NUMERICALLY OPTIMIZED MARKOVIAN COUPLING AND MIXING IN ONE-DIMENSIONAL MAPS

KALVIS M. JANSONS AND PAUL D. METCALFE

Abstract. Algorithms are introduced that produce optimal Markovian couplings for large finite-state-space discrete-time Markov chains with sparse transition matrices; these algorithms are applied to some toy models motivated by fluid-dynamical mixing problems at high Peclét number. An alternative definition of the time-scale of a mixing process is suggested. Finally, these algorithms are applied to the problem of coupling diffusion processes in an acute-angled triangle, and some of the simplifications that occur in continuum coupling problems are discussed.

§1. *Introduction*. The mixing by flow of a passive scalar in the limit of weak molecular diffusion is a well-established problem with much relevance to applications in the real world. Estimates of the time-scales of such processes are essential in geophysical fluid dynamics and in chemical engineering. A common approach to mixing processes is to study the Lagrangian dynamics of the system (see, for instance, the review by [**6**]), and in many cases this reduces to the problem of studying the separation of two infinitesimally close particles. These fluid-dynamical problems provide the motivation for this work. This motivation is taken a step closer to the fluid mechanical applications in [**7**], where we consider the optimal coupling of the Kolmogorov diffusion and some related optimal control problems.

Mixing problems also appear in probability theory, where they govern the convergence to stationarity of a stochastic process. By running two copies of the process at the same time, but imposing a dependence between the two copies to make them collide quickly, it is possible to obtain rigorous bounds on the convergence to stationarity of probability distributions (in the L_1 norm). This technique – the coupling method [**5**] – is often used to prove convergence to stationarity by proving that collision is guaranteed in finite time. However, we can improve this rigorous coupling bound by choosing the dependence between our two copies of the underlying stochastic process optimally.

When our underlying stochastic process models physical particle motion, we can view these two copies of the process as a two-particle system. The motion of each particle, viewed separately, is governed by the one-particle law, and we choose the remaining freedom in the two-particle problem so as to drive the two particles together quickly. After the particles collide, they do not contribute to the L_1 norm of the difference of the probability distributions of the two particles.

In this study, we are particularly interested in finding very efficient numerical methods to investigate such systems, which we hope will be useful in applications and improve our understanding of optimal couplings in general. We display some of our results in pictorial forms, which highlight the spatial

[MATHEMATIKA, **54** (2007), 161–181]

structures of the optimal coupling strategies. Where possible, we have tried to make the text understandable to readers from probability theory, numerical analysis and applied mathematics, but it is impossible to keep everyone completely happy all of the time. Our hope is that coupling methods will be used more widely and, in particular, will be increasingly used in applied mathematics to study mixing.

We work largely in the context of a finite-state-space discrete-time Markov chain X, with state-space D, and transition probabilities

$$P_{ip} = \mathbb{P}[X_{n+1} = p | X_n = i]. \tag{1}$$

Subject to technical conditions [**5**], satisfied in all problems discussed here, it is found that $P^N \to P_\infty$ as $N \to \infty$ and, further, that all the rows of the limit P_∞ are constant. This constant row is the stationary probability distribution of the Markov chain, and gives the probability of finding the system in a given state in the well-mixed limit. Our aim is to derive bounds on this convergence to stationarity. We study the long-time behaviour of such processes, and calculate asymptotic decay rates. Furthermore, we derive bounds that govern the whole history of convergence, as these are more likely to be useful in fluid-dynamical applications.

We consider that the main contributions of this study are (i) the efficient numerical algorithms themselves (see §3), which are easily generalized to solve many related problems, and (ii) the structure of the optimal coupling strategies for our toy problems of §4. We also comment in §5 on what our numerical experiments reveal about how good coupling bounds actually are. In §6, we apply our method to analyse diffusion in an acute-angled triangle with reflecting boundaries. Finally, in §7, we discuss the application of these ideas to continuum problems, in which several simplifications are possible (see, for example, [**7**]).

§2. *Optimal Markovian coupling*. The coupling inequality is a standard tool in probability theory [**5**], allowing the derivation of bounds on the convergence to stationarity of a Markov process. (For reference, a Markov process is a process for which the past and the future, given the present, are independent.) We rederive it here for the benefit of those who have not seen it before.

We construct a Markov process – a Markovian coupling – $(X_n, Y_n)_{n \geqslant 0}$, so that X and Y are individually Markov processes with identical transition laws, and require $X_n = Y_n$ for $n > T$, where $T = \inf\{n : X_n = Y_n\}$ (*i.e.*, the first time that X and Y collide). Then

$$\begin{aligned}\mathbb{P}[X_n \in A] - \mathbb{P}[Y_n \in A] &= (\mathbb{P}[X_n \in A, n < T] + \mathbb{P}[X_n \in A, n \geqslant T]) \\ &\quad - (\mathbb{P}[Y_n \in A, n < T] + \mathbb{P}[Y_n \in A, n \geqslant T]) \\ &= (\mathbb{P}[X_n \in A \,|\, n < T] - \mathbb{P}[Y_n \in A \,|\, n < T])\mathbb{P}[n < T]. \end{aligned} \tag{2}$$

Now

$$\begin{aligned}&\sum_i \big|\mathbb{P}[X_n = i] - \mathbb{P}[Y_n = i]\big| \\ &\qquad = \mathbb{P}[n < T] \sum_i \big|\mathbb{P}[X_n = i \,|\, n < T] - \mathbb{P}[Y_n = i \,|\, n < T]\big| \end{aligned} \tag{3}$$

and, since

$$\begin{aligned}\big|\mathbb{P}[X_n = i \,|\, n < T] - \mathbb{P}[Y_n = i \,|\, n < T]\big| \\ \leqslant \mathbb{P}[X_n = i \,|\, n < T] + \mathbb{P}[Y_n = i \,|\, n < T],\end{aligned} \tag{4}$$

we find that

$$\sum_i \big|\mathbb{P}[X_n = i] - \mathbb{P}[Y_n = i]\big| \leqslant 2\mathbb{P}[T > n]. \tag{5}$$

This inequality is often used in probability theory to bound the convergence to stationarity of a Markov chain. It is often relatively simple to find some coupling scheme for which the right side of (5) decays to zero as $n \to \infty$, and this then gives a bound on the ℓ_1 distance from stationarity of the probability distribution. In this work we consider optimal couplings, highlighting fundamental limitations of the coupling method.

For reference, the coupling inequality (5) becomes

$$\int \big|\mathbb{P}[X_t \in dx] - \mathbb{P}[Y_t \in dx]\big| \leqslant 2\mathbb{P}[T > t]. \tag{6}$$

for processes in continuous space and time.

There is no *a priori* reason to restrict ourselves to Markovian coupling processes, but non-Markovian coupling processes are less physically relevant, and we restrict ourselves to Markovian coupling processes throughout.

We still have a great deal of freedom in the choice of this Markovian coupling. In essence, we are free to impose any dependence we choose between the processes X and Y, provided that we preserve the laws of X and Y. We can therefore choose the transition law of the coupling process to minimize functionals such as the probability $\mathbb{P}[T > n]$, or the expected coupling time $\mathbb{E}[T]$. We will refer to consistent transition laws for the coupling process (X, Y) as coupling strategies; it is these objects over which we optimize.

Note that the coupling bound may not be achievable, and that, if estimates of the convergence of two distributions are needed, then one must also carry out an optimization over the distribution of (X_0, Y_0). This optimization over the starting law is a postprocessing step of no essential interest, and we will restrict ourselves to deterministic starting laws throughout, for which this optimization over the distribution of (X_0, Y_0) is trivial.

As the underlying system is symmetrical in X and Y, we can, without loss of generality, restrict our coupling process (X, Y) to $D^{(2)} \equiv D \times D/\sim$, where $\sim$ denotes the particle-particle symmetry, which all optimal fields have (where by the term 'field' we just mean a function of state or position). If the optimal strategy is not unique, then there exist optimal strategies that do not have the particle-particle symmetry (see **[2]**). However, from such strategies we can always construct an optimal strategy with the particle-particle symmetry and, given all optimal strategies in $D^{(2)}$, we can construct all optimal strategies in $D \times D$; there is no loss in generality in restricting ourselves to $D^{(2)}$. Note that all two-particle sums of field variables are taken over $D^{(2)}$.

2.1. *Operator version.* From the Chapman–Kolmogorov equations,

$$\mathbb{P}[X_n = j \,|\, X_0 = i] = (P^n)_{ij}, \tag{7}$$

so that

$$\mathbb{P}^i[X_n = k] - \mathbb{P}^j[Y_n = k] = (P^n)_{ik} - (P^n)_{jk}, \tag{8}$$

and hence

$$\sum_k \left|(P^n)_{ik} - (P^n)_{jk}\right| \leqslant 2\mathbb{P}^{ij}[T > n], \tag{9}$$

where superscripts on the probability law $\mathbb{P}[\cdot]$ and the expectation $\mathbb{E}[\cdot]$ always denote the initial condition.

The expected coupling time $\mathbb{E}[T]$ also gives a bound on the convergence of the system; summing (9) over n we find that

$$2\mathbb{E}^{ij}[T] \geqslant \sum_{n \geqslant 0} \sum_k \left|(P^n)_{ik} - (P^n)_{jk}\right|. \tag{10}$$

We will refer to the right side of (10) as an *operator sum*.

This equation shows how the expected coupling time $\mathbb{E}^{ij}[T]$ provides a bound on the time-scale of convergence that takes into account the whole history of convergence. Typically, $(-\log|\lambda|)^{-1}$, where λ is the leading non-trivial eigenvalue of P, is used to characterize the time-scale of convergence to stationarity. However, there may be a significant lag time before the response of the system is well approximated by an eigenmode with this decay rate. The expected coupling time is a more reliable bound on the time-scale of mixing that takes this lag time into account.

Although probability densities (when measured in the ℓ_1 norm) cannot exhibit the large transient growth that may be found in some linear stability problems (see [**12**] for much discussion, and for further references into a large literature), there are many probabilistic systems in which there are long waiting times before the asymptotic behaviour of the system may be observed [**4**].

Finally, the comparison between coupling methods and operator bounds in continuum problems works analogously, but such problems are not the focus of this paper, and are not discussed here. We do make brief mention of some continuum problems in §7.

§3. *Algorithm.* Recall that we aim to generate a Markov process (X, Y), such that $\mathbb{P}[X_{n+1} = p | X_n = i]$ and $\mathbb{P}[Y_{n+1} = p | Y_n = i]$ are fixed, and are both equal to P_{ip}, with the property that some functional of the process is minimal. The systems that we consider will have sparse transition matrices, and here we introduce algorithms for the computation of optimal couplings for Markov chains with sparse transition matrices. We seek a sequence of transition probabilities

$$Q^{(n)}_{ijpq} = \mathbb{P}[(X_{n+1}, Y_{n+1}) = (p), q) \,|\, (X_n, Y_n) = (i,j)]. \tag{11}$$

To obtain the transition probabilities for one particle, we sum over all possible destinations of the other particle, and hence the constraints on the one-particle dynamics require that

$$\sum_q Q^{(n)}_{ijpq} = P_{ip}, \qquad \sum_p Q^{(n)}_{ijpq} = P_{jq} \tag{12}$$

in $D \times D$, which are folded onto $D^{(2)}$. We also need $Q^{(n)}_{ijpq} \geqslant 0$.

The optimal expected coupling time $\tilde{\mathbb{E}}^{ij}[T]$ (we denote optimal fields with a tilde) satisfies a time-homogeneous problem, with only one coupling strategy. By conditioning on the first step of the process, we see that

$$\tilde{\mathbb{E}}^{ij}[T] = \inf_{Q}\left\{(1-\delta_{ij})\left(1+\sum_{pq} Q_{ijpq}\tilde{\mathbb{E}}^{pq}[T]\right)\right\}. \tag{13}$$

We can also study the convergence to stationarity in more detail by computing $\tilde{\mathbb{P}}^{ij}[T > n]$. This will give us a set of time-dependent transition probabilities for the coupled problem, and we note that

$$\tilde{\mathbb{P}}^{ij}[T > n] = \inf_{Q^{(n)}}\left\{(1-\delta_{ij})\sum_{pq} Q^{(n)}_{ijpq}\tilde{\mathbb{P}}^{pq}[T > n-1]\right\}. \tag{14}$$

The optimal tail-probability field $\tilde{\mathbb{P}}^{ij}[T > n]$ takes the asymptotic form

$$\tilde{\mathbb{P}}^{ij}[T > n] \sim \lambda^n \tilde{V}^{ij} \qquad \text{as } n \to \infty. \tag{15}$$

That is, we aim to pick out the slowest decaying mode in the system, and λ is the corresponding eigenvalue, and so does not depend on i and j.

We impose this *Ansatz* on the probability problem (14) to obtain a time-homogeneous tail problem

$$\tilde{V}^{ij} = \lambda^{-1}\inf_{Q}\left\{(1-\delta_{ij})\sum_{pq} Q_{ijpq}\tilde{V}^{pq}\right\}, \tag{16}$$

where

$$\lambda = \sum_{ijpq}(1-\delta_{ij})Q_{ijpq}\tilde{V}^{pq}. \tag{17}$$

This allows us to pass directly to the asymptotic tail without computing intermediate steps – and gives a great saving in both computational time and storage. These are, however, numerical observations, and we have not performed a rigorous analysis of this method.

3.1. *The structure of the numerical methods.* The structure of the numerical method for the expected-time problem, without improved convergence is as follows.

ALGORITHM 1. *The expected-time algorithm.*
1. *Initialize objective function.*
2. *Solve the linear programming to get the new objective function.*
3. *If converged, STOP.*
4. *Go to 2.*

However, this sort of successive relaxation method tends to be very slow. Thus, we normally add an improved convergence step to speed up the code massively, which leads to the follow structure.

ALGORITHM 2. *The accelerated expected-time algorithm.*
1. *Initialize objective function.*
2. *Solve the linear programming to obtain the new objective function.*

3. *If converged, STOP.*
4. *If it is time to accelerate*
   EITHER *solve the GMRES linear system (see [11])*
   OR *do RRE (see [10])*
5. *Go to 2.*

The two options for improved convergence both gave very similar results, but the results presented in this study were computed using GMRES.

Similarly, the tail-probability problem can be solved as follows.

ALGORITHM 3. *The tail-probability algorithm.*
1. *Initialize objective function.*
2. *Solve the linear programming to obtain the new objective function.*
3. *Normalize objective function to find* λ.
4. *If converged, STOP.*
5. *Go to 2.*

The improved-convergence version has the following structure.

ALGORITHM 4. *The accelerated tail-probability algorithm.*
1. *Initialize objective function.*
2. *Solve the linear programming to obtain the new objective function.*
3. *Normalize objective function to find* λ.
4. *If converged, STOP.*
5. *If it is time to accelerate*
   EITHER *solve the eigenvalue problem*
   OR *do RRE*
6. *Go to 2.*

In the following subsection, we consider these algorithms in more detail.

3.2. *Optimizing coupling probabilities.* We note that $\tilde{\mathbb{P}}^{ij}[T > 0] = 1 - \delta_{ij}$, and that, given $\tilde{\mathbb{P}}^{ij}[T > n-1]$, it is simple to calculate the strategy $Q^{(n)}_{ijpq}$ and thence $\tilde{\mathbb{P}}^{ij}[T > n]$. For each point (i,j) we must minimize

$$\sum_{pq} Q^{(n)}_{ijpq} \tilde{\mathbb{P}}^{pq}[T > n-1] \tag{18}$$

subject to the constraints (12). This is just a linear programming problem which can easily be solved by standard algorithms (see, for example, [3]). Note that, if the transition matrix P_{ip} is sparse, then most entries of $Q^{(n)}_{ijpq}$ are known to be zero. This reduces the size of the linear programming problems that must be considered and gives a considerable saving in computational time.

3.3. *Minimizing expected time* Minimizing the expected coupling time is a little more subtle. At first sight we are faced with a large – and costly – non-linear constrained optimization problem. However, on noting that

$$\tilde{\mathbb{E}}^{ij}[T \wedge n] = (1 - \delta_{ij})\left(1 + \sum_{pq} Q^{(n)}_{ijpq} \tilde{\mathbb{E}}^{pq}[T \wedge (n-1)]\right), \tag{19}$$

where $a \wedge b \equiv \min\{a, b\}$, an algorithm becomes clear. Once given $\tilde{\mathbb{E}}^{ij}[T \wedge (n-1)]$, it is easy to construct $\tilde{\mathbb{E}}^{ij}[T \wedge n]$ using linear programming; we then proceed inductively (see Algorithm 1). The limit as $n \to \infty$ of this sequence gives $\tilde{\mathbb{E}}^{ij}[T]$, and from this we can construct an optimal expected-time strategy Q_{ijpq} by linear programming. Even in the case of a finite-state-space Markov chain with a unique stationary density there may be no unique optimal strategy (note that the solution of a linear programming problem is not necessarily unique), although the optimal expected time field $\tilde{\mathbb{E}}^{ij}[T]$ is unique. We found that the numerical method, outlined in Algorithm 2, never produced multiple fixed points, despite varying the parameters of the convergence acceleration methods, and using bizarre initial conditions. Note that the linear programming problems at each iteration need not be solved consecutively; they may be solved concurrently on a parallel computer.

The sequence $\tilde{\mathbb{E}}^{ij}[T \wedge n]$ converges as $n \to \infty$, but this convergence is frequently very slow. The rate-limiting step is the ability of the basic iteration (19) to propagate information over large distances, and the convergence may be accelerated using a number of standard methods to enhance the propagation of information over large distances. The easiest acceleration method is to note that (19) is analogous to the Jacobi iterative scheme for solving linear systems. Replacing (19) by the analogous Gauss-Seidel scheme doubles the convergence rate, as might be expected. It is, however, possible to do much better (see Algorithm 2); the easiest effective method of convergence acceleration is occasionally to apply the Incomplete Reduced Rank Extrapolation (IRRE) method of [**10**] to the generated sequences, but the most convenient method that we have found is occasionally to solve the linear system

$$\mathbb{E}^{ij}[T] = (1 - \delta_{ij})\left(1 + \sum_{pq} Q^{(n)}_{ijpq} \mathbb{E}^{pq}[T]\right), \tag{20}$$

using a linear system solver; the preconditioned GMRES method of [**11**] was found to be very effective. An optimally tuned (and carefully coded, using the `DGELSD` routine from LAPACK) IRRE was found to be as effective as a more conventional linear system solver, and did not need the extra matrix multiplications required by GMRES. However, the optimal tuning parameters for IRRE may vary throughout the convergence of the system, and a poor choice of fixed tuning parameters could generate instability or limit cycles. See [**8**] for details of the IRRE method.

Both of these methods produce an expected-time field $\mathbb{E}^{ij}[T]$ that is closer to the limit, and which can be fed into (19) as an initial condition. The basic iteration is then able to find an improved strategy from this accelerated field, and adjusts the field with a small-scale relaxation. The field must be iterated through the basic iteration (19) until these transients have died away, after which time this process may be repeated to accelerate the convergence to the limit greatly. Unfortunately, this destroys the probabilistic interpretation of the iterants, but if $\tilde{\mathbb{E}}^{ij}[T]$ is all that is required, this loss is more than compensated by the rapid convergence to the limit produced by these methods. This convergence acceleration may be seen in Figure 1, and in particular a slight

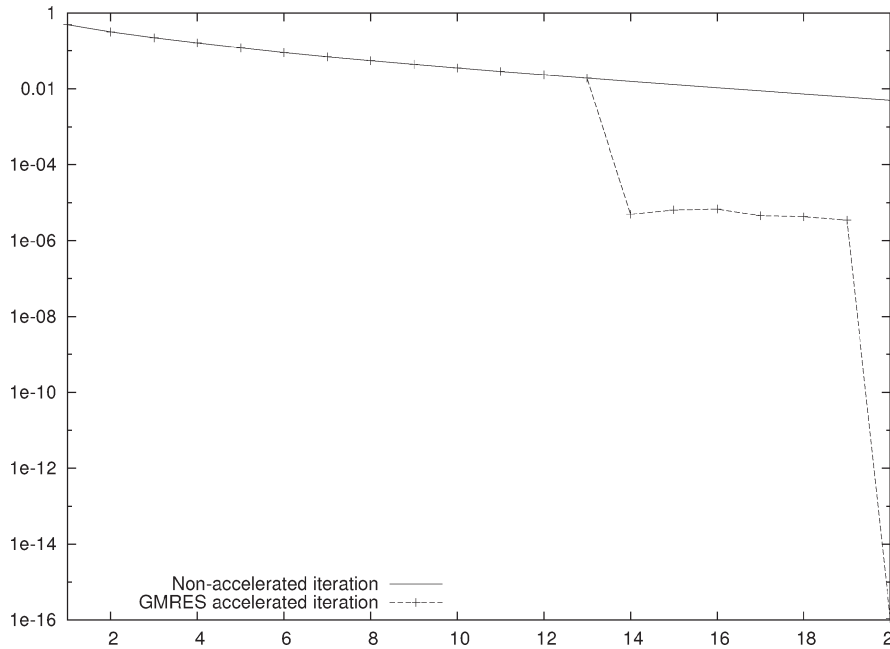


Figure 1: Expectation convergence histories for the integer-logistic problem with $N = 111$, $k = 2$, and $\mu = 3.9$. This is a typical convergence history – see §4 for the definition of the problem. The relative difference between successive iterants is plotted against the iteration number.

increase in the error may be seen after the convergence acceleration step, as the basic iteration locally adjusts the expected-time field.

3.4. *The tail-probability problem.* To find the solution of the tail-probability problem (see Algorithm 3), we start with $V^{ij} = 1 - \delta_{ij}$ and iterate

$$V^{ij} := \lambda^{-1} \inf_{Q} \left\{ (1 - \delta_{ij}) \sum_{pq} Q_{ijpq} V^{pq} \right\}, \tag{21}$$

where λ is chosen to set the normalization $\sum_{ij} V^{ij} = 1$.

The convergence of the iteration (21) can be greatly accelerated by using the occasional exact solution of the eigenvalue problem

$$\lambda V^{ij} = (1 - \delta_{ij}) \sum_{pq} Q_{ijpq} V^{pq} \tag{22}$$

to find the eigenvector corresponding to the eigenvalue with greatest real part. This eigenvector is then normalized to set $\sum_{ij} V^{ij} = 1$, and passed through the basic iteration (21) to allow transients to die away. After these transients have died away, this acceleration process can be repeated if necessary, and is summarized in Algorithm 4.

This eigenproblem (22) may be solved using a sparse solver such as ARPACK [**9**]. The efficacy of this convergence acceleration may be clearly

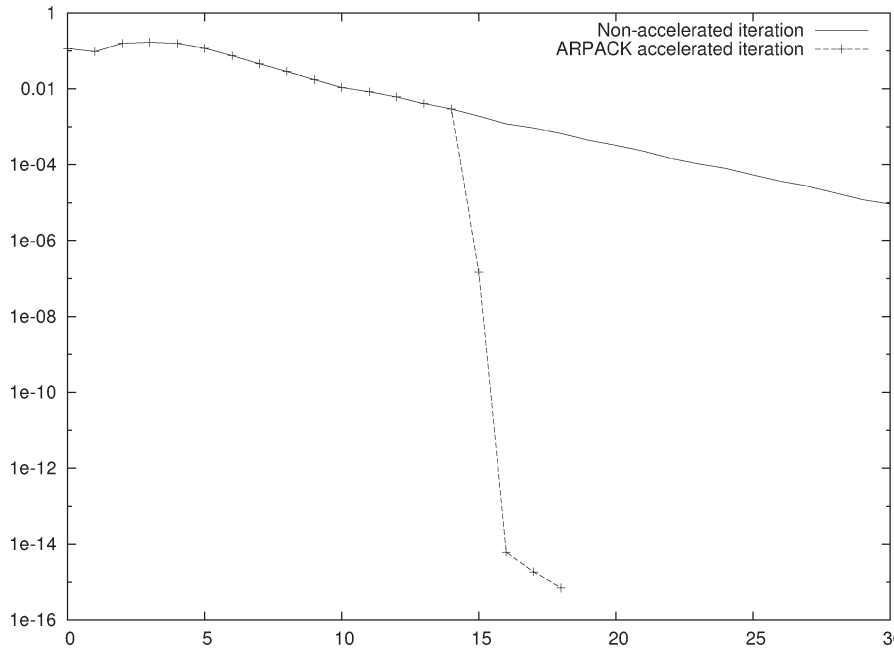


Figure 2: Tail-probability convergence histories for the integer-logistic problem with $N = 111$, $k = 2$, and $\mu = 3.9$. Again, this is a typical convergence history – see §4 for the definition of the problem. The relative difference between successive iterants is plotted against the iteration number.

seen in Figure 2. Again, optimal tail-probability fields $\tilde{V}^{ij}$ are unique, and we found that the numerical algorithm never produced multiple fixed points.

§4. *Integer-logistic and other maps.* We consider N-state systems, with the states labelled by i, where $0 \leqslant i < N$, and consider Markov chains of the form

$$X_{n+1} = R(F(X_n) + U_n), \tag{23}$$

where U_n is a random variable on $[-k, k]$, and R imposes boundary conditions, for example, reflection at 0 and $N - 1$. This may be thought of as a crude discretization of a continuum problem in which particles jump and then undergo diffusive motion. When $k \ll N$, our Markov chain loosely models a strongly mixing flow with some weak diffusion at molecular scales. This means that the transition matrix $P_{ip} = \mathbb{P}[X_1 = p \,|\, X_0 = i]$ is sparse, which leads to considerable numerical savings.

The detailed behaviour of such Markov chains depends on the numerical rounding of floating-point numbers to integers; the qualitative behaviour of these systems may be largely unchanged by different roundings, but quantitative results may vary greatly between different computers and different implementations of rounding. The precise details of the implementation of "reflecting boundary conditions" may also affect the results, and in systems with small k and N may prevent the existence of a unique stationary distribution. In sufficiently large problems, the details of the boundary condition are less important and, in general, affect only quantitative results.

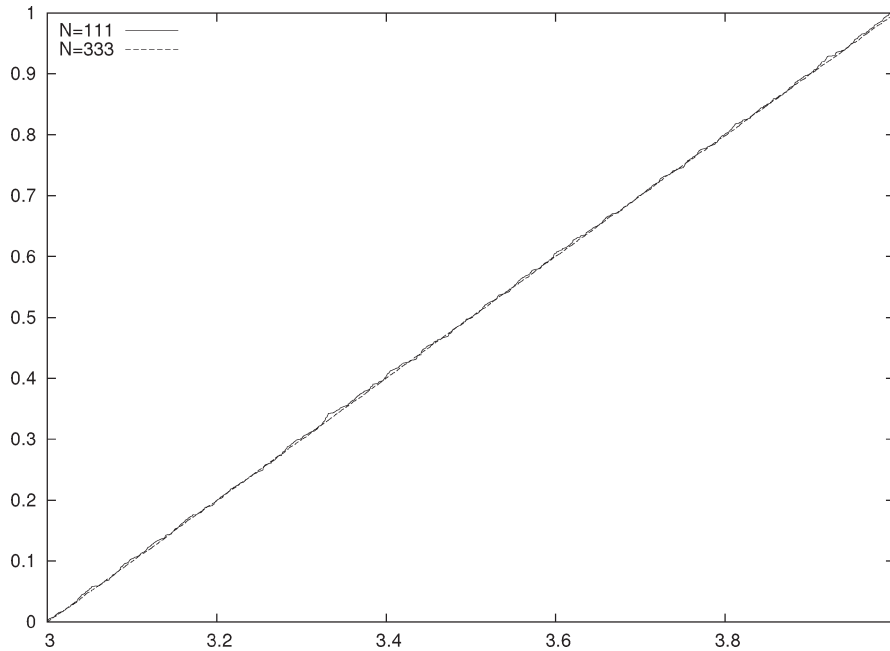


Figure 3: The proportion of distinct transition matrices in $[3, \mu]$ for the integer-logistic map is plotted against μ; plots are shown for $N = 111$ (973 transition matrices in total) and $N = 333$ (8720 transition matrices in total).

In this section, we work largely with integer versions of the logistic map, given by

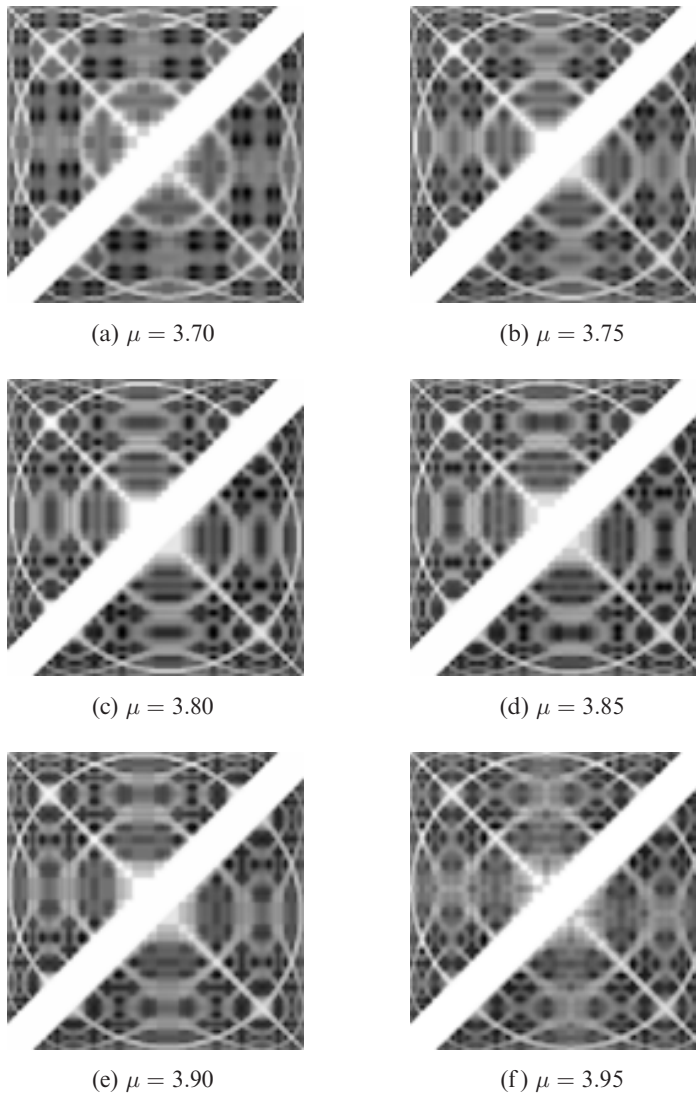
$$F(i) = \left\lfloor \frac{\mu}{N-1} i(N-1-i) + \frac{1}{2} - \varepsilon \right\rfloor, \tag{24}$$

where $i \in [0, N-1]$, $\lfloor x \rfloor$ is the integer part of x, and ε is a small floating-point number that is used to ensure that the map F takes the interval into itself. We will work with μ in the range $[3.7, 4]$, which we found to give irreducible transition matrices; when μ was smaller, this was not so. We impose reflecting boundary conditions on the system by setting

$$R(i) = \begin{cases} i, & i \in [0, N-1], \\ R(-i), & i < 0, \\ R(2(N-1)-i)), & i > N-1. \end{cases} \tag{25}$$

Finally, we use uniform random variables U_n in the "diffusive" step in (23). This is not necessary for our optimization scheme, but it does not reduce the connectivity of the underlying network or the qualitative results, and it does significantly reduce the number of parameters in the system.

Owing to the integer nature of the map (24), we obtain the same transition matrix over a small interval in μ. The values of μ at which the transition matrix changes are approximately uniformly distributed, as may be seen in Figure 3. We see that this class of models provides a range of problems that may be



(a) $\mu = 3.70$ (b) $\mu = 3.75$

(c) $\mu = 3.80$ (d) $\mu = 3.85$

(e) $\mu = 3.90$ (f) $\mu = 3.95$

Figure 4: Expectation fields $\mathbb{E}^{ij}[T]$ and tail-probability fields $\tilde{V}^{ij}$ for integer-logistic maps with $N = 111$ and $k = 2$ (recall that k is the maximal 'diffusive' step). In each of the subfigures, the expectation field is in the bottom right corner, the tail-probability field is in the top left corner. Light colours correspond to small values, dark to large values.

expected to have some degree of similarity for neighbouring values of μ. Usually, functionals of transition matrices with close μ values are similar, although large jumps do exist.

A range of expectation and tail-probability fields may be seen in Figure 4, which shows the complexity of the fields that result from this optimization. The expectation and tail-probability fields are clearly similar – correlations between the two fields are typically around 95% – but are not identical. Similarly, the optimization selects essentially different strategies for the two

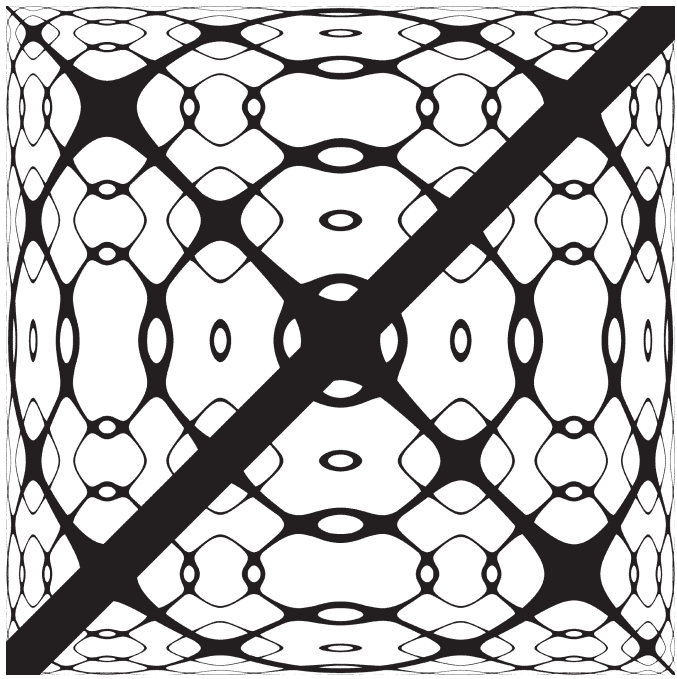


Figure 5: $\mathrm{Pi}_{\varepsilon}(\cdot)$ chart for the logistic map with $\mu = 3.9$. Points (x, y) in $\mathrm{Pi}_{0.05}$ (5) are plotted in black.

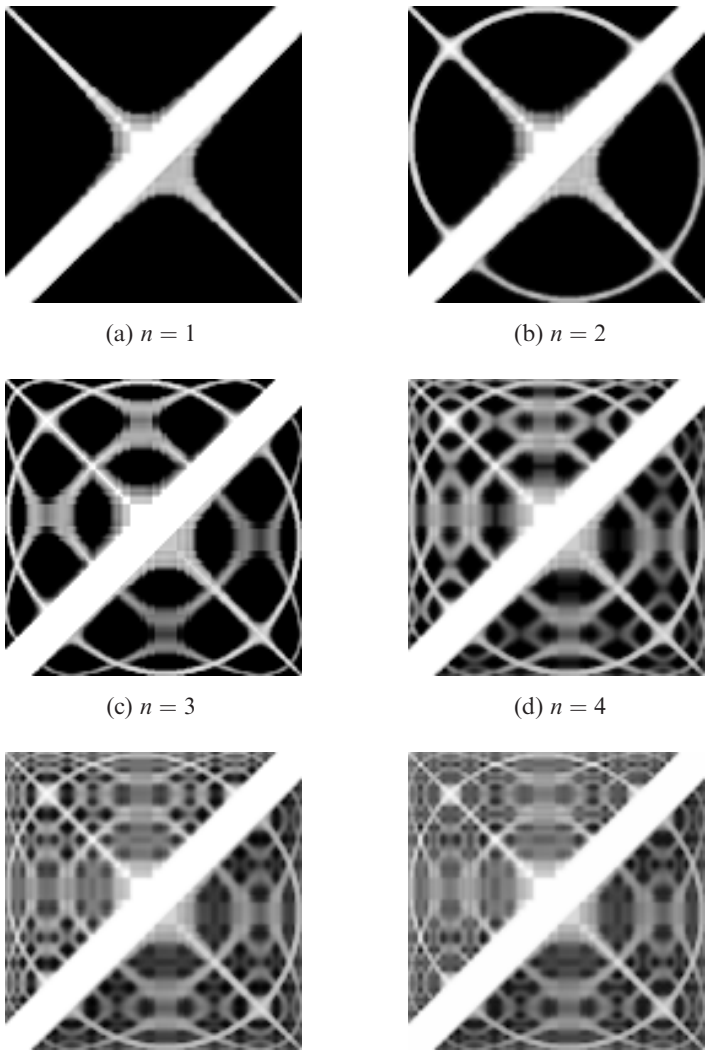
different problems. Usually, the optimal tail-probability strategy generates a suboptimal expectation field, and the optimal expectation strategy generates a suboptimal tail-probability field with a slower long-time decay rate. The only counterexamples that we have to this are in very small systems, in which the strategies are heavily constrained by (12).

It is possible to understand much of the behaviour shown in Figure 4 by plotting the pre-images of exit points (*i.e.*, the points in $D \times D$ where the two particles can collide) in $D \times D$. We define

$$\mathrm{Pi}_{\varepsilon}(n) \equiv \Big\{(x, y) \in D \times D : \min_{0 \leqslant i \leqslant n} |f^i(x) - f^i(y)| < \varepsilon\Big\}, \tag{26}$$

for a continuum map f. A plot of $\mathrm{Pi}_{0.05}$ (5) for the continuum logistic map $f(x) = \mu x(1 - x)$ is shown in Figure 5. It may be clearly seen that the pre-images of exit points in $D \times D$ provide the framework around which the optimal expected-time and tail-probability fields are formed.

As may be seen in Figure 6, when n is small the fields $\tilde{\mathbb{E}}^{ij}[T \wedge n]$ and $\tilde{\mathbb{P}}^{ij}[T > n]$ are non-trivial. Only regions of $D \times D$ near low-order pre-images of the exit line have any possibility of coupling in a few steps. These are the only regions of $D \times D$ in which the coupling can alter the fields $\tilde{\mathbb{E}}^{ij}[T \wedge n]$ and $\tilde{\mathbb{P}}^{ij}[T > n]$ at small n. However, as n increases, the pre-images of the exit line form a fine web over $D \times D$; all points of $D \times D$ have a route to the exit line, and the coupling can exert control everywhere.



(a) $n = 1$ (b) $n = 2$

(c) $n = 3$ (d) $n = 4$

(e) $n = 5$ (f) $n = 6$

Figure 6: Buildup of expectation fields $\mathbb{E}^{ij}[T \wedge n]$ and probability fields $\mathbb{P}^{ij}[T > n]$ for the integer-logistic map with $N = 111$, $k = 2$ and $\mu = 3.9$. In each of the subfigures, the expectation field is in the bottom right corner, the probability field is in the top left corner.

Figure 7 shows that this pre-image structure persists over all scales until the "diffusive" cutoff. At points of conflict between different orders of pre-image there are marked differences between couplings that optimize the expectation field and couplings that optimize the probability field. Subtle trade-offs occur between potentially fast but risky strategies and slow but safer strategies, and these trade-offs mean that it is hard to construct optimum strategies analytically for all but small problems.

In Figure 8, we see the decay of the distribution difference and the coupling bound $\tilde{\mathbb{P}}^{ij}[T > n]$ in a typical case. Both exhibit a cutoff phenomenon,

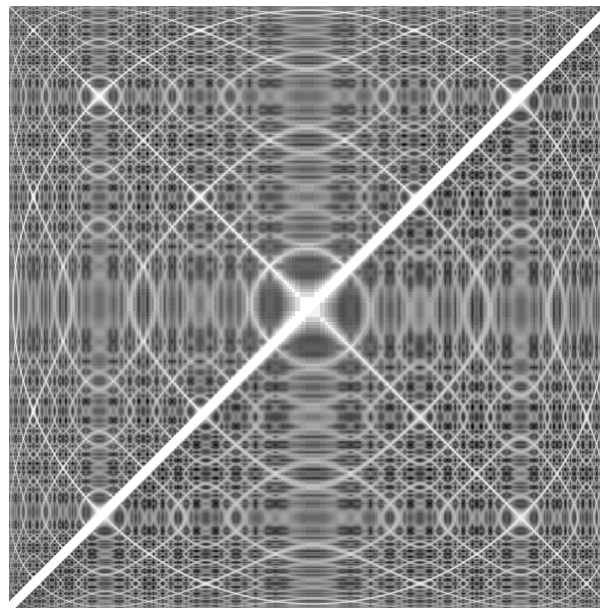


Figure 7: Expectation and tail-probability fields for the integer-logistic map, with $\mu = 3.90$, $N = 1000$, and $k = 3$. The expectation field is in the bottom right corner; the tail-probability field is in the top left corner. The correlation between the expectation field and the tail-probability field is 0.987.

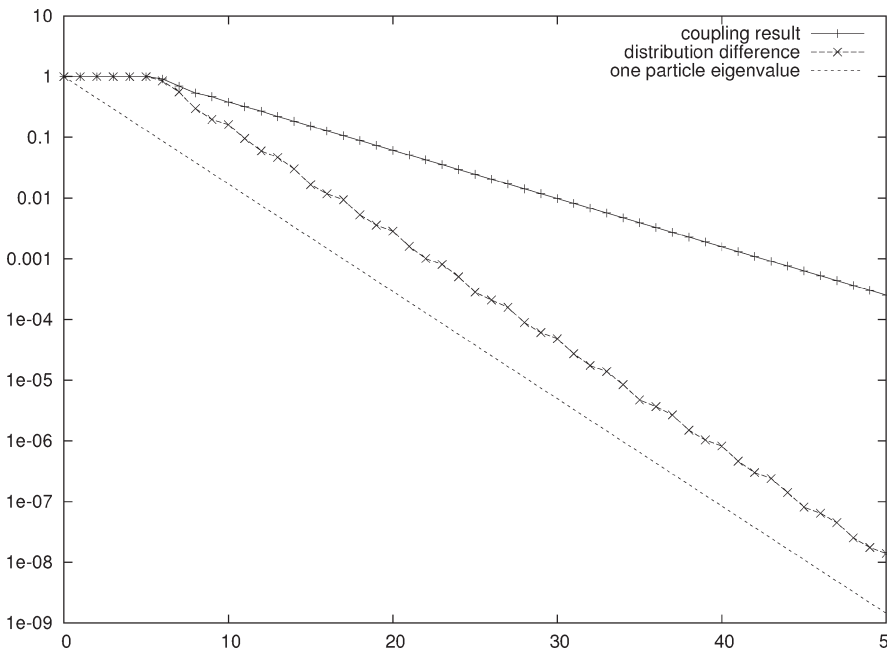
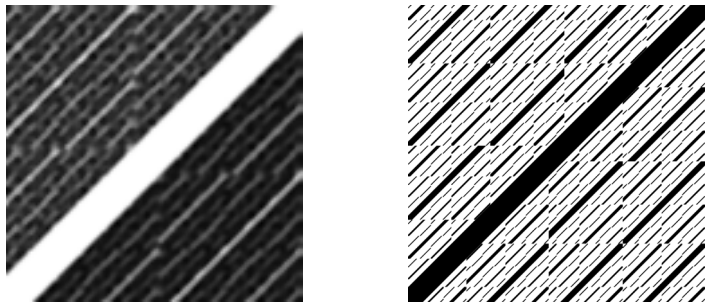


Figure 8: A plot of $\sup_{ij} \tilde{\mathbb{P}}^{ij}[T > n]$ against n for the integer-logistic map with $N = 111$, $k = 2$, and $\mu = 3.9$. Also shown is $\sup_{ij} \sum_k |\mathbb{P}^i(X_n = k) - \mathbb{P}^j(Y_n = k)|/2$, and a line decaying with the modulus of the leading eigenvalue of the one-particle transition matrix.



(a) fields (b) $\mathrm{Pi}_{0.05}$ (5)

Figure 9: Expectation and tail-probability fields for the periodic times-μ map (27) with $\mu = 1.9$, $N = 111$ and $k = 2$. The expectation field is in the bottom right corner; the tail-probability field is in the top left corner. The correlation between the two fields is 0.969. The sets $\mathrm{Pi}_{0.05}$ (5) are also shown.

in which there is a delay during which no decay occurs, before switching to their long-time asymptotic decay. We also see that the decay rate of the coupling bound is significantly less than the decay rate of the distribution difference.

We also investigated continuous logistic maps using genetic algorithm optimization. As with all Monte Carlo methods, convergence was poor and, as far as could be seen, the behaviour of the continuous problems is qualitatively similar to that of these discrete problems for large N. In particular, the sets $\mathrm{Pi}_{\varepsilon}(\cdot)$ could be clearly seen, and subtle trade-offs were also necessary at points of conflict between different generations of the pre-image sets.

4.1. *Other maps.* We considered a large number of different maps; all behaved analogously to the integer-logistic map, except that the fields $\tilde{\mathbb{E}}^{ij}[T]$, $\tilde{\mathbb{P}}^{ij}[T > n]$ and $\tilde{V}^{ij}$ were based on the relevant pre-image sets. A plot of the expectation and tail-probability fields for the map $f(x) = \lfloor \mu x \rfloor$, with periodic boundary conditions, is shown in Figure 9. The integer version of this map is

$$F(i) = \mu\, i \qquad \text{for } i \in [0, N-1], \tag{27}$$

and we use $R(i) = (i \,|\, N)$. Here, the pre-image sets are approximately parallel lines, and are clearly visible in the figure.

§5. *How good is coupling?* In these mixing problems, coupling schemes are rarely able to reproduce either the one-particle decay rate, or the operator sum bounded by the expected coupling time, although we find that the tail-probability decay rate is rarely less than a third of the one-particle decay rate, and the expected coupling time is rarely more than twice the operator sum. As may be seen in Figures 10 and 11, in these mixing problems, the expected coupling time is more sensitive than the tail-probability decay rate to changes in the underlying one-particle problem.

The coupling bound (5) deviates from equality only if there exists a point i for which both $\mathbb{P}[X_n = i \,|\, n < T]$ and $\mathbb{P}[Y_n = i \,|\, n < T]$ are non-zero. It is clear that preventing the overlap of the support of these two sequences of distributions is very difficult with just a single strategy – such as in the expectation or

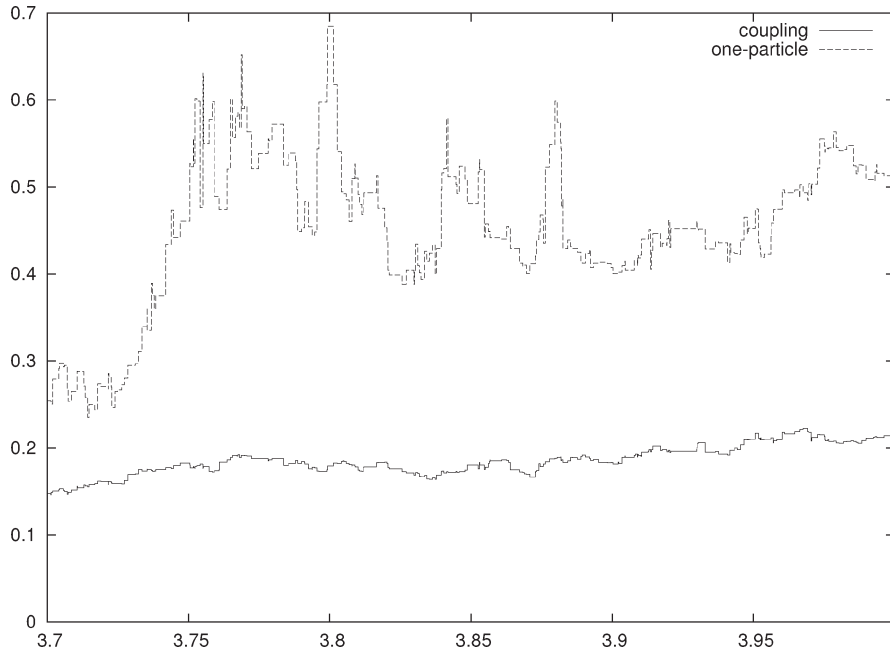


Figure 10: Long-time decay rates for integer-logistic maps with $N = 111$ and $k = 2$, for *all* μ between 3.7 and 4. We plot $-\log|\lambda|$ for both the tail-probability decay and the leading non-trivial eigenvalue of the one-particle transition matrix P.

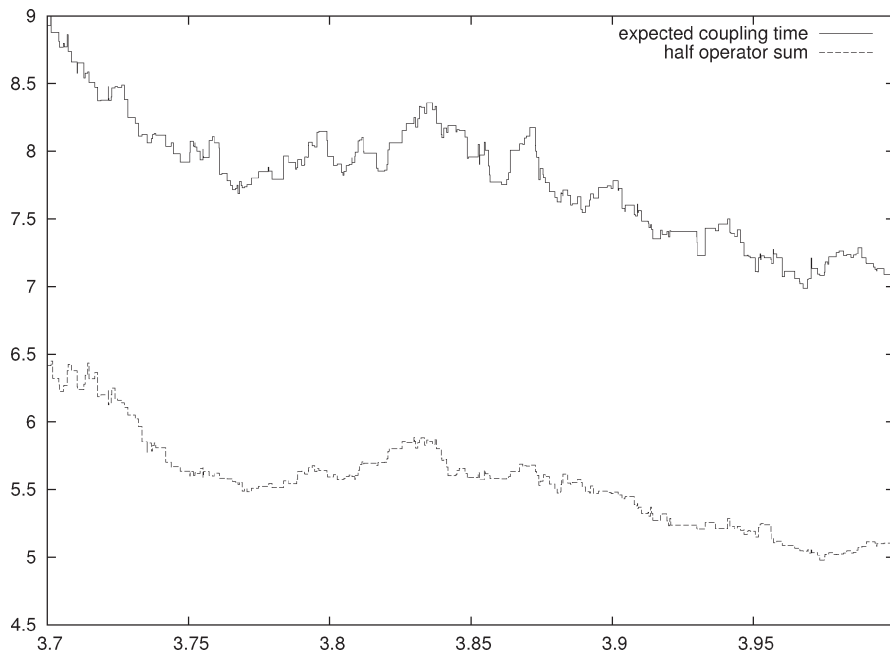


Figure 11: Comparison of $\rho\sum_{ij}\sum_{kn}|(P^n)_{ik} - (P^n)_{jk}|/2$ with $\rho\sum_{ij}\mathbb{E}^{ij}[T]$, where $\rho \equiv (N(N+1)/2)^{-1}$, for integer-logistic maps with $N = 111$ and $k = 2$, for *all* μ between 3.7 and 4.

tail-probability problems – but even in the full probability problem, for which the strategy is time-dependent, allowing much more freedom, it is a difficult requirement, and rarely achieved except in small problems. The bound was found to be exact only for some small Markov chains, with fewer than 15 states and little complexity.

Coupling methods are more able to match the one-particle decay rate (a property that [**2**] call "efficiency") – *i.e.*, the leading non-trivial eigenvalue of the one-particle transition matrix – than to match the full operator sum, but the ability to match the one-particle decay rate rarely survives much longer than the ability to match the full operator sum as the parameters of the system are varied.

It is possible for problems with the same one-particle decay rate to have different two-particle decay rates. A variant of our times-μ map (27), with reflecting rather than periodic boundary conditions, with $N = 111$ and $k = 2$, exhibits this curious behaviour when $\mu \in \{1.7, 1.8, 1.9\}$. Network problems exist with tunable parameters, in which the error in the decay rate may be made arbitrarily small.

Although a coupling strategy may not reproduce the decay rate of the one-particle problem, most of the spectrum of the one-particle transition matrix P_{ip} does appear in the two-particle problem. If the one-particle Markov chain is irreducible and aperiodic, its transition matrix P_{ip} has a complete set of right eigenvectors $(x^{(k)})_{1 \leqslant k \leqslant N}$ with corresponding eigenvalues (λ_k) [**5**]. Under these conditions, only one of the eigenvalues, λ_1, is equal to 1 in modulus, and all of the other eigenvalues have modulus less than 1. Note that the first eigenvector, $x^{(1)}$, is constant.

Using our constraints on the two-particle dynamics (12), we can show that $\lambda_2, \lambda_3, \ldots, \lambda_N$ appear in the spectrum of the tail decay of any coupling strategy. Consider the tail-probability matrix

$$\tilde{Q}_{ijpq} \equiv (1 - \delta_{ij}) Q_{ijpq}, \tag{28}$$

where Q_{ijpq} is a coupling strategy. Since $x_q^{(1)}$ is independent of q, the coupling constraints (12) imply that

$$\sum_{pq} \tilde{Q}_{ijpq} (x_p^{(1)} x_q^{(k)} - x_p^{(k)} x_q^{(1)}) = \lambda_k (x_p^{(1)} x_q^{(k)} - x_p^{(k)} x_q^{(1)}). \tag{29}$$

For $k > 1$, we have produced an eigenvector of the tail-probability matrix with eigenvalue λ_k.

We also note that the largest eigenvalue of the tail-probability matrix is real and positive, as $\mathbb{P}^{ij}[T > n] \sim \lambda^n V^{ij}$ for large n. If the one-particle problem has a negative leading eigenvalue, or a complex-conjugate pair, then the coupling method must create a leading eigenvalue of the tail-probability matrix that *is* real and positive, but it is only occasionally able to create a leading eigenvalue with the correct modulus.

Finally, if we suppose that the coupling is trivial, namely, that the Markov chains X and Y are independent, we may generate many more eigenvalues of the tail-probability matrix, since then the field $x_p^{(k)} x_q^{(l)} - x_p^{(l)} x_q^{(k)}$, for $p > q$, is an eigenvector of the tail-probability matrix with eigenvalue $\lambda_k \lambda_l$. This construction generates $N(N-1)/2$ of the eigenvalues of the $(N(N+1)/2) \times (N(N+1)/2)$

tail-probability matrix; the remaining N eigenvalues are not simply fixed by the one-particle problems.

§6. *Coupling diffusion processes in an acute-angled triangle with reflecting boundaries.* To be effective, our algorithms of §3 require only sparsity of the underlying one-particle transition matrix, and may be applied to a variety of problems which satisfy this condition. The coupling of two-dimensional Brownian motion in various domains is such a problem, and is actively studied in probability theory [**1**, **2**]. In particular, [**2**] study the coupling of two-dimensional Brownian motions in triangular domains. They used the particular (suboptimal) strategy of mirror coupling,

$$dX_t = dB_t + dL_t, \qquad dY_t = -dB_t + dM_t, \tag{30}$$

where B is a standard Brownian motion, and L and M are local-time processes that enforce reflecting boundary conditions. They proved that mirror coupling obtains the correct one-particle decay rate in obtuse-angled triangles, but is unable to obtain the one-particle decay rate in triangles with distinct acute angles.

A natural question to ask is whether an optimal coupling is able to obtain the correct one-particle decay rate in triangles in which mirror coupling cannot. We are unable to study directly the continuum problem using the algorithms of §3, but we are able to consider a sequence of triangular lattice approximations where the Brownian motion is replaced by a random walk. At each interior point of the lattice, a particle can move to any of the 6 nearest neighbours, or remain in its current position. We select the probabilities of these transitions to reproduce the first two moments of standard Brownian motion. The natural modifications of this scheme are made at boundary points to approximate reflection. For simplicity, our triangular lattices had equal numbers of points on each side, which limited these models to near-equilateral triangles.

In lattice models of an equilateral triangle T_0, we found that optimal coupling strategies were able to reproduce the one-particle decay rates in the continuum limit, even in the presence of simple flow fields such as straining motion, source flow and shear flow. We found no flow fields that prevented the coupling method from obtaining the one-particle decay rate in the continuum limit. However, the coupling method was unable to obtain the correct one-particle decay rate if the diffusivity was inhomogeneous but isotropic.

We also studied a slightly sheared equilateral triangle, with a 55° angle. The relative rate error for these triangles is plotted in Figure 12 as a function of M, the number of lattice points on each side of the triangle. Note that the number of states in the underlying one-particle Markov chain is $M(M+1)/2$. For the largest triangles we considered, we had 1540 states in the one-particle Markov chain and had 1186570 states in the two-particle problem.

We see that, in this sequence of models, as M increases the relative error in the decay rate seems to tend to a non-zero constant. This suggests that there may be no coupling strategies that obtain the one-particle decay rate in this 55° triangle T_1. This begs the question of whether optimized coupling can ever obtain the one-particle decay rate in a triangle where mirror coupling fails to do so, and remains an open question.

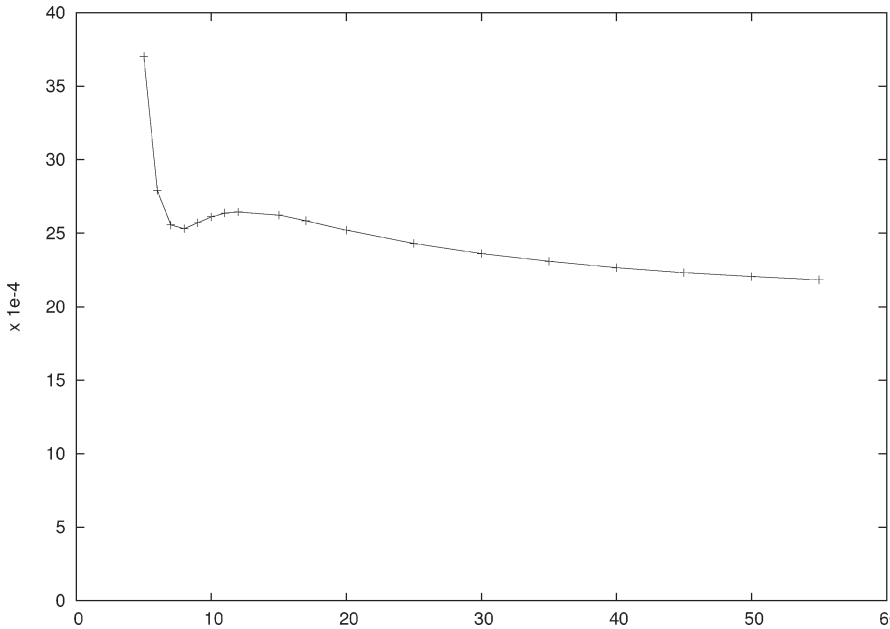


Figure 12: Relative rate error for triangle T_1, plotted against M.

§7. *Continuum problems.* Our discretization of the triangle in the previous section was constrained by computer resources. We were unable to increase the number of points beyond $N = 1540$, and were unable to obtain clear convergence of the relative error in the decay rate. It is preferable not to discretize continuum problems as a finite-state-space discrete-time Markov chain, but rather to discretize directly the underlying partial differential equations. Some simplifications occur here which remove the need for a numerical optimization step.

We consider in detail the optimal coupling of the Kolmogorov diffusion, and some related optimal control problems in [7], but we briefly consider some of the corresponding continuum problems here.

We consider Itô stochastic differential equations

$$\begin{aligned} dX_t &= u(X_t)\,dt + \sigma(X_t)\cdot dB_t + dL_t, \\ dY_t &= u(Y_t)\,dt + \sigma(Y_t)\cdot dW_t + dM_t, \end{aligned} \tag{31}$$

where B and W are standard Brownian motions, and L and M are local-time processes that enforce boundary conditions. In this continuum problem, we control the system by the choice of a field of orthogonal matrices $c(t,x,y)$, so that

$$dW_t = c(t,x,y)\cdot dB_t. \tag{32}$$

For $\phi(t,x,y) \equiv \mathbb{E}^{xy}[T \wedge t]$, we obtain the PDE

$$\begin{aligned} \frac{\partial\phi}{\partial t} &= 1 + u(x)\cdot\nabla_x\phi + u(y)\cdot\nabla_y\phi + \tfrac{1}{2}\big(\big(\sigma(x)\cdot\sigma^T(x)\big) : \nabla_x\nabla_x\phi\big) \\ &\quad + \tfrac{1}{2}\big(\big(\sigma(y)\cdot\sigma^T(y)\big) : \nabla_y\nabla_y\phi\big) + \big(\sigma(x)\cdot c(t,x,y)\cdot\sigma^T(y)\big) : \nabla_y\nabla_x\phi; \end{aligned} \tag{33}$$

we have used dyadic notation in which "$\cdot$" denotes contraction on nearest free indices. We have the initial condition $\phi(0, x, y) = 0$, and the boundary condition $\phi(t, x, x) = 0$ (which comes from successful coupling when the two particles collide), and other boundary conditions that come from the underlying single-particle problem.

At each time, and at each spatial position, we must choose an orthogonal matrix $c(t, x, y)$ that minimizes

$$c : (\sigma^T(y) \cdot (\nabla_x \nabla_y \phi) \cdot \sigma(x)), \tag{34}$$

the counterpart to (19) for continuous systems. This can most easily be done using the singular value decomposition

$$\sigma^T(y) \cdot (\nabla_x \nabla_y \phi) \cdot \sigma(x) = u \cdot s \cdot v^T, \tag{35}$$

in which u and v are orthogonal matrices, and s is a diagonal matrix with positive elements. The minimizing c is then $-vu^T$, and the value of (35) at this optimum is then $-\sum_i s_{ii}$. For these continuum problems, the optimization step can therefore be done in closed form, and we obtain a non-linear PDE that must be solved for $\phi(t, x, y)$. We note that we recover mirror coupling if the diffusivity is homogeneous and isotropic, if we can neglect the flows $u(x)$ and $u(y)$, and if $\phi(t, x, y) = \psi(t, \|x - y\|_2)$.

With this optimization step done, the steady-state version of (33) is only weakly elliptic, and its solutions are not as smooth as solutions to the Poisson equation. This may be clearly seen in the expected coupling time for standard Brownian motion in the unit interval with reflecting boundary conditions at the endpoints. We find that

$$\mathbb{E}^{xy}[T] = \begin{cases} -\frac{1}{4}\eta^2 + \frac{1}{2}\eta\xi(1 - \log\xi), & \text{if } \xi \leqslant 1, \\ -\frac{1}{4}\eta^2 + \frac{1}{2}\eta(2 - \xi)(1 - \log(2 - \xi)), & \text{if } \xi > 1, \end{cases} \tag{36}$$

where $\eta = x - y$ and $\xi = x + y$, and we represent $D^{(2)}$ with $y \leqslant x$. This is a steady solution of (33) with the optimization step built in, and the strategy used is mirror coupling. Note that the third derivative of the solution is discontinuous across the line $\xi = 1$.

§8. *Conclusion.* Optimal coupling schemes can be calculated for a variety of Markov chains with sparse transition matrices. We see that the bounds on convergence derived from these coupling schemes are rarely tight for problems with no obvious simplifying features, but in problems of physical interest these bounds are rarely worse than a factor of 2 from the one-particle values. (One particularly simple problem is that of advection-diffusion on the interval, with reflecting boundary conditions. Under mild regularity conditions, all coupling strategies are able to recover the one-particle decay rate, a result that is a simple application of corollary 2.7(ii) of [**2**].)

Continuum problems have some simplifying features that reduce the derivation of optimal coupling strategies to the solution of a non-linear parabolic equation, although the shapes of the domain $D^{(2)}$ on which these equations hold are typically numerically challenging. Two and three dimensional fluid

problems turn into four and six dimensional coupling problems. These will be studied elsewhere.

References

1. K. Burdzy and Z.-Q. Chen, Coalescence of synchronous couplings. *Probab. Theory Related Fields* **123** (2002), 553–578.
2. K. Burdzy and W. S. Kendall, Efficient Markovian couplings: examples and counterexamples. *Ann. Appl. Probab.* **10** (2000), 362–409.
3. V. Chvatal, *Linear Programming*. Freeman (New York, 1983).
4. P. Diaconis, The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.* **93** (1996), 1659–1664.
5. R. Durrett, *Probability: Theory and Examples* (2nd edition). Duxbury Press (Belmont, CA, 1996).
6. G. Falkovich, K. Gawędzki and M. Vergassola, Particles and fields in fluid turbulence. *Rev. Modern Phys.* **73** (2001), 913–975.
7. K. M. Jansons and P. D. Metcalfe, Optimally coupling the Kolmogorov diffusion, and related optimal control problems. *LMS J. Comput. Math.* **10** (2007), 1–20.
8. K. Jbilou and H. Sadok, Analysis of some vector extrapolation methods for solving systems of linear equations. *Numer. Math.* **70** (1995), 73–89.
9. R. B. Lehoucq, D. C. Sorensen and C. Yang, Solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods. *ARPACK Users' Guide: Software, Environments and Tools*, SIAM (Philadelphia, PA, 1998).
10. M. Mešina, Convergence acceleration for the iterative solution of the equations $X = AX + f$. *Comput. Methods Appl. Mech. Engrg.* **10** (1977), 165–173.
11. Y. Saad and M. H. Schultz, GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.* **7** (1986), 856–869.
12. L. N. Trefethen, Pseudospectra of linear operators. *SIAM Rev.* **39** (1997), 383–406.

Kalvis M. Jansons,
Department of Mathematics,
University College London,
Gower Street,
London WC1E 6BT.
E-mail: coupling@kalvis.com

Paul D. Metcalfe,
Cyprotex Discovery Ltd.,
15 Beech Lane,
Macclesfield SK10 2DR.

MSC (2000): *Primary*, 60J10; *Secondary*, 60J22, 93E20.

Received on the 28rd of February, 2007.