# 17 Dynamical synchronization of truth and model as an approach to data assimilation, parameter estimation, and model learning

Gregory S. Duane and Joseph J. Tribbia

National Center for Atmospheric Research, Boulder, Colorado,
gduane@ucar.edu

## 1 Synchronized Chaos and Data Assimilation

### 1.1 Synchronized chaos

Synchronization of weakly coupled oscillators has been known since the time of Huygens (1673), who observed that pendulum clocks hung on a common wall tend to fall into antisynchronized motion. Synchronization of regular oscillators with limit cycle attractors is in fact ubiquitous in Nature (Strogatz 2003). Only recently however, has the synchronization of chaotic oscillators become known. The phenomenon was first brought to light by Fujisaka and Yamada (1983) and independently by Afraimovich, Verichev, and Rabinovich (1987), but extensive research on the subject in the '90s was spurred by the seminal work of Pecora and Carroll (1991), who considered configurations such as the following combination of Lorenz systems:

$$\begin{aligned} \dot{X} &= \sigma(Y - X) & & \\ \dot{Y} &= \rho X - Y - XZ & \dot{Y}_1 &= \rho X - Y_1 - XZ_1 \\ \dot{Z} &= -\beta Z + XY & \dot{Z}_1 &= -\beta Z_1 + XY_1 \end{aligned} \tag{1}$$

which synchronizes rapidly, slaving the $Y_1, Z_1$-subsystem to the master $X, Y, Z$-subsystem, as seen in Fig. 17.1, despites differing initial conditions and despite sensitive dependence on initial conditions.

If we imagine that the first Lorenz system represents the world, and that the second Lorenz system is a predictive model, then synchronization effects *data assimilation* (Daley 1991) of observed variables into the running model. The only observed variable in the foregoing example is $X$, but that is sufficient to cause the desired convergence of model to truth. Synchronization is known to be tolerant of reasonable levels of noise, as might arise in the observation channel, and occurs with partial coupling schemes that do not completely replace a model variable with a variable of the observed system.

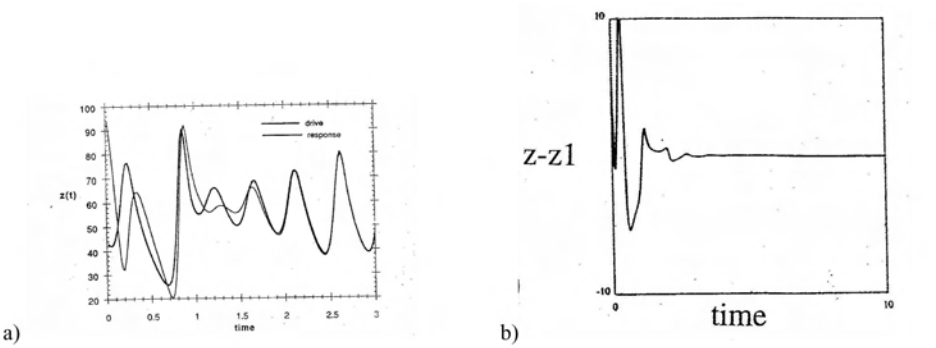


**Fig. 17.1.** The trajectories of the synchronously coupled Lorenz systems in the Pecora-Carroll complete replacement scheme (1) rapidly converge (a). Differences between corresponding variables approach zero (b).

Specifically, systems can also synchronize when coupled *diffusively*, as with a pair of bidirectionally coupled Rossler systems:

$$\begin{aligned} \dot{X} &= -Y - Z + \alpha(X_1 - X) & \dot{X}_1 &= -Y_1 - Z_1 + \alpha(X - X_1) \\ \dot{Y} &= X + aY & \dot{Y}_1 &= X_1 + aY_1 \\ \dot{Z} &= b + Z(X - c) & \dot{Z}_1 &= b + Z_1(X_1 - c) \end{aligned} \tag{2}$$

where $\alpha$ parametrizes the coupling strength. Contrary to naive expectations, the tendency to synchronize does not increase monotonically with $\alpha$, but synchronization will indeed occur for sufficiently large $\alpha$. The diffusive coupling also works unidirectionally.

For a pair of coupled systems that are not identical, synchronization may still occur, but the correspondence between the states of the two systems in the synchronized regime is different from the identity. In this situation, known as *generalized synchronization*, we have two different dynamical systems

$$\dot{\mathbf{x}} = F(\mathbf{x}) \tag{3}$$

$$\dot{\mathbf{y}} = G(\mathbf{y}) \tag{4}$$

with $\mathbf{x} \in R^N$ and $\mathbf{y} \in R^N$. If the dynamics are modified so as to couple the systems:

$$\dot{\mathbf{x}} = \hat{F}(\mathbf{x}, \mathbf{y}) \tag{5}$$

$$\dot{\mathbf{y}} = \hat{G}(\mathbf{y}, \mathbf{x}) \tag{6}$$

the systems are said to be generally synchronized iff there is some invertible function $\Phi : R^N \rightarrow R^N$ such that $||\Phi(\mathbf{x}) - \mathbf{y}|| \rightarrow 0$ as $t \rightarrow \infty$. Identical synchronization may be transformed to generalized synchronization simply by a change of variables

in one system, but not the other, i.e. a change in the description of one of the systems (Rulkov, Sushchik, and Tsimring 1995). In this situation, the correspondence function $\Phi$ is known *a priori*. Generalized synchronization may be difficult to detect without prior knowledge of $\Phi$.

Synchronization reduces the effective dimension of the phase space by half. Once synchronized, the 6-dimensional system (2) evolves in an 3-dimensional hyperplane, and is further constrained, as $t \to \infty$, to evolve on a strange attractor within that hyperplane. (For the coupled Lorenz systems (1), an auxiliary variable $X_1$ could be added, satisfying $\dot{X}_1 = \sigma(Y_1 - X_1)$, to define a system that is 6-dimensional before synchronization.) With *generalized synchronization* of nonidentical systems, the hyperplane becomes a *synchronization manifold* defined by an invertible correspondence function $\Phi : R^N \to R^N$. The $N$-dimensional manifold in $2N$-dimensional space is $\mathcal{M} \equiv \{(p, \Phi(p)) | p \in R^N\}$. The synchronization manifold is dynamically invariant: If $x(t)$ is a trajectory of a system such as (1) or (2), for $x \in R^{2N}$, and $x(t_1) \in \mathcal{M}$, then $x(t_2) \in \mathcal{M}$ for all $t_2 > t_1$. That is, a perfectly synchronized system remains synchronized.

It is commonly not the existence, but the stability of the synchronization manifold that distinguishes coupled systems exhibiting synchronization from those that do not (such as (2) for different values of $\alpha$). $N$ Lyapunov exponents can be defined for perturbations in the $N$-dimensional space that is transverse to the synchronization manifold $\mathcal{M}$. If the largest of these, $h^{\perp}_{max}$, is negative, then motion in the synchronization manifold is stable against transverse perturbations. In that case, the coupled systems will synchronize for some range of differing initial conditions. However, since $h^{\perp}_{max}$ only determines *local* stability properties, the size of the basin of attraction for the synchronized regime remains unknown. As $h^{\perp}_{max}$ is increased through zero, the system undergoes a *blowout bifurcation*. For small positive values of $h^{\perp}_{max}$, on-off synchronization occurs (a special case of on-off intermittency), as illustrated in Fig. 17.2b, where degradation results from a time lag in the coupling.

Synchronization is surprisingly easy to arrange, occurring for a wide range of coupling types. Synchronization degrades through on-off intermittency or through generalized synchronization. In the former case, vestiges of synchronization are discernible even far from the blowout bifurcation point (Duane 1997). Generalized synchronization is known to occur even when the systems are very different, as in the case of a Lorenz system diffusively coupled to a Rossler system. The two systems with attractors of different dimension are known to synchronize, but the correspondence function is not smooth (Pecora, Carroll, Johnson, Mar, and Heagy 1997).

The phenomenon of chaos synchronization is not restricted to low-dimensional systems. It is known, for instance, that two $D$-dimensional Generalized Rossler systems (each equivalent to a Rossler system for $D = 3$) will synchronize for any $D$, no matter how large, when coupled via only one of the $D$ variables:

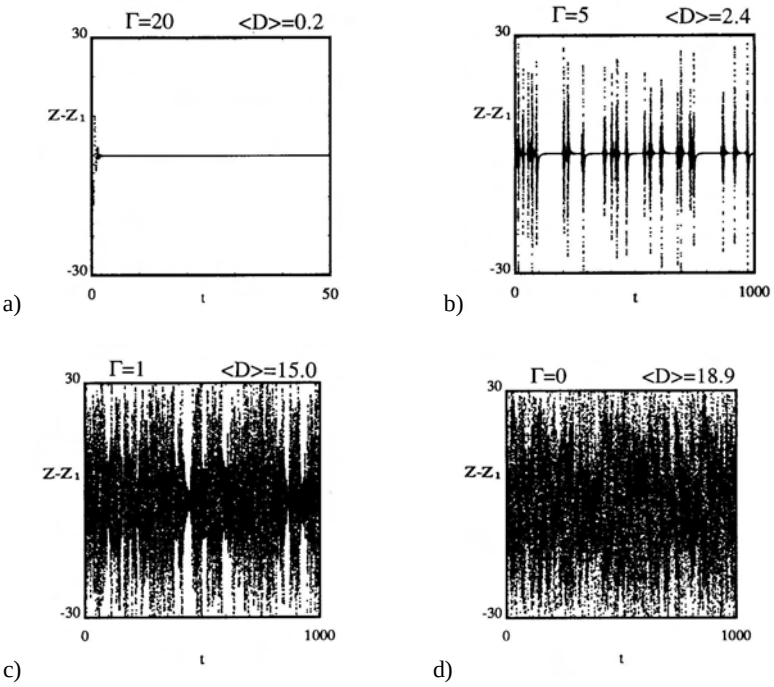


**Fig. 17.2.** The difference between the simultaneous states of two Lorenz systems with time-lagged coupling, represented by $Z(t) - Z_1(t)$ vs. t for various values of the inverse time-lag $\Gamma$ illustrating complete synchronization (a), intermittent or "on-off" synchronization (b), partial synchronization (c), and de-coupled systems (d). Average euclidean distance $\langle D \rangle$ between the states of the two systems in $X, Y, Z$-space is also shown. The trajectories are generated by adaptive Runge-Kutta numerical integrations with $\sigma = 10.$, $\rho = 28.$, and $\beta = 8/3$.

$$\begin{aligned} \dot{x}_1^A &= -x_2^A + \alpha x_1^A + x_1^B - x_1^A & \dot{x}_1^B &= -x_2^B + \alpha x_1^B + x_1^A - x_1^B \\ \dot{x}_i^A &= x_{i-1}^A - x_{i+1}^A & \dot{x}_i^B &= x_{i-1}^B - x_{i+1}^B \qquad i = 2 \ldots D-1 \\ \dot{x}_D^A &= \epsilon + \beta x_D^A (x_{D-1}^A - d) & \dot{x}_D^B &= \epsilon + \beta x_D^B (x_{D-1}^B - d) \end{aligned} \tag{7}$$

Each system has an attractor of dimension $\approx D - 1$, for $D$ greater than about 40, and a large number of positive Lyapunov exponents that increases with $D$.

The existence of synchronized chaos in naturally occuring systems was made more plausible by demonstrations of synchronization in spatially extended systems governed by PDE's (Kocarev, Tasev, and Parlitz 1997) Synchronization in geophysical fluid models was demonstrated by Duane and Tribbia (2001), originally with a view toward predicting and explaining new families of long-range teleconnections (Duane and Tribbia, 2004). Their findings are discussed in Section 3.2.

Please provide citation for fig 17.3

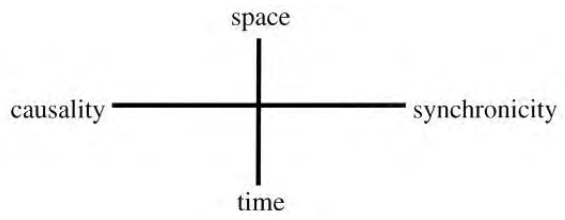


**Fig. 17.3.** Diagram constructed by Carl Jung, later modified by Wolfgang Pauli, to suggest relationships based on synchronicity as an "acausal connecting principle", existing alongside causal relationships.

### 1.2 Synchronization-based data assimilation and Jungian synchronicity

Since the problem of data assimilation arises in any situation requiring a computational model of a parallel physical process to track that process as accurately as possible based on limited input, it is suggested here that the broadest view of data assimilation is that of machine perception by an artificially intelligent system. Like a data assimilation system, the human mind forms a model of reality that functions well, despite limited sensory input, and one would like to impart such an ability to the computational model. In the artificial intelligence view of data assimilation, the additional issue of model error can be approached naturally as a problem of machine learning, as discussed in the concluding section.

In this more general context, the role of synchronism is reminiscent of the psychologist Carl Jung's notion of synchronicity in his view of the relationship between mind and the material world. Jung had noted uncanny coincidences or "synchronicities" between mental and physical phenomena. In collaboration with Wolfgang Pauli (Jung and Pauli, 1955), he took such relationships to reflect a new kind of order connecting the two realms. The new order was taken to explain relationships between seemingly unconnected phenomena in the objective world as well, exising alongside the familiar order based on causality (3). It was important to Jung and Pauli that synchronicities themselves were distinct, isolated events, but as described in Sect. 1.1, such phenomena can emerge naturally as a degraded form of chaos synchronization.

A principal question that is addressed in this chapter is whether the synchronization view of data assimilation is merely an appealing reformulation of standard treatments, or is different in substance. The first point to be made is that all standard data assimilation approaches, if successful, do achieve synchronization, so that synchronization defines a more general family of algorithms that includes the standard ones. It remains to determine whether there are synchronization schemes that lead to faster convergence than the standard data assimilation algorithms. It has been shown analytically that optimal synchronization is equivalent to Kalman filtering when the dynamics change slowly in phase space, so that the same linear approximation is valid at each point in time for the real dynamical system and its model. When the dynamics change rapidly, as in the vicinity of a regime transition, one must consider the full nonlinear equations and there are better synchronization strategies than the one given by Kalman filtering or ensemble Kalman filtering. The deficiencies of the standard methods, which are well

known in such situations, are usually remedied by ad hoc corrections, such as "covariance inflation"(Anderson, 2001). In the synchronization view, such corrections can be derived from first principles.

## 2 Synchronization vs. conventional data assimilation in the nonlinear realm

### 2.1 Optimal coupling for synchronization

To compare synchronization to standard data assimilation, Duane, Tribbia, and Weiss (2006) inquired as to the coupling that is optimal for synchronization, so that this coupling could be compared to the gain matrix used in the standard 3dVar and Kalman filtering schemes. The general form of coupling of truth to model that we consider in this section is given by a system of stochastic differential equations:

$$\begin{aligned} \dot{\mathbf{x}}_T &= f(\mathbf{x}_T) \\ \dot{\mathbf{x}}_B &= f(\mathbf{x}_B) + \mathbf{C}(\mathbf{x}_T - \mathbf{x}_B + \boldsymbol{\xi}) \end{aligned} \tag{8}$$

where true state $\mathbf{x}_T \in R^n$ and the model state $\mathbf{x}_B \in R^n$ evolve according to the same dynamics, given by $f$, and where the noise $\boldsymbol{\xi}$ in the coupling (observation) channel is the only source of stochasticity. The form (8) is meant to include dynamics $f$ described by partial differential equations, as in the last section. The system is assumed to reach an equilibrium probability distribution, centered on the synchronization manifold $\mathbf{x}_B = \mathbf{x}_T$. The goal is to choose a time-dependent matrix $\mathbf{C}$ so as to minimize the spread of the distribution.

Note that if $\mathbf{C}$ is a projection matrix, or a multiple of the identity, then Eq. (8) effects a form of nudging. But for arbitrary $\mathbf{C}$, the scheme is much more general. Indeed, continuous-time generalizations of 3DVar and Kalman filtering can be put in the form (8).

Let us assume that the dynamics vary slowly in state space, so that the Jacobian $\mathbf{F} \equiv Df$, at a given instant, is the same for the two systems

$$Df(\mathbf{x}_B) = Df(\mathbf{x}_T) \tag{9}$$

where terms of $O(\mathbf{x}_B - \mathbf{x}_T)$ are ignored. Then the difference between the two Eqs. (8), in a linearized approximation, is

$$\dot{\mathbf{e}} = \mathbf{F}\mathbf{e} - \mathbf{C}\mathbf{e} + \mathbf{C}\boldsymbol{\xi} \tag{10}$$

where $\mathbf{e} \equiv \mathbf{x}_B - \mathbf{x}_T$ is the synchronization error.

The stochastic differential equation (10) implies a deterministic partial differential equation, the Fokker-Planck equation, for the probability distribution $\rho(\mathbf{e})$:

$$\frac{\partial \rho}{\partial t} + \nabla_{\mathbf{e}} \cdot [\rho(\mathbf{F} - \mathbf{C})\mathbf{e}] = \frac{1}{2}\delta \nabla_{\mathbf{e}} \cdot (\mathbf{C}\mathbf{R}\mathbf{C}^T \nabla_{\mathbf{e}} \rho) \tag{11}$$

where $\mathbf{R}=<\boldsymbol{\xi}\boldsymbol{\xi}^T>$ is the observation error covariance matrix, and $\delta$ is a time-scale characteristic of the noise, analogous to the discrete time between molecular kicks in a Brownian motion process that is represented as a continuous process in Einstein's well known treatment. Equation (11) states that the local change in $\rho$ is given by the divergence of a probability current $\rho(\mathbf{F}-\mathbf{C})\mathbf{e}$ except for random "kicks" due to the stochastic term.

The PDF can be taken to have the Gaussian form $\rho=N\exp(-\mathbf{e}^T\mathbf{K}\mathbf{e})$, where the matrix $\mathbf{K}$ is the inverse spread, and $N$ is a normalization factor, chosen so that $\int \rho d^n\mathbf{e}=1$. For background error covariance $\mathbf{B}$, $\mathbf{K}=(2\mathbf{B})^{-1}$. In the one-dimensional case, $n=1$, where $C$ and $K$ are scalars, substitution of the Gaussian form in Eq. (11), for the stationary case where $\partial\rho/\partial t=0$ yields:

$$2B(C-F)=\delta RC^2 \tag{12}$$

Solving $dB/dC=0$, it is readily seen that $B$ is minimized ($K$ is maximized) when $C=2F=(1/\delta)B/R$.

In the multidimensional case, $n>1$, the relation (12) generalizes to the *fluctuation-dissipation relation*

$$\mathbf{B}(\mathbf{C}-\mathbf{F})^T+(\mathbf{C}-\mathbf{F})\mathbf{B}=\delta\mathbf{C}\mathbf{R}\mathbf{C}^T \tag{13}$$

that can be obtained directly from the stochastic differential equation (10) by a standard proof that is reproduced in Appendix A. $\mathbf{B}$ can then be minimized element-wise. Differentiating the matrix equation (13) with respect to the elements of $\mathbf{C}$, we find

$$\begin{aligned} &d\mathbf{B}(\mathbf{C}-\mathbf{F})^T+\mathbf{B}(d\mathbf{C})^T+(d\mathbf{C})\mathbf{B}+(\mathbf{C}-\mathbf{F})d\mathbf{B} \\ &=\delta[(d\mathbf{C})\mathbf{R}\mathbf{C}^T+\mathbf{C}\mathbf{R}(d\mathbf{C})^T] \end{aligned} \tag{14}$$

where the matrix $d\mathbf{C}$ represents a set of arbitrary increments in the elements of $\mathbf{C}$, and the matrix $d\mathbf{B}$ represents the resulting increments in the elements of $\mathbf{B}$. Setting $d\mathbf{B}=0$, we have

$$[\mathbf{B}-\delta\mathbf{C}\mathbf{R}](d\mathbf{C})^T+(d\mathbf{C})[\mathbf{B}-\delta\mathbf{R}\mathbf{C}^T]=0 \tag{15}$$

Since the matrices $\mathbf{B}$ and $\mathbf{R}$ are each symmetric, the two terms in Eq. (15) are transposes of one another. It is easily shown that the vanishing of their sum, for arbitrary $d\mathbf{C}$, implies the vanishing of the factors in brackets in Eq. (15). Therefore $\mathbf{C}=(1/\delta)\mathbf{B}\mathbf{R}^{-1}$, as in the 1D case.

### 2.2 Synchronization vs. conventional data assimilation in the linear realm

Turning now to the standard methods, so that a comparison can be made, it is recalled that the analysis $\mathbf{x}_A$ after each cycle is given by:

$$\begin{aligned} \mathbf{x}_A&=\mathbf{R}(\mathbf{R}+\mathbf{B})^{-1}\mathbf{x}_B+\mathbf{B}(\mathbf{R}+\mathbf{B})^{-1}\mathbf{x}_{\rm obs} \\ &=\mathbf{x}_B+\mathbf{B}(\mathbf{R}+\mathbf{B})^{-1}(\mathbf{x}_{\rm obs}-\mathbf{x}_B) \end{aligned} \tag{16}$$

In 3dVar, the background error covariance matrix $\mathbf{B}$ is fixed; in Kalman filtering it is updated after each cycle using the linearized dynamics. The background for the next cycle is computed from the previous analysis by integrating the dynamical equations:

$$\mathbf{x}_B^{n+1} = \mathbf{x}_A^n + \tau f(\mathbf{x}_A^n) \tag{17}$$

where $\tau$ is the time interval between successive analyses. Thus the forecasts satisfy a difference equation:

$$\mathbf{x}_B^{n+1} = \mathbf{x}_B^n + \mathbf{B}(\mathbf{R}+\mathbf{B})^{-1}(\mathbf{x}_{\rm obs}^n - \mathbf{x}_B^n) + \tau f(\mathbf{x}_A^n) \tag{18}$$

We model the discrete process as a continuous process in which analysis and forecast are the same:

$$\begin{aligned}\dot{\mathbf{x}}_B = f(\mathbf{x}_B) + 1/\tau \mathbf{B}(\mathbf{B}+\mathbf{R})^{-1}(\mathbf{x}_T - \mathbf{x}_B + \boldsymbol{\xi}) \\ + O[(\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1})^2]\end{aligned} \tag{19}$$

using the white noise $\boldsymbol{\xi}$ to represent the difference between observation $\mathbf{x}_{\rm obs}$ and truth $\mathbf{x}_T$. The continuous approximation is valid so long as $f$ varies slowly on the time-scale $\tau$.

It is seen that when background error is small compared to observation error, the higher order terms $O[(\mathbf{B}(\mathbf{B}+\mathbf{R})^{-1})^2]$ can be neglected and the optimal coupling $\mathbf{C}=1/\delta\mathbf{B}\mathbf{R}^{-1}$ is just the form that appears in the continuous data assimilation equation (19), for $\delta=\tau$. Thus under the linear assumption that $Df(\mathbf{x}_B)=Df(\mathbf{x}_T)$, the synchronization approach is equivalent to 3dVar in the case of constant background error, and to Kalman filtering if background error is dynamically updated over time. The equivalence can also be shown for an exact description of the discrete analysis cycle, by comparing it to a coupled pair of synchronized maps. See Duane et al. (2006) Appendix B.

### 2.3 Universal covariance inflation factors in the nonlinear realm

In the fully nonlinear case, the optimal coupling scheme for synchronization may differ from that used in standard data assimilation methods.

In a region of state space where nonlinearities are strong and Eq. (9) fails, the prognostic equation for error (10) must be extended to incorporate nonlinearities. Additionally, model error due to processes on small scales that escape the digital representation should be considered. While errors in the parameters or the equations for the explicit degrees of freedom require deterministic corrections, the unresolved scales, assumed dynamically independent, can only be represented stochastically. The physical system is governed by:

$$\dot{\mathbf{x}_T} = f(\mathbf{x}_T) - \boldsymbol{\xi}_M \tag{20}$$

in place of Eq. (8a), where $\boldsymbol{\xi}_M$ is model error, with covariance $\mathbf{Q}\equiv<\boldsymbol{\xi}_M\boldsymbol{\xi}_M^T>$. The error equation (10) becomes

$$\dot{\mathbf{e}} = (\mathbf{F}-\mathbf{C})\mathbf{e} + G\mathbf{e}^2 + H\mathbf{e}^3 + \mathbf{C}\boldsymbol{\xi} + \boldsymbol{\xi}_M \tag{21}$$

where we have included terms up to cubic order in $\mathbf{e}$, with $H<0$ to prevent divergent error growth for large $||\mathbf{e}||$. In the multi-dimensional case, Eq. (21) is shorthand for a tensor equation in which $G$ and $H$ are tensors of rank three and rank four (and the restrictions on $H$ are more complex). In the one-dimensional case, which we shall analyze here, $G$ and $H$ are scalars.

The Fokker-Planck equation is now:

$$\frac{\partial \rho}{\partial t} + \nabla_e \cdot \{\rho[(F-C)e + Ge^2 + He^3]\} = \frac{1}{2}\delta \nabla_e \cdot [(CRC^T + Q)\nabla_e \rho] \tag{22}$$

Using the ansatz for the PDF $\rho$:

$$\rho(e) = N \exp(-Ke^2 - Le^3 - Me^4) \tag{23}$$

with a normalization factor $N = [\int_{-\infty}^{\infty} de\ \exp(-Ke^2 - Le^3 - Me^4)]^{-1}$, we obtain from Eq. (22) the following relations between the dynamical parameters and the PDF parameters:

$$\begin{aligned} F - C &= \frac{1}{2}\tau(C^2R + Q)(-2K) \\ G &= \frac{1}{2}\tau(C^2R + Q)(-3L) \\ H &= \frac{1}{2}\tau(C^2R + Q)(-4M) \end{aligned} \tag{24}$$

The goal is to minimize the background error:

$$B(K, L, M) = \frac{\int_{-\infty}^{\infty} de\ e^2 \exp(-Ke^2 - Le^3 - Me^4)}{\int_{-\infty}^{\infty} de\ \exp(-Ke^2 - Le^3 - Me^4)}. \tag{25}$$

Using Eq. (24) to express the arguments of $B$ in terms of the dynamical parameters, we find $B(K, L, M){=}B(K(C), L(C), M(C)){\equiv}B(C)$, and can seek the value of $C$ that minimizes $B$, for fixed dynamical parameters $F, G, H$.

The coupling that gives optimal synchronization can again be compared with the coupling used in standard data assimilation, as for the linear case. In particular, one can ask whether the "covariance inflation" scheme that is used as an ad hoc adjustment in Kalman filtering (Anderson 2001) can reproduce the $C$ values found to be optimal for synchronization. The form $C{=}\tau^{-1}B(B{+}R)^{-1}$ is replaced by the adjusted form

$$C = \frac{1}{\tau}\frac{\mathcal{F}B}{\mathcal{F}B + R} \tag{26}$$

where $\mathcal{F}$ is the covariance inflation factor.

The optimization problem was solved numerically in the one-dimensional case (Duane et al. 2006) with results as plotted in Table 17.1. If the function $f$ in the dynamical equation describes motion in a two-well potential, with minima at distances $d_1$ and $d_2$ from the unstable fixed point, it can be shown that the dynamical parameters in (21) are s $G{=}1/d_2{-}1/d_1$ and $H{=}{-}1/(d_1 d_2)$. Results are displayed in the table for

**Table 17.1.** Covariance inflation factor vs. bimodality parameters $d_1, d_2$, for 50% model error in the resolved tendency.

| | $d_1$ | | | | | |
|---|---|---|---|---|---|---|
| | .75 | 1. | 1.25 | 1.5 | 1.75 | 2. |
| .75 | 1.26 | 1.26 | 1.28 | 1.30 | 1.32 | 1.34 |
| 1. | 1.26 | 1.23 | 1.23 | 1.25 | 1.27 | 1.29 |
| $d_2$ 1.25 | 1.28 | 1.23 | 1.22 | 1.23 | 1.24 | 1.25 |
| 1.5 | 1.30 | 1.25 | 1.23 | 1.22 | 1.23 | 1.24 |
| 1.75 | 1.32 | 1.27 | 1.24 | 1.23 | 1.23 | 1.23 |
| 2. | 1.34 | 1.29 | 1.25 | 1.24 | 1.23 | 1.23 |

a range of values of the bimodality parameters $d_1$ and $d_2$. The model error in Eq. (20) is chosen to be about 50% of the resolved tendency $\dot{x}_T$, with the resulting model error covariance $Q{=}0.04$ approximately one-fourth of the background error covariance $B$. The covariance inflation factors are remarkably constant over a wide range of parameters and agree with typical values used in operational practice.

## 3 Automatic parameter estimation and model learning in synchronization-based data assimilation

### 3.1 Identical synchronization implies parameter estimation for non-identical systems

Machine learning might also be realized in the synchronization context, so as to correct for deterministic model error in the resolved degrees of freedom. By allowing model parameters to vary slowly, generalized synchronization would be transformed to more nearly identical synchronization. Indeed, parameter adaptation laws can be added to a synchronously coupled pair of systems so as to synchronize the parameters as well as the states. Parlitz (1996) showed for example that two unidirectionally coupled Lorenz systems with different parameters:

$$\begin{aligned} \dot{X} &= \sigma(Y - X) & \dot{X}_1 &= \sigma(Y - X_1) \\ \dot{Y} &= \rho X - Y - XZ & \dot{Y}_1 &= \rho_1 X_1 - \nu Y_1 - X_1 Z_1 + \mu \\ \dot{Z} &= -\beta Z + XY & \dot{Z}_1 &= -\beta Z_1 + X_1 Y_1 \end{aligned} \tag{27}$$

could be augmented with parameter adaptation rules:

$$\begin{aligned} \dot{\rho}_1 &= (Y - Y_1)X_1 \\ \dot{\nu} &= (Y_1 - Y)Y_1 \\ \dot{\mu} &= Y - Y_1 \end{aligned} \tag{28}$$

so that the Lorenz systems would synchronize, and additionally $\rho_1 \to \rho$, $\nu \to 1$, and $\mu \to 0$.

Equations for a synchronously coupled pair of systems can in fact always be augmented to allow parameter adaptation as well, provided that relevant dynamical variables are observed, as shown by Duane, Yu, and Kocarev (2007). Consider a "real system" given by ODE's:

$$\dot{\mathbf{x}} = \boldsymbol{f}(\mathbf{x}, \mathbf{p}), \tag{29}$$

$$\dot{\mathbf{p}} = 0, \tag{30}$$

where $\mathbf{x} \in I\!R^N$, $\boldsymbol{f} : I\!R^N \rightarrow I\!R^N$, and $\mathbf{p} \in I\!R^m$ is the vector of (unknown, constant) parameters of the system. We further assume that $\mathbf{s} = \boldsymbol{h}(\mathbf{x})$, where $\boldsymbol{h} : I\!R^N \rightarrow I\!R^n$, $n \leq N$, is an $n$ dimensional vector representing the experimental measurement output of the system. A "computational model" of the system (29) is given by:

$$\dot{\mathbf{y}} = \boldsymbol{f}(\mathbf{y}, \mathbf{q}) + \boldsymbol{u}(\mathbf{y}, \mathbf{s}), \tag{31}$$

$$\dot{\mathbf{q}} = \boldsymbol{N}(\mathbf{y}, \mathbf{x} - \mathbf{y}), \tag{32}$$

where $\boldsymbol{N}(\mathbf{y}, 0) = 0$, and $\boldsymbol{u}$ is the control signal. Let $\mathbf{e} \equiv \mathbf{y} - \mathbf{x}$ and $\mathbf{r} \equiv \mathbf{q} - \mathbf{p}$. Choose a positive definite Lyapunov function $L_o(\mathbf{e})|_{\mathbf{q}=\mathbf{p}}$. Assume that the control signal $\boldsymbol{u}$ is designed such that there is some time $t_0$ for which $\dot{L}_o(\mathbf{e}(t))|_{\mathbf{q}=\mathbf{p}} < 0$ when $\mathbf{e}(t) \neq 0$ and $\dot{L}_o(\mathbf{e}(t))|_{\mathbf{q}=\mathbf{p}} = 0$ when $\mathbf{e}(t) = 0$, for all $t > t_0$. That is, after time $t_0$, the system proceeds monotonically toward synchronization. Let $\boldsymbol{h} \equiv \boldsymbol{f}(\mathbf{y}, \mathbf{r} + \mathbf{p}) - \boldsymbol{f}(\mathbf{y} - \mathbf{e}, \mathbf{p})$. Duane et al. (2007) established the following theorem:

**Theorem 1.** *Assume that (i) the control law $\boldsymbol{u}$ in (31) is designed such that the synchronization manifold $\mathbf{x} = \mathbf{y}$ is globally asymptotically stable, (ii) $\boldsymbol{f}$ is linear in the parameters* **p***, and (iii) the parameter estimation law (32) is designed such that*

$$N_j = -\delta_j \sum_i \left( \frac{\partial L_o}{\partial e_i} \right) \left( \frac{\partial h_i}{\partial r_j} \right),$$

*where $\delta_j$ are positive constants. Then the synchronization manifold $\mathbf{y} = \mathbf{x}$, $\mathbf{p} = \mathbf{q}$ is globally asymptotically stable.*

The theorem ensures the stability of the synchronization manifold $\mathbf{y} = \mathbf{x}, \mathbf{p} = \mathbf{q}$. It says that if the two systems synchronize for the case of identical parameters, then the parameters of the "real system" can be estimated when they are not known *a priori*, provided that each partial derivative $\partial L_o / \partial e_i$ is known for which the vector $\partial h_i / \partial r_j$ $(j = 1, ....)$ is not zero. For the usual form $L_o \equiv \sum_i (e_i)^2$, the requirement is that $x_i$ be known if the equation for $\dot{y}_i$ contains parameters that one seeks to estimate. By considering a more general Lyapunov function that is defined in terms of some subset $S$ of the state variables, or their indices, $L_o \equiv \sum_{i \in S} c_i (e_i)^2$ for positive coefficients $c_i$, one obtains the looser requirement for each desired parameter, that $x_i$ be known for at least some $i$ for which the $\dot{y}_i$ equation contains that parameter. (Convergence may be slower if fewer $x_i$ are known.)

### 3.2 Parameter estimation in geophysical fluid models

Generalizations to PDEs would allow parameters in geophysical models to automatically adapt. In general, some of the partial derivatives $\partial L_o/\partial e_i$ may not be known and Theorem 1 may be inapplicable. The theorem gives us no way to estimate the parameters $\rho$ or $b$ for the Lorenz system, for instance, if only $x_1$ is coupled, while $x_2$ and $x_3$ (hence $\partial L_o/\partial e_2$ and $\partial L_o/\partial e_3$) are unknown. But in the case of translationally invariant PDE's, the parameters are the same at each point in space. In general, they can be estimated from a limited amount of information about the state at a discrete set of points (or for a finite set of Fourier components), if such information is also sufficient to give identical synchronization globally or locally when coupled to a "model system". Consider a pair of spatially extended systems that are synchronously coupled in one direction:

$$\begin{aligned} \frac{\partial \phi^A}{\partial t} + \Gamma(\phi^A) &= f^A F(\phi^A, \phi^A) \\ \frac{\partial \phi^B}{\partial t} + \Gamma(\phi^B) + C(\phi^A, \phi^B) &= f^B F(\phi^B, \phi^A), \end{aligned} \tag{33}$$

where $\Gamma$ is some general form in the field $\phi$, possibly involving spatial derivatives, $C$ is a similarly general form coupling $\phi^B$ to $\phi^A$, such that $C(\phi, \phi) = 0$, and the form $F$ with coefficient $f$ specifies a "forcing", which in general may also couple to the other system. The two systems are dynamically identical when $\phi^A = \phi^B$ and $f^A = f^B$.

Define a core Lyapunov function $L_o(\phi^A - \phi^B)|_{f^A = f^B} \equiv \int d^3x\ (\phi^A - \phi^B)^2$. Then a ready generalization of Theorem 1, to the case of a continuum of state variables tells us that if we impose the parameter estimation law

$$\dot{f}^B = \int d^3x\ (\phi^A - \phi^B) F(\phi^B, \phi^A), \tag{34}$$

we will find $f^B \to f^A$ as in the ODE case.

The procedure outlined above estimates a coefficient of forcing for a wide class of synchronously coupled PDE's, and is readily generalized for limited measurements of $\phi^A$. The quasigeostrophic potential vorticity equation, for instance, used to describe the large-scale atmospheric circulation, has been shown to exhibit high-quality synchronization when two copies are coupled via partial exchange of only mid-range Fourier components of the flow field (Duane and Tribbia 2001, Duane and Tribbia 2004). The synchronization manifold is either globally attracting or locally attracting in a very wide basin. Synchronization of such systems may be useful for meteorological data assimilation (Duane 2003). It is important that forcing parameters in the equation can also be estimated.

Consider a two-layer channel model (derived from one described by Vautard et al. (1988)), with flow evolution given by the quasigeostrophic equation for potential vorticity $q$ on a $\beta$-plane:

$$\frac{Dq_l}{Dt} \equiv \frac{\partial q_l}{\partial t} + J(\psi_l, q_l) = F_l + D_l \tag{35}$$

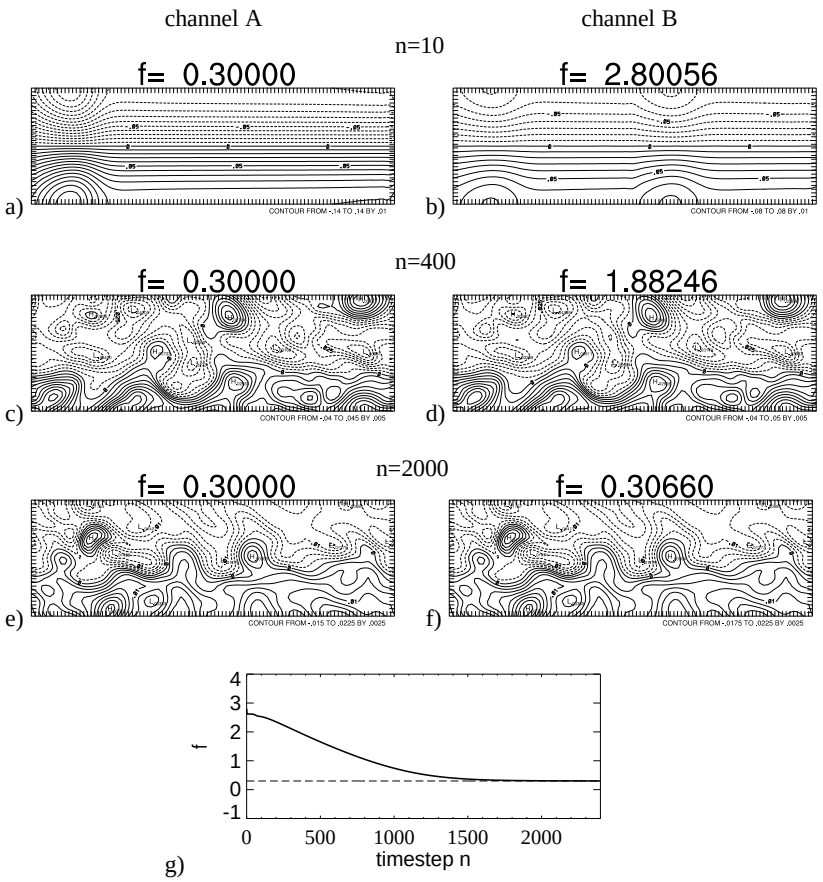


**Fig. 17.4.** The evolving flow $\psi$ (a-f) for two quasigeostrophic channel models that are synchronously coupled as in (Duane and Tribbia 2001; Duane and Tribbia 2004) (but in one direction only), and with the forcing parameter $f^B$ for the second channel (denoted $\mu_o$ in the reference) allowed to vary according to the truncated parameter adaptation rule (37) with $S = \{\mathbf{k} : \mathbf{k}_x, \mathbf{k}_y \leq 12\}$ (in waves per channel- length). Starting from the initial value $f^B = 3.0$ at time step $n = 0$ (not shown), $f^B$ converges (g) to the value of the corresponding parameter $f^A = 0.3$ (dashed line) in the first channel, as the flows synchronize. (An average of the two layers $l = 1, 2$ is shown.)

where the layer $l$=1, 2, $\psi$ is streamfunction, and the Jacobian $J(\psi, \cdot)$=$\frac{\partial \psi}{\partial x}\frac{\partial \cdot}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \cdot}{\partial x}$ gives the advective contribution to the Lagrangian derivative $D/Dt$. Equation (35) states that potential vorticity is conserved on a moving parcel, except for forcing $F_l$ and dissipation $D_l$. The discretized potential vorticity is $q_l = f_0 + \beta y + \nabla^2 \psi_l + R_l^{-2}(\psi_1 - \psi_2)(-1)^i$ where $f(x, y)$ is the vorticity due to the Earth's rotation at each point $(x, y)$, $f_0$ is the average $f$ in the channel, $\beta$ is the constant $df/dy$ and $R_l$ is the Rossby radius of deformation in each layer. The forcing $F$ is a relaxation term designed to induce a jet-like flow near the beginning of the channel: $F_l$=$f(q_l^* - q_l)$ for $q_l^*$ corresponding to a choice of $\psi^*$ that resembles the flow in Fig. 17.4a, for example.

The dissipation terms $D$, boundary conditions, and other parameter values are given in Duane and Tribbia (2004). One might seek to estimate the coefficient $f$.

Consider two systems, "A" and "B", both given by equations of the form (35), but with the forcing in the B system defined differently, in terms of its spectral components $F^B_{\mathbf{k}}$, so as to effect a unidirectional coupling:

$$F^B_{\mathbf{k}} = f^B \sum_{\mathbf{k}} a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + f^B \sum_{\mathbf{k}} b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}}) \tag{36}$$

where the layer index $l$ is suppressed and the coefficients $a_{\mathbf{k}}, b_{\mathbf{k}}$ are slightly smoothed step functions of $\mathbf{k}$, so that each spectral component is either coupled to the corresponding component in the A system or to the background flow $q^*$ or niether. The coefficients are chosen so as to couple only the medium-scale components:

$$b_{\mathbf{k}} = \begin{cases} 0 & \text{if } |k_x| \le k_{x0} \text{ and } |k_y| \le k_{y0} \\ (k_n/|\mathbf{k}|)^4 & \text{if } |\mathbf{k}| > k_n \\ 1 - (k_0/|\mathbf{k}|)^4 & \text{otherwise} \end{cases}$$

and to force only the large-scale components:

$$a_{\mathbf{k}} = \begin{cases} 1 - b_{\mathbf{k}} & \text{if } |\mathbf{k}| \le k_n \\ 0 & \text{if } |\mathbf{k}| > k_n \end{cases}$$

as in (Duane and Tribbia 2004), where the constants $k_0, k_{x0}, k_{y0}$ and $k_n$ are defined. The systems thus coupled synchronize, without bursting, as shown in Fig. 17.4. (The forcing for the A system is correspondingly truncated: $F^A_{\mathbf{k}} = f^A \sum_{\mathbf{k}} a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^A_{\mathbf{k}})$.)

The parameter estimation rule in spectral space:

$$\dot{f}^B = \sum_{\mathbf{k} \in S} (q^A_{\mathbf{k}} - q^B_{\mathbf{k}})[a_{\mathbf{k}}(q^*_{\mathbf{k}} - q^B_{\mathbf{k}}) + b_{\mathbf{k}}(q^A_{\mathbf{k}} - q^B_{\mathbf{k}})] \tag{37}$$

is the Fourier transform of (34) if $S$ is universal. But even for a restricted range of wavenumbers in $S$, as in the figure, the rule (37) causes $f^B$ to converge to $f^A$ as would follow from the use of a correspondingly restricted Lyapunov function.

As a more realistic example than the channel model described above, the Weather Research and Forecasting (WRF) model was considered, as adapted for weather prediction over military test ranges for the Army Test and Evaluation Command (ATEC). The ATEC application is based on observations that are so frequent that they can be assumed to occur at every numerical time step, so that a continuously coupled differential equation system of the form (8) or (19) can be taken to reflect the actual data assimilation scenario.

At a relevant level of model detail, the prognostic equation for humidity (water vapor mixing ratio) $q$ is:

$$\frac{\partial q}{\partial t} = \frac{\partial}{\partial z}\{K(\frac{\partial q}{\partial z} - Mf(u_0, T_0, ...))\} \tag{38}$$

where $K$ is a moisture diffusivity, and $M = M(x, y)$ quantifies the impact of soil moisture at each location $(x, y)$, which is a function $f$ of state variables $u_0, T_0$, etc. at the surface. To study the estimation of $M$ using the synchronization method, attention is restricted to a single vertical column $(x, y) = (x_0, y_0)$ and a model is introduced that is diffusively coupled to ("nudged"by) the true state. The model humidity $q_m$, for instance, is governed by:

$$\frac{\partial q_m}{\partial t} = \frac{\partial}{\partial z}\{K(\frac{\partial q_m}{\partial z} - Mf(u_{m0}, T_{m0}, ...))\} + c(q_{obs} - q_m) \tag{39}$$

where $q_{obs}$ is the observed humidity (at any level $z$ where an observation is taken) that is the sum of the true $q$ and observational noise. $c$ is a coupling ("nudging") coefficient. Similar equations govern the evolution of temperature $T$, wind speed $u$, and other model variables, but the parameter $M$ is thought to enter only the humidity equation (39).

In accordance with the general rule (34), $M$ for the model was made to vary with observational input as:

$$\dot{M} \sim -\frac{\partial K}{\partial z} f(u_{m0}, T_{m0}, ...)(q_{obs}(z) - q_m(z)) \tag{40}$$

for the case of observations taken at just one level $z$.

Repeated convergence of $M$ to its true value, each time followed by a burst away from synchronization, is seen in Fig. 17.5.

### 3.3 Stochasticity for global optimization and model learning: a heuristic example

In complex cases, where there are multiple zeros of $\dot{L}$ corresponding to local optima, a stochastic component (in the parameters) may be necessary to allow parameters to jump among multiple basins of attraction so as to reach the global optimum (Fig. 17.6).

To illustrate the utility of added noise in optimizing synchronization patterns, we turn to a synchronized oscillator ("cellular neural network (CNN)") representation of the travelling salesman problem. The representation generalizes an older one due to Hopfield and Tank (1985) a traditional neural network, fully interconnected, with fixed weights chosen so that the globally optimal state corresponds to a shortest-distance "tour"among a collection of cities with a pre-specified distance for each pair of cities. In this original representation, for a collection of 5 cities, one considers a $5 \times 5$ matrix of binary values, where the rows correspond to cities and the columns correspond to slots, 1 through 5, in the tour schedule. A tour is any pattern of 0's and 1's, such that there is exacly one 1 in each row (exactly one city visited at a time) and one 1 in each column (each city visited exactly once). For instance the tour depicted in Fig. 17.7a is ECABD. There is a 10-fold degeneracy in optimal patterns (shortest-distance cyclic tours) due to arbitrariness in the selection of the starting city and the direction of the tour. The travelling salesman problem, in this representation or any other, is difficult because of the multiplicity of local optima.

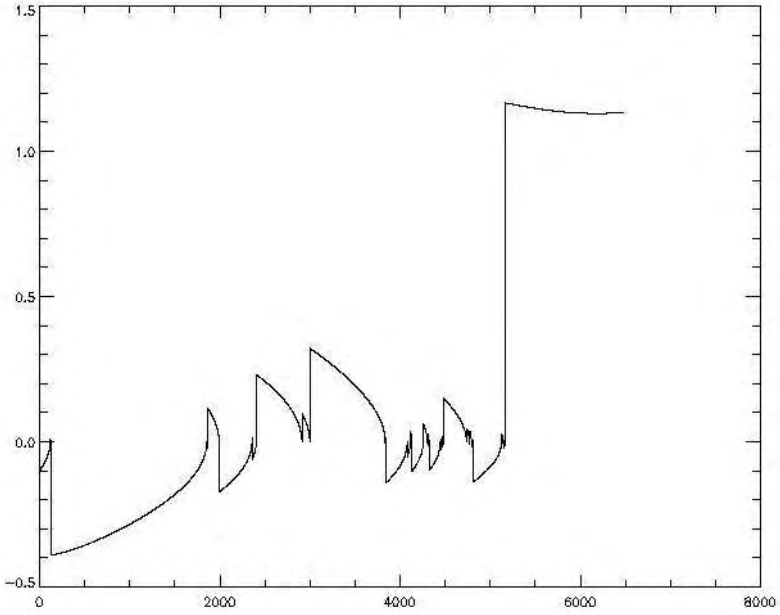


**Fig. 17.5.** The variable model parameter $M$ converges to the true value $M_T$ with repeated "bursting", in the dynamical parameter adaptation scheme that only requires a single realization ($M - M_T$ is plotted). Results are unstable, but even for this case, where observations are only assimilated at one point, the correct value of the parameter can be identified. (State variables also do not converge completely over the time interval shown.)

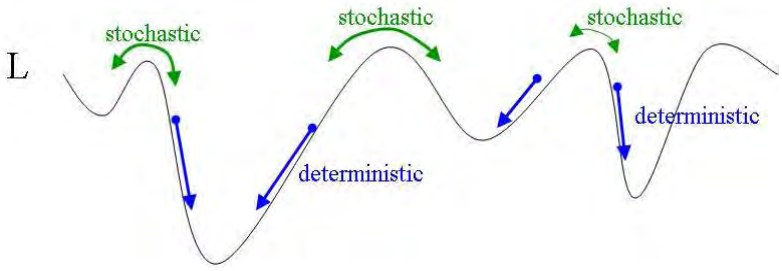


**Fig. 17.6.** Deterministic parameter estimation rules cause parameters to reach local optima of the Lyapunov function $L$. Stochasticity (e.g. in a simulated annealing algorithm) allows jumps among different basins of attraction.

The representation considered here is a $5 \times 5$ array of coupled periodic oscillators. A tour is specified by a synchronization pattern in which all oscillators in each row and each column are desynchronized, and such that for each oscillator in each column (row), there is exactly one oscillator in each column (row) that is synchronized with it. In Fig. 17.7b, the same tour ECABD is depicted in the new representation, simultaneously

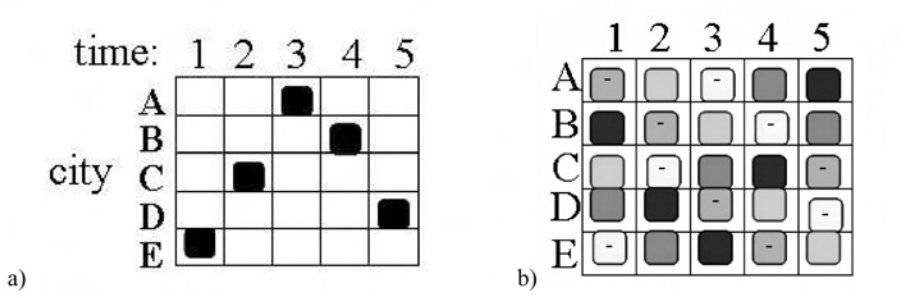


**Fig. 17.7.** In the original Hopfield representation (a), a tour (here ECABD) is specified by the collection of units that are "on", provided there is only one "on" unit in each row and only one in each column. In the CNN representation (b) equivalent cyclic permutations of the same tour are simultaneously represented as sets of 5 units that are synchronized, with an analogous proviso. Units with the same relative phase are shown in the same color. (Dash marks are included to distinguish ambiguous colors in the grayscale version.)

with equivalant tours given by cyclic permutations of cities: CABDE, ABDEC, etc. There is now only a two-fold degeneracy in optimal patterns, due to arbitrariness in direction.

To solve the travelling salesman problem in the representation we have described, let each oscillator be given by a complex number $z_{ij}$ ($i = 1, \ldots 5, j = A, \ldots E$) that contains both a phase $\arg(z_{ij})$ and an amplitude $|z_{ij}|$. Assume all oscillators have the same frequency $\omega$ and make the replacement $z_{ij} \rightarrow \exp(-i\omega t) z_{ij}$, so that only the relative phases are represented in the complex quantities $z_{ij}$. The Lyapunov function we seek to minimize is:

$$L = A \sum_{ij} (|z_{ij}|^2 - 1)^2 + B \sum_{ij} (\left( \frac{z_{ij}}{|z_{ij}|} \right)^5 - 1)^2 - C \sum_i \sum_{jj'} \left| \frac{z_{ij}}{|z_{ij}|} - \frac{z_{ij'}}{|z_{ij'}|} \right|^2$$
$$-D \sum_j \sum_{ii'} \left| \frac{z_{ij}}{|z_{ij}|} - \frac{z_{i'j}}{|z_{i'j}|} \right|^2 + E \sum_i \sum_j \sum_{j'} d_{jj'} \mathrm{Re}(\frac{z_{ij}}{|z_{ij}|} \frac{z^*_{i+1,j'}}{|z_{i+1,j'}|}) \qquad (41)$$

where asterisks denote complex conjugates, and we close the tours by defining $z_{6j} \equiv z_{1j}$. The first term, with coefficient $A$, tends to force $|z| = 1$. The second term, with coefficient $B$ tends to force $z$ to one of five phase states, corresponding to the fifth roots of unity. The terms with coefficients $C$ and $D$ penalize for synchronization within each row and within each column, respectively. The last term, expressed in terms of distances $d_{jj'}$ between cities $j$ and $j'$, is easily seen to be minimized when each partial sum $\sum^{sync}$ over synchronized oscillators at $i, j$ and $i+1, j'$ (i.e. for which $\arg(z_{ij}) = \arg(z_{i+1,j'})$), $\sum^{sync}_{ijj'} d_{jj'}$ is minimized, that is when the specified tour has shortest distance.

If the oscillators are governed by equations

$$\dot{z}_{ij} = -\frac{\partial L}{\partial z^*_{ij}}, \qquad \dot{z}^*_{ij} = -\frac{\partial L}{\partial z_{ij}} \qquad (42)$$

**Table 17.2.** Relative phases (from $-\pi$ to $\pi$) of oscillators in the $5 \times 5$ CNN representation of the 5-city travelling salesman problem, after reaching a steady state. No tours are evident.

| | slot in schedule | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| A | -2.47002 | -2.18725 | -0.14433 | 1.28413 | 1.28601 |
| B | 2.49589 | -1.22075 | 1.40448 | -2.48385 | -0.01186 |
| city C | 1.23951 | 2.64314 | -2.61880 | 0.00870 | -1.22135 |
| D | 0.00406 | 1.08844 | -2.63690 | -1.27138 | 2.53321 |
| E | -1.29235 | 1.11239 | 0.04947 | 2.51944 | -2.52968 |

**Table 17.3.** Relative phases as in Table 17.2, but with noise of steadily decreasing amplitude included in the oscillator dynamics (43). The tour found, ADEBC, is seen by focussing attention on oscillators with relative phase $-2\pi/5 \approx -1.28$. Permutations of the same tour are represented by other collections of oscillators with approximately equal relative phases. ADEBC is indeed the shortest-distance tour for the pre-specifi ed table of distances ($d_{jj'}$) that was arbitrarily selected.

| | slot in schedule | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| A | -1.28206 | 0.03570 | 1.23972 | 2.49383 | -2.52120 |
| B | 1.23280 | 2.48519 | -2.49996 | -1.31980 | -0.01645 |
| city C | 0.03447 | 1.27973 | 2.52424 | -2.53673 | -1.25935 |
| D | -2.49808 | -1.29343 | -0.01308 | 1.19502 | 2.50160 |
| E | 2.49330 | -2.42328 | -1.19018 | 0.04572 | 1.28507 |

following Hoppensteadt (1996), then the derivative of the Lyapunov function $\dot{L} = \sum_{ij}[(\partial L/\partial z_{ij})\ \dot{z}_{ij} + (\partial L/\partial z^*_{ij})\ \dot{z}^*_{ij}] = -2\sum_{ij}|z_{ij}|^2 \leq 0$, so that the cost function is monotonically decreasing and must reach at least a local minimum, since $L$ can be easily seen to be bounded below.

The point of this lengthy description is that the network as described almost never reaches a global minimum, or even a state defining a tour. For a given set of prespecified distances (not shown) the network reaches a phase state shown in Table 17.2. However, if gaussian noise is added to the dynamics

$$\dot{z}_i = -\frac{\partial L}{\partial z^*_{ij}} + \xi, \qquad \dot{z}^*_{ij} = -\frac{\partial L}{\partial z_{ij}} + \xi^* \tag{43}$$

where $\xi$ represents complex gaussian noise, then the system attains the tour state shown in Table 17.3, which furthermore is the shortest-distance tour, provided that the amplitude of the noise (an analog "temperature") must be decreased slowly enough. This technique of "simulated annealing", which had also been applied in the original Hopfield representation, is effective in selecting one of a small number of global optima (precisely $2 \times 5! = 240$) out of $5^{25}$ phase states. (The term in (41) restricting the phase of each oscillator to one of five values, can be dropped, but the the temperature must be lowered even more slowly to attain a globally optimal pattern.)

The complexity of this combinatorial example far exceeds that of most parameter estimation problems, in which there are usually only a small number of local optima.

But the example suggests that simulated annealing could perhaps be extended to a genetic algorithm that would make random qualitative changes in the model as well, until synchronization is achieved. The approach to model learning thus defined is in keeping with the suggestion that synchronicity is a useful paradigm for the relationship between reality and a computational model.

**Acknowledgments**

The authors thank Jeff Weiss, Dongchuan Yu, Ljupco Kocarev, and Josh Hacker for useful discussions. This work was partially supported under NSF Grant 0327929.

## References

Afraimovich, V. S., Verichev, N. N., and Rabinovich, M. I. (1986) Stochastic synchronization of oscillation in dissipative systems. Radiophys. Quantum Electron. 29, 795-803.

Anderson, J. L. (2001) An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev. 129, 2884-2903.

Anderson, J. L. (2003) A local least-squares framework for ensemble filtering. Mon. Wea. Rev. 131, 634-642.

Daley, R. (1991) *Atmospheric Data Analysis*, Cambridge Univ. Press, Cambridge.

Duane, G.S. (1997) Synchronized chaos in extended systems and meteorological teleconnections. Phys. Rev. E 56, 6475-6493.

Duane, G.S. (2003) Synchronized chaos in climate dynamics, in: In, V., Kocarev, L., Carroll, T.L. et al. (Eds.), *AIP Conference Proceedings 676*, Melville, New York, pp. 115-126.

Duane, G.S. and Tribbia, J.J. (2001) Synchronized chaos in geophysical fluid dynamics. Phys. Rev. Lett. 86, 4298-4301.

Duane, G.S. and Tribbia, J.J. (2004) Weak Atlantic-Pacific teleconnections as synchronized chaos. J. Atmos. Sci. 61, 2149-2168.

Duane, G.S., Tribbia, J.J., and Weiss, J.B. (2006) Synchronicity in predictive modelling: a new view of data assimilation. Nonlin. Processes in Geophys. 13, 601-612.

Duane, G.S., Yu, D., and Kocarev, L. (2007) Identical synchronization, with translation invariance, implies parameter estimation. submitted to Phys. Lett. A.

Fujisaka, H. and Yamada, T. (1983) Stability theory of synchronized motion in coupled-oscillator systems. Prog. Theor. Phys. 69, 32-47.

Hopfield, J.J. and Tank, D.W. (1985)"Neural" computation of decisions in optimization problems. Biol. Cybern. 52, 141-152.

Hoppensteadt, F.C. (1996) Synaptic organizations and dynamical properties of weakly connected neural oscillators 2. Learning phase information. Biol. Cybern. 75, 129-135.

Huygens, C. (1673)*Horoloqim Oscillatorium.* Apud. F. Muget.

Jung, C. G. and Pauli, W. (1955) The interpretation of nature and the psyche, Pantheon, New York.

Kocarev, L., Tasev, Z., and Parlitz, U. (1997) Synchronizing spatiotemporal chaos of partial differential equations. Phys. Rev. Lett. 79, 51-54.

Lorenz, E. N. (1963) Deterministic nonperiodic flows. J. Atmos. Sci. 20, 130-141.

Parlitz, U. (1996) Estimating model parameters from time series by autosynchronization. Phys. Rev. Lett. 76, 1232-1235.

Parlitz, U. and Kocarev, L. (1997) Using surrogate data analysis for unmasking chaotic communication systems. Int. J. Bifurcations and Chaos 7, 407-413.

Pecora, L. M. and Carroll, T. L. (1990) Synchronization in chaotic systems. Phys. Rev. Lett. 64, 821-824.

Pecora, L. M., Carroll, T. L., Johnson, G. A., Mar, D. J., and Heagy, J. F. (1997) Fundamentals of synchronization in chaotic systems, concepts, and applications. Chaos 7, 520-543.

Rulkov, N. F., Sushchik, M. M., and Tsimring, L. S. (1995) Generalized synchronization of chaos in directionally coupled chaotic systems. Phys. Rev. E 51, 980-994.

So, P., Ott, E., and Dayawansa, W. P. (1994) Observing chaos – deducing and tracking the state of a chaotic system from limited observation. Phys. Rev. E 49, 2650-2660.

Strogatz, S. H. (2003) *Sync: The Emerging Science of Spontaneous Order*. Theia, New York, 338 pp., 2003.

Vautard, R., Legras, B., and Déqué, M. (1988) On the source of midlatitude low-frequency variability. Part I: A statistical approach to persistence. J. Atmos. Sci. 45, 2811-2843.

Von Der Malsburg, C. and Schneider, W. (1986) A neural cocktail-party processor. Biol. Cybernetics 54, 29-40.

Yang, S.-C., Baker, D., Cordes, K., Huff, M., Nagpal, G., Okereke, E., Villafañe, J., and Duane, G. S. (2006) Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system. J. Atmos. Sci. 63, 2340-2354.