MEMORY LOSS FOR TIME-DEPENDENT DYNAMICAL SYSTEMS

WILLIAM OTT, MIKKO STENLUND, AND LAI-SANG YOUNG

ABSTRACT. This paper discusses the evolution of probability distributions for certain time-dependent dynamical systems. Exponential loss of memory is proved for expanding maps and for one-dimensional piecewise expanding maps with slowly varying parameters.

1. Introduction

This paper is about statistical properties of nonautonomous dynamical systems, such as flows defined by time-dependent vector fields or their discrete-time counterparts described by compositions of the form $f_n \circ \cdots \circ f_2 \circ f_1$ where all the $f_i : X \to X$ are self-maps of a space X. The topic to be discussed is the degree to which such a system retains its memory of the past as it evolves with time.

Memory is lost when the initial state of a system is quickly forgotten. Conceptually, this can happen in two very different ways. The first is for trajectories to merge, so that in time, they evolve effectively as a single trajectory independent of their points of origin. This happens in systems that are contractive. Consider for example a system defined by the composition of a sequence of maps f_i of a compact metric space X to itself, and assume that all the f_i have a uniform Lipschitz constant $L < 1$, *i.e.*, for all $x, y \in X$, $d(f_i x, f_i y) \leqslant L d(x, y)$. Since the diameter of the image of X decreases exponentially with time, all trajectories eventually coalesce into an exponentially small blob, which in general continues to evolve with time (except when all the f_i have the same fixed point). A similar phenomenon is known to occur in random dynamical systems. An SDE of the form

$$dx_t = a(x_t)\, dt + \sum_{i=1}^{n} b_i(x_t) \circ dW_t^i \tag{1.1}$$

gives rise to a stochastic flow of diffeomorphisms, in which almost every Brownian path defines a time-dependent flow (see *e.g.* [10]). When all of the Lyapunov exponents are strictly negative, trajectories are known to coalesce into *random sinks* (see [3, 13]). This phenomenon occurs naturally in applications, such as the Navier-Stokes system with sufficiently large viscosity (see *e.g.* [17, 18]), and in certain neural oscillator networks (see *e.g.* [14]).

In chaotic systems (autonomous or not), memory is lost quickly not through the coalescing of trajectories but for a diametrically opposite reason, namely their sensitive dependence on initial conditions. Small errors multiply quickly with time, so that in practice it is virtually impossible to track a specific trajectory in a chaotic system. For this reason, a statistical approach is often taken. Let ρ_0 denote an initial probability density with respect to a reference measure m, and suppose its time evolution is given by ρ_t. As with individual trajectories, one may ask if these probability distributions retain memories of their pasts. We will say a system loses its memory in the statistical sense if for two initial distributions ρ_0 and $\hat\rho_0$, $\int |\rho_t - \hat\rho_t|\, dm \to 0$ as $t \to \infty$. It is this form of memory loss that is studied in the present paper. Of particular interest is when memory is lost quickly: we say a system has *exponential statistical loss of memory* if there is a number $\alpha > 0$ such that for any ρ_0 and $\hat\rho_0$, $\int |\rho_t - \hat\rho_t|\, dm < Ce^{-\alpha t}$. Such memory loss may happen over a finite time interval, *i.e.*, for $t \leqslant T$, or for all $t \geqslant 0$.

Date: November 11, 2008.

2000 *Mathematics Subject Classification.* 37C60, 37C40.

Key words and phrases. memory loss, time-dependent dynamical systems, coupling, expanding maps, piecewise expanding maps.

William Ott is partially supported by NSF postdoctoral fellowship DMS 0603509.

Mikko Stenlund is partially supported by a fellowship from the Academy of Finland.

Lai-Sang Young is partially supported by NSF grant DMS 0600974.

Observe that while the two forms of memory loss described above are quite different on the phenomenological level, the latter can be seen mathematically as a manifestation of the first: By viewing $\{\rho_t\}_{t\geqslant 0}$ as a trajectory in the space of probability densities, statistical loss of memory is equivalent to ρ_t and $\hat{\rho}_t$ having a common future. The results of this paper are based on this point of view.

Before proceeding to specific results, we first describe a model that we think is very useful to keep in mind, even though the analysis of this model is somewhat beyond the scope of the present work.

Example 1.1. Lorentz gas with slowly moving scatterers. The 2-dimensional periodic Lorentz gas is usually modeled by the uniform motion of a particle in a domain $X = \mathbb{T}^2 \setminus \bigcup_i \Gamma_i$ where the Γ_i are pairwise disjoint convex subsets of $\mathbb{T}^2$ and the particle bounces off the "walls" of this domain (equivalently the boundaries of the scatterers) according to the rule that the angle of incidence is equal to the angle of reflection. In this model, the scatterers represent very heavy particles or ions, which move so slowly relative to the light particle (the one whose motion is described by the billiard flow) that one generally assumes they are fixed. This is the traditional setup in billiard studies. In reality, however, these large particles are bombarded by many light particles, and we focus on only one tagged light particle. The bombardments do cause the large particles to move about, though very slowly, and effectively independently of the motion of the tagged particle. Thus one can argue that it is more realistic to model the situation as a billiard flow in a *slowly varying environment*, *i.e.*, where the positions of the scatterers change very slowly with time. (See the recent work [8], which attempts to model the motion of a single heavy particle.)

In this paper, we prove exponential loss of memory in the statistical sense discussed above for time-dependent systems defined by expanding and piecewise expanding maps, the latter in one dimension only. Expanding maps (time-dependent or not) provide the simplest paradigms for exponential loss of memory in the statistical sense; we use them to illustrate our ideas on the most basic level as their analysis requires few technical considerations. Piecewise expanding maps, on the other hand, begin to exhibit some of the characteristics of the time-dependent billiard maps in the guiding example above. Our results can therefore be seen as a first step toward this physically relevant system.

The results of this paper apply to finite as well as infinite time, and our setting extends not only that of iterations of single maps (for which results on correlation decay for expanding maps and $1D$ piecewise expanding maps are not new), but it also includes skew products in which fiber dynamics are of these types as well as random compositions. What is different and new here is that the stationarity of the process is entirely irrelevant. Nor do the constituent maps have to belong to a bounded family, in which case the rates of memory loss may vary accordingly. A study which is closest to ours in spirit is [12].

Coupling methods are used in this paper, although we could have used spectral arguments, the Hilbert metric, or other techniques (see *e.g.* [6, 7, 15, 19, 20, 22, 23]). We do not claim that our methods are novel. On the contrary, one of the points of this paper is that under suitable conditions, existing methods for autonomous systems can be adapted to give results for this considerably broader class of dynamical settings, and we identify some of these conditions. Finally, even though coupling arguments have been used in more sophisticated settings, see *e.g.* [4, 5, 7, 23], we were unable to locate a coupling-based proof for single expanding maps. Section 2 will include this as a special case.

Notation. The following notation is used throughout: given $f_i : X \to X$ for $i \in \mathbb{N}$,

(1) for $n \geqslant m$, we write $F_{n,m} = f_n \circ \cdots \circ f_m$;
(2) for $n \geqslant 1$, we write $F_n = F_{n,1}$.

2. Time-dependent expanding maps

2.1. **Results.** Let M be a compact, connected Riemannian manifold without boundary. A smooth map $f : M \to M$ is called *expanding* if there exists $\lambda > 1$ such that

$$|Df(x)v| \geqslant \lambda |v|$$

for every $x \in M$ and every tangent vector v at x. Expanding maps provide the simplest examples of systems with exponential loss of statistical memory.

First we introduce some frequently-used notation. If ν is a Borel probability measure on M, then we let $f_*\nu$ denote the measure obtained by transporting ν forward using f, *i.e.*, $f_*\nu(E) = \nu(f^{-1}E)$ for all Borel sets E. If $d\nu = \varphi\, dm$ where m is the Riemannian measure on M, then the density of $f_*\nu$ is given by $\mathcal{P}_f(\varphi)$ where

$$\mathcal{P}_f(\varphi)(x) := \sum_{y \in f^{-1}x} \frac{\varphi(y)}{|\det Df(y)|}.$$

Here $\mathcal{P}_f$ is the transfer operator associated with the map f; $\mathcal{P}_{F_n}$ is defined similarly.

In order to have a uniform rate of memory loss, we need to impose some bounds on the set of mappings to be composed. For $\lambda \geqslant 0$ and $\Gamma \geqslant 0$, define

$$\mathcal{E}(\lambda, \Gamma) := \big\{ f : M \to M : \|f\|_{C^2} \leqslant \Gamma,\ \ |Df(x)v| \geqslant \lambda |v| \ \ \forall\, (x, v) \big\}$$

and let

$$\mathcal{D} := \big\{ \varphi > 0 : \int \varphi\, dm = 1,\ \ \varphi \text{ is Lipschitz} \big\}.$$

Theorem 1. *Given λ and Γ with $\lambda > 1$, there exists a constant $\Lambda = \Lambda(\lambda, \Gamma) \in (0, 1)$ such that for any sequence $f_i \in \mathcal{E}(\lambda, \Gamma)$ and any $\varphi, \psi \in \mathcal{D}$, there exists $C_{(\varphi,\psi)}$ such that*

$$\int |\mathcal{P}_{F_n}(\varphi) - \mathcal{P}_{F_n}(\psi)|\, dm \leqslant C_{(\varphi,\psi)} \Lambda^n \quad \forall\, n \geqslant 0. \tag{2.1}$$

Remark 2.1. We have assumed in Theorem 1 that all of the f_i are in a single $\mathcal{E}(\lambda, \Gamma)$. It will become clear that more general results in which λ and Γ are allowed to vary with i can be formulated and proved by concatenating the arguments below.

Remark 2.2. Correlation decay for expanding maps has been studied before. For a single map, see *e.g.* [19, 21]. For random compositions, see *e.g.* [1, 2]. For time-dependent maps, [12] proves that $\int |\mathcal{P}_{F_n}(\varphi) - \mathcal{P}_{F_n}(\psi)|\, dm \to 0$ as $n \to \infty$ without discussing the rate of convergence.

2.2. **Outline of proof.** Let $\varepsilon > 0$ be a small number to be determined, and fix $\lambda_0 > 1$ so that for all $f \in \mathcal{E} = \mathcal{E}(\lambda, \Gamma)$, we have $d(fx, fy) \geqslant \lambda_0 d(x, y)$ whenever $d(x, y) < \varepsilon$. Here $d(\cdot, \cdot)$ denotes Riemannian distance. For $L > 0$, we define

$$\mathcal{D}_L := \left\{ \varphi > 0 : \int \varphi\, dm = 1,\ \ \left| \frac{\varphi(x)}{\varphi(y)} - 1 \right| \leqslant L d(x, y) \ \text{if}\ d(x, y) < \varepsilon \right\}.$$

Notice that $\mathcal{D} = \bigcup_{L>0} \mathcal{D}_L$: For $\varphi \in \mathcal{D}$,

$$\left| \frac{\varphi(x)}{\varphi(y)} - 1 \right| = \frac{1}{\varphi(y)} |\varphi(x) - \varphi(y)| \leqslant \frac{\mathrm{Lip}(\varphi)}{\min(\varphi)} d(x, y);$$

functions in $\mathcal{D}_L$ are clearly locally Lipschitz. Key to the proof is the following observation:

Proposition 2.3. *There exists $L^* > 0$ for which the following holds. For any $L > 0$, there exists $\tau(L) \in \mathbb{Z}^+$ such that for all $\varphi \in \mathcal{D}_L$ and $f_i \in \mathcal{E}$, $\mathcal{P}_{F_n}(\varphi) \in \mathcal{D}_{L^*}$ for all $n \geqslant \tau(L)$.*

As our proof in Section 2.3 will show, the choice of L^* is arbitrary, provided it is greater than a number determined by λ and Γ.

Now let $f_i \in \mathcal{E}$ and $\varphi, \psi \in \mathcal{D}$ be given. Then there exists $N_0 = N_0(\varphi, \psi)$ such that both $\mathcal{P}_{F_{N_0}}(\varphi)$ and $\mathcal{P}_{F_{N_0}}(\psi)$ are in $\mathcal{D}_{L^*}$. This waiting period is the reason for the prefactor $C_{(\varphi,\psi)}$ on the right side of (2.1). With this out of the way, we may assume we start with two densities $\varphi, \psi \in \mathcal{D}_{L^*}$ from here on.

Notice that all functions in $\mathcal{D}_{L^*}$ are $\geqslant \kappa$ for some constant $\kappa > 0$; it is easy to see from the definition of $\mathcal{D}_{L^*}$ that they have uniform lower bounds on ε-disks. We think of the measures $\varphi\, dm$ and $\psi\, dm$ as having a part, namely $\kappa\, dm$, in common. Since $(F_n)_*(\kappa\, dm)$ will also be common to both $(F_n)_*(\varphi\, dm)$ and $(F_n)_*(\psi\, dm)$, we regard this part of the two measures as having been "matched". In order to retain control of distortion bounds, however, we will "match" only half of what is permitted, and renormalize the "unmatched part" as follows: Let

$$\hat{\varphi} = \frac{\varphi - \frac{1}{2}\kappa}{1 - \frac{1}{2}\kappa \cdot m(M)} \quad \text{and} \quad \hat{\psi} = \frac{\psi - \frac{1}{2}\kappa}{1 - \frac{1}{2}\kappa \cdot m(M)}\ . \tag{2.2}$$

Lemma 2.4. *For $\varphi \in \mathcal{D}_{L^*}$, if $\hat{\varphi}$ is as above, then $\hat{\varphi} \in \mathcal{D}_{2L^*}$.*

Let $N = \tau(2L^*)$ be given by Proposition 2.3. Then $\bar\varphi_N := \mathcal{P}_{F_N}(\hat\varphi)$ and $\bar\psi_N := \mathcal{P}_{F_N}(\hat\psi)$ are in $\mathcal{D}_{L^*}$. We subtract off $\frac{1}{2}\kappa$ from each of $\bar\varphi_N$ and $\bar\psi_N$ and renormalize as in (2.2), obtaining $\hat\varphi_N$ and $\hat\psi_N$ respectively. By Lemma 2.4, they are in $\mathcal{D}_{2L^*}$. In general, given $\hat\varphi_{(k-1)N}, \hat\psi_{(k-1)N} \in \mathcal{D}_{2L^*}$, we let

$$\bar\varphi_{kN} := \mathcal{P}_{F_{kN,(k-1)N+1}}(\hat\varphi_{(k-1)N}) \quad \text{and} \quad \bar\psi_{kN} := \mathcal{P}_{F_{kN,(k-1)N+1}}(\hat\psi_{(k-1)N}).$$

By Proposition 2.3, $\bar\varphi_{kN}, \bar\psi_{kN} \in \mathcal{D}_{L^*}$. We subtract off $\frac{1}{2}\kappa$ and renormalize to obtain $\hat\varphi_{kN}$ and $\hat\psi_{kN}$ in $\mathcal{D}_{2L^*}$ (Lemma 2.4), completing the induction.

Since a fraction of $\frac{1}{2}\kappa \cdot m(M)$ of the not-yet-matched parts of the measures is matched every N steps, we obtain

$$\int |\mathcal{P}_{F_n}(\varphi) - \mathcal{P}_{F_n}(\psi)|\, dm \leqslant 2(1 - \frac{1}{2}\kappa \cdot m(M))^k \quad \text{for} \quad kN \leqslant n < (k+1)N.$$

This leads directly to the asserted exponential estimate. ■

Remark 2.5. Theorem 1 also holds for initial densities that are not strictly positive provided one is able to guarantee that they eventually evolve into densities that are strictly positive. One way to make this happen is to have sufficiently many of the initial f_i remain in a small enough neighborhood of some fixed $f \in \mathcal{E}$, and take advantage of the fact that every expanding map f has the property that given any open set $U \subset M$, there exists $N(U) \in \mathbb{N}$ such that $f^n(U) \supset M$ for all $n \geqslant N(U)$.

2.3. **Details of proof.** We begin with an essential distortion estimate.

Lemma 2.6. *There exists a constant C_0 depending on λ_0 and Γ such that*

$$\frac{|\det DF_n(x)|}{|\det DF_n(y)|} \leqslant e^{C_0 d(F_n(x), F_n(y))}$$

for all $x, y \in M$ and $n \in \mathbb{Z}^+$ with the property that $d(F_k(x), F_k(y)) < \varepsilon$ for all $k < n$.

Proof of Lemma 2.6. We have

$$\begin{aligned}
\log \frac{|\det DF_n(x)|}{|\det DF_n(y)|} &= \sum_{k=1}^{n} (\log|\det Df_k(F_{k-1}(x))| - \log|\det Df_k(F_{k-1}(y))|) \\
&\leqslant \sum_{k=0}^{n-1} C_1 d(F_k(x), F_k(y)) \leqslant \sum_{k=0}^{n-1} C_1 \lambda_0^{-(n-k)} d(F_n(x), F_n(y)) \\
&\leqslant \frac{C_1}{\lambda_0 - 1}\, d(F_n(x), F_n(y)),
\end{aligned}$$

where C_1 is an upper bound on the Lipschitz constant of the $\mathcal{C}^1$ function $\log|\det Df|$ for any function f in the family $\mathcal{E}$. ■

We are in position to prove Proposition 2.3, which asserts the existence of $L^* > 0$ such that $\mathcal{D}_{L^*}$ attracts densities.

Proof of Proposition 2.3. Let $y \in D(x, \varepsilon)$ where $D(x, \varepsilon)$ is the disk of radius ε centered at x. We let $G_{n,i}$ be the i^{th} branch of $F_n^{-1}|D(x, \varepsilon)$, and let

$$\varphi_n^i := \frac{\varphi \circ G_{n,i}}{|\det DF_n \circ G_{n,i}|}.$$

Then φ_n^i is the contribution to the density $\varphi_n := \mathcal{P}_{F_n}(\varphi) = \sum_i \varphi_n^i$ obtained by pushing along the i^{th} branch. Estimating distortion one branch at a time, we have

$$\frac{\varphi_n^i(x)}{\varphi_n^i(y)} = \left(\frac{\varphi(G_{n,i}(x))}{\varphi(G_{n,i}(y))}\right) \cdot \left(\frac{|\det DF_n(G_{n,i}(y))|}{|\det DF_n(G_{n,i}(x))|}\right).$$

To estimate the first factor on the right, we use $d(G_{n,i}(x), G_{n,i}(y)) < \lambda_0^{-n} d(x, y)$ and $\varphi \in \mathcal{D}_L$. To estimate the second factor, we use Lemma 2.6. Combining the two, we obtain

$$\left|\log \frac{\varphi_n^i(x)}{\varphi_n^i(y)}\right| \leqslant (L\lambda_0^{-n} + C_0) d(x, y).$$

Exponentiating, moving $\varphi_n^i(y)$ to the right side, and summing over i before dividing by φ_n again, we obtain

$$\frac{\varphi_n(x)}{\varphi_n(y)} \leqslant e^{(L\lambda_0^{-n}+C_0)d(x,y)}.$$

By taking ε small enough, we may assume

$$\left|\frac{\varphi_n(x)}{\varphi_n(y)} - 1\right| \leqslant 2\left|\log\frac{\varphi_n(x)}{\varphi_n(y)}\right| \leqslant 2(L\lambda_0^{-n} + C_0)d(x,y). \tag{2.3}$$

Finally, we choose $\tau(L)$ large enough so that $L\lambda_0^{-\tau(L)} \leqslant C_0$, and conclude that

$$\left|\frac{\varphi_n(x)}{\varphi_n(y)} - 1\right| \leqslant L^* d(x,y)$$

for all $n \geqslant \tau(L)$, where $L^* = 4C_0$. ■

Only the proof of Lemma 2.4 remains.

Proof of Lemma 2.4. The distortion of $\hat{\varphi}$ satisfies

$$\left|\frac{\hat{\varphi}(x)}{\hat{\varphi}(y)} - 1\right| = \left|\frac{\varphi(x) - \frac{1}{2}\kappa}{\varphi(y) - \frac{1}{2}\kappa} - 1\right| = \left|\frac{\frac{\varphi(x)}{\varphi(y)} - \frac{\frac{1}{2}\kappa}{\varphi(y)}}{1 - \frac{\frac{1}{2}\kappa}{\varphi(y)}} - 1\right| = \frac{\left|\frac{\varphi(x)}{\varphi(y)} - 1\right|}{1 - \frac{\frac{1}{2}\kappa}{\varphi(y)}}.$$

Since $\varphi \geqslant \kappa$, the rightmost quantity above is $\leqslant 2\left|\frac{\varphi(x)}{\varphi(y)} - 1\right|$. We conclude that $\hat{\varphi} \in \mathcal{D}_{2L^*}$ if $\varphi \in \mathcal{D}_{L^*}$. ■

The proof of Theorem 1 is now complete.

3. Time-dependent $1D$ piecewise expanding maps

3.1. **Statement of results.** We consider in this section piecewise $\mathcal{C}^2$ expanding maps of the circle. More precisely, we let $\mathcal{S}^1$ be the interval $[0,1]$ with end points identified, and say $f : \mathcal{S}^1 \to \mathcal{S}^1$ is *piecewise $\mathcal{C}^2$ expanding* if there exists a finite partition $\mathcal{A}_1 = \mathcal{A}_1(f)$ of $\mathcal{S}^1$ into intervals such that for every $I \in \mathcal{A}_1$,

(1) $f|I$ extends to a $\mathcal{C}^2$ mapping in a neighborhood of I;
(2) there exists $\lambda > 1$ such that $|f'(x)| \geqslant \lambda$ for all $x \in I$.

It simplifies the analysis slightly to assume $\lambda > 2$, and we will do that (if $\lambda \leqslant 2$, we replace f by a suitable power of f and adjust the assumptions below accordingly).

Unlike the case of expanding maps (with no discontinuities), compositions of piecewise expanding maps do not necessarily have exponential loss of memory. Indeed, systems defined by a single piecewise expanding map may not even be ergodic, and decay of correlations (loss of memory) in that context is equivalent to mixing. Some additional conditions are therefore needed for results along the lines of Theorem 1. Let $\mathcal{A}_n := \bigvee_{i=1}^n f^{-(i-1)}\mathcal{A}_1$ be the join of the pullbacks of the partition $\mathcal{A}_1$ and let $\mathcal{A}_n|I$ be the restriction of $\mathcal{A}_n$ to the set I. For $J \subset \mathcal{S}^1$, let $\mathrm{int}(J)$ denote the interior of J.

Definition 3.1. We say f is *enveloping* if there exists $N \in \mathbb{Z}^+$ such that for every $I \in \mathcal{A}_1$, we have

$$\bigcup_{J \in \mathcal{A}_N|I} f^N(\mathrm{int}(J)) = \mathcal{S}^1.$$

The smallest such N is called the *enveloping time*.

If the enveloping time of f is N, then starting from any $I \in \mathcal{A}_1$, $f^N|I$ *overcovers* $\mathcal{S}^1$, in the sense that every $z \in \mathcal{S}^1$ lies in $f^N(J)$ for some $J \in \mathcal{A}_N|I$, and more than that: it is a positive distance from $f^N(\partial J)$. From here on, our universe $\mathcal{E}$ is comprised of piecewise $\mathcal{C}^2$ expanding, enveloping maps.

For the same reason that many (individual) piecewise expanding maps are not mixing, one cannot expect the arbitrary composition of piecewise expanding maps to produce exponential loss of memory – even when the constituent maps have good mixing properties: this is because such properties do not necessarily manifest themselves in a single step. To effectively leverage the mixing properties of individual maps, we may need a number of consecutive f_i to be near a single map. We will formulate two sets of results: *a local result*, which assumes that all the f_i are near a single piecewise expanding map g, and *a global result*, which allows the f_i to wander far and wide but slowly.

3.1.1. *Local result.* Let $g \in \mathcal{E}$ be fixed. We let $\Omega(g) = \{x_1 = x_{k+1}, x_2, \ldots, x_k\} \subset \mathcal{S}^1$ be the set of discontinuity points of g labeled counterclockwise, and let $d_\Omega(g) := \min_i |x_{i+1} - x_i|$. For $\varepsilon < \frac{1}{4} d_\Omega(g)$, we say $f \in \mathcal{E}$ is ε-near g, written $f \in \mathcal{U}_\varepsilon(g)$, if the following hold:

(1) $\Omega(f) = \{y_1 = y_{k+1}, y_2, \ldots, y_k\}$ where $|y_i - x_i| < \varepsilon$;
(2) if ξ_{fg} maps each interval $[x_i, x_{i+1}]$ affinely onto $[y_i, y_{i+1}]$, then on each $[x_i, x_{i+1}]$,

$$\|f \circ \xi_{fg} - g\|_{\mathcal{C}^2} < \varepsilon \ .$$

As in the case of single $1D$ piecewise expanding maps, a natural class of densities to consider is

$$\mathcal{D} = \left\{ \varphi \in \mathrm{BV}(\mathcal{S}^1, \mathbb{R}) : \varphi \geqslant 0, \ \int_{\mathcal{S}^1} \varphi(x)\, dx = 1 \right\}.$$

Recall the definitions of F_n and $\mathcal{P}_{F_n}$ from the end of Section 1 and the beginning of Section 2, respectively.

Theorem 2. *Let $g \in \mathcal{E}$. There exist $\Lambda < 1$ and $\varepsilon > 0$ sufficiently small (depending on g) such that for all $f_i \in \mathcal{U}_\varepsilon(g)$ and $\varphi, \psi \in \mathcal{D}$, there exists $C_{(\varphi,\psi)} > 0$ such that for all $n \in \mathbb{Z}^+$, we have*

$$\int_{\mathcal{S}^1} \left|\mathcal{P}_{F_n}(\varphi) - \mathcal{P}_{F_n}(\psi)\right| dx \leqslant C_{(\varphi,\psi)} \Lambda^n. \tag{3.1}$$

There exists an extensive literature on correlation decay for $1D$ piecewise expanding maps in the contexts of a single map and random i.i.d. compositions. See, *e.g.*, [1, 2, 9, 16].

3.1.2. *Global result.* It is straightforward to verify that the collection of sets $\mathcal{S} := \{\mathcal{U}_\varepsilon(f) : f \in \mathcal{E}, \ \ \varepsilon < \frac{1}{4} d_\Omega(f)\}$ generates a topology on $\mathcal{E}$.[1] Consider now a continuous map $\gamma : [a,b] \to \mathcal{E}$ (see Figure 1) and a finite or infinite sequence of f_i of the form $f_i = \gamma(t_i)$ where $a \leqslant t_1 \leqslant t_2 \leqslant t_3 \leqslant \cdots \leqslant b$. Let $\Delta := \max_i(t_{i+1} - t_i)$. If we think of the closed interval $[a,b]$ as time, then decreasing Δ corresponds to decreasing the 'velocity' at which the curve $\gamma([a,b])$ is traversed.

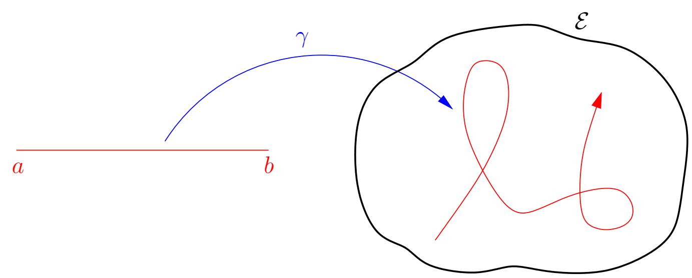

FIGURE 1. The picture we envision is that of "driving" the system along a curve γ in $\mathcal{E}$ and losing memory of past density distributions at *variable rates* depending on local characteristics. That, we submit, is the true nature of *memory loss in dynamical systems with slowly varying parameters.*

Theorem 3. *Let $\gamma : [a,b] \to \mathcal{E}$ be a continuous map. Then there exist $\delta_0 > 0$ and $\Lambda < 1$ (depending on γ) for which the following holds: For every $\{t_i\}$ as above with $\Delta \leqslant \delta_0$ and $\varphi, \psi \in \mathcal{D}$, there exists $C_{(\varphi,\psi)} > 0$ such that for all relevant n,*

$$\int_{\mathcal{S}^1} |\mathcal{P}_{F_n}(\varphi) - \mathcal{P}_{F_n}(\psi)|\, dx \leqslant C_{(\varphi,\psi)} \Lambda^n.$$

[1] To prove $\mathcal{S}$ forms the basis of a topology, it suffices to check that for $f_1, f_2 \in \mathcal{E}$, $\varepsilon_1, \varepsilon_2 > 0$, and $g \in \mathcal{U}_{\varepsilon_1}(f_1) \cap \mathcal{U}_{\varepsilon_2}(f_2)$, there exists $\varepsilon > 0$ such that $\mathcal{U}_\varepsilon(g) \subset \mathcal{U}_{\varepsilon_1}(f_1) \cap \mathcal{U}_{\varepsilon_2}(f_2)$.

Remark 3.2. We have tried not to overburden the formulation of Theorem 3, but as will be clear from the proofs, various generalizations are possible: The curve can be defined on an infinite interval and can traverse various subregions of $\mathcal{E}$ with nonuniform derivative bounds, leading to variable rates of memory loss. One does not, in fact, have to start with a prespecified curve and occasional long distance jumps can be accommodated.

Remark 3.3. Finally, we note that Theorem 3 – together with its generalizations mentioned in Remark 3.2 – is a simplified version of the Lorentz gas example in Section 1, an important difference being the absence of the stable directions.

3.2. **Proof of local result.** The following is an outline of the main steps of our proof:

Step 1. As in the expanding case (in Section 2), we represent the set of densities $\mathcal{D}$ as $\mathcal{D} = \bigcup_a \mathcal{D}_a$ where the conditions on $\mathcal{D}_a$ are more relaxed as a increases, and show that there is an a^* for which $\mathcal{D}_{a^*}$ is an attracting set under $\mathcal{P}_{F_n}$ for any sequence of f_i in a subset of $\mathcal{E}$ with uniform bounds. The time it takes to enter $\mathcal{D}_{a^*}$ from each $\mathcal{D}_a$ is shown to be bounded.

Step 2. Unlike the expanding case, where all functions in this attracting set are uniformly bounded away from 0, and coupling (or matching of densities) can be done immediately, we do not have such a bound here. Instead, we guarantee the matching of a fixed fraction of the measures *a finite number of steps later* using the enveloping property of g.

Step 3. To complete the cycle, we must show that after subtracting off the amount matched and renormalizing as in Section 2, functions in $\mathcal{D}_{a^*}$ are in $\mathcal{D}_a$ for some bounded a.

We now carry out these steps in detail.

Step 1. For $\varphi \in \mathrm{BV}(\mathcal{S}^1, \mathbb{R})$, we let $\bigvee_0^1 \varphi$ denote the total variation of φ, and let

$$\mathcal{D}_a := \left\{ \varphi \in \mathrm{BV}(\mathcal{S}^1, \mathbb{R}) : \varphi \geqslant 0, \ \int \varphi = 1, \ \bigvee_0^1 \varphi \leqslant a \right\}.$$

Clearly, $\mathcal{D} = \bigcup_a \mathcal{D}_a$. Let

$$\lambda(f) := \min_{I \in \mathcal{A}_1(f)} \inf_{x \in I} |f'(x)| \ ,$$

and recall the following well-known inequality originally due to Lasota and Yorke.

Lemma 3.4 (**Lasota-Yorke inequality** [11])**.** *Let f be a piecewise $\mathcal{C}^2$ expanding map. For $\varphi \in \mathrm{BV}(\mathcal{S}^1, \mathbb{R})$, we have*

$$\bigvee_0^1 \mathcal{P}_f(\varphi) \leqslant 2\lambda(f)^{-1} \bigvee_0^1 \varphi + A(f) \|\varphi\|_{L^1} \tag{3.2}$$

where

$$A(f) = \sup_{z \in \mathcal{S}^1} \frac{|f''(z)|}{|f'(z)|^2} + 2 \sup_{I \in \mathcal{A}_1(f)} \frac{\sup_{z \in I} |f'(z)|^{-1}}{|I|}.$$

We now fix $\mathcal{E}_0 \subset \mathcal{E}$ with uniform $\mathcal{C}^2$ bounds and with g well inside $\mathcal{E}_0$. Let

$$\lambda_0 := \inf_{f \in \mathcal{E}_0} \lambda(f) \quad \text{and} \quad A_0 := \sup_{f \in \mathcal{E}_0} A(f).$$

We assume $\lambda_0 > 2$. Upon repeated applications of (3.2), for $f_i \in \mathcal{E}_0$ and $\varphi \in \mathcal{D}$ we obtain

$$\bigvee_0^1 \mathcal{P}_{F_n}(\varphi) \leqslant (2\lambda_0^{-1})^n \bigvee_0^1 \varphi + \frac{A_0}{1 - 2\lambda_0^{-1}}, \tag{3.3}$$

which is the analog of the distortion estimate (2.3) in Section 2.

Our main result in Step 1 is

Proposition 3.5. *Fix any $a^* > \frac{A_0}{1-2\lambda_0^{-1}}$. Then for every $a > 0$, there exists $\tau(a) \in \mathbb{Z}^+$ such that for all $f_i \in \mathcal{E}_0$, $\varphi \in \mathcal{D}_a$ and $n \geqslant \tau(a)$, $\mathcal{P}_{F_n}(\varphi) \in \mathcal{D}_{a^*}$.*

Proof. This is an immediate consequence of (3.3). In fact, it is enough to choose

$$\tau(a) \geqslant \ln\left(\left(a^* - \frac{A_0}{1-2\lambda_0^{-1}}\right)a^{-1}\right) \Big/ \ln(2\lambda_0^{-1}) \tag{3.4}$$

among nonnegative integers. ∎

Step 2. The second step is perturbative. We will first work with iterates of g before extending our results to f_i in some suitable $\mathcal{U}_\varepsilon(g)$.

Lemma 3.6. *There exist $n_0 \in \mathbb{Z}^+$ and $\kappa_0 > 0$ (depending on g) such that for all $\varphi \in \mathcal{D}_{a^*}$, $\mathcal{P}_{g^{n_0}}(\varphi) \geqslant \kappa_0$.*

Proof. Let $\mathcal{A}_1$ be the partition for g, and let n_1 be such that all elements of $\mathcal{A}_{n_1}$ have length $< \frac{1}{2a^*}$. We will show that for every $\varphi \in \mathcal{D}_{a^*}$ there exists $J = J(\varphi) \in \mathcal{A}_{n_1}$ such that $\varphi|J \geqslant \frac{1}{2}$. Suppose, to derive a contradiction, that for each $J \in \mathcal{A}_{n_1}$, there exists $z_J \in J$ with $\varphi(z_J) < \frac{1}{2}$. Then

$$\int_J \varphi \leqslant |J|\left(\varphi(z_J) + \bigvee_J \varphi\right) < \frac{|J|}{2} + \frac{1}{2a^*}\bigvee_J \varphi.$$

Summing over J, this gives $\int_{\mathcal{S}^1} \varphi < \frac{1}{2} + \frac{1}{2} = 1$.

Next we claim that for every $J \in \mathcal{A}_{n_1}(g)$, there exists $s = s(J)$ and a subinterval $J_s \subset J$ such that $g^s|J_s$ is $\mathcal{C}^2$ and $g^s(J_s) \supset I$ for some $I \in \mathcal{A}_1(g)$. To prove this, we inductively define a nested sequence of intervals $J = J_1 \supset J_2 \supset J_3 \supset \cdots$ as follows. Assume that J_k has been defined. If $g^k(J_k) \supset I$ for some $I \in \mathcal{A}_1(g)$, set $s = k$. If not, then either $g^k(J_k) \subset I$ for some $I \in \mathcal{A}_1(g)$ or $g^k(J_k)$ intersects 2 elements of $\mathcal{A}_1(g)$. In the former case, set $J_{k+1} = J_k$, and in the latter, let J_{k+1} be the longer of the 2 intervals in $\mathcal{A}_1(g)|g^k(J_k)$. This process must terminate in a finite number of steps because $\inf |g'| > 2$.

Let $n_0 := s_0 + N$ where $s_0 = \max\{s(J) : J \in \mathcal{A}_{n_1}\}$ and $N = N(g)$ is the enveloping time for g. We now produce the κ_0 with the asserted property in the lemma. Fix arbitrary $\varphi \in \mathcal{D}_{a^*}$. Let $J = J(\varphi) \in \mathcal{A}_{n_1}$ be such that $\varphi|J \geqslant \frac{1}{2}$, and let $I \in \mathcal{A}_1$ be such that $g^{s(J)}(J_s) \supset I$. Then $\mathcal{P}_{g^{s(J)}}(\varphi)|I \geqslant \frac{1}{2}M_0^{-s(J)}$ where $M_0(g) := \sup |g'|$. From $g^N(I) = \mathcal{S}^1$, it follows that $\mathcal{P}_{g^{s(J)+N}}(\varphi) \geqslant \frac{1}{2}M_0^{-(s(J)+N)}$ on $\mathcal{S}^1$. We still have some steps to go if $s(J) < s_0$, but g is onto (as all enveloping maps are necessarily onto), and even in the worst-case scenario, we still have $\mathcal{P}_{g^{n_0}}(\varphi) \geqslant \frac{1}{2}M_0^{-n_0} := \kappa_0$ everywhere on $\mathcal{S}^1$. ∎

Define

$$\mathcal{A}(F_n) := \bigvee_{i=1}^{n} (F_{i-1})^{-1}\mathcal{A}_1(f_i)$$

where F_0 is the identity map. Now let $f_i \in \mathcal{U}_\varepsilon(g)$. From the one-to-one correspondence between elements of $\mathcal{A}_1(f_i)$ and $\mathcal{A}_1(g)$, one deduces that provided ε is sufficiently small, there is a well-defined mapping $\Phi_n : \mathcal{A}_n(g) \to \mathcal{A}(F_n)$ where for $J \in \mathcal{A}_n(g)$, $\Phi_n(J) \in \mathcal{A}(F_n)$ has the same itinerary as J. (In general, Φ_n need not be onto.) For $J = (a, b)$, let $J_\delta := (a + \delta, b - \delta)$.

Lemma 3.7. *Let n_0 be as in Lemma 3.6. Then there exist $\varepsilon > 0$ with $\mathcal{U}_\varepsilon(g) \subset \mathcal{E}_0$ and $\kappa > 0$ such that for all $f_i \in \mathcal{U}_\varepsilon(g)$, $\mathcal{P}_{F_{n_0}}(\varphi) \geqslant \kappa$ for all $\varphi \in \mathcal{D}_{a^*}$.*

Proof of Lemma 3.7. Let $\varphi \in \mathcal{D}_{a^*}$ be fixed. In the argument below, $\varepsilon > 0$ and $\delta > 0$ will be taken to be as small as is needed (ε and δ depend on g and on a^* but not on φ). We let n_1, $J = J(\varphi) \in \mathcal{A}_{n_1}(g)$, $s(J) \in \mathbb{Z}^+$, and $I \in \mathcal{A}_1(g)$ be as in the proof of Lemma 3.6. In particular, ε is small enough (depending on g and n_1) so that $\Phi_n : \mathcal{A}_n(g) \to \mathcal{A}(F_n)$ is well defined for all $f_i \in \mathcal{U}_\varepsilon(g)$ and the following 2 values of n: $n = n_1$ and $n = N$, where $N = N(g)$ is the enveloping time for g.

We claim that for every $I \in \mathcal{A}_1(g)$ and $f_i \in \mathcal{U}_\varepsilon(g)$, we may assume that $F_N(I_\delta) = \mathcal{S}^1$. For each $I' \in \mathcal{A}_N(g)|I$, $g^N(I')$ and $F_N(\Phi_N(I'))$ can be made arbitrarily close. This conclusion remains true if we shrink I' by a small amount, *i.e.*, δ (we need only do this for the leftmost and rightmost $I' \in \mathcal{A}_N(g)|I$). The assertion follows from this and the enveloping property of g.

Now let $f_i \in \mathcal{U}_\varepsilon(g)$ be fixed, and let $J' = \Phi_{n_1}(J)$. Assuming ε is chosen sufficiently small, $J'' = J' \cap J \neq \emptyset$, and $F_{s(J)}(J'') \supset I_\delta$ where δ is as in the previous paragraph. Thus $F_{s(J)+N}(J'') = \mathcal{S}^1$, and since $F_{n_0,s(J)+N+1}$

is onto, it follows that $F_{n_0}(J'') = \mathcal{S}^1$. Noting that $\varphi|J'' \geqslant \frac{1}{2}$, we conclude that

$$\mathcal{P}_{F_{n_0}}(\varphi) \geqslant \frac{1}{2}(M_0+\varepsilon)^{-n_0} := \kappa.$$

■

Step 3. The matching process introduces, for $\varphi \in \mathcal{D}_{a^*}$ with $\varphi \geqslant \kappa$, a new density

$$\hat{\varphi} := \frac{\varphi - \kappa}{1-\kappa}.$$

(We may subtract off any amount $\leqslant \kappa$, the only requirement being that $\hat{\varphi}$ remains $\geqslant 0$.) Since subtracting a constant does not diminish variation, and magnifying it by a constant c magnifies the variation by at most c, it follows that $\hat{\varphi} \in \mathcal{D}_{a^*(1-\kappa)^{-1}}$.

Proof of Theorem 2. We first iterate φ and ψ until $\mathcal{P}_{F_n}(\varphi) \in \mathcal{D}_{a^*}$ and $\mathcal{P}_{F_n}(\psi) \in \mathcal{D}_{a^*}$. This accounts for the prefactor $C_{(\varphi,\psi)}$ in (3.1). We then follow the matching scheme in the proof of Theorem 1, obtaining $\Lambda = (1-\kappa)^{(n_0+\tau(a^*(1-\kappa)^{-1}))^{-1}}$. ■

3.3. **Proof of global results.** Since $\gamma([a,b])$ is compact, we may assume it lies in a subset $\mathcal{E}_0$ of $\mathcal{E}$ with uniformly bounded derivatives and a minimum expansion $\lambda_0 > 2$ as in Section 3.2. This implies in particular that the set $\mathcal{D}_{a^*}$ can be taken to be uniform for all $g \in \gamma([a,b])$.

For each $g \in \mathcal{E}_0$, there are three numbers that are relevant:

(1) $\varepsilon(g)$, which describes the size of the neighborhood in which our local results apply;
(2) $\kappa(g) > 0$ as given by Lemma 3.7;
(3) $n(g) := n_0(g) + \tau(a^*(1-\kappa)^{-1})$ where n_0 is as in Lemma 3.6 and involves the enveloping time of g.

These quantities depend not just on the derivatives of g but on its geometry, *i.e.* how $\mathcal{A}_n(g)$ partitions $\mathcal{S}^1$, how quickly the covering property takes hold, and so on. Our local results imply that for all $f_i \in \mathcal{U}_{\varepsilon(g)}(g)$ and $\varphi, \psi \in \mathcal{D}_{a^*}$, $n(g)$ is the number of steps at the end of which we are guaranteed that the pair of densities has been matched once, and that their unmatched parts, renormalized, are returned to $\mathcal{D}_{a^*}$. Moreover, the amount matched is $\geqslant \kappa(g)$.

Proof of Theorem 3. For each $t \in [a,b]$, let $V_\alpha(t)$ denote the α-neighborhood of t in $\mathbb{R}$, and let $\alpha(t) > 0$ be such that $\gamma(V_{\alpha(t)}(t) \cap [a,b]) \subset \mathcal{U}_{\varepsilon(\gamma(t))}(\gamma(t))$. By compactness, there exist $z_1 < z_2 < \cdots < z_D$ such that $\bigcup_j V_{\frac{1}{2}\alpha(z_j)}(z_j)$ covers $[a,b]$. Let $g_j = \gamma(z_j)$, $V_j = V_{\alpha(z_j)}(z_j)$, and $\frac{1}{2}V_j = V_{\frac{1}{2}\alpha(z_j)}(z_j)$. Define $\delta_0 := \min_j \frac{\alpha(z_j)}{2n(g_j)}$.

We claim that if t_i defines a partition on $[a,b]$, the mesh Δ of which is $\leqslant \delta_0$, then the $f_i = \gamma(t_i)$ will have the desired properties. Consider f_i for arbitrary i, and let $\varphi, \psi \in \mathcal{D}_{a^*}$. Since $t_i \in \frac{1}{2}V_j$ for some j, our choice of δ_0 assures that $f_i, f_{i+1}, \cdots, f_{i+n(g_j)-1} \in \mathcal{U}_{\varepsilon(g_j)}(g_j)$. Thus a matching will take place, and the process can be repeated again at the end of $n(g_j)$ steps. Since $\max_j n(g_j) < \infty$ and $\min_j \kappa(g_j) > 0$, exponential loss of memory is proved. ■

The argument above applies to γ defined on a compact interval. If the curve in $\mathcal{E}$ is infinite, one simply divides it up into suitably short segments and treats them one at a time (see Remark 3.2).

References

1. V. Baladi and L.-S. Young, *On the spectra of randomly perturbed expanding maps*, Comm. Math. Phys. **156** (1993), no. 2, 355–385. MR MR1233850 (94g:58172)
2. ———, *Erratum: "On the spectra of randomly perturbed expanding maps" [Comm. Math. Phys. **156** (1993), no. 2, 355–385; MR1233850 (94g:58172)]*, Comm. Math. Phys. **166** (1994), no. 1, 219–220. MR MR1309547 (95k:58125)
3. P. H. Baxendale, *Stability and equilibrium properties of stochastic flows of diffeomorphisms*, Diffusion processes and related problems in analysis, Vol. II (Charlotte, NC, 1990), Progr. Probab., vol. 27, Birkhäuser Boston, Boston, MA, 1992, pp. 3–35. MR MR1187984 (93h:58167)
4. X. Bressaud, R. Fernández, and A. Galves, *Decay of correlations for non-Hölderian dynamics. A coupling approach*, Electron. J. Probab. **4** (1999), no. 3, 19 pp. (electronic). MR MR1675304 (2000j:60049)
5. X. Bressaud and C. Liverani, *Anosov diffeomorphisms and coupling*, Ergodic Theory Dynam. Systems **22** (2002), no. 1, 129–152. MR MR1889567 (2003e:37032)
6. L. A. Bunimovich, Ya. G. Sinaĭ, and N. I. Chernov, *Statistical properties of two-dimensional hyperbolic billiards*, Uspekhi Mat. Nauk **46** (1991), no. 4(280), 43–92, 192. MR MR1138952 (92k:58151)

7. N. Chernov, *Advanced statistical properties of dispersing billiards*, J. Stat. Phys. **122** (2006), no. 6, 1061–1094. MR MR2219528 (2007h:37047)
8. N. Chernov and D. Dolgopyat, *Brownian Brownian motion - I*, Memoirs of the American Mathematical Society (2008), to appear.
9. F. Hofbauer and G. Keller, *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z. **180** (1982), no. 1, 119–140. MR MR656227 (83h:28028)
10. H. Kunita, *Stochastic flows and stochastic differential equations*, Cambridge Studies in Advanced Mathematics, vol. 24, Cambridge University Press, Cambridge, 1997, Reprint of the 1990 original. MR MR1472487 (98e:60096)
11. A. Lasota and J. A. Yorke, *On the existence of invariant measures for piecewise monotonic transformations*, Trans. Amer. Math. Soc. **186** (1973), 481–488 (1974). MR MR0335758 (49 #538)
12. ———, *When the long-time behavior is independent of the initial density*, SIAM J. Math. Anal. **27** (1996), no. 1, 221–240. MR MR1373154 (97a:47043)
13. Y. Le Jan, *On isotropic Brownian motions*, Z. Wahrsch. Verw. Gebiete **70** (1985), no. 4, 609–620. MR MR807340 (87a:60090)
14. K. Lin, E. Shea-Brown, and L.-S. Young, *Reliability of coupled oscillators*, to appear.
15. C. Liverani, *Decay of correlations*, Ann. of Math. (2) **142** (1995), no. 2, 239–301. MR MR1343323 (96e:58090)
16. ———, *Decay of correlations for piecewise expanding maps*, J. Statist. Phys. **78** (1995), no. 3-4, 1111–1129. MR MR1315241 (96d:58077)
17. N. Masmoudi and L.-S. Young, *Ergodic theory of infinite dimensional systems with applications to dissipative parabolic PDEs*, Comm. Math. Phys. **227** (2002), no. 3, 461–481. MR MR1910827 (2003g:37148)
18. J. C. Mattingly, *Ergodicity of 2D Navier-Stokes equations with random forcing and large viscosity*, Comm. Math. Phys. **206** (1999), no. 2, 273–288. MR MR1722141 (2000k:76040)
19. D. Ruelle, *The thermodynamic formalism for expanding maps*, Comm. Math. Phys. **125** (1989), no. 2, 239–262. MR MR1016871 (91a:58149)
20. ———, *Thermodynamic formalism*, second ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2004, The mathematical structures of equilibrium statistical mechanics. MR MR2129258 (2006a:82008)
21. M. Rychlik, *Regularity of the metric entropy for expanding maps*, Trans. Amer. Math. Soc. **315** (1989), no. 2, 833–847. MR MR958899 (90a:28027)
22. L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998), no. 3, 585–650. MR MR1637655 (99h:58140)
23. ———, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188. MR MR1750438 (2001j:37062)

(William Ott) Courant Institute of Mathematical Sciences, New York, NY 10012, USA
URL: `http://www.cims.nyu.edu/∼ott`

(Mikko Stenlund) Courant Institute of Mathematical Sciences, New York, NY 10012, USA; Department of Mathematics and Statistics, P.O. Box 68, Fin-00014 University of Helsinki, Finland.
URL: `http://www.math.helsinki.fi/mathphys/mikko.html`

(Lai-Sang Young) Courant Institute of Mathematical Sciences, New York, NY 10012, USA
URL: `http://www.cims.nyu.edu/∼lsy`