MANAGEMENT OF COMPLEX DYNAMICAL SYSTEMS

R.S.MACKAY

Abstract. Complex dynamical systems are systems with many interdependent components which evolve in time. One might wish to control their trajectories, but a more practical alternative is to control just their statistical behaviour. In many contexts this would be both sufficient and a more realistic goal, e.g. climate and socio-economic systems. I refer to it as "management" of complex dynamical systems. In this paper, some mathematics for management of complex dynamical systems is developed in the weakly dependent regime, and questions are posed for the strongly dependent regime.

1. Introduction

The last 30 years have seen a surge of interest in the dynamics of complex systems, e.g. [BY, NBW, NN, BBV]. By a "complex dynamical system" I mean a deterministic or stochastic dynamical system with many interdependent components. I will often think of the components as associated with spatial locations, but space could just be a space of labels for the components.

This interest has extended to *control* of complex dynamical systems, e.g. [Si, A+, CDK, IP, VJ, LB].

Here, I concentrate on *management* of complex dynamical systems. By "management" I mean control of the spatiotemporal statistics of the system rather than individual trajectories. For example, management of the weather means control of the climate as opposed to control of the weather. It is close to what [LB] call "control of collective behaviour" and [W3] call "probability density function shaping".

Some potential areas of application, albeit ambitious, are geo-engineering the climate, managing vegetation patterns in the Sahel, electricity demand management, routing in telecommunications networks, high confinement mode in magnetically confined nuclear fusion plasmas, epidemic prevention, famine relief, demographics, health service organisation, and other domains of government policy.

The paper follows the programme proposed in [M1] and Section 9 of [SM+], of which some aspects were developed in [M2, DM, M3, BuM, SM].

2. Setting

Many deterministic dynamical systems exhibit well defined temporal statistics, known variously as Gibbs states and Sinai-Ruelle-Bowen measures [Y]. The theory extends to some classes of spatially extended system, e.g. [J].

Statistical behaviour is simpler to study in stochastic systems, however. Furthermore, technicalities are reduced by studying discrete-time systems rather than continuous-time ones.

Date: October 5, 2017.

Thus I focus on probabilistic cellular automata (PCA). These are Markov chains on the product X of a set of state spaces X_s, one for each site s in a countable set S, which is called "space". The local state spaces X_s are assumed to be Polish spaces (i.e. complete, separable metric spaces) of bounded diameter. The metric on X_s is denoted d_s. For illustration, it is enough to consider $X_s = \{0, 1\}$ with $d_s(0, 1) = 1$. The Markov chain updates the state x_s at the sites $s \in S$ independently, given the current state $x = (x_r)_{r\in S}$ of the whole system. Thus it is specified by a family of probabilities[1] p^x_s for the next state x'_s at site s given the current state $x \in X$. For many examples in the literature, p^x_s is independent of the components of x outside some finite neighbourhood $N_s \subset S$ of s, but that is not an essential feature for us. All that is needed is that the dependence outside large finite sets in S should go to zero sufficiently fast that the transition operator P maps continuous functions on X (in product topology) to continuous functions (the Feller property). Actually, results can be extended to some global interactions, e.g. [FH], but I shall not consider that situation here.

The step I want to take here is to allow p^x_s to depend on time $t \in \mathbb{Z}$. This can represent the effects of external forces and of control.

Let us recall some basic results for Markov chains. If (i) the system is autonomous, (ii) X is finite, (iii) there is a unique communicating component, and (iv) it is aperiodic, then there is a unique stationary probability and it attracts all initial probabilities exponentially (in any metric, since all are equivalent in the finite case). The system is said to be *geometrically ergodic* (though "exponentially mixing" is a better term). It may be, however, that the time to approach the stationary probability goes to infinity as the system size goes to infinity. Thus even in the finite case it is of interest to derive quantitative bounds.

One nice way to derive quantitative bounds for geometric ergodicity is to use what I call *Dobrushin metric*, defined as follows. Give X the product topology and let $\mathcal{P}(X)$ be the space of Borel probability measures on X. For $\rho, \sigma \in \mathcal{P}(X)$, let

$$D(\rho, \sigma) = \|\rho - \sigma\|_Z, \tag{1}$$

where Z is the space of zero-charge Borel measures on X (μ *zero-charge* means $\mu(X) = 0$) and for $\mu \in Z$,

$$\|\mu\|_Z = \sup_{f\in F\setminus C} \frac{\mu(f)}{\|f\|_F},$$

where F is Dobrushin's [D2] space of functions $f : X \to \mathbb{R}$ which are continuous and have finite

$$\|f\|_F = \sum_{s\in S} \Delta_s(f),$$

where

$$\Delta_s(f) = \sup\left\{\frac{f(x) - f(x')}{d_s(x_s, x'_s)} : x_r = x'_r \ \forall r \neq s, x_s \neq x'_s\right\},$$

C denotes the constant functions, and $\mu(f) = \int f \, d\mu$. The metric D makes $\mathcal{P}(X)$ into a complete metric space (of diameter $\sup_{s\in S} \operatorname{diam}_s(X_s)$) [M2].

There are other ways to metrise $\mathcal{P}(X)$. Many are not useful in the present context, as described in [M2]. To these can be added the Lévy-Prokhorov metrics with respect

[1] I follow probability theorists in using "probability" as a shorthand for "probability distribution".

to the ℓ_∞ or ℓ_1 metrics on X, because they are at least 3 times the corresponding transportation metrics (Thm 1.1.3 of [BK]), which were shown in [M2] to be of no use. One metric on $\mathcal{P}(X)$ that is close in spirit to D was proposed and used by Steif [St]. In an Appendix I show that Dobrushin metric is the same as Steif's in the finite context, so I conjecture they are equal in general.

I explained in [M2] how to use metric (1) to derive Dobrushin's results on ergodicity for autonomous PCA [D2] (the derivation of Wasserstein's very similar results [Va] can also be streamlined using this metric). In this paper, I will show how to use it in the non-autonomous context.

3. Natural probability for non-autonomous PCA

Denote the time-dependent transition operator by $P_t, t \in \mathbb{Z}$, acting on functions $f : X \to \mathbb{R}$ to the right, or probabilities $\rho \in \mathcal{P}(X)$ to the left. Suppose for each time $t \in \mathbb{Z}$ there is a matrix $k(t) = (k_{sr})_{s,r \in S}$ such that for all $r, s \in S$ and states x, x' with $x_q = x'_q$ for all $q \neq s$, then

$$D_r(p_r^x, p_r^{x'}) \leq d_s(x_s, x'_s) k_{sr},$$

where D_r is transportation (Kantorovich-Rubinstein-Wasserstein-Hutchinson) metric on $\mathcal{P}(X_r)$, which can be defined by

$$D_r(p, p') = \|p - p'\|_r,$$

with, for zero-charge measures μ on X_r,

$$\|\mu\|_r = \sup \frac{\mu(g)}{\|g\|_{Lip}} \tag{2}$$

over non-constant Lipschitz functions $g : X_r \to \mathbb{R}$, where

$$\|g\|_{Lip} = \sup_{x \neq y \in X_r} \frac{g(x) - g(y)}{d(x,y)}$$

is the best Lipschitz constant for g. Furthermore, suppose there are $C \geq 1$ and $\gamma \in [0, 1)$ such that for all $t \in \mathbb{Z}$ and $n \geq 1$,

$$\|k(t) \ldots k(t+n-1)\|_1 \leq C\gamma^n. \tag{3}$$

I say such systems are in the "weakly dependent regime". Note that it suffices to require the left-hand side less than 1 for a single $n \geq 1$ to achieve a bound of this form. Note also that I have used the transpose of the common definition of k and the ℓ_1-norm ($\|k\|_1 = \sup_{r \in S} \sum_{s \in S} k_{sr}$), because the time-dependent generalisation of a key step in the proof of Theorem 3 of [M2] becomes

$$\Delta(P_t \ldots P_{t+n-1} f) \leq k(t) \ldots k(t+n-1) \Delta(f)$$

for the vectors of Lipschitz constants $\Delta_s(f)$ of a function $f \in F$, using component-wise partial order.

Consider sequences $\rho = (\rho_t)_{t \in \mathbb{Z}}$ of probabilities in $\mathcal{P}(X)$ with Dobrushin metric on each component and supremum metric over $t \in \mathbb{Z}$. Define an operator $\mathbb{P}$ on such sequences by

$$(\mathbb{P}\rho)_t = P_{t-1}\rho_{t-1}. \tag{4}$$

Then under the weak dependence condition (3), $\mathbb{P}^n$ is a contraction for all large enough $n \geq 1$. It follows that there is a unique time-dependent probability $\rho = (\rho_t)_{t\in\mathbb{Z}}$ such that for any initial time u, initial probability σ at time u, and $t \geq u$, then

$$D(\sigma P_u \dots P_{t-1}, \rho_t) \leq C\gamma^{t-u} D(\sigma, \rho_u).$$

I call the sequence ρ the *natural probability* for the process.

This is an extension of a result of Kolmogorov (reviewed in [BuM]) to the network context (Kolmogorov called ρ "absolute probability").

To obtain existence, uniqueness and exponential convergence to a natural probability, it is enough to assume slightly weaker conditions than above, namely existence of $C > 0$, $\gamma \in (0,1)$ such that for all $t > u$,

$$\|P_u \dots P_{t-1}\|_Z \leq C\gamma^{t-u}. \tag{5}$$

I call this the "exponentially mixing" regime (others would describe it as a spectral gap condition). It is implied by weak dependence. Condition (5) implies that the dynamics on probabilities are eventually contracting and hence the above results.

One framework into which it is helpful to put the system is the Banach space approach to uniformly hyperbolic deterministic dynamical systems. It was developed by Hirsch, Pugh & Shub, Coppel and many others in the 1970s, though can be traced back to Massera & Schaeffer (1958) and Perron (1928). For an exposition that treats the context of spatially extended dynamical systems, see [M3]. In the present context, consider the space of sequences $\rho = (\rho_t) + t \in \mathbb{Z}$ of probabilities with norm on tangent vectors being the supremum over $t \in \mathbb{Z}$ of the Z-norm at each t. The equation to be solved is $\rho_{t+1} = \rho_t P_t$, which can be written as

$$\rho(I - \mathbb{P}) = 0,$$

with $\mathbb{P}$ the operator (4). This has a unique bounded solution (in the subspace where $\rho_t(X) = 1$) iff (5). Note that in general an operator of the form $(I - \mathbb{P})$ having bounded inverse implies a splitting into backwards contracting and forwards contracting subspaces for the sequence $P = (P_t)_{t\in\mathbb{Z}}$, but when the P_t are probability transition operators there can be no backwards contracting subspace because it would imply some difference between probabilities would grow larger than the diameter of $\mathcal{P}(X)$ at some future time.

Next we consider how the natural probability ρ changes in response to small changes in the control, i.e. in P. I use supremum norm over $t \in \mathbb{Z}$ for both ρ and P. The answer is that ρ is C^1 with respect to the change in P, with derivative

$$\rho'_t = \rho_{t-1}P'_{t-1} + \rho_{t-2}P'_{t-2}P_{t-1} + \rho_{t-3}P'_{t-3}P_{t-2}P_{t-1} + \dots, \tag{6}$$

where $'$ denotes infinitesimal changes. The series converges geometrically in the Z-norm under the assumption (5) on P and bounded P' (which is implied by working with supremum norm on P).

The continuous differentiability of ρ and the formula (6) can be proved using the Banach space approach above: if $\rho(I - \mathbb{P}) = 0$ and $(I - \mathbb{P})$ has bounded inverse on Z then the implicit function theorem[2] gives that for nearby $\mathbb{P}(\varepsilon)$ there is a unique solution $\rho(\varepsilon)$, it depends as smoothly on ε as does $\mathbb{P}$, and its derivative is given by the chain rule: $\rho'(I - \mathbb{P}) - \rho\mathbb{P}' = 0$, so $\rho' = \rho\mathbb{P}'(I - \mathbb{P})^{-1}$. This can be expanded out to give formula (6).

[2] actually its affine version suffices

Note that under the assumption (5), equation (6) shows that the effect of control at a given time decays as time advances from that moment. If the dependence of the PCA is also local in space then one may expect the effect of spatially localised control to decay in space too. This is the topic of the next section.

Before treating that, we note that the natural probability is not a complete description of the spatiotemporal statistics of realisations of the PCA. To obtain the complete probability for space-time realisations one can use the natural probability as starting point at any time and then evolve realisations forward using the PCA. It is the same distinction as between solutions of the Fokker-Planck equation for a stochastic process in $\mathbb{R}^n$ and the probability over trajectories.

4. Spatiotemporal effects of control in the weakly dependent regime

Starting points for this section are the results of [BaM] on the response of a network to localised forces, and of [Gr, Ku] on decay of correlations for the classical statistical mechanics of spin systems.

We suppose that the update probability distribution at site s depends exponentially weakly on the state at site r, with respect to a metric on S, in a way that I shall define (the modified dependency matrix $\tilde{k}$ (8) is small enough that $I - \tilde{k}$ has bounded inverse). Bounded-range dependence is a special case.

As the first step, consider the set of sites in space-time, rather than just space or time. This was the approach of [M3], for example.

The results of [BaM] apply to a direct product of state spaces over the set of sites, and under an invertibility condition show that the response to an exponentially localised force is exponentially localised. In the present context, however, we are interested in a tensor product, namely the dual to the space of functions from the direct product to the reals. So some more work is required.

To allow more generality, we extend from PCA to random fields on space-time $M = S \times \mathbb{Z}$ with values in $\prod_s X_s$ for each $t \in \mathbb{Z}$. This permits for example, dependent updates like queue-swapping[3], and non-Markovian time-dependence like Gibbsian[4]. We even allow sets M with no interpretation as space-time, like statistical mechanical lattices as long as they are countable.

Thus as in [D1, Gr], the model is specified by conditional probabilities μ_m^x on Polish spaces $X_m, m \in M$, given the state $x \in X = \prod_{n \in M} X_n$, independent of x_m. A consistency condition is required, however, for the μ_m^x to be conditionals of a probability on X (not mentioned in [Gr]). One way to specify it is to require that the operators τ_m

[3]Think of being in a supermarket and the state is the number of people in each checkout queue; if someone changes queue then the updates to the numbers are dependent.

[4]A Gibbsian process in discrete time (autonomous and discrete state space) is a stochastic process for which

$$P(x_t | x_{t-1}, \ldots) \propto \prod_{n \geq 0} f_n(x_t, \ldots, x_{t-n})$$

as a function of x_t, with $\prod f_n^{n+1} < \infty$. The Nth-order Markov case is usually presented with $f_n = 1$ unless $n = N$ and $\sum_{x_t} f_N(x_t, \ldots, x_{t-N}) = 1$, but there is freedom to rewrite this to put the bulk of the variation into lower order f_n, including f_0. Gibbsian allows infinite history dependence as long as the dependence decays fast enough.

commute, where

$$(\tau_m f)(x) = \int_{X_m} f(x \vee_m \xi)\ d\mu_m^x(\xi),$$

on continuous functions $f : X \to \mathbb{R}$, and $x \vee_m \xi$ denotes the state obtained from $x \in X$ by changing x_m to ξ. Then the job is to find probabilities σ on X such that $\sigma(\tau_m f) = \sigma(f)$ for all $m \in M$ and continuous functions $f : X \to \mathbb{R}$.

As in [Gr], choose an enumeration of M and define T to be the limit of performing τ_m in sequence:

$$T = \dots \tau_2 \tau_1.$$

Then the job boils down to solving $\sigma T = \sigma$ for probabilities σ. This has a unique solution in probabilities σ if $I - T$ is invertible on zero-charge measures.

Dobrushin's condition $\|k\|_1 < 1$ makes T a contraction, using the metric (1). We outline a proof of this. Let k_{mn} be such that

$$D_n(\mu_n^x, \mu_n^{x'}) \leq d_m(x_m, x_m')k_{mn}$$

when x, x' agree off m. Then, generalising [Va] or [Gr], who work with only the discrete metrics on the X_m, but their proofs go through to our Polish spaces,

$$\Delta_m(Tf) \leq \sum_n k_{mn}\Delta_n(f).$$

Thus $\|T\| \leq \|k\|_1$, so if $\|k\|_1 < 1$ then $(I-T)$ is invertible with $\|(I-T)^{-1}\| \leq (1-\|k\|_1)^{-1}$. More generally, $I - T$ has bounded inverse if there exist $C > 0, \gamma \in (0, 1)$ such that $\|k^n\| \leq C\gamma^n$.

Whenever $I-T$ has bounded inverse we can deduce unique continuation of σ for small changes in T. Furthermore, σ changes smoothly with respect to parameters if T does, with

$$\sigma' = \sigma T'(I - T)^{-1}, \tag{7}$$

as can be deduced by applying the chain rule to $\sigma T = \sigma$.

The point of this section is to deduce bounds on the extent of the changes to σ visible at sites distant from those where changes to T are made. Suppose we can choose a semi-metric δ on M and $z > 1$ such that the matrix $\tilde{k}$ defined by

$$\tilde{k}_{mn} = z^{\delta(m,n)}k_{mn} \tag{8}$$

is bounded in ℓ_1-norm. We say k is "exponentially local". Suppose we make a change to T on only site b or a set of sites near b in the semi-metric δ. Making a change on only one site is feasible in the case of PCA, but in general the consistency condition on conditional probabilities requires one to make changes at other sites simultaneously, thus we treat the more general case. Now

$$T' = \sum_n \dots \tau_{n+1}\tau_n'\tau_{n-1}\dots\tau_1.$$

For a function $g \in F$,

$$|\tau_n' g(x)| \leq \|\tau_n'\|_n \Delta_n(g),$$

using the transportation norm (2) on zero-charge measures on X_n. Also

$$\Delta_n(\tau_{n-1}\dots\tau_1 g) \leq \Delta_n(g)$$

and

$$|\dots\tau_{n+1}g(x)| \leq \sup_{y\in X}|g(y)|.$$

Thus

$$|T'g(x)| \leq \sum_n \|\tau_n'\|_n \Delta_n(g). \tag{9}$$

Suppose we observe σ via functions $f \in F$ which are independent of the state off site a, or outside a neighbourhood of a in the semi-metric δ. We wish to apply (9) to $g = (I-T)^{-1}f$. Let

$$\tilde{\Delta}_n(g) = z^{\delta(n,a)}\Delta_n(g),$$

and $\tilde{\Delta}(g)$ denote the vector with these components. Then adapting [Gr] again to the more general metrics d_s,

$$\tilde{\Delta}_m(Tf) \leq \sum_n \tilde{k}_{mn}\tilde{\Delta}_n(f).$$

As in the lemma of section 5 of [BaM], $\tilde{k}$ depends continuously on z. If $\|k^n\|_1 \leq C\gamma^n$ for some $\gamma \in (0,1)$, then $I-k$ has bounded inverse (viz. $\sum_{n\geq 0} k^n$). Then as in Theorem 3 of [BaM], for all z near enough 1, $\|\tilde{k}-k\|_1 < \|(I-k)^{-1}\|^{-1}$, and for any such z, $(I-\tilde{k})$ has bounded inverse and

$$\|\tilde{\Delta}((I-T)^{-1}f)\|_1 \leq D^{-1}\|\tilde{\Delta}(f)\|_1, \tag{10}$$

where

$$D = \|(I-k)^{-1}\|^{-1} - \|\tilde{k}-k\|.$$

Apply σ to $T'g$ to deduce that

$$|\sigma'f| = |\sigma T'(I-T)^{-1}f| \leq \sum_n \|\tau_n'\|_n \Delta_n(g),$$

with, from (10),

$$\sum_n \Delta_n(g) z^{\delta(n,a)} \leq D^{-1}\|\tilde{\Delta}(f)\|_1.$$

Hence

$$|\sigma'f| \leq \sup_n \left(z^{-\delta(n,a)}\|\tau_n'\|_n\right)\; D^{-1}\|\tilde{\Delta}(f)\|_1,$$

which is of order $z^{-\delta(b,a)}\Delta_a(f)$.

This result shows that large weakly dependent systems can not be managed by control from one site.

Ideally I would like to define exponentially local for T rather than k. Then one could take the weaker hypothesis that $I-T$ is invertible rather than $I-k$ invertible. But I did not find a way to do this.

Note that applying [Gr, Ku] to space-time gives exponential spatiotemporal decay of correlations in the Dobrushin regime, using the discrete metric on each X_n. Their proofs can be extended straightforwardly to the regime where $I-k$ is invertible and with arbitrary metrics on the X_n (subject to keeping Polish and bounded diameter). Again it would be good to achieve this under the weaker hypothesis of $I-T$ invertible.

5. Steering space-time phases in the weakly dependent regime

Although the previous section shows that the (unique) space-time phase in a weakly dependent system with exponentially local dependence can not be controlled on a large scale from one point in space-time, one can use the formulae (6, 7) for derivatives with respect to parameters to determine how to change the update rules everywhere or in a sufficiently dense set of space-time to improve the statistical behaviour.

This section sets out a research programme.

Think first about improving the stationary distribution for an autonomous system. The formula $\rho' = \rho P'(I-P)^{-1}$ of [M2] tells us that if a government's objective functions are Φ_j, then $(\Phi_j\rho)' = \Phi_j'\rho P'(I-P)^{-1}$. The changes to the transition operator P may be of open-loop form, i.e. independent of current state, or of feedback form, i.e. taking into account the current state (or recent history, via the Gibbsian generalisation). The formula is deceptively simple. The hard part is understanding $(I-P)^{-1}$ sufficiently to work out the effects.

Next consider improving the spatiotemporal statistics, which can include things like reducing the mean time to be seen in accident and emergency. Then we can use the formula (7). Again, the hard part is to understand $(I-T)^{-1}$.

In practice, one has to think about the costs of management, not just the benefits. Suppose the system obtains a benefit $b(x)$ from being in state x but has to pay a cost $c_P(x)$ for using transition operator P. One measure of cost for changing the transition operator from P to $\tilde{P}$ could be $\sup_x \sum_s D_s(p_s^x, \tilde{p}_s^x)$. Then the mean net benefit per unit time in stationary probability ρ_P for P is $\rho_P(b) - \rho_P(c_P)$. Optimal control would consist in maximising this over feasible transition operators P. The first order condition for a maximum over P (preserving $P1 = 1$, where 1 is the function taking value 1 everywhere) is

$$\rho_P P'(I-P)^{-1}(b - c_P) = \rho_P(c_P'),$$

where $'$ denotes variations with respect to parameters of P. It would be good to develop ways of solving this.

Another context is a system subject to external forcing, so P is time-dependent, and we want to design time-dependent control to maximise the net benefit averaged over time. The response depends on how well we can forecast. A simple context in which the optimal behaviour can be decided with a finite look-ahead time is studied in [FMW]. A case of optimal control of a system without any forecasting is in [AV]. One might instead wish to optimise the predicted discounted future net benefit $\sum_{t\geq 0} \delta^t \rho_t(b(t) - c(t))$ for some discount factor $\delta \in (0,1)$.

One might also wish the benefits and costs to be more general functions of the sequence of states than just a sum for the state at each time.

There are clearly many directions to explore here.

6. Questions for strongly dependent systems

This paper has obtained results on management of PCA in the weakly dependent regime, extended to the exponentially mixing regime. There are spatially infinite PCA, however, for which there is more than one stationary probability. Toom's NEC majority voter PCA is a prime proved example [T] (for extensions, see [DM, SM]). There are also spatially infinite autonomous PCA for which there are time-dependent phases, therefore non-unique. Systems with non-unique phase are not in the exponentially mixing regime. For their finite versions, even if C and γ exist, C may be so large and γ so close to 1 that the results of the previous sections are useless.

In this section, a few suggestions are given about how one might address the big challenge of theory for the management of strongly dependent systems. Outside the exponentially mixing regime, there is the possibility that control from one site could have long-range effect in space-time and indeed this is what is observed in numerical simulations of Toom's PCA and variants. As argued in [M1], for stochastic systems, "nonlinearity" means "dependence" between components. Thus the issue is like the difference between weakly nonlinear systems (where for example a local Lyapunov function can be constructed to deduce asymptotic stability from an asymptotically stable linearisation) and strongly nonlinear systems (where multistability and chaos may occur).

In Toom's NEC majority voter PCA, the units are arranged in a square lattice and have state space $\{+,-\}$. At each time step each unit adopts the majority state of its North-East-Centre neighbourhood with probability $1-p$, and the opposite state with the remaining probability p. Toom proved that for p small enough there are at least two stationary probabilities, one with a density of $+$ larger than $\frac{1}{2}$, the other with density less than $\frac{1}{2}$. Numerically the threshold appears to be around $p = 0.09$. Furthermore, the phenomenon is robust to breaking the symmetry between $+$ and $-$, for example making the "error" probability p into $p_\pm$ according to the current state of the unit, or the majority state of its NEC neighbourhood. The region of two phases appears numerically to be bounded by a cusp curve, with vertex at the symmetric case with $p = 0.09$, the mainly $-$ phase jumping into the mainly $+$ one as one crosses the upper boundary, and the opposite for the lower boundary, e.g. [SM]. The boundary can be described as a "tipping point".

In particular, simulations of Toom's PCA starting in the mainly $-$ phase indicate that if one site is made a zealot, voting $+$ every time regardless of its neighbours, a plume of $+$ emerges from it, growing in a SE direction, as in Figure 1. This is an example of "pinning control".

Similar effects can be achieved with "boundary control". For example, taking the set of sites to be $\mathbb{Z}_-^2$ ($\mathbb{Z}_- = \{n \in \mathbb{Z} : n \leq 0\}$) and imposing various boundary conditions on the N and E borders can cause interesting patterns of invasion of $+$ phase into what was initially mainly $-$.[5]

Alternatively, starting in the mainly $-$ phase near the upper boundary of the region with two stationary probabilities, if one artificially turns a block of sites $+$ at one time and then lets the PCA run freely again, the block of $+$ can grow and turn the system into the mainly $+$ phase. This is an example of "nucleation" of a more stable phase. It was discussed extensively in [SM].

"Pinning control" and "nucleation" are two examples of what can generally be described as "nudges": relatively small and low density changes in the rules which can lead to large effects. They have been adopted by the UK government via the prime minister's Behavioural Insights team.

[5] Unfortunately the demonstration of this that I made in 2005 no longer works so I can not show a figure.

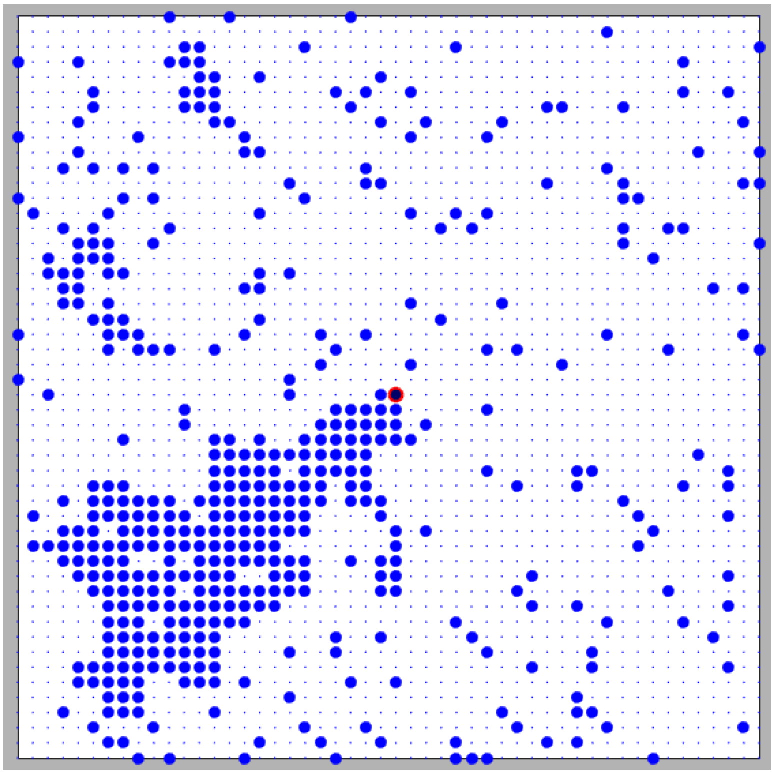

FIGURE 1. A plume formed by a zealot at the centre of Toom's NEC majority voter PCA. The error rates are 0.036 if the majority state for the NEC neighbourhood is blue, 0.067 if empty, and the system was started in the mainly empty phase.

Producing some theory on the effects of control on strongly dependent systems is a challenge. It is not even clear how to understand their dynamics in the autonomous case. Here is a starting point on the latter. Restrict attention to the extremal phases, i.e. those which are not a convex combination of phases. They are mutually singular, i.e. for each phase μ there is a subset A_μ of state space with full μ-measure and measure 0 for all other phases. Define the attractor for μ to be the smallest such A_μ. Define the basin B_μ of a phase μ to be the set of states from which realisations go to μ with probability 1. There can be states from which realisations go to some phases with probabilities strictly between 0 and 1 (so we can say they go to a convex combination of phases), or which never settle into any phase. What is the way to make sense of this picture?

Now consider the question of starting in attractor A_μ and designing control to push the system into the basin B_ν for some other more desirable phase ν. Some cost constraint should be imposed, else one could just solve it by flipping the state en masse to one in A_ν. So let us ask how to achieve it with minimal cost for some cost function of applying control. This is strongly reminiscent of Wentzell-Freidlin theory [FW], which treats deterministic dynamics subject to small noise and computes the rate of switching from one attractor to another in the regime of small noise. The transition paths are clustered around those of minimal cost, measured by the time-integral of the square of the noise force in the Gaussian case (in general, the negative logarithm of the noise probability density). Wentzell-Freidlin theory was extended to Markov chains on compact manifolds with some slow transitions [Ki], and to noisy globally coupled maps [Ha]. What is needed is to extend it to spatially extended Markov chains.

Another type of control question is how to prevent external inputs from tipping the system into a less desirable phase.

There are also questions about large finite systems, for which there may in principle be a unique phase but in practice it has several metastable phases mimicking an infinite system. Can the finite system size be viewed as providing noise that allows the system to transit from one phase to another?

7. Directions for other extensions

One can view the results of this paper as being about the spectral projection for a transition operator corresponding to spectrum near +1. Can one extend them to other spectral projections? This might allow the programme of [M0] for many-body quantum systems to be developed.

Another direction to pursue is that of mobile agents, whose interdependence is a function of proximity in real space. Thus the agents can be labelled in label-space but part of their state is position (and attitude) in real space and their interaction is via real space.

Next, one can consider agents who seek approximately to optimise some individual benefit but may also explore and learn (the exploration-exploitation trade-off). It may be reasonable to model them probabilistically and not necessarily making the optimal move at each time.

In the engineering world, one could also consider the question of optimal design of a system, including its response to external influences.

Lastly, I mention a curious aspect of control of probability distributions, namely that they can be affected by observations. In general, observing a system tightens up the probability distribution for its state. Indeed, for state x, model M, and observations O, Bayes' rule gives

$$P(x|M,O) = P(O|M,x)P(x|M)/Z(M,O).$$

This is the basis for data assimilation, which began with the Kalman filter and is now an immense field, e.g. [ABN].

Acknowledgements

This work was supported by the Alfred P. Sloan Foundation and by a Royal Society Wolfson Research Merit Award. I am grateful to Piotr Slowinski and Manuela Bujorianu for discussions on these issues and to Piotr for making the demonstration from which I printed out Figure 1.

Appendix: Comparison of Dobrushin metric with Steif's metric

Steif [St] defined the metric

$$\bar{d}(\mu,\nu) = \inf_{m\in\Pi(\mu,\nu)} \sup_{s\in S} m(d_s)$$

on probabilities on a product space $X = \prod_{s \in S} X_s$, where $\Pi(\mu, \nu)$ is the space of joinings of μ to ν (also known as "couplings"), i.e. probabilities on $X \times X$ whose marginal on the first factor is μ and on the second factor is ν, and $d_s(x, y) = 1$ if $x_s \neq y_s$, 0 otherwise. It extends a metric that Ornstein defined on translation invariant probabilities on $A^{\mathbb{Z}}$ for a finite alphabet A, to probabilities that are not translation invariant, and extends to product spaces which have no symmetry group like translations. Steif used it to prove exponential mixing of interacting particle systems in a weakly dependent regime.

Steif took the discrete (or "indicator") metric $d_s(x_s, y_s) = 1$ if $x_s \neq y_s$ on each component, but his metric can be generalised to probabilities on a product X of metric spaces (X_s, d_s) with more general metrics d_s, where $d_s(x, y)$ for $x, y \in X$ is interpreted as $d_s(x_s, y_s)$. One can allow any metrics d_s such that the X_s are Polish and of bounded diameter.

What is the relation between $\bar{d}$ thus defined, and D of (1)? I conjecture that they are not only equivalent but equal. Here is a proof in the case that S is finite and each X_s is finite. The general case is likely to require technicalities along the lines of the Kantorovich-Rubinstein theorem (for simpler proofs than the original, see [Vi, Ed11]).

Recall basic results of linear programming, e.g. Ch.4 of [BHM]. Firstly, to the primal problem

$$\max_x \sum_j c_j x_j$$

over vectors x with components $x_j \geq 0$ for j in some subset J and $x_j \in \mathbb{R}$ ("unrestricted") for the other j, subject to constraints $\sum_j a_{ij} x_j \leq b_i$ for i in a subset I and $=$ for the other i, is associated the dual problem

$$\min_y \sum_i y_i b_i$$

over vectors y with components $y_i \geq 0$ for $i \in I$, unrestricted otherwise, subject to constraints $\sum_i y_i a_{ij} \geq c_j$ for $j \in J$, $=$ otherwise. Secondly, if either problem has a feasible point (i.e. satisfying the constraints) and the objective function is bounded (above for the primal problem, below for the dual), then the same applies to the other and they have the same optimising value.

For two probabilities μ, ν on finite X, the definition of $\bar{d}(\mu, \nu)$ can be reformulated as the minimising value for

$$\min_{\phi, m} \phi$$

over $\phi \in \mathbb{R}$, $m = (m_{x,y})_{x,y \in X} \geq 0$, subject to

$$\phi - \sum_{x,y} m_{x,y} d_s(x_s, y_s) \geq 0$$

for each $s \in S$, $\sum_y m_{x,y} = \mu_x$ for each $x \in X$, and $\sum_x m_{x,y} = \nu_y$ for each $y \in X$. There is a feasible point, e.g. $m_{x,y} = \mu_x \nu_y$, $\phi = \max_s \operatorname{diam}_s(X_s)$. The objective function is bounded below by 0. The dual problem is

$$\max_{f,g,e} \sum_x \mu_x f_x + \sum_y \nu_y g_y$$

over $f = (f_x)_{x\in X}$, $g = (g_y)_{y\in X}$ unrestricted and $e = (e_s)_{s\in S} \geq 0$, subject to $\sum_s e_s \leq 1$ and

$$f_x + g_y - \sum_s e_s d_s(x_s, y_s) \leq 0 \tag{11}$$

for all $x, y \in X$. For fixed e, the maximum is attained by $g = -f$, by the Kantorovich-Rubinstein theorem applied to cost function $\sum_s e_s d_s(x_s, y_s)$. But then the constraints (11) are equivalent to $f_x - f_y - e_s d_s(x_s, y_s) \leq 0$ for all x agreeing with y off a single site s. Thus the problem is equivalent to

$$\max_{f,e} \sum_{x\in X} (\mu_x - \nu_x) f_x$$

over $f = (f_x)_{x\in X}$ unrestricted and $e = (e_s)_{s\in S} \geq 0$, subject to $\sum_{s\in S} e_s \leq 1$ and

$$f_x - f_y - e_s d_s(x_s, y_s) \leq 0$$

for each configuration y agreeing with x off a single site s. The maximising value is equivalent to the definition (1) of $D(\mu, \nu)$, hence the result.

References

[AV] Anulova SV, Veretennikov A Yu, Exponential convergence of multi-dimensional stochastic mechanical systems with switching impacts, Dokl Math 91 (2015) 60–63.

[ABN] Asch M, Bocquet M, Nodet M, Data assimilation (SIAM, 2017)

[A+] Aström KJ et al (eds), Control of complex systems (Springer, 2001)

[BaM] Baesens C, MacKay RS, Exponential localization of linear response in networks with exponentially decaying coupling, Nonlinearity 10 (1997) 931–940

[BY] Bar-Yam Y, Dynamics of Complex Systems (Perseus, 1997)

[BBV] Barrat A, Barthélemy M, Vespignani A, Dynamical processes on complex networks (Cambridge, 2008)

[BK] Bogachev VI, Kolesnikov AV, The Monge-Kantorovich problem: achievements, connections and perspectives, Russ Math Surv 67:5 (2012) 785–890.

[BHM] Bradley SP, Has AC, Magnanti TL, Applied Mathematical Programming (Addison-Wesley, 1977)

[BuM] Bujorianu L-M, MacKay RS, Perturbation and Sensitivity of Inhomogeneous Markov Chains in Dynamic Environments, IEEE Proc 21st International Symposium on Mathematical Theory of Networks and Systems, University of Groningen, The Netherlands, 7-11 July 2014, pp 184–191

[CDK] Chornei RK, Daduma H, Knopov PS, Control of spatially structured random processes and random fields with applications (Springer, 2006)

[DM] Diakonova M, RS MacKay RS, Mathematical examples of space-time phases, Int J Bif Chaos 21 (2011) 22972304.

[D1] Dobruschin PL, Description of a stochastic field by conditional probabilities, and conditions for its regularity, Theory Probab Appl 13 (1968) 197–224

[D2] Dobrushin RL, Prescribing a System of Random Variables by Conditional Distributions, Theory Probab Appl 15 (1970) 458–486

[Ed11] Edwards DA, On the Kantorovich-Rubinstein theorem, Expos Math 29 (2011) 387–398.

[FMW] Flatley L, MacKay RS, Waterson M, Optimal strategies for operating energy storage in an arbitrage or smoothing market, J Dyn Games 3 (2016) 371–398.

[FH] Föllmer H, Horst U, Convergence of locally and globally interacting Markov chains, Stoch Proc Appln 96 (2001) 99–121.

[FW] Freidlin MI, Wentzell AD, Random perturbations of dynamical systems (Springer, 1984, 2012).

[Gr] Gross L, Decay of correlations in classical lattice models at high temperature, Commun Math Phys 68 (1979) 9–27.

[Ha] Hamm A, Large deviations from the thermodynamic limit in globally coupled maps, Physica D 142 (2000) 4169.

[IP] Ioannou PA, Pitsillides A (eds), Modeling and control of complex systems (CRC, 2008)

[J] Jiang M, Sinai-Ruelle-Bowen Measures for Lattice Dynamical Systems, J Stat Phys 111 (2003) 863–902.

[Ki] Kifer Y, A discrete-time version of the Wentzell-Freidlin theory, Ann Probab 4 (1990) 1676–92.

[Ku] Künsch H, Decay of correlations under Dobrushin's uniqueness condition and its applications, Commun Math Phys 84 (1982) 207–22.

[LB] Liu YY, Barabasi A-L, Control principles of complex systems, Rev Mod Phys 88 (2016) 035006

[M0] MacKay RS, Discrete Breathers: classical and quantum, Physica A 288 (2000) 174198.

[M1] MacKay RS, Nonlinearity in Complexity Science, Nonlinearity 21 (2008) T273–81.

[M2] MacKay RS, Robustness of Markov processes on large networks, J Difference Eqns & Applns 17 (2011) 1155–67.

[M3] MacKay RS, Space-time phases, in: Ball RC, Kolokoltsov VN, MacKay RS (eds), Complexity Science: The Warwick Masters Course, London Mathematical Society Lecture Notes vol 408 (Cambridge Univ Press, 2013)

[NBW] Newman M, Barabasi A-L, Watts DJ (eds), The structure and dynamics of networks (Princeton, 2006)

[NN] Nicolis G, Nicolis C, Foundations of complex systems (World Sci, 2007)

[Si] Siljak DD, Decentralized control of complex systems (Academic, 1991)

[SM+] San Miguel M, Johnson J, Kertesz J, Kaski K, Diaz-Guilera A, MacKay RS, Loreto V, Erdi P, Kondor I, Helbing D, Challenges in complex systems science, Euro Physical J Special Topics 214 (2012) 245–271

[SM] Slowinski P, MacKay RS, Phase diagrams of majority voter probabilistic cellular automata, J Stat Phys 159 (2015) 43–61.

[St] Steif JE, Convergence to equilibrium and space-time Bernoullicity for spin systems in the $M < \varepsilon$ case, Ergod Th Dyn Sys 11 (1991) 547–575.

[T] Toom AL, Stable and attractive trajectories in multicomponent systems, in: Multicomponent random systems, ends Dobrushin R, Sinai Ya, Adv Probab Related Topics 6 (1980) 549–576

[VJ] Vamvoudakis K, Jagannathan S (eds), Control of complex systems (Elsevier, 2016)

[Va] Vasershtein LN, Markov processes over denumerable products of spaces, describing large systems of automata, Probl Info Transm 5 (1969) 47–52

[Vi] Villani C, Optimal transport – old and new (Springer, 2009)

[W3] Wang A, Wang Y, Wang H, Shape control of conditional output probability density functions for linear stochastic systems with random parameters, Int J Sys Contr Comm 3 (2011) 82–94

[Y] Young L-S, What are SRB measures, and which systems have them?, www.cims.nyu.edu/∼lsy/papers/SRBsurvey.ps

Mathematics Institute and Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK

E-mail address: `R.S.MacKay@warwick.ac.uk`