
Synchronization can Control Regularization in Neural Systems via Correlated Noise Processes: Supplementary Material

Jake Bouvrie
Department of Mathematics
Duke University
Durham, NC 27708
jvb@math.duke.edu

Jean-Jacques Slotine
Nonlinear Systems Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02138
jjs@mit.edu

1 Proof of Theorem 3.1

Recall the definitions $\alpha := \|\mathbf{x}\|^2 + m\gamma^2$, $\beta := \langle \mathbf{x}, \mathbf{y} \rangle$, and $\boldsymbol{\mu} := (\beta/\alpha)\mathbf{1}$. Viewing $\mathbf{w}(t)$ as a collection of Gaussian random variables indexed by t , expressions for $\bar{w}(t)$ and $\tilde{\mathbf{w}}(t)$ can be obtained as manipulations of Gaussians:

$$\begin{aligned}\bar{w}(t) &\sim \mathcal{N}\left(\frac{1}{n}\mathbf{1}^\top \boldsymbol{\mu}_w(t), \frac{1}{n^2}\mathbf{1}^\top \Sigma_w(t)\mathbf{1}\right) \\ &= \mathcal{N}\left(e^{-\alpha t} \mathbb{E}[\bar{w}(0)] + (1 - e^{-\alpha t})\frac{\beta}{\alpha}, e^{-2\alpha t} \mathbb{E}[(\bar{w}(0))^2] + \frac{\sigma^2}{2\alpha n}(1 - e^{-2\alpha t})\right)\end{aligned}$$

where $\bar{w}(0) = \frac{1}{n}\mathbf{1}^\top \mathbf{w}(0)$. Turning to the fluctuations, let $Q = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ denote the orthogonal projection onto the zero-mean subspace of \mathbb{R}^n . Note that $\tilde{\mathbf{w}} = Q\mathbf{w}$ and $\bar{w}\mathbf{1} = (I - Q)\mathbf{w}$. We have

$$\begin{aligned}\tilde{\mathbf{w}}(t) &\sim \mathcal{N}(Q\boldsymbol{\mu}_w(t), Q\Sigma_w(t)Q^\top) \\ &= \mathcal{N}\left(e^{-(L+\alpha I)t} \mathbb{E}[\tilde{\mathbf{w}}(0)], e^{-(L+\alpha I)t} \mathbb{E}[\tilde{\mathbf{w}}(0)\tilde{\mathbf{w}}(0)^\top]e^{-(L+\alpha I)t} \right. \\ &\quad \left. + \frac{\sigma^2}{2}(QLQ^\top + \alpha I)^{-1}(I - e^{-2(L+\alpha I)t})\right).\end{aligned}$$

We can now consider the error

$$\mathbb{E}\left[\frac{1}{n}\|\mathbf{w}(t) - \boldsymbol{\mu}\|^2\right] = \mathbb{E}\left[\frac{1}{n}\|\tilde{\mathbf{w}}(t)\|^2\right] + \mathbb{E}\left[\frac{1}{n}\|\bar{w}(t)\mathbf{1} - \boldsymbol{\mu}\|^2\right].$$

In general if $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \Sigma_x)$ then $\mathbb{E}[\|\mathbf{x} - \mathbf{c}\|^2] = \text{tr}(\Sigma_x) + \|\boldsymbol{\mu}_x - \mathbf{c}\|^2$ for any (non-random) vector \mathbf{c} . The first error term on the right-hand side can be estimated as

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\|\tilde{\mathbf{w}}(t)\|^2\right] &\leq \frac{1}{n} \sum_{i>0} \lambda_i(\Sigma_w(0))e^{-2(\lambda_i(L)+\alpha)t} + \frac{\sigma^2}{2n} \sum_{i>0} \frac{1 - e^{-2(\lambda_i(L)+\alpha)t}}{\lambda_i(L) + \alpha} \\ &\quad + \mathbb{E}[\mathbf{w}(0)]^\top Q^\top e^{-2(L+\alpha)t} Q \mathbb{E}[\mathbf{w}(0)] \\ &\leq \lambda_{\max}(\Sigma_w(0))e^{-2(\underline{\lambda}+\alpha)t} + \frac{\sigma^2(1 - e^{-2(\underline{\lambda}+\alpha)t})}{2(\underline{\lambda} + \alpha)} + e^{-2(\underline{\lambda}+\alpha)t} \|\mathbb{E}[\mathbf{w}(0)]\|^2\end{aligned}$$

where $\underline{\lambda}$ is the smallest non-zero eigenvalue of L and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of its argument. The first term on the right-hand side of the first inequality follows from Von Neumann's trace inequality. The second error term is given by

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\|\bar{w}(t)\mathbf{1} - \boldsymbol{\mu}\|^2\right] &= \mathbb{E}\left[\left(\bar{w}(t) - \frac{\beta}{\alpha}\right)^2\right] \\ &= e^{-2\alpha t} \mathbb{E}[(\bar{w}(0))^2] + \frac{\sigma^2}{2\alpha n}(1 - e^{-2\alpha t}) + e^{-2\alpha t} \left(\mathbb{E}[\bar{w}(0)] - \frac{\beta}{\alpha}\right)^2.\end{aligned}$$

Note that $\mathbb{E}[(\bar{w}(0))^2] = \frac{1}{n^2} \mathbf{1}^\top \Sigma_w(0) \mathbf{1}$ and $\mathbb{E}[\bar{w}(0)] = \frac{1}{n^2} \mathbf{1}^\top \boldsymbol{\mu}_w(0)$. Defining the constants

$$\tilde{C} := \lambda_{\max}(\Sigma_w(0)) - \frac{\sigma^2}{2(\lambda + \alpha)} + \|\mathbb{E}[\mathbf{w}(0)]\|^2$$

$$\bar{C} := \mathbb{E}[(\bar{w}(0))^2] - \frac{\sigma^2}{2\alpha n} + \left(\mathbb{E}[\bar{w}(0)] - \frac{\beta}{\alpha}\right)^2$$

and combining with the above, we obtain the Theorem.

2 Proof of Theorem 5.1

The OU process (16b) is ergodic and has stationary distribution $\mu_\infty = \mathcal{N}(\mathbf{0}, \frac{1}{2}\sigma^2 I)$. Furthermore, the system (17)-(16b) satisfies the conditions of Theorem 2.1. Homogenizing (17) requires the averaged vector field

$$F(u) = \int_{\mathbb{R}^d} (2u^\top Az + z^\top Az) z \mu_\infty(dz) = 2 \mathbb{E}[zz^\top] Au + \mathbb{E}[zz^\top Az] = \sigma^2 Au$$

(using that odd moments of a zero-mean Gaussian are zero), and leads to the averaged system

$$\dot{U} = -\gamma\sigma^2 AU, \quad U(0) = \mathbf{u}(0).$$

The solution to this ODE is easily found to be $U(t) = e^{-\gamma\sigma^2 At} U(0)$. Theorem 2.1 then provides that $\mathbf{u}(t)$ converges in distribution to $U(t)$ as $\varepsilon \rightarrow 0$. Since $U(t)$ is deterministic for all $t \geq 0$ in this case, $\mathbf{u}(t)$ also converges to $U(t)$ in probability. Let $(e_i)_{i=1}^d$ denote the canonical basis of \mathbb{R}^d and let x^i denote the i -th coordinate of a vector $x \in \mathbb{R}^d$. The projection function $\pi_i(x) = \langle x, e_i \rangle = x^i$ is clearly continuous, so by the continuous mapping theorem, $\pi_i(\mathbf{u}_t) \rightarrow U^i(t)$ in probability.

Let $\mathbf{u}_\varepsilon(t)$ denote the (strong) solution to (17) for some fixed $\varepsilon \in (0, 1]$. If the family $\{\mathbf{u}_\varepsilon^i(t)\}_{\varepsilon \in (0, 1]}$ is uniformly integrable (for each $t < \infty$), then together with $\mathbf{u}^i(t) \rightarrow U^i(t)$ i.p., we would have that $\mathbb{E}[\mathbf{u}_\varepsilon^i(t)] \rightarrow \mathbb{E}[U^i(t)] = U^i(t)$ as $\varepsilon \rightarrow 0$ (by way of convergence in L_1). We establish uniform integrability by showing that $\sup_{\varepsilon \in (0, 1]} \mathbb{E}[\pi_i^2(\mathbf{u}_\varepsilon(t))] < \infty$. First note that for any $\varepsilon > 0$, the OU process (16b) is a Gaussian process $\mathbf{Z}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \Sigma_t)$ with bounded moments $\mathbb{E}[\|\mathbf{Z}_t\|^p] < \infty, p \geq 1$, for all $t \leq T < \infty$: Suppose $X \sim \mathcal{N}(0, I_{d \times d})$. Then for each t , $\mathbf{Z}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2} X$ in law. Because the standard Normal moments $\mathbb{E}[\|X\|^p]$ are bounded for all p , we have, restricting our attention to p even, that

$$\begin{aligned} \mathbb{E}[\|\mathbf{Z}_t\|^p] &\leq 2^{p-1} (\|\boldsymbol{\mu}_t\|^p + \mathbb{E}[\|\Sigma_t^{1/2} X\|^p]) \\ &\leq C_p (e^{-pt/\varepsilon} + (\text{tr } \Sigma_t)^{p/2} \mathbb{E}[\|X\|^p]) \\ &\leq C_p (1 + e^{-pt/\varepsilon}) < \infty \end{aligned}$$

where C_p is a constant depending on p that changes from instance to instance, and where $\boldsymbol{\mu}_t, \Sigma_t$ are given by (13a), (13b) (resp.) with $L_z = 0, \eta = 1, \gamma = \sigma/\sqrt{2}$. Returning to the second moment of \mathbf{u} , define the norm $\|x\|_A \triangleq \sqrt{\langle x, Ax \rangle}$, where A is the symmetric strictly positive definite matrix appearing in (17). Note that $\lambda_{\min}(A)\|x\|^2 \leq \|x\|_A^2 \leq \lambda_{\max}(A)\|x\|^2$ for any $x \in \mathbb{R}^d$, where $\lambda_{\min}(A) > 0$ is the smallest eigenvalue of A and $\lambda_{\max}(A) < \infty$ is the largest eigenvalue of A . Applying Ito's lemma to the map $\mathbf{u} \mapsto \|\mathbf{u}\|_A^2$, we have for any $\varepsilon > 0$ and $0 \leq t \leq T < \infty$,

$$\begin{aligned} \mathbb{E}[\|\mathbf{u}_\varepsilon(t)\|_A^2] &= -2\gamma \int_0^t \mathbb{E} \left[(2\mathbf{u}_\varepsilon(s)^\top A \mathbf{Z}_s + \mathbf{Z}_s^\top A \mathbf{Z}_s) (\mathbf{u}_\varepsilon(s)^\top A \mathbf{Z}_s) \right] ds + \|\mathbf{u}(0)\|_A^2 \\ &\leq -2\gamma \int_0^t \mathbb{E} \left[(\mathbf{Z}_s^\top A \mathbf{Z}_s) (\mathbf{u}_\varepsilon(s)^\top A \mathbf{Z}_s) \right] ds + \|\mathbf{u}(0)\|_A^2 \\ &\leq 2\gamma \int_0^t \mathbb{E} \left[\|\mathbf{u}_\varepsilon(s)\|_A \|\mathbf{Z}_s\|_A^3 \right] ds + \|\mathbf{u}(0)\|_A^2 \\ &\leq C \int_0^t \mathbb{E} \left[\|\mathbf{u}_\varepsilon(s)\|_A^2 + \|\mathbf{Z}_s\|_A^6 \right] ds + \|\mathbf{u}(0)\|_A^2 \\ &\leq C \int_0^t \mathbb{E}[\|\mathbf{u}_\varepsilon(s)\|_A^2] ds + C(t + \varepsilon(1 - e^{-t/\varepsilon})) + \|\mathbf{u}(0)\|_A^2 \\ &\leq C(1 + \varepsilon + \|\mathbf{u}(0)\|_A^2) e^{Ct} < \infty \end{aligned}$$

where C is a constant independent of ε that changes from line to line. The second inequality follows using that $-(\mathbf{u}_\varepsilon^\top A \mathbf{Z})^2 \leq 0$, the third from Cauchy-Schwarz, and the fourth follows from Young's inequality. The fifth line follows from substituting and integrating the estimate for $\mathbb{E}[\|\mathbf{Z}_t\|^p]$ computed above, and the final line follows from an application of Gronwall's inequality.

Since $\mathbb{E}[\|\mathbf{u}_\varepsilon(t)\|^2] \leq (1/\lambda_{\min}(A)) \mathbb{E}[\|\mathbf{u}_\varepsilon(t)\|_A^2]$, the coordinates of $\mathbf{u}_\varepsilon(t)$ individually have bounded second moments for all $\varepsilon > 0$, and $\sup_{\varepsilon \in (0,1]} \mathbb{E}[\pi_i^2(\mathbf{u}_\varepsilon(t))] \leq C(1 + \sup_{\varepsilon \in (0,1]} \varepsilon e^{Ct}) < \infty$ for all i . Hence, $\mathbb{E}[\mathbf{u}_\varepsilon^i(t)] \rightarrow U^i(t)$ for each i , and so $\mathbb{E}[\mathbf{u}_\varepsilon(t)] \rightarrow U(t)$ as $\varepsilon \rightarrow 0$.