Distributed Field Estimation Algorithms in Vehicular Sensor Networks

Dietmar Schabus, Thomas Zemen, Michael Pucher
Telecommunications Research Center Vienna (FTW), Vienna, Austria
E-Mail: {schabus, pucher, zemen}@ftw.at

***Abstract*—This paper deals with cooperative reconstruction of environmental variables (e.g., temperature) along a road by a vehicular sensor network using wireless communication. Vehicles take repeated measurements and approximate the environment using a set of basis functions. We investigate the applicability and performance of popular averaging techniques (gossiping and consensus propagation) on the basis coefficients, and propose a simpler approach to avoid divergence problems. We have developed a graphical simulation environment to study the behavior of different algorithms in this scenario and we show simulation results which support our simplified approach.**

***Index Terms*—wireless sensor network, mobile sensors, gossip algorithm, consensus propagation**

I. Introduction

Considering recent developments in the field of wireless sensor networks and cooperative communication systems for traffic telematics, this paper addresses the problem of cooperatively measuring environmental variables in a network of vehicles. Variables of interest are air and road temperature, precipitation, road condition, etc. — anything for which in-vehicle sensors do already exist or are likely to be introduced in the near future. Through the exchange of information among vehicles in ad-hoc networks, we aim at spreading knowledge about the environment while reducing the effects of noisy sensor measurements and at the same time keeping the communication overhead low.

Recent works have studied algorithms using the inherent broadcast nature of the wireless medium based on gossiping [1] and consensus propagation [2], [3], in the latter case even for scenarios with time-varying spatial fields [4]. Now it might seem at first thought that reconstructing a (static or slowly changing) field using a network of mobile sensors might not pose different challenges than reconstructing a time-varying field with a fixed network. This is not the case, however, mainly due to the constant change of the network adjacencies.

Furthermore, nodes in a vehicular network sometimes are within communication range for only a very short time due to high cruising speeds, and hence the chance for exchanging information should not be missed even if it is brief.

The scientific contributions of this paper are as follows: We identify the problems arising from sensor mobility when trying to apply existing sensor network algorithms to the vehicular scenario (Section III). Then we describe a very simple approach to the scenario which avoids these problems (Section IV). In particular, our approach has the following advantages:

- No knowledge about the network topology is required.
- It is robust to frequent topology changes.
- The quality of the results degrades gracefully under adverse conditions.
- The overall communication costs are kept low.

For the sake of comparison, we also describe versions of our algorithm which correspond to broadcast gossiping and consensus propagation. To study and evaluate different communication algorithms, we have developed a graphical simulation environment, which we describe in Section V. Finally, we evaluate our approach in comparison to the other two under different conditions in our simulation environment (Section VI). The results are in favor of our approach, and show that it takes advantage of high traffic density particularly well. The now following Section II introduces how we model the scenario we consider, along with the assumptions made. We also introduce some notation for the rest of the paper.

II. Models for Traffic, Environment and Communication

We consider a two-dimensional field with a smoothly varying environmental parameter (e.g., air termperature) with a single road of length L running through the field. Since vehicles can only measure the environment along the road (and this is also what we are interested in), it is sufficient to model location with one-dimensional real values $x \in [0, L)$. We assume each vehicle to know its own location at any time (via satellite, maps and, e.g., inertial sensors). Any vehicle a can measure the environmental parameter $e(x)$ at position x using its on-board sensor, which is affected by additive Gaussian sensor noise $n_a(x)$ with standard deviation σ_n, resulting in noisy measurements $\hat{e}_a(x)$:

$$\hat{e}_a(x) = e(x) + n_a(x), \tag{1}$$

$$n_a(x) \sim \mathcal{N}(0, {\sigma_n}^2). \tag{2}$$

Vehicles enter the road section in question at both end points (positions 0 and L) and make their way through the field to the other endpoint, where they leave the scenario. Fig. 2 illustrates the concept of repeated noisy measurements of an

This research was supported by the FTW project COCOMINT funded by Vienna Science and Technology Fund (WWTF). The Telecommunications Research Center Vienna (FTW) is supported by the Austrian Government and the City of Vienna within the competence center program COMET.

environmental parameter at changing locations. We assume the wireless communication between vehicles to be bounded by some communication radius R in the Euclidean plane.

III. Issues Arising from Mobility

The most important consequence of mobility in the scenario we are considering is that vehicles collect measurements from different locations as time passes and they move through the scene. Hence they need to keep track of the location of their measurements, and exchanged messages also need to contain location information. This also means, however, that as time passes each vehicle accumulates more and more knowledge about the environment from its own measurements alone.

A second important consequence is that the network topology constantly changes. While some vehicles might be able to communicate over a long period of time because they travel in the same direction with similar speeds, vehicles traveling in opposite directions will have only a short time window to exchange messages. The information provided by those vehicles is especially useful, however, as it typically concerns areas of the environment the other vehicle has not yet seen.

Due to these reasons, successful algorithms for static networks like broadcast gossiping [1] and consensus propagation [2] cannot be applied in a straightforward way to the mobile scenario: In the case of consensus propagation the algorithm makes direct use of the set of neighbors in excluding the message received from a node in the previous iteration:

$$K_{i\to j}^{(n)} = \frac{1+\sum_{u\in\mathcal{N}(i)\backslash j} K_{u\to i}^{(n-1)}}{1+\frac{1}{\beta}(1+\sum_{u\in\mathcal{N}(i)\backslash j} K_{u\to i}^{(n-1)})} \quad (3)$$

$$\mu_{i\to j}^{(n)} = \frac{y_i+\sum_{u\in\mathcal{N}(i)\backslash j} \mu_{u\to i}^{(n-1)}\, K_{u\to i}^{(n-1)}}{1+\sum_{u\in\mathcal{N}(i)\backslash j} K_{u\to i}^{(n-1)}}, \quad (4)$$

where i and j are the source and destination nodes, respectively, $\mathcal{N}(i)$ is the set of neighbors of i, y_i is the measurement taken by node i, and $\beta > 0$ is an attenuation parameter. Nodes exchange value pairs (μ, K), where μ and K can be interpreted as iterative estimates of the average and cardinality, respectively.

This exclusion of j is also performed in the broadcast-adapted version of consensus propagation [3], where the respective values are subtracted in the receiving node as a post-processing step. However, with the topology changing between iterations, assumptions about known past messages might not hold – neither at the sending nor the receiving node.

Broadcast gossiping does not make explicit use of knowledge about neighbors, however the time model used for convergence analysis makes the implicit assumption that each node receives equally many messages from all its neighbors, which does not hold in the mobile scenario. The impact of mobility on gossiping has been studied before [5], but without considering multiple measurements at different times and locations. Also, consensus among moving agents (and hence switching topology) has been studied before [6], however under the assumption of a fixed set of nodes in a strongly connected and balanced graph, where edges are added or removed as the agents move. This is very different from the scenario considered here, where new nodes (with little knowledge) enter the network, and older nodes (with much knowledge) leave the network, and connections are sparse.

IV. Proposed Approach

To deal with the aforementioned issues arising from sensor mobility, we propose the following approach to the vehicular scenario. As vehicles move, they repeatedly take measurements and store them along with the current location. The road of length L is divided into sections $S_1, S_2, ..., S_{\lceil L/K\rceil}$ of length K where the section borders are globally defined and known to all vehicles. Whenever a vehicle a completes collecting measurements for a section S_i, it compresses the information by approximating the environment in the section as a linear combination of basis functions $u_1, ..., u_D$, i.e.,

$$e(i,x) \approx \sum_{d=1}^{D} c_{i,d}^{(a)}\, u_d(x), \quad (5)$$

with $1 \le i \le \lceil L/K\rceil$ and $1 \le x \le K$, resulting in D basis coefficients $c_{i,d}^{(a)}$ for each section that are determined by vehicle a. The section length K, the subspace dimensionality D and the basis functions themselves should be chosen appropriately for the environmental variable in question, such that approximating the noisy measurements already reduces the sensor noise in the resulting model, but still avoiding underfitting of the measurement data, i.e., minimizing the overall error considering the tradeoff between model bias and model variance [7].

At regular time intervals, vehicles broadcast the coefficients they have determined to other vehicles in range. Receiving vehicles collect such messages, and update their belief state on the corresponding sections to the average of all available approximations, thus further reducing the effects of sensor noise (and also sensor bias). More formally, let $V_{a,i} = \{b, c, d, ...\}$ denote the set of vehicles from which a message concerning section S_i has been received by vehicle a. $V_{a,i}$ may or may not contain a itself, depending on whether or not a has already traveled through S_i. Vehicle a can then compute the average coefficients

$$\hat{c}_{i,d}^{(a)} = \frac{1}{|V_{a,i}|} \sum_{v\,\in\, V_{a,i}} c_{i,d}^{(v)}\ , \quad (6)$$

$1 \le d \le D$, and use these to reconstruct the environment in section S_i as in (5).

Note that the coefficients broadcasted by each vehicle are still the ones based on its own measurements only, unaffected by incoming messages. Only when exchanging data, e.g., with a road-side unit of the road operator, vehicles provide their best estimate of the environment in the form of the averaged coefficients. This will of course lead to duplicate messages being received, but they can be easily discarded based on vehicle and section identifiers (which of course need to be

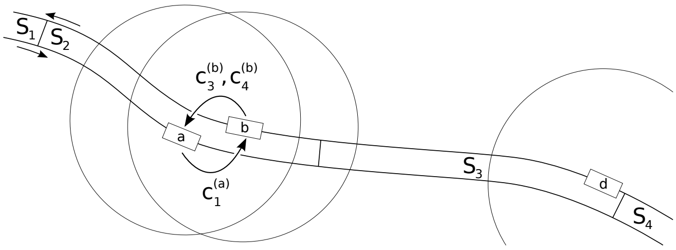

Fig. 1. Illustration of the proposed approach. Vehicle a provides vehicle b with its coefficient vector for section S_1, and vehicle b provides vehicle a with its coefficient vectors for sections S_3 and S_4. Both a and b are currently collecting measurements for section S_2. Vehicle d has just completed collecting measurements for section S_4 and is now calculating the corresponding coefficients. It will then discard its measurements and will later provide its $c_4^{(d)}$ to a, when they are within communication range.

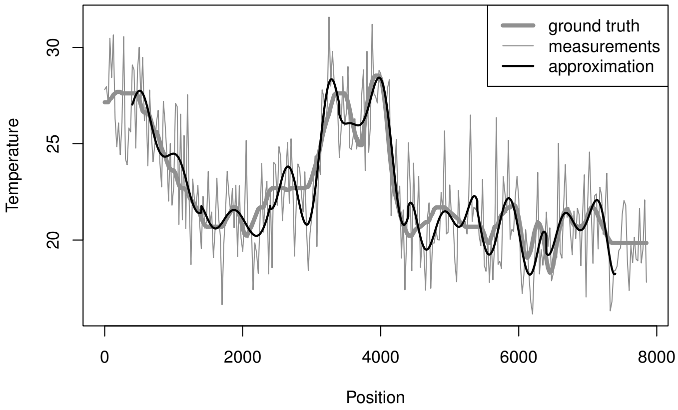

Fig. 2. An exemplary environmental parameter $e(x)$ (thick gray line), one vehicle's noisy measurements $\hat{e}(x)$ (thin gray line) and that vehicles approximation (back line).

part of the transmitted messages). The protocol is illustrated in Fig. 1. One more thing to note is that vehicles need to eventually stop sending out information when it is outdated, e.g., broadcast coefficients only for the last n sections they traveled through.

Keeping the protocol this simple has the following advantages:

- No knowledge about the network topology is required.
- The protocol is robust to (frequent) topology changes. In particular, receiving many messages from a vehicle b (traveling in the same direction with similar speed) but only very few messages from a vehicle c (traveling in the opposite direction) does not result in a bias towards b's model in the receiving vehicle.
- The quality of the results degrades gracefully under adverse conditions like message loss, communication delays and decreasing traffic density.
- Choosing an appropriate set of basis functions for the environmental parameter and application under consideration keeps the messages small in size, and as it is often sufficient that two given vehicles exchange information only once, the number of messages is also kept small. This results in overall low communication costs.

We will refer to the algorithm described so far as "proposed algorithm".

For the sake of comparison, we also consider a version of the protocol where vehicles always broadcast and update their current belief state (i.e., coefficients are averaged on each incoming message). This roughly corresponds to broadcast gossiping, with the difference that vehicles "wake up" at regular intervals rather than randomly (to ensure an equal amount of messages for a fair comparison), and that additional measurements enter the algorithm during runtime. We will refer to this version as "gossiping".

Furthermore, we also consider a version of consensus propagation, in which vehicle a broadcast the values

$$K_{a,i} = 1 + \sum_{v \in V_{a,i}} K_{v,i} \qquad (7)$$

$$\hat{c}_{i,d}^{(a)} = \frac{1}{K_{a,i}} \left(1 + \sum_{v \in V_{a,i}} \hat{c}_{i,d}^{(v)} \, K_{v,i} \right), \qquad (8)$$

$1 \leq d \leq D$ (compare (3) and (4)), which differs from original consensus propagation [2] in that we do not exclude the receiving node from the sums – not even in post-processing, as it is done in the broadcast-based version [3]. The reason for this is that in a constantly changing network on the one hand the sending node does not know which other nodes will receive its message, and on the other hand a receiving node cannot assume that the incoming message has been influenced by its own last outgoing message. This means that without introducing some additional handshake or acknowlegement functionality to the protocol, we have no other choice than summing over all neighbors. Note also that for the concept of "neighbors" we use the set $V_{a,i}$, which we earlier defined as the set of vehicles from which a has received any messages.

Also, we use the unattenuated variant of consensus propagation ($\beta = \infty$), as this has shown the best performance in our experiments (compare (3)). We will refer to this algorithm as "consensus propagation" in the following.

V. Simulation Environment

We have developed a graphical simulation environment[1] to experiment with different sensor network algorithms for the scenario we are interested in. In addition to the properties already stated for the scenario, the simulation environment makes the following simplifications and assumptions:

Vehicles enter the road at both endpoints according to a rate λ Poisson process and hence the inter-car times at the endpoints are exponentially distributed. The parameter λ can also be understood as specifying the (inverse) traffic density. We assume each vehicle a to travel at a constant

[1] Demonstration videos at http://userver.ftw.at/~schabus/vtc2011spring/

velocity v_a which is drawn from a Gaussian distribution, $v_a \sim \mathcal{N}(\mu_v, {\sigma_v}^2)$. So far, we do not use any concept of lanes or overtaking, therefore faster vehicles pass slower ones as if they were not there. As long as the Euclidean distance between two vehicles is smaller than the communication radius R, we assume error- and noise-free communication between them.

The exemplary environmental parameter profile we used is depicted in Fig. 2, where a thick gray line indicates the true environment $e(x)$, a thin gray line indicates one vehicle's noisy measurements $\hat{e}(x)$ and a black line indicates that vehicles approximation using basis functions, determined from its measurements. We take the road length to be roughly 7800m in length, and divide it into seven sections of 1km length, leaving 400m at the beginning and the end as padding. We assume vehicles can take a measurement every 25m, which requires a measuring frequency of $\leq$ 2Hz for speeds up to 180km/h. For the sake of smooth section connections, we add an overlap of 200m to both section ends. This results in 56 measurements per vehicle per section. We interpret the field values as air temperature varying between 18°C and 28°C. The parameters λ, μ_v, σ_v, R, σ_n (traffic density, mean and standard deviation of velocity, communication radius and standard deviation of sensor noise) can be adjusted at runtime taking immediate effect. Additionally, the environment allows on-line switching between different algorithms to study their behavior. A live performance charting component, which continuously updates the current measurement and approximation error graphs, has proven particularly useful.

VI. Simulation Results

For comparison of the three described algorithms, we have conducted experiments in our simulation environment with fixed parameters

$$\begin{aligned} \mu_v &= \text{130km/h} \\ \sigma_v &= \text{20km/h,} \\ R &= \text{300m} \end{aligned}$$

and varying

$$\begin{aligned} \lambda &\in \{10, 20, 30, ..., 200\}, \\ \sigma_n &\in \{0.00, 0.25, 0.50, ..., 5.00\}. \end{aligned}$$

As basis functions we used the first 10 Legendre polynomials, and the basis coefficients were computed from the noisy measurements in a least squares fashion. We have measured the mean squared error (MSE) per vehicle over 100 vehicles that traveled entirely through the scene, each vehicle reporting its estimates upon leaving the road. The results are shown in Fig. 3. As expected, the error increases with increasing sensor noise as well as with decreasing traffic density, and we can see that gossiping performs worse than the other two overall, which are rather similar. Interestingly, gossiping seems to be unable to take as much advantage of high traffic density (small λ) to cope with strong sensor noise as the other two.

For a more direct comparison, we show a selection of the same data in two-dimensional plots in Fig. 4 and 5. Fig. 4

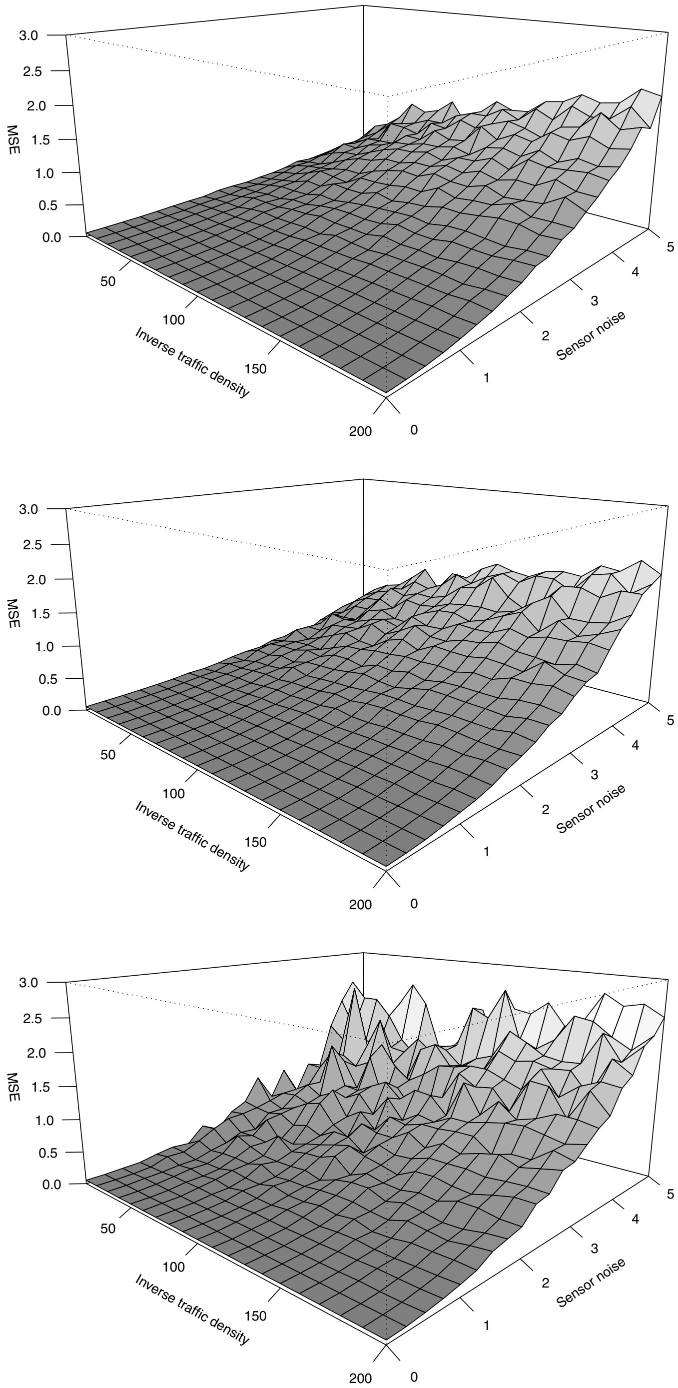

Fig. 3. Mean squared error of the proposed algorithm (top), consensus propagation (middle) and gossiping (bottom) over varying values of inverse traffic density (λ) and sensor noise (σ_n).

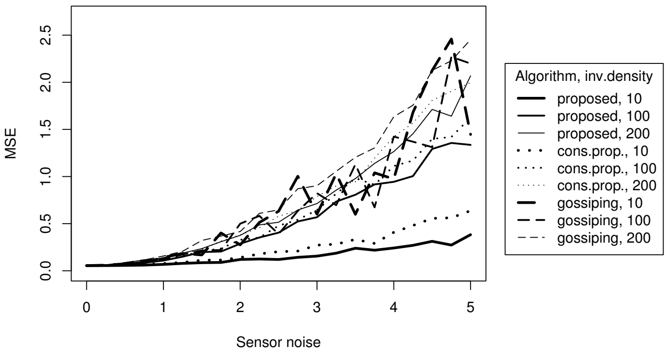

Fig. 4. Performance of the three algorithms for selected values of inverse traffic density.

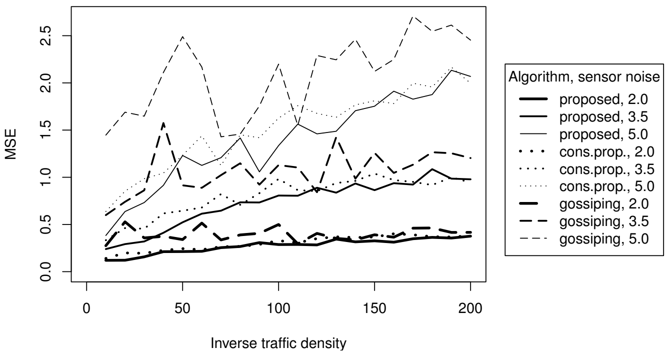

Fig. 5. Performance of the three algorithms for selected values of sensor noise.

shows the influence of sensor noise on three selected densities for all three algorithms. It reveals that the proposed algorithm (solid lines) outperforms consensus propagation (dotted lines), and that gossiping (dashed lines) does not degrade as gracefully and generally shows higher error. Fig. 5 shows the influence of (inverse) traffic density on three selected sensor noise levels for the three algorithms. We can see that the biggest difference between the algorithms is observed in high traffic density (on the left side of the graph), suggesting that the good performance of the proposed algorithm is due to its ability to take advantage of many incoming messages particularly well.

VII. Conclusion and Outlook

We have investigated cooperative distributed estimation of an environmental field by a network of sensor-equipped vehicles, which exchange messages in ad-hoc wireless networks in a broadcast fashion. Considering two successful protocols for static wireless sensor networks, namely broadcast gossiping and consensus propagation, we have identified the additional challenges and problems arising from sensor mobility. Especially the constant change of the network topology brings a quite drastic change to the situation. Therefore we have proposed a somewhat simpler approach to the problem, which does not make explicit or implicit assumptions about the node neighborhoods and is hence robust to changes in network topology.

The idea is that each vehicle computes an approximation of a section of the environment based on its (noisy) measurements from that section, using a linear combination of basis functions. Then the vehicle begins to broadcast its findings about the environment in the form of the basis coefficients, and receiving vehicles collect such messages. When the current best estimate of the environment is required, for example when the information is provided to the driver or to a road-side unit of the road operator, an approximation based on the average of all received coefficients for the section in question is used.

We have developed a simulation environment to experiment with different communication algorithms and to evaluate the proposed approach. Simulations with varying levels of sensor noise and traffic density have shown that it performs better than similar protocols roughly corresponding to broadcast gossiping and consensus propagation. Its superiority is particularly well visible under high sensor noise (an adverse condition) and high traffic density (a favorable condition), suggesting that the proposed approach takes good advantage of the high number of messages received in dense traffic.

As future work, we want to consider fields with various properties and matching basis functions, in particular discrete prolate spheroidal sequences (also known as Slepian sequences) [8], which are of high interest because they would eliminate the need for matrix inversion in the least squares step of determining the coefficients. We would also like to develop our models and our simulation environment in the direction of more and more realism, i.e., use a more realistic movement and communication model, allow the field to vary over time, etc., and study the behavior of cooperative estimation techniques under these more realistic conditions.

References

[1] T.C. Aysal, M.E. Yildiz, A.D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *Trans. Sig. Proc.*, vol. 57, no. 7, pp. 2748–2761, 2009.

[2] C.C. Moallemi and B. Van Roy, "Consensus propagation," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 4753 –4766, Nov. 2006.

[3] V. Schwarz, C. Novak, and G. Matz, "Broadcast-based dynamic consensus propagation in wireless sensor networks," in *Proc. 43rd Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove (CA), USA, Nov. 2009.

[4] V. Schwarz and G. Matz, "Distributed reconstruction of time-varying spatial fields based on consensus propagation," in *Proc. ICASSP 2010*, Dallas (TX), USA, March 2010, pp. 2926–2929.

[5] A.D. Sarwate and A.G. Dimakis, "The impact of mobility on gossip algorithms," in *Proc. 28th Conf. Computer Communications (INFOCOM)*, Rio de Janeiro, Brazil, April 2009.

[6] R. Olfati-Saber and R.M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1520 – 1533, 2004.

[7] L.L. Scharf and D.W. Tufts, "Rank reduction for modeling stationary signals," vol. ASSP-35, no. 3, pp. 350–355, March 1987.

[8] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty - V: The discrete case," *The Bell System Technical Journal*, vol. 57, no. 5, pp. 1371–1430, May-June 1978.