MEAN-SQUARE OPTIMAL WEIGHT DESIGN FOR AVERAGE CONSENSUS

Valentin Schwarz and Gerald Matz

Institute of Telecommunications, Vienna University of Technology
Gußhausstraße 25/389, A-1040 Vienna, Austria; email: {vschwarz,gmatz}@nt.tuwien.ac.at

ABSTRACT

In this work we investigate the best possible convergence of average consensus in the mean-square sense with weights that change over time. Although this convergence may be hard to achieve in practice, it provides useful practical insights into the behavior of average consensus. In particular, we show that the correlation between the states plays an important role in the design of the weights and that the weight-morphing scheme proposed in our previous work is close to optimal both for correlated and uncorrelated measurements.

Index Terms— Average consensus, wireless sensor networks, distributed inference

1. INTRODUCTION

Distributed algorithms have received considerable attention in recent years since they avoid many of the weaknesses of centralized processing. An example for such a weakness is the vulnerability of centralized systems to failure and attack of the central processor. In this paper we consider distributed inference where a sensor network computes in a distributed manner the average (arithmetic mean) of the measurements obtained by the individual nodes.

While there are also other methods for distributed averaging [1, 2], this work deals with the average consensus (AC) algorithm which was initially proposed in the thesis of Tsitsiklis [3]. During the last years, many extensions and modifications of AC were published. In [4], a promising dynamic version of AC was presented; [5] and [6] study asynchronous AC, which is particularly attractive in the context of wireless sensor networks (WSN). In [7], the authors provide an extensive overview of AC and [8] compares AC and gossip averaging for WSN applications.

In this work we consider the discrete-time version of AC which is suited for applications in WSN. Here, the weights of AC can be tuned to optimize the convergence behavior of AC. In [9] the authors introduce different objectives and methods to optimize the weights; one of the weight designs achieves the best asymptotic convergence speed. However, the asymptotically best weight design can be inferior to other weight designs during the initial averaging phase. This disadvantage is avoided with nonlinear AC schemes as proposed in [10]. In our previous work [11], we introduced a simple weight morphing technique that extends [10] and achieves faster convergence. Weight morphing can also be applied in dynamic AC to improve the tracking performance [11].

In this work, we propose a weight design method that optimizes the overall convergence of AC in the mean-square sense. This establishes an ultimate performance benchmark for other weight designs, specifically the one in [11]. The result is related to the work of [12] where the authors optimize weights to achieve a maximal MSE decrease for uncorrelated states. We additionally provide results for correlated states, which is practically useful and necessary since the AC iterations render the states increasingly correlated. In contrast to [12], we further derive closed-form solutions for the optimal weights for the case of correlated and uncorrelated states. Finally, we show numerical simulations that illustrate the performance of the proposed weight design.

Funded by WWTF Grant ICT-044 and FWF Grant S10606.

2. PRELIMINARIES

We consider a scenario in which AC is applied on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ that models a WSN. The set $\mathcal{V}$ represent the sensors and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of undirected edges that models the communication links between nodes. In the case of random geometric graphs [13] (a suitable graph model for WSN), a link between two nodes exists if they lie within a prescribed communication range. We will describe further graph models in Section 4.1. In what follows, the number of nodes is denoted as $I = |\mathcal{V}|$ and we assume without loss of generality that $\mathcal{V} = \{1, \ldots, I\}$ and that $\mathcal{G}$ is connected, i.e., there exists a path between any two nodes in the graph. We further consider only graphs whose topology does not change over time.

Initially, every sensor obtains a measurement s_i, $i = 1, \ldots, I$. These measurements are to be averaged by all sensors in a non-centralized manner, i.e., the entire network should know the mean $\bar{s} = \frac{1}{I}\sum_{i=1}^{I} s_i$. AC provides a simple method to achieve this distributed averaging task. Each sensor stores the measurement as initial state $x_i[0] = s_i$. Then, at each iteration, the state is broadcasted to the neighboring nodes and the following state update is performed at each node:

$$x_i[k+1] = w_{ii}\, x_i[k] + \sum_{j:\,(i,j)\in\mathcal{E}} w_{ij}\, x_j[k]\,.$$

The state updates of all nodes can be compactly rewritten as

$$\boldsymbol{x}[k+1] = \boldsymbol{W}\boldsymbol{x}[k]\,, \tag{1}$$

where $\boldsymbol{x}[k] = \big(x_1[k]\ x_2[k]\ \ldots\ x_I[k]\big)^T$ is the state vector and $\boldsymbol{W}$ denotes the weight matrix whose elements are given by

$$[\boldsymbol{W}]_{ij} = \begin{cases} w_{ij}, & (i,j) \in \mathcal{E}, \\ 0, & \text{else.} \end{cases}$$

The state vector converges to the true mean, $\lim_{k\to\infty} \boldsymbol{x}[k] = \bar{s}\mathbf{1}$, if the weight matrix satisfies the conditions $\boldsymbol{W}\mathbf{1} = \mathbf{1}$ and $\rho(\boldsymbol{W} - \boldsymbol{J}) < 1$; here, $\mathbf{1}$ is the all-ones vector, $\rho(\cdot)$ denotes the spectral radius, and $\boldsymbol{J} \triangleq \frac{\mathbf{1}\mathbf{1}^T}{\mathbf{1}^T\mathbf{1}} = \frac{1}{I}\mathbf{1}\mathbf{1}^T$ is the orthogonal rank-one projection matrix on the subspace spanned by the all-ones vector. We refer the reader to [7] for further details.

The convergence speed of AC depends on the weight matrix $\boldsymbol{W}$. In [9] different weight design methods for optimal asymptotic convergence speed are presented. Specifically, $\boldsymbol{W}$ is chosen to minimize the spectral gap of $\boldsymbol{W}-\boldsymbol{J}$ since this gap determines the asymptotic convergence speed. This minimization problem can be solved using convex optimization techniques. The resulting weights will be referred to as CVX weights in what follows. The drawback of this approach is that it requires to solve an optimization problem on the entire graph. A popular method to obtain weights in a distributed fashion is inspired by Markov chain theory and determines the weights according to the maximum degree of two neighboring nodes. The weights are called maximum degree or Metropolis-Hastings (MH) weights. In [11], it is shown that adaptive weight morphing between MH and CVX weights achieves better performance than both of these designs individually. With this weight morphing method, the (nonlinear) state update reads

$$\boldsymbol{x}[k+1] = \boldsymbol{W}(\boldsymbol{x}[k])\,\boldsymbol{x}[k]\,, \tag{2}$$

where the weight matrix depends on the current state (and hence is implicitly time-varying).

3. PROPOSED METHOD

The main contribution of this work is a stepwise MSE-optimal weight design method. In this section we first compute the MSE and then derive a closed-form solution for the weights that are MSE-optimal. We allow for a different weight matrix $\boldsymbol{W}[k]$ to be used in each AC update (cf. (1)). We assume $\boldsymbol{W}^T[k] = \boldsymbol{W}[k]$ and impose the side-constraints $\boldsymbol{W}[k]\mathbf{1} = \mathbf{1}$ that ensures that the spatial mean is preserved during the iterations, i.e., $\mathbf{1}^T\boldsymbol{x}[k] = I\bar{s}$ and hence $\boldsymbol{J}\boldsymbol{x}[k] = \boldsymbol{J}\boldsymbol{x}[0] = \bar{s}\mathbf{1}$.

3.1. MSE Derivation

A reasonable metric for the difference between the actual mean and the AC estimate $\boldsymbol{x}[k]$ is given by the MSE

$$\epsilon^2[k] \triangleq \mathrm{E}\left\{\left\|\boldsymbol{x}[k] - \bar{s}\mathbf{1}\right\|^2\right\},$$

where $\mathrm{E}\{\cdot\}$ denotes the expectation operator. The MSE will be seen below to have the advantage of leading to a framework amenable to closed-form solution. Using $\boldsymbol{x}[k+1] = \boldsymbol{W}[k]\boldsymbol{x}[k]$ and $\bar{s} = \frac{1}{I}\mathbf{1}^T\boldsymbol{x}[0]$, the MSE at time $k+1$ can be developed as

$$\begin{aligned}
\epsilon^2[k+1] &\triangleq \mathrm{E}\left\{\left\|\boldsymbol{W}[k]\boldsymbol{x}[k] - \boldsymbol{J}\boldsymbol{x}[0]\right\|^2\right\} \\
&= \mathrm{E}\left\{\left\|\boldsymbol{W}[k]\boldsymbol{x}[k] - \boldsymbol{J}\boldsymbol{x}[k]\right\|^2\right\} \\
&= \mathrm{E}\left\{\left\|\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)\boldsymbol{x}[k]\right\|^2\right\} \\
&= \mathrm{E}\left\{\mathrm{tr}\left\{\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)\boldsymbol{x}[k]\boldsymbol{x}^T[k]\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)^T\right\}\right\} \\
&= \mathrm{tr}\left\{\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)\boldsymbol{R}_{\boldsymbol{x}}[k]\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)^T\right\}
\end{aligned} \tag{3}$$

where in the second step we used the mean-preservation property and $\mathrm{tr}\{\cdot\}$ denotes the matrix trace; furthermore, in the last step we defined $\boldsymbol{R}_{\boldsymbol{x}}[k] = \mathrm{E}\{\boldsymbol{x}[k]\boldsymbol{x}^T[k]\}$. The expression (3) shows that the MSE only depends on the correlation matrix and the weight matrix (and, via the weight matrix, implicitly on the graph topology). If the states are i.i.d., i.e., $\boldsymbol{R}_{\boldsymbol{x}}[k] = \boldsymbol{I}$, the MSE simplifies to

$$\epsilon^2[k+1] = \|\boldsymbol{W}[k] - \boldsymbol{J}\|_{\mathrm{F}}^2 = \sum_i \lambda_i[k]^2\,,$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm and λ_i denotes the ith eigenvalue of $\boldsymbol{W}[k] - \boldsymbol{J}$.

3.2. MSE Minimization

Our goal is to choose the weight matrix $\boldsymbol{W}[k]$ to minimize the MSE in (3), thereby achieving the best step-by-step performance in the MSE sense. Repeating this optimization for every iteration step k yields a sequence of weight matrices that optimize the overall performance. The optimization problem of interest can be written as

$$\boldsymbol{W}[k] = \arg\min_{\boldsymbol{W}} \epsilon[k+1]^2 \quad \text{subject to } \boldsymbol{W}\mathbf{1} = \mathbf{1},\ \boldsymbol{W}^T = \boldsymbol{W}. \tag{4}$$

This constrained optimization problem is convex and can be solved with standard numerical optimization tools (e.g. [14]) as is done in [12] for the uncorrelated case. We next derive a closed-form solution for the general case.

We take the side constraints into account by optimizing only the north-east off-diagonal elements of $\boldsymbol{W}[k]$; the south-west off-diagonal elements and the the diagonal elements as are determined as $w_{ij}[k] = w_{ji}[k]$ and $w_{ii}[k] = 1 - \sum_{j\neq i} w_{ij}[k]$. Arranging the north-east off-diagonal elements of the weight matrix into a vector $\boldsymbol{w}$ of length $|\mathcal{E}|$, we can write

$$\boldsymbol{W} = \boldsymbol{I} - \boldsymbol{B}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{B}^T, \tag{5}$$

where $\boldsymbol{B}$ denotes the $I \times |\mathcal{E}|$ incidence matrix of $\mathcal{G}$ and $\mathrm{diag}\{\cdot\}$ converts a vector into a diagonal matrix or extracts the diagonal of a matrix into a vector. The incidence matrix has size $I \times E$ and $[\boldsymbol{B}]_{ij} = 1$ if edge j starts at node i and $[\boldsymbol{B}]_{ij} = -1$ if edge j ends at node i. Since we only consider undirected graphs the sign is irrelevant. With this definition, the side constraints are satisfied for any choice of $\boldsymbol{w}$ and thus the constrained optimization problem (4) with respect to $\boldsymbol{W}$ is thereby converted into an unconstrained optimization problem with respect to $\boldsymbol{w}$.

We next rewrite the cost function $\epsilon^2[k]$ as

$$\begin{aligned}
\epsilon^2[k+1] &= \mathrm{tr}\left\{\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)\boldsymbol{R}_{\boldsymbol{x}}[k]\big(\boldsymbol{W}[k]-\boldsymbol{J}\big)^T\right\} \\
&= \mathrm{tr}\Big\{\boldsymbol{W}[k]\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{W}^T[k] - \boldsymbol{J}\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{W}^T[k] \\
&\qquad - \boldsymbol{W}[k]\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{J}^T + \boldsymbol{J}\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{J}^T\Big\} \\
&= \mathrm{tr}\left\{\boldsymbol{W}[k]\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{W}^T[k]\right\} - c^2
\end{aligned} \tag{6}$$

where

$$c^2 = \mathrm{tr}\big\{\boldsymbol{J}\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{J}^T\big\} = \mathrm{E}\left\{\mathrm{tr}\big\{\boldsymbol{J}\boldsymbol{x}[k]\boldsymbol{x}^T[k]\boldsymbol{J}^T\big\}\right\} = \mathrm{E}\{\bar{s}^2\}\,I$$

is independent of $\boldsymbol{W}$. Inserting (5) into (6) we obtain

$$\begin{aligned}
\epsilon^2[k+1] &= \mathrm{tr}\left\{\boldsymbol{I}\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{I}\right\} - 2\,\mathrm{tr}\left\{\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{B}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{B}^T\right\} \\
&\quad + \mathrm{tr}\left\{\boldsymbol{B}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{B}^T\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{B}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{B}^T\right\} - c^2 \\
&= \mathrm{tr}\left\{\,\mathrm{diag}\{\boldsymbol{w}\}\,\boldsymbol{G}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{H}\right\} \\
&\quad - 2\,\mathrm{tr}\{\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{G}\} + \mathrm{tr}\{\boldsymbol{R}_{\boldsymbol{x}}[k]\} - c^2,
\end{aligned}$$

where we defined $\boldsymbol{G} \triangleq \boldsymbol{B}^T\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{B}$ and $\boldsymbol{H} \triangleq \boldsymbol{B}^T\boldsymbol{B}$ (the iteration index k of $\boldsymbol{G}$ is omitted for simplicity reasons). This last expression can be minimized by setting its first-order derivative to zero, i.e.,

$$\frac{d\epsilon^2[k+1]}{d\boldsymbol{w}} = 0\,,$$

Writing out the MSE in terms of the elements of $\mathrm{diag}\{\boldsymbol{w}\}$, $\boldsymbol{G}$, and $\boldsymbol{H}$ yields

$$\epsilon^2[k+1] = \sum_{i,j} w_i[\boldsymbol{G}]_{ij} w_j[\boldsymbol{H}]_{ji} - 2\sum_i (w_i[\boldsymbol{G}]_{ii} + [\boldsymbol{R}_{\boldsymbol{x}}[k]]_{ii}) - c^2,$$

which, after differentiation, results in the linear equations

$$\frac{d\epsilon^2[k+1]}{dw_i} = 2\sum_j [\boldsymbol{G}]_{ij}[\boldsymbol{H}]_{ij} w_j - 2[\boldsymbol{G}]_{ii} = 0\,.$$

Equivalently, we have

$$(\boldsymbol{G} \circ \boldsymbol{H})\boldsymbol{w} = \mathrm{diag}\{\boldsymbol{G}\}\,, \tag{7}$$

where $\circ$ denotes the direct (element-wise) product. This is a system of linear equations whose solution

$$\boldsymbol{w}^\star = (\boldsymbol{G} \circ \boldsymbol{H})^{-1}\,\mathrm{diag}\{\boldsymbol{G}\}$$

can be computed e.g. using the QR factorization. The optimum weight matrix is then obtained by inserting $\boldsymbol{w}^\star$ into (5).

We note that the AC state update $\boldsymbol{x}[k+1] = \boldsymbol{W}[k]\boldsymbol{x}[k]$ implies $\boldsymbol{R}_{\boldsymbol{x}}[k+1] = \boldsymbol{W}[k]\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{W}^T[k]$ and hence by induction

$$\boldsymbol{R}_{\boldsymbol{x}}[k+1] = \boldsymbol{W}[k]\ldots\boldsymbol{W}[0]\boldsymbol{R}_{\boldsymbol{x}}[0]\boldsymbol{W}^T[0]\ldots\boldsymbol{W}^T[k].$$

Thus, successive optimization of the weight matrices requires only the initial correlation $\boldsymbol{R}_{\boldsymbol{x}}[0] = \boldsymbol{R}_{\boldsymbol{s}}$ of the measurements to be known.

Since the states $x_i[k]$ will become more and more equal as k increases, their correlation matrix tends towards a scalar multiple of the rank-one matrix $\boldsymbol{J}$ and hence (7) becomes numerically unstable. To mitigate numerical problems, the weight optimization scheme is modified by replacing $\boldsymbol{R}_{\boldsymbol{x}}[k]$ with $\boldsymbol{R}_{\boldsymbol{x}}[k] - \mathrm{tr}\{\boldsymbol{R}_{\boldsymbol{x}}[k]\boldsymbol{J}\}\boldsymbol{J}$. This replacement causes only an additive shift in the MSE cost function (6) and hence does not change the solution for the optimum weights.

In general, $\boldsymbol{W}^*[k]$ is not guaranteed to be a stable solution for AC, i.e., freezing the weight matrix $\boldsymbol{W}[k'] = \boldsymbol{W}^*[k]$, $k' > k$, may lead to unstable behavior. While it is not possible to compute stable MSE-optimal weights in closed form, they can in principle be obtained by solving the following constrained convex optimization problem:

$$\begin{aligned} &\min \|(\boldsymbol{G} \circ \boldsymbol{H})\boldsymbol{w} - \mathrm{diag}\{\boldsymbol{G}\}\|^2 \\ &\text{s.t.} \quad \rho\big(\boldsymbol{W} - \boldsymbol{J}\big) < 1\,. \end{aligned} \tag{8}$$

3.3. Uncorrelated Measurements

In the following we consider the special case of uncorrelated measurements, i.e., $\boldsymbol{R}_{\boldsymbol{x}}[0] = \boldsymbol{R}_{\boldsymbol{s}} = \gamma\boldsymbol{I}$. Here, it can be shown that (7) simplifies to

$$(\boldsymbol{A}(\mathcal{L}) + 4\boldsymbol{I})\boldsymbol{w} = 2\boldsymbol{1}\,, \tag{9}$$

where $\boldsymbol{A}(\mathcal{L})$ denotes the adjacency matrix of the line graph $\mathcal{L}$ (also called adjoint graph) of $\boldsymbol{A}$ (see [15] for more details). It is seen that by minimizing the MSE (6), the weights are invariant to the variance γ of the measurements, which is intuitive. In practice uncorrelated measurements can only occur in the initial state, but the resulting weights can be also used as traditional constant weights for AC.

In our simulations we observed that for the uncorrelated case we mostly obtained positive weights. For some graphs this can be proved by requiring diagonal dominance in (9). In particular, [16] implies that a sufficient condition for the solution of (9) to be positive is that the sum of the non-diagonal elements in each row of $(\boldsymbol{A}(\mathcal{L})+4\boldsymbol{I})$ is smaller than the diagonal element (which here equals 4). This is guaranteed for graphs where the sum of the degrees of any two neighboring nodes is less than 5 and hence the maximum degree in the associated line graph is less than 4. We found empirically that this sufficient condition is rather conservative, i.e., the majority of graphs that does not satisfy this condition still leads to positive weights. So far, we are aware of only one type of graph for which $\boldsymbol{w}$ has non-positive elements sufficiently large star graphs with additional edges between leaf nodes or a graph which consist of two star graphs where the center nodes are connected).

A positive $\boldsymbol{w}$ ensures that $\boldsymbol{B}\,\mathrm{diag}\{\boldsymbol{w}\}\boldsymbol{B}^T$ is positive-semidefinite. Hence, for positive weights the eigenvalues of $\boldsymbol{W}-\boldsymbol{J}$ are all smaller than one. Initial results and simulations suggest that the smallest eigenvalue of $\boldsymbol{W}-\boldsymbol{J}$ is always larger than -1 (thereby rendering $\boldsymbol{W}$ stable). However, we do not have a general proof for this conjecture so far. As an example, the smallest eigenvalue of the star graph can be shown to equal $\frac{2-I}{2+I} > -1$ (this is the eigenvalue closest to -1 that we observed for graphs that result in positive weights).

4. NUMERICAL RESULTS

In this section we provide results of numerical simulations illustrating the MSE optimal convergence of AC. Before showing error curves, we introduce different graph models for which we then show the amount of state correlation over time.

4.1. Graph Models

In all our examples, we considered graphs with $I = 100$ nodes; the associated edges were constructed using the following models.

A very simple random graph model is the Erdös-Renyi (ER) model, denoted by $\mathcal{G}(I, p)$ where I denotes the number of nodes and p is the probability of an edge between two nodes. Hence, the expected number of edges is $pI(I-1)$.

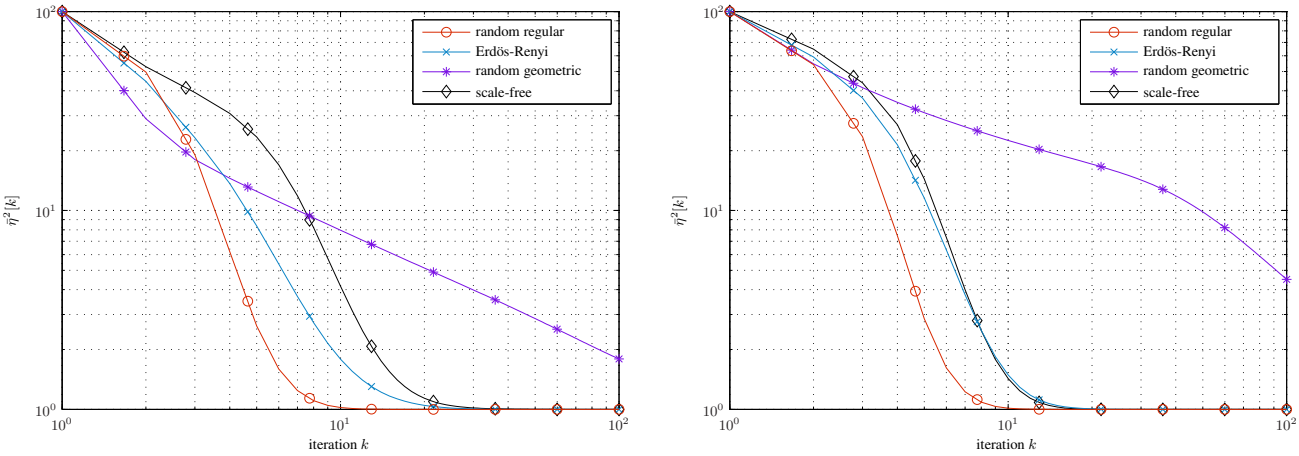
Another type of random graphs are random d-regular graphs [17], where each node has r randomly chosen neighbors. The number of nodes I has to be even for the graph construction to work.

Social networks and the Internet can be modeled by a different class of graphs, so called scale-free graphs. These graphs incorporate an exponential degree distribution, i.e., a few "hub" nodes have many neighbors and most of the nodes have only a few neighbors. The model which we used for scale-free graphs is from [18].

In contrast to the other random graphs, [13] incorporates the geometry (positions) of the nodes in the construction process of the graph. For a random geometric graph $\mathcal{G}(I, r, \mathcal{A})$, I nodes are randomly distributed in a region $\mathcal{A}$ and nodes within distance r are connected. Ignoring boundary effects, this yields an expected number of $\frac{r^2\pi(I-1)}{|A|}$ edges. Since random geometric graphs are well suited to model WSN, we concentrate on such graphs in the subsequent simulations. The region $\mathcal{A}$ was chosen as $[0,1]\times[0,1]$ and the radius was $r = \frac{\nu}{\sqrt{I}}$ with $\nu = 2$.

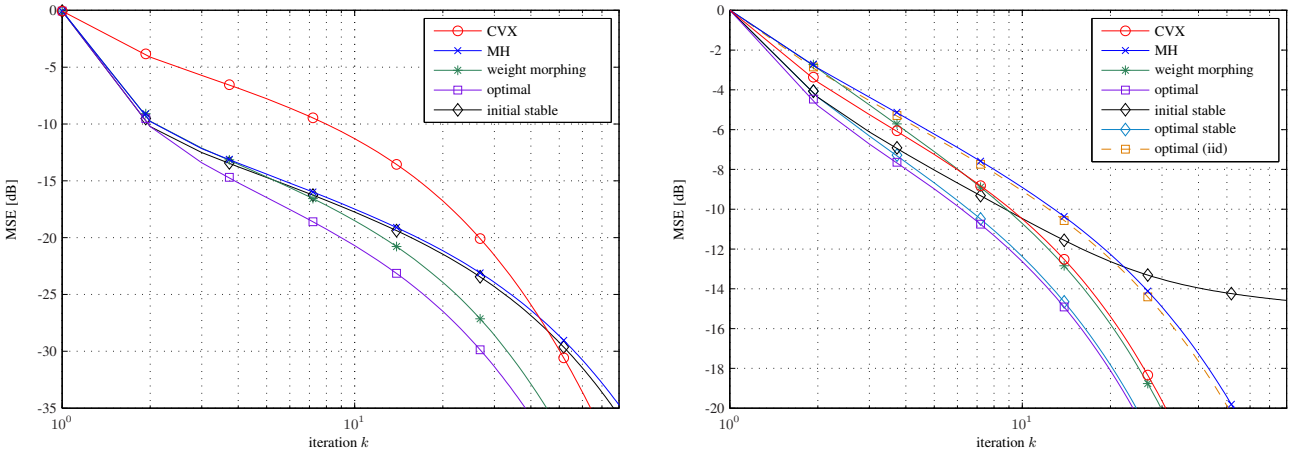
4.2. State Correlation

With AC, the states $x_i[k]$ become more and more equal as the iterations progress. This implies that the correlation between the state variables increases. Fast convergence is indicated by a rapid buildup of strong correlation. We investigate this aspect using uncorrelated



(a) MH weights (b) CVX weights

Fig. 1. *State correlation $\bar{\eta}^2[k]$ versus the number of AC iterations for different graph types: (a) MH weights, (b) CVX weights.*



(a) uncorrelated states (b) correlated states

Fig. 2. *MSE convergence for (a) initially uncorrelated and (b) correlated states for different weight design schemes.*

measurements, i.e., $\boldsymbol{R_s} = \boldsymbol{I}$. As a correlation measure we use the eigenvalue spread of the correlation matrix $\boldsymbol{R_x}[k]$,

$$\eta^2[k] \triangleq \frac{\sum_{i=1}^{I} \lambda_i^2[k]}{\left(\sum_{i=1}^{I} \lambda_i[k]\right)^2}$$

with $\lambda_i[k]$ denoting the ith eigenvalue of $\boldsymbol{R}_x[k]$. The metric $\eta^2[k]$ equals I in case all states are uncorrelated and 1 in case of full correlation. To suppress outliers, we averaged $\eta^2[k]$ over 100 scenarios, resulting in $\bar{\eta}^2[k]$.

Fig. 1(a) shows $\bar{\eta}^2[k]$ versus the number of AC iterations. For this setting we used MH weights for all graph types discussed in the previous section. All graphs have 100 nodes and approximately 300 edges. It is seen that random regular graphs are first to achieve almost full correlation. This can be explained by the regular structure and the small diameter of such graphs. ER graphs and scale-free graphs behave very similarly because both have small diameters, but are less homogenous than random regular graphs. The correlation decrease for random geometric graphs scales as $\frac{1}{k}$, i.e., has slowest convergence to full correlation. This is because these graphs have large diameters and hence require many hops to traverse the network.

A slightly different behavior is obtained for CVX weights as shown in Fig. 1(b). For all graphs, $\bar{\eta}^2[k]$ converges to full correlation faster than with MH weights, but the initial decrease is slower than for MH weights. Otherwise the observations made for MH weights apply to CVX weights as well.

4.3. MSE Convergence

Our next simulations consider the MSE convergence for the case of random geometric graphs. The empirical MSE is defined as

$$\epsilon^2[k] \triangleq \frac{\|\boldsymbol{x}[k] - \bar{s}\mathbf{1}\|^2}{\mathbb{E}\{s_i^2[k]\}},$$

averaged over 100 scenarios. The MSE versus the number of iteration k for initially uncorrelated states (i.e., the nodes measure Gaussian noise) is plotted in Fig. 2(a). It is seen that the the CVX weights initially perform poorly but outperform MH the weights after about 40 iterations. The best performance is of course achieved by the optimum weight design according to (7). After the first iteration the MSE of MH weights are almost as good as for the optimum weights and then the weight morphing method proposed in [11] obtains the

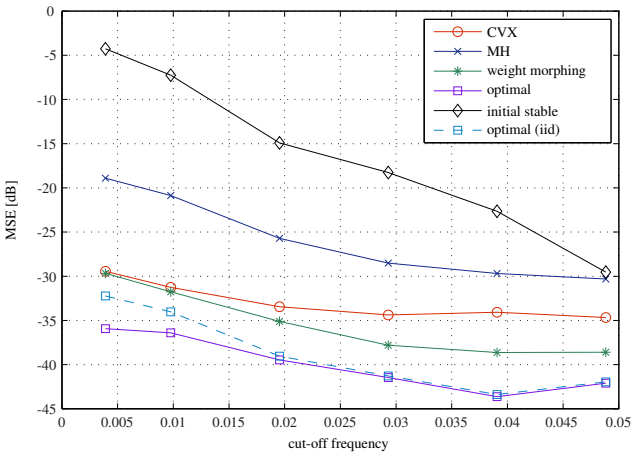


Fig. 3. *MSE achieved with different different weight designs after 80 iterations versus spatial cut-off frequency of the measured field.*

second best performance. Additionally we plotted the convergence of AC where the weights are optimized in the first iteration according to (7) and then kept constant. The performance is slightly better than that of MH weights.

To study the case of correlated initial states, we consider measurements of a spatially correlated field. For the MSE curves in Fig. 2(b) we used a normalized spatial cut-off frequency of 0.02. It is seen that in contrast to the uncorrelated case, here the CVX weights perform always better than the MH weights. Again, the optimum weights perform best, but these weights are not stable. Therefore, we also plot the MSE curve for the optimum stable weights obtained via (8) (the optimization needs to be done at each iteration) and the performance is a bit worse than for the optimum with unstable weights. The scheme labeled "initial stable" performs the stable optimization of (8) only in the first iteration and then leaves the weights constant; it achieves fast convergence initially but then saturates and performs poorly. If the correlation of the initial state vector is not known and an uncorrelated setting is assumed and the weights are only optimized in the first step, then the performance is only slightly better than with MH weights.

4.4. Field Averaging

Finally, we provide results for different spatial measurement correlations by varying the spatial cut-off frequency of the field. The MSE averaged over 1000 simulations is plotted in Fig. 3. It is seen that for all cut-off frequencies the weight morphing method of [11] lies 5dB above the best possible MSE curve, which is much better than the MH curve and slightly better than the CVX curve. Assuming uncorrelated initial states instead of correlated states only deteriorates the performance for highly correlated states/low cut-off frequencies. Performing an ideal optimization at the beginning only provides good results in the transient phase (see Fig. 2), but after 80 iterations it has the worst MSE.

5. CONCLUSION

We have shown how to design the weights in average consensus in order to achieve the best MSE performance. We have provided a closed-form solution to this problem in terms of linear system of equations. Optimizing the weights in each steps may be practically challenging but provides useful performance bounds for the convergence of other designs like the weight morphing scheme of [11]. In our simulations we have shown how the correlations of the states change over time and that the graph topology plays an important role. Moreover the choice of constant weights is dependent on the initial correlation of the states. The distributed implementation of the MSE-optimal weight design is an interesting topic for future work.

REFERENCES

[1] C. C. Moallemi, *A Message-Passing Paradigm For Optimization*, Ph.D. thesis, Stanford University, Sept. 2007.

[2] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508–2530, June 2006.

[3] J.N. Tsitsiklis, *Problems in Decentralized Decision making and Computation*, Ph.D. thesis, Massachusetts Institute of Technology, Dec. 1984.

[4] M. Zhu and S. Martínez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, issue 2, pp. 322–329, Feb. 2009.

[5] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[6] M. Mehyar, D. Spanos, J. Pongsajapan, S.H. Low, and R.M. Murray, "Asynchronous distributed averaging on communication networks," *IEEE/ACM Trans. on Networking*, vol. 15, no. 3, pp. 512–520, June 2007.

[7] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[8] Patrick Denantes, Florence Benezit, Patrick Thiran, and Martin Vetterli, "Which distributed averaging algorithm should I choose for my sensor network?," in *Proc. IEEE INFOCOM-2008*, Phoenix, AZ, April 2008, pp. 986–994.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[10] L. Georgopoulos and M. Hasler, "Nonlinear average consensus," in *Proceedings of the 2009 International Symposium on Nonlinear Theory and its Applications*, Oct. 2009, p. 10.

[11] V. Schwarz and G. Matz, "Nonlinear average consensus based on weight morphing," in *Proc. IEEE ICASSP-2012*, Kyoto, JP, March 2012.

[12] D. Jakovetiï£¡c, J. Xavier, and J. M. F. Moura, "Consensus in correlated random topologies: weights for finite time horizon," in *Proc. IEEE ICASSP-2010*, Dallas, TX, March 2010, pp. 2974–2977.

[13] Mathew Penrose, *Random Geometric Graphs*, Oxford University Press, 2003.

[14] M. Grant and S. Boyd, ," CVX: Matlab software for disciplined convex programming (web page and software), Stanford University, CA (http://stanford.edu/∼boyd/cvx).

[15] G. Royle C. Godsil, *Algebraic Graph Theory*, Springer, 2001.

[16] M. Kaykobad, "Positive solutions of positive linear systems," *Linear Algebra and its Applications*, vol. 64, pp. 133–140, 1985.

[17] B. Bollobás, *Random Graphs*, Cambridge Univ Press, 2001.

[18] Réka Albert and Albert-László Barabási, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.