Optimal strategies in the average consensus problem

Jean-Charles Delvenne   Ruggero Carli   Sandro Zampieri

Abstract— We prove that for a set of communicating agents to compute the average of their initial positions (average consensus problem), the optimal topology of communication is given by a de Bruijn's graph. Consensus is then reached in a finitely many steps. A more general family of strategies, constructed by block Kronecker products, is investigated and compared to Cayley strategies.

I. INTRODUCTION

Coordination algorithms for multiple autonomous vehicles and decentralized estimation techniques for handling data coming from distributed sensor networks have attracted large attention in recent years. This is mainly motivated by that both coordinated control and distributed estimation have applications in many areas, such as coordinated flocking of mobile vehicles [26], [27], cooperative control of unmanned air and underwater vehicles [4], [3], multi-vehicle tracking with limited sensor information [19], monitoring very large scale areas with fine resolution and collaborative estimation over wireless sensor networks [24].

Typically, both in coordinated control and in distributed estimation the agents need to communicate data in order to execute a task. In particular they may need to agree on the value of certain coordination state variables. One expects that, in order to achieve coordination, the variables shared by the agents, converge to a common value, asymptotically. The problem of designing controllers that lead to such asymptotic coordination is called *coordinated consensus*, see for example [12], [20], [15] and references therein. Generalisation to high order consensus [22] and nonholonomic agents [18], [11], [28] have also been explored. One of the simplest consensus problems that has been mostly studied consists in starting from systems described by an integrator and in finding a feedback control yielding consensus, namely driving all the states to the same value [20]. The information exchange is modeled by a directed graph describing in which pair of agents the data transmission is allowed. The situation mostly treated in the literature is when each agent has the possibility of communicate its state to the other agents that are positioned inside a neighborhood [26], [15] and the communication network is time-varying [27], [15]. Robustness to communication link failure [8] and the effects of time delays [20] has been considered recently. Randomly time-varying networks have also been analyzed in [14]. Moreover, a first analysis involving quantized data

J.-C. Delvenne is with the Institute for Mathematical Sciences, Imperial College, 53 Prince's Gate, London SW7 2PG, United Kingdom, jc.delvenne@imperial.ac.uk. R. Carli and S. Zampieri are with the Department of Information Engineering, Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy {carlirug|zampi}@dei.unipd.it.

transmission has been proposed in [7], [16]. In this paper we consider the consensus problem from a different perspective. We are interested to characterize the relationship between the amount of information exchanged by the agents and the achievable control performance. More precisely we assume that N agents are given initial positions in the euclidean space, and move in discrete-time in order to reach the average of their initial positions. This problem is also called *average coordinated consensus*. Every agent asks several agents their position before taking a decision to modify its own position. We impose that, in order to limit costs of communication, every agent communicates with only ν agents (including itself), where $\nu < N$. This means that in the graph describing the communications between agents, the max in-degree is at most ν. In this paper, we exhibit a family of strategies for solving this problem based on de Bruijn's graphs and we prove that according to a suitable criteria this is the best that one can do. Precisely we compute its performances according two criteria: rate of convergence to the average of their initial positions and an LQR criterion. We find that a deadbeat strategy is optimal according to the rate of convergence, and nearly optimal according to the LQR criterion. Finally, we compare it with an another strategy having limited communication and exhibiting symmetries: the Cayley strategies [6], [5]. It should be noted however that our strategy is limited to the case where the number of agents is an exact power of ν. Whether it is possible to build a linear time-invariant deadbeat strategy for any number of agents (for a given ν) remains an open problem.

The paper is organized as follows. In Section II we provide some basic notions of graph theory and some notational conventions. In Section III we formally define the average consensus problem. In Section IV we introduce the block Kronecker strategy. In Section V we show that the block Kronecker strategy is the quickest possible strategy and we compare it with the Cayley strategy. In section VI we evaluate the performance of the block Kronecker strategy according to suitable quadratic criteria. Finally we gather our conclusions in Section VII.

II. PRELIMINARIES ON GRAPH THEORY

Before defining the problem we want to solve, we summarize some notions on graph theory that will be useful throughout the rest of the paper.

Let $\mathcal{G} = (V, \mathcal{E})$ be a directed *graph* where $V = (1, \ldots, N)$ is the set of *vertices* and $\mathcal{E} \subset V \times V$ is the set of *arcs* or *edges*. If $(i, j) \in \mathcal{E}$ we say that the arc (i, j) is outgoing from i and incoming in j. The *adjacency matrix* A is a $\{0, 1\}$-valued square matrix indexed by the elements in V

defined by letting $A_{ij} = 1$ if and only if $(i,j) \in \mathcal{E}$. Define the *in-degree* of a vertex j as $\sum_i A_{ij}$ and the *out-degree* of a vertex i as $\sum_j A_{ij}$. In our setup we admit the presence of self-loops. A graph is called *in-regular* (*out-regular*) of degree k if each vertex has in-degree (out-degree) equal to k. A *path* in $\mathcal{G}$ consists of a sequence of vertices $i_1 i_2 \ldots \ldots i_r$ such that $(i_\ell, i_{\ell+1}) \in \mathcal{E}$ for every $\ell = 1, \ldots, r-1$; i_1 (resp. i_r) is said to be the *initial* (resp. *terminal*) vertex of the path. A *cycle* is a path in which the initial and the terminal vertices coincide. A vertex i is said to be *connected* to a vertex j if there exists a path with initial vertex i and terminal vertex j. A directed graph is said to be *connected* if, given any pair of vertices i and j, either i is connected to j or j is connected to i. A directed graph is said to be *strongly connected* if, given any pair of vertices i and j, i is connected to j.

Finally some notational conventions. Let A any matrix belonging to $\mathbb{R}^{N \times N}$. With $\mathrm{Tr}\, A$ we denote the trace of A, i.e. the sum of the diagonal entries. We say that A is nonnegative, denoted $A \geq 0$, or positive, denoted $A > 0$, if the entries of A are respectively nonnegative or positive.

III. PROBLEM FORMULATION

We suppose that the positions of all N agents are listed into one vector of dimension N. If the agents move, say, in $\mathbb{R}^3$, it seems that we would need a $3N$-dimensional vector. However we will suppose that the positions are scalar, as every linear strategy on scalar positions, if applied separately on every component of the position, trivially extends to strategies for higher dimensions.

More precisely the problem of our interest can be formalized in the following way. Consider $N > 1$ identical systems whose dynamics are described by the following discrete time state equations

$$x_i^+ = x_i + u_i \qquad i = 1, \ldots, N$$

where $x_i \in \mathbb{R}$ is the state of the i-th system, x_i^+ represents the updated state and $u_i \in \mathbb{R}$ is the control input. More compactly we can write

$$x^+ = x + u \qquad (1)$$

where $x, u \in \mathbb{R}^N$. The goal is to design a feedback control law

$$u = Kx, \qquad K \in \mathbb{R}^{N \times N}$$

yielding the average consensus, namely a control such that all the x_i's become asymptotically equal to the average of the initial condition. More precisely, our objective is to obtain K such that, for any initial condition $x(0) \in \mathbb{R}^N$, the closed loop system

$$x^+ = (I + K)x,$$

yields

$$\lim_{t \to \infty} x(t) = \alpha \mathbf{1} \qquad (2)$$

where $\mathbf{1} := [1, \ldots, 1]^T$ and

$$\alpha = \frac{1}{N} \mathbf{1}^T x(0). \qquad (3)$$

Writing $x(t)$ as a linear combination of the eigenvectors of $I + K$, it is almost immediate to see that the average consensus problem is solved if and only if the following three conditions hold:

(A) Every row and every column of $I + K$ sums to one. Hence it has eigenvalue 1 with $\mathbf{1}$ as left and right eigenvector.

(B) The eigenvalue 1 of $I + K$ has algebraic multiplicity one (namely it is a simple root of the characteristic polynomial of $I + K$).

(C) All the other eigenvalues are strictly inside the unit circle.

For nonnegative matrices, namely for matrices having all the components nonnegative, condition (A) is called double stochasticity, condition (B) is ergodicity and condition (C) is a consequence of double stochasticity. We do not require our matrices to be nonnegative, even though it will appear that the optimal matrices are.

Observe now that the fact that the element in position i, j of the matrix $I + K$ is different from zero, means that the system i needs to know exactly the state of the system j in order to compute its feedback action. This implies that the j-th agent must communicate his state x_j to i-th agent. In this context a good description of the communication effort required by a specific feedback K is given by the directed graph $\mathcal{G}_{I+K}$ with set of vertices $\{1, \ldots, N\}$ in which there is an arc from j to i whenever in the feedback matrix K the element $(I + K)_{ij} \neq 0$. The graph $\mathcal{G}_K$ is said to be the *communication graph* associated with K. Conversely, given any directed graph $\mathcal{G}$ with set of vertices $\{1, \ldots, N\}$, a feedback K is said to be *compatible* with $\mathcal{G}$ if $\mathcal{G}_{I+K}$ is a subgraph of $\mathcal{G}$ (we will use the notation $\mathcal{G}_{I+K} \subseteq \mathcal{G}$).

In the sequel, we will impose the following constraint on the communication graph: the max in-degree of the nodes is ν. This models the fact that communication lines are costly to establish or operate, and every agent has the right to talk to a limited number of other agents. Note that for compatibility with usual conventions we consider that ν counts all arcs entering a node, including self-loops (which could be considered as 'free communication' in most technological situations).

Without this constraint, the problem becomes trivial: choose the complete graph, and the consensus is reached in one step. We therefore add the following constraint on $I + K$:

(D) Every row of $I + K$ contains at most ν non-zero elements.

From this point of view we would like to obtain a matrix $I + K$ satisfying (A),(B),(C),(D) and minimizing a suitable performance index. The simplest control performance index is the exponential rate of convergence to the average consensus. When we are dealing with average consensus controllers it is meaningful to consider the displacement from the average of the initial condition

$$\Delta(t) := x(t) - \left(\frac{1}{N} \mathbf{1}^T x(0) \right) \mathbf{1}.$$

It is immediate to check that, $\Delta(t) = x(t) - \left(\frac{1}{N}\mathbf{1}^T x(t)\right)\mathbf{1}$ (since the average position $\frac{1}{N}\mathbf{1}^T x(t)$ is the same at all times t) and that it satisfies the closed loop equation

$$\Delta^+ = (I + K)\Delta. \qquad (4)$$

Notice moreover that the initial conditions $\Delta(0)$ are such that

$$\mathbf{1}^T \Delta(0) = 0. \qquad (5)$$

Hence the asymptotic behavior of our consensus problem can equivalently be studied by looking at the evolution (4) on the hyperplane characterized by the condition (5). The speed of convergence toward the average of the initial condition can be defined as follows. Let P any matrix satisfying conditions (A),(B),(C). Define

$$\rho(P) = \begin{cases} 1 & \text{if } \dim \ker(P - I) > 1 \\ \max_{\lambda \in \sigma(P)\setminus\{1\}} |\lambda| & \text{if } \dim \ker(P - I) = 1, \end{cases}$$

which is called the *essential spectral radius* of P. As the dominant eigenvalues of P^t is one and the others are smaller in magnitude than $\rho(P)^t$, the essential spectral radius says how quickly P^t converges to the rank-one matrix $1/N\mathbf{1}\mathbf{1}^T$, where N is the dimension of P. In this context the index $\rho(I + K)$ seems quite appropriate for analyzing how performance is related to the communication effort associated with a graph. The smaller the essential spectral radius, the quicker the system will converge to the average of the initial condition.

However in control theory, strategies that converge in finite time or very quickly are sometimes dismissed on the ground that they lead to large values of update values $u(t) = x(t + 1) - x(t)$, that can be physically impossible or very costly to implement. Hence a strategy is often required to optimize an *LQR cost*, taking into account both the quickness convergence and the norm of updates values. Therefore another suitable measure of performance could be the following quantity:

$$J = \mathbb{E}(\sum_{t \geq 0} ||x(t) - x(\infty)||^2 + \gamma ||u(t)||^2), \qquad (6)$$

where $x(t)$ is the vector of positions at time t, $x(\infty) = \lim_\infty x(t)$ is the vector whose every entry is the average of initial positions, $u(t) = x(t + 1) - x(t)$ is the update vector at time t, the initial positions are supposed to be uncorrelated random variables with unit variance, $\mathbb{E}$ denotes the expectation, $||x||^2 = x^T x$ is the euclidean norm and γ is a nonnegative real.

We will prove that the optimal topology of communication (in the meaning of speed of convergence) is given by a de Brujin's graph. We will call the control strategies based on such graph block Kronecker strategies, as explained in the next section. For these strategies we will evaluated (6) and we will compare them to another family of strategies based on a regular communication graph having the same degree ν: the Cayley strategies [6], [5].

IV. BLOCK KRONECKER STRATEGIES

In this section, we define block Kronecker strategies. Let A be a $n \times n$ matrix satisfying (A),(B),(C),(D) and k be a nonnegative integer. Note that if A is full then $n \leq \nu$ (since the number of non-zero elements cannot exceed ν). Then we build an $n^k \times n^k$ matrix M in the following way. Let

$$A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix}$$

be a row-partition of the matrix A, where $a_i \in \mathbb{R}^{1\times n}$. Then M is the matrix

$$M = \begin{bmatrix} I_{n^{k-1}} \otimes a_0 \\ I_{n^{k-1}} \otimes a_1 \\ \vdots \\ I_{n^{k-1}} \otimes a_{n-1} \end{bmatrix}. \qquad (7)$$

For example, if

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

(with $\alpha + \beta = 1$) and $k = 3$, then

$$M = \begin{pmatrix} \alpha & \beta & & & & & & \\ & & \alpha & \beta & & & & \\ & & & & \alpha & \beta & & \\ & & & & & & \alpha & \beta \\ \beta & \alpha & & & & & & \\ & & \beta & \alpha & & & & \\ & & & & \beta & \alpha & & \\ & & & & & & \beta & \alpha \end{pmatrix}$$

This is a kind of block Kronecker product. A general theory of block Kronecker product is built in [17]. We only need a more restricted definition, detailed below. The new matrix M is a matrix of larger dimension than A and satisfying conditions (A),(B),(C),(D): (A) and (D) follow from the definition, while (B) and (C) are proved below. Hence it can play the role of the matrix $I + K$ in Section III. We start by some reminders on Kronecker product, define the block Kronecker product and explore the properties of the latter.

A. Kronecker product

We recall that the *Kronecker product* $A \otimes B$ of the matrices A and B is the matrix $[a_{ij}B]_{i,j}$, whose dimensions are the product of dimensions of A and B. Some useful properties of the Kronecker product are the following:

- $AB \otimes CD = (A \otimes C)(B \otimes D)$;
- $\text{Tr } A \otimes B = \text{Tr } A \text{Tr } B$;
- the eigenvalues of $A \otimes B$ are all possible products of an eigenvalue of A with an eigenvalue of B;
- the eigenvectors of $A \otimes B$ are all possible Kronecker products of an eigenvector of A with an eigenvector of B.

The Kronecker product is sometimes called tensor product. Let us see why. For instance consider the matrices B, C, D

of sizes $m_B \times n_B$, $m_C \times n_C$, $m_D \times n_D$. The Kronecker product has size $m_B m_C m_D \times n_B n_C n_D$, and an arbitrary element of $B \otimes C \otimes D$ can be denoted $(B \otimes C \otimes D)_{abc,def} = B_{ad} C_{be} D_{cf}$, where the index written as abc denotes the number $c + bm_D + am_C m_D$ and the index def is the number $f + en_D + dn_C n_D$; we suppose that the indices start form zero: $a = 0, \ldots, m_B - 1$, etc. If B, C, D happen to be square matrices of size n, this notation coincides with the usual notation in base n of an index running from 0 to $n^3 - 1$. This notation of the Kronecker product is very close to the tensor product used in algebra and differential geometry. The only difference is that $B \otimes C \otimes D$, viewed as a tensor product, is considered as a 6-dimensional array with a, b, c, d, e, f as separate indices, instead of a matrix (i.e., a 2-dimensional array). All this immediately extends to more than three matrices.

B. Block Kronecker product

Let us now consider the following variant of Kronecker product, that we call *block Kronecker product*. Consider for instance two matrices B (of size $n^3 \times n^3$) and C (of size $n^2 \times n^2$). The block Kronecker product of B and C is defined as follows: its element of index $abcde, ghijk$ is the element $B_{cde,ghi} C_{ab,jk}$ (notice the shift of the first indices by two places). We will denote this matrix by $B \odot C$. This definition applies to any two square matrices whose dimensions are powers of n. In general, we can write $(B \odot C)_{p,q} = (B \otimes C)_{\sigma^t(p),q}$, where σ operates a cyclic permutation by one place to the left on the digits of p in base n, and C is of size n^t.

The matrix M defined by Equation (7) can be expressed as $M = (I \otimes \cdots \otimes I) \odot A$ (where the $n \times n$ identity matrix I is repeated $k - 1$ times). If we write the index of M in base n, then $M_{i_1 \ldots i_k, j_1 \ldots j_k} = I_{i_2,j_1} I_{i_3,j_2} \ldots I_{i_{k-1},j_k} A_{i_1,j_0}$.

This form is particularly useful to compute the behavior of M from the properties of the block Kronecker product, which we now explore. As a first property, we can easily see that

$$(B \odot C)^T = C^T \odot B^T. \tag{8}$$

We can also prove the following lemma.

Lemma 4.1: For any matrices A, B, C, D, E, F for which all the products below are meaningful, we have

$$((A \otimes B) \odot C)((D \otimes E) \odot F) = BD \odot (CE \otimes AF). \tag{9}$$

Proof: We write, using Einstein's convention (indices repeated twice in an expression are implicitly summed over),

$$
\begin{aligned}
&[((A \otimes B) \odot C)((D \otimes E) \odot F)]_{u,w} = \\
&= ((A \otimes B) \odot C)_{u,v}((D \otimes E) \odot F)_{v,w} \\
&= A_{u_2,v_1} B_{u_3,v_2} C_{u_1,v_3} D_{v_2,w_1} E_{v_3,w_2} F_{v_1,w_3},
\end{aligned}
$$

where u, v, and w, interpreted as sequences of digits in base n, have been partitioned into $u_1 u_2 u_3$, $v_1 v_2 v_3$, and $w_1 w_2 w_3$ in an appropriate way. This is possible if B and D have same size, as well as C and E, and A and F. Then the expression above can be regrouped as

$$
\begin{aligned}
(BD)_{u_3,w_1}(CE)_{u_1,w_2}(AF)_{u_2,w_3} = \\
= (BD \odot (CE \otimes AF))_{u,w},
\end{aligned}
$$

which ends the proof. ∎

In particular, if $B = D = 1$ we have

$$(A \odot C)(E \odot F) = (CE \otimes AF). \tag{10}$$

If we choose $C = E = 1$ instead, we have

$$(A \otimes B)(D \odot F) = BD \odot AF. \tag{11}$$

The following proposition provides an interesting characterization of the powers of any order of the matrix M.

Proposition 4.1: For A a square matrix, M defined by Equation (7), and any integers $r \geq 0$ and $0 \leq s < k$,

$$M^{rk+s} = \underbrace{(A^r \otimes \cdots \otimes A^r)}_{k-s} \odot \underbrace{(A^{r+1} \otimes \cdots \otimes A^{r+1})}_{s},$$

where the exponents in the right-hand side sum to $rk + s$.

Proof: We prove the claim by induction on $rk + s$. It is true by definition for $rk + s = 1$. The induction step is easily proved by applying Equation (9). Indeed, $[(A^r \otimes (A^r \otimes \cdots \otimes A^r)) \odot (A^{r+1} \otimes \cdots \otimes A^{r+1})][((I \otimes \cdots \otimes I) \otimes (I \otimes \cdots \otimes I)) \odot A)]$ can be written as $(A^r \otimes \cdots \otimes A^r) \odot (A^{r+1} \otimes \cdots \otimes A^{r+1} \otimes (A^r A))$. The argument is correct also for limit cases $s = 0$ and $s = k - 1$. ∎

In particular we have the following.

Corollary 4.1: For A a square matrix and M defined by Equation (7),

$$M^k = A \otimes \cdots \otimes A.$$

Moreover, if A satisfies (A),(B),(C) the essential spectral radius of M is the kth root of the essential spectral radius of A.

Proof: The first part is a particular case of Proposition 4.1. From the properties of Kronecker product, we know the spectrum of M^k is composed of all possible products of k eigenvalues of A. Hence the largest eigenvalue in absolute value, different from 1, of the matrix M^k results to be $1^{k-1}\lambda$, where λ denotes the largest eigenvalue in absolute value, different from 1, of the matrix A. ∎

This also proves also that conditions (B) and (C) are verified for M when they are for A. If we take

$$A = 1/n \mathbf{1}\mathbf{1}^T, \tag{12}$$

of size n, then M^k is the matrix $1/n^k \mathbf{1}\mathbf{1}^T$ of size n^k with all identical elements. Thus we have a strategy converging exactly in k steps. We comment further on this example in the next section. Another property of M that will prove useful is stated in the next proposition.

Proposition 4.2: For A a square matrix, M defined by Equation (7), and any integers $r \geq 0$ and $0 \leq s < k$,

$$
M^{T^{rk+s}} M^{rk+s} = \underbrace{A^{T^r} A^r \otimes \cdots \otimes A^{T^r} A^r}_{k-s} \otimes
$$

$$
\otimes \underbrace{A^{T^{r+1}} A^{r+1} \otimes \cdots \otimes A^{T^{r+1}} A^{r+1}}_{s},
$$

where the sums of exponents is $rk + s$.

Proof: From Proposition 4.1, we know that $M^{rk+s} = (A^r \otimes \cdots \otimes A^r) \odot (A^{r+1} \otimes \cdots \otimes A^{r+1})$. Hence, by Equation (8), $M^{T^{rk+s}} = (A^{T^{r+1}} \otimes \cdots \otimes A^{T^{r+1}}) \odot (A^r \otimes \cdots \otimes A^r)$. These two expressions are multiplied using Equation (10). ∎

Now we would like to compute $\operatorname{Tr} M^{T^t} M^{t+1}$. This will be useful later when we will evaluate the performance of the block Kronecker strategy. We first need the following lemma.

Lemma 4.2: Let $B_0, B_1, \ldots, B_{k-1}$ be k square matrices of same dimensions. If $l \leq k$ is relatively prime to k, then

$$\operatorname{Tr} (B_0 \otimes B_1 \otimes \cdots \otimes B_{l-1}) \odot (B_l \otimes \cdots \otimes B_{k-1}) = \operatorname{Tr} B_0 B_l B_{2l} B_{3l} \cdots B_{(k-1)l},$$

where the indices are understood modulo k.

Proof: If we use Einstein's convention (repeated indices are summed over), we can write

$$
\begin{aligned}
&\operatorname{Tr} (B_0 \otimes B_1 \otimes \cdots \otimes B_{l-1}) \odot (B_l \otimes \ldots \otimes B_{k-1}) \\
=\ & [(B_0 \otimes B_1 \otimes \cdots \otimes B_{l-1}) \odot (B_l \otimes \cdots \otimes B_{k-1})]_{p,p} \\
=\ & (B_0)_{p_{k-l},p_0} (B_1)_{p_{k-l+1},p_1} \cdots (B_{l-1})_{p_{k-1},p_{l-1}} \\
& \quad (B_l)_{p_0,p_l} \cdots (B_k)_{p_{k-l-1},k} \\
=\ & (B_0)_{p_{k-l},p_0} (B_l)_{p_0,p_l} (B_{2l})_{p_l,p_{2l}} \\
& \quad (B_{3l})_{p_{2l},p_l} \cdots (B_{(k-1)l})_{p_{(k-2)l},p_{(k-1)l}} \\
=\ & \operatorname{Tr} B_0 B_l B_{2l} \ldots B_{(k-1)l},
\end{aligned}
$$

where $p = p_0 p_1 \ldots p_{k-1}$. ∎

Proposition 4.3: For A and M as defined above, if A is normal (i.e., $A^T A = A A^T$) then

$$\operatorname{Tr} M^{T^t} M^{t+1} = \operatorname{Tr} A^{T^t} A^{t+1}.$$

Proof: We know that $M^{T^t} M^t = A^{T^r} A^r \otimes \ldots \otimes A^{T^{r+1}} A^{r+1}$, if $t = rk + s$ for some $0 \leq s < k$. Hence

$$M^{T^t} M^{t+1} = (A^{T^r} A^r \otimes \cdots \otimes A^{T^{r+1}} A^{r+1})((I \otimes \cdots \otimes I) \odot A),$$

which by Equation (11) is equal to $(A^{T^r} A^r \otimes \cdots \otimes A^{T^{r+1}} A^{r+1}) \odot A^{T^r} A^{r+1}$. By Lemma 4.2, this matrix has the same trace as $A^{T^r} A^r \ldots A^{T^{r+1}} A^{r+1} A^{T^r} A^{r+1}$. As A^T and A commute, this is also the trace of $A^{T^t} A^{t+1}$. ∎

C. De Bruijn's graph

The communication graph of M is (a subgraph of) a de Bruijn graph, which has n^k vertices and arcs from any i to $ni, ni+1, ni+2, \ldots$ and $ni + k - 1$ (all modulo n^k). In particular, if A is given by Equation (12), then M is the adjacency matrix of a de Bruijn graph, normalized so as for every row to sum to one. This graph was introduced by de Bruijn [10] in 1946 and has been considered for efficient distribution of information in different context such as in parallel computing [23] and peer-to-peer networks [13]. This paper can be seen as an extension of this idea to consensus problems.

D. Design decentralisation

The process itself of convergence to consensus is decentralised, in the sense that every agent acts on its own. However the communication strategy (who talks to whom?) must be designed once for all beforehand. This can be done in centralised way, where a new external agent dispatch to every other agent their own strategy. This can also be done in a decentralised way, where every agent is attributed a number i between 0 and $N - 1$ and then finds the agents of number $\nu i, \nu i + 1, \ldots, \nu i + \nu - 1$. Achieving this in the most effective way is a problem of its own, and is not treated in this paper.

V. THE QUICKEST POSSIBLE STRATEGY

We have seen that starting from A with all identical entries, we get arbitrarily large matrices M converging in finite time k. If we write $N = n^k$ the dimension of M, this convergence time is $\log N / \log n = \log N / \log \nu$, where ν is the maximal in-degree of the graph of communication for M. We can see that no strategy, whether linear or not, whether time-invariant or not, can converge more rapidly. Indeed, to reach the average of the initial conditions, every agent must have information about all other agents, but it can only know ν other positions in one step of time, ν^2 in two steps of time, etc. Hence the propagation of information needs around $\log N / \log \nu$ steps to connect all agents. This reasoning is made rigorous in the following proposition.

Proposition 5.1: Let $M \in \mathbb{R}^{N \times N}$ such that $M \geq 0$. Let ν be defined as above. Then $M^k > 0$ implies $\nu^k \geq N$.

Proof: The fact $M^k > 0$ implies that for any pair of nodes (i, j) there exists in the graph $\mathcal{G}_M$ a path connecting i to j of length k. Hence there are at least N^2 paths of length k. Let now P_i denote the number of paths having length i. The previous consideration implies that $P_k \geq N^2$. On the other hand it is easy to see that $P_1 \leq \nu N$ and in general that $P_i \leq \nu^i N$ from which we get that $P_k \leq \nu^k N$. Hence $\nu^k N \geq N^2$ from which it results that $\nu^k \geq N$. ∎

The above proposition considers only the time-invariant case. An identical result can be found for the time-varying case, showing that there is no difference, in terms of speed of convergence toward the meeting point, between the time-invariant and the time-varying cases. This can be seen an *a posteriori* justification of our interest in the class of the time-invariant strategies.

A linear time-invariant strategy converges in finite time iff its essential spectral radius is 0. For a strategy converging in infinite time, the essential spectral radius is a good measure of the convergence to the average of the initial conditions, as already mentioned.

A. Comparison between block Kronecker strategy and Cayley strategy

In this subsection we propose a comparison of the block Kronecker strategy with another strategy whose underlying communication graph has limited max in-degree and exhibits strong symmetries: the Cayley strategy.
First we recall the concept of Cayley graph defined on Abelian groups [2], [1]. Let G be any finite Abelian group

(internal operation will always be denoted $+$) of order $|G| = N$, and let S be a subset of G containing zero. The Cayley graph $\mathcal{G}(G, S)$ is the directed graph with vertex set G and arc set

$$\mathcal{E} = \{(g, h) : h - g \in S\}.$$

Notice that a Cayley graph is always in-regular, namely the in-degree of each vertex is equal to $|S|$. Notice also that strong connectivity can be checked algebraically. Indeed, it can be seen that a Cayley graph $\mathcal{G}(G, S)$ is strongly connected if and only if the set S generates the group G, which means that any element in G can be expressed as a finite sum of (not necessarily distinct) elements in S. If S is such that $-S = S$ we say that S is inverse-closed. In this case the graph obtained is undirected.

A notion of Cayley structure can also be introduced for matrices. Let G be any finite Abelian group of order $|G| = N$. A matrix $P \in \mathbb{R}^{G \times G}$ is said to be a Cayley matrix over the group G if

$$P_{i,j} = P_{i+h, j+h} \qquad \forall\, i, j, h \in G.$$

It is clear that for a Cayley matrix P there exists a $\pi : G \to \mathbb{R}$ such that $P_{i,j} = \pi(i - j)$. The function π is called the generator of the Cayley matrix P. Notice how, if P is a Cayley matrix generated by π, then $\mathcal{G}_P$ is a Cayley graph with $S = \{h \in G : \pi(h) \neq 0\}$. Moreover, it is easy to see that for any Cayley matrix P we have that $P\mathbf{1} = \mathbf{1}$ if and only if $\mathbf{1}^T P = \mathbf{1}^T$. This implies that a Cayley stochastic matrix is automatically doubly stochastic. In this case the function π associated with the matrix P is a probability distribution on the group G. Among the multiple possible choices of the probability distribution π, one is particularly simple, namely $\pi(g) = 1/|S|$ for every $g \in S$.

Example 1: Let us consider the group $\mathbb{Z}_N$ of integers modulo N and the Cayley graph $\mathcal{G}(\mathbb{Z}_N, S)$ where $S = \{-1, 0, 1\}$. Notice that in this case S is inverse-closed. Consider the uniform probability distribution

$$\pi(0) = \pi(1) = \pi(-1) = 1/3$$

The corresponding Cayley stochastic matrix is given by

$$P = \begin{bmatrix} 1/3 & 1/3 & 0 & 0 & \cdots & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1/3 & 0 & 0 & 0 & \cdots & 1/3 & 1/3 \end{bmatrix}. \quad (13)$$

Notice that in this case we have two symmetries. The first is that the graph is undirected and the second that the graph is circulant. These symmetries can be seen in the structure of the transition matrix P that, indeed, turns out to be both symmetric and circulant [9].

Example 2: Let us now consider the group $\mathbb{Z}_N \times \mathbb{Z}_N$ and the Cayley graph $\mathcal{G}(\mathbb{Z}_N \times \mathbb{Z}_N, S)$ where $S = \{(1, 0); (0, 0); (0, 1)\}$. Again consider the uniform probability distribution

$$\pi((0, 0)) = \pi((1, 0)) = \pi((0, 1)) = 1/3$$

The corresponding Cayley stochastic matrix is given by the following block circulant matrix belonging to $\mathbb{R}^{N^2 \times N^2}$

$$P = \begin{bmatrix} P_1 & P_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & P_1 & P_2 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & P_1 & P_2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ P_2 & 0 & 0 & 0 & \cdots & 0 & 0 & P_1 \end{bmatrix} \quad (14)$$

where $P_1, P_2 \in \mathbb{R}^{N \times N}$ are such that

$$P_1 = \begin{bmatrix} 1/3 & 1/3 & 0 & \cdots & 0 & 0 \\ 0 & 1/3 & 1/3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1/3 & 0 & 0 & \cdots & 0 & 1/3 \end{bmatrix}, \; P_2 = \frac{1}{3}I. \quad (15)$$

This example can be generalized to the more general case of the discrete d-dimensional tori $\mathbb{Z}_N^d$, extensively studied in the literature regarding the peer-to-peer networks [21], [25].

Now we recall an interesting result regarding the essential spectral radius of the Cayley stochastic matrices. Assume that $P \in \mathbb{R}^{N \times N}$ is a Cayley stochastic matrix generated by a suitable π and assume that $|S| = \nu$, where S is as previously defined. Moreover assume that $0 \in S$. Notice that this last fact implies that $P_{ii} > 0, \forall i : 1 \leq i \leq N$. Then it follows that $\rho \geq 1 - C/N^{2/(\nu-1)}$, where $C > 0$ is a constant independent of S and N the number of agents. This result was proved in [6].

On the other side, the block Kronecker strategy constructed from any matrix A has essential spectral radius $|\lambda|^{1/k}$, where $|\lambda|$ is the essential spectral radius of A, as stated in Corollary 4.1.

If $0 < |\lambda| < 1$, then $|\lambda|^{1/k}$ behaves like $1 - \mu/k$ for large k and some μ. Recall that k is $\log N / \log n$. Hence this is better than abelian Cayley strategies.

In conclusion, block Kronecker strategies have a better essential spectral radius, hence a quicker convergence speed, than Cayley strategies. For the particular choice given by Equation (12), we converge in finite time, and this time is the smallest possible over all linear strategies with the same constrained degree.

B. Simulation result

As an illustration, we present a simulative comparison between the Cayley strategy and the block Kronecker strategy. The network considered consists of $N = 81$ agents. The matrix P for the Cayley strategy is the matrix (13), whereas the matrix M for the block Kronecker strategy is built starting from

$$A = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

with $n = 3$ and $k = 4$. The initial conditions has been chosen randomly inside the interval $[-50, 50]$. In both cases the in-degree is 3. Notice that, as depicted in Figure 1, the block Kronecker strategy reaches the average of the initial conditions in a finite number of steps whereas, the Cayley

strategy, after the same numbers of steps, is still far from converging toward the meeting point.
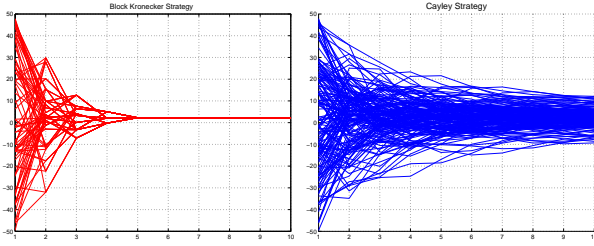

Fig. 1. The block Kronecker strategy (left) converges in finite time, while the Cayley strategy (right) has a relatively slow convergence

VI. LQR COST

In this section we want to evaluate the performance of the block Kronecker strategy according to the quadratic cost $J = J_1 + \gamma J_2$, where $J_1 = \mathbb{E} \sum_{t \geq 0} (x(t) - x(\infty))^T (x(t) - x(\infty))$ accounts for the quickness of convergence, $J_2 = \mathbb{E} \sum_{t \geq 0} (x(t+1) - x(t))^T (x(t+1) - x(t))$ limits the norm of the updates, and γ weights the respective importance of those two factors. Precisely we evaluate J for any block Kronecker strategy derived from a normal matrix A. Remember that the initial state $x(0)$ is supposed to be characterized by a identity covariance matrix. We start with a lemma which provides an upper and a lower bound for J_1.

Lemma 6.1: If A is a normal $n \times n$ matrix satisfying conditions (A),(B),(C), and ρ is the essential spectral radius of A, then the J_1 cost of the corresponding block Kronecker strategy of size n^k satisfies:

$$J_L \leq J \leq J_U,$$

where $J_L = N \frac{1 - (\text{Tr } (A^T A)/n)^k}{1 - \text{Tr } (A^T A)/n} - k$ and $J_U = J_L + \frac{k}{1-\rho^2}(\text{Tr } A^T A - 1)$.

Proof: Classical arguments lead to write:

$$
\begin{aligned}
J_1 &= \mathbb{E} \sum_{t \geq 0} (x(t) - x(\infty))^T (x(t) - x(\infty)) \\
&= \sum_{t \geq 0} \mathbb{E} \ (x(t) - x(\infty))^T (x(t) - x(\infty)) \\
&= \sum_{t \geq 0} \mathbb{E} \ \text{Tr } (x(t) - x(\infty))^T (x(t) - x(\infty)) \\
&= \sum_{t \geq 0} \mathbb{E} \ \text{Tr } (x(t) - x(\infty))(x(t) - x(\infty))^T \\
&= \sum_{t \geq 0} \text{Tr } (M^t - E)\mathbb{E}(x(0)x(0)^T)(M^t - E)^T \\
&= \sum_{t \geq 0} \text{Tr } (M^t - E)(M^t - E)^T
\end{aligned}
$$

with $E = 1/n^k \mathbf{1}\mathbf{1}^T$.

Now, $\text{Tr } (M^t - E)(M^t - E)^T = \text{Tr } M^t M^{t^T} - \text{Tr } E = \text{Tr } M^t M^{t^T} - 1$. When M is derived by block Kronecker product from a normal matrix A, this is equal to $(\text{Tr } (A^T A)^r)^{k-s}(\text{Tr } (A^T A)^{r+1})^s$ if $t = rk + s$, according to Proposition 4.2.

We get a lower bound on J_1 by summing only the first k terms:

$$
\begin{aligned}
J_1 &= \sum_{r \geq 0} \sum_{s=0}^{k-1} ((\text{Tr } (A^T A)^r)^{k-s}(\text{Tr } (A^T A)^{r+1})^s - 1) \\
&\geq \sum_{s=0}^{k-1} (\text{Tr } (A^T A)^0)^{k-s}(\text{Tr } (A^T A)^1)^s - k \qquad (16) \\
&= \sum_{s=0}^{k-1} n^{k-s}(\text{Tr } A^T A)^s - k,
\end{aligned}
$$

The last summation is a geometric series that can be evaluated, leading to the bound

$$J_1 \geq N \frac{1 - (\text{Tr } A^T A/n)^k}{1 - \text{Tr } A^T A/n} - k$$

This proves the left inequality in the claim.

For the right inequality, we find an upper bound on the terms neglected in the lower bound (16). As normal matrices can be diagonalized by a unitary transformation, the eigenvalues of $A^T A$, which we denote $1, \lambda_1, \lambda_2, \ldots, \lambda_{n-1}$, are precisely the squared module of the eigenvalues of A. In particular, $\rho^2 = \lambda_1$, and the trace of $(A^T A)^t$ is $1 + \sum \lambda_i^t$.

The terms neglected in the lower bound (16) are

$$\sum_{r \geq 1} \sum_{s=0}^{k-1} ((1 + \sum_i \lambda_i^r)^{k-s}(1 + \sum_i \lambda_i^{r+1})^s).$$

For every r, we bound every of the k terms by the highest (for which $s = 0$). Hence the neglected terms are bounded from above by:

$$\sum_{r \geq 1} ((1 + \sum_i \lambda_i^r)^k - 1) = k \sum_{r \geq 1} P(\lambda_1^r, \ldots, \lambda_{n-1}^r),$$

where P is a polynomial in the variables $\lambda_1, \ldots, \lambda_{n-1}$ with no independent term: all monomials have degree at least one. Now we can sum all corresponding monomials for $r = 1, 2, \ldots$: this is a geometric series of progression at most λ_1. Hence $\sum_{r \geq 1} P(\lambda_1^r, \ldots, \lambda_k^r)$ is at most $\frac{1}{1-\lambda_1} P(\lambda_1, \ldots, \lambda_k) = \frac{1}{1-\lambda_1}(\text{Tr } A^T A - 1)$.

Hence J_1 differs from our lower bound by at most $k \frac{1}{1-\lambda_1}(\text{Tr } A^T A - 1)$. ∎

Thus $J_1 = N \frac{1 - (\text{Tr } (A^T A)/n)^k}{1 - \text{Tr } (A^T A)/n} + \mathcal{O}(\log N)$. Now we estimate J_2.

Lemma 6.2: Under the assumptions of Lemma 6.1, if ρ_i denote the eigenvalues of A different from one,

$$2J_1 - N - \sum_i \frac{1}{1 - |\rho_i|^2} \leq J_2 \leq 2J_1 - N.$$

Proof: First we notice, adapting the first steps of the proof of the preceding proposition, that $J_2 = \sum \text{Tr } (M - I)^T (M^T)^t M^t (M - I)$, with I the identity. This involves terms of the form $(M^T)^{t+1} M^t$. More precisely,

$$J_2 = \sum_{t \geq 0} \text{Tr } (M^{T^{t+1}} M^{t+1} - M^{T^{t+1}} M^t -$$
$$- M^{T^t} M^{t+1} + M^{T^t} M^t)$$
$$= 2 \sum_{t \geq 0} (\text{Tr } M^{T^t} M^t - 1) - N -$$
$$- 2 \sum_{t \geq 0} (\text{Tr } M^{T^t} M^{t+1} - 1)$$

The first term of the last member is precisely $2J_1$, the last term is, thanks to Proposition 4.3, $2 \sum_{t \geq 0} (\text{Tr } A^{T^t} A^{t+1} - 1)$.

From Cauchy-Schwartz inequality applied to Frobenius norm, $\text{Tr } A^{T^t} A^{t+1} \leq \sqrt{\text{Tr } A^{T^{t+1}} A^{t+1} \text{Tr } A^{T^t} A^t} \leq \text{Tr } A^{T^t} A^t$.

Hence $\sum_{t \geq 0} (\text{Tr } A^{T^t} A^{t+1} - 1) \leq \sum_{t \geq 0} \sum_i \lambda_i^t = \sum_i \frac{1}{1 - \lambda_i}$, where, as argued in the proof of Lemma 6.1, $\lambda_i = |\rho_i|^2$. ∎

Hence,

$$J = N \Big((1 + 2\gamma) \frac{1 - (\text{Tr } (A^T A)/n)^k}{1 - \text{Tr } (A^T A)/n} - \gamma + \mathcal{O}(\log N/N) \Big).$$

Since the trace of $A^T A$ is the sum of squares of elements of A, we see that the coefficient of N (neglecting the $\mathcal{O}(\log N/N)$ term) is optimized by the matrix $A = 1/n \mathbf{1}\mathbf{1}^T$, whatever the value of γ is. In this case, the lower bound obtained on J_1 is exact, since only k terms are non-zero. The optimal cost is then

$$J = N \Big((1 + 2\gamma) \frac{1 - 1/N}{1 - 1/n} - \gamma + \mathcal{O}(\log N/N) \Big),$$

with $\nu = n$.

Hence there is here no trade-off between J_1 and J_2 among the family of block Kronecker strategies, in contrast with the general LQR theory.

Note that the optimal control strategy for unconstrained degree (every agent knows every position) is easily solved by a scalar algebraic Riccati equation, leading to the optimal cost $J = N(1 + \sqrt{1 + 4\gamma})/2$. If γ is small and n is large, then the optimal finite-time block Kronecker approaches the unbounded degree optimal solution with a cost approximately equal to $(1 + \gamma)N$.

VII. Conclusions

We have introduced a family of strategies for a consensus problem, whose graph of communication is de Bruijn's graph. We have shown that they can converge in finite logarithmic time, which is optimal. We have evaluated the LQR cost of these strategies, proving their quasi-optimality if the cost of update is small and the degree of the graph not too low.

This work can be extended in several directions, including:
- designing strategies valid for any N, not only exact powers of n;
- tackling the continuous-time case, where no deadbeat strategy can exist;
- estimating the LQR cost for Cayley strategies;
- finding strategies that minimize the LQR cost for any cost γ of the update;

VIII. Acknowledgements

References

[1] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures and Algorithms*, 5:271–284, 1994.

[2] L. Babai. Spectra of Cayley graphs. *Journal of Combinatorial Theory, Series B*, 27:180–189, 1979.

[3] R. W. Beard, J. Lawton, and F. Y. Hadaegh. A coordination architecture for spacecraft formation control. *IEEE Transaction on Control Systems Technology*, 9:777–790, 2001.

[4] P. Bhatta and N. E. Leonard. Stabilization and coordination of underwater gliders. In *41st IEEE Conference on Decision and Control*, 2002.

[5] R. Carli, F. Fagnani, A. Speranzon, and S.Zampieri. Communication constraints in coordinated consensus problem. In *American Control Conference (ACC '06)*, June 2006.

[6] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in the average consensus problem. *Accepted for publication to Automatica*.

[7] R. Carli, F. Fagnani, and S. Zampieri. On the state agreement with quantized information. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS), Kyoto, Japan*, pages 1500–1508, July 2006.

[8] J. Cortes, S. Martinez, and F. Bullo. Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions. *IEEE Trans. Automat. Control*.

[9] P. J. Davis. *Circulant matrices*. A Wiley-Interscience Publication, Pure and Applied Mathematics. John Wiley & Sons, New York-Chichester-Brisbane, 1979.

[10] N. G. de Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 49:758–764, 1946.

[11] D. V. Dimarogonas and K. J. Kyriakopoulos. On the rendezvous problem for multiple nonholonomic agents. *IEEE Transactions on automatic control*, 52(5):916–922, 2007.

[12] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Trans. Automat. Control*, 49(9):1465–1476, 2004.

[13] P. Fraigniaud and P. Gauron. D2b: A de Bruijn based content-addressable network. *Theor. Comput. Sci.*, 355(1):65–79, 2006.

[14] Y. Hatano and M. Meshabi. Agreement of random networks. In *IEEE Conference on Decision and Control*, 2004.

[15] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Automat. Control*, 48(6):988–1001, 2003.

[16] A. Kashyap, T. Basar, and R. Srikant. Consensus with quantized information updates. In *Proceedings of CDC Conference, san Diego*, 2006.

[17] R. H. Koning, H. Neudecker, and T. Wansbeek. Block kronecker products and the vecb operator. *Linear algebra and its applications*, 149:165–184, 1991.

[18] Z. Lin, B. Francis, and M. Maggiore. Necessary and sufficient graphical conditions for formation control of unicycles. *IEEE Transactions on automatic control*, 50(1):121–127, 2005.

[19] M. Mazo, A. Speranzon, K. H. Johansson, and X. Hu. Multi-robot tracking of a moving object using directional sensors. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2004.

[20] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Automat. Control*, 49(9):1520–1533, 2004.

[21] S. Ratnasamy, P. Francis, M. Handley, R. M. Karp, and S. Shenker. A scalable content-addressable network. In *SIGCOMM.*, pages 161–172, 2001.

[22] W. Ren, K. L. Moore, and Y. Chen. Necessary and sufficient graphical conditions for formation control of unicycles. *ASME Journal of Dynamic Systems, Measurement, and Control*, to appear, 2007.

[23] M. R. Samatham and D. K. Pradhan. The de Bruijn multiprocessor network: A versatile parallel processing and sorting network for vlsi. *IEEE Trans. Comput.*, 38(4):567–581, 1989.

[24] A. Speranzon, C. Fischione, and K. Johansson. Distributed and collaborative estimation over wireless sensor networks. In *Proceedings of the IEEE Conference on Decision and Control (CDC'06)*, pages 1025–1030, December 2006.

[25] I. Stoica, R. Morris, D. R. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM.*, pages 161–172, 2001.

[26] H. G. Tanner, A. Jadbabaie, and G. J. Pappas. Stable flocking of mobile agents, part i: fixed topology. In *IEEE Conference on decision and control*, 2003.

[27] H. G. Tanner, A. Jadbabaie, and G. J. Pappas. Stable flocking of mobile agents, part ii: dynamic topology. In *IEEE Conference on decision and control*, 2003.

[28] H. G. Tanner, S. G. Loizou, and K. J. Kyriakopoulos. Nonholonomic navigation and control of cooperating mobile manipulators. *IEEE Transactions on robotics and automation*, 19(1):53–64, 2003.