WHAT TO DO ABOUT NOISY CONSENSUS?

Victor Solo, Marc Piggott

School of Electrical Engineering and Telecommunications
University of New South Wales
Sydney, AUSTRALIA
email: v.solo@unsw.edu.au

ABSTRACT

There is an achilles heel underlying the consensus literature. It has been known for some time that measurement noise causes the explosive growth of the consensus mode. Here we state the noise problem; characterise the behaviour of the noisy consensus system including its drift to ∞; critique the existing remedies and develop a new remedy.

Index Terms— consensus, adaptive.

1. INTRODUCTION

While there has been a considerable interest in consensus algorithms over the past decade [1],[2] the effect of noise has not received a lot of treatment.

Following early work of [3] interest in consensus type algorithms was stimulated e.g. by [4],[2],[5],[6] with related developments in gossip algorithms, rendezvous problems, and adaptive distributed estimation [1],[7].

But early on a problem was noted when there is noise [8]. In that case the consensus mode explodes. No resolution was presented in [8] however. Subsequently [9],[10] showed how to deal with the problem by using decaying gains.

However in the adaptive signal processing and adaptive control literatures [11],[12],[13] decaying gains have been dismissed since they cause the adaptive algorithm to lose the ability to adapt i.e. to lose the ability to track.

So the question is how to deal with the problem of noise when gains are fixed. [14] present a solution which assumes a very special measurement structure and only applies to very special topologies. In this paper we propose a very general solution; we develop a modification to the consensus control law which produces bounded behaviour.

In section 2 we review the deterministic consensus theory. In section 3 we formulate the problem of noise and characterise its impact. In section 4 we present our new approach. Conclusions are in section 5.

Notation & Acronyms. $iid(0, \sigma^2)$ denotes a sequence of independent identically distributed random variables each with zero mean and variance σ^2. *AR(p)* denotes an autoregressive process of order p. 1_* denotes a vector of $1s$.

This work was partly supported by the Australian Research Council

2. DETERMINISTIC CONSENSUS

Consider a network of N nodes. At node i agent i updates a state $x_i(t)$ according to dynamics dependent on the (noise free measurements of the) states of nearby agents

$$\dot{x}_i(t) = \beta\Sigma_j a_{ij}(x_j - x_i)$$

where $A = [a_{ij}]$ is a symmetric adjacency matrix of non-negative weights. We exclude self loops by taking $a_{kk} = 0$. If the adjacency matrix is binary and undirected then it is symmetric [15].

Introduce the 'degree' vector $d_i = \Sigma_j a_{ij} = (A1)_i$ and $D = diag(d_i)$. We then have a graph Laplacian $L = D - A$ obeying $L1 = D1 - A1 = d - d = 0$ and which is symmetric.

Then we can write the equations in vector form as follows

$$\begin{aligned} \dot{x}_i(t) &= \beta\Sigma_j a_{ij}x_j - \beta d_i x_i \\ \Rightarrow \dot{x} &= \beta Ax - \beta Dx \\ &= -\beta Lx \end{aligned} \quad (2.1)$$

Note the following positivity property of the Laplacian.

$$\begin{aligned} & \Sigma_i\Sigma_j(x_i - x_j)^2 a_{ij} \\ = \; & \Sigma_i\Sigma_j a_{ij}(x_i^2 - 2x_i x_j + x_j^2) \\ = \; & \Sigma_i d_i x_i^2 - 2\Sigma_i\Sigma_j x_i x_j a_{ij} + \Sigma_j d_i x_j^2 \\ = \; & 2(x^T Dx - x^T Ax) \\ = \; & 2x^T Lx \end{aligned}$$

This shows that the Laplacian is positive semi-definite [15] which means its eigenvalues are ≥ 0.

We introduce the following assumption.

Assumption **C**. The network is connected.

Then the 0 eigenvalue of L has multiplicity 1 [15].

To analyse the dynamic system we introduce the eigenvector decomposition (EVD) of L. Let L have possibly repeated eigenvalues $\lambda_o = 0, 0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{N-1}$ and introduce $\Lambda = diag[\lambda_i]$. Then the EVD is

$$L = Q\Lambda Q^T = \Sigma_0^{N-1}\lambda_i q_i q_i^T = \Sigma_1^{N-1}\lambda_i q_i q_i^T, Q^T Q = I$$

$Q = [q_o, \cdots, q_{N-1}]$ is the matrix of eigenvectors. Since $QQ^T = I = Q^T Q$ we have $q_i^T q_j = \delta_{i,j}$. Note that $L1_* = 0 \Rightarrow q_o = 1_* \frac{1}{\sqrt{N}}$.

The solution to (2.1) is then

$$\begin{aligned} x(t) &= e^{-\beta L t}x(0) \\ &= \Sigma_0^{N-1} e^{-\beta\lambda_i t} q_i q_i^T x(0) \\ &\rightarrow q_o q_o^T x(0) \text{ as } t \rightarrow \infty \\ &= 1_* \frac{1_*^T x(0)}{N} \end{aligned}$$

This demonstrates consensus in that all nodes converge to the same solution $1_*^T x(0)/N$.

We sum this up as follows.

Result O. Consider the system (2.1) with L symmetric and connected. Then all states $x_i(t) \rightarrow 1_*^T x(0)/N$ as $t \rightarrow \infty$.

Various versions of this type of result can be found in the references cited earlier.

3. NOISY CONSENSUS

Now we introduce measurement noise. The update becomes

$$\dot{x}_i(t) = \beta\Sigma_j a_{ij}(x_j + \epsilon_{ji} - x_i) \tag{3.1}$$

where ϵ_{ji} are node-wise independent white noises and at each node are $iid(0, \sigma_{ji}^2)$. More formally we introduce iid Brownian motions $W_{ji}(t)$ with increment variances $Var(dW_{ji}(t)) = var(\epsilon_{ji}(t)dt) = \sigma_{ji}^2 dt$. Since $a_{ii} = 0$ there is no self noise ϵ_{ii}.

In vector form we get

$$\begin{aligned} \dot{x} &= -\beta L x + \beta n \qquad (3.2) \\ n_i &= \Sigma_j a_{ij}\epsilon_{ji} \\ var(n_i dt) &= \Sigma_j a_{ij}^2 \sigma_{ji}^2 dt = \sigma_{n,i}^2 dt \end{aligned}$$

So ndt has diagonal covariance matrix $D_n dt = diag[\sigma_{n,i}^2]dt$.

Note that if we have a binary adjacency matrix and all measurement noise variances are the same i.e. $\sigma_{ji}^2 = \sigma^2$ then

$$\sigma_{n,i}^2 = \Sigma_j a_{ij}\sigma^2 = d_i \sigma^2$$

It is crucial to point out here that our noise model differs in an important way from that in [8]. Certainly they work in discrete time whereas we work in continuous time. But that is not the issue here. We model measurement noise in a natural way by adding noise directly to the states. This is not what [8] does. They do not model measurement noise but rather just add a noise directly to (2.1) giving indeed (3.2). But their noise is unaffected by the structure of the graph; whereas our noise is intimately affected by the graph structure. This not only makes their system different but makes their analysis simpler whereas our analysis is more difficult.

To analyse what happens in our case we take a modal view. Introduce the modes $\xi_j = q_j^T x$. Multiplying through the state equation by q_j^T we get

$$\begin{aligned} \dot{\xi}_j &= -\beta\lambda_j \xi_j + \beta\nu_j \\ \nu_j &= q_j^T n \\ var(\nu_j dt) &= q_j^T D_n q_j dt = \sigma_{\nu,j}^2 dt \end{aligned}$$

This means that for $j \neq 0$ the modes are continuous time AR(1) processes (aka Ornstein Uhlenbeck processes). We note that since $\lambda_o = 0$ the 0th mode $\xi_o = \frac{1_*^T x}{\sqrt{N}}$ obeys

$$\dot{\xi}_o = \beta\nu_o = \beta 1_*^T n/\sqrt{N}$$

This leads to the following result.

Result I. Noisy Modal Behaviour. For the noisy consensus system (3.1) with network state $x(t)$ and modes $\xi_j(t) = q_j^T x(t)$,

(a) ξ_o is an integrated white noise i.e. a Brownian motion and so explodes e.g. its variance is

$$var(\xi_o) = \beta^2\sigma_{\nu,o}^2 t \text{ where } \sigma_{\nu,o}^2 = \frac{\Sigma_i \sigma_{n,i}^2}{N}$$

(b) The other modes do not explode but are continuous time AR(1) processes and have steady state variances

$$var(\xi_j) = \frac{\beta\sigma_{\nu,j}^2}{2\lambda_j}$$

To understand what happens to the network in the presence of this explosion we need to shift our focus and look at consensus from a different point of view. In the presence of noise what we can hope is that the nodes stay nearby each other. To measure this we compute the consensus distance (c.f. [8] in discrete time)

$$\Delta = \sqrt{\Sigma_i \Sigma_j (x_i - x_j)^2}$$

Now we have the following result.

Lemma D.

$$\Delta^2 = 2N(x^T x - \xi_o^2) = 2N\Sigma_1^{N-1}\xi_j^2$$

Proof. We have

$$\begin{aligned} \Sigma_i \Sigma_j (x_i - x_j)^2 &= \Sigma_i \Sigma_j (x_i^2 - 2x_i x_j + x_j^2) \\ &= 2N\Sigma_i x_i^2 - 2N\xi_o^2 \\ &= 2N(x^T x - \xi_o^2) \\ &= 2N(x^T Q Q^T x - \xi_o^2) \\ &= 2N(\xi^T \xi - \xi_o^2) \\ &= 2N\Sigma_1^{N-1}\xi_j^2 \end{aligned}$$

where we have used the fact that $\xi_o = \Sigma_i x_i/\sqrt{N}$.

Now we have the following result.

Result II. Noisy consensus. For the noisy consensus system (3.1) with network state $x(t)$ and modes $\xi_j(t) = q_j^T x(t)$,

(a) The consensus distance has finite mean: $E(\Delta) \leq \sqrt{E(\Delta^2)}$ where

$$E(\Delta^2) = \beta\Sigma_1^{N-1}\frac{\sigma_{\nu,j}^2}{2\lambda_j}$$

(b) The network mean $\bar{x}(t) = 1_*^T x/N = \xi_o/\sqrt{N}$ is a Brownian motion.

(c) $var(x_i - \bar{x}(t)) \leq \beta(\Sigma_1^{N-1}\frac{\sigma_{\nu,j}}{\sqrt{2\lambda_j}})^2$.

Proof. (a),(b) follow from the calculations above. For (c) consider that

$$\begin{aligned}
& E(x - \frac{1_*1_*^T x}{N})(x - \frac{1_*1_*^T x}{N})^T \\
= \; & E(\Sigma_0^{N-1} q_j q_j^T x - q_o q_o^T x)(\cdot)^T \\
= \; & E(\Sigma_1^{N-1} q_j \xi_j)(\Sigma_1^{N-1} q_j \xi_j)^T \\
= \; & \Sigma_1^{N-1}\Sigma_1^{N-1} q_j q_k E(\xi_j \xi_k)
\end{aligned}$$

Let u_i be a vector of $0s$ but with a 1 in position i. Then note that $|u_i^T q_j| \leq \| u_i \| \| q_j \| = 1$. Using this and noting $x_i = u_i^T x$ we find

$$\begin{aligned}
var(x_i - \bar{x}(t)) &= \Sigma_1^{N-1}\Sigma_1^{N-1} u_i^T q_j u_i^T q_k E(\xi_j \xi_k) \\
&\leq \Sigma_1^{N-1}\Sigma_1^{N-1} |u_i^T q_j||u_i^T q_k||E(\xi_j \xi_k)| \\
&\leq \Sigma_1^{N-1}\Sigma_1^{N-1}\sqrt{E(\xi_j^2)}\sqrt{E(\xi_k^2)} \\
&= (\Sigma_1^{N-1}\sqrt{E(\xi_j^2)})^2
\end{aligned}$$

and using Result Ib delivers the result.

Note that the size of $E(\Delta)$ is determined by measurement noise and mode time constants the only free parameter being β.

We can now summarize the behaviour of the noisy network.

(i) (c) shows that the nodes fluctuate with bounded variance around the network mean

(ii) (b) shows the network mean is a Brownian motion and drifts randomly and unboundedly.

(iii) (a) is consistent with (b),(c) in that the relative distance between the nodes remains bounded.

This overall behaviour is clearly unsatisfactory.

4. REVERTING CONSENSUS

The main previous remedy has been to abandon fixed gains and consider decaying gains [9],[10].

As mentioned in the introduction this is not an acceptable solution since the algorithm then loses it ability to track. Also as noted earlier [14] have found a special kind of noisy measurement regime suited to a few applications where fixed gains can be used and drifting can be avoided.

Here we seek a general resolution. In particular we now investigate the following very simple scheme. We add a self reverting term to each update.

The new scheme is

$$\dot{x}_i(t) = -\alpha\beta x_i + \beta\Sigma_j a_{ij}(x_j + \epsilon_{ji} - x_i) \qquad (4.1)$$

where $\alpha > 0$ and other quantities are as in section 3. Note the scaling of the reversion gain by β. This simplifies design and subsequent interpretation. In vector form this becomes

$$\dot{x} = -\alpha\beta x + \beta L x + \beta n$$

We now proceed to analyse this system in a similar manner to section 3.

The modal equations are now

$$\dot{\xi}_j = -\beta(\alpha + \lambda_j)\xi_j + \beta\nu_j$$

Note that the 0th mode is no longer a Brownian motion but an AR(1) with bounded variance.

We now have the following results.

Result III. Noisy Modal Behaviour. For the reverting consensus system (4.1), the modes do not explode but have steady state variances

$$var(\xi_j) = \beta\frac{\sigma_{\nu,j}^2}{2(\alpha + \lambda_j)}, 0 \leq j \leq N-1$$

Result IV. Noisy consensus. For the reverting consensus system (4.1),

(a) The consensus distance has finite mean

$$E(\Delta) \leq \sqrt{E(\Delta^2)} = \sqrt{\beta\Sigma_1^{N-1}\frac{\sigma_{\nu,j}^2}{2(\alpha + \lambda_j)}}$$

(b) The network mean $\bar{x}(t) = 1_*^T x/N = \xi_o/\sqrt{N}$ is an AR(1) process with variance $\beta\frac{\sigma_{\nu,o}^2}{2\alpha}$ where again $\sigma_{\nu,o}^2 = \frac{1}{N}\Sigma_i\sigma_{n,i}^2$.

(c) $var(x_i - \bar{x}(t)) \leq \beta(\Sigma_1^{N-1}\frac{\sigma_{\nu,j}}{\sqrt{2(\alpha+\lambda_j)}})^2$.

Proof. Follows much as before.

Now that we have stopped the network drifting unboundedly the question now is how to choose α.

Consider the case of a binary adjacency matrix where all node measurement variances are the same $= \sigma^2$. Then $\sigma^2_{\nu,o} = \sigma^2 \frac{1}{N}\Sigma d_i - d_{av}$ the average degree. Also $\sigma^2_{\nu,j} \leq d_{max}\sigma^2$.

We should like the average consensus distance to be much smaller than the fluctuation standard deviation of the mean i.e.

$$\frac{1}{N}\Sigma_1^{N-1}\frac{d_{av}\sigma^2}{2(\alpha+\lambda_j)} = \epsilon^2\frac{d_{max}\sigma^2}{2\alpha}$$

where e.g. $\epsilon = .1$ is a small fraction. Cancelling σ^2 and subsuming d_{max}/d_{av} into ϵ gives

$$\frac{1}{N}\Sigma_1^{N-1}\frac{\alpha}{\alpha+\lambda_j} = \epsilon^2$$

We illustrate with some examples.

Example RING. Ring with N nodes; [15][section 5.3],[8]. The eigenvalues of the Laplacian are given by

$$\lambda_k = 2(1 - cos(\frac{2\pi k}{N})), k = 0, 1, 2, \cdots$$

So we require

$$\frac{1}{N}\Sigma_1^{N-1}\frac{\alpha}{\alpha + 2(1 - cos(\frac{2\pi k}{N})} = \epsilon^2$$

For large N the sum can be approximated by an integral giving

$$\int_0^1 \frac{\alpha}{\alpha + 2(1 - cos(2\pi\theta))}d\theta = \frac{\alpha}{2\pi}\int_0^{2\pi}\frac{d\phi}{\alpha + 2 - 2cos(\phi)}$$

This integral is known, so we get

$$\frac{\alpha}{\sqrt{(\alpha+2)^2 - 2^2}} = \epsilon^2 \equiv \alpha^2 = \epsilon^4[\alpha^2 + 2\alpha] \equiv \alpha \approx 2\epsilon^4$$

5. CONCLUSIONS

We have reviewed the deterministic consensus algorithm and then characterised its behaviour in the presence of measurement noise. This behaviour is catastophic. While the nodal states remain near each other their centre is a random walk and drifts off to ∞. We have presented a simple local remedy which stops the unbounded drift by adding a reverting term at each node. In future work we will investigate the possibility of an optimal remedy.

6. REFERENCES

[1] A.H. Sayed, *Adaptation, Learning and Optimization over Networks*, Now, Hanover, MA, 2014.

[2] W. Ren, R.W. Beard, and E.M. Atkins, "A survey of consensus problems in multi-agent coordination", in *2005 American Control Conference*, 2005, pp. 1859–1864.

[3] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms", *IEEE Trans. Automat. Contr.*, vol. 31, pp. 803–812, 1986.

[4] D. P. Spanos and R. M. Murray, "Distributed sensor fusion using dynamic consensus.", in *Proc. 16th IFAC World Congress*. IFAC, 2005, pp. –.

[5] R.O. Saber and R.M. Murray, "Consensus protocols for networks of dynamic agents", in *Proc. Amer. Control Conf.*, 2003, pp. 951 – 956.

[6] A. Jadbabaie and S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules", *IEEE Trans. Autom. Control*, vol. 48, pp. 988–1001, 2003.

[7] I.D. Schizas, G. Mateos, and G.B. Giannakis, "Distributed lms for consensus-based in-network adaptive processing", *IEEE Trans. Signal Proc.*, vol. 57, pp. 2365–, 2009-2382.

[8] S. Boyd L. Xiao and S.-J. Kim, "Distributed average consensus with least-mean-square deviation", *J. Parallel and Distributed Computing*, vol. 67, pp. 33–46, 2007.

[9] M. Huang and J. Manton, "Coordination and consensus of networked agents with noisy measurements: Stochastic algorithms and asymptotic behaviour", *SIAM Jl. Control*, vol. 48, pp. 134–161, 2009.

[10] S. Kar and J.M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs", *IEEE Journal on Selected Topics on Signal Processing*, vol. . 5:4, pp. 674–690, 2011.

[11] O. Macchi, *Adaptive Signal Processing, The Least Mean Squares Approach*, J. Wiley, New York, 1995.

[12] A.H. Sayed, *Adaptive Filters*, J. Wiley, New York, 2008.

[13] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance*, Prentice Hall, New Yrok, 1995.

[14] T. Li, F. Wu, and J-F. Zhang, "Multi-agent consensus with relative-state-dependent measurement noises", *IEEE Trans. Autom. Contr.*, vol. 59, pp. 2463– 2468, 2014.

[15] P.van Mieghem, *Graph Spectra for Complex Networks*, Cambridge Univeristy Press, Cambridge, UK, 2011.