(\mathcal{V},\mathcal{W})$, where $\mathcal{V} = \{v_1,\ldots,v_{N_{tot}}\}$, is a non-empty set, which contains the nodes of the network, and $\mathcal{W}$ is the adjacency matrix with

entries w_{ij}. We consider undirected graphs, therefore $\mathcal{W}$ is a symmetric matrix, $w_{ij} = w_{ji}$ with w_{ij} defined as:

$$w_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

In the case that the network evolves over time then $\mathcal{W}$ depends on time and will be denoted by $\mathcal{W}^t$ with entries w_{ij}^t.

The creation or the drop of a cooperative link and the choices made for the best coordination level, can be cast as a game, where nodes are the players and their action is the cooperation level and creation or the drop of a link. In particular, Continuous Action Iterative Prisoners' Dilemma (CAIPD) [22] can be used in order to define this process. In CAIPD a cooperative node has to pay a cost c and the benefit of his neighbour node will be b. Thus a node that chooses to defect, pays no cost but can be benefited by connecting to its cooperative neighbours. Thus in CAIPD each node i at iteration t has chosen a cooperation level x_i^t (indicating the degree of involvement in all cooperation of node i at time t, which proportionately measures the cost and benefits of cooperation with cost coefficients c and benefit coefficient b), which results in a vector of cooperation levels $\mathbf{x^t} = \left(x_1^t, \ldots, x_{N_{tot}^t}\right)^T$. In particular $\forall i \in \mathcal{I} = \{1, 2, \ldots, N_{tot}\}$, $0 \leq x_i^t < 1$, with $x_i^t = 0$ denoting a purely defecting node and $x_i^t = 1$ denoting a purely cooperative node. The *success* of a node i at time t is measured by its fitness $f_i^t(\mathbf{x})$ as:

$$f_i^t(\mathbf{x}^t) = -deg[v_i]^t c x_i^t + b \sum_{j \in \mathcal{N}_i^t} x_j^t \tag{4}$$

where $\mathcal{N}_i^t = \{j \in \mathcal{I} : w_{ij} > 0\}$, is the set of neighbour nodes of node i, $deg[v_i]^t = |\mathcal{N}_i^t|$ is the cardinality of $\mathcal{N}_i^t$, and b, c are constants with $b > c$.

A. Probability to create or drop a link

In this subsection we define the probability that a node creates or drops a link according to our proposed model, which are introduced in terms of *regret learning*. Each node will maintain a history of regret that they had either because they didn't create or they didn't drop a link with other nodes. Hence we assume that the nodes are able to assess the cost and benefits of hypothetical cooperation with nodes they had not been not yet connected with. In particular, the regret of a node i not to connect with node j when t iterations of CAIPD had been played, $R_i^{t+1}(a_{i,c}(j))$, is defined as:

$$\begin{aligned} R_i^{t+1}(a_{i,c}(j)) = \quad & \tilde{w}_{ij}^t(\tfrac{1}{t}(f_{i,c}^{t+1}(\mathbf{x}^t) - f_i^t(\mathbf{x}^t)) \\ & +(1-\tfrac{1}{t})R_i^t(a_{i,c}(j))) \end{aligned} \tag{5}$$

where $\tilde{w}_{ij}^t = \begin{cases} 1 & \text{if nodes i and j are not connected} \\ 0 & \text{otherwise} \end{cases}$, $f_{i,c}^t(\mathbf{x}_c^{t-1})$ is the fitness of player/node i when the new node is added in $\mathcal{N}_i^{t-1}$, and $\tilde{\alpha}_1$.

The probability then node i will create a link with a node j can be computed using (5) as follows:

$$p_{ijc}^{t+1} = \frac{e^{\beta R_i^{t+1}(a_{i,c}(j))}}{\sum_{\tilde{j} \in \mathcal{N}_{ic}^t} e^{\beta R_i^{t+1}(a_{i,c}(\tilde{j}))}} \tag{6}$$

where $\mathcal{N}_{ic}^t = \{j \in \mathcal{I} : w_{ij}^t = 0\}$, is the set of nodes that is not connected with node i and β is a learning rate.

Similarly the regret of a node i keeping a connection with node j when t iterations of CAIPD have been played, $R_i^{t+1}(a_{i,d}(j))$, is defined as:

$$\begin{aligned} R_i^{t+1}(a_{i,d}(j)) = \quad & w_{ij}^t(\tfrac{1}{t}(f_{i/j,d}^{t+1}(\mathbf{x}_d^t) - f_{i/j}^{t+1}(\mathbf{x}^t)) \\ & +(1-\tfrac{1}{t})R_i^t(a_{i,d}(j))) \end{aligned} \tag{7}$$

where $f_{i,d}^{t+1}(\mathbf{x}_d^t)$ is the fitness of player i when j is removed from $\mathcal{N}_i^{t-1}$.

The probability of node i dropping the link with a node j will be defined as:

$$p_{ijd} = \frac{e^{\beta R_i^t(a_{i,d}(j))}}{\sum_{\bar{j} \in \mathcal{N}_i^t} e^{\beta R_i^t(a_{i,d}(\bar{j}))}} \tag{8}$$

where β is a learning rate.

B. Evolution of nodes' cooperation levels

In [20] the updates of the nodes' coordination level were based on the tendency each node has, because of (4), to align with the the coordination levels of its neighbour nodes. More formally the change in cooperation level in SEE model for an agent i is computed as:

$$\dot{x}_i^t = \frac{1}{deg[v_i]^t} \left(\sum_{j=1}^{N_{tot}} p_{ij}^t (x_j^t - x_i^t) \right) \tag{9}$$

where p_{ij}^t is defined as $p_{ij}^t = \frac{1}{1+e^{f_j^t(\mathbf{x}^t) - f_i^t(\mathbf{x}^t)}}$.

A direct extension of the SEE model (9), is to take also into account the evolution in the hypothetical cooperation levels which are due to the nodes which node i want to be connected or want to drop its existing links. Thus the updates the change in cooperation level for an agent i is computed as:

$$\begin{aligned} \dot{x}_i^t = \quad & \tfrac{1}{deg[v_i]^t} \left(\textstyle\sum_{j=1}^{N_{tot}} p_{ij}^t (x_j^t - x_i^t) \right) \\ & +\kappa_1 \left(\textstyle\sum_{j=1}^{N_{tot}} p_{ijc}^t (x_j^t - x_i^t) \right) \\ & -\kappa_2 \left(\textstyle\sum_{j=1}^{N_{tot}} p_{ijd}^t (x_j^t - x_i^t) \right) \end{aligned} \tag{10}$$

where κ_1 and κ_2 are constants. The first term of (10) is the same as in (9), the second term represents the hypothetical cooperation level which is due to the nodes which node i want to be connected with. Finally the third term, is represents the tendency of agent i to deviate from the coordination level of the agents which he wants to drop links with.

The proposed model for the evolution of cooperation level, in matrix form can be written as:

$$\dot{\mathbf{x}}^t = D^t L^t \mathbf{x}^t + (\kappa_1 \tilde{L}^t - \kappa_2 \bar{L}^t)\mathbf{x}^t \tag{11}$$

where $\dot{\mathbf{x}}^t$ and $\mathbf{x}^t$ are $N_{tot} \times 1$ vectors, D^t is a diagonal $N_{tot} \times N_{tot}$ matrix, with $D^t[i,i] = \frac{1}{deg[v_i]^t}$ and L^t $\tilde{L}^t$ and $\bar{L}^t$ are $N_{tot} \times N_{tot}$ stochastic matrices which are defined as:

- $L^t[i,i] = \sum_{j \in \mathcal{N}^t} p_{ij}^t$ and $L^t[i,j] = -p_{ij}^t$
- $\tilde{L}^t[i,i] = \sum_{j \in \mathcal{N}^t} p_{ijc}^t$ and $L^t[i,j] = -p_{ijc}^t$
- $\bar{L}^t[i,i] = \sum_{j \in \mathcal{N}^t} p_{ijd}^t$ and $L^t[i,j] = -p_{ijd}^t$

Note here that if only the first term of the model will be considered, $\dot{\mathbf{x}}^t = D^t L^t \mathbf{x}^t$, this is the evolution of coordination levels in SEE model.

SEE and (11) are deterministic models. In order to take into account that new connections and link drops are not occurred deterministically, the following stochastic version of (11) can be used:

$$\dot{\mathbf{x}}^t = D^t L^t \mathbf{x}^t + \kappa_3 \epsilon. \tag{12}$$

where $\kappa_3 = t^{3/2}(\kappa_1 \tilde{L}^t - \kappa_2 \bar{L}^t)\mathbf{x}$ and each element of ϵ is defined as $\epsilon[i] \sim N(0,1)$. Note here that the disturbances that are introduced are increasing over the time. In order to bound the effects of disturbances we will set $kappa_1$ and κ_2 to be equal to $\frac{1}{t}$.

IV. MAIN RESULTS

A. Stochastic approach

Based on the above result, let us now substitute the expression of the mean-field equilibrium strategy $u^* = -R^{-1}B^T[PX + \Psi]$ as in (**??**) in the open-loop microscopic dynamics $dX(t) = (AX(t) + Bu(t) + C)dt + \Sigma d\mathcal{B}(t)$ given in (**??**) so to obtain the closed-loop microscopic dynamics

$$dx_i^t = [D^t L^t]_{i.} x^t dt + \sigma d\mathcal{B}^t \tag{13}$$

Now, let $\mathcal{X}$ be the set of equilibrium points for (13), namely, the set of X such that

$$\mathcal{X} = \{x \in ??| [D^t L^t]_{i.} x = 0\}$$

and let $V(x^t) = dist(x^t, \mathcal{X})$.[1] The next result establishes a condition under which the above dynamics converges asymptotically to the set of equilibrium points.

Theorem 1: **(2nd moment boundedness)** Let a compact set $\mathcal{M} \subset \mathbb{R}^2$ be given. Suppose that for all $x \notin \mathcal{M}$

$$\partial_x V(x)^T \Big([D^t L^t]_{i.} x \Big) < -\tfrac{1}{2}\sigma^2(x)\partial_{xx} V(x) \tag{14}$$

then dynamics (13) is a stochastic process with 2nd moment bounded.

Proof:

Let x^t be a solution of dynamics (13) with initial value $x^0 \notin \mathcal{X}$. Set $t = \{\inf t > 0 | \, x^t \in \mathcal{X}\} \leq \infty$ and let $V(x^t) = dist(x^t, \mathcal{X})$. For all $t \in [0, t]$ [2]

$$V(x^{t+dt}) - V(x^t) = \|x^{t+dt} - \Pi_{\mathcal{X}}(x^t)\| - \|x^t - \Pi_{\mathcal{X}}(x^t)\|$$

$$= \|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\| - \|x^t - \Pi_{\mathcal{X}}(x^t)\|$$

$$= \tfrac{1}{\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|^2 -$$

$$\tfrac{1}{\|x^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t - \Pi_{\mathcal{X}}(x^t)\|^2.$$

[1] define distance

[2] Define projection

From the definition of infinitesimal generator

$$\mathcal{L}V(x^t) = \lim_{dt\to 0} \tfrac{\mathbb{E}V(x^{t+dt}) - V(x^t)}{dt}$$

$$= \lim_{dt\to 0} \tfrac{1}{dt}\Big[\mathbb{E}\Big(\tfrac{1}{\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|^2\Big) - \tfrac{1}{\|x^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t - \Pi_{\mathcal{X}}(x^t)\|^2\Big]$$

$$\leq \tfrac{1}{\|x^t - \Pi_{\mathcal{X}}(x^t)\|}\Big[\partial_x V(x^t)^T \Big([D^t L^t]_{i.} x^t\Big) + \tfrac{1}{2}\sigma^2(x^t)\Big].$$

From (14) the above implies that $\mathcal{L}V(x^t) < 0$, for all $x^t \notin \mathcal{M}$ and this concludes our proof. ■

B. Worst-case approach

Let us note that by substituting the mean-field equilibrium strategies $u^* = -R^{-1}B^T[PX + \Psi]$ and $w^* = \frac{1}{\gamma^2}D^T[PX + \Psi]$ as given in (**??**) in the open-loop microscopic dynamics $\dot{X}(t) = AX(t) + Bu(t) + C + Dw$ as defined in (**??**), the closed-loop microscopic dynamics is

$$\dot{x}_i^t = [D^t L^t]_{i.} x^t + k_1[\tilde{L}^t]_{i.} x^t + k_2[\bar{L}^t]_{i.} x^t \tag{15}$$

Now, let $\mathcal{X}$ be the set of equilibrium points for (**??**), namely, the set of X such that

$$\mathcal{X} = \{x \in ??| [D^t L^t]_{i.} x^t + k_1[\tilde{L}^t]_{i.} x^t + k_2[\bar{L}^t]_{i.} x^t = 0\},$$

and let $V(x^t) = dist(x^t, \mathcal{X})$. The next result establishes a condition under which the above dynamics converges asymptotically to the set of equilibrium points.

Theorem 2: **(worst-case stability)** If it holds

$$\begin{aligned} &\partial_x V(x^t)^T \Big([D^t L^t]_{i.} x^t + k_1[\tilde{L}^t]_{i.} x^t + k_2[\bar{L}^t]_{i.} x^t \Big) \\ &< -\|x^t - \Pi_{\mathcal{X}}(x^t)\|^2 \end{aligned} \tag{16}$$

then dynamics (15) is asymptotically stable, namely, $\lim_{t\to\infty} dist(x^t, \mathcal{X}) = 0$.

Proof: Let x^t be a solution of dynamics (15) with initial value $x^0 \neq \mathcal{X}$. Set $t = \{\inf t > 0 | \, x^t \in \mathcal{X}\} \leq \infty$ and let $V(x^t) = dist(x^t, \mathcal{X})$. For all $t \in [0, t]$

$$V(x^{t+dt}) - V(x^t) = \|x^{t+dt} - \Pi_{\mathcal{X}}(x^t)\| - \|x^t - \Pi_{\mathcal{X}}(x^t)\|$$

$$= \|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\| - \|x^t - \Pi_{\mathcal{X}}(x^t)\|$$

$$= \tfrac{1}{\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|^2$$

$$- \tfrac{1}{\|x^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t - \Pi_{\mathcal{X}}(x^t)\|^2.$$

From the definition of infinitesimal generator

$$\begin{aligned} \dot{V}(x^t) &= \lim_{dt\to 0} \tfrac{V(x^{t+dt}) - V(x^t)}{dt} \\ &= \lim_{dt\to 0} \tfrac{1}{dt}\Big[\tfrac{1}{\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|}\|x^t + dx^t - \Pi_{\mathcal{X}}(x^t)\|^2 \\ &\leq \tfrac{1}{\|x^t - \Pi_{\mathcal{X}}(x^t)\|}\Big[\partial_x V(x^t)^T \Big([D^t L^t]_{i.} x^t \\ &\qquad + k_1[\tilde{L}^t]_{i.} x^t + k_2[\bar{L}^t]_{i.} x^t\Big) \leq 0 \end{aligned}$$

which implies $\mathcal{L}V(x^t) < 0$, for all $x^t \neq \mathcal{X}$ and this concludes our proof.

■

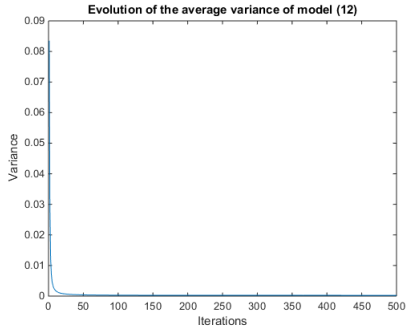

Fig. 1. Average variance of 100 simulation instances

Figure 1 shows the evolution of the variance of the model in 12, for 100 simulation instances of networks with 1000 nodes. Therein the variance is not only constant but also tends to zero as the number of iterations increases.

V. Simulation results

This section is divided in three parts. The first part illustrates the results of sample network when is generated either by SEE model, the deterministic model in (11), or by the stochastic model in (12). The second part present the average results in the cooperation level of each node of 100 network instances. Finally the third part contains simulation results and comparisons of the cooperation level and the distance between cars in a simulation scenario of cooperative autonomous cars moving in a heavy traffic road. In all parts we followed [20] and set the parameters for all models of cooperation level update to $b = 4$, $c = 1$ and $\beta = 1$.

A. Sample networks

In this section numerical results of a sampled network for the SEE model and the two models we propose, are presented. The results of the three models are present for a network evolution case where 1000 nodes, uniformly chosen initial cooperation level in the interval $0 \leq x_i \leq 1$, and initial degree of each node 10.

In all figures the top left plot depicts the clusters of coordination level in the last iteration of the evolution process, the evolution of the coordination level of all nodes as a function of time is depicted in the top right plots. The histogram on the bottom left depicts the distribution of the number connections between the nodes. Finally the bottom right plot depicts the variance of the cooperation level.

As it is depicted in Figures 2,3 and 4 the variance of the three coordination level vanishes very fast in all three models. On the other hand there are difference in the all the measures we report. In particular we can see that in the SEE model rang of the coordination levels among the nodes have greater range than the other two models, with the deterministic model in (11) having the smallest range. This is also depicted in the histogram of the clusters of coordination levels of these three Figures. In addition the number of clusters significantly differs between the SEE and the other two models, since in the two models we propose

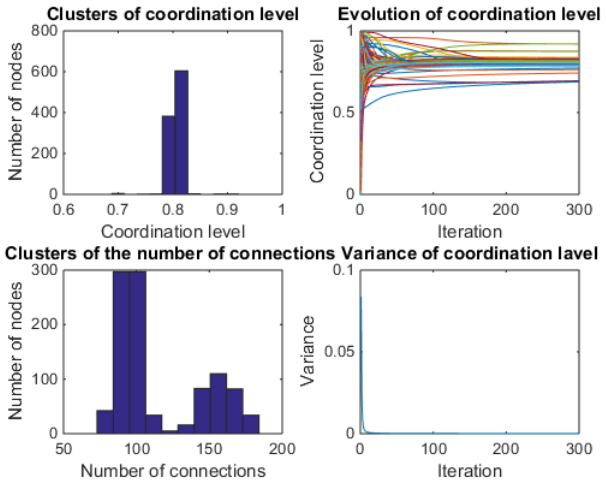

Fig. 2. Results for the SEE model.

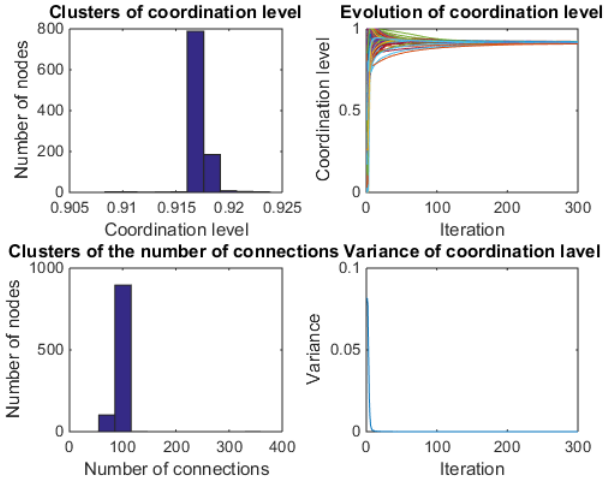

Fig. 3. Results for the deterministic model in 11

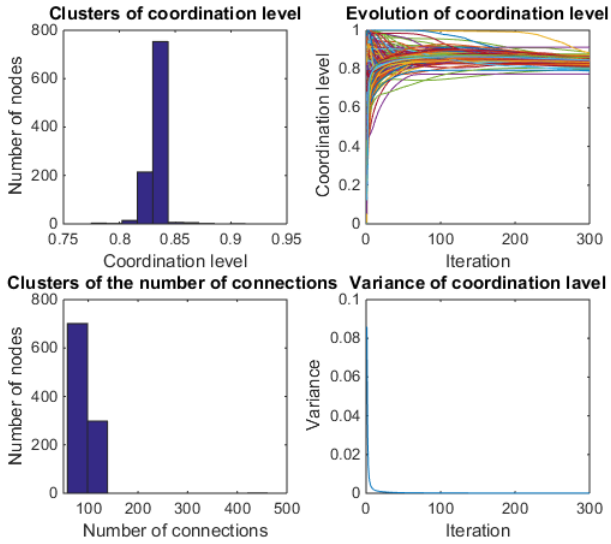

Fig. 4. Results for the deterministic model in 12

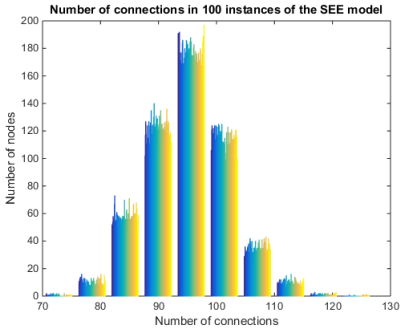

Fig. 5. Number of connections levels of 100 simulation instances of SEE model

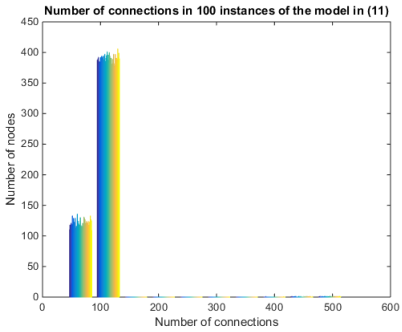

Fig. 6. Number of connections levels of 100 simulation instances of the model in (11)

there is a node that has significantly more connections than the others.

B. Results of 100 simulation instances.

This section illustrates the average results for 100 instance of network structures. In each instance a network of 1000 nodes was considered, with uniformly chosen initial cooperation levels in the interval $0 \leq x_i \leq 1$, and initial degree of each node 10.

Histograms 5, 6 and 7 depict the number of connections that was observed in each of the 100 networks simulated. The models in (11) and (12) in all simulations appear to generate a cluster of few nodes with significantly more connections than the majority of the nodes. The range of

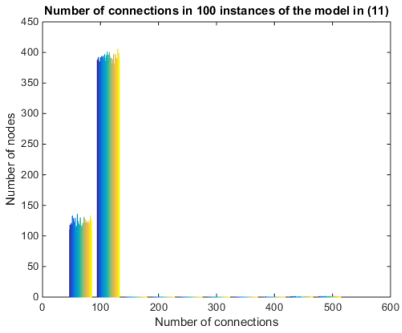

Fig. 7. Number of connections levels of 100 simulation instances of the model in 12

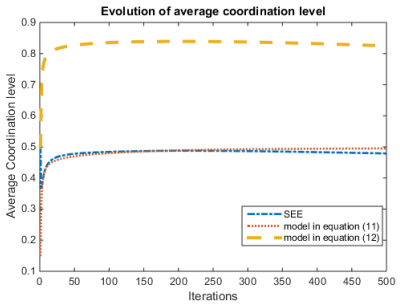

Fig. 8. Average coordination level of the three models as a function of time.

the number of connections in these two models is from 70 to 500 connections. On the other hand in the SEE model less nodes appear with the maximum number of connections to be less than 130 nodes. This difference is due to the different definitions of the probability to create links between the two proposed models and SEE model. Additionally the number of the links that was dropped in each instance didn't affect the nodes with the high connectivity in the two proposed models.

Figure 8 depicts the average coordination level of the three models. The two deterministic models, SEE and (11) had similar behaviour, with both models having a coordination level around 0.5. On the other hand the coordination level of the stochastic model after 50 iterations increased to 0.8.

C. Cooperative autonomous cars in heavy traffic

The optimal velocity model [17] and the Nagel-Schrekenberg model [19] are studying the changes in traffic flow in congested roads. In this article these two models in conjunction with models of network evolution in order to study the impact of coordination level in vehicles' behaviour, i.e. their average velocity and their coordination levels. In order to use the network evolution models each vehicle is assumed to be a node in the network and spatial restrictions are imposed to the nodes that can be connected. Thus a vehicle can influence with its behaviour, coordination level, only the vehicles that its near to it.

We will use two variants of the single lane congested road that used in both [17] and [19], to take into account the continuous and discrete nature of the velocities in each case. In the first variance we will place 100 vehicles in a single lane road, with the distances between vehicles to be uniformly chosen from the interval $[1, 30]$ meters. The initial velocities of the cars were also chosen uniformly from the interval $(0, 30]$ m/s. Each vehicle i adjusts its coordination levels x_i, based on the coordination levels of the vehicles in a range of 30 meters in front and behind it. Note that it is not necessary that a vehicle will be "connected", i.e. will take into account, with all the vehicles in its range. In order to define the cooperation levels of each vehicle the following two variables are introduced $d_{i,j}$ and d_{min} which denote the distance between vehicle i and j and the minimum safe distance when a vehicle is fully cooperative.

	SEE	Model in (11)	Model in (12)
Average Velocity	24.12 m/s	25.1077m/s	25.022m/s
Standard error 0.0259	0.0316	0.0413	

TABLE I

AVERAGE VELOCITY OF EACH AUTONOMOUS VEHICLE AND ITS CORRESPONDING STANDARD ERROR FOR THE THREE MODELS, WHEN THE OPTIMAL VELOCITY MODEL IS USED.

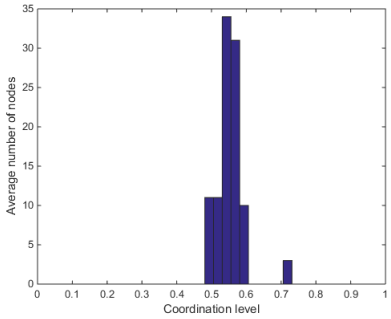

Fig. 9. Clusters of coordination level for the SEE model

The coordination level of a vehicle i can be defined then as $x_i = \frac{d_{i,j}}{d_{min}}$, where $d_{i,j}$ is the distance of vehicle i from the vehicle in front of it and $d_{min} = 30$ meters. If $d_{i,j} > 30$ meters or there is no vehicle in front of vehicle i, then $x_i = 1$. The cost function in (2) can be used in order to update the velocities. The range of the control function, acceleration is defined as $-u_{max} \leq u_i \leq u_{max}$, with $u_{max} = 3.75\ m/s^2$ and $t_h = 1.8$ seconds.

Table I illustrates the average velocity of a car for each of the three coordination models that were tested. When the deterministic model in (11) the maximum average velocity observed. This can be explained by its coordination levels, Figure 10, that had concentrated in the area of 0.5. When the other two models are considered their range of coordination level, Figures 9 and 10, was wider than the one in Figure 10. This can be explained by the spatial restrictions that are imposed and limits the number of possible vehicles that can be "connected" and therefore influence other vehicles behaviour.

In the second variant the single lane was discretised in cells of 7.5 meters, with each cell being occupied by one vehicle at each time. Similarly to the continuous case 100 vehicles were used, with the distances between vehicles to be uniformly chosen from the interval $[0, 4]$ cells. The initial velocities of the cars were also chosen uniformly from the interval $[1, 3]$ cells per time instance. Each vehicle i adjusts its coordination levels x_i, based on the coordination levels of the vehicles in a range of 4 cells in front and behind it.

ovmm1.png

Fig. 10. Clusters of coordination level for the model in (11)

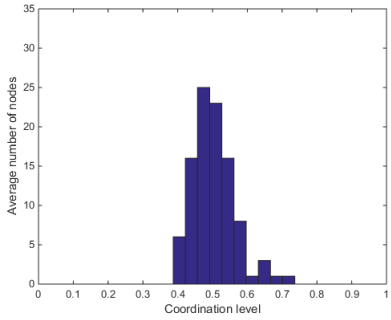

Fig. 11. Clusters of coordination level for the model in (12)

	SEE	Model in (11)	Model in (12)
Average Velocity	18.9847 m/s	19.152m/s	19.41m/s
Standard error	0.1725	0.2550	0.2175

TABLE II

AVERAGE VELOCITY OF EACH AUTONOMOUS VEHICLE AND ITS CORRESPONDING STANDARD ERROR FOR THE THREE MODELS, WHEN NAGEL-SCHREKENBERG MODEL IS USED.

again it is not necessary that a vehicle will be "connected" with all the vehicles in its range. The coordination level of a vehicle i can be defined then as $x_i = \frac{d_{i,j}}{d_{min}}$, where $d_{i,j}$ is the distance of vehicle i from the vehicle in front of it, number of empty cells in front of it $times7.5$, and $d_{min} = 30$ meters. If $d_{i,j} > 30$ meters or there is no vehicle in front of vehicle i, then $x_i = 1$. Then the variant of Nagel-Schrekenberg model presented in Section II-D can be used to update vehicles' velocities.

Similarly to the results of the optimal velocity model's, as it is shown in Table II, the absolute difference in the average velocity between the three models are small, but when the two proposed models are used we observe an improvement in average velocity.

Histograms 12, 13 and 14 depict the cluster of coordination levels that created over the 100 simulation instances and the average number of vehicles that belong to each cluster. The coordination level for the SEE model has smallest range than the proposed ones. The difference from the optimal velocity model can be explained from the extra restriction in the coordination levels which is now a discrete variable.

VI. CONCLUSIONS AND FUTURE WORK

In this article two variants of the SEE model for network evolution were presented. A feature of these variants that is absent of SEE model is the ability to drop links as well. In both variants regret based learning was used in order to define the probability to create a new link or drop an existing one. The second order convergence of the stochastic model was shown. Simulations were employed in order to study the properties of the proposed models. In addition simulations in order to identify the effect of the proposed models in the velocity of autonomous cars were employed.

In future work the effects that various parameters can have to the proposed variants of SEE will be studied. Additionally

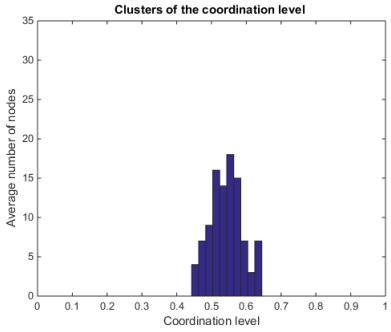

Fig. 12. Clusters of coordination level for the SEE model

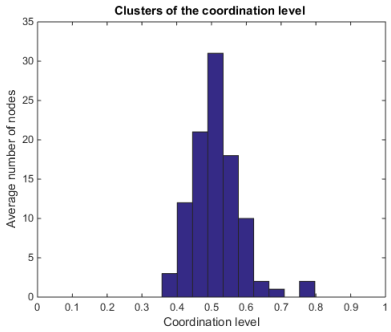

Fig. 13. Clusters of coordination level for the model in (11)

more complicated traffic models wil be considered in order to take into account multiple lanes or junctions.

REFERENCES

[1] T. Tatarenko, "Log-linear learning: Convergence in discrete and continuous strategy potential games," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 2014, pp. 426–432.

[2] L. Ng and M. Reza Emami, "A concurrent approach to robot team learning," in *Robotic Intelligence In Informationally Structured Space (RiiSS), 2013 IEEE Workshop on*, 2013, pp. 50–57.

[3] E. Uchibe, "Cooperative behavior acquisition by learning and evolution in a multi-agent environment for mobile robots," Ph.D. dissertation, Doctoral dissertation, Osaka University, 1999.

[4] J. Dai and H. Lin, "Learning-based design of fault-tolerant cooperative multi-agent systems," in *American Control Conference (ACC), 2015*. IEEE, 2015, pp. 1929–1934.

[5] C.-F. Juang, M.-G. Lai, and W.-T. Zeng, "Evolutionary fuzzy control and navigation for two wheeled robots cooperatively carrying an object in unknown environments," 2014.

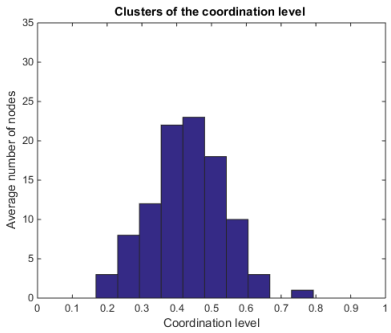

Fig. 14. Clusters of coordination level for the model in (12)

[6] S. Murata, Y. Yamashita, H. Arie, T. Ogata, J. Tani, and S. Sugano, "Generation of sensory reflex behavior versus intentional proactive behavior in robot learning of cooperative interactions with others," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*. IEEE, 2014, pp. 242–248.

[7] L. Rozo, S. Calinon, and D. G. Caldwell, "Learning force and position constraints in human-robot cooperative transportation," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on*. IEEE, 2014, pp. 619–624.

[8] X. Chen, B. Fu, Y. He, and M. Wu, "Timesharing-tracking framework for decentralized reinforcement learning in fully cooperative multi-agent system," *Automatica Sinica, IEEE/CAA Journal of*, vol. 1, no. 2, pp. 127–133, 2014.

[9] J. Dai and H. Lin, "Automatic synthesis of cooperative multi-agent systems," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE, 2014, pp. 6173–6178.

[10] S. Hart and A. Mas-Colell, "A general class of adaptive strategies," *Journal of Economic Theory*, vol. 98, no. 1, pp. 26–54, 2001.

[11] A. C. Chapman, A. Rogers, and N. R. Jennings, "Benchmarking hybrid algorithms for distributed constraint optimisation games," *Autonomous Agents and Multi-Agent Systems*, vol. 22, no. 3, pp. 385–414, 2011.

[12] J. M. Maestre and R. R. Negenborn, Eds., *Distributed Model Predictive Control Made Easy*. Springer, 2014.

[13] J. B. Rawlings and B. T. Stewart, "Coordinating multiple optimization-based controllers: New opportunities and challenges," *Journal of Process Control*, vol. 18, pp. 839–845, 2008.

[14] B. T. Stewart, A. N. Venkat, J. B. Rawlings, S. J. Wright, and G. Pannocchia, "Cooperative distributed model predictive control," *Systems & Control Letters*, vol. 59, pp. 460–469, 2010.

[15] Y. Kuwata and J. P. How, "Cooperative distributed robust trajectory optimization using receding horizon MILP," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 2, pp. 423–431, 2011.

[16] P. A. Trodden and A. G. Richards, "Cooperative distributed MPC of linear systems with coupled constraints," *Automatica*, vol. 49, no. 2, pp. 479–487, February 2013.

[17] C.-C. Chien, Y. Zhang, and P. A. Ioannou, "Traffic density control for automated highway systems," *Automatica*, vol. 33, no. 7, pp. 1273–1285, 1997.

[18] M. A. Samad Kamal, J.-i. Imura, T. Hayakawa, A. Ohata, and K. Aihara, "Smart driving of a vehicle using model predictive control for improving traffic flow," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 2, pp. 878–888, 2014.

[19] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.

[20] B. Ranjbar-Sahraei, D. Bloembergen, H. B. Ammar, K. Tuyls, and G. Weiss, "Effects of evolution on the emergence of scale free networks," in *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, vol. 14, 2014, pp. 376–383.

[21] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[22] B. Ranjbar-Sahraei, H. Bou Ammar, D. Bloembergen, K. Tuyls, and G. Weiss, "Evolution of cooperation in arbitrary complex networks," in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2014, pp. 677–684.