# Cooperative learning in multi-agent systems from intermittent measurements

Naomi Ehrich Leonard Alex Olshevsky

***Abstract*— Motivated by the problem of decentralized direction-tracking, we consider the general problem of cooperative learning in multi-agent systems with time-varying connectivity and intermittent measurements. We propose a distributed learning protocol capable of learning an unknown vector $\mu$ from noisy measurements made independently by autonomous nodes. Our protocol is completely distributed and able to cope with the time-varying, unpredictable, and noisy nature of inter-agent communication, and intermittent noisy measurements of $\mu$. Our main result bounds the learning speed of our protocol in terms of the size and combinatorial features of the (time-varying) network connecting the nodes.**

## I. INTRODUCTION

Widespread deployment of mobile sensors is expected to revolutionize our ability to monitor and control physical environments. However, for these networks to reach their full range of applicability they must be capable of operating in uncertain and unstructured environments. Realizing the full potential of networked sensor systems will require the development of protocols that are fully distributed and adaptive in the face of persistent faults and time-varying, unpredictable environments.

Our goal in this paper is to initiate the study of cooperative multi-agent learning by distributed networks operating in unknown and changing environments, subject to faults and failures of communication links. While our focus here is on the basic problem of learning an unknown vector, we hope to contribute to the development of a broad theory of cooperative, distributed learning in such environments, with the ultimate aim of designing sensor network protocols capable of adaptability.

We will study a simple, local protocol for learning a vector from intermittent measurements and evaluate its performance in terms of the number of nodes and the (time-varying) network structure. Our direct motivation is the problem of direction tracking from chemical gradients. A network of mobile sensors needs to move in a direction $\mu$ (understood as a vector on the unit circle), which none of the sensors initially knows; however, intermittently some sensors are able to obtain a sample of $\mu$. The sensors can observe the velocity of neighboring sensors but, as the sensors move, the set of neighbors of each sensor changes; moreover, new sensors occasionally join the network and current sensors sometimes permanently leave the network. The challenge is to design a protocol by means of which the sensors can adapt their velocities based on the measurements of $\mu$ and observations of the velocities of neighboring sensors so that every node's velocity converges to $\mu$ as fast as possible. This challenge is further complicated by the fact that all estimates of $\mu$ as well as all observations of the velocities of neighbors are assumed to be noisy.

We will consider a natural generalization in the problem, wherein we abandon the constraint that $\mu$ lies on the unit circle and instead consider the problem of learning an arbitrary vector $\mu$ by a network of mobile nodes subject to time-varying (and unpredictable) inter-agent connectivity, and intermittent, noisy measurements. We will be interested in the speed at which local, distributed protocols are able to drive every node's estimate of $\mu$ to the correct value. We will be especially concerned with identifying the salient features of network topology that result in good (or poor) performance.

## II. COOPERATIVE MULTI-AGENT LEARNING

We begin by formally stating the problem for a fixed number of nodes. We consider $n$ autonomous nodes engaged in the task of learning a vector $\mu \in \mathbb{R}^l$. At each time $t = 0, 1, 2, \ldots$ we denote by $G(t) = (V(t), E(t))$ the graph of inter-agent communications at time $t$: two nodes are connected by an edge in $G(t)$ if and only if they are able to exchange messages at time $t$. Note that by definition the graph $G(t)$ is undirected. If $(i, j) \in G(t)$ then we will say that $i$ and $j$ are neighbors at time $t$. We will adopt the convention that $G(t)$ contains no self-loops. We will assume the graphs $G(t)$ satisfy a standard condition of uniform connectivity over a long-enough time scale: namely, there exists some constant positive integer $B$ (unknown to any of the nodes) such that the graph sequence $G(t)$ is $B$-connected, i.e. the graphs $(\{1, \ldots, n\}, \bigcup_{kB+1}^{(k+1)B} E(t))$ are connected for each integer $k \geq 0$. Intuitively, the uniform connectivity condition means that once we take all the edges that have appeared between times $kB$ and $(k+1)B$, the graph is connected.

Each node maintains an estimate of $\mu$; we will denote the estimate of node $i$ at time $t$ by $v_i(t)$. At time $t$, node $i$ can update $v_i(t)$ as a function of the noise-corrupted estimates $v_j(t)$ of its neighbors. We will use $o_{ij}(t)$ to denote the noise-corrupted estimate of the offset $v_j(t) - v_i(t)$ available to neighbor $i$ at time $t$:

$$o_{ij}(t) = v_j(t) - v_i(t) + w_{ij}(t)$$

Here $w_{ij}(t)$ is a zero-mean random vector every entry of which has variance $(\sigma')^2$, and all $w_{ij}(t)$ are assumed to

This work was supported in part by AFOSR grant FA9550-07-1-0-0528, ONR grant N00014-09-1-1074 and ARO grant W911NG-11-1-0385.

Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ, 08544, USA (naomi@princeton.edu).

Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, USA (aolshev2@illinois.edu).

be independent of each other, as well as all other random variables in the problem (which we will define shortly). These updates may be the result of a wireless message exchange or may come about as a result of sensing by each node. Physically, each node is usually able to sense (with noise) the relative difference $v_j(t) - v_i(t)$, for example if $v_i(t)$ represent velocities and measurements by the agents are made in their frame of reference; alternatively, it may be that nodes are able to measure the absolute quantities $v_j(t), v_i(t)$ and then $w_{ij}(t)$ is the sum of the noises in these two measurements.

Occasionally, some nodes have access to a noisy measurement

$$\mu_i(t) = \mu + w_i(t),$$

where $w_i(t)$ is a zero-mean random vector every entry of which has variance $\sigma^2$; we assume all vectors $w_i(t)$ are independent of each other and of all $w_{ij}(t)$. In this case, node $i$ incorporates this measurement into its updated estimate $v_i(t+1)$. We will refer to a time $t$ when at least one node has a measurement as a *measurement time*. For the rest of the paper, we will be making an assumption of uniform measurement speed, namely that no more than $T$ steps pass between successive measurement times; more precisely, letting $t_k$ be the times when at least one node makes a measurement, we will assume that $t_0 = 0$ and $|t_{k+1} - t_k| \leq T$ every all nonnegative integer $k$.

It is useful to think of this formalization in terms of our motivating scenario, which is a collection of nodes - vehicles, UAVs, mobile sensors, or underwater gliders - which need to learn and follow a direction. Updated information about the direction arrives from time to time as one or more of the nodes takes measurements, and the nodes need a protocol by which they update their velocities $v_i(t)$ based on the measurements and observations of the velocities of neighboring nodes.

This formalization also describes the scenario in which a moving group of animals must all learn which way to go based on intermittent samples of a preferred direction and social interactions with near neighbors. An example is collective migration where high costs associated with obtaining measurements of the migration route suggest that the majority of individuals rely on the more accessible observations of the relative motion of their near neighbors when they update their own velocities $v_i(t)$ [22].

## III. Our results

We now describe the protocol which we analyze for the remainder of this paper. If at time $t$ node $i$ does not have a measurement of $\mu$, it moves its velocity in the direction of its neighbors:

$$v_i(t+1) = v_i(t) + \frac{\Delta(t)}{4} \sum_{j \in N_i(t)} \frac{o_{ij}(t)}{\max(d_i(t), d_j(t))}. \quad (1)$$

where $N_i(t)$ is the set of neighbors of node $i$, $d_i(t)$ is the cardinality of $N_i(t)$, and $\Delta(t)$ is a stepsize which we will specify later.

On the other hand, if node $i$ does have a measurement $\mu_i(t)$, it updates as

$$v_i(t+1) = v_i(t) + \frac{\Delta(t)}{4} (\mu_i(t) - v_i(t)) + \frac{\Delta(t)}{4} \sum_{j \in N_i(t)} \frac{o_{ij}(t)}{\max(d_i(t), d_j(t))}. \quad (2)$$

Intuitively, each node seeks to align its estimate $v_i(t)$ with both the measurements it takes and estimates of neighboring nodes. As nodes align with one another, information from each measurement slowly propagates throughout the system.

Our protocol is motivated by a number of recent advances within the literature on multi-agent consensus. On the one hand, the weights $(1/4)/\max(d(i), d(j))$ we accord to neighboring nodes are based on Metropolis weights (first introduced within the context of multi-agent control in [9]) and are chosen because they lead to a tractable Lyapunov analysis as in [41]. On the other hand, we introduce a stepsize $\Delta(t)$ which we will later choose to decay to zero with $t$ at an appropriate speed by analogy with the recent work on multi-agent optimization [42], [55], [60].

The use of a stepsize $\Delta(t)$ is crucial for the system to be able to successfully learn the unknown vector $\mu$. Intuitively, as $t$ gets large, the nodes should avoid overreacting by changing their estimates in response to every new noisy sample. Rather, the influence of every new sample on the estimates $v_1(t), \ldots, v_n(t)$ should decay with $t$: the more information the sensors have collected in the past, the less they should be inclined to revise their estimates in response to a new sample. This is accomplished by ensuring that the influence of each successive new sample decays with the stepsize $\Delta(t)$.

Our protocol is also motivated by models used to analyze collective decision making and collective motion in animal groups [20], [33]. Our time varying stepsize rule is similar to models of context-dependent interaction in which individuals reduce their reliance on social cues when they are progressing towards their target [57].

We now proceed to set the background for our main result, which bounds the rate at which the estimates $v_i(t)$ converge to $\mu$. We first state a proposition which assures us that the estimates $v_i(t)$ do indeed converge to $\mu$ almost surely. A proof may be found in the technical report [32].

**Proposition 1.** *If the stepsize $\Delta(t)$ is nonnegative, nonincreasing and satisfies*

$$\sum_{t=1}^{\infty} \Delta(t) = +\infty, \qquad \sum_{t=1}^{\infty} \Delta^2(t) < \infty,$$

$$\sup_{t \geq 1} \frac{\Delta(t)}{\Delta(t+c)} < \infty \qquad \textit{for any integer } c$$

*then for any initial values $v_1(0), \ldots, v_n(0)$, we have that with probability* 1

$$\lim_{t \to \infty} v_i(t) = \mu \qquad \textit{for all } i.$$

We remark that this proposition may be viewed as a generalization of earlier results on leader-following, which achieved similar conclusions either without the assumptions of noise, or on fixed graphs, or with the assumption of a fixed leader (see [23], [43], [38], [40]). Our protocol is very much in the spirit of this earlier literature. All the previous protocols (as well as ours) may be thought of consensus protocols driven by noisy inputs, and we note there are a number of other possible variations on this theme which can accomplish the task of learning the unknown vector $\mu$.

Our main result in this paper is a strengthened version of Proposition 1 which provides quantitative bounds on the rate at which convergence to $\mu$ takes place. We are particularly interested in the scaling of the convergence time with the number of nodes and with the combinatorics of the interconnection graphs $G(t)$. We will adopt the natural measure of how far we are from convergence, namely the sum of the squared distances from the final limit:

$$Z(t) = \sum_{i=1}^{n} ||v_i(t) - \mu||_2^2.$$

We will refer to $Z(t)$ as the variance at time $t$.

Before we state our main theorem, we introduce some notation. First, we define the the notion of the lazy Metropolis walk on an undirected graph: this is the random walk which moves from $i$ to $j$ with probability $1/(4\max(d(i), d(j)))$ whenever $i$ and $j$ are neighbors. Moreover, given a random walk on a graph, the hitting time from $i$ to $j$ is defined to be the expected time until the walk visits $j$ starting from $i$. We will use $d_{\max}$ to refer to the largest degree of any node in the sequence $G(t)$ and $M$ to refer to the largest number of nodes that have a measurement at any one time; clearly both $d_{\max}$ and $M$ are at most $n$. Finally, $\lceil x \rceil$ denotes the smallest integer which is at least $x$. With this notation in place, we now state our main result. A proof of it may be found in the technical report [32].

**Theorem 2.** *Let the stepsize be $\Delta(t) = 1/(t+1)^{1-\epsilon}$ for some $\epsilon \in (0,1)$. Suppose each of the graphs $G(t)$ is connected and let $\mathcal{H}$ be the largest hitting time from any node to any node in a lazy Metropolis walk on any of the graphs $G(t)$. If t satisfies the lower bound*

$$t \geq 3T \left[ \frac{432T\mathcal{H}}{\epsilon} \ln \left( \frac{144T\mathcal{H}}{\epsilon} \right) \right]^{1/\epsilon},$$

*then we have that Eq. (3) holds.*

*In the general case when each $G(t)$ is not necessarily connected but the sequence $G(t)$ is $B$-connected, we have that if t satisfies the lower bound*

$$t \geq 2 \left[ \frac{576Tn^2 d_{\max}}{\epsilon} \ln \left( \frac{192Tn^2 d_{\max}}{\epsilon} \right) \right]^{1/\epsilon},$$

*then we have that Eq. (4) holds.*

Our theorem provides a quantitative bound on the convergence time of the repeated alignment process of Eq. (1) and Eq. (2). We believe this is the first time a convergence time result has been demonstrated in the setting of time-varying (not necessarily connected) graphs, intermittent measurements by possibly different nodes, and noisy communications among nodes. The convergence time expressions are somewhat unwieldy, and we pause now to discuss some of their features.

First, observe that the convergence times are a sum of two terms: the first which decays with $t$ as $O(1/t^{2-2\epsilon})$ and the second which decays as $O(e^{-t^{\epsilon}})$ (here $O$-notation hides all terms that do not depend on $t$). In the limit of large $t$, the second will be negligible and we may focus our attention solely on the first. Thus our finding is that it is possible to achieve a nearly quadratic decay with time by picking a stepsize $1/(t+1)^{1-\epsilon}$ with $\epsilon$ close to zero.

Moreover, we find that for every choice of $\epsilon \in (0,1)$, the scaling with the number of nodes $n$ is polynomial. Moreover, in analogy to some recent work on consensus [41], better convergence time bounds are available when the largest degree of any node is small. This is somewhat counter-intuitive since higher degrees are associated with improved connectivity. A plausible intuitive explanation for this mathematical phenomenon is that low degrees ensure that the influence of new measurements on nodes does not get repeatedly diluted in the update process.

Furthermore, while it is possible to obtain a nearly quadratic decay with the number of iterations $t$ as we just noted, such a choice of $\epsilon$ blows up the bound on the transient period before the asymptotic decay bound kicks in. Every choice of $\epsilon$ then provides a tradeoff between the transient size and the asymptotic rate of decay. This is to be contrasted with the usual situation in distributed optimization (see e.g., [49], [55]) where a specific choice of stepsize usually results in the best bounds.

Moreover, the constant in front of asymptotically dominant $1/t^{2-2\epsilon}$ decay in Theorem 2 scales as some function of the graph sequence (i.e., as either maximum hitting time or as $n^2 d_{\max}$) raised to the power $1/\epsilon$. Thus the choice of the constant $\epsilon$ may additionally be thought to tradeoff scaling between scaling with time (where $\epsilon$ close to zero gives the better scaling) and network features (where $\epsilon$ close to one gives the better scaling).

Finally, in the case when all graphs are connected, the effect of network topology on the convergence time comes through the maximum hitting time $\mathcal{H}$ in all the individual graphs $G(t)$. There are a variety of results on hitting times for various graphs which may be plugged into Theorem 2 to obtain precise topology-dependent estimates. We first mention the general result that $\mathcal{H} = O(n^2)$ for an arbitrary connected graph from [44]. On a variety of reasonably connected graphs, hitting times are considerably smaller. A recent preprint [59] shows that for many graphs, hitting times are proportional to the inverse degrees. In a 2D grid, hitting time is $O(n \log^2 n)$ while in the 3D grid hitting time is $O(n \log n)$ [15].

We illustrate the convergence times of Theorem 2 with a concrete example. Suppose we have a collection of nodes

$$E[Z(t) \mid v(0)] \leq 9lT^2 \left(M\sigma^2 + nT(\sigma')^2\right) \frac{(24T\mathcal{H})^{1/\epsilon}}{t^{2-2\epsilon}} + Z(0)e^{-((t/T-1)^{\epsilon}-1)/(24T\mathcal{H}\epsilon)}. \tag{3}$$

$$\begin{aligned} E[Z(t) \mid v(0)] \quad \leq \quad & 9l\max(T,B)^2(M\sigma^2 + n\max(T,B)(\sigma')^2)\frac{(32\max(T,B)n^2 d_{\max})^{1/\epsilon}}{t^{2-2\epsilon}} \\ & + Z(0)e^{-((t/\max(T,B)-1)^{\epsilon}-1)/(32n^2 d_{\max}\max(T,B)\epsilon)}. \end{aligned} \tag{4}$$

interconnected in (possibly time-varying) 2D grids with a single (possibly different) node sampling at every time. Let us further assume that communication among nodes is noiseless ($\sigma' = 0$) while the dimension $l$ of the vector we are learning as well as the noise variance $\sigma^2$ are constants independent of the number of nodes. Choosing a step size $\Delta(t) = 1/\sqrt{t}$, we have that Theorem 2 implies that variance $E[Z(t) \mid Z(0)]$ will fall below $\delta$ after

$$\max\left(O\left(\frac{n^2\log^4 n}{\delta}\right),\ O\left(n^2\log^4 n\left[\log\frac{Z(0)}{\delta}\right]^2\right)\right)$$

steps of the protocol.

## IV. Related work

We believe that our paper is the first to derive rigorous convergence time results for the problem of cooperative multi-agent learning by a network subject to unpredictable communication disruptions and intermittent measurements. The key features of our model are 1) its cooperative nature (many nodes working together) 2) its reliance only on distributed and local observations 3) the incorporation of time-varying communication restrictions.

Naturally, our work is not the first attempt to fuse learning algorithms with distributed control or multi-agent settings. Indeed, the study of learning in games is a classic subject which has attracted considerable attention within the last couple of decades due in part to its applications to multi-agent systems. We refer the reader to the recent papers [2], [7], [8], [17], [16], [10], [21], [39], [1], [37], [24] as well as the classic works [34], [19] which study multi-agent learning in a game-theoretic context. Moreover, the related problem of distributed reinforcement learning has attracted some recent attention; we refer the reader to [34], [56], [50]. We make no attempt to survey these literatures here and refer the reader to the references in the above papers, as well as the surveys [54], [47]. Moreover, we note that much of the recent literature in distributed robotics has focused on distributed algorithms robust to faults and communication link failures. We refer the reader to the representative papers [4], [36].

Our work here is very much in the spirit of the recent literature on distributed filtering [45], [46], [52], [3], [53], [13], [35], [14], [28], [30], [51] and especially [12]. These works consider the problem of tracking a time-varying signal from local measurements by each node, which are then repeatedly combined through a consensus-like iteration. The above-referenced papers consider a variety of schemes to this effect and obtain bounds on their performance, usually stated in terms of solutions to certain Lyapunov equations. Our work is also related to a number of recent papers on distributed detection [28], [25], [5], [6], [30], [29], [26], [27] which seek to evaluate protocols for networked cooperative hypothesis testing and related problems. Like the previously mentioned work on distributed filtering, these papers use the idea of local iterations which are combined through a distributed consensus update, termed "consensus plus innovations"; a similar idea is called "diffusion adaptation" in [51]. The focus of the distributed detection literature has been on obtaining large deviation bounds for choosing the right hypothesis in fixed and i.i.d. random networks.

In this work, we consider the related (and often simpler) question of learning a static unknown vector. However, we derive results which are considerably stronger compared to what is available in the filtering literature, obtaining convergence rates in settings when the network is time-varying and measurements are intermittent. Most importantly, we are able to explicitly bound the speed of convergence to the unknown vector $\mu$ in these unpredictable settings in terms of network size and combinatorial features of the networks.

## V. Conclusion

We have proposed a model for cooperative learning by multi-agent systems facing time-varying connectivity and intermittent measurements. We have proved a protocol capable of learning an unknown vector from independent measurements in this setting and provided quantitative bounds on its learning speed. Crucially, these bounds have a dependence on the number of agents $n$ which grows only polynomially fast, leading to reasonable scaling for our protocol. On sequences of connected graphs, the largest hitting time turned out to be the most relevant combinatorial primitivie.

Our research points to a number of intriguing open questions. Our results are for undirected graphs and it is unclear whether there is a learning protocol which will achieve similar bounds (i.e., a learning speed which depends only polynomially on $n$) on directed graphs. It appears that our bounds on the learning speed are loose by several orders of magnitude when compared to simulations, so that the learning speeds we have presented in this paper could potentially be further improved. Moreover, it is further possible that a different protocol provides a faster learning speed compared to the one we have provided here.

Finally, and most importantly, it is of interest to develop a general theory of decentralized learning capable of handling situations in which complex concepts need to be learned by distributed network subject to time-varying connectivity and intermittent arrival of new information. Consider, for example, a group of UAVs all of which need to learn a new strategy to deal with an unforeseen situation, for example, how to perform formation maintenance in the face of a particular pattern of turbulence. Given that selected nodes can try different strategies, and given that nodes can observe the actions and the performance of neighboring nodes, is it possible for the entire network of nodes to collectively learn the best possible strategy? A theory of general-purpose decentralized learning, designed to parallel the theory of PAC (Provably Approximately Correct) learning in the centralized case, is warranted.

## References


[1] D. Acemoglu, M. Dahleh, I. Lobel, A. Ozdaglar, "Bayesian learning in social networks," *Review of Econoics Studies*, vol. 78, no. 4, pp. 1201-1236, 2011.

[2] S. Adlakha, R. Johari, G. Weintraub, and A. Goldsmith, "Oblivious equilibrium: an approximation to large population dynamic games with concave utility," *Game Theory for Networks*, pp. 68-69, 2009.

[3] P. Alriksson, A. Rantzer, "Distributed Kalman filtering using weighted averaging," *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, 2006.

[4] R. Alur, A. D'Innocenzo, K. H. Johansson, G. J. Pappas, G. Weiss, "Compositional modeling and analysis of multi-hop control networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2345-2357, 2011.

[5] D. Bajovic, D. Jakovetic, J.M.F. Moura, J. Xavier, B. Sinopoli "Large deviations performance of consensus+innovations distributed detection with non-Gaussian observations," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5987-6002, 2012.

[6] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, J.M.F. Moura, "Distributed detection via Gaussian running consensus: large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4381-4396, 2011.

[7] M.-F. Balcan, F. Constantin, G. Piliouras, and J.S. Shamma, "Game couplings: Learning dynamics and applications," *Proceedings of IEEE Conference on Decision and Control*, 2011.

[8] B. Banerjee, J. Peng, "Efficient no-regret multiagent learning," *Proceedings of The 20th National Conference on Artificial Intelligence,* 2005.

[9] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing markov chain on a graph," *SIAM Review*, 46(4):667-689, 2004.

[10] M. Bowling, "Convergence and no-regret in multiagent learning," *Advances in Neural Information Processing Systems*, pp. 209, 2005.

[11] G. Brightwell, P. Winkler, "Maximum hitting time for random walks on graphs," *Random Structures and Algorithms*, vol. 1, no. 3, pp. 263-276, 1990.

[12] R. Carli, A. Chiuso, L. Schenato, S. Zampieri, "Distributed Kalman filtering based on consensus strategies," *IEEE Journal on Selected Areas in Communication*, vol. 26, no. 4, pp. 622-633, 2008.

[13] F.S. Cativelli, A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Transactions on Automatic Control*, vol. 55, no. 9, pp. 2069-2084, pp. 2069-2084, 2010.

[14] F.S. Cativelli, A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, no. 55, pp. 1917-1932, 2011.

[15] A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, P. Tiwari, "The electrical resistance of a graph captures its commute and cover time," *Computational Complexity*, vol. 6, no. 4, pp. 312-340, 1996.

[16] G.C. Chasparis, J.S. Shamma, and A. Arapostathis, "Aspiration learning in coordination games", *Proceedings of the 49th IEEE Conference on Decision and Control*, 2010.

[17] G.C. Chasparis, J.S. Shamma, and A. Rantzer, "Perturbed learning automata for potential games," *Proceedings of the 50th IEEE Conference on Decision and Control*, 2011.

[18] K.L. Chung, "On a stochastic approximation method," *Annals of Mathematical Statistics,* vol. 25, no. 3, pp. 463-483, 1954.

[19] C. Claus, C. Boutellier, "The Dynamics of reinforcement learning in cooperative multiagent systems," *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, 1998.

[20] I. D. Couzin, J. Krause, N. R. Franks and S. A. Levin, "Effective leadership and decision-makingin animal groups on the move," *Nature*, 433, 2005.

[21] R. Gopalakrishnan, J. R. Marden, and A. Wierman, "An architectural view of game theoretic control," *ACM Sigmetrics Performance Evaluation Review*, 38(3), 2011.

[22] V. Guttal and I. D. Couzin, "Social interactions, information use, and the evolution of collective migration," *Proceedings of the National Academy of Sciences*, 107(37), 2010.

[23] A. Jadbabaie, J. Lin, A.S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988-1001, 2003.

[24] A. Jadbabaie, P. Molavi, A. Sandroni, A. Tahbaz-Salehi, *Games and Economic Behavior*, vol. 76, pp. 210-225, 2012.

[25] D. Jakovetic, J.M.F. Moura, J. Xavier, "Distributed detection over noisy networks: large deviations analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4306 4320, 2012

[26] D. Jakovetic, J.M.F. Moura, J. Xavier, "Consensus+innovations detection: phase transition under communication noise," *Proceedings of the 50th Annual Allerton Conference on Communication, Control and Computing*, 2012.

[27] D. Jakovetic, J.M.F. Moura, J. Xavier, "Fast cooperative distributed learning,"*Proceedings of the 46th Asilomar Conference on Signals, Systems, and Computers*, 2012.

[28] S. Kar, J.M.F. Moura, "Gossip and distributed Kalman filtering: weak consensus under weak detectability," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1766-1784, 2011.

[29] S. Kar, J.M.F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and tradeoffs," *IEEE Journal on Selected Topics on Signal Processing*, vol. 5, no. 4, pp.: 674-690, 2011.

[30] U. Khan, J.M.F. Moura, "Distributing the Kalman Filters for Large-Scale Systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919-4935, 2008.

[31] H.J. Landau, A. M. Odlyzko, "Bounds for eigenvalues of certain stochastic matrices," *Linear Algebra and Its Applications,* vol. 38, pp. 5-15, 1981.

[32] N. E. Leonard, A. Olshevsky, "Cooperative learning in multi-agent systems from intermittent measurements," preprint, available at `http://arxiv.org/abs/1209.2194`.

[33] N. E. Leonard, T. Shen, B. Nabet, L. Scardovi, I. D. Couzin and S. A. Levin, "Decision versus compromise for animal groupsin motion", *Proceedings of the National Academy of Sciences*, 109(1), 2012.

[34] M.L. Littman, "Markov games as a framework for multi-agent reinforcement learning, *Proceedings of the Eleventh International Conference on Machine Learning*, 1994.

[35] C. G. Lopes, A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, 2007.

[36] S. Martinez, F. Bullo, J. Corts, E. Frazzoli, "On synchronous robotic networks - Part I: Models, tasks and complexity," *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2199-2213, 2007.

[37] S. Mannor, J. Shamma, "Multi-agent learning for engineers," *Artificial Intelligence*, vol. 171, no. 7, pp. 417-422, 2007.

[38] C. Ma, T. Li, J. Zhang, "Consensus control for leader-following multi-agent systems with measurement noises," *Journal of Systems Science and Complexity*, vol. 23, no. 1, pp. 35-49, 2010.

[39] J.R. Marden, G. Arslan, J.S. Shamma, "Cooperative control and potential games," *IEEE Transacions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(6), pp. 1393-1407, 2009.

[40] Z. Meng, W. Ren, Y. Cao, Z. You, "Leaderless and leader-following consensus with communication and input delays under a directed network topology," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, vol. 41, no. 1, 2011.

[41] A. Nedic, A. Olshevsky, A. Ozdaglar, J.N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Transactions on Automatic Control,* vol. 54, no. 11, pp. 2506-2517, 2009.

[42] A. Nedic, A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48-61, 2009.

[43] W. Ni, D. Zhang, "Leader-following consensus of multi-agent systems

under fixed and switching topologies," *Systems and Control Letters*, vol. 59, no. 3-4, pp. 209-217, 2010.

[44] Y. Nonaka, H. Ono, K. Sadakane, M. Yamashita, "The hitting and cover times of Metropolis walks," *Theoretical Computer Science*, vol. 411, no. 16-18, pp. 1889-1894, 2010.

[45] R. Olfati-Saber, "Distributed Kalman filter with embeddable consensus filters," *Proceedings of the 44th IEEE Conference on Decision and Control and European Control Confernece*, 2005.

[46] R. Olfati-Saber, "Distributed Kalman filtering for sensor networks," *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007.

[47] L. Panais, S. Luke, "Cooperative multi-agent learning: the state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387-434, 2005.

[48] B. Polyak, *Introduction to Optimization*, New York: Optimization Software Inc, 1987.

[49] S.S. Ram, A. Nedic, V.V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications,* vol. 147, no. 3, pp. 516-545, 2010.

[50] T.W. Sandholm, R.H. Crites, "Multiagent reinforcement learning in the iterated prisoner's dilemma," *Biosystems*, vol. 37, no. 1-2, pp. 147-166, 1996.

[51] A.H. Sayed, "Diffusion adaptation over networks," available at `http://arxiv.org/abs/1205.4220`.

[52] D. Spanos, R. Olfati-Saber, R. M. Murray, "Approximate distributed Kalman filtering in sensor networks with quantifiable performance," *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks*, 2005.

[53] A. Speranzon, C. Fischione, K. Johansson, "Distributed and collaborative estimation over wireless sensor networks," *Proceedings of the IEEE Conference on Decision and Control*, 2006.

[54] P. Stone, M. Veloso, "Multiagent systems: a survey from a machine learning perspective," *Autonomous Robots*, vol. 8, no. 3, pp. 345-383, 2000.

[55] K. Srivastava and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772-790, 2011.

[56] M. Tan, "Multi-agent reinforcement learning: independent vs. cooperative agents," *Proceedings of the Tenth International Conference on Machine Learning*, 1993.

[57] C. Torney, Z. Neufeld and I. D. Couzin, "Context-dependent interaction leads to emergent search behavior in social aggregates," *Proceedings of the National Academy of Sciences*, 106(52), 2009.

[58] B. Touri and A. Nedic, "On existence of a quadratic comparison function for random weighted averaging dynamics and its implications," *Proceedings of the 50th IEEE Conference on Decision and Control,* 2011.

[59] U. von Luxburg, A. Radl, M. Hein, "Hitting and commute times in large graphs are often misleading," preprint, available at `http://arxiv.org/abs/1003.1266`.

[60] F. Yousefian, A. Nedic, U.V. Shanbhag, "On stochastic gradient and subgradient methods with adaptive steplength sequences," *Automatica*, vol. 48, no. 1, pp. 56-67, 2012.