Conditional Entropy of Nonlinear Dynamical Systems

Olaf Stefan Göhrs and Joachim Holzfuss

Institut für Angewandte Physik, TU Darmstadt
Schloßgartenstr. 7, 64289 Darmstadt, Germany

Abstract. We present a method for detecting and quantifying relationships between two ore more time series. It is based on the estimation of conditional entropies. In the case of continuous systems an appropriate conditional probability is calculated.

Our method can be used for deciding, if one observable Y depends on another one X. If Y can be expressed as a continuous function $f(X)$, the conditional entropy vanishes.

These facts help to find optimal delay times in the reconstruction of nonlinear dynamics. Furthermore, we show the advantage of variable delay times.

INTRODUCTION

For the analysis of nonlinear dynamical systems interdependences between different time sequences X and Y are of special interest. Covariance and mutual information are frequently used for quantifying such relationships [1], [2], [3]. But they are symmetric and therefor one cannot decide if an observable is completely determined by another one.

The conditional entropy $H(Y|X)$ seems to be an appropriate measure [4]. It is defined as the expection

$$H(Y|X) := -E(\log P(Y = y_j|X = x_i)) \tag{1}$$
$$= -\sum_{i,j} P(Y = y_j, X = x_i) \log P(Y = y_j|X = x_i) \tag{2}$$

where $P(Y = y_j|X = x_i)$ denotes the probability that Y is in the state y_j under the condition that $X = x_i$. There are some advantageous properties of $H(Y|X)$:

(A) It vanishes, if Y is a function of X:
$H(Y|X) = 0$, if $Y = f(X)$.

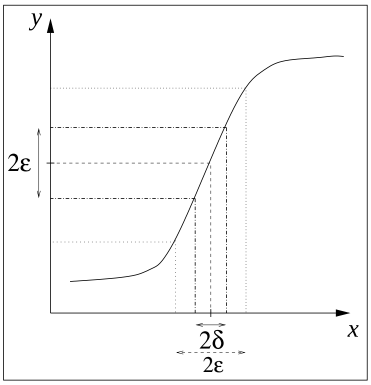
(B) It is not increased by adding other observables:
$H(Y|(X_1, \cdots, X_i)) \leq H(Y|(X_1, \cdots, X_{i-1}) \leq \cdots \leq H(Y|X_1) \leq H(Y)$.

The problem is that the definition of $H(Y|X)$ requires discrete valued observables. For the continuous case one has to find an equivalent conditional probability.

A simple way of defining such probabilities in continuous state spaces is to identify the probability of a single point with the probability of its ϵ-environments:

$$P_\epsilon(y|x) := P(Y \in B_\epsilon(y)|X \in B_\epsilon(x)), \tag{3}$$

$B_\epsilon(c)$ being a ball with center c and radius ϵ. In the one-dimensional case these environments are the intervals $I_\epsilon(x) = [x-\epsilon, x+\epsilon]$, respectively $I_\epsilon(y) = [y-\epsilon, y+\epsilon]$.



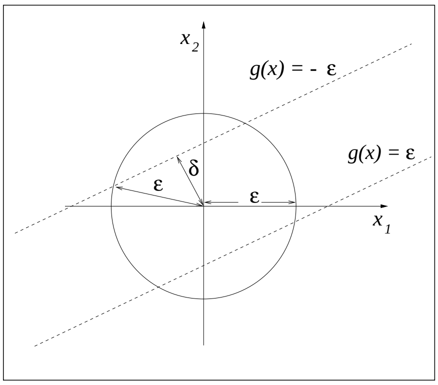


FIGURE 1. Left: An interval is stretched by a continuous function. Right: The part of the ball, which is mapped to the interval $I_\epsilon(y)$ is between the two dotted lines.

Let $f : I \to I$ be a nonlinear, differentiable function that maps an interval $I \subset \mathbf{R}$ to itself; the interval that is mapped by f to $I_\epsilon(y)$ is of size 2δ (Fig. 1, left). For small ϵ one can linearize f by using its derivative f' and calculate $\delta \simeq \frac{\epsilon}{|f'(x)|}$. That means if $|f'(x)| > 1$ the conditional probability becomes $P_\epsilon(y|x) < 1$, even if all points in $I_\epsilon(y)$ are completely determined by the points in $I_\epsilon(x)$. If one requires that in this case the conditional probability becomes 1, $P_\epsilon(y|x)$ has to be adapted by a factor $\eta^{-1}(x,y)$ where $\eta(x,y)$ denotes the fraction of the interval $I_\epsilon(x)$, that is mapped to $I_\epsilon(y)$. Assuming an equidistribution in $I_\epsilon(x)$ one finds $\eta(x,y) = |f'(x)|^{-1}$.

Now consider that $\mathbf{x} \in \mathbf{R}^n$. A linearized map can be calculated by the total differential, $g(\mathbf{x}) = \mathrm{d}f(\mathbf{x})$. The part of $B_\epsilon(\mathbf{x})$, which is mapped to $I_\epsilon(y)$, is between the hyperplanes $g(\mathbf{x}) = \pm\epsilon$ (Fig. 1, right). Assuming equidistribution again the fraction $\eta(\mathbf{x}, y)$ can be estimated by the quotient of the ball's volume between those hyperplanes $V_{\delta,n}$ and the entire volume $V_{\epsilon,n}$, i.e.:

$$\eta_{(n)} = \frac{V_\delta^n}{V_\epsilon^n} = \frac{\int\limits_0^\delta \left(\sqrt{\epsilon^2 - \delta^{*2}}\right)^{n-1} \mathrm{d}\delta^*}{\int\limits_0^\epsilon \left(\sqrt{\epsilon^2 - \delta^{*2}}\right)^{n-1} \mathrm{d}\delta^*}. \tag{4}$$

The distance 2δ between the hyperplanes is given by the coefficients $a_i := \frac{\partial f(\mathbf{x})}{\partial x_i}$ of the linear map $g(\mathbf{x})$: $\delta = \frac{\epsilon}{\sqrt{\sum_{i=1}^{n} a_i^2}}$. Using Fubini's theorem the volumes in (4) can be calculated as integrals over $(n$-1)-dimensional balls.

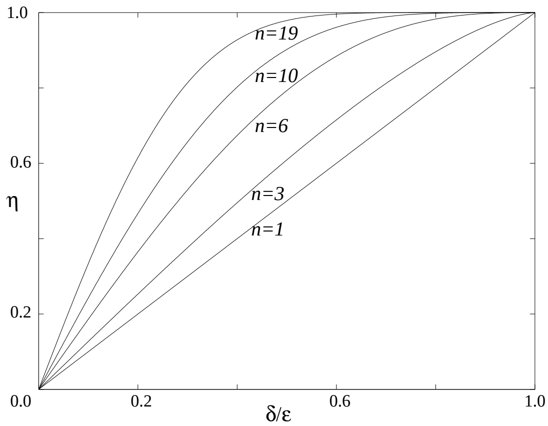


FIGURE 2. η is a function of δ/ϵ and depends on the dimension n

In Fig. 2 η is shown as function of δ/ϵ. It is calculated for five different dimensions n from one to nineteen. One can see that η increases with higher dimension. One can finally define an adapted conditional probability by

$$P^*_\epsilon(y|\mathbf{x}) = \begin{cases} P_\epsilon(y|\mathbf{x}) & : \quad \delta > \epsilon \\ \max\left(\frac{1}{\eta^{(n-1)}(\mathbf{x},y)}, \frac{1}{\eta^{(n)}(\mathbf{x},y)}\right) P_\epsilon(y|\mathbf{x}) & : \quad \text{else} \end{cases} \tag{5}$$

The maximium provides that $P^*_\epsilon(y|\mathbf{x})$ does not increase by adding redundant information.

With these considerations one is able to compute the conditional entropy from time series $\{x_s\}_{s=1}^N$ and $\{y_s\}_{s=1}^N$. First one has to estimate the probability $P_\epsilon(y|\mathbf{x})$ by counting all pairs $(\mathbf{x}_t, y_t)$ with $\mathbf{x}_t \in B_\epsilon(\mathbf{x})$ and $y_t \in I_\epsilon(y)$ and dividing this number by the number of points in $\{\mathbf{x}_t : \mathbf{x}_t \in B_\epsilon(\mathbf{x})\}$. Then the linear map $g(\mathbf{x})$ is fitted using singular value decomposition. If no local independence between the x-components can be assumed, all $\mathbf{x}_t$ are projected to the principal axes before, and η is estimated directly by counting the points with $|g(\mathbf{x}_t)| \leq \epsilon$. Otherwise equation (4) can be used. Now the conditional probability (5) is computed for all times $s = 1, \ldots, N$. Using

$$H^*_\epsilon(Y|X) = -E(\log P^*_\epsilon(y|\mathbf{x})) \tag{6}$$

a conditional entropy can be calculated. The limit

$$\tilde{H}(Y|X) := \lim_{\epsilon \to 0} \lim_{N \to \infty} H^*_\epsilon(Y|X) \tag{7}$$

suffices condition (A), and (B) in the sense that $\tilde{H}(Y|X)$ vanishes if there is a differentiable function $y = f(x)$.

INTERDEPENDENCES IN TIME SERIES

A lot of methods (e.g.: [3,5]) have been proposed to determine optimal delay times for the reconstruction of nonlinear dynamics from scalar time series [6,7]. The idea of calculating autocorrelations or mutual information is to find delay coordinates that are as independent as possible. Our intention is a little bit different. We try to find those delay vectors carrying the largest amount of information about future states. Therefor we calculate the conditional entropies

$$H_t^*(\Delta\tau) := H_\epsilon^*(X_{T+t}|(X_T, X_{T-\Delta\tau})) \tag{8}$$

for some prediction time t. X_T denotes the present state, X_{T+t} and $X_{T-\Delta\tau}$ are states in the future, resp. past. By averaging (8) over some future times t we determine the delay time $\Delta\tau_1$ which minimizes $H_{av}^*(\Delta\tau)$. The delay vectors $(X_T, X_{T-\Delta\tau_1})$ carry the most information about the dynamics. At smaller $\Delta\tau$ most of the information of $X_{T-\Delta\tau}$ is already contained in X_T, for later times one loses information about the future.

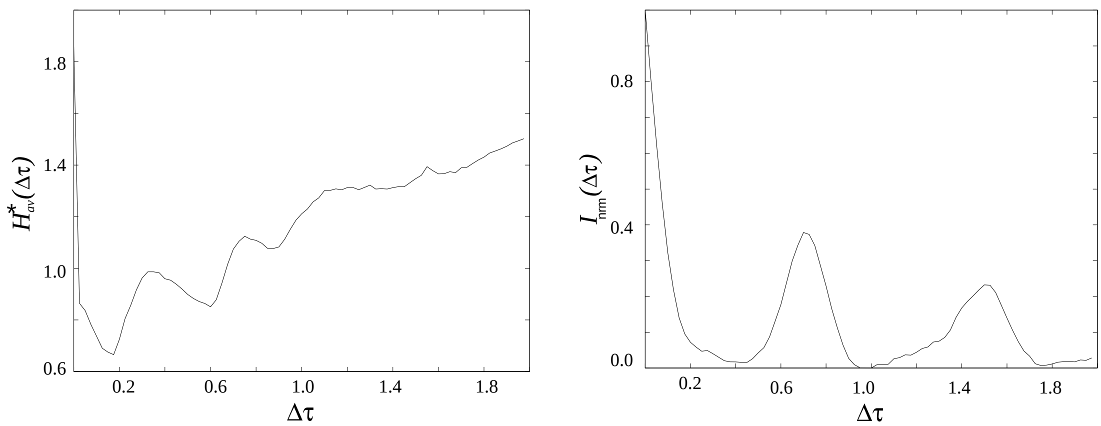


FIGURE 3. Conditional entropy (left) and mutual information (right) functions for the Lorenz system

For the well known Lorenz system

$$\dot{x} = 10(y - x), \quad \dot{y} = 28x - y - xz \quad \dot{z} - \frac{8}{3}z + xy$$

both mutual information and conditional entropy were calculated for delay times $\Delta\tau = 0 \dots 2$ (Fig. 3). One can see easily that both curves are not minimal at the same time. The minimum of $H_{av}^*(\Delta\tau)$ is just before the beginning of overfolding of the tracjectories at $\Delta\tau_1 \simeq 0.2$, while the maximal independence, described by the first minimum of the mutual information function is reached some time later ($\Delta\tau \simeq 0.45$). This means, that independence between the delay coordinates does not guarantee that a system's dynamics is optimally described.

Finally we analyze the Lorenz driven Duffing oscillator:

$$\dot{u} = v, \quad \dot{v} = 1.4x - 0.2v - u - u^3,$$

where x ist the x-component of a retarded version of the Lorenz System, whose components of the vector field are multiplicated by 0.2 to adjust the time scales. The projection to the u-v-plane is shown in Fig. 4 (left).

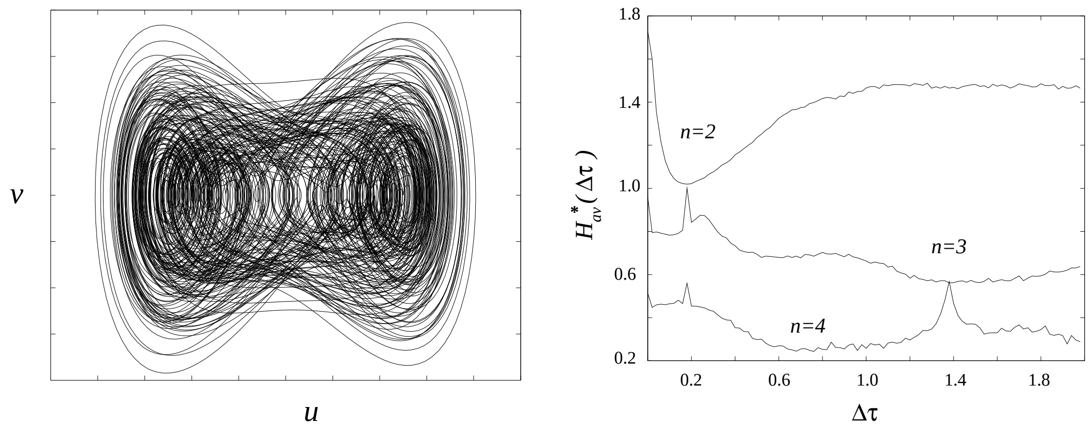


FIGURE 4. Left: The Lorenz driven Duffing oscillator (u-v-plane). Right: Conditional entropy $H_{av,n}(\Delta\tau)$ for the u-component at embedding dimensions $n = 2 \ldots 4$

The best delay time for the u-component was estimated in the same way as before. We found a delay time $\Delta\tau_1 \simeq 0.2$ (Fig. 4, right, upper curve). Usually this time is kept constant and delay vectors

$$\mathbf{x}(T) = (x(T), x(T - \Delta\tau_1), x(T - 2\Delta\tau_1) \ldots, x(T - (n-1)\Delta\tau_1)$$

are constructed. But is that the best choice? We kept $\Delta\tau_1$ and computed the conditional entropy

$$H^*_{av,n}(\Delta\tau) := \Big\langle H^*_\epsilon(X_{T+t}|(X_T, X_{T-\Delta\tau_1}, X_{T-\Delta\tau})) \Big\rangle_t .$$

Obviously the found minimum is not $\Delta\tau_2 = 2\Delta\tau_1$ (Fig. 4, right, middle curve), but $\Delta\tau_2 \simeq 1.4$. One can see that adding the same information again does not reduce the conditional entropy (at $\Delta\tau = \Delta\tau_1$).

Keeping the best delay time, the method can be iterated (Fig. 4, right, lower curve). If the conditional entropy is not decreased anymore by adding further time delayed components, one stops. This way one can determine a sufficient embedding dimension.

Another advantage is the possibility of using different observables of one dynamical system. The algorithm is the same as above, but the conditional entropies $H_{av}(\Delta\tau)$ have to be calculated not only for the observable X itself but for other ones Y or Z, too e.g. $H^*_\epsilon(X_{T+t}|(X_T, Y_{T-\Delta\tau_1}, Z_{T-\Delta\tau}))$.

CONCLUSIONS

We have presented a new method for the detection and quantification of interdepences in time series by calculating a modified conditional entropy. Especially the problem of estimating conditional probabilities in continuous systems has been discussed. We can detect not only relationships between different observables, but also temporal interdependences in scalar or multivariate time series. Optimal parameters for the time delay reconstruction of nonlinear dynamics can be found this way. As the conditional entropy is not symmetric, we get information about the direction of dependence.

REFERENCES

1. Li W., *J. Stat. Phys.* **60**, 5/6, 823 (1990).
2. Pompe B., *J. Stat. Phys.* **73**, 587 (1993).
3. Fraser A.M., and Swinney H.L., *Phys. Rev. A* **33**, 1134 (1986).
4. Gray, R.M., *Entropy and Information Theory*, New York, Springer Verlag, 1990.
5. Liebert, W., Pawelzik, K., Schuster, H.G., *Europhys. Lett.* **14**, 521 (1991).
6. Sauer, T., Yorke, J.A., Casdagli, M., *J. Stat. Phys.* **65**, 579 (1991).
7. Takens, F., in: *Dynamical Systems and Turbulence*, Warwick, Lecture Notes in Mathematics, 898, Springer Verlag, 1980, pp. 230.