Convergence of a Class of Decentralized Beamforming Algorithms

James Bucklew and William A. Sethares *

March 22, 2007

Abstract

One of the key issues in decentralized beamforming is the need to phase-align the carriers of all the sensors in the network. Recent work in this area has shown the viability of certain methods that incorporate single-bit feedback from a beacon. This paper analyzes the behavior of the method (showing conditions for convergence in distribution and also giving a concrete way to calculate the final distribution of the convergent ball) and then generalizes the method in three ways. First, by incorporating both negative and positive feedback it is possible to double the convergence rate of the algorithm without adversely effecting the final variance. Second, a way of reducing the amount of energy required (by reducing the number of transmissions needed for convergence) is shown; its convergence and final variance can also be conveniently described. Finally, a wideband analog is proposed that operates in a decentralized manner to align the time delay (rather than the phase) between sensors.

EDICS: SAS-ADAP, COM-NETW

Permission to publish this abstract separately is given.

Keywords: adaptive algorithms, sensor networks, phase alignment, time delay estimation, wideband, final variance, small stepsize algorithms

*Both authors are with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI 53706-1691 USA. 608-262-5669 bucklew@ece.wisc.edu, sethares@ece.wisc.edu

1 Introduction

A collection of sensors are scattered in unknown locations. The sensors wish to cooperatively transmit a common message signal as efficiently as possible using a beamforming method in order to be energy efficient. Significant gains occur when exploiting distributed beamforming [7, 9] because of the improved signal-to-noise ratio at the receiver: while the received signal magnitude increases with the number of transmitters J, the SNR increases with J^2. Since the total amount of power transmitted increases linearly in J, this represents an J-fold increase in energy efficiency. A key issue in the use of distributed beamforming systems [1] is that the phases of the carriers must be synchronized throughout the network.

A recently proposed scheme [11] accomplishes this phase synchronization using single-bit feedback from a base station. Each sensor broadcasts within each timeslot, perturbing the phase of its carrier slightly from the previous timestep. The base station replies with a signal that indicates whether the received signal is more (or less) coherent than the previous time. The sensors respond in the obvious way: if the signal is improved they keep the new phase while if the signal worsened they revert to their old phase. This scheme is shown, under certain conditions in [12], to asymptotically achieve perfect phase coherence in the noise-free case. However, even tiny disturbances (which may arise physically from thermal noise, from unmodeled dynamics, or from interference with other nearby communications systems) cause convergence to a ball about the correct answer (and not to the correct answer itself). Analysis in the present paper is able to concretely describe both the rate of convergence and the distribution of this convergent ball.

This paper begins by showing in Sect. 2 how the single-bit feedback mechanism for phase alignment can be written as a "small stepsize" μ-dependent al-

gorithm with a discontinuous update term [4, 10]. The discontinuity arises because the sensors either accept or reject the most recent phase change based on the single-bit feedback. Sect. 3 applies the analytical techniques of [5] and [6] to examine the convergence of the algorithm in terms of a related ordinary differential equation (ODE). An extension of these results, detailed in the Appendix, allows derivation of the asymptotic variance. This concretely describes the final distribution of the algorithm about its equilibrium.

The method of [11, 12] either updates or freezes the estimated phases at each timestep. Sect. 4 observes that it may be possible to do better than to freeze the updates: if adding a small number makes things worse, then most likely subtracting a small number would improve things. This is an old idea [3, 8] in signal processing, and a "signed" algorithm for the decentralized phase alignment problem that uses both positive and negative feedback is described and analyzed: it is shown to converge twice as fast as the original, with the same final distribution.

One key requirement in a sensor network system is energy efficiency. Sect. 5 proposes the ρ-percent method in which only a subset of the sensors transmit at each timestep. Analysis shows that the savings in the number of transmissions (and hence in the energy) can be significant. Since the subset is chosen randomly at each epoch, there is no need to coordinate the sensors, and the method remains decentralized.

Finally, the above methods are inherently limited to aligning the phases of narrowband transmissions with a carrier of known frequency. Sect. 6 proposes and analyzes an analogous algorithm that operates with wideband signals by aligning the received signal in time. The method remains decentralized and incorporates only single-bit feedback from the beacon.

2 Algorithm Statement

Let $\hat{\phi}_{j,n}$ be the phase of the carrier signal at sensor/transmitter j at timestep n and let α_j be the phase difference due to the (unknown) distance between the base station and sensor j. At each timeslot, each sensor randomly perturbs its phase by a small amount $\mu\gamma_{j,n}$. Further suppose that the received signal at the base station at iteration n is corrupted by a Gaussian noise g_n with mean zero and variance σ_t^2.

The algorithm described above can be written

$$\hat{\phi}_{j,n+1} = \hat{\phi}_{j,n} + \mu\gamma_{j,n} 1_{\{g_{n+1}+\sum_{r=1}^{J}\cos(\hat{\phi}_{r,n}+\mu\gamma_{r,n}-\alpha_r) > g_n+\sum_{r=1}^{J}\cos(\hat{\phi}_{r,n}-\alpha_r)\}} \qquad (1)$$

for $j = 1, 2, \ldots, J$ where $1_{\{A\}}$ is an indicator function taking on value one if A is true and is zero otherwise. The sum of cosines terms represent the received carrier wave and take on maxima when the $\hat{\phi}$ are phase-aligned. Thus the indicator function is unity if the perturbed phases $\hat{\phi}_{r,n} + \mu\gamma_{r,n}$ are better aligned than the unperturbed phases, and is zero otherwise. The goal of the algorithm is to drive the $\hat{\phi}$ terms to a value at which the sum is maximum, which occurs when all of the terms are maximized, i.e., when $\hat{\phi}_r$ is equal to α_r.

For the purpose of analysis, it is more convenient to rewrite the algorithm in "error system" form by letting $\phi_{j,n} = \hat{\phi}_{j,n} - \alpha_j$. Subtracting α_j from both sides of (1) gives

$$\phi_{j,n+1} = \phi_{j,n} + \mu\gamma_{j,n} 1_{\{g_{n+1}+\sum_{r=1}^{J}\cos(\phi_{r,n}+\mu\gamma_{r,n}) > g_n+\sum_{r=1}^{J}\cos(\phi_{r,n})\}}. \qquad (2)$$

Suppose that the i.i.d. perturbation random variables $\{\gamma_{j,n}\}$ are chosen to have a symmetric distribution (about zero) with finite variance σ_γ^2. Then for small μ (keeping just the first terms in the Taylor series), $\cos(\phi_{r,n} + \mu\gamma_{r,n})$ can be approx-

imated by $\cos(\phi_{r,n}) - \mu\gamma_{r,n}\sin(\phi_{r,n})$. The algorithm is then

$$\phi_{j,n+1} = \phi_{j,n} + \mu\gamma_{j,n}1_{\{\tilde{g}+\sum_{r=1}^{J}\gamma_{r,n}\sin(\phi_{r,n})<0\}} \tag{3}$$

where $\tilde{g}$ is normal with mean zero and variance $2\sigma_t^2/\mu^2$.

In order to investigate the behavior of the algorithm, observe that convergence of ϕ_r to zero is equivalent to convergence of the phase estimates $\hat{\phi}_r$ to their unknown values α_r. The analysis requires some technical machinery, which is described fully in Appendix A. The basis of the analytical approach is to find an ordinary differential equation (ODE) that accurately mimics the behavior of the algorithm for small values of μ. Studying the ODE then gives information regarding the behavior of the algorithm. For example, if the ODE is stable, the algorithm is convergent (at least in distribution). If the ODE is unstable, the algorithm is divergent. The approach grows out of results in [4] and [5], which are themselves based on the techniques of [6]. The approach is conceptually similar to stochastic approximation but its assumptions (and hence conclusions) are somewhat different. First, the stepsize μ in (3) is fixed, unlike in stochastic approximations where the stepsize is required to converge to zero [10]. Thus the algorithms do not necessarily converge to a fixed vector; rather, they converge in distribution. Moreover, the analysis is capable of delivering concrete values for the convergent distribution; as far as we know, this is not possible with other methods. Second, no continuity assumptions need to be made on the update terms; this is crucial because of the discontinuity caused by the indicator function in (3), and is also more general than other methods that require differentiability of the update term.

3 Convergence Analysis

This section applies the weak convergence analysis of Appendix A to the algorithm (3). If J is of reasonable size ($J = 10$ is used in many of the simulations), the Central Limit Theorem shows that

$$\frac{1}{\sqrt{J-1}} \sum_{r=1(r\neq j)}^{J} \gamma_{r,n} \sin(\phi_{r,n}) \approx Z$$

where Z is a zero mean normal random variable with variance $\frac{\sigma_\gamma^2 ||\sin(\phi_{n/(j)})||^2}{J-1}$ where $||\sin(\phi_{n/(j)})||^2 = \sum_{r=1(r\neq j)}^{J} \sin^2(\phi_{r,n})$. To calculate the limiting ODE (the $\hat{H}$ of (10)), it is necessary to smooth the update by taking the expectation

$$\begin{aligned}
&E[\gamma_{j,n} 1_{\{\tilde{g}+\sum_{r=1}^{J} \gamma_{r,n} \sin(\phi_{r,n})<0\}}] \\
&= E[\gamma_{j,n} 1_{\{\frac{1}{\sqrt{J-1}}\tilde{g}+\frac{1}{\sqrt{J-1}}\sum_{r=1,r\neq j}^{J} \gamma_{r,n} \sin(\phi_{r,n})+\frac{1}{\sqrt{J-1}}\gamma_{j,n} \sin(\phi_{j,n})<0\}}] \\
&\approx E[\gamma_{j,n} 1_{\{\frac{1}{\sqrt{J-1}}\tilde{g}+Z+\frac{1}{\sqrt{J-1}}\gamma_{j,n} \sin(\phi_{j,n})<0\}}] \\
&= E_\gamma E[\gamma_{j,n} 1_{\{\frac{1}{\sqrt{J-1}}\tilde{g}+Z+\frac{1}{\sqrt{J-1}}\gamma_{j,n} \sin(\phi_{j,n})<0\}} | \gamma_{j,n} = \gamma]] \\
&= E_\gamma[\gamma E[1_{\{\tilde{Z}+\frac{1}{\sqrt{J-1}}\gamma \sin(\phi_{j,n})<0\}}]]
\end{aligned}$$

where $\tilde{Z} = Z + \frac{1}{\sqrt{J}}\tilde{g}$ is normal with zero mean and variance $\frac{1}{J-1}(\sigma_\gamma^2 ||\sin(\phi_{n/(j)})||^2 + 2\sigma_t^2/\mu^2)$. Also,

$$\begin{aligned}
&E[1_{\{\tilde{Z}+\frac{1}{\sqrt{J-1}}\gamma \sin(\phi_{j,n})<0\}}] \\
&= P(\tilde{Z} + \frac{1}{\sqrt{J-1}}\gamma \sin(\phi_{j,n}) < 0) \\
&\approx \frac{1}{2} - \frac{\gamma \sin(\phi_{j,n})}{\sqrt{2\pi(J-1)}\sqrt{\sigma_\gamma^2 ||\sin(\phi_{n/(j)})||^2/(J-1) + 2\sigma_t^2/(\mu^2(J-1))}} \\
&= \frac{1}{2} - \frac{\gamma \sin(\phi_{j,n})}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2 ||\sin(\phi_{n/(j)})||^2 + 2\sigma_t^2/\mu^2}}.
\end{aligned}$$

The last approximation for large J arises because when Φ is a normal $N(0,\sigma^2)$ random variable, $P(\Phi + x < 0) \approx (1/2) - (x/(\sigma\sqrt{2\pi}))$ for small positive x. Thus,

$$\begin{aligned} & E[\gamma_{j,n} 1_{\{\tilde{g}+\sum_{r=1}^{J}\gamma_{r,n}\sin(\phi_{r,n})<0\}}] \\ &\approx E_\gamma[\gamma[\frac{1}{2} - \frac{\gamma\sin(\phi_{j,n})}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2||\sin(\phi_{n/(j)})||^2 + 2\sigma_t^2/\mu^2}}] \\ &= -\frac{\sigma_\gamma^2\sin(\phi_{j,n})}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2||\sin(\phi_{n/(j)})||^2 + 2\sigma_t^2/\mu^2}}. \end{aligned}$$

Thus the limiting differential equation (11) is

$$\frac{d\phi_j(t)}{dt} = -\frac{\sigma_\gamma^2\sin(\phi_j(t))}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2||\sin(\phi_{/(j)}(t))||^2 + 2\sigma_t^2/\mu^2}}. \tag{4}$$

A straightforward linearization argument shows that this ODE is stable about zero. Simulations in Sect. 7 show that the ODE accurately tracks the trajectories of the algorithm.

Once the algorithm has converged, it is important to be able to characterize the final distribution. As in the Appendix, define a random process $V(t)$ by

$$V(t) = v_0 + M(t) + L(t) - \theta\int_0^t V(s)ds,$$

where $M(t)$ is a Wiener process with variance $\sigma_M^2 t$, $L(t)$ is another Wiener process independent of M with variance $\sigma_L^2 t$, then $V(t)$ is an Orstein-Uhlenbeck process (an asymptotically stationary Gaussian process) with mean zero and variance $(\sigma_L^2 + \sigma_M^2)/2\theta$.

Suppose that $\sigma_t^2/\mu^2 = \tilde{\sigma}_t^2$. Then $\theta = \partial_w\hat{H}(W(s))|_{W(s)=0} = -\sigma_\gamma^2/\sqrt{4\pi\tilde{\sigma}_t^2}$, $G(0,y,u) = (\gamma 1_{\{\tilde{g}<0\}} - \gamma/2)^2$, and $\hat{G}(0) = \sigma_\gamma^2/4$ where G and $\hat{G}$ are defined in (16)-(17). Thus $\sigma_M^2 = \sigma_\gamma^2/4$. For the variance of L, note that its terms are

$\bar{H}(0,Y) - \hat{H}(0) = \bar{H}(0,Y) = \gamma/2$. Thus the squared values are $\sigma_\gamma^2/4$ and $\sigma_L^2 = \sigma_\gamma^2/4$. The asymptotic variance is therefore given by

$$\mu \frac{\sigma_\gamma^2/4 + \sigma_\gamma^2/4}{2[\sigma_\gamma^2/\sqrt{4\pi\tilde{\sigma}_t^2}]} = \mu \frac{\sqrt{\pi\tilde{\sigma}_t^2}}{2}. \tag{5}$$

Somewhat surprisingly, the asymptotic variance is independent of the size of the phase perturbations σ_γ^2. Sect. 7 shows that this calculated variance matches closely to the empirical variance derived from simulations.

4 The Signed Algorithm

In the distributed phase alignment algorithm (1), the phase of each sensor is updated by the perturbed value (if the feedback from the beacon says that the overall alignment improved) or else it remains fixed. Accordingly, in many iterations, no changes are made. Since each of the individual phase updates are scalar, it seems reasonable that when the feedback indicates no improvement, an update in the opposite direction might be useful. Effectively, this replaces the indicator function in the update with a signum function $\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$. The algorithm is

$$\hat{\phi}_{j,n+1} = \hat{\phi}_{j,n} + \mu\gamma_{j,n}\text{sgn}(g_{n+1} + \sum_{r=1}^{J} \cos(\hat{\phi}_{r,n} + \mu\gamma_{r,n} - \alpha_r) - g_n - \sum_{r=1}^{J} \cos(\hat{\phi}_{r,n} - \alpha_r))$$

for $j = 1, 2, \ldots, J$. Following the logic of (1)-(3) leads to the error system which is valid for small μ

$$\phi_{j,n+1} = \phi_{j,n} - \mu\gamma_{j,n}\text{sgn}(\tilde{g} + \sum_{r=1}^{J} \gamma_{r,n} \sin(\phi_{r,n})). \tag{6}$$

Carrying out the same calculations as in Sect. 3, the expectation of the update term is

$$-\frac{2\sigma_\gamma^2 \sin(\phi_{j,n})}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2 ||\sin(\phi_{n/(j)})||^2 + 2\sigma_t^2/\mu^2}}. \tag{7}$$

This is exactly twice the value in (4), and so the corresponding ODE for the signed algorithm converges twice as fast as when using the indicator function.

The final variance can also be calculated as before: $\theta = \partial_w \hat{H}(W(s))|_{W(s)=0} = -2\sigma_\gamma^2/\sqrt{4\pi\tilde{\sigma}_t^2}$, $G(0,y,u) = (\gamma \text{sgn}(\tilde{g}))^2$, and $\hat{G}(0) = \sigma_\gamma^2$. Thus $\sigma_M^2 = \sigma_\gamma^2$. For the variance of L, note that $\bar{H}(0,Y) - \hat{H}(0) = 0 - 0 = 0$ which implies $\sigma_L^2 = 0$ and the asymptotic variance is

$$\mu \frac{\sigma_\gamma^2}{2[2\sigma_\gamma^2/\sqrt{4\pi\tilde{\sigma}_t^2}]} = \mu \frac{\sqrt{\pi\tilde{\sigma}_t^2}}{2},$$

which is identical to (5). Thus, the signed algorithm converges twice as fast as (1) yet has the same residual error variance.

5 The "ρ % Solution" Algorithm

One of the key requirements in a sensor network system is energy efficiency. The energy consumed in phase alignment using the algorithms of Sects. 2 and 4 is proportional to the number of transmissions per timestep times the number of iterations to convergence. This section proposes a variation that can help to reduce the total number of transmissions needed to achieve convergence.

The key observation is that when there are fewer sensors operating, the convergence tends to be faster. This is because all the sensor perturbations occur in synchrony: with fewer sensors the feedback more directly reflects the contribution of any given sensor. This observation is exploited by having only a subset of the

sensors operate at each timestep. If the subset is chosen randomly, then there is no need to coordinate all the sensors, and the method remains decentralized.

Suppose that at each epoch, sensors independently transmit to the beacon with probability $p = \rho/100$ using its current phase value. This transmission is then immediately followed by another transmission of a "perturbed" phase angle. The beacon then feeds back a single bit which specifies which of the two transmissions had greater power. Each transmitting sensor then updates its current phase value based on the feedback. Thus, in each transmission epoch, only pJ sensors transmit on average; but they must transmit twice.

This strategy can also be written as a small stepsize μ-dependent algorithm. Let $\{B_{1,n}, B_{2,n}, \ldots, B_{J,n}\}$ be independent zero-one Bernoulli random variables with $P(B_{j,n} = 1) = p$. The event $\{B_{j,n} = 1\}$ indicates that at time n, sensor j will transmit. The algorithm for phase convergence at sensor j is then

$$\hat{\phi}_{j,n+1} = \hat{\phi}_{j,n} + \mu B_{j,n} \gamma_{j,n} 1_{\{g_{n+1} + \sum_{r=1}^{J} B_{r,n} \cos(\hat{\phi}_{r,n} + \mu\gamma_{r,n} - \alpha_r) > g_n + \sum_{r=1}^{J} B_{r,n} \cos(\hat{\phi}_{r,n} - \alpha_r)\}}$$

for $j = 1, 2, \ldots, J$. Assuming that the product Jp is large enough to invoke a Central Limit Theorem, it is possible to mimic the analysis given in (1)-(4) to obtain the limiting ODE

$$\frac{d\phi_j(t)}{dt} = -pE\left[\frac{\sigma_\gamma^2 \sin(\phi_j(t))}{\sqrt{2\pi}\sqrt{\sigma_\gamma^2 \sum_{r=1(r\neq j)}^{J} B_r[\sin(\phi_r(t))]^2 + 2\sigma_t^2/\mu^2}}\right]. \qquad (8)$$

The presence of the Bernoulli random variables in the denominator makes a simple closed form solution impossible (though an infinite power series could be developed). Sect. 7 shows that the total number of transmissions (and hence the total energy consumed in the phase alignment process) can be decreased when following this strategy.

In addition, it is possible to combine the idea of the signed update from Sect. 4 with the ρ-percent algorithm. Following the same procedure shows that this algorithm has the same ODE, but multiplied by a factor of two, indicating a doubling of the convergence rate.

The variance analysis of the ρ-percent algorithm is straightforward. In this setting, directly from (8), $\theta = \partial_w \hat{H}(W(s))|_{W(s)=0} = -p\sigma_\gamma^2/\sqrt{4\pi\tilde{\sigma}_t^2}$, $G(0,y,u) = (B\gamma 1_{\tilde{g}<0} - B\gamma/2)^2$, and $\hat{G}(0) = p\sigma_\gamma^2/4 = \sigma_M^2$. For the variance of L, observe that its terms are $\bar{H}(0,Y) - \hat{H}(0) = \bar{H}(0,Y) = B\gamma/2$. Thus the mean squared value is $p\sigma_\gamma^2/4 = \sigma_L^2 = \sigma_\gamma^2 p^2/4$ and the asymptotic variance is

$$\mu\frac{\sigma_L^2 + \sigma_M^2}{2\theta} = \mu\frac{\sqrt{\pi\tilde{\sigma}_t^2}}{2}.$$

Importantly, this is independent of p (and hence ρ).

6 Wideband Time Delay Algorithm

In pseudo-noise (PN) code-division spread-spectrum sensor networks, the analogous problem is to align the time delays of the PN aquisition waveform. This section shows that the same kind of reasoning that led to the narrowband phase alignment algorithms of Sects. 2-5 can also be applied in the more general wideband setting by developing a decentralized algorithm (based on single-bit feedback from a beacon) and the associated limiting ODE.

Assume there is a "target" timeshift τ^* and let $\tilde{\tau}_{j,n}$ be the time delay perceived at the beacon from the jth sensor at the nth signaling epoch. The algorithm

$$\tilde{\tau}_{j,n+1} = \tilde{\tau}_{j,n} + \mu\gamma_{j,n} 1_{\{\sum_{r=1}^J (\tilde{\tau}_{r,n}+\mu\gamma_{r,n}-\tau^*)^2 < \sum_{r=1}^J (\tilde{\tau}_{r,n}-\tau^*)^2\}}$$

acts to align the received signals in time. Define the time delay error at the $j$the

sensor at time n as $\tau = \tilde{\tau}_{r,n} - \tau^*$. The error system for the algorithm is

$$\tau_{j,n+1} = \tau_{j,n} + \mu\gamma_{j,n} 1_{\{\sum_{r=1}^{J}(\tau_{r,n}+\mu\gamma_{r,n})^2 < \sum_{r=1}^{J}(\tau_{r,n})^2\}}.$$

Expanding the square in the indicator function allows the algorithm to be rewritten

$$\tau_{j,n+1} = \tau_{j,n} + \mu\gamma_{j,n} 1_{\{\sum_{r=1}^{J}\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}\gamma_{r,n}^2 < 0\}}.$$

To compute the ODE, it is necessary to first compute the expectation of the indicator term conditioned on $\gamma_{j,n} = \gamma$. Thus

$$\begin{aligned}
&P(\sum_{r=1}^{J}\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}\gamma_{r,n}^2 < 0 | \gamma_{j,n} = \gamma) \\
&= P(\sum_{r\neq j}(\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}\gamma_{r,n}^2) + \tau_{j,n}\gamma + \frac{\mu}{2}\gamma^2 < 0) \\
&= P(\frac{1}{\sqrt{J-1}}\sum_{r\neq j}(\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}(\gamma_{r,n}^2 - \sigma_\gamma^2)) + \frac{\tau_{j,n}\gamma + \frac{\mu}{2}\gamma^2}{\sqrt{J-1}} + \sqrt{J-1}\frac{\mu}{2}\sigma_\gamma^2 < 0).
\end{aligned}$$

Since the term $(\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}(\gamma_{r,n}^2 - \sigma_\gamma^2))$ has zero mean and variance $\tau_{r,n}^2\sigma_\gamma^2 + \frac{\mu^2}{4}\text{Var}(\gamma^2)$,

$$\approx P(\hat{Z} + \frac{1}{\sqrt{J-1}}(\tau_{j,n}\gamma + \frac{\mu}{2}\gamma^2) + \sqrt{J-1}\frac{\mu}{2} < 0)$$

where $\hat{Z}$ is Gaussian with zero mean and variance $\sum_{r\neq j}\tau_{r,n}^2\sigma_\gamma^2 + \frac{\mu^2}{4}\frac{\text{Var}(\gamma^2)}{J-1}$. When $\mu << 1$, the above expression is approximately

$$P(\tilde{Z} < -\frac{\tau_{j,n}\gamma}{\sqrt{J-1}})$$

where $\tilde{Z}$ is normal, mean zero, and has variance $\frac{\sigma_\gamma^2}{J-1}\sum_{r\neq j}\tau_{r,n}^2$. Define $||\tau||_{n/(j)}^2 = \sum_{r\neq j}\tau_{r,n}^2$. Then

$$P(\tilde{Z} < -\frac{\tau_{j,n}\gamma}{\sqrt{J-1}}) = P(Z < -\frac{\tau_{j,n}\gamma}{\sigma_\gamma||\tau||_{n/(j)}})$$

where Z is a standard normal random variable. Thus, letting $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal, the limiting ODE is

$$\frac{d\tau_j(t)}{dt} = E[\gamma\Phi(-\frac{\tau_j(t)\gamma}{\sigma_\gamma||\tau(t)||_{/(j)}})],$$

where $||\tau(t)||^2_{/(j)} = \sum_{r\neq j}\tau_r(t)^2$. For small positive x, $\Phi(-x) \approx .5 - x/\sqrt{2\pi}$. Therefore when the argument is small, the ODE is

$$\frac{d\tau_j(t)}{dt} = -\frac{\tau_j(t)\sigma_\gamma}{\sqrt{2\pi}||\tau(t)||_{/(j)}}.$$

A thermal noise component in the received signals causes the beacon to observe $\tilde{\tau}_{r,n} + g_{r,n}$ instead of $\tilde{\tau}_{r,n}$. As in the previous phase alignment algorithms μ must be chosen taking into account the size of the thermal noise. Thus, assume that $g_{r,n} = \mu\tilde{g}_{r,n}$ where $\tilde{g}_{r,n}$ has a symmetric probability density and variance $\tilde{\sigma}_t^2$. The limiting ODE then requires computing

$$\begin{aligned} & P(\sum_{r=1}^{J}\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}\gamma_{r,n}^2 + t_{r,n} < 0|\gamma_{j,n} = \gamma) \\ = \; & P(\sum_{r\neq j}[\tau_{r,n}\gamma_{r,n} + \frac{\mu}{2}\gamma_{r,n}^2 + t_{r,n} - \frac{\mu}{2}\sigma_\gamma^2] + \tau_{j,n}\gamma + \frac{\mu}{2}\gamma^2 + t_{j,n} + (J-1)\frac{\mu}{2}\sigma_\gamma^2 < 0) \end{aligned}$$

where $t_{k,n} = (\tau_{k,n} + \mu\gamma_{k,n})\tilde{g}_{k,n+1} - \tau_{k,n}\tilde{g}_{k,n} + \frac{\mu}{2}(\tilde{g}_{k,n+1}^2 - \tilde{g}_{k,n}^2)$. When $\mu << 1$, the above expression is approximately

$$P(\hat{Z} + \frac{1}{\sqrt{J-1}}\tau_{j,n}(\gamma + \tilde{g}_{j,n+1} - \tilde{g}_{j,n}) < 0),$$

where $\hat{Z}$ is normal, mean zero, with variance $\frac{\sigma_\gamma^2 + 2\sigma_t^2}{J-1}||\tau||^2_{n/(j)}$. Assuming that the noise expressions are Gaussian allows rewriting this as

$$P(\tilde{Z} + \frac{1}{\sqrt{J-1}}\tau_{j,n}\gamma < 0)$$

where $\tilde{Z}$ is normal, mean zero, with variance $\frac{||\tau||^2_{n/(j)}(\sigma_\gamma^2+2\sigma_t^2)+2\tau_{j,n}^2\sigma_t^2}{J-1}$. Letting Z denote a standard normal, the above expression simplifies to

$$P(Z + \frac{\tau_{j,n}\gamma}{\sqrt{||\tau||^2_{n/(j)}(\sigma_\gamma^2+2\sigma_t^2)+2\tau_{j,n}^2\sigma_t^2}} < 0)$$

and the ODE is

$$\frac{d\tau_j(t)}{dt} = E[\gamma\Phi(-\frac{\tau_j(t)\gamma}{\sqrt{||\tau(t)||^2_{/(j)}(\sigma_\gamma^2+2\sigma_t^2)+2\tau_j(t)^2\sigma_t^2}})].$$

7 Simulations

This section illustrates the relationship between the trajectories of the algorithm(s) and the behavior of the ODE(s) and shows that the calculated variances accurately reflect the behavior of the algorithm. Fig. 1 shows the behavior of the error system (2) in a simple configuration with $J = 10$ sensors randomly initialized in $(-\pi, \pi)$. The jagged lines are the trajectories of the phase estimates while the smooth lines show the trajectories of the ODE (4) starting from the same set of initial locations. Observe that the phase estimates follow the ODE quite closely as they converge to zero. Observe also that locally (near zero) the ODEs converge exponentially, as expected from the linearization argument. When far from zero (the bottom-most trajectories in Fig. 1) the motion appears to be slower than exponential. Recall that convergence of the error system to a region about zero is equivalent to convergence of the actual trajectories of the phase estimates to a region about their (unknown) values. The final error variance can be calculated directly from the simulation; for Fig. 1 this is 0.0015, which compares well with the predicted error variance from (5), which is 0.0014.

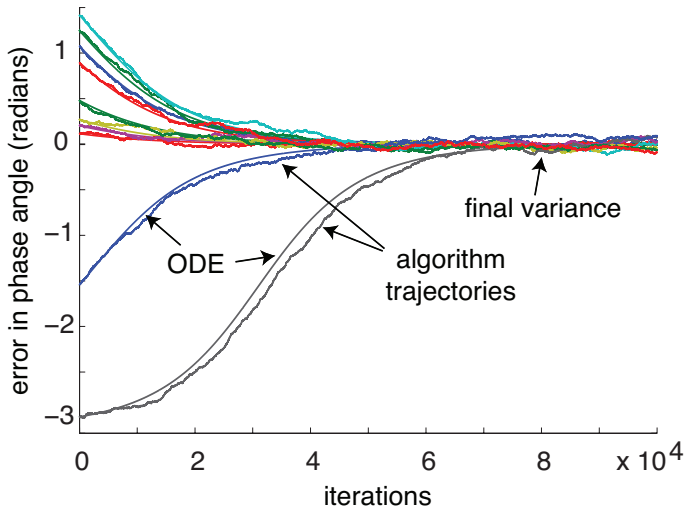

Figure 1: Trajectories of the error system for the decentralized phase alignment algorithm (2) and the corresponding trajectories of the ODEs of (5).

Similarly. Fig. 2 shows trajectories of the error systems for the indicator algorithm (with error system (2) and ODE (4)) and for the signed version (with error system (6) and ODE (7)) . $J = 10$ sensors were used though only six are shown in the figure to reduce clutter. The two algorithms were initialized at the same values and allowed to iterate. Observe that in all cases, the signed algorithm converges faster, at about twice the rate of the indicator version, as suggested by the corresponding ODEs. Parameters for the simulation are $\mu = .0005$, $\tilde{\sigma}_t^2 = 10$, and $\sigma_\gamma^2 = 2$. The final variance, calculated to be $.0014$, agrees with the empirical value (measured from the simulations) to four decimal places.

Fig. 3 shows the trajectories of the ρ-percent error system 5 and the corresponding ODE (8). Again, the algorithm follows the ODE as it converges exponentially towards its stable point. In these simulations the stepsize $\mu = 0.01$ and the standard deviation of the thermal noise was 0.001. The predicted final vari-

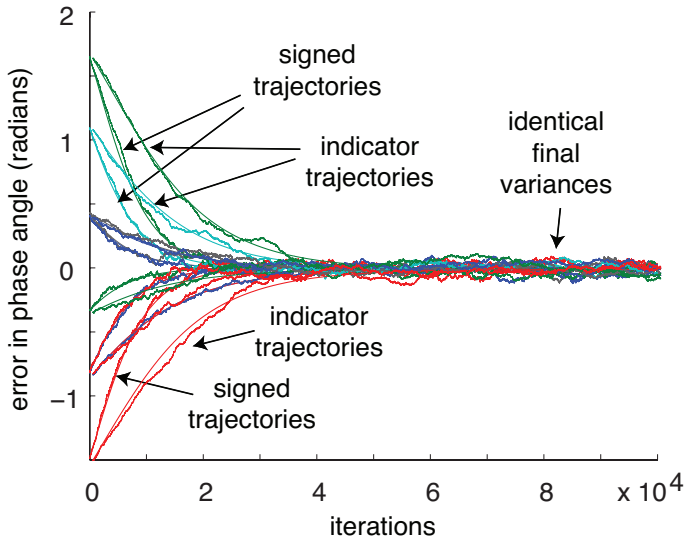

Figure 2: Comparison of the convergence of the indicator algorithm (2) and the signed variation (6) from identical initial conditions. As expected from the ODEs, the signed algorithm converges with a rate twice that of the indicator algorithm. As expected from the variance calculation, the final variances are the same.

ance for the $\rho = 30$ case was $8.862E^{-4}$ while the actual variance, computed over all sensors, was $8.587E^{-4}$. The predicted final variance for the $\rho = 15$ case was $8.862E^{-4}$ (the same as for $\rho = 30$) while the actual variance was $9.32E^{-4}$.

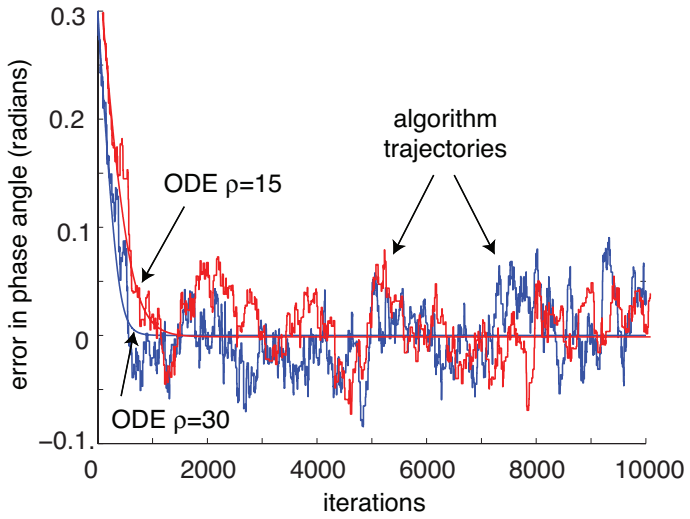

Figure 3: A typical trajectory of the ρ% algorithm (with $\rho = 15$ and $\rho = 30$) and the corresponding ODEs. Smaller ρ take more iterations to converge, but use significantly fewer transmissions (and hence less energy) per iteration.

It is also necessary to verify that the convergence of the ρ-percent algorithm is rapid enough that the total number of transmissions needed is less than for the corresponding algorithm where all sensors transmit at every time step (which is essentially the $\rho = 100$ case). With $J = 10$ sensors, a thermal noise with standard deviation 0.001, and a stepsize of $\mu = 0.01$, Table 1 shows how many iterations are needed for convergence as a function of the ρ value. The experiment is conducted by setting the phase error for sensor $\#1$ at 1.0 radian, and checking how many iterations are needed before the sensor converges 95% of the way to zero.

As might be expected, the number of iterations decreases as ρ increases, but so

percent ρ	iterations to convergence	# of (double) transmissions
5	2695	134.9
10	1670	166.6
20	1233	245.3
30	1137	338.7
40	1070	427.0
50	998	497.0
60	919	554.1
70	864	604.8
80	813	650.0
90	768	690.8
95	749	711.1
99	734	726.7

Table 1: Convergence experiment for the ρ-percent method of Sect. 5

do the number of transmissions. For this simulation, which is fairly typical, about 720 transmissions are needed for the $\rho = 100$ case. Since the algorithm requires two transmissions in each epoch, any ρ that requires fewer than half this number (i.e., 360) transmissions will be more efficient. In this case, the crossover occurs at about $\rho = 30$. The purpose here is not to try and elucidate the best parameters to use, only to demonstrate that significant gains in energy usage, as reflected in the number of transmissions required, are possible when using the ρ-percent algorithm.

8 Conclusions

This paper has analyzed a recently proposed algorithm for the decentralized beamforming problem, demonstrating concrete expressions for the rate of convergence and for the final variance of the algorithm about its converged values. Moreover,

the algorithm has been extended and improved in three ways: adapting with both positive and negative feedback results in an algorithm with twice the rate of convergence and the same final variance, adjusting only ρ-percent of the sensors at each timestep reduces the energy requirements of the algorithm, and an analogous method suitable for use with wideband transmissions is proposed and analyzed.

References

[1] G. Barriac, R. Mudumbai, U. Madhow, "Distributed beamforming for information transfer in sensor networks," IPSN'04, April 26–27, 2004, Berkeley, California, USA.

[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[3] D. L. Duttweiler, "Adaptive filter performance with nonlinearities," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 30, pp. 578–586, Aug 1982.

[4] J. A. Bucklew, T. Kurtz, and W. A. Sethares, "Weak convergence and local stability of fixed step size recursive algorithms,"*IEEE Trans. on Info. Theory*, Vol. 39, No. 3, pp. 966–978, May, 1993.

[5] J.A. Bucklew, W.A. Sethares, "The covering problem and μ-dependent adaptive algorithms," *IEEE Trans. Sig. Proc.*, Vol. 42, No. 10, pp. 2616–2627, October, 1994.

[6] S. Ethier and T. Kurtz, *Markov Processes - Characterization and Convergence*, Wiley-Interscience, New York, 1986.

[7] M. Gastpar and M. Vetterli, "On the capacity of large gaussian relay networks," *IEEE Trans. on Inform. Theory*, vol. 51, no. 3, pp. 765–779, 2005.

[8] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Trans. on Info. Theory*, Vol. IT-30, No. 2, pp. 191–198, March 1984.

[9] B. Hassibi and A. Dana, "On the power efficiency of sensory and ad-hoc wireless networks," *IEEE Trans. on Inform. Theory*, July 2006.

[10] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, Cambridge, MIT Press, 1983.

[11] R. Mudumbai, B. Wild, U. Madhow and K. Ramchandran, "Distributed beamforming using 1 bit feedback: from concept to realization," *Proc. of 44'th Allerton Conference on Communication Control and Computing*, Sept. 2006.

[12] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed Transmit Beamforming using Feedback Control," submitted to *IEEE Trans. on Info. Theory* (see http://www.ece.ucsb.edu/∼raghu/research/pubs.html)

[13] H. Ochiai, P. Mitran, H. V. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks, *IEEE Trans. Signal Proc.*, Vol. 53, No. 11, Nov. 2005.

A Appendix

The general form of a discrete-time iteration process is

$$W_{k+1} = W_k + \mu H(W_k, Y_k, U_{k+1}, \mu) \tag{9}$$

where W_k is a vector of parameters, Y_k is a random perturbation, U_{k+1} is a (random) input vector, and μ is the algorithm stepsize. The function H represents the update term of the algorithm and is in general discontinuous. This form (9) is called μ-dependent since the step size μ appears both inside and outside the update function H. What is the nature of the random process W_k? In typical operation, it converges to a region about some special state and then bounces randomly near that state. This Appendix shows one way to characterize this convergence, and demonstrates that under mild assumptions, the final distribution is normal with parameters that can be described in terms of the distributions of the inputs and noises.

The analysis begins by relating the behavior of the algorithm (9) for small μ to the behavior of the associated deterministic integral equation

$$W(t) = W_0 + \int_0^t \hat{H}(W(s))ds \tag{10}$$

or equivalently, to the associated deterministic ordinary differential equation (ODE)

$$\dot{W}(t) = \hat{H}(W(t)) \tag{11}$$

where $\hat{H}$ is smoothed by the distribution of the inputs U_{k+1} and the noises Y_k. Speaking loosely, the ODE $W(t)$ of (11) represents the "averaged" behavior of the parameters W_k in (9) and this smoothed version is often differentiable even if H itself is discontinuous.

A time scaled version of W is defined as

$$W_\mu(t) = W_{\lfloor t/\mu \rfloor} \;\; t \in [0, \infty) \tag{12}$$

where $\lfloor z \rfloor$ means the integer part of z. Note that W_k represents the discrete iteration process, while $W_\mu(t)$ represents a continuous time-scaled version. $W(t)$

(with no subscript) is the ODE (11) to which $W_\mu(t)$ converges weakly. In a previous paper [5], we had analyzed in some detail the conditions necessary to guarantee weak convergence of $W_\mu(\cdot)$ to $W(\cdot)$.

This appendix focuses on the final convergent distribution of W_k by finding conditions under which the error

$$V_\mu(t) = \frac{1}{\sqrt{\mu}}(W_\mu(t) - W(t)) \tag{13}$$

converges weakly to a solution of a particular stochastic differential equation (SDE) V_μ. In many interesting cases, it is possible to calculate the steady state variance of this SDE and make concrete predictions about the residual mean squared error of the algorithm.

The random sequence $\{W_k, Y_k, U_k\}$ is defined on some probability space (Ω, F, P) and takes values in $\Re^d \times E_1 \times E_2$, where d is the length of W_k and E_1 and E_2 are measurable state spaces on which Y_k and U_k evolve. $\{W_k, Y_k, U_k\}$ is adapted to a filtration $\{\mathcal{F}_k\}$, (usually one takes $\mathcal{F}_k$ = the σ-algebra generated by the random variables $\{W_l, Y_l, U_l\}_{l=-\infty}^{k}$). Let $\mathcal{P}(A)$ denote the collection of probability measures on the space A. Assume the following:

C.1 $\{Y_k\}$ is stationary, ergodic[1] and there is a sequence of i.i.d. E_3-valued random variables $\{\psi_k\}$, independent of $\{Y_k\}$, and a measurable function $q : \Re^d \times E_1 \times E_3 \to E_2$ such that $U_{k+1} = q(W_k, Y_k, \psi_k)$, and W_0 is independent of $\{(Y_k, \psi_k)\}$. Define $P(U_{k+1} \in C|\mathcal{F}_k) = P(q(W_k, Y_k, \psi_k) \in C|\mathcal{F}_k) = \eta(W_k, Y_k, C)$ and assume that H is integrable with respect to $\eta(w, y, \cdot)$ for each $(w, y) \in \Re^d \times E_1$. Let

[1]Stationarity and Ergodicity imply that $\sum_{k=0}^{\lfloor t/\mu \rfloor} 1_B(Y_k) \to t \times \nu_Y(B)$ a.s., where ν_Y denotes the (asymptotic) distribution of the $\{Y_k\}$. This convergence is the essential assumption needed about the $\{Y_k\}$ sequence. Hence some sort of asymptotic stationarity/ergodicity could be assumed.

$\nu_Y \in \mathcal{P}(E_1)$ denote the distribution of Y_k. Define

$$\bar{H}(w, y, \mu) = \int_{E_2} H(w, y, u, \mu)\eta(w, y, du). \tag{14}$$

C.2 For every $K \in \Re^+$, $\bar{H}(w, y, \mu)$ is continuous and converges uniformly on $\{\{|w| \leq K\} \times E\}$ to a continuous function $\tilde{H}(w, y)$ on $\{|w| \leq K\}$. Furthermore for some $\mu_0 > 0$,

$$\begin{aligned} E\{\sup_{|w|\leq K, 0<\mu\leq\mu_0} |H(w, Y_k, q(w, Y_k, \psi_k), \mu)|\} &< \infty, \\ E\{\sup_{|w|\leq K, 0<\mu\leq\mu_0} |\bar{H}(w, Y_k, \mu)|\} &< \infty. \end{aligned}$$

Note that there are no assumptions on the autocorrelations of the inputs or disturbances. H is allowed to be discontinuous, provided that the expectation over η is smooth enough to make $\bar{H}$ continuous and the limit operation in the μ step size variable is uniform which leads to a continuous $\tilde{H}$. Just as $\tilde{H}$ is an averaging and limiting process for H, the distribution of Y_k is used to average $\tilde{H}$ over the inputs Y_k, and the doubly averaged quantity

$$\hat{H}(w) = \int \tilde{H}(w, y)\nu_Y(dy) \tag{15}$$

is the key ingredient in the ODE.

The mathematical framework in which this work is imbedded is described comprehensively in [2] and [6]. Let (E, r) denote a metric space with associated Borel field $\mathcal{B}(E)$. $D_E[0, \infty)$ is the space of right continuous functions with left limits mapping from the interval $[0, \infty)$ into E, and $D_E[0, \infty)$ is assumed to be endowed with the Skorohod topology.

Let $\{X_\alpha\}$ (where α ranges over some index set) be a family of stochastic processes with sample paths in $D_E[0, \infty)$ and let $\{P_\alpha\} \subset \mathcal{P}(D_E[0, \infty))$ be the

family of associated probability distributions (i.e. $P_\alpha(B) = P\{X_\alpha \in B\}$ for all $B \in \mathcal{B}(E)$). $\{X_\alpha\}$ is said to be relatively compact if $\{P_\alpha\}$ is relatively compact in the space of probability measures $\mathcal{P}(D_E[0,\infty))$ endowed with the topology of weak convergence. The symbol $\Rightarrow$ denotes weak convergence while $\rightarrow$ denotes convergence under the appropriate metric.

Theorem 1 *Let $W_\mu(t) = W_{\lfloor t/\mu \rfloor}$, and for $K \in \Re^+$, define $\tau_\mu^K = inf\{t : |W_\mu(t)| \geq K\}$, and $W_\mu^{\tau_\mu^K}(\cdot) = W_\mu(\cdot \wedge \tau_\mu^K)$ define the "stopped" process. Assume C.1,C.2, and that $W_\mu(0) \rightarrow w_0$ in probability as $\mu \rightarrow 0$. Then for each K, $\{W_\mu^{\tau_\mu^K}, \mu > 0\}$ is relatively compact, and every limit point (as $\mu \rightarrow 0$) satisfies (10) for $t < \tau^K = inf\{t : |W(t)| \geq K\}$.*

The stopping time τ_μ^K measures how long it takes the time scaled process $W_\mu(t)$ to reach K in magnitude. The stopped process $\{W_\mu^{\tau_\mu^K}(t)\}$ is defined to be equal to $W_\mu(t)$ from time zero to the stopping time τ_μ^K and is then held constant for all $t > \tau_\mu^K$. The theorem asserts that for any $K \in \Re^+$, every possible sequence (as $\mu \rightarrow 0$) of the stopped process $\{W_\mu^{\tau_\mu^K}(t)\}$ contains a weakly convergent subsequence, and that every limit of these subsequences is a process that satisfies the ODE (10), at least up until the stopping time. If the solution to the differential equation is unique, then the sequence actually converges in probability (not just has a weakly convergent subsequence). The limiting quantity (the solution of the ODE) is continuous. The Skorohod topology for continuous functions corresponds to uniform convergence on bounded time intervals. Hence, convergence in probability means that for every $T > 0,\ \epsilon > 0$, $\lim_{\mu \rightarrow 0} P(\sup_{0 \leq t < \tau^k \wedge T} |W_\mu^{\tau_\mu^K}(t) - W(t)| > \epsilon) = 0$. Note that if no solution of the ODE becomes unbounded in finite time, then $\tau_\mu^K \rightarrow \infty$ as $K \rightarrow \infty$. In this case, $\{W_\mu\}$ is relatively compact without needing to restrict attention to the stopped processes.

Theorem 1 is a kind of "law of large numbers" for discrete time iterative processes such as (9). The corresponding "central limit theorem" describes the weak convergence of the error process (13) where the scaling factor $\frac{1}{\sqrt{\mu}}$ expands V_μ to compensate for the time compression of $W_\mu(t)$. The next theorem shows that the error process V_μ converges to a forced ODE that is driven by the sum of two independent, mean zero Brownian motions. The driving term $H - \bar{H}$ accounts for the error introduced by the smoothing with the disturbance while $\hat{H} - \bar{H}$ accounts for the error when averaging over the inputs.

Let

$$G(w, y, u, \mu) = (H(w, y, u, \mu) - \bar{H}(w, y, \mu))(H(w, y, u, \mu) - \bar{H}(w, y, \mu))^T \quad (16)$$

be the matrix that represents the deviation of H from its smoothed version $\bar{H}$. If H is square integrable with respect to $\eta(w, y, \cdot)$ for each pair $(w, y) \in \Re^d \times E_1$, the smoothed version of G is

$$\bar{G}(w, y, \mu) = \int_{E_2} G(w, y, u, \mu)\eta(w, y, du).$$

Suppose $\bar{G}(w, y, \mu)$ converges as $\mu \to 0$ to some $\tilde{G}(w, y)$. Averaging over all inputs yields

$$\hat{G}(w) = \int \tilde{G}(w, y)\nu_Y(dy). \quad (17)$$

The various G's play a similar role in the central limit theorem that the H's play in Theorem 1. In addition to C.1 and C.2, further assume:

C.3 H is square integrable with respect to $\eta(w, y, \cdot)$ for each pair $(w, y) \in \Re^d \times E_1$. $\bar{H}$ is continuously differentiable as a function of w, the continuous $\partial_w \bar{H}$ converge

uniformly to $\partial_w \tilde{H}$. $\bar{G}$ is continuos and converges uniformly to $\tilde{G}$. For all $K \in \Re^+$

$$E\{\sup_{|w|\leq K, 0<\mu\leq\mu_O} |H(w, Y_k, q(w, Y_k, \psi_k), \mu)|^2\} < \infty$$
$$E\{\sup_{|w|\leq K, 0<\mu\leq\mu_O} |\bar{G}(w, Y_k, \mu)|\} < \infty$$
$$E\{\sup_{|w|\leq K, 0<\mu\leq\mu_O} |\partial_w \bar{H}(w, Y_k, \mu)|\} < \infty$$

Note that C.3 implies $\hat{H}$ is locally Lipschitz (in fact continuously differentiable), so the solution of (10) is unique and hence $V_\mu(t)$ is well defined (on any interval of which the solution of the ODE is bounded). For simplicity (so that it is not necessary to stop the process outside of a compact set), assume that the solution exists for all $t \geq 0$. Define

$$\tilde{M}_\mu(t) = \sum_{k=0}^{[t/\mu]-1} (H(W_k, Y_k, U_{k+1}, \mu) - \bar{H}(W_k, Y_k, \mu))\sqrt{\mu}$$

and

$$L_\mu(t) = \sum_{k=0}^{[t/\mu]-1} (\bar{H}(W(k\mu), Y_k, \mu) - \hat{H}(W(k\mu)))\sqrt{\mu}.$$

There are a variety of different conditions (for example, mixing conditions on $\{Y_k\}$) that imply $\{L_\mu\}$ converges weakly to a (time inhomogeneous) Brownian motion. We simply assume this convergence.

C.4 $L_\mu \Rightarrow L$.

Given the assumptions C.1-C.4, the proof of the theorem follows the same logic as that in Theorem 2.2 of [4].

Theorem 2 *Assume C.1-C.4, that $W_\mu(0) \to w_0$ in probability, that the solution of (10) exists for all $t \geq 0$, and that $V_\mu(0) \to v_0$ in probability as $\mu \to 0$. Then $\tilde{M}_\mu \Rightarrow \tilde{M}$ where $\tilde{M}$ is a mean zero Brownian motion independent of L with*

$$E\{\tilde{M}(t)\tilde{M}(t)^T\} = \int_0^t \hat{G}(W(s))ds$$

and $V_\mu \Rightarrow V$ *satisfying*

$$V(t) = v_0 + \tilde{M}(t) + L(t) + \int_0^t \partial_w \hat{H}(W(s))V(s)ds.$$