USING MARGINAL MEDIAN AND BLOCK LEAST MEAN SQUARES IN MOBILE ADAPTIVE NETWORKS

Fotios Mandanas and Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki, 54124, Greece
Email: {fmandan@gmail.com, costas@aiia.csd.auth.gr}

ABSTRACT

Here, we are interested in the case where a randomly chosen subset of agents share deliberately with their neighbors local information corrupted with gross errors, which is a common phenomenon in strategic communication. Within this framework, we propose a twofold extension of the state-of-the-art diffusion adaptation techniques over mobile networks. First, the agents apply the block least mean squares algorithm exploiting a limited memory of prior regression vectors and prior estimates of distance to the target in the time domain. Second, the agents estimate the combination weights based on negative exponentials of the scaled distance of intermediate location estimates within their neighborhood from the marginal median of these estimates in the spatial domain. The experimental learning curves demonstrate that, at the expense of some additional computational requirements, the enhanced adapt-then-combine (ATC) diffusion strategy can cope with the outliers and yield better results than the standard ATC strategy.

Index Terms— Adapt-Then-Combine, diffusion, marginal median, block least mean squares, strategic communication

1. INTRODUCTION

Adaptive networks [1] constitute a class of dynamical networks whose understanding has enabled researchers to decode several phenomena emerging in a broad range of applications not limited to the engineering domain, but extending to social science as well [2]. Nearly all real world networks are adaptive to some extent, because dynamical changes of their state and the underlying network topology take place, enabling self-organization.

Diffusion adaptation techniques incorporate the real-time cooperation of the nodes and the diffusion of information among them. These approaches have been applied to model complicated patterns of comportment detected in the biological networks, such as bird flight formations [3], trail formation of ants or the foraging behavior of fish schools [4], where fishes proceed in extraordinary consistency, speed similarity, ensuring a safe distance from their neighborhood in order to avoid collisions. Cooperation between agents can enhance network attitude due to a) the sharp quality distinction of information obtained by each agent; b) the inadequate information possessed by some agents rendering them unable to convalesce the ambiguity. On the other hand, link failures can discombobulate adaptation and the learning capabilities of the network, provoking limited or even deterrent cooperation between the agents. In either case, it has been proven that node cooperation outperforms the case when no cooperation takes place among network nodes [5]. In these instances, each node acquires estimates from its neighbourhood and finally employs a specific combination of them.

Diverse strategies of cooperation have been developed. For example, *diffusion techniques*, which are single time-scale implementations, such as the Adapt-Then-Combine (ATC) techniques, where information exchange takes place followed by aggregation [4] and the Combine-Then-Adapt (CTA) techniques, where aggregation precedes the information exchange [6]. A unifying general diffusion model, whose special cases are ATC and CTA was also developed [7]. Within the context of ATC, a distributed adaptive algorithm for sparsity-aware learning was proposed in [8]. Alternatively, *consensus techniques* can be used where all agents shall attain identical consensus state that ensures equipoise [9], [10], [11]. The initial consensus implementations were based on the adoption of two time-scales [12], [13], one for measurements assortment and one for the equilibrium fulfillment, preventing, however, the real-time adaptation process. Latter approaches were single time-scale [14], [15] and were based on distributed optimization [9].

Here, we are interested in *strategic communication* in which agents must decide what to say to others and how to react to what others say to them [2]. It has been found that communication expands the behavior of agents yielding new and potentially productive forms of interaction to prevail. For example, evolving automata were used to model endogenous, strategic communication in [16]. By allowing the agents to exchange communication tokens prior to playing a single-shot Prisoner's Dilemma game, where defection was the dominant strategy, occasional outbreaks of cooperation between agents had emerged. By applying the identical framework to a game of coordination (e.g., the Stag Hunt game), communication had proven beneficial as well [17]. In the aforementioned studies, the critical parameter of interest in models of strategic communication is the amount of processing power of the agents.

Motivated by the seminal work on diffusion networks by Ali H. Sayed and his colleagues, who have demonstrated that diffusion networks are more stable, converge faster and demonstrate lower mean-square deviation than consensus ones [6, 18], we build on the work on mobile adaptive networks [4], aiming at shedding light on their performance within a framework of strategic communication. In particular, we assume that a randomly chosen subset of agents share deliberately very noisy information with their neighbors. We allow the agents (i) to possess a limited memory of prior regression vectors and prior estimates of distance to the target in the time domain so that a block least mean squares (LMS) is applied and (ii) to estimate the combination weights based on negative exponentials of the scaled distance between the intermediate location estimates within their neighborhood and their marginal median in the spatial domain. In summary, the contributions of the paper are threefold: 1) The development of three novel diffusion algorithms for mobile adaptive networks, employing either block LMS, or the aforementioned combination weights, or the combination of both techniques; 2) The detailed demonstration of the proposed algorithms merits against the state-of-the-art algorithm [4]; 3) The study of the impact of various parameters in the performance of the proposed algorithms.

The remainder of this paper is structured as follows: An overview of the state-of-the-art diffusion algorithm for a mobile adaptive network, proposed in [4], is presented in Section 2. The developed algorithms are detailed in Section 3. Section 4 discloses experimental evidence for the advantages of the proposed algorithms against the standard ATC diffusion strategy. Finally, Section 5 concludes the paper and indicates future research directions.

2. MOBILE ADAPTIVE NETWORKS

Let N be the number of agents in a network. The k-th agent, represented as a node of a mobile network, updates its location from $\mathbf{x}_{k,i}$ to $\mathbf{x}_{k,i+1}$ using

$$\mathbf{x}_{k,i+1} = \mathbf{x}_{k,i} + \Delta t \cdot \mathbf{v}_{k,i+1} \tag{1}$$

where Δt is a time step and $\mathbf{v}_{k,i+1}$ is its updated velocity. The velocity of the k-th agent $\mathbf{v}_{k,i+1}$ should ensure that: 1) the agent moves to locations of higher SNR; 2) the agent moves towards the target $\mathbf{w}^0$; 3) the agent moves in coordination with its neighbors; 4) collisions between agents are avoided. The first objective is fulfilled if the network of agents moves in a direction that reduces $\sum_{k=1}^{N} \sigma_k^2(i)$, where $\sigma_k^2(i)$ is the noise variance estimated by the k-th agent at its location at time i. This can be achieved by means of [4]:

$$\mathbf{v}^{'}_{k,i+1} = \mathbf{g}_{k,i} = -\sum_{l \in \mathcal{N}_{k,i} \setminus \{k\}} \left(\sigma_l^2(i) - \sigma_k^2(i)\right) \frac{\mathbf{x}_{l,i} - \mathbf{x}_{k,i}}{\|\mathbf{x}_{l,i} - \mathbf{x}_{k,i}\|} \tag{2}$$

where $\mathcal{N}_{k,i}$ is the neighborhood of node k at time i and $\|\cdot\|$ denotes the ℓ_2 norm. To meet the second objective, the location of the center of mass of the network at time i defined as $\mathbf{x}_i^g = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_{k,i}$ should converge toward $\mathbf{w}^0$. This can be achieved if $\mathbf{v}_{k,i+1}$ points to the direction $\mathbf{w}^0 - \mathbf{x}_{k,i}$, [4], i.e.,

$$\mathbf{v}^{''}_{k,i+1} = h(\mathbf{w}^0 - \mathbf{x}_{k,i}) = \begin{cases} \mathbf{w}^0 - \mathbf{x}_{k,i} & \text{if } \left\|\mathbf{w}^0 - \mathbf{x}_{k,i}\right\| \leq s \\ s\,\frac{\mathbf{w}^0 - \mathbf{x}_{k,i}}{\left\|\mathbf{w}^0 - \mathbf{x}_{k,i}\right\|} & \text{otherwise} \end{cases} \tag{3}$$

where the positive scaling factor s bounds the speed in pursuing the target. Since $\mathbf{w}^0$ is unknown, one replaces $\mathbf{w}^0$ by a proper estimate $\mathbf{w}_{k,i}$. Let $\boldsymbol{\delta}_{k,i} = \sum_{l \in \mathcal{N}_{k,i} \setminus \{k\}} \left(\|\mathbf{x}_{l,i} - \mathbf{x}_{k,i}\| - r\right) \frac{\mathbf{x}_{l,i} - \mathbf{x}_{k,i}}{\left\|\mathbf{x}_{l,i} - \mathbf{x}_{k,i}\right\|}$, where r is the safe distance to be respected by each agent from its neighbors to avoid collisions. The third and the fourth objectives, i.e., the coordinated movement of the mobile network and the avoidance of agents collisions is achieved by [4]:

$$\mathbf{v}^{'''}_{k,i+1} = \mathbf{v}^g_{k,i} + \frac{1}{N}\,\boldsymbol{\delta}_{k,i} \tag{4}$$

where $\mathbf{v}^g_{k,i}$ is an appropriate local estimate of the velocity of the center of mass of the network $\mathbf{v}_i^g = \frac{1}{N}\sum_{k=1}^{N} \mathbf{v}_{k,i}$. By combining (2)-(4), the following update of the velocity vector is proposed in [4]:

$$\mathbf{v}_{k,i+1} = \lambda\, h(\mathbf{w}_{k,i} - \mathbf{x}_{k,i}) + \alpha \frac{\mathbf{g}_{k,i}}{\left\|\mathbf{g}_{k,i}\right\|} + \beta\, \mathbf{v}^g_{k,i} + \gamma\, \boldsymbol{\delta}_{k,i} \tag{5}$$

for non-negative weighting factors λ, α, β, and γ. To prevent singularities, $\frac{\mathbf{x}}{\|\mathbf{x}\|} \triangleq \mathbf{0}$ whenever $\mathbf{x} = \mathbf{0}$.

In (5), distributed local estimates of the target location $\mathbf{w}_{k,i}$ and the velocity of the center of the gravity $\mathbf{v}^g_{k,i}$ need to be specified. Let us suppose that at every time instant i, every node has access to a scalar measurement $d_k(i)$ from a random process $\underline{d}_k(i)$ and a (column) regression vector $\mathbf{u}_{k,i}$ of size $M = 2$ from another random process $\underline{\mathbf{u}}_{k,i}$ related via $\underline{d}_k(i) = \underline{\mathbf{u}}_{k,i}^T \mathbf{w}^0 + \underline{n}_k(i)$, where $\underline{n}_k(i)$ is a zero mean white random process independent of all other processes. The regression vector $\mathbf{u}_{k,i} = (\cos\theta_k(i), \sin\theta_k(i))^T$ is a unit direction vector employing the azimuth angle of the line connecting $\mathbf{x}_{k,i}$ with $\mathbf{w}^0$. Let $\mathbb{E}\{\cdot\}$ denote the expectation operator. A distributed adaptive estimate of $\mathbf{w}^0$, which minimizes $J^{glob}(\mathbf{w}) = \sum_{k=1}^{N} \mathbb{E}\{|\underline{d}_k(i) - \underline{\mathbf{u}}_{k,i}^T \mathbf{w}|^2\}$ can be obtained by the ATC diffusion algorithm [7] under certain assumptions set in [4]:

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_{k,i}} c^w_{l,k} \left(\mathbf{q}_{l,i} - \mathbf{w}_{k,i-1}\right) \tag{6}$$

$$\mathbf{w}_{k,i} = \sum_{l \in \mathcal{N}_{k,i}} a^w_{l,k}\, \boldsymbol{\psi}_{l,i} \tag{7}$$

where the weights $a^w_{l,k}$ and $c^w_{l,k}$ satisfy the properties

$$\sum_{l=1}^{N} a^w_{l,k} = \sum_{l=1}^{N} c^w_{l,k} = 1, \quad c^w_{l,k} = a^w_{l,k} = 0 \text{ if } l \notin \mathcal{N}_k \tag{8}$$

and

$$\mathbf{q}_{l,i} = \mathbf{x}_{l,i} + \mathbf{u}_{l,i}\, d_l(i). \tag{9}$$

Similarly, seeking for an adaptive distributed estimate of $\mathbf{v}^g$, which minimizes the cost function $J^{glob}(\mathbf{v}^g) = \sum_{k=1}^{N} \mathbb{E}\left\{\|\mathbf{v}_{k,i} - \mathbf{v}^g\|^2\right\}$, one arrives at the ATC diffusion algorithm:

$$\boldsymbol{\phi}_{k,i} = \mathbf{v}^g_{k,i-1} + \nu_k \sum_{l \in \mathcal{N}_{k,i}} c^v_{l,k} \left(\mathbf{v}_{l,i} - \mathbf{v}^g_{k,i-1}\right) \tag{10}$$

$$\mathbf{v}^g_{k,i} = \sum_{l \in \mathcal{N}_{k,i}} a^v_{l,k}\, \boldsymbol{\phi}_{l,i} \tag{11}$$

where $c^v_{l,k}$ and $a^v_{l,k}$ satisfy (8) and every node k is assumed to have access to the local data $\mathbf{v}_{l,i}$ for $l \in \mathcal{N}_{k,i}$. The performance of the aforementioned diffusion strategy has been thoroughly analyzed in [7].

3. PROPOSED METHODS

In (7), $|\mathcal{N}_{k,i}|$ intermediate location estimates $\boldsymbol{\psi}_{l,i}$ are combined. Let P_k be the length of a buffer monitoring $\boldsymbol{\psi}_{l,i}$, $l \in \mathcal{N}_{k,i}$, where $P_k \leq |\mathcal{N}_{k,i}|$. As a reference intermediate location estimate, we propose to use the marginal median $\boldsymbol{\psi}^{MM}_{k,i}$ of the aforementioned P_k intermediate location estimates. That is, the vector-valued observations are ordered along each of the $M = 2$ channels independently, and the median of the element-wise observations is found [19]. Let $\zeta_{l,k} = \left\|\boldsymbol{\psi}_{l,i} - \boldsymbol{\psi}^{MM}_{k,i}\right\|$. Then, the following combination weights are used:

$$a^w_{l,k} = \rho \exp\left(-\frac{\zeta_{l,k}}{2\,\sigma_a}\right) \tag{12}$$

where σ_a is the kernel size and ρ is a normalization parameter, guaranteeing that (8) is satisfied. The motivation behind using (12) is to exploit the robust properties of marginal median as a location estimator of multivariate observations within strategic communication, where the intermediate location estimates may be heavily corrupted by noise. In the following, we refer to this technique as ATC-mmed.

Instead of only using the regression vector $\underline{\mathbf{u}}_{k,i}$ at time instant i, one may employ the regression vectors $\underline{\mathbf{u}}_{k,j}$ at time instants $j =$

$i, i-1, \ldots, i-L+1$. By doing so, an $M \times L$ random matrix $\underline{\mathbf{U}}_{k,i} = \left[\underline{\mathbf{u}}_{k,i} | \underline{\mathbf{u}}_{k,i-1} |, \ldots, | \underline{\mathbf{u}}_{k,i-L+1}\right]$ results. Similarly, each agent may keep track of the L most recent estimates of distance $\underline{d}_k(j)$ in the random vector $\underline{\mathbf{d}}_{k,i}$. The relation between $\underline{\mathbf{d}}_{k,i}$ and $\underline{\mathbf{U}}_{k,i}$ is $\underline{\mathbf{d}}_{k,i} = \underline{\mathbf{U}}_{k,i}^T \mathbf{w}^0 + \underline{\mathbf{n}}_{k,i}$, where $\underline{\mathbf{n}}_{k,i}$ is a zero mean white random noise vector. By applying the first order optimality condition to $\mathbb{E}\left\{\left\|\underline{\mathbf{d}}_{k,i} - \underline{\mathbf{U}}_{k,i}^T \mathbf{w}^0\right\|^2\right\}$ with respect to (w.r.t.) $\mathbf{w}^0$, the following solution is obtained:

$$\mathbf{w}^0 = \mathbb{E}\left\{\underline{\mathbf{U}}_{k,i} \underline{\mathbf{U}}_{k,i}^T\right\}^{-1} \mathbb{E}\left\{\underline{\mathbf{U}}_{k,i} \underline{\mathbf{d}}_{k,i}\right\}. \quad (13)$$

In (13), the covariance matrix $\mathbf{R}_{u,k} = \mathbb{E}\left\{\underline{\mathbf{U}}_{k,i} \underline{\mathbf{U}}_{k,i}^T\right\}$ is employed. By dropping the expectation operator, an estimate of $\mathbf{R}_{u,k}$ results, i.e., $\mathbf{U}_{k,i} \mathbf{U}_{k,i}^T = \sum_{j=i}^{i-L+1} \mathbf{u}_{k,j} \mathbf{u}_{k,j}^T$. Following similar lines to [7], the intermediate estimates, that arise from local adaptation, are obtained by

$$\boldsymbol{\psi}_{k,i} = \mathbf{w}_{k,i-1} + \mu_k \sum_{l \in \mathcal{N}_{k,i}} c_{lk}^w \, \mathbf{U}_{k,i} \left[\mathbf{d}_{k,i} - \mathbf{U}_{k,i}^T \mathbf{w}_{k,i-1}\right]. \quad (14)$$

The distributed estimate $\mathbf{w}_{k,i}$ of $\mathbf{w}^0$ is obtained through (7). The proposed technique, coined as ATC-block, incorporates the same steps with the ATC diffusion algorithm for mobile networks with the only differentiation being focused on the intermediate estimates $\boldsymbol{\psi}_{k,i}$, which are estimated taking into account the last L pairs of $\mathbf{u}_{k,i-j}$ and $d_k(i-j)$, with $j = 0, 1, \ldots, L-1$.

The third technique, abbreviated as ATC-block-mmed, combines the aforementioned methods. Firstly, by applying the block LMS, the intermediate estimates $\boldsymbol{\psi}_{k,i}$ are determined according to (14). Secondly, the combination coefficients $a_{l,k}^w$ are estimated by employing (12), which resorts to the marginal median estimate.

4. EXPERIMENTAL EVALUATION

The proposed algorithms and the ATC diffusion algorithm for mobile networks were implemented and tested using the setting in [4]. The number of agents in the network is $N = 100$. The target location is set at (100, 100). The initial estimation of target location is (0,0) for all agents, their initial velocity is zero, and the velocity estimation of the center of the mass is set initially zero. The initial locations of the $N = 100$ agents are uniformly distributed over two square regions whose coordinates are $[20-30, 0-10]$ and $[0-10, 20-30]$. The learning step sizes μ_k and ν_k are set to 0.5 $\forall\, k$. The combination coefficients are $a_{l,k}^w = \frac{1}{|\mathcal{N}_k|}$ if $l \in \mathcal{N}_k$, while $a_{l,k}^v = c_{l,k}^w = c_{l,k}^v = \delta_{lk}$, where δ_{lk} denotes the Kronecker delta. The impact of noise level is bypassed by setting the weighting factor $\alpha = 0$. Regarding velocity control, the coefficients are $\lambda = 0.8, \beta = 1-\lambda = 0.2, \gamma = 0.1$. The safe distance r among neighbours is equal to 1. The scaling factor is $s = 1$, the time step is $\Delta t = 0.5s = 0.5$, the maximum distance among neighbors is considered to be $R = 10$, while the maximum buffer size of regression vectors is $P_k = 11$. The total number of time instants is $I = 250$.

Two modes of operation are studied. In the first mode, there are no outliers. The local distance estimates $d_k(i)$ are contaminated with zero-mean Gaussian noise with standard deviation equal to 20, while the elements of $\mathbf{u}_{k,i}$ are corrupted with noise independently drawn from a zero-mean Gaussian distribution having standard deviation equal to 10. In the second mode, a case of strategic communication is simulated. The repercussion of outliers due to deliberate agent decisions or faulty processes of agents is studied. In particular,

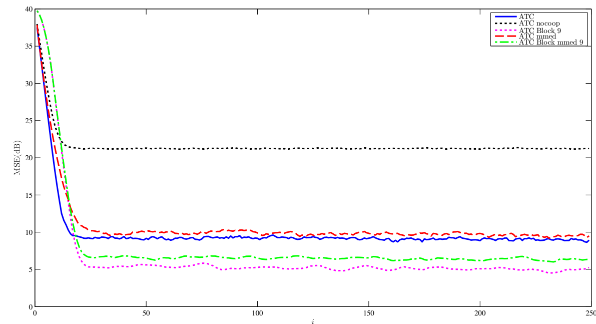


Fig. 1: Ensemble average MSE of the various ATC variants versus iterations in the presence of nominal errors.

we would like to see how robust the network is to the propagation of false information. Toward this goal, let q be the percentage of randomly selected agents of the network that contaminate the local data $\{d_k(i), \mathbf{u}_{k,i}\}$ by adding on the top of nominal errors, occurring in the first mode of operation, zero-mean gross errors following the Gaussian distribution with standard deviation equal to 100.

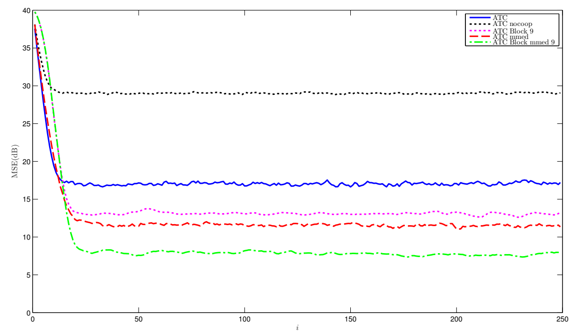
When there is not any cooperation between the nodes of the network $a_{l,k}^w = a_{l,k}^v = \delta_{lk}$ and $v_{k,i}^g = \frac{1}{|\mathcal{N}_{k,i}|} \sum_{l \in \mathcal{N}_{k,i}} v_{l,i}$ is the average velocity of the neighborhood.

The quality of the diffusion algorithms was evaluated w.r.t. the learning curve, i.e., the ensemble averaged network mean-squared-error (MSE) for estimating the target location versus iterations. To assess the ATC algorithm and the proposed techniques, 100 Monte Carlo simulations of all algorithms were run using: a) different initial agent location; b) different randomly chosen agents which deliberately share local information contaminated by outliers; and c) different realizations of nominal and gross errors added to the local data $\{d_k(i), \mathbf{u}_{k,i}\}$. Then, the ensemble average of the MSE for estimating the target location over the 100 Monte Carlo independent trials of the experiment gives an approximation of the ensemble-averaged learning curve of the network.

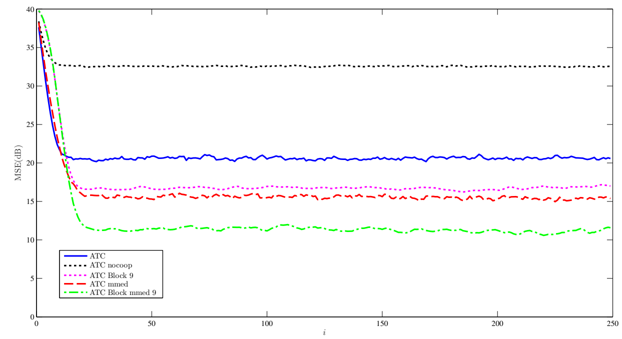
The ensemble average of the MSE versus iterations for the various algorithms is plotted in Figure 1, where it is attested that the ATC-block algorithm outperforms all the others. The kernel size was $\sigma_a = 1$, with the number of the past data being $L = 9$ in block LMS. In this case, the use of combination weights, which resort to marginal median, deteriorates slightly the performance of the network. This is confirmed by the comparison between a) ATC and ATC-mmed, b) ATC-block and ATC-block-mmed algorithms. However, both the ATC-block and ATC-block-mmed algorithms yield a better performance, reducing the network MSE. Absence of cooperation between agents deteriorates the network performance significantly.

When a percentage of agents deliberately share local information corrupted with gross errors on the top of nominal errors, the ensemble average MSE of ATC, ATC-mmed, ATC-block, and ATC-block-mmed algorithms versus iterations is overlaid in Figures 2a and 2b for $q = 20\%$ and $q = 50\%$, respectively. In both cases, the kernel size was $\sigma_a = 1$, while the number of the past data was $L = 9$ in block LMS. It is crystal clear that all the proposed techniques improve the performance of the network, with ATC-block-mmed algorithm outperforming all the competing techniques. The ATC-mmed and ATC-block algorithms yield a comparable performance, although ATC-mmed exhibits a smaller MSE than ATC-block.

Taking into account that the ATC-block-mmed algorithm was proven to be the most efficient algorithm in the framework of strate-

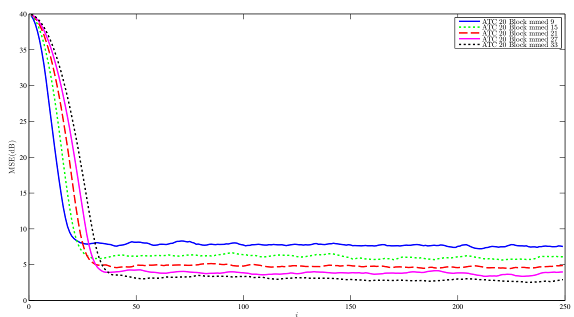


(a) $q = 20\%$

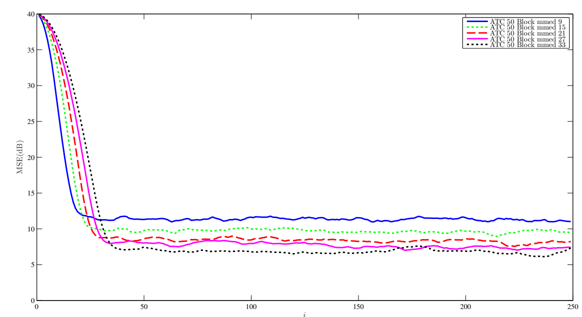


(b) $q = 50\%$

Fig. 2: Ensemble average MSE of the various ATC variants versus iterations, when a percentage q of agents deliberately share local information corrupted with gross errors.



(a) $q = 20\%$



(b) $q = 50\%$

Fig. 3: Ensemble average MSE of ATC-block-mmed versus iterations when a percentage q of agents deliberately share local information corrupted with gross errors for $L = 9, 15, 21, 27, 33$ in block LMS.

gic communication, we study next the influence of the length of past data L employed in the block LMS. The ensemble average MSE of ATC-block-mmed versus iterations for $L = 9, 15, 21, 27, 33$ and $q = 20\%$ or $q = 50\%$ is plotted in Figure 3. It can be seen that the adoption of more past data enhances the network performance.

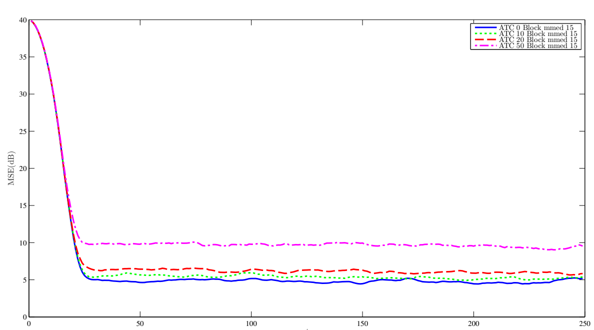


Fig. 4: ATC-block-mmed MSE for $q = 0, 10, 20, 50\%$.

Next, we fix $L = 15$ and we study the influence of the percentage q of agents, sharing local information corrupted with gross errors, in the performance of the ATC-block-mmed. The ensemble average MSE of ATC-block-mmed versus iterations is shown in Figure 4 for $q = 0, 10, 20, 50\%$. The kernel size was $\sigma_a = 1$. It is self-evident that the performance of the network deteriorates when q increases, but still the MSE level is acceptable.

Finally, it has been proven that the use of $a^v_{l,k} = \frac{1}{|\mathcal{N}_k|}$, instead of $a^v_{l,k} = \delta_{lk}$, improves the performance of the ATC diffusion algorithm and its proposed variants.

5. CONCLUSIONS AND FUTURE WORK

Three variants of the ATC diffusion algorithm for mobile adaptive networks have been proposed and tested in the context of strategic communication. The use of combination weights, which are negative exponentials of the distance between the intermediate location estimates and their marginal median, and the replacement of LMS with block LMS yield a better performance than the standard ATC diffusion algorithm, when a percentage of agents share local information corrupted with gross errors. When gross errors are not present, the best performance is accomplished with the ATC-block algorithm. The employment of more past data improves the performance of both ATC-block and ATC-block-mmed algorithms. The same applies when P_k increases in the computation of the marginal median. The aforementioned results confirm the importance of a) proper weighting and b) suitable derivation of intermediate estimates in diffusion strategies and complement the discussions made in [7]. The promising experimental results justify the investment of effort toward analyzing theoretically the performance of these algorithms.
Acknowledgment. The authors acknowledge the preliminary work conducted by Mr. Theodoros Theodoridis.

6. REFERENCES

[1] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Found. Trends Mach. Learn.*, vol. 7, no. 4-5, pp. 311–801, Jul 2014.

[2] J. H. Miller and S. E. Page, *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, Princeton University Press, 2007.

[3] F. S. Cattivelli and A. H. Sayed, "Modeling bird flight formations using diffusion adaptation," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2038–2051, May 2011.

[4] S.-Y. Tu and A. H. Sayed, "Mobile adaptive networks," *IEEE Journal Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 649–664, Aug 2011.

[5] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1035–1048, March 2010.

[6] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, "Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.

[7] A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, vol. 102, no. 4, pp. 460–497, April 2014.

[8] S. Chouvardas, G. Mileounis, N. Kalouptsidis, and S. Theodoridis, "A greedy sparsity-promoting LMS for distributed adaptive learning in diffusion networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, May 2013, pp. 5415–5419.

[9] J. N. Tsitsiklis, J. N. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Automatic Control*, vol. 31, no. 9, pp. 803–812, Sep 1986.

[10] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automatic Control*, vol. 48, no. 6, pp. 988–1001, June 2003.

[11] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Automatic Control*, vol. 50, no. 5, pp. 655–661, May 2005.

[12] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 26–35, May 2007.

[13] B. Johansson, T. Keviczky, M. Johansson, and K. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. IEEE Conf. Decision and Control*, Dec 2008, pp. 4185–4190.

[14] I. Schizas, G. Mateos, and G. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2365–2382, Jun 2009.

[15] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.

[16] J. H. Miller, C. Butts, and D. Rode, "Communication and cooperation," *Journal of Economic Behavior and Organization*, vol. 47, no. 2, pp. 179–195, Feb 2002.

[17] J. H. Miller and S. Moser, "Communication and coordination," *Complexity*, vol. 9, no. 5, pp. 31–40, May 2004.

[18] S.-Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6217–6234, Dec 2012.

[19] C. Kotropoulos and I. Pitas, "Multichannel L filters based on marginal data ordering," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2581–2595, Oct 1994.