Assessing different norms in nonlinear analysis of noisy time series

D. Kugiumtzis

Department of Informatics, University of Oslo, PO Box 1080 Blindern, N-0316 Oslo, Norway

Received 3 June 1996; revised 10 October 1996; accepted 12 November 1996

Communicated by A.M. Albano

Abstract

Many methods dealing with the analysis of multivariate data involve computations of point interdistances in state space. In time series analysis, the points are reconstructed from scalar data, most commonly using the method of delay coordinates. While, theoretically, all norms are considered equivalent in measuring interdistances, they perform differently in the presence of noise. The statistical description of three of the most popular norms, $L_1$, $L_2$, and $L_{\infty}$, revealed certain shortcomings of each, depending on the corresponding noise-free interdistances. For chaotic time series, the effect of noise on the different measures also varies with the reconstruction. Estimating the correlation dimension for simulated noisy data using the three norms confirmed the statistical analysis. Generally, the $L_2$ norm turns out to be the most robust over a range of time series types and reconstructions, whereas the $L_1$ and $L_{\infty}$ norms seem to deteriorate markedly for specific types of systems or reconstructions. Furthermore, the investigation gave some more insight into the role of the reconstruction parameters, showing in particular, the importance of the time window length $\tau_w$.

1. Introduction

Most of the methods of nonlinear time series analysis (e.g. estimation of invariants and prediction) operate with the interdistance vectors of reconstructed points in a pseudo-state space [1-3]. Traditionally, the metric used to compute the interdistance $^1$ is chosen either as the standard Euclidean norm, $L_2$, or as the maximum norm, $L_{\infty}$, to enhance the computation time. A third common norm, the “taxicab norm”, $L_1$, is another possibility hitherto little used in the time series context [4]. These three norms have different properties that are well studied in statistical settings. For a given data set, the $L_1$ norm is related to the median, the $L_2$ norm to the mean and the $L_{\infty}$ norm to the midrange of the data. For a goodness of fit problem, different criteria lead to different norms; thus minimizing the average absolute error is a $L_1$ problem, minimizing the average squared error is a $L_2$ problem, and minimizing the maximum error is a $L_{\infty}$ problem [5].

For infinitely many noise-free data, all three metrics are equivalent, while in applications with finite noisy data sets, they can give different results. In the setting of chaotic time series, this point has received little attention. We are aware of only two papers on the evaluation of norms in nonlinear analysis of time series [6,7], but to the best

---

$^1$ We use the term interdistance vector for the difference vector between two points and we refer to the length of this vector computed with some norm simply as interdistance.
of our knowledge nothing has been published on the performance of the norms under noisy conditions. Here, an attempt is made to find the most robust metric to be used with nonlinear methods (operating with interdistance vectors) when the time series is noisy.

Since probability density functions are hard to derive in all cases, a statistical approach emphasizing first and second order moments is followed. The performances of the norms are discussed here with respect to noisy trajectories reconstructed from chaotic time series. The theoretical findings are tested using estimation of the correlation dimension.

In Section 2, we establish statistical descriptions of the three norms, and in Section 3 the norms are evaluated using correlation dimension estimation of deterministic time series corrupted by white noise. Then in Section 4 the results from the simulations are discussed followed by some remarks on the impact of different norms on other nonlinear methods.

2. Interdistances of noisy data measured with different norms

The statistical analysis is done with regard to reconstructed time series but the results are valid, in general, for noisy multidimensional data corrupted by white noise.

2.1. Notation and definitions

The noise-free scalar time series of length $N$ is denoted $x_i = x(i \tau_s)$, $i = 1, \ldots, N$, where $\tau_s$ is the sampling time, set to 1 for maps. The noisy counterpart is $y_i = x_i + v_i$, $i = 1, \ldots, N$, where $v_i$ is the noise element. The noise is assumed to be white and normally distributed with 0 mean value and variance $\sigma^2$, i.e. $v_i \sim N(0, \sigma^2)$, but the findings below hold qualitatively for any distribution symmetric about 0.

The noise-free time series is assumed to originate from a deterministic process, either continuous or discrete. Then the noisy time series is the outcome of a univariate stochastic process $y_i = x_i + v_i$, with $y_i \sim N(x_i, \sigma^2)$.

The standard method of delays (MOD) is used to represent the data in $\mathbb{R}^m$, that is the reconstructed vectors are $x_i = [x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}]^T$ for $i = 1, \ldots, N - (m - 1)\tau$, where $m$ is the so-called embedding dimension and $\tau$ is the delay time. The interdistance vector of two reconstructed points $x_i$ and $x_j$ is

$$\delta x_{ij} = [x_j - x_i, \ldots, x_j + (m-1)\tau - x_i + (m-1)\tau]^T = [\delta x_{ij,1}, \ldots, \delta x_{ij,m}]^T.$$ 

The interdistance vector of the corresponding noisy points $y_i$ and $y_j$ is

$$\delta y_{ij} = [y_j - y_i, \ldots, y_j + (m-1)\tau - y_i + (m-1)\tau]^T = [\delta y_{ij,1}, \ldots, \delta y_{ij,m}]^T,$$

where $\delta y_{ij,k} = \delta x_{ij,k} + \delta v_{ij,k}$, for $k = 1, \ldots, m$ and $\delta v_{ij,k} = v_{j+k-1} - v_{i+k-1}$. The above notation is simplified by dropping the indices $i$ and $j$ whenever the two points are not specified explicitly. For the componentwise differences one has $\delta v_k \sim N(0, 2\sigma^2)$ and $\delta y_k \sim N(\delta x_k, 2\sigma^2)$.

The three norms that are used to compute the interdistance of two noise-free points can be considered as scalar functions of the interdistance vector: $L_1(\delta x) = \sum_{k=1}^m |\delta x_k|$, $L_2(\delta x) = (\sum_{k=1}^m \delta x_k^2)^{1/2}$, and $L_\infty(\delta x) = \max_{k=1,\ldots,m} |\delta x_k|$. The corresponding norms for the noisy points are scalar functions of $m$ stochastic variables, the $m$ components of $\delta y$. In the following, we refer to them as “noisy norms”. The objective is to find the distribution of the noisy norms with emphasis on the first and second moments. Then the most robust noisy norm $L(\delta y)$ is the one being less biased and having the least variance. For comparison across dimensions, the norms are scaled, so that $L_1(\delta x) = (1/m) \sum_{k=1}^m |\delta x_k|$ and $L_2(\delta x) = ((1/m) \sum_{k=1}^m \delta x_k^2)^{1/2}$. 
2.2. Statistics of the norms

The stochastic variable \( z_k = |\delta y_k| \) follows the folded normal distribution \([8]\) and has expectation

\[
E\{z_k\} = \frac{2\sigma}{\sqrt{\pi}} e^{-\delta x_k^2/4\sigma^2} + \delta x_k \text{erf}\left(\frac{\delta x_k}{2\sigma}\right),
\]

where \( \text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_0^x e^{-y^2/2} \, dy \). The variance is given in terms of the first and second moment as

\[
\text{Var}\{z_k\} = E\{z_k^2\} - E\{z_k\}^2,
\]

where the second moment is

\[
E\{z_k^2\} = 3\sigma^2 + 2\sigma^2.
\]

In Appendix A a detailed analysis of the statistics of \( z_k \) is made for any noise with a distribution symmetric about 0.

It can be shown that \( E\{z_k\} \geq |\delta x_k| \) (see Appendix A) and consequently \( \text{Var}\{z_k\} \leq 2\sigma^2 \) for all \( |\delta x_k| \). Moreover, as \( |\delta x_k| \) increases, \( E\{z_k\} \) converges exponentially to \( |\delta x_k| \) (the first term on the right-hand side of Eq. (1) goes exponentially to 0 and the second exponentially to \( |\delta x_k| \)). It can thus be concluded that, for \( |\delta x_k| \) larger than the noise amplitude (here \( \sqrt{2}\sigma \)), \( E\{z_k\} \approx |\delta x_k| \) and therefore the bias occurs only for “small” values of \( |\delta x_k| \).

Furthermore, \( \text{Var}\{z_k\} \approx \text{Var}\{\delta y_k\} \) above the noise amplitude.

In the following derivation of the moments of the noisy norms all the stochastic variables \( z_k, \ k = 1, \ldots, m \) are independent, because the noise is assumed white. They are also non-identically distributed, because each \( z_k \) follows the same distribution but with a different mean, which makes the statistics of the norms more complicated.

**L₁ norm.** The noisy \( L_1 \) norm is the average of the \( m \) stochastic variables \( z_k = |\delta y_k|, \ k = 1, \ldots, m \), i.e.

\[
L_1(\delta y) = \frac{1}{m} \sum_{k=1}^m z_k.
\]

Each \( z_k \) has a different probability density function (pdf) \( f_{z_k} \) and the pdf of \( L_1 \) is defined as the convolution of all the \( m \) pdf’s, which for \( m > 2 \) has a rather complicated form. However, the mean and variance of \( L_1 \) are given simply as

\[
E\{L_1(\delta y)\} = \frac{1}{m} \sum_{k=1}^m E\{z_k\}
\]

and

\[
\text{Var}\{L_1(\delta y)\} = \frac{1}{m^2} \sum_{k=1}^m \text{Var}\{z_k\}.
\]

According to the properties of the mean of \( z_k \), when all \( |\delta x_k| \) are greater than the noise amplitude then \( E\{L_1(\delta y)\} \approx L_1(\delta x) \), while for each \( |\delta x_k| \) less than the noise amplitude, \( E\{L_1(\delta y)\} \) deviates from \( L_1(\delta x) \) with a magnitude \((1/m)(E\{z_k\} - |\delta x_k|)\). Thus the noisy \( L_1 \) is a biased estimate of the noise-free \( L_1 \) whenever some differences \( \delta x_k \) fall below the noise amplitude and the bias effect gets larger as more differences get closer to 0. Geometrically, the bias occurs when the interdistance vector \( \delta x \) is close to one of the \( m \) coordinate axes, or to a hyperplane formed by some coordinate axes. The variance of \( L_1 \) is quite stable with a maximum value \( 2\sigma^2/m \) when all \( |\delta x_k| \) are greater than the noise amplitude. Note that the variance decreases with increasing \( m \).

**L₂ norm.** A different approach is followed to find the moments of the noisy \( L_2 \) norm, \( L_2(\delta y) = ((1/m) \sum_{k=1}^m \delta y_k^2)^{1/2}, \) described in Appendix B.

The expectation of \( L_2(\delta y) \) is found approximately as

\[
E\{L_2(\delta y)\} \approx \sqrt{w_x + 2\sigma^2} - \frac{\sigma^2(w_x + \sigma^2)}{m(w_x + 2\sigma^2)^{3/2}}.
\]
and the variance

\[ \text{Var}[L_2(\delta y)] = \frac{2\sigma^2 (w_x + \sigma^2)}{m(w_x + 2\sigma^2)}, \]  

where \( w_x = (1/m) \sum_{k=1}^m \delta x_k^2 \).

Both the expectation and variance of the noisy \( L_2 \) norm do not depend on the components of \( \delta x \) explicitly but on the average of their squares \( w_x \), i.e. the first and second order moments of \( L_2(\delta y) \) have a so-called circular dependency on the vector \( \delta x \). Contrary to the noisy \( L_1 \), the noisy \( L_2 \) is not sensitive to any direction in state space. The expectation is always less than \( \sqrt{w_x + 2\sigma^2} \) because the second term of the right-hand side of Eq. (4) is always positive. In fact, the effect of the second term on the magnitude of \( E\{L_2(\delta y)\} \) is rather limited because it is much less than the first term, especially when \( m \) or \( \sqrt{w_x} \) increases. This gives approximately \( E\{L_2(\delta y)\} \approx \sqrt{w_x + 2\sigma^2} \) which

![Fig. 1. Expectation and standard deviation of \( L_\infty(\delta y) \) when normal noise with \( \sigma^2 = 1 \) is added to the components of \( \delta x = [\delta x_1, \delta x_2, \delta x_3] \). Here, the statistics are computed numerically from 10 000 realizations of \( \delta y \) adding normal noise to \( \delta x \). For each plot, \( |\delta x_2| \) and \( |\delta x_3| \) are kept constant, and \( E(L_\infty(\delta y)) \) and \( \pm \sqrt{\text{Var}(L_\infty(\delta y))} \) are computed for varying \( |\delta x_1| \): (a) \( |\delta x_2| = 0.5, |\delta x_3| = 0.5 \), (b) \( |\delta x_2| = 0.5, |\delta x_3| = 5 \), and (c) \( |\delta x_2| = 5, |\delta x_3| = 5 \). They are displayed, together with \( L_\infty(\delta x) \) for comparison, with lines according to the legend.](https://example.com/fig1.png)
Table 1
Bias and variance of the three noisy norms according to different values of the components of $\delta x$, increasing relative to the noise amplitude $2\sqrt{2\sigma}$. (The components of $|\delta x|$ are indexed in ascending order to facilitate the comparison.)

| $|\delta x_1| \leq \cdots \leq |\delta x_m|$ | $L_1$ | $L_2$ | $L_\infty$ |
|---|---|---|---|
| (1) $|\delta x_k| \leq 2\sqrt{2\sigma}$, $k = 1, \ldots, m$ | Bias is | Bias is about | Bias is about |
| | $\frac{1}{m} \sum_{k=1}^{m} (E(z_k) - |\delta x_k|) > 0$, $\text{Var}(L_1(\delta y)) < \frac{1}{m} 2\sigma^2$ | $\sqrt{w_x + 2\sigma^2} - \sqrt{w_x} > 0$, $\text{Var}(L_2(\delta y)) < \frac{1}{m} 2\sigma^2$ | $2\sqrt{2\sigma} - L_\infty(\delta x) > 0$, $\text{Var}(L_\infty(\delta y)) < 2\sigma^2$ |
| (2) $|\delta x_k| \leq 2\sqrt{2\sigma}$, $k = 1, \ldots, r$ | Bias is | Same as above | No bias and |
| | $\frac{1}{m} \sum_{k=1}^{m} (E(z_k) - |\delta x_k|) > 0$, $\text{Var}(L_1(\delta y)) < \frac{1}{m} 2\sigma^2$ | Same as above | $\text{Var}(L_\infty(\delta y)) = 2\sigma^2$ |
| $|\delta x_k| > 2\sqrt{2\sigma}$, $k = r + 1, \ldots, m$ | No bias | Same as above | Bias increases as |
| | $\text{Var}(L_1(\delta y)) = \frac{1}{m} 2\sigma^2$ | Same as above (for (2)) | $|\delta x_m| - |\delta x_{m-1}| \leq 2\sqrt{2\sigma}$ |

indicates a systematic bias, $E[L_2(\delta y)] - L_2(\delta x) \simeq \sqrt{w_x + 2\sigma^2} - \sqrt{w_x}$. The bias decreases with the increasing $\sqrt{w_x}$. The variance of the noisy $L_2$ norm decreases with $m$ as for $L_1$, as indicated from Eq. (5). Moreover, Eq. (5) shows that the variance is always less than $2\sigma^2/m$.

$L_\infty$ norm. The moments of the noisy $L_\infty$ norm defined as $L_\infty(\delta y) = \max_{k=1, \ldots, m} |z_k| = \max_{k=1, \ldots, m} |\delta y_k|$ cannot be determined analytically for $m > 2$ as discussed in Appendix C and therefore $E[L_\infty(\delta y)]$ and $\text{Var}(L_\infty(\delta y))$ have to be computed numerically. In Fig. 1, the expectation and standard deviation of the noisy $L_\infty$ norm are shown for different values of the components of $|\delta x|$ and $m = 3$. The bias of $E[L_\infty(\delta y)]$ depends on the difference between the largest components of $|\delta x|$ rather than on their magnitude. Bias occurs whenever the largest components are “close”, in the sense that their difference is less than the noise amplitude, and increases as the difference goes to zero. Geometrically, the condition for bias implies that $\delta x$ is close either to the origin (see Fig. 1(a)) or to the diagonal or antidiagonal of some $k$-dimensional plane if the $k$ largest components of $|\delta x|$ are clustered within the noise amplitude (Figs. 1(b) and (c)). The variance of $L_\infty(\delta y)$ is at the level of $\text{Var}(z_k) = 2\sigma^2$ for any $|\delta x|$.

The three noisy norms have different properties with regard to different values of the components of the noise-free vector $|\delta x|$ and in Table 1, we summarize the main differences. The norms deteriorate when all components are smaller than the noise amplitude (case (1) in Table 1). The bias for $L_2$ decreases as the radius $\sqrt{w_x}$ increases. The $L_1$ norm is biased when some components are small (compared to the noise amplitude), and the $L_\infty$ norm is biased when the largest components are close.

3. Estimation of the correlation dimension using the three norms

3.1. The correlation dimension

In nonlinear time series analysis, the correlation dimension is well suited to test the different norms because it is based on computing point interdistances and comparing them with some given distance $r$ [9]. The starting point is the probability $P(L(\delta x) \leq r)$, where $L$ stands for some norm. This probability is estimated by the so-called correlation integral

$$C(r) = \frac{2}{N(N-1)} \sum_{i,j>i}^{N} \Theta(r - L(\delta x_{i,j}))$$
where $\Theta(x)$ is the Heaviside function, defined as $\Theta(x) = 1$ for $x \geq 0$ and $\Theta(x) = 0$ for $x < 0$ and $K$ is the time correlation [10]. Then the correlation dimension $v$ is computed by fitting the scaling law $C(r) \propto r^v$ over a range of $r$-values. There are different ways to evaluate the scaling law from the correlation integral [11]. In the computations below we chose to compute the local slope of the graph of $\log C(r)$ vs. $\log r$ for each $r$ using a smoothing procedure. The estimate of $v$ was computed from the average slope in an interval $[r_1, r_2]$ of length defined by $r_2/r_1 = 4$ showing the least variance. When the data are corrupted by noise the scaling interval is often observed for small $r$ with $r_1$ close to the noise amplitude. Thus the reduction of the noise effect in the computations using a proper norm can give better estimates of $v$. However, the focus in this work is not on the reliability of the estimate of $v$ itself, but on the variations of the estimates with the use of different norms.

3.2. Bias in the correlation integral

The correlation integral $C(r)$ for noisy time series is computed by averaging a large number of noisy interdistances (compared to some length $r$). Due to this averaging the bias of the noisy interdistances, particularly those close to $r$, generate bias in the computed $C(r)$. Moreover, the variance of the correlation integral $C(r)$ is considered insignificant due to the averaging as well as the limited variance of the noisy norms themselves, as discussed in Section 2. Thus, the focus is on the bias.

In order to estimate the bias of $C(r)$ due to noise, we will identify the points with interdistances shorter than $r$ when they are noise-free, $L(\delta x) \leq r$, and larger than $r$ on the average, when noise is added, $E[L(\delta y)] > r$. (The opposite is unlikely according to the statistical analysis in Section 2). The distribution of such “sensitive” points is different for the three norms, as shown in Fig. 2. In Figs. 2(b)-(d), the two-dimensional areas of points with distances from a center point smaller than $r$ and $r$ are drawn. The darker regions show the sensitive areas, and points in them are expected to jump out of the $r$ or $r$ limit with the addition of noise. The three figures give insight into the change of the bias of the norms with the distance $r$. However, it is hard to find an analytic form for the bias of $C(r)$ for each norm as it is dependent on the distribution of the set of the noise-free points in the state space. In some cases, the bias can be estimated qualitatively according to the given point distribution. For example, the points displayed in Fig. 2 with dots and crosses are reconstructed with MOD, $m = 2$ and $\tau = 5$, from noise-free measurements of the $x$ variable of the Henon map [12]. Note that the points appear to be distributed along vertical stripes as shown in Fig. 2(a) where the reconstructed attractor is displayed. Therefore, for this data set, more points are expected to lie in the sensitive region for the $L_1$ norm (see Fig. 2(b)) and less in the sensitive region of the $L_\infty$ norm (see Fig. 2(d)). On the other hand, the sensitive region of the $L_2$ norm seems to be less dependent on the point distribution.

3.3. Different systems and reconstructions

The distribution of points in state space is determined by the underlying system that generated the sampled trajectory (the observed scalar data) and the reconstruction parameters. When using MOD, the components of each reconstructed point are actually samples from a segment of the time series and the components of an interdistance vector $\delta x$ are differences between samples of two such segments of the time series.

In the following we discuss the performance of the norms using estimation of $v$ based on time series from different types of systems. In the results shown for the noisy data, only one realization is utilized. Due to the small variance of $C(r)$ the results from one realization should be representative. However, in order to verify this, we have estimated

\[2\] To compute the slope for each $r$ we used the best fit slope for three distance values, the current $r$, the previous and the next.
Fig. 2. (a) The Henon data [12] reconstructed with MOD, $m = 2$ and $r = 5$. The displayed rhombus, circle and square indicate the equidistant points from a center point with a distance $r$ measured with the three norms. (b) The points with distances from the center point given by $L_1 (dx) \leq \frac{1}{2} r$ and $L_1 (dx) \leq r$ are shown with the light area inside the two rhombus. The darker regions correspond to points with $E\{L_1 (dy)\} > \frac{1}{2} r$ and $E\{L_1 (dy)\} > r$, respectively, after normal noise 5% of the standard deviation of the data was added to the points. The reconstructed points that fall in these regions are denoted with crosses. (c) Same as in (b) but for the $L_2$ norm. (d) Same as in (b) but for the $L_\infty$ norm. The expectation of the noisy norms in (b) and (c) is computed from Eqs. (3) and (4) but for (d) is approximated numerically.

$v$ also for 40 realizations of the noisy data and no significant difference was found (the difference was of the order 0.01).

3.3.1. Regular deterministic systems

Regular deterministic systems generate periodic or quasiperiodic trajectories in the original state space, that evolve similarly, i.e. the initial discrepancy of two trajectories remains unchanged under their evolution. The same will generally be true for their sampled projections (at least for parts of them). Thus the samples of two time series segments that start from two close samples $x_i$ and $x_j$ will differ by the same amount $|x_i - x_j|$. Therefore, many
interdistance vectors $\delta x$ are expected to have similar components which gives rise to bias in the computations with noisy data points when the $L_\infty$ norm is used, especially when the embedding dimension $m$ is large.

In Fig. 3 we show the estimation of the correlation dimension using the three norms for data from a 2-torus. The estimates are computed for reconstructions with fixed $\tau = 5\tau_s$ and increasing $m$. Other selections of $\tau$ gave similar results. The error bars in the figure show the standard deviation of the local slope in the selected interval $[r_1, r_2]$. (In the search for the interval giving best scaling, large values of $r$ were excluded because they gave an incorrect scaling of about $v \approx 1.4$, a phenomenon often referred to as “the knee problem” (see [13,14]), which is not related to noise but to the intrinsic dynamics.) The correct scaling appears for very small values of $r$ indicating that small regions of this size have to be well populated in order to achieve scaling and therefore we used as many as 50 000 data. For the noise-free data, all three norms give the same slight overestimation of the correct $v$ (plateau at about 2.1) and for a large range of $m$ (Fig. 3(a)). The scaling is partly masked by the addition of normal white noise with a standard deviation equal to 2% of the standard deviation of the data. The estimates of $v$ computed with $L_1$ and $L_2$ seem to be little affected by noise (the plateau rises to $\approx 2.25$) while the estimates with the $L_\infty$ norm increase constantly with $m$, indicating a large noise effect.

### 3.3.2. Chaotic systems

For chaotic systems, the presence of a positive Lyapunov exponent causes trajectories to diverge and converge in state space. Consequently, the differences of their projections, which are actually the components $\delta x_k$, vary also from small to large values, according to the length of the time series segment, which is determined from the parameters of the reconstruction.

**Maps.** For chaotic time series from discrete systems, the reconstruction is determined solely by $m$ as $\tau$ is usually set to 1. For such systems, the divergence–convergence phenomenon is pronounced within few iterations, so that successive $\delta x_k$ components tend to differ significantly, which is in favor of the $L_\infty$ norm. For larger $m$, the attractor is often stretched parallel to some coordinate axes (see Fig. 2(a)), and for such a structure the bias of the noisy $L_1$ norm is amplified.
Fig. 4. The slope of $\log C(r)$ vs. $\log r$ is computed for noisy time series from the Henon map of three different lengths as shown in the legend of each plot. The results in (a) are for the $L_1$ norm, in (b) for the $L_2$ norm, and in (c) for the $L_\infty$ norm. The reconstruction was made with $r = 1$ and $m = 7$ and the noise level was 2%.

The effect of noise is independent of the length of the time series as long as there are enough data to support the statistics for $r$-scales larger than the noise amplitude. For example, for the Henon data, it seems that the slope of $\log C(r)$ vs. $\log r$ stabilizes for $r$ larger than the noise amplitude after some thousands of samples are given at a 2% noise level. In Fig. 4, slopes for different time series lengths are shown for reconstructions with $m = 7$. Obviously, the slope computed from few data differs distinctly from the slope computed with some or many thousands of points (compare the slopes for $N = 500$ with those for $N = 4000$ and $N = 50000$ in Fig. 4). Moreover, the curves for $N = 4000$ and $N = 50000$ in Figs. 4(a)–(c), show that the slopes computed with the $L_\infty$ norm are closer to the correct plateau while the slopes computed with the $L_1$ norm seem to form the highest plateau above the correct value.

The common feature of the noisy slope for large enough $m$ is that it converges from above to the noise-free slope for $r$ scales larger than the noise amplitude. However, the convergence may vary according to the different bias effect using different norms. This is shown in Fig. 5 for the Henon data corrupted with noise of different levels and
reconstructed with $m = 7$. For the $L_1$ norm in Fig. 5(a), the convergence to the noise-free slope is very slow. For the 0.5% noise level the noisy and noise-free slope coincide at a stage where the scaling is no longer maintained giving significant overestimation of $\nu$. For the $L_2$ norm in Fig. 5(b), the convergence is a bit faster giving better estimation of $\nu$, e.g. for the 0.5% noise level the noisy slope lies longer on the correct plateau. However, the best results are obtained with $L_\infty$ in Fig. 5(c), where the convergence is the fastest. Even for 2% noise level, the noisy slope converges to the noise-free slope when it still gives good scaling. It seems that the addition of 5% noise completely masks the scaling and for this noise level there is significant overestimation of $\nu$ independently of the chosen norm. The number of data required to establish scaling or the level of noise masking completely the scaling may vary for different maps.

However, the general features of the estimation of $\nu$ with different norms were found to be similar for the maps we tested including the logistic, Henon, and Ikeda map. Representative results for the estimated $\nu$ are shown for
the Henon data in Fig. 6. The estimates are computed for reconstructions of increasing $m$ with $\tau = 1$. The nice plateau obtained for the noise-free data in Fig. 6(a) disappears when 2% normal white noise is added to the data. As shown in Fig. 6(b), all estimates of $\nu$ increase with $m$, apart from $m = 2, 3$, but the $L_1$ norm gives the most biased estimates while the $L_\infty$ gives the closest estimates to those of the noise-free data, and $L_2$ being the intermediate case.

**Flows.** In applications with continuous processes, the time series are often sampled densely ($\tau_s$ is small). The reconstruction from such time series depends on both $m$ and $\tau$. If $\tau$ is small, the successive components of $\delta x$ will not differ significantly due to continuity. Then also the largest components will be close, regardless of the value of $m$. For such reconstructions, $L_\infty$ is expected to give biased results. Using a larger $\tau$, reduces this shortcoming of the $L_\infty$ norm. The parameters $\tau$ and $m$ both affect $L_2$ and $L_1$ but in a way that is best explained by the combined parameter $\tau_w = (m - 1)\tau$, the time window length. For constant $\tau_w$, $m$ and $\tau$ may vary without significantly affecting the results. Apart from the differences between the norms for reconstructions with small $\tau$, other systematic differences do not appear when estimating $\nu$ and the results vary according to particular characteristics of the attractor.

We used here the Lorenz data [15] for comparison. In Fig. 7, the estimates of $\nu$ are plotted against the time window length $\tau_w$ for noise-free and noisy Lorenz data. We use $\tau_w$ instead of $m$ to stress the role of $\tau_w$ for reconstruction and we estimate $\nu$ for two different combinations of $m$ and $\tau$ that give the same $\tau_w$ (fixed $\tau = 10$ and $m_1 = 2, \ldots, 10$, and fixed $\tau = 2$ and $m_2 = 2, \ldots, 100$). For noise-free data, all norms give approximately the same good estimates with small variance for a large range of $\tau_w$. For $L_1$ and $L_2$ this holds regardless of the combinations of $(m, \tau)$ but for $L_\infty$ it does not hold for large $\tau_w$ when $\tau$ is too small. For noisy data, the estimates increase for all norms, but in different ways. The increase is much larger for the $L_1$ norm as $\tau_w$ increases, as shown in Fig. 7(d). For the Lorenz data, as $\tau_w$ increases, the attractor expands from the diagonal and folds back in a way that the trajectories are formed quite in parallel to some coordinate axes (corresponding to the components of large delays), giving rise to more bias in the computation of $C(r)$ with $L_1$ norm. However, results from other systems showed better performance of the $L_1$ norm for noisy data, suggesting that the insufficiency of the $L_1$ norm here is basically due to the structure of
Fig. 7. The correlation dimension as a function of the time window length $\tau_w$ for the Lorenz data computed with different norms. In (a), (b), and (c), the estimates of $v$ are computed with $L_1$, $L_2$ and $L_\infty$ norm, respectively, for 4000 noise-free data sampled from the $x$-variable with $t_s = 0.02s$. In (d)-(f), the same results of the same computations are shown for the noisy data derived by adding 5% normal noise to the noise-free data. For each $\tau_w$, the estimates of $v$ are computed for two combinations of reconstruction parameters, one for $t = 2$ and another for $t = 10$, and the results in each plot are shown with a grey thin line and a black thick line, respectively. The error bars denote the standard deviation of the estimates of $v$. The correct $v \approx 2.06$ is shown with a horizontal thin line.

The trajectories of the Lorenz system.\textsuperscript{3} The same figure shows the independence of the $L_1$ norm to the particular combination of $m$ and $\tau$ because the formation of the attractor is related to $\tau_w$ and not to $m$ and $\tau$ explicitly. The same is valid for the $L_2$ norm. However, the circular dependence of the noisy $L_2$ norm on the interdistance vector gives rise to a systematic bias in the computation of $C(r)$ which seems to be the least biased of the three norms compared here. As shown in Fig. 7(e), the estimates of $v$ from the $L_2$ norm are not significantly increased after the addition of noise, and they have comparatively small variance over a large range of $\tau_w$. The estimates with the

\textsuperscript{3} This is supported from simulations with many other systems, as Rössler [16], Rabinovich–Fabrikant [17], Rössler hyperchaos [18] and Mackey–Glass with different delays [19].
$L_\infty$ norm are slightly worse for large $\tau$ and they are significantly deteriorated for small $\tau$. As $\tau_w$ increases, more interdistance vectors have the largest components closer and thus more noise is encountered by the $L_\infty$ measure. This effect is stronger for reconstructions with small $\tau$, where the $m$ is much larger for the same $\tau_w$.

4. Discussion

In applications of the correlation dimension to real data where the nature of the data often is not known, one would like to use the norm that shows robustness to different systems and reconstructions, and for such a purpose $L_2$ seems to be the most appropriate choice. However, if one deals with data assumed to be generated by chaotic maps, the $L_\infty$ would give better estimates of $v$ for large $m$. For data assumed to be generated by chaotic flows, the $L_1$ norm can also be used, especially if one wants to avoid using $L_2$ norm due to computational efficiency. It should be noted though that for some types of continuous systems (as the Lorenz system shown above) it performs worse than $L_2$.

The results in Fig. 7 with $L_1$ and $L_2$ confirm that, for a given $\tau_w$, any combination of $\tau$ and $m$ (above the fractal dimension limit for $m$) should give equivalent reconstructions at least for the estimation of the correlation dimension, as we argued in [14]. Certainly, a small $m$ is desired to expedite the computations, and this cannot be chosen arbitrarily as the fractal dimension is not known. When the $L_\infty$ norm was used, the claim about the independence of $\tau_w$ and the interdependence of $m$ and $\tau$ was not supported in the simulations, due to the shortcoming of this norm when applied to special reconstructions with very small $\tau$. But this seems to be a consequence of the use of $L_\infty$ and it turns out that the claim holds as long as the applied norm includes all the components of the vector, as is the case with $L_1$ and $L_2$. In that respect, it should hold for any other norm of this type, including generalized norms such as the Mahalanobis norm [20], or norms with exponential weighting of the components as the one implied by the reconstruction suggested by Farmer in [21] and tested for the prediction of noise-free data in [6]. This reconstruction as well as the reconstruction using Singular Value Decomposition (SVD) [22] are expected to be less norm-dependent due to the decreasing variance of the components of the interdistance vectors. Particularly, the first components, having the largest variance, vary significantly from the other components. Therefore we believe that for computations with reconstructions that weight the components, e.g. using SVD, the $L_\infty$ norm can be selected.
Fig. 8. Prediction error as a function of the embedding dimension $m$ in (a) and the prediction time $T$ in (b) for 4000 data from the Henon map. The three curves in each plot correspond to the three different norms used to locate the nearest neighbors in the local linear prediction and to compute their distances from the target point (see the legend). For each target point in the test set of the last 1000 samples, a local linear map is constructed from the 15 nearest neighbors in the training set of the first 3000 samples. This map is used to compute the predicted value for the target point. The prediction error is measured with the Normalized Root Mean Square Error (NRMSE) over all points in the test set (see [27] for details). The results in (a) are for one step ahead predictions ($T = 1$). In (b), local maps are constructed for each $T$ (direct prediction) and $m = 3$. The neighbors are not weighted, so that discrepancies in the neighbor distances due to the chosen norm be better pronounced.

In [23], MOD reconstructions with large $m$ and small $\tau$ were used together with the $L_\infty$ norm in a correction scheme for the estimation of $\nu$ from noisy data. We believe the efficiency of this correction may be limited because $C(r)$ is strongly biased due to the particular MOD parameter setting in combination with $L_\infty$. Improved correction of the estimated $C(r)$ from noisy data may be possible if the $L_2$ norm is considered instead [24].

The results on the performance of the three norms for noisy data assessed here using the correlation dimension are valid also for other methods that deal with scaling laws, such as the estimation of the $K_2$-entropy [25]. However, other nonlinear methods utilizing point interdistances may be less sensitive to the choice of the norm. For example, linear local prediction models [26,27] are not particularly norm-dependent because the computation of interdistances involves only the determination of the local neighborhoods. Simulations indicated that small discrepancies in the selection of neighbors due to the chosen norm do not affect significantly the predictive properties of the method. For illustration, the prediction error as a function of the embedding dimension and the prediction time is shown in Fig. 8 for the Henon data. The choice of the norm does not seem to affect the prediction error. Simulations with other types of data gave similar results.

5. Conclusion

The study of the statistical performance of the norms on interdistance vectors of points reconstructed from noisy chaotic time series showed that the theoretical equivalence of the norms does not hold when the points are corrupted with noise. Methods dealing with scaling laws, such as the estimation of the correlation dimension used here, are sensitive to the choice of the norm. Moreover, in the presence of noise, biased estimates are obtained due to the bias of the norm itself which may depend on the distribution of the points on the attractor. From the simulations with different systems and parameter settings, none of the norms showed distinctive superiority. However, it turns out
that the $L_2$ norm is the most consistent of the three norms tested and most robust to noise for different types of time series and reconstructions. More precisely, the $L_2$ norm does not depend on the distribution of interdistance vectors in state space, while the other two norms are sensitive in specific directions (along coordinate axes for $L_1$ and along diagonal and antidiagonal axes for $L_\infty$). However, for chaotic maps, the $L_2$ norm is outperformed by $L_\infty$. For data generated from continuous processes, the problem of finding the reconstruction parameters is simplified when $L_1$ or $L_2$ is used because only the time window length $\tau_w$ has to be determined. For such data, $L_\infty$ gives poor estimates when the delay time $\tau$ is small.

The selection of the norm in applications has gained little attention, and the norms are often used uncritically with concern only for the computational expense. In that respect, the choice of $L_\infty$, that seems to be the most popular norm in applications so far, is not always suitable. Hopefully, the results here will stimulate more research towards new metrics that may be more robust to noise.

Acknowledgements

This work has been supported by the Norwegian Research Council (NFR). The author is indebted to Nils Christophersen who followed and supported the work presented in this paper.

Appendix A. Statistics of $z_k = |\delta y_k|$

Let $f$ be the probability density function (pdf) and $F$ the cumulative density function (cdf) of the noise element $\delta v_k$, following a distribution symmetric about 0. Then the pdf of $\delta y_k = \delta x_k + \delta v_k$ is $f_{\delta y_k}(\delta y_k) = f(\delta y_k - \delta x_k)$ and the pdf for $z_k = |\delta y_k|$ is the pdf of the folded distribution of $\delta y_k$

$$f_{z_k}(z_k) = f_{\delta y_k}(z_k) + f_{\delta y_k}(-z_k) = f(z_k - \delta x_k) + f(z_k + \delta x_k), \quad z_k > 0.$$ Proceeding with the integral form of the expectation of $z_k$, it can be found

$$E\{z_k\} = 2 \int_{-\delta x_k}^{\infty} u f(u) \, du + 2\delta x_k F(\delta x_k) - \delta x_k. \quad (A.1)$$

Similarly, the second moment can be found

$$E\{z_k^2\} = \delta x_k^2 + \text{Var}\{\delta v_k\} \quad (A.2)$$

where $\text{Var}\{\delta v_k\}$ is the noise variance. The variance of $z_k$ can be found directly from Eqs. (A.1) and (A.2).

The explicit formula of $E\{z_k\}$ for normal noise in Eq. (1) is derived directly by substituting the normal pdf and cdf in Eq. (A.1). Further, note that

$$E\{z_k\} - \delta x_k = 2 \int_{-\delta x_k}^{\infty} u f(u) \, du - \delta x_k (1 - F(\delta x_k)) = 2 \int_{\delta x_k}^{\infty} (u - \delta x_k) f(u) \, du,$$

which establishes that $E\{z_k\} \geq |\delta x_k|$ when noise is symmetric about 0. Consequently, we have also that $\text{Var}\{z_k\} \leq \text{Var}\{\delta y_k\} = \text{Var}\{\delta v_k\}$. For uniform noise, $v_k \sim U[-a, a]$, the inequalities above become equalities when $|\delta x_k| > 2a$, the amplitude of noise. For normal noise, this holds approximately.
Appendix B. Statistics of $L_2(\delta y)$

The noise is assumed to be normal, thus $\delta x_k \sim N(\delta x_k, 2\sigma^2)$. First the distribution of $w_y = (1/m) \sum_{k=1}^{m} \delta x_k^2$ is found. According to [8] $w_y \sim (2\sigma^2/m) \chi^2_m(\lambda)$, i.e. $w_y$ follows the non-central distribution $\chi^2$ with $m$ degrees of freedom and non-centrality parameter $\lambda = (1/2\sigma^2) \sum_{k=1}^{m} \delta x_k^2$, multiplied by a factor $2\sigma^2/m$. The moments for $\chi^2$ are $E[\chi^2] = m + \lambda$ and $\text{Var}[\chi^2] = 2(m + 2\lambda)$, which for $w_y$ gives

$$E[w_y] = 2\sigma^2 + \frac{1}{m} \sum_{k=1}^{m} \delta x_k^2 = 2\sigma^2 + w_x$$

(B.1)

and

$$\text{Var}[w_y] = \frac{8\sigma^4}{m} + \frac{8\sigma^2}{m^2} \sum_{k=1}^{m} \delta x_k^2 = \frac{8\sigma^2}{m} (E[w_y] - \sigma^2).$$

(B.2)

where $w_x = (1/m) \sum_{k=1}^{m} \delta x_k^2$.

Based on the moments of $w_y$, the moments of $L_2$ can be found approximately. The noisy norm $L_2$ can be defined as a simple smooth function of $w_y$, $L_2 = g(w_y) = \sqrt{w_y}$. The expectation of $L_2$ can be then approximated as [28]

$$E[L_2(\delta y)] \approx g(E[w_y]) + g''(E[w_y])\frac{\text{Var}[w_y]}{2}.$$  

(B.3)

After the substitution of the function $g$ with the square root and the expressions of $E[w_y]$ and $\text{Var}[w_y]$ from Eqs. (B.1) and (B.2) in Eq. (B.3), the final approximate expression for $E[L_2(\delta y)]$ is as in Eq. (4).

Similarly, the variance of $L_2$ can be approximated as [28]

$$\text{Var}[L_2(\delta y)] \approx |g'(E[w_y])|^2 \text{Var}[w_y]$$

(B.4)

and after the substitutions of $g$, $E[w_y]$, and $\text{Var}[w_y]$, the final expression for $\text{Var}[L_2]$ is as in Eq. (5).

Appendix C. Statistics of $L_\infty(\delta y)$

The maximum norm can be defined in terms of the order statistics $z_{1:m}, z_{2:m}, \ldots, z_{m:m}$ of the $m$ non-identically distributed variables $z_1, z_2, \ldots, z_m$ [29], thus $L_\infty(\delta y) = z_{m:m}$. The pdf of $z_{m:m}$ is [30]

$$f_{z_{m:m}}(z) = \frac{1}{(m-1)!} \left| \begin{array}{c} \begin{array}{cccc} \cdots & F_{z_1}(z) & F_{z_2}(z) & \cdots \\ F_{z_1}(z) & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ F_{z_1}(z) & F_{z_2}(z) & \cdots & F_{z_m}(z) \end{array} \end{array} \right|^{m-1}$$

for $m$ rows,

where the $f_{z_k}(z)$ and $F_{z_k}(z)$ are the pdf and cdf of each $z_k$, and $|A|^+$ is the permanent of matrix $A$ (the permanent is defined like the determinant except that all signs in the expansion are positive). It is hard to compute $f_{z_{m:m}}(z)$ analytically for $m > 2$, and even harder to derive the moments with direct integration from the pdf. In the literature, recurrence relations are given for the moments of the order statistics [31] but they are not of any use here because it is required that one of the $m$ order statistics is known.
References


