THE TRANSFER OPERATOR IS THE MEASURE-THEORETIC PUSHFORWARD

LINAS VEPSTAS <LINAS@LINAS.ORG>

ABSTRACT. This paper provides a simple proof that the Ruelle transfer operator (the Frobenius-Perron operator) is one and the same thing as the measure-theoretic pushforward. Here, the measure-theoretic pushforward is simply the idea of a pushforward applied to a measureable function between sigma algebras. The transfer operator, by contrast, an important tool for studying the behaviour of a measure-preserving dynamical system. It is normally defined as an operator in the sense of functional analysis, representing the time evolution of a dynamical system in terms of a Banach space of functions on the coordinate-space manifold of the dynamical system.

By identifying the transfer operator with the pushforward, a variety of curious conjectures suggest themselves, including the promotion of the Frobenius-Perron theorem to a near-categorical theorem on the category of pushforwards acting on measure spaces. Some additional conjectues regarding connections between the KAM theorem, homogeous spaces and number theory suggest themselves.

1. INTRODUCTION

The method of the transfer operator was introduced by David Ruelle[need ref] as a powerful mechanism for studying the nature of iterated maps. The transfer operator, sometimes called the Perron-Frobenius operator, or the Ruelle-Frobenius-Perron operator, provides a means to escape the narrow confines of point-set topology when considering an iterated function, and instead explore the function using wildly different topologies. In its most concrete form, it is a linear operator acting on a Banach space of functions. However, the structure and the properties of the operator depend very much on which space of functions one considers.

It is easiest to begin with the concrete definition. Consider a function $g : [0,1] \to [0,1]$, that is, a function mapping the unit interval of the real number line to itself. Upon iteration, the function may have fixed points or orbits of points. These orbits may be attractors or repellors, or may be neutral saddle points. The action of g may be ergodic or chaotic, strong-mixing or merely topologically mixing. In any case, the language used to discuss g is inherently based on either the point-set topology of the unit interval, or the "natural" topology on the unit interval, the topology of open sets.

A shift in perspective may be gained not by considering how g acts on points or open sets, but instead by considering how g acts on distributions on the unit interval. Intuitively, one might consider a dusting of points on the unit interval, with a local density given by $\rho(x)$ at point $x \in [0,1]$, and then consider how this dusting or density evolves upon iteration by g. This verbal description may be given form as

$$\rho'(y) = \int_0^1 \delta\left(y - g(x)\right) \rho(x)\, dx \tag{1.1}$$

Date: 18 October 2006.

where $\rho'(y)$ is the new density at point $y = g(x)$ and δ is the Dirac delta function.

In this viewpoint, g becomes an operator that maps densities ρ to other densities ρ', or notationally,

$$\mathscr{L}_g \rho = \rho' \tag{1.2}$$

The operator $\mathscr{L}_g$ is called the transfer operator or the Ruelle-Frobenius-Perron operator. It is not hard to see that it is a linear operator, in that

$$\mathscr{L}_g (a\rho_1 + b\rho_2) = a\mathscr{L}_g \rho_1 + b\mathscr{L}_g \rho_2 \tag{1.3}$$

for constants a, b and densities ρ_1, ρ_2.

When the function g is differentiable, and doesn't have a vanishing derivative, the integral formulation of the transfer operator above can be rephrased in a more convenient form, as

$$\left[\mathscr{L}_g \rho\right](y) = \sum_{x:y=g(x)} \frac{\rho(x)}{|dg(x)/dx|} \tag{1.4}$$

where the sum is presumed to extend over at most a countable number of points. If these conditions do not hold, a transfer operator can still be defined, although more care must be taken in its definition.

One generalization should be immediately apparent: although the word "density" implies that ρ is a smooth map from the unit interval to the non-negative reals, no such requirement need to be enforced: ρ may be a map from the unit interval to any ring R, and it need not be smooth, differentiable or even continuous. This generalization gives a very rich structure to $\mathscr{L}_g$: the precise form of $\mathscr{L}_g$ will take will depend very strongly on R, whether its the reals $\mathbb{R}$, the complex numbers $\mathbb{C}$, or some other field or ring. It will also depend strongly on whether one restricts oneself to smooth functions, continuous functions, square-integrable functions, or some other function space. An adequate study requires reference to the specific topology that the function space is endowed with; many different topologies may be considered. That is, in general, one must consider $\mathscr{L}_g$ to be an operator acting on a topological space endowed with multiplication and addition, that is, a topological vector space. A precise definition of the transfer operator is given in the next section, as being the pushforward onset of measureable, sigma-additive functions.

The structure of $\mathscr{L}_g$ also depends on the topology applied to the unit interval. Besides the natural topology on the real number line, the unit interval can be given several other topologies. The most important of these is the Cantor set topology, or the p-adic topology. Here, one considers the unit interval $[0,1]$ to consist of the set of strings

$$\Omega = \left\{ \sigma = (\sigma_0, \sigma_1, \sigma_2, \ldots) : \sigma_k \in \{0, 1, \ldots, p-1\}, x = \sum_{k=0}^{\infty} \sigma_k p^{-(k+1)}, x \in [0,1] \right\} \tag{1.5}$$

Intuitively, this set is simply the set of all the digits of a base-p expansion of the real numbers $x \in [0,1]$. The connection with physics comes from the realization that this set can be understood to be the collection of all field configurations of a one-dimensional, one-sided lattice, where each lattice location can take on one of p values. Such lattices are commonly given the product topology, where the open sets are the cylinder sets consisting of substrings of sequences of letters. The topology also has a natural measure, derived from the length of letter sequences. Aside from the p-adic expansion above, one also has the continued fraction expansion, where one considers the sequence of integers making up the continued fraction

$$x = [0; \sigma_1, \sigma_2, \sigma_3, \ldots] = \cfrac{1}{\sigma_1 + \frac{1}{\sigma_2 + \frac{1}{\sigma_3 + \ldots}}} \tag{1.6}$$

where each σ_k is a positive integer; the entire sequence again be given a product topology, although the measure is constructed differently.

2. Measure-theoretic description

The clearest and most precise, although also the most abstract formulation of the transfer operator is in the language of measure theory, where it is seen to be same thing as the push-forward. In this language, it gains a topological setting, which helps clarify how the transfer function behaves on spaces of functions. This definition is formulated, from first principles, below.

Consider a topological space X, and a field F over the reals $\mathbb{R}$. Here, F may be taken to be $\mathbb{R}$ itself, or $\mathbb{C}$ or some more general field over $\mathbb{R}$. The restriction of F being a field over the reals is necessitated by the measure-theoretic manipulations below, where functions will need to be multiplied by the real-valued measure.

One may then define the algrbra of functions $\mathscr{F}(X)$ on X as the set of functions $f \in \mathscr{F}(X)$ such that $f : X \to F$. The algebra of functions is a vector space, in that given two functions $f_1, f_2 \in \mathscr{F}(X)$, their linear combination $af_1 + bf_2$ is also an element of $\mathscr{F}(X)$; thus f_1 and f_2 may be interpreted to be the vectors of a vector space. An algebra is a vector space for which the multiplication of vectors is defined. In the case of $\mathscr{F}(X)$, the multiplication is the pair-wise multiplication of functions; that is, the product $f_1 f_2$ is defined as the function $(f_1 f_2)(x) = f_1(x) \cdot f_2(x)$, and so $f_1 f_2$ is again an element of $\mathscr{F}(X)$. Sine one clearly has $f_1 f_2 = f_2 f_1$, multiplication is commutative, and so $\mathscr{F}(X)$ is also a commutative ring.

The space $\mathscr{F}(X)$ may be endowed with a topology. The coarsest topology on $\mathscr{F}(X)$ is the *weak topology*, which is obtained by taking $\mathscr{F}(X)$ to be the space that is dual to X. As a vector space, $\mathscr{F}(X)$ may be endowed with a norm $\|f\|$. For example, one may take the norm to be the L^p-norm

$$\|f\|_p = \left(\int |f(x)|^p \, dx \right)^{1/p} \tag{2.1}$$

For $p = 2$, this norm converts the space $\mathscr{F}(X)$ into the Hilbert space of square-integrable functions on X. Other norms are possible, in which case $\mathscr{F}(X)$ has the structure of a Banach space rather than a Hilbert space.

Consider now a homomorphism of topological spaces $g : X \to Y$. This homomorphism induces the pullback $g^* : \mathscr{F}(Y) \to \mathscr{F}(X)$ on the algebra of functions, by mapping $f \mapsto g^*(f) = f \circ g$ so that $f \circ g : Y \to F$. The pullback is a linear operator, in that

$$g^*(af_1 + bf_2) = ag^*(f_1) + bg^*(f_2) \tag{2.2}$$

This is very easily demonstrated by considering how $g^* f$ acts at a point: $(g^* f)(x) = (f \circ g)(x) = f(g(x))$ and so the linearity of g^* on $af_1 + bf_2$ follows trivially.

One may construct an analogous mapping, but going in the opposite direction, called the push-forward: $g_* : \mathscr{F}(X) \to \mathscr{F}(Y)$. There are two ways of defining a push-forward. One way is to define it in terms of the sheaves of functions on subsets of X and Y. The sheaf-theoretic description is more or less insensitive to the ideas of measurability, whereas this is important to the definition of the transfer operator, as witnessed by the appearance of the Jacobian determinant in equation 1.4. By contrast, the measure-theoretic push-forward captures this aspect. It may be defined as follows.

One endows the spaces X and Y with sigma-algebras $(X,\mathscr{A})$ and $(Y,\mathscr{B})$, so that $\mathscr{A}$ is the set of subsets of X obeying the axioms of a sigma-algebra, and similarly for $\mathscr{B}$. A mapping $g : X \to Y$ is called "measurable" if, for all Borel sets $B \in \mathscr{B}$, one has the pre-image $g^{-1}(B) \in \mathscr{A}$ being a Borel set as well. Thus, a measurable mapping induces a push-forward on the sigma-algebras: that is, one has a push-forward $g_* : \mathscr{F}(\mathscr{A}) \to \mathscr{F}(\mathscr{B})$ given by $f \mapsto g_*(f) = f \circ g^{-1}$, which is defined by virtue of the measurability of g. The push-forward is a linear operator, in that $g_*(af_1 + bf_2) = ag_*(f_1) + bg_*(f_2)$.

One regains the transfer operator as defined in equation 1.4 by considering the limiting behavior of the push-forward on progressively smaller sets. That is, one has

Theorem 2.1. *The transfer operator is the point-set topology limit of the measure-theoretic push-forward.*

Proof. The proof that follows is rather informal, so as to keep it simple. It is aimed mostly at articulating the language and terminology of measure theory. The result is none-the-less rigorous, if taken within the confines of the definitions presented.

Introduce a measure $\mu : \mathscr{A} \to \mathbb{R}^+$ and analogously $\nu : \mathscr{B} \to Y$. The mapping g is measure-preserving if ν is a push-forward of μ, that is, if $\nu = g_*\mu = \mu \circ g^{-1}$. The measure is used to rigorously define integration on X and Y. Elements of $\mathscr{F}(\mathscr{A})$ can be informally understood to be integrals, in that $f(A)$ for $A \in \mathscr{A}$ may be understood as

$$f(A) = \int_A \tilde{f}(z)d\mu(z) = \int_A \tilde{f}(z)\left|\mu'(z)\right|dz \tag{2.3}$$

where $|\mu'(x)|$ is to be understood as the Jacobean determinant at a point $x \in X$. Here, $\tilde{f}$ can be understood to be a function that is being integrated over the set A, whose integral is denoted by $f(A)$. The value of $\tilde{f}$ at a point $x \in X$ can be obtained by means of a limit. One considers a sequence of $A \in \mathscr{A}$, each successively smaller than the last, each containing the point x. One then has

$$\lim_{\overrightarrow{A \ni x}} \frac{f(A)}{\mu(A)} = \tilde{f}(x) \tag{2.4}$$

which can be intuitively proved by considering A so small that $\tilde{f}$ is approximately constant over A:

$$f(A) = \int_A \tilde{f}(z)d\mu(z) \approx \tilde{f}(x)\int_A d\mu = \tilde{f}(x)\mu(A) \tag{2.5}$$

To perform the analogous limit for the push-forward, one must consider a point $y \in Y$ and sets $B \in \mathscr{B}$ containing y. In what follows, it is now assumed that $g : X \to Y$ is a multi-sheeted countable covering of Y by X. By this it is meant that for any y that is not a branch-point, there is a nice neighborhood of y such that its pre-image consists of the union of an at most countable number of pair-wise disjoint sets. That is, for y not a branch point, and for $B \ni y$ sufficiently small, one may write

$$g^{-1}(B) = A_1 \cup A_2 \cup \cdots = \bigcup_{j=1}^{k} A_j \tag{2.6}$$

where k is either finite or stands for ∞, and where $A_i \cap A_j = \varnothing$ for all $i \neq j$. A branch points, such a decomposition may not be possible. The axiom of sigma-additivity guarantees that such multi-sheeted covers behave just the way one expects integrals to behave: in other words, one has

$$\mu\left(g^{-1}(B)\right)=\mu\left(\bigcup_{j=1}^{k}A_j\right)=\sum_{j=1}^{k}\mu\left(A_j\right) \tag{2.7}$$

whenever the collection of A_j are pair-wise disjoint. Similarly, in order to have the elements $f\in\mathscr{F}(\mathscr{A})$ behave as one expects integrals to behave, one must restrict $\mathscr{F}(\mathscr{A})$ to contain only sigma-additive functions as well, so that

$$f\left(g^{-1}(B)\right)=f\left(\bigcup_{j=1}^{k}A_j\right)=\sum_{j=1}^{k}f\left(A_j\right) \tag{2.8}$$

As the set B is taken to be smaller and smaller, the sets A_j will become smaller as well. Denote by x_j the corresponding limit point of each A_j, so that $g(x_j)=y$ and the pre-image of y consists of these points: $g^{-1}(y)=\left\{x_1,x_2,\cdots\,\middle|\,g(x_j)=y\right\}$. One now combines these provisions to write

$$\begin{aligned}
\left[g_*\tilde{f}\right](y) &= \lim_{\overrightarrow{B\ni y}}\left[\frac{\left(g_*f\right)(B)}{\nu(B)}\right]\\
&= \lim_{\overrightarrow{B\ni y}}\left[\frac{\left(f\circ g^{-1}\right)(B)}{\nu(B)}\right]\\
&= \lim_{\overrightarrow{A_j\ni g^{-1}(y)}}\frac{f\left(A_1\cup A_2\cup\cdots\right)}{\nu(B)}\\
&= \lim_{\overrightarrow{A_j\ni g^{-1}(y)}}\frac{\sum_{j=1}^{k}f\left(A_j\right)}{\nu(B)}\\
&= \sum_{j=1}^{k}\tilde{f}\left(x_j\right)\lim_{\overrightarrow{A_j\ni x_j}}\frac{\mu\left(A_j\right)}{\nu(B)} \qquad (2.9)
\end{aligned}$$

The limit in the last line of this sequence of manipulations may be interpreted in two ways, depending on whether one wants to define the measure ν on Y to be the push-forward of μ, or not. If one does take it to be the push-forward, so that $\nu=g_*\mu$, then one has

$$\lim_{\overrightarrow{A_j\ni x_j}}\frac{\mu\left(A_j\right)}{g_*\mu(B)}=\frac{1}{\left|g'\left(x_j\right)\right|} \tag{2.10}$$

where $\left|g'\left(x_j\right)\right|$ is the Jacobian determinant of g at x_j. This last is a standard result of measure theory, and can be intuitively proved by noting that $g\left(A_j\right)=B$, so that

$$\nu\left(B\right)=\int_{A_j}g'(z)\,d\mu(z)\approx g'\left(x_j\right)\mu\left(A_j\right) \tag{2.11}$$

for "small enough" B. Assembling this with the previous result, one has

$$\left[g_*\tilde{f}\right](y)=\sum_{x_j\in g^{-1}(y)}\frac{\tilde{f}\left(x_j\right)}{\left|g'\left(x_j\right)\right|} \tag{2.12}$$

which may be easily recognized as equation 1.4. This concludes the proof of the theorem, that the transfer operator is just the point-set topology limit of the push-forward. □

Some curious lemmas show up along the way.

Lemma 2.2. *One has*

$$\sum_{j=1}^{k} \frac{1}{\left|g'(x_j)\right|} = 1 \tag{2.13}$$

Proof. This follows by taking the limit $\overrightarrow{A_j \ni x_j}$ of

$$\frac{\mu(A_j)}{g_*\mu(B)} = \frac{\mu(A_j)}{\sum_{i=1}^{k}\mu(A_i)}$$

and then summing over j. □

Corollary 2.3. *The constant function is an eigenvector of the transfer operator, associated with the eigenvalue one.*

Proof. This may be proved in two ways. From the viewpoint of point-sets, one simply takes $\tilde{f} = \text{const.}$ in equation 2.12, and applies the lemma above. From the viewpoint of the sigma-algebra, this is nothing more than the seemingly vacuous statement that $\nu = g_*\mu$. This becomes slightly less vacuous if one takes the space $Y = X$, so that $g : X \to X$ is a measure-preserving map: $g_*\mu = \mu$. □

The last corollary suggests two interesting conjectures.

Conjecture 2.4. *(Ruelle-Perron-Frobenius theorem). All transfer operators are compact, bounded operators.*

This conjecture is of course just the Frobenius-Perron theorem, recast in the context of measure theory. I suppose it has a simple-enough proof that the author hasn't yet seen or re-discovered. XXX Don't let this sit there. XXX, Ref:Dunford and Schwartz. A corollary to the Perron-Frobenius theorem is then:

Corollary 2.5. *(Haar measure) For any homomorphisms $g : X \to X$, one may find a measure μ such that $g_*\mu = \mu$; that is, every homomorphism g of X induces a measure μ on X such that g is a measure-preserving map. If g is ergodic, then the measure is unique.*

By the Frobenius-Perron theorem, this measure is unique. It is, of course the Haar measure, although this seems to be an unusual way of finding the Haar measure. Sketch of semi-proof: μ is a fixed point of g_*. The fixed point exists because $g_*$is a bounded operator, and the space of measures is compact, and so a bounded operator on a compact space will have a fixed point. Ref. some appropriate fixed-point thm (Ref Dunford and Schwartz). To prove uniqueness, we have to show that g is ergodic, i.e. that there are no invariant subspaces, that orbit of g is the whole space. Else, if g is not ergodic on the whole space, then show that orbit of g splits or foliates the measure space into a bunch of pairwise disjoint leaves, and its ergodic on each leaf, there's a distinct fixed point μ in each leaf, and there's a symmetry group that takes μ in one leaf to that in another.

XXX Make note that there is a certain sense of duality floating here: if there are "discrete" symmetries of g, then the space is foliated into ergodic leaves. One can then presumably state the KAM theorem in this language. The KAM theorem appears to state that small perturbations cause the leaves to touch with one another at certain points, while also leaving large "islands of stability", i.e. non-zero-measure, compact, simply-connected regions of a leaf which do not touch any other leaf. That is, an "island of stability" is what would be called a "nice neighborhood" when discussing the action of a discrete group, e.g. the Fuchsian group. From this viewpoint, an "island of stability" is just a "homogenous space". An "island of stability" is a "fundamental domain"; just as the action of a discrete group carries one fundamental domain into another, so here, we have thanks to the KAM theorem, the appeareance of a discrete group, which carries one "island of stability" to another. This now opens the dorr to the application of concepts from homogenous spaces to dyanmical systems: What is the "j-invariant" of the island of stability? Just as the hypergeometric functions are a kind of "quantization" of homogenous spaces, so it would seem that the analog of a hypergeometric function on an island of stability would provide a certain quantization of a KAM system. Just as there are rigidity theorems for homogenous spaces, are there corresponding rigidity theorems for dynamical systems?

XXX argue that push-forward and pullback are opposites, that pullback is the Koopman operator.

XXX Make the note that ρ is an element of the dual topology. That is, given a topology, and a function on topology that blah dual...

3. Bibliography

References