뇌파 분석을 위한 LTS 추정기법을 이용한 시계열 데이터의 효율적인 프랙탈 차원 추정

이광호
목포대학교 공과대학 컴퓨터공학교육과

Efficient Estimation of the Fractal Dimension from Time Series Data Using LTS (Least Trimmed Squares) Estimator for EEG (Encephalogram) Analysis

Kwangho Lee
Dept. of Computer Eng. Education, Col. of Engineering
Mokpo National University

요 약

본 논문은 일차원의 시계열 데이터를 입력으로 하여 위상공간 재구성 과정을 거쳐 다차원 위상공간상에서 프랙탈 차원을 계산하는 효율적인 방법을 제안한다. 프랙탈 차원의 추정에 소요되는 계산량을 줄이기 위해 로그 연산을 비트 연산으로 대치하고, 거리계산의 순서를 바꿈으로써 위상공간의 차원에 무관한 상수 시간의 계산복잡도를 가지는 알고리즘을 구현하였다. 또한 최소절단자승 추정기법을 적용하여 로그-로그 그래프 상에서의 기울기 추정을 함으로써 프랙탈 차원의 추정치에 대한 정확도를 높였다. 참값이 알려진 시계열 데이터에 대한 차원 추정 실험을 통하여 제안된 방법의 정확성을 보였다.

1. 서론

자연계에는 기존의 선형적 과학으로 해석되지 않는 여러 가지 혼돈(chaos) 현상이 존재한다. 카오스와 비선형 동력학, 그리고 프랙탈 과학은 컴퓨터 그래픽스 물리학, 수학 등의 여러 학문 분야에 걸쳐 연구되고 있는 분야이며 난류 해석, 기상 예측, 비선형계 해석, 생체 신호 해석 등의 많은 분야에서 연구가 행해지고 있다. 자연계의 혼돈 현상을 나루는 이른바 혼돈이론(chaos theory)은 최근의 뇌-과학 연구와 발맞추어 학계에 보고되고 있다 [1,2]. 인간의 뇌파도 여러 연구진들의 연구에 의해 혼돈현상의 일부라고 보고된 바 그 실체를 밝히고자 하는 연구도 있었다 [3,4]. 특히 의학계에서는 환자의 뇌파로부터 프랙탈 차원을 계산하여 이를 병리 진단에 이용하려는 움직임이 있다 [5,6]. 그 동안의 여러 연구진들에 의해 뇌파를 분석하기 위한 수단으로 푸리에 해석처럼 전통적인 방식이 시도되었으나 뇌파의 비선형적인 특성 때문에 푸리에 해석과 같은 선형 해석에 근거한 분석방법으로는 바람직한 결과를 얻지 못했다.

본 논문은 뇌파 데이터 및 심전도 등의 데이터 해석을 통한 의료진단 지원 시스템의 한 부분으로서 의학계에의 적용을 목적으로 개발된 시계열 데이터로부터 프랙탈 차원을 추정하는 방법을 제안한다. 제안된 방법은 기존의 방법보다 계산시간을 단축하고 추정치의 정확도를 개선하였다. 본 논문은 대표적인 끌개들을 이용하여 비선형계의 끌개의 복잡도에 대한 정량적인 척도로서의 프랙탈 차원을 구하는 방법론을 제시한다.

2. 프랙탈 차원의 여러 가지 정의

카오스 신호의 복잡도 척도로서의 프랙탈 차원은 다음과 같은 여러 가지 정의를 가지고 있다 [7,8,9].

2.1 용적 차원 (capacity dimension)

선의 차원은 1, 면의 차원은 2, 세상의 차원은 3인 일반적인 유클리디안 차원을 실수 차원으로 확장하기 위해 임의의 차원의 위상 공간에 점들이 모여 있는 상황을 가정하자. 한 축의 길이가 L 인 $d-$차원의 공간을 채우기 위하여 한 축의 길이가 ε 인 하이퍼큐브를 사용했을 때, $N(\varepsilon)$ 개의 하이퍼큐브가 사용되었을 때의 관계식은 다음과 같다.

$$N(\varepsilon) = L^{d}(1/\varepsilon)^{d} \tag{1}$$

이 식에서 선, 면, 체적의 경우에 d 는 각각 1, 2, 3에 해당하고, 하이퍼큐브는 각각 사분, 정사각형, 정육면체에 해당한다. 따라서 식(1)은 유클리디안 차원을 포함하는 일반식이다. 식(1)에 로그를 취하면

$$d = \frac{\log N(\varepsilon)}{\log L + \log(1/\varepsilon)} \tag{2}$$

이 되며, 여기에서 ε 을 작은 값으로 극한시키면 L 을 포함하는 항을 무시할 수 있으므로 용적 차원 d_C 는 다음과 같이 정의된다.

$$d_C = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)} \tag{3}$$

2.2 정보 차원 (information dimension)

용적차원은 하이퍼큐브 내에 끌개상의 원소가 단지 존재하는가의 여부만을 고려하는 반면 정보 차원은 하이퍼큐브 안에 존재하는 끌개상의 원소의 빈도까지 고려하여 다음과 같이 정의된다.

$$d_I = \lim_{\varepsilon \to 0} \frac{-\sum_{i=1} p_i \log p_i}{\log(1/\varepsilon)} \tag{4}$$

여기에서 분자의 항은 엔트로피를 의미하므로 결국 정보 차원은 하이퍼큐브의 크기 ε 의 변화에 따른 엔트로피의 변화(지수적)에 대한 척도로 해석된다. p_i 는 i 번째 하이퍼큐브가 포함하는 끌개상의 원소들의 전체 끌개 원소에 대한 비율을 의미한다.

2.3 상관 차원 (correlation dimension)

상관 차원은 다음과 같이 정의된다 [7].

$$D_G = \lim_{r \to 0} \frac{\log C(r)}{\log(r)} \tag{5}$$

여기에서 $C(r)$ 은 상관적분(correlation integral)이라고 불리며 다음과 같이 계산된다.

$$C(r) = \lim_{n \to \infty} \frac{1}{N^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \theta(r - |x_i - x_j|) \tag{6}$$

여기에서 $\theta(\cdot)$ 은 계단함수(Heaviside function)로, $x<0$ 일 때 $\theta(x)=0$ 이고 $x>0$ 일 때 $\theta(x)=1$ 이다. 즉, 상관적분 $C(r)$ 은 임의의 두 점 x_i 와 x_j 사이의 거리가 r 보다 작은 확률을 의미한다. 식(5)를 다시 쓰면 $C(r) \propto r^{D_G}$ 이고, 이는 r 을 증가시킴에 따라 그 거리값보다 작은 점들의 쌍의 개수가 증가하는데 그 증가율을 결정짓는 지수값이 차원에 해당한다는 기하학적 해석이 가능하다.

본 논문은 다른 차원에 비해 그 계산량이 적고 구현이 비교적 쉬운 상관차원을 구현하여 끌개의 차원에 대한 추정 실험에 이용하였다. 상관차원을 구하는 문제는, 식(5)에 의해, $\log C(r)$ 과 $\log(r)$ 을 플로팅한 점들을 선형으로 회귀하였을 때의 구해지는 직선 식에서 기울기에 해당하는 파라메터를 추정하는 기울기 추정 문제로 바뀐

다

3 제안된 상관 차원 추정 방법

3.1 위상 공간 재구성 (phase space reconstruction)

혼돈 신호의 프랙탈 차원은 일반적으로 정수가 아닌 실수이나 추정하고자 하는 차원의 크기에 대해서도 알려져 있지 않은 경우가 대부분이다. 또한 측정 가능한 비선형계의 신호는 보통 1-차원적인 신호이기 때문에 원래의 측정된 차원(1-차원)하에서 차원을 측정하면 항상 1 이하의 추정치를 얻게 될 것이며 이는 분명히 모순이다. 따라서 여러 개의 차원을 구하기 위해서는 측정된 1-차원의 신호로부터 다차원 신호로 변환하여 그 변환된 데이터를 이용하여 차원을 구하는 과정이 필요하다. 이에 대한 수학적 기초는 타켄스(Takens)에 의해 확립되었다 [10]. 이를 위해 기존의 연구에서는 채워넣기(embedding) 과정을 거쳐 1-차원 신호를 d-차원 신호로 다음과 같이 변환한다. 1-차원 신호를 $s(n)$ 이라고 할 때, 새로운 d-차원 신호 $y(n)$ 은 다음과 같이 재구성된다.

$$y(n)=[s(n), s(n+T), \ldots, s(n+T(d-1))] \tag{7}$$

여기에서 d 는 채워넣기 차원이라고 일컫고 신호를 해석하기 위한 기저공간의 차원을 말하며, T 는 하나의 d-차원 데이터를 구성하는 1-차원 신호들 간의 시간간격을 의미한다. 한편, T 를 너무 작게 설정하면 식(7)을 구성하는 d 개의 1-차원 신호들이 시간적으로 가까워서 서로 간의 상관도가 크므로, 재구성되는 다차원 신호는 d-차원 공간상에서 대각선 방향으로 놓이게 되어, 이와 같이 재구성된 d-차원 위상공간 상에서는 매우 작은 차원의 추정치가 얻어진다. 이와 반대로 T 값이 클 때는 d-차원 데이터를 구성하는 요소들의 상관도가 급격히 작아지므로, 상관성을 이용한 차원의 계산은 그 의미를 잃게 된다. 본 논문에서는 자기상관(autocorrelation)을 이용한 방식보다 우수하다고 알려진 평균 상호 정보(average mutual information)에 의하여 T 를 결정하였다 [11].

3.2 시간복잡도의 개선

식(5,6)에 의해 프랙탈 차원 추정을 하기 위해서는 두 개의 d-차원 데이터 x_i 와 x_j 간의 거리 r 을 구하는 과정이 포함되어 있다. 표기의 편의상 x_i 와 x_j 를 각각 $y(i)$ 와 $y(j)$ 라고 놓으면 식(5,6)에서 $|x_i - x_j|$ 의 제곱은 다음과 같이 표현된다.

$$\begin{aligned} r^2 &= |x_i - x_j|^2 = |y(i)-y(j)|^2 \\ &= [s(i)-s(j)]^2 + \cdots + [s(i+T(d-1)) - s(j+T(d-1))]^2 \end{aligned} \tag{8}$$

따라서, $\log(r)$ 은 식(8)에 로그를 취한 값의 1/2이다. 위상공간 재구성에 의해 n 개의 d-차원 데이터가 생성되었을 때, $\log(r)$ 거리 계산이 필요한 횟수는 $C_2^n = n(n-1)/2$ 이므로, 5,000개의 d-차원 시계열 데이터로부터는 천만번 이상의 거리계산, 즉, 로그 연산이 필요하다. 그러므로 이 과정의 계산시간 단축이 전체 수행시간의 단축에 필수적이다. 계산시간을 단축을 위하여, 로그연산을 비트 연산으로 바꾼으로써 계산시간을 1/5로 단축시켰다. 이는 컴퓨터에서 실수를 저장하는 변수는 8바이트 크기의 배정도(double precision floating point number)로 선언되는데, 배정도 실수가 1비트의 부호비트, 11 비트의 지수(exponent), 그리고 52 비트의 가수(mantissa)로 구성된 점에 착안하여 지수부분을 비트 연산에 의해 추출함으로써 로그연산의 필요성을 제거하였다. 그리고 C_2^n 번의 거리 계산시 d-차원 데이터 쌍을 짝짓는 순서를 변경함으로써 약 d 배의 계산속도 향상을 이루었다. 그 과정은 다음과 같다. d-차원 데이터 쌍을 $y(i)$ 과 $y(j)$ 라 하고 이 데이터 쌍의 거리제곱 r^2 이 식(8)에 의해 이미 계산된 후라고 가정하면, $y(i+T)$ 와 $y(j+T)$ 간의 거리제곱 r'^2 은 다음과 같다.

$$r'^2 = r^2 - [s(i)-s(j)]^2 + [s(i+Td) - s(j+Td)]^2 \tag{9}$$

즉, 이미 r^2 값을 알고 r'^2 을 구하는데 추가되는 연산은 덧셈 3회와 곱셈 1회이며 이 연산량은 채워넣기 차원 d 에 무관한 양이다. 반면에, 이와 같은 계산상의 순서를 고려를 하지 않은 기존의 방식으로 계산했을 때 덧셈 $(2\times d-1)$회 및 곱셈 d 회의 연산이 필요하므로 채워넣기 차원에 비례하여 계산량이 증가한다. 따라서 기존의 방식보다 약 $5\times d$ 배의 계산시간의 향상이 있었다. 채워넣기 차원이 클수록 계산시간의 단축은 상대적으로 더 크다.

3.3 프랙탈 차원 추정 정확도의 개선

방대한 실험을 통하여 로그-로그 그래프를 관찰하면 일반적으로 그래프가 이상적인 직선 형태를 보이지 않는다. 따라서 이와 같은 상황에서 직선의 기울기를 구하는 문제는 그래프의 어느 부분으로부터 기울기를 산출하는가에 따라 그 추정치의 변동이 매우 크다. 뇌파의 프랙탈 차원을 정확하게 추정해내는 것은 그 정보가 의학 및 그 밖의 응용분야에의 적용의 타당성과 관련되는 문제이기도 하므로 추정치의 정확도는 매우 중요하다. 그러나 구현된 실험결과에 따르면 데이터로 사용된 뇌파 신호들 뿐 아니라 이상적인 혼돈 신호들(로렌쯔 끌개, 로지스틱 끌개 등)에 대해서도 완벽한 직선을 보이는 로그-로그 그래프는 발견되지 않았다. 지금까지의 여러 연구진들도 이와 같은 사실을 보고하여 왔으나, 이와 같은 상황에서의 직선의 기울기를 구하는 방식이 연구자들마다 각기 다르고 휴리스틱한 면이 많았다. 본 논문에서는 프랙탈 차원의 추정치에 대한 정확도를 높이는 측면에서 일반적으로 적용되는 최소자승추정(LS, Least Squares estimator) 외에 최소절단자승추정(LTS, Least Trimmed Squares estimator)을 적용하였다 [12]. 최소절단자승 추정기법은 다음 식을 최소화하는 회귀 방법이다.

$$\min \sum_{i=1}^{h} (r^2)_{i:n} \tag{10}$$

여기에서 $(r^2)_{1:n} \le \cdots \le (r^2)_{n:n}$ 은 정렬이 된 잔차의 제곱을 의미하며 $h = \lfloor n/2 \rfloor + 1$ 이다. 식(10)은 값이 큰 잔차제곱들이 합산에 사용되지 않는다는 점만 제외하면 최소자승추정과 같으며, 큰 잔차들이 회귀 과정에 기여하지 않는다는 점이 로그-로그 그래프의 직선에서 많이 벗어난 곡선 부분을 무시하는 방향으로 기울기를 찾아내는 원리이다. 실험에 사용되었던 입력데이터인 뇌파 신호뿐 아니라 수식적으로 생성된 이상적인 끌개 데이터를 사용한 실험에서도 로그-로그 그래프가 완전한 직선을 보이지 않는다는 그 동안의 경험에 비추어 볼 때, 최소자승오차를 최소화하는 최소자승추정법보다 최소절단자승추정법이 프랙탈 차원의 추정을 위해 보다 적합한 회귀 모형임을 발견할 수 있었다.

3.4 차원 추정치의 수렴

채워넣기 차원을 증가시키면서 프랙탈 차원을 구하는 과정을 반복하다가 추정치가 일정한 값에 수렴할 때의 추정치를 해당 끌개의 프랙탈 차원으로 결정한다. 채워넣기 차원이 주어진 입력 데이터의 고유한 차원에 비해 작을 때에는 채워넣기 차원이 충분히 끌개를 펼쳐서 표현할 수 없으므로 (실例로 5.5의 차원을 가지는 어떤 끌개를, 채워넣기 차원을 3으로 하여 위상공간을 재구성한 후 차원을 추정하였다면 실제 끌개를 3차원 위상공간에 투영시킨 그림자(projection)에 대한 차원을 구하게 되므로), 최소한 실제의 차원보다 큰 채워넣기 차원에서의 다차원 데이터의 재구성이 요구된다. 참값에 해당하는 차원의 약 2배 크기의 채워넣기 차원에서의 다차원 데이터 재구성의 필요성은 여러 연구진들에 의해 주장된 바 있다 [10, 13, 14]. 그러나 미지의 신호에 대해 프랙탈차원을 구하는 것이 목적인 상황에서는 참값의 차원을 알 수 없는 상황이므로, 채워넣기 차원을 점차 늘려가면서 직선의 기울기를 구하다가 그 값이 수렴하는 단계에서의 기울기 추정치를 해당 끌개의 차원으로 결정하였다.

4 실험결과 및 분석

4.1 로렌쯔 끌개

로렌쯔에 의해 알려진 로렌쯔 끌개는 기상 변화에 대한 간단한 수학적 모형으로서, 기이한 끌개 중에서도 가장 널리 알려진 다음과 같은 미분방정식으로 표현된다 [15].

$$\begin{aligned} \frac{dx}{dt} &= \sigma \times (y - x) \\ \frac{dy}{dt} &= R \times x - y - x \times z \\ \frac{dz}{dt} &= x \times y - B \times z \end{aligned} \tag{11}$$

이 식에서 σ , R , B 는 상수이며 로렌쯔는 각각 10, 28, 8/3 으로 고정시켰으며 본 논문에서도 이 상수를 이용하였고, 초기값으로 (x,y,z) = (1,1,1)에서 시작하여 4차 룽게-쿠타 (Runge-Kutta) 방법 [16]에 의해 시간간격 0.01씩 증가시키면서 얻은 시계열 데이터를 얻었다. 이 과정에서 처음에 생성되는 10,000개의 시계열 데이터는 버리고 그 이후의 16,384개의 데이터를 실험에 이용하였다. 초기의 데이터를 버리고 이후의 데이터를 취한 이유는 불안정 상태를 지나 충분히 끌개 안으로 수렴된 후의 순수한 끌개 상의 데이터를 얻기 위함이었다. 룽게-쿠타 방법에 의해 생성된 x, y, z 의 3차원 시계열

중에서 위상공간 재구성 단계에서의 입력 데이터로 x 좌표의 시계열을 사용하였다 (그림 1)은 채워넣기 차원이 6일 때의 최소자승추정 및 최소절단자승추정 기법을 적용하여 커브의 기울기, 즉, 차원의 추정치를 구한 그래프이다 최소절단자승추정법에 의해 로렌쯔 끌개의 차원이 2.05로 추정되었는데, 이는 로렌쯔 끌개의 알려진 참값 2.05 ±0.01에 오차범위내로 정확한 추정치를 발견했음을 보여준다 (그림 2)는 채워넣기 차원을 1에서 10까지 변화시키면서 차원 추정치의 변화를 도시한 그래프로서, 채워넣기 차원이 4이상일 때 참값에 수렴이 시작됨을 볼 수 있다

4.2 로지스틱 끌개

로지스틱 끌개상의 점들은 다음 사상에 의해 표현된다 [17]

$$x_{n+1} = a \times x_n \times (1 - x_n) \tag{12}$$

이 식에서 상수값은 $a=3.56994$ 로 고정하고 시계열 데이터를 생성하였다 초기값은 $x=0.5$ 로 하였고 처음에 생성되는 10,000개의 시계열 데이터는 버리고 그 이후의 16,384개의 데이터를 실험에 이용하였다 채워넣기 차원은 1부터 7까지 변화시키면서 구한 차원의 추정치의 변화 양상은 (그림 3)에 보였으며, 추정 과정 중의 전형적인 예로써 채워넣기 차원 2일 때의 로지스틱 끌개의 차원의 추정치를 최소자승추정 및 최소절단자승추정 기법에 의해 구한 결과를 (그림 4)에 보였다. 로지스틱 끌개의 최소절단자승추정 기법에 의한 추정치는 0.50으로 추정되었는데 이는 알려져 있는 로지스틱 끌개의 차원인 0.4926-0.5024에 오차 범위 내로 정확히 추정하였다

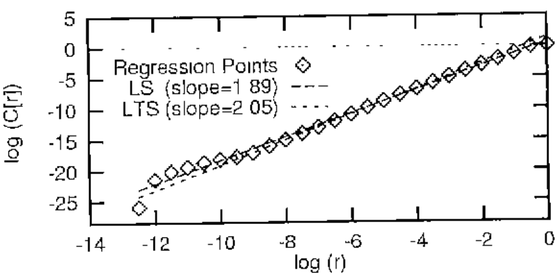
5 결론

본 논문에서는 기존의 프랙탈 차원 추정 방법보다 빠르고 정확하게 혼돈 신호의 프랙탈차원을 구하는 계산 방식을 제안하고, 실험을 통하여 제안된 프랙탈 차원 추정 방식의 정확성 및 계산 속도의 증가 효과를 확인하였다 이미 프랙탈 차원의 값이 알려진 혼돈 데이터들에 제안된 방식을 적용하여 오차 범위 내에서 참값을 추정하는 정확성을 보였으며, 실제 환자의 뇌파 데이터 및 정상인의 뇌파 데이터에 적용함으로써 정신병 환자의 진단에 적용할 수 있는 가능성을 타진하고 있는 바 이에 관한 상세한 보고는 추후의 연구 과제로 삼는다

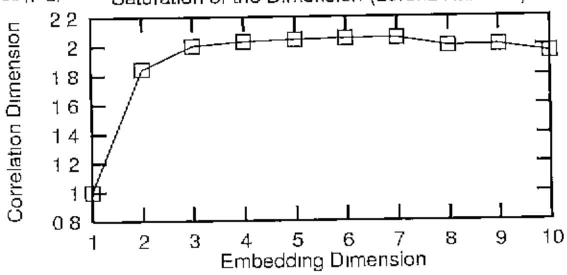
참고문헌

[1] 심주진, 심흥기, "카오스 동성 시뮬레이터 개발에 관한 연구," 한국정보과학회 '98 봄 학술발표논문집(B), 제25권 1호, 1998년 4월 24일-25일, 충남대학교, pp 231-233
[2] 이광호 "미분방정식으로 표현된 기이한 끌개의 차원 분석," 한국정보과학회 호남제주지부, 학술발표논문집, 1998년 8월 29일, 목포대학교
[3] D W Duke and W S. Pritchard, Proceedings of the Conference on Measuring Chaos in the Human Brain, Florida State University World Scientific, April 3-5, 1991
[4] J P Pijn, J V Neerven, A Noest, and F H Lopes da Silva, "Chaos or noise in EEG signals, dependence on state and brain site Electroencephalography and clinical Neurophysiology, 79, 371 381 1991
[5] C Besthorn, H Sattel, C G Kabisch, R Zerfass, and H Forstl, "Parameters of EEG dimensional complexity in Alzheimer's disease," Electroencephalography and clinical Neurophysiology, 95, 84-89, 1995
[6] C J Stam, B Jelles, H A M Achtereekte, S. A R B Rombouts, J P J Slaets and R W M Keunen, "Investigation of EEG non linearity in dementia and Parkinson's disease," Electroencephalography and clinical Neurophysiology, 95, 309 317, 1995
[7] P Grassberger and I Procaccia "Measuring the Strangeness of Strange Attractors," Physica Vol D, No 9, 1983, pp 189-208.
[8] G L Baker and J P Gollub, *Chaotic Dynamics an introduction*, Cambridge Univ Press, 1996, pp 110-119
[9] F C Moon Chaotic and Fractal Dynamics, John Wiley & Sons, 1992, pp 329-339
[10] F Takens "Detecting Strange Attractants in Turbulence" in D A Rand and L S Young (eds) *Dynamical Systems and Turbulence*, Warwick 1980, Springer, Berlin Heidelberg New York, pp 366-381 (*Lecture Notes in Mathematics*, Vol 898)
[11] Henry D I Abarbanel, *Analysis of Observed Chaotic Data*, Springer, 1996
[12] P J Rousseeuw and A M Leroy, "Robust Regression and Outlier Detection," John Wiley & Sons, New York, 1987
[13] N Packard et al, *Phys Rev Lett*, Vol 45, 1980, pp 712-716
[14] H Whitney, *Annals of Mathematics*, Vol 37, 1936, pp 645-680
[15] E N Lorenz "Deterministic Nonperiodic Flow," Journal on the Atmospheric Science, Vol 20, No 2, March, 1963 pp 130-141
[16] W H Press et al. *Numerical Recipes in C*, Cambridge University Press, 1992, pp 710-714
[17] R M May, Nature vol 261, 1976, pp 459

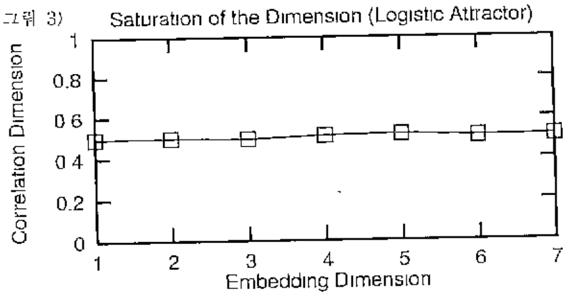
(그림 1) Log-Log Graph using Lorenz Attractor, Emb Dim=6



(그림 2) Saturation of the Dimension (Lorenz Attractor)



(그림 3) Saturation of the Dimension (Logistic Attractor)



(그림 4) Log-Log Graph using Logistic Attractor, Emb.Dim=2