패턴의 과학 [1]: 패턴의 자기닮음꼴과 프랙탈 차원

2019년 12월 6일
권석준

1. 왜 패턴에 관심을 갖는가

자연에는 다양한 종류의 패턴이 존재한다. 작게는 나노미터 스케일에서 관측되는 단결정-single crystal의 격자-lattice 패턴부터 애니메이션<니모를 찾아서>의 주인공이자 남태평양에서 서식하는 열대어 환경에서의 줄무늬 패턴과 공게 구름의 변화무쌍한 형태까지, 스케일을 더 확장하면 복잡한 두꺼운 대기 순환이 만들어낸 지구보다 수 배나 큰 대적점 같은 소용돌이 패턴-vortex pattern에서 우리 은하에 분포하는 수천억 개의 별들이 이루는 나선형 별 모양의 패턴까지, 더 나아가면 우주 마이크로파 배경-cosmic microwave background, CMB의 미세하지만 분명히 존재하는 열적 요동-thermal fluctuation의 패턴까지 우주의 거의 모든 스케일에 걸쳐 갖가지 종류의 패턴이 관측된다.

물론 패턴이 자연에만 존재하는 것은 아니다. 시시각각 변동하며 투자자들의 마음을 줄이게 만드는 주가 변동 시계열 데이터에도 패턴이 감춰져 있고 나일강의 범람 규모를 수천 년 동안 추적한 지리 관측 데이터에도 패턴이 숨어 있다. 히말라야 산맥에 위치한 고산준봉들의 높이와 면적 분포에도 지배적인 패턴이 있고 야구 경기에서 타자를 속
이기 위해 유인구를 배합하는 투수의 투구에도 패턴은 존재한다. 소셜네트워크SNS 상에서 사람들의 의견이 형성되고 전파되는 과정에도 특정한 종류의 패턴이 발견되며 도시의 도로가 연결되는 방식을 추상화한 데이터에도 패턴이 있고 인류 문명사에 겪어가 수 놓인 전쟁의 역사에도 반복되는 패턴이 있는 것처럼 보인다.

자연에서든 사회에서든 역사에서든 그리고 인공적인 것이든, 인간이 패턴을 인식하는 과정은 감각 기관에서 관찰 대상의 신호를 처리함으로써 시작된다. 예를 들어 전자현미경으로 관찰한 나노미터 스케일의 결정 구조는 일면의 신호 처리과정을 거쳐 2차원 혹은 3차원 이미지로 인간의 시각 기관이 인지할 수 있는 스펙트럼, 즉, 가시광선 대역에서 시각화된다. 격자 같은 패턴은 질서정연하다는 특징이 있기 때문에 재료의 결정 구조를 잘 모르는 사람이 봤도 한눈에 규칙성 있는 패턴의 존재를 알아차리는 것은 어렵지 않다.

예를 들면 가정 주부수 범위에 섞어 있는 다양한 주부수의 진동하는 공기 밀도는 파동의 형태로 인간의 귀로 전달되고 달팽이관 속에서 청각 신호로 바뀌어 그것이 아름다운 하모니의 패턴인지 아니면 불규칙한 소음인지 구분된다. 필자의 장남이 다섯 살 무렵, 필자의 노트북에서 블록공중합체\textit{block copolymer} 바막의 미세한 자기조립 패턴\textit{self-assembly pattern}의 전자현미경 사진을 비교한 데이터를 보고, 열개의 샘플 사진 중 어떤 샘플이 가장 예쁜 패턴을 가지고 있는지 (다른 말로 하면 어떤 샘플이 가장 반듯한지) 단번에 알아맞춘 적이 있다. 이에 꼭 놀리는 필자는 어린 아이도 패턴의 규칙적인 수준을 상대적으로 비교하고 그로부터 심미적 가치도 느끼는 것을 보고 인간에게 모종의 정보 처리 회로 혹은 그에 해당하는 기능이 있는 것은 아닐까 생각할 적이 있다.

하지만 특정 신호 속에 숨어 있는 패턴의 특징을 명확하게 감지하는 것은, 실제 혼란받은 연구자라고 해도 인간의 감각 기관에만 의존하는 것만으로는 충분하지 않은 경우가 대부분이다. 예를 들어 컴퓨터의 도움 없이 거대한 데이터 속에 숨어 있는 신호의 주기성\textit{periodicity}이나 특이성을 본격적으로 알아내는 쉽지 않으며 신호의 구조가 복잡할수록 관찰만으로 패턴을 단번에 인식하기가 어려워진다. 그럼에도 불구하고 인간의 호기심과 미적 가치를 탐하는 본능은 주변 데이터에 내재된 패턴을 무시하기보다는 찾는 쪽을 선호하는 것으로 보인다. 예로 인간이 패턴을 인식하고 이해하는 이유 중 하나가 예측 가능성을 높여주며 주변 환경 변화에 대비하거나 정보 처리에 필요한 자원의 소모를 줄이려는 목적이라고 추정하는 심리학 혹은 인지 과학 이론도 있다.\textsuperscript{[1]}

인간은 대칭 구조에서 아름다운을 느끼는 경우가 많는데 이 역시 대칭으로 대표되는 아름다움이라는 가치 속에 패턴이 갖는 정보 처리의 효율성이 내재되었다기 때문이라고 생각할 수 있다. 따라서 아무리 복잡한 신호와 데이터라고 해도 어떤 패턴이 숨어 있다면 그것이 무엇인지, 얼마나 복잡한 것인지, 다른 것과 어떻게 구분이 되는지 알아내려고 하는 인간의 노력을 계속될 것이다. 불안하게도 그것이 개인의 암묵적이나 경험에만 의존할 수는 없으므로 결국 패턴에 대한 이해는 과학적인 방법, 특히, 주로 수학과 물리학적 방법에 의존할 수밖에 없다.

필자의 연구 분야는 나노미터 수준에서 관측되는 갖가지 재료의 자기조립, 마이크로미터 스케일에서 관측되는 빛과 물질의 상호작용, 밀리미터나 센티미터 수준의 현상을 해석하는 계산 과학까지 비교적 넓은 스케일에 걸쳐 다양한 편이지만, 전체를 관통하는 핵심 기워드가 있다면 그 중 하나는 아마도 ‘패턴’일 것이다. 나노미터 스케일의 격자나 자기조립 패턴부터 뇌파 등 생체 신호의 복잡도 계산까지 필자가 관심있는 것은 바로 데이터나 신호에 내재된 패턴의 수학적 특성을 명확히 밝혀내고 그것을 공학적으로 응용하는 일이다. 이를 다른 말로 씌우면 분석 대상인 데이터에 내재된 질서도\textit{order} 혹은 정보를 알아내는 것이고 조금 더 상세하게 이야기한다면 이를 정량적으로 측정하고 다른 신호와 구분하는 것이다.
단도직일적으로 필자가 연구하는 분야의 패턴 분석에 대한 이야기를 바로 시작하는 것도 나쁘지 않겠으나, 그에 앞서 패턴 분석이 어떠한 방식을 거치고 있는지에 대해 살펴보는 것도 좋을 것이다. 본 글은 주로 패턴의 복잡도 분석에 초점을 맞추고 있고 1부에서는 패턴에 내재된 자기덜음률과 프랙탈 차원 분석, 2부에서는 패턴의 정보량을 직접적으로 측정할 수 있게 해 주는 엔트로피의 개념과 실제 측정 사례 등에 대한 내용을 살펴볼 것이다.

2. 패턴의 자기덜음률

자연에 존재하는 패턴 중에는 축적을 달리해도 비슷한 특성이 계속 나타나는 경우가 자주 보인다. 예를 들어 산에 가면 흔히 볼 수 있는 고사리 같은 얕은 식물은 적어도 세 번 정도 일부분만 계속 확대해도 원래의 이파리 배치 패턴이 그대로 유지되는 것을 볼 수 있다. 적란은 같은 구름 역시 멀리서 볼 때나 가까이에서 확대해서 볼 때나 축적을 가리면 구분할 수 없는 비슷한 모양의 구름을 가진 경우가 있으며 산의 지형적 특성 역시 관찰하는 스키어의 상관없이 서로 닿아 보이는 형태가 유지되는 경우를 자주 볼 수 있다. 이렇게 관찰하는 축적에 상관없이 특정한 모양이 지속적으로 유지되는 특성을 자기덜음률이라고 하며, 자연에서 이러한 덜음률 유지가 자주 관찰되는 것은 우연이 아니다. 고사리나 브로콜리 같은 식물은 예로 들어 보자. 자기덜음률만 유지한다면 유전 정보를 여러 종류로 갖고 있어도 걸린한 유전 정보의 조합만으로도 제한된 공간에서 최대한 복잡한 형태를 만들고 나아가 다양한 기능을 갖출 수 있다. 이것이 왜 생명체에 유리한 것이냐? 예상할 수 있다. 이는 생명체가 오랜 시간 동안 살아남고 또한 다음 세대에 무사히 유전 정보를 넘기는 과정에 더 유리하기 때문이다.

지난 6월, 영국의 과학잡지 네이처에는 두뇌가 외부의 정보를 인지하는 과정에서 가능한 단위 시간 동안 많은 시각 정보를 처리하고 외부의 노이즈에 유연하게 대처하기 위해 시각 중추의 시각 반응이 신경 지극에 대해 면합수함수, power-law의 조건을 보인다는 연구 논문이 게재되었다.[2] 정보 처리 과정에서 관측된 면합수관계는 시신경 세포들의 신경 반응에 자기덜음률 패턴이 있다는 것을 의미하는 것이기도 한다. 앞서 언급한 고사리의 경우 대칭의 기준이 되는 한 선분에서 양쪽으로 이파리가 맞게 나열 것인지 정도만 유전자의 속에 암호화 해 놓으면 (나중에 살펴 보겠지만 이는 L-system이라는 자기덜음률 생성을 코딩하는 컴퓨터 형식 문법으로 재현이 가능하다.) 별다른 어려움 없이 제한된 면적에 최대한 많은 이파리들을 제어 얻어서 단위 면적당 광합성의 효율을 높일 수 있다.

반면 자기덜음률을 통해 유전 정보가 절약되는 이점이 없다면, 고사리는 주어진 면적을 복잡한 모양으로 채우기 위해 매번 분기할 때마다 서로 다른 유전 정보에 대한 참조가 반복적으로 필요할 것이고, 다음 세대로 복잡한 유전 정보를 전달하는 과정에서 오류는 error rate도 누적될 것이다. 결국 고사리의 유전 정보를 수백 수천 세대에 걸쳐 안정적으로 전달할 수 있을 정도로 오래 존속할 수 없었던 것이다. 물론 생명체의 유전 정보 활용성이 생명체에서 관측되는 자기덜음률의 근본 원인은 아니다. 다만 왜 자기덜음률이 생명체에 나타날 수 있는지 주요 원인 중 하나로 해석하면 될 것이다.
자기달음꼴은 비단 생명체에서만 관측되는 것은 아니다. 자기달음꼴이 향상 동반하는 특징이 척도불변성 scale-invariance라는 사실을 상기하면 자기달음꼴은 통계물리학에서 말하는 임계 현상 critical phenomena에서도 자주 관측된다. 예를 들어 계가 상전이를 겪을 때 계의 질서의 파라미터 order parameter는 상전이를 일으키는 변수(예를 들어 온도나 압력, 자기장의 세기 등)에 대해 역함수 의존 관계를 갖는다. 상전이는 계가 열역학적으로 불안정한 상태에 있다는 것이므로 시간에 따라 계는 상태 변화를 겪는다. 예를 들어 작은 입자가 액상에 분산되어 있다가 분산 안정성을 잃어 버리면 시간이 지남에 따라 입자들이 서로 붙게 된다. 이 현상의 동적 특징이 주로 확산 diffusion에 의해 지배받아 발생하며 확산한 계에 집중한 곳으로서 이를 확산한계응집 diffusion-limited aggregation, DLA라고 한다. [그림1]처럼 작은 클러스터가 생기는 지점을 중심으로 바람 방향을 향해 점점 응집 구조가 확산하는 양상이 DLA의 대표적인 특징이다. 이는 나뭇가지의 가지차기, 혹은 번개의 갈라짐 형상(분기 패턴)과도 매우 닮았다. DLA를 비롯하여 나무와 번개의 패턴 등은 공통적으로 척도불변성을 보이는 데, 따라서 이들은 모두 자기달음꼴을 갖는다.

또한 지진이 발생하는 빈도와 지진의 강도 사이 혹은 산불의 빈도와 규모 사이, 인구 규모 별로 줄세운 도시의 규모 분포 등에서도 척도불변성이 자주 관측되고, 따라서 이러한 동적 현상의 이면에도 자기달음꼴이 숨어 있음을 알 수 있다. 바흐의 첼로 조곡 ceol suites에서도 자기달음꼴이 발견되는데[3], 바흐가 의도적으로 수학 이론을 상정하여 자기달음꼴을 갖게끔 인위적으로 화성을 설계했을 것 같지는 않지만 복잡한 멜로디 속에 단순한 원칙이 자리잡고 있는 것은 사뭇 흥미로운 부분이다.

자기달음꼴을 찾는 과정에서 빈도와 확인상의 크기 사이에는 역함수 관계가 자주 관측된다. 예를 들어 앞서 언급한 고가리 이파리의 경우 길이가 반으로 줄어들면 이파리 개수는 서너 배로 늘어나는 관계, DLA 현상의 경우 입자들이 동치는 과정에서 관측되는 클러스터의 크기가 출현 빈도의 관계, 지진의 진도와 빈발 횟수의 관계 등 다양한 현상, 다양한 패턴에서 역함수 의존성이 관측된다. 사실 역함수가 척도불변성을 만들어 낼 수 있는 유일한 수학적 관계라는 사실을 알게 되면 이는 그렇게 신기한 사실이 아닐 수 있다. 척도불변성을 동반하는 자기달음꼴은 어떤 대상의 부분
여기서 한 가지 유의할 부분은 프랙탈이 반드시 ‘완벽한’ 자기닫음꼴을 가지야만 한다는 것은 아니라는 것이다. 엄밀한 의미에서 완벽한 자기유사성을 갖는 프랙탈은 기하학적 프랙탈 밖에 없으며, 이는 이론적으로는 반복함수계 iterate function systems, IFS를 통해서만 생성될 수 있다. 여기서 말하는 반복함수계는 이론적으로, 어떤 기하학적 특성을 만들어내는 함수를 재귀적으로 반복 적용하는 개념을 말한다. 통계적인 유사성을 만족하는 패턴(망델브로 집합 Mandelbrot set이나 채리아 집합Julia set이나 대략적인/느슨한 유사성을 갖는, 자연에서 발견되는 자기닫음꼴 패턴 등) 무작위적 프랙탈 stochastic fractal은 넓은 의미에서는 프랙탈로 분류되지만, 체도에 무관한 자기닫음꼴이 완벽하게 유지되는 것은 아니다. 그렇지만 본 글에서는 완벽한 자기유사성과 느슨한 자기유사성을 엄밀하게 구분하지 않고 넓은 의미에서 자기닫음꼴-프랙탈 차원의 관계를 다룰 것이다.
정수가 아닌 차원이 어떤 의미를 갖는지 알아보자. 바삭바삭한 튜igham을 잘 입혀진 칫한 한 조각을 생각해 보자. 그 리고 돈금없는 질문을 던져 보자. 이 칫한 조각의 표면적은 얼마나인가? 넓게 잡아 봐야 100 cm²도 안 될 것 같아 보인다. 그렇지만 사실 이 문제는 칫한이 지극히 천축한 간식이라는 사실과는 달리 만만한 문제가 아니다. 예를 들어 누군가는 칫한의 튜igham을 전부 빼내 내 그의 부피를 측정하려는 시도를 할 수도 있을 것이고 누군가는 칫한을 다양 한 각도에서 사진 찍어서 칫한이 차지하는 핀셀 개수의 평균을 구해 보려는 사람도 있을 것이다. 만약 칫한을 전자현 미경으로 관찰한다면 튜igham을 이루는 튜igham가루와 기름 그리고 점가물이 만들어 난 미세한 다공성의 거친 구조까지 관찰할 수 있을 것이고 더 자세히 관찰하면 관찰할수록 칫한 조각의 표면적은 점점 증가할 것이다. 즉, 칫한 조각 의 표면적은 어떤 스케일에서 얼마나 자세하게 관찰하느냐에 따라 달라지는 것이다. 불행히도 전자현미경으로 관찰 하여 표면적이 엄청나게 넘어졌다고 해서 칫한의 양이 더 늘어나는 것은 아니다.

앞서 이미지 핀셀을 이야기했으니 이 개념을 계속 자세히 보자. 칫한의 표면적을 계산하기 위해 낮은 해상도의 핀셀 개수를 세었을 때와 높은 해상도의 핀셀의 개수를 세었을 때, 당연히 후자의 핀셀 개수가 더 많을 것이다. 앞서 살펴본

$$\lim_{x \to 0} \frac{\log(x)}{x} = a$$

의 관계에서 $x$를 정사각형 모양 핀셀 한 개의 한 변의 길이라고 하고 $f(x)$를 칫한 조각을 덮기 위한 핀셀의 수라고 한다면 만약 칫한 조각에 자기달음꿀이 존재하는 경우 앞서 증명한 자기달음꿀을 지배하는 역함수의 범위에 따라 $log(x)$와 $log(1/x)$의 비율은 상수로 고정되어야 한다. 만약 평평한 형태의 칫한 부위를 쪼개 이미지의 넓이가 $A$고 이 를 한 변의 길이 $x$의 정사각형으로 덮기 위해서는 대략 $f(x) \sim A/x^2$개의 핀셀이 필요할 것이다. 따라서 양변에 로그 를 취하면 $a = 2$가 되고 이는 평평한 칫한 조각은 원래 주어진 유플리드 공간의 차원($D_{EUCL}$), 즉, 이 경우 $D_{EUCL} = 2$차 원을 가진다는 뜻이 된다.

그렇게 칫한 조각이 날개나 다리 같은 복잡한 부위라면 필요한 핀셀 개수가 더 늘어날 수 있고, 이 경우 $a$는 2보다 커 질 수 있다. 이를 일반화하여 결국 특정 대상을 덮기 위한 선분 (1차원 대상일 경우), 정사각형(2차원 대상일 경우), 정육 면체(3차원 대상일 경우), 하이퍼 큐브(4차원 이상인 경우) 등 각 차원에 맞는 단위 블록의 크기가 점점 작아질 때마다 블 록 개수가 얼마나 더 필요한지의 로그 비율을 분석할 수 있고 이 값이 만약 상수라면 그 값을 실질적 ‘차원’이라고 정 의할 수 있다. 물론 그 상수가 꼭 자연수일 필요는 없으므로 이를 일반화하여 프랙탈 차원이라고 정의할 수 있다.
예를 들어 [그림2]에 보인 칸토어 집합의 경우 자기있다음꼴을 만들어내는 과정을 반복할 때마다 중간의 1/3이 없어지므로 p번 되풀이하면 길이는 1/3^p가 되고 토막의 개수는 2^p개가 된다. 따라서 p가 계속 커진다면 이 패턴의 프랙탈 차원은 \frac{\log 2}{\log 3} = 0.631로 계산될 것이다. 이 숫자가 차원을 나타내는 값이라고 직관적으로 이해하기는 쉽지 않지만 원래 1차원에서 시작한 선분이 무한한 1/3 토막내기에서 살아 남아 0차원의 점보다는 정보량이 보존된 패턴으로 변모함을 고려하면 이 집합의 실질적 차원은 0과 1 사이에 있어야 할 것이고 칸토어 집합은 그 값이 0.631로 주어졌다는 것으로 이해할 수 있다.

[그림3]의 코흐 눈송이의 경우 자기있다음꼴 되풀이 과정에서 선분 길이는 1/3로 줄여드는 대신 선분 개수는 4개로 늘어나므로 프랙탈 차원은 \frac{\log 4}{\log 3} = 1.262이다. 앞서 보인 칸토어 집합과는 달리 1차원 선분의 차원이 오히려 1을 초과한 것을 볼 수 있는데 이는 선분의 개수 증가 속도가 선분의 길이 감소 속도보다 빠르기 때문이다.
한 가지 유의할 부분은 모든 자기달음꼴 도형이나 패턴의 프랙탈 차원이 항상 비자연수가 되는 것은 아니라는 것이다. 그 대표적인 예가 [그림4]와 같은 바로 공간전극선 space-filling curves의 경우 이들은 기본적으로 자기달음을 갖는 것으로 일반적으로 유메리치 차원(D_{EUC})과 동일한 프랙탈 차원, 즉 자연수 차원을 갖는다. 예를 들어 [그림4]의 페아노 곡선Peano curve type 1이나 힐버트 곡선Hilbert curve의 프랙탈 차원은 2다. 참고로 힐버트 곡선류의 공간전극선에서 말하는 '곡선'은 기하학적으로 곡률을 정의할 수 있다는 의미에서의 곡선이 아닌, 연속적으로 움직이는 점들이 그는 경적으로 이해해야 한다.

3. 패턴의 프랙탈 차원 계산

앞서 살펴 본 것처럼 패턴의 자기달음꼴 특성은 곧 프랙탈 차원을 계산하는 것으로부터 분석될 수 있다. 수학자 모란P. Moran의 이론에 의하면 일반적으로 자기달음꼴을 갖는 패턴은 다음 정리가 성립한다[4].

\[ 1 = \sum_{k=1}^{N} \lambda_k^D S \]

위 식에서 \(N\)은 한 번 반복될 때마다 생기는 자기달음꼴의 개수, \(\lambda_k\)는 \(k\)번째 자기달음꼴 성분의 자기달음꼴 scale, 그리고 \(D_S\)는 패턴 전체의 자기달음꼴 차원similarity dimension이다. (자기달음꼴 차원과 프랙탈 차원은 혼동하면 큰관하하다. 프랙탈은 자기달음꼴을 가지고 있기 때문에 프랙탈 차원과 자기달음꼴 차원이 일치하지만 자기달음꼴 형태라고 해서 는 틀리기는 되는 것은 아니기 때문에 자기달음꼴 패턴의 프랙탈 차원이라는 이야기는 성립하지 않는다.) 예를 들어 앞서 살펴 본 칸토어 집합은 \(N = 2\)이고 \(\lambda_1 = \lambda_2 = \frac{1}{3}\)이므로 \(1 = 2(\frac{1}{3})^D S\), 즉, \(D_S = \frac{\log 2}{\log 3} = 0.631\)이다. 이는 앞서 계산했던 프랙탈 차원과 일치하는 값이다. 이를 응용하면 [그림5]와 같은 비대칭 칸토어 집합asymmetric Cantor set의 프랙탈 차원도 계산할 수 있는데 이
경우 되풀이 과정에서 원쪽 붉은 선분은 0.4배, 오른쪽 녹색 선분은 0.5배로 축소되어 자기님음꼴이 형성되므로 이 패턴의 프랙탈 차원은 위 관계식에서 $1 = 0.4D_s + 0.5D_s$, 즉, $D_s = 0.867$이라는 것을 알 수 있다. 이는 대칭적 칸토어 프랙탈 패턴의 프랙탈 차원 0.631보다는 크게, 더 충분한 자기님음꼴 패턴임을 고려하면 직관적으로 이해할 수 있는 부분이다.

![그림5 비대칭적 칸토어 집합 Asymmetric Cantor set에서 만들어지는 자기님음꼴 패턴](image)

답음비가 알려져 있거나 직관적으로 관찰할 수 있는 대상이 아닌 경우 프랙탈 차원을 계산하기 위해서는 앞서 설명한 단위 블록의 척도 비율에 따른 개수 증가 비율에 대한 추적이 필요하다. 인공적인 것이든, 자연에서 관찰되는 것이든, 주어진 패턴의 프랙탈 차원은 원리적으로는 일명 박스개수세기 방법 box-counting method으로 추정할 수 있다. 이 방법은 이름 그대로, 어떤 형태나 패턴을 단위 블록(예를 들어, $m = 1$차원은 단위 선분, $m = 2$차원은 단위 정사각형, $m = 3$차원은 단위 정육면체(즉, 박스), $m \geq 4$차원은 하이퍼큐브 Hypercube)으로 조밀하게 덮고자 할 때 필요한 블록의 개수를 세고, 그 개수와 블록의 크기 사이의 관계를 역함수로 가정하여 지수 $power$를 계산하고, 그 지수를 프랙탈 차원으로 정의하는 방식이다. 참고로 박스개수세기 방법 box-counting method를 확장하면 정사각형 외에도 2차원의 경우 정육각형, 정삼각형 등, 3차원의 경우 정사면체, 정십이면체 등 다양한 다각형 혹은 다면체를 단위 박스로 활용할 수 있는데, 그중 가장 작은 차원을 갖는 경우를 채택하여 계산된 차원을 하우스도르프 차원 $\text{Hausdorff dimension, } FD_{\text{haus}}$이라 고도 한다.
하우스도르프 차원은 용량 차원 capacity dimension 혹은 콜모고로프 용량 Kolmogorov capacity이라고도 불린다. [그림6-1]은 앞서 살펴본 DLA 패턴의 프랙탈 차원을 박스개수세기 방법으로 계산한 사례다. 그림의 오른쪽 그래프는 이중로그 척도로 그린 박스크기-박스개수 관계인데, 이 그래프에서 선형성이 관측되면 박스크기-박스개수 관계에 역함수 의존성이 있다고 말할 수 있다. 따라서, 선형성이 관측되는 경우, 그래프를 선형회귀하여 기울기를 계산하면 그 절대값이 프랙탈 차원이 된다. 그리고 선형성이 보존되는 척도까지만 실제 자기단일곱이 유지되는 척도라고 할 수 있다.

DLA 같이 컴퓨터로 생성된 패턴이 아닌, 임의의 1차원 데이터나 2차원 이상의 이미지도 적절한 처리 과정을 거치면 프랙탈 차원을 계산할 수 있다. [그림6-2]처럼 주변에서 흔히 볼 수 있는 나무 사진 같은 경우, 나무의 형태만 이진화 필터 binary filter로 추출한 형상의 프랙탈 차원을 박스개수세기 방법으로 계산할 수 있다. 주어진 이미지의 픽셀 개수가 충분하다면 박스 크기를 변화시킬 수 있는 범위가 넓어져 더 정확한 프랙탈 차원 계산이 가능하다.
다만 한 가지 주의 깊게 고려해야 할 사실은 주어진 이미지로부터 프랙탈 차원을 계산하는 것은 이미지에서 측정하기 
고차 하는 대상의 복잡도에 의존한다는 사실이다. 이는 완전 이미지를 이진화 이미지 혹은 회색조 이미지 gray-scale 
image로 변환하여 형태를 구분짓는 경계선(엣지 edge) 정보 등을 추출하는 필터를 어떻게 쓰느냐에 따라 프랙탈 차원 
이 다르게 측정될 수 있다는 것을 의미한다.

프랙탈 차원은 수학적으로 더 확장된 의미에서 정의되고 계산될 수 있다. 박스개수세기 방법은 한 단위 박스 안에 독 
립된 데이터가 있는지 여부만 따져서 필요한 박스의 최소 개수를 세기 때문에 박스 안에 데이터가 몇 개나 있는지 혹은 
각 데이터의 중요도가 가중치를 고려할 수 있는지 여부는 고려되지 않는다. 따라서 이를 반영할 수 있는 개념이 필요해, 정보 이론 information theory에서 도입한, ‘일반화된 q차 엔트로피 generalized entropy at order q’ 혹은 ‘레니 엔 
트로피 Renyi entropy’라는 수학적 개념을 활용할 수 있다. 프랙탈 차원과 관련된 정보를 측정하는데 활용되는 레니 엔 
트로피에서 정의되는 일반화된 q차 프랙탈 차원은 다음과 같이 정의된다.

\[ D_q = \lim_{r \to 0} \frac{1}{q-1} \frac{\log \sum_{i=1}^{M(r)} p_i^q}{\log r}, \quad p_i = N_i / N, \quad q = 0, 1, 2, \ldots \]

위 방정식에서 \( M(r) \)은 주어진 데이터나 기하학적 패턴을 한 변의 길이가 \( r \)인 \( m \)차원(\( m \geq 1 \))의 단위 블록으로 덮기 위해 필요한 개수, \( p_i \)는 \( r \)변의 단위 블록에 데이터나 패턴의 치수가 갖게 되는 확률을 의미한다. (여기서 \( N \)은 데이터에 포함된 독립된 포인트의 총 개수, \( N_i \)는 \( r \)변의 단위 블록에 포함된 데이터 포인트의 개수를 의미한다.) 면저 \( q = 0 \)인 경우

\[ D_0 = \lim_{r \to 0} \frac{\log M(r)}{\log(1/r)} \]

이므로 \( D_0 \)는 앞서 살펴본 박스개수세기 방법에서 얻어지는 프랙탈 차원(즉, 용량 차원 capacity dimension 혹은 하우스트로 
프 차원)과 같다. \( q \geq 1 \)인 경우 개별 하이퍼큐브가 덮고 있는 데이터 포인트의 분율 fraction, 즉, 정보량 information이 중 
요해지는데 \( D_1 \)은

\[ D_1 = \lim_{r \to 0} \lim_{q \to 1} \frac{1}{q-1} \frac{\log \sum_{i=1}^{M(r)} p_i^q}{\log r} = \lim_{r \to 0} \frac{\sum_{i=1}^{M(r)} p_i \log p_i}{\log(1/r)} \]

으로 표현된다. \( D_1 \)은 ‘정보 차원 information dimension’이라고 부르기도 하는데 이는 위식의 분자, 즉,

\[ S = - \sum_{i=1}^{M(r)} p_i \log p_i \]

가 사는 C. Shannon의 정보 엔트로피 information entropy에 해당하는 양이기 때문이다. 카폴란-요크 추측 Kaplan-Yorke 
conjecture에 따르면 정보 차원은 특히 동역학계의 라프노프 차원 Lyapunov dimension에 근접할 것이며, 즉, 정보 차원은
동역학계의 실질적인 프랙탈 차원에 근접할 것임을 알려져 있다[5]. \( q = 2 \)인 경우의 레니 차원, \( D_2 \)는

\[
D_2 = \lim_{{r \to 0}} \frac{\log C(r)}{\log r}, \quad C(r) = \sum_{{i=1}}^{M(r)} p_i^2
\]

그림7 하스트 지수 \( H \)가 주어졌을 때 인공적으로 생성된 2D 프랙탈 표면의 \( D_2 \) 결정. 2D fractional Brownian motion의 프랙탈 차원을 응용 차원으로 측정한다. 정보 차원, Capacity dimension \( F_{D_{cap}} \) 정

정보 차원, Information dimension \( F_{D_{inf}} \) 상관 차원, Correlation dimension \( F_{D_{cor}} \) 으로 측정한 결과.

그림7 하스트 지수 \( H \)가 주어졌을 때 인공적으로 생성된 2D 프랙탈 표면의 \( D_2 \) 결정. 2D fractional Brownian motion의 프랙탈 차원을 응용 차원으로 측정한다. 정보 차원, Capacity dimension \( F_{D_{cap}} \) 정

정보 차원, Information dimension \( F_{D_{inf}} \) 상관 차원, Correlation dimension \( F_{D_{cor}} \) 으로 측정한 결과.

라고 표현할 수 있으며 이를 상관 차원 correlation dimension이라 한다. 위 식에서 \( C(r) \)은 상관합 correlation sum이라고 부르며 해석적으로 이를 계산할 방법이 없으므로 수치해석적 방법이 필요하다. 그 중 대표적인 방법은 Grassberger와 Proccacia가 개발한 방법으로, 샘플링한 데이터들의 체비세프 거리Chebyshev distance를 계산하여 이를 진화시키는 해비사이드 함수Heaviside function의 변수로 삼아 상관합, 나아가 상관 차원을 계산하는 방법이다 [6]. 수학적 정의로부터 예상할 수 있다시피, 일반적으로 \( q \)가 커질수록 프랙탈 차원은 조금씩 감소하는 경향, 즉, \( D_0 \geq D_1 \geq D_2 \)의 관계가 관찰된다. 이들 레니 엔트로피 기반의 프랙탈 차원을 이용하여 \( 1 \)차원뿐 아니라 자기 NavLink을 갖는 2차원 이상의 프랙탈 패턴의 차원을 측정할 수 있다.

그림7 하스트 지수 Hurst exponent \( H \)가 주어졌을 때 생성된 2차원 자기NavLink 자형인 프랙탈 브라운 운동 패턴의 \( D_2 \) 결정. 2D fractional Brownian motion pattern의 프랙탈 차원을 측정한 결과를 나타낸다. 하스트 지수는 데이터에서 샘플링한 일부 데이터의 변동 범위가 샘플 크기에 대해 역함수 관계를 보일 때, 그 지수를 의미하는 통계 지표로서, 보통 프랙탈 브라운 운동 패턴의 복잡도를 측정하는 1차적인 지표로 활용된다. 이에 대해서는 2부에서 더 자세히 설명할 것이지만, 일반적으로 프랙탈 브라운 운동 패턴에 자기NavLink 특성이 있는 경우,

\[
F_{D_{haus}} = 1 + D_{EUC} - H
\]
의 관계를 보인다. 그림에서 볼 수 있다시피 허스터 지수로 예측되는 이론적인 프랙탈 차원, 즉 $FD_{\text{haus}} = 3 - H$와 비교할 때 세 측정 지표는 이론적 예측값과 비슷한 값을 보여주고 있다. 다시 차이는 있으나 세 지표가 거의 같은 양상으로 생성된 프랙탈 지형의 차원 값을 보여주는 것을 알 수 있다.

4. 프랙탈 패턴의 생성

패턴의 프랙탈 차원을 계산할 수 있다면 반대로 주어진 프랙탈 특성을 만족시키는 기하학적 형태, 물리적인 대상, 혹은 추상적인 데이터를 인공적으로 생성하는 것도 가능할 것이다. 먼저 수학적으로는, 복소평면 상에서 평면의 각 점이 발산하는 속도를 표현한, 시간매개형 프랙탈 패턴Escape-time fractals 생성 방법이 있다. 이에 대한 유명한 사례로, 복소 공간에서 시간매개형 재귀 사상Recursive mapping으로 생성되는 망델브로 집합이나 칼리아 집합이 있다. 여기서 말하는 재귀 사상이란 추상화된 함수 개념으로서, 간단하게 말해 바로 이전에 생성된 계의 산출값이 다음 계계에서 다시 계의 입력값이 되는 수학적 구조다. 19세기 후반~20세기 초반에 줄영문 유명한 푸앵카레-벤디슨 정리Poincaré-Bendixson theorem에 따르면, 연속 시공간 상에서는 3차원 이상의 동역학계에서만 혼돈계가 발생할 수 있다. 따라서 2차원 이하의 동역학계에서는 연속 시간 상에서의 시간매개형 프랙탈 패턴이 나타날 수 없지만[7], 이산 공간에서라면 1차원 동역학계에서도 프랙탈 패턴이 발생될 수 있다. 망델브로 집합을 예로 들어 보자. 복소수

$$c = r + is, \quad r, s \in \mathbb{R}$$

를 계의 파라미터로 갖는 복소 사상은

$$z_{n+1} = f(z_n) = z_n^2 + c, \quad z_0 = 0$$

이라고 쓸 수 있고, 복소수 $z_n$를 실수부 $x_n$와 허수부 $y_n$으로 분리하면

$$x_{n+1} = x_n^2 - x_n^2 + r$$
$$y_{n+1} = 2x_n y_n + s$$

로 풀어 쓸 수 있다. 이때 주어진 복소수 파라미터 $c$에 대해 되풀이 순서인 $n$이 충분히 커질 때 $z_n$의 점대값이 발생하지 않는다면 $c$는 망델브로 집합의 원소가 된다. $r$은 $x$축, $s$는 $y$축에 배치하면 망델브로 집합의 원소 $c$의 분포를 복소평면 상에 2차원 이미지로 시각화할 수 있고, 각 $c$에 대해 $z_n$이 수렴하는 속도를 색깔별로 구분하면 [그림8-1]처럼 잘 알려진 망델브로 집합의 자기다짐을 패턴이 관찰된다.
홍미롭게도 [그림8-1]처럼 망델브로 집합 패턴의 스케일을 네 배로 계속 확대해도 원래의 자기작용물들이 유지되는 척도불변성을 관찰할 수 있다.

참고로 1991년 일본의 수학자 시시쿠라M. Shishikara가 수치해석 방법을 사용해 망델브로 집합의 경계선([그림8-2] 참조)의 프랙탈 차원은 정확히 2임을 증명하였다. (사실 이 때문에, 망델브로 집합의 경계선 패턴은 비치연수의 프랙탈 차원을 가지지 않는다는 말력에서는, 프랙탈에 속하지 않는다고 볼 수 있다.)[8] 이는 망델브로 집합의 경계선도 페아노 곡선이나 헐버트 곡선 같은 과간충전곡선(명변충전곡선)에 속함을 의미하는 것이기도 하다. 만약 \( c \)가 고정되었다면 변수 \( z_n \), 복소 공간에서 어떤 경로를 보일까? 그것은 초기값 \( z_0 \)가 어떤 값인지에 따라 바뀐다. 즉, 초기값의 실수부를 \( x \)축으로 허수부를 \( y \)축으로 놓고 발산하지 않는 \( z_0 \)의 분포를 복소 평면 상에서 2차원 이미지로 시각화할 수 있으며, 이를 점리아 집합이라고 한다. 망델브로 집합과 마찬가지로 [그림9]의 점리아 집합에서도 초기값이 수렴하는 속도를 표현할 수 있다. 홍미로운 부분은 \( c \)가 망델브로 집합의 원소인지는 여부가 점리아 집합의 프랙탈 패턴 특성을 결정한다는 것이 다. \( c \)가 망델브로 집합에 속할 경우 점리아 집합은 비교적 단순한 패턴을 보이는데 이들의 공통점은 경계선 안쪽 영
역이 모두 채워져 있다는 것이고 이는 프랙탈 차원이 1과 2 사이라는 것을 의미한다. 반대로 \(c\)가 망델브로 집합에 속하지 않을 경우, 쥐리아 집합은 복잡한 패턴을 보이는데 흥미롭게도 이들의 프랙탈 차원은 망델브로 집합과 마찬가지로 모두 정확히 2, 즉, 평면충전곡선에 속하게 된다.

그림9 쥐리아 집합\(Julia\ set\)의 다양한 프랙탈 패턴. 왼쪽은 파라미터 \(c\)가 망델브로 집합의 원소가 아닌 경우(이때 프랙탈 차원은 모두 정확히 2), 오른쪽은 원소인 경우(이때 프랙탈 차원은 모두 1과 2 사이에 존재).

복소 공간에서의 사상을 이용하여 망델브로 집합이나 쥐리아 집합 류의 패턴을 생성하는 것은 수학적으로는 반복함수계를 사용하여 자기 då말음꼴 패턴을 생성하는 것에 해당한다. 예를 들어 [그림8-2]의 고사리를 닮은 것 같은 자기 따른음꼴 패턴은

\[
x_{n+1} = ax_n + by_n + e \\
y_{n+1} = cx_n + dy_n + f
\]

같이 단순한 선형 결합과 상수 연산을 활용한 반복함수계를 통해 생성할 수 있다. 물론 어떠한 파라미터 조합을 어떤 순서로 택할 것인지에 따라 패턴의 세부 정보가 달라진다. 예를 들어 [그림10]의 자기 따른음꼴 패턴은 네 가지 서로 다른 파라미터 조합 \((a, b, c, d, e, f)\)을 변가하여이며 적용하여 생성된 결과물이다.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>set1</th>
<th>set2</th>
<th>set3</th>
<th>set4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.00</td>
<td>0.20</td>
<td>-0.15</td>
<td>0.75</td>
</tr>
<tr>
<td>b</td>
<td>0.00</td>
<td>-0.26</td>
<td>0.28</td>
<td>0.04</td>
</tr>
<tr>
<td>c</td>
<td>0.00</td>
<td>0.23</td>
<td>0.26</td>
<td>-0.04</td>
</tr>
<tr>
<td>d</td>
<td>0.16</td>
<td>0.22</td>
<td>0.24</td>
<td>0.85</td>
</tr>
<tr>
<td>e</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>f</td>
<td>0.00</td>
<td>1.60</td>
<td>0.44</td>
<td>1.60</td>
</tr>
<tr>
<td>p</td>
<td>0.10</td>
<td>0.08</td>
<td>0.08</td>
<td>0.74</td>
</tr>
</tbody>
</table>

그림 10 반복함수계IFS를 활용하여 생성된 인공적인 자기답음꼴 고사리 이미지와 되풀이 과정에 활용된 파라미터 조합.

반복함수계는 형식 문법formal grammar을 재귀적으로 반복 적용할 수 있는 컴퓨터 코드를 사용해 만들 수도 있다. 대표적인 형식 문법은 헝가리 식물학자 린덴마이어A. Lindenmeyer가 1960년대에 고안한 L-system으로서, 이는 점과 선을 잇는 방식을 문법 요소로 정리한 알고리즘이다. 식물의 성장 과정에서 보이는 가지치기 패턴에 대한 관심이 묻었던 린덴마이어는 재귀 사양recursion mapping을 이용하여 가능한 단순한 문법의 조합만으로 상대적으로 더 복잡한 자기답음꼴 분기 패턴을 만들 수 있음을 보였다. L-system 같은 형식 문법은 비단 식물의 분기 패턴뿐만 아니라 다양한 자기답음꼴 패턴을 비교적 쉽게 생성할 수 있다는 장점이 있다. 예를 들어 앞서 보았던 코흐의 눈송이Koch curve 패턴은 $F \rightarrow F - F + +F - F$ 같은 문법 규칙으로 [그림 10]의 과정을 거쳐 생성할 수 있다. 이 규칙에서 $F$는 한 점에서 다른 한 점으로 정해진 길이만큼 직진으로 이동하면서 선분을 그려리는 뜻이며, $-\text{ 혹은 } +$는 반시계 혹은 시계 방향으로 정해진 각도만큼 방향을 바꾸라는 뜻이다. 만약 각도를 60°로 설정해 둔다면 (불록 각도는 임의로 설정할 수 있다) 이 규칙의 의미는 [그림 10]처럼

1. 일반 짧은 선분을 그리고
2. 60°만큼 반시계 방향으로 방향을 틀 다음
3. 다시 짧은 선분을 그리고
4. 이번에는 60° + 60° = 120°만큼 시계 방향으로 방향을 틀 다음
5. 다시 짧은 선분을 그리기
6. 다시 60°만큼 반시계 방향으로 방향을 틀 다음
7. 다시 짧은 선분을 그리고

는 뜻이다.


![Fractal trees generated by L-system](image)

그림11 L-system 문법을 활용하여 생성한 프랙탈 나무의 비대칭적 분기 패턴과 관련 문법.
$L - system$ 외에도 자기답음꼴 패턴을 만드는 방법이 있다. 컴퓨터 그래픽에 많이 활용되는 일명 ‘마름모-정사각형 방법(Diamond-square method)’이라는 방법인데(random midpoint displacement fractal, the cloud fractal or the plasma fractal)라고도 부른다. 원리는 간단하다. 먼저 [그림12-1]에 보인 것 같이 바둑판 모양의 격자 grid를 준비한다. 그리드 평면의 한 변에는 점을 $2^N + 1$개 배치한다. 예를 들어 [그림12-1]에는 한 변에 $2^2 + 1 = 5$개씩, 총 $(2^2 + 1)(2^2 + 1) = 25$개 점이 있다. 처음에는 격자의 네 코너를 세웠으로 삼아 미리 정해둔 범위의 값 중 하나를 임의로 선택하여 배치한다.

다음 단계에서는 네 코너를 두 개의 대각선으로 잇고 두 대각선이 만나는 점을 찾은 후, 그 점에 네 코너 값의 평균과 일정한 범위 내에서 선택한 난수 random number를 합친 값을 대응시킨다. 이제 가운데 점의 값이 생성되었으므로 세 번째 그림처럼 각 씩들을 둘러싸인 또 다른 씩들을 격자 평면의 각 변의 중심 점에 대응시키고, 각 중심 점의 값은 앞의 방식과 비슷한 방식으로 생성한다. 이를 반복하면 정사각형과 마름모의 크기가 변 AndAlso며 축소되면서 격자 위에서 계속 생성된다.

![그림12-1 입체도 Fractal landscape 생성을 위한 마름모-정사각형 알고리듬(Diamond-square algorithm)의 원리.](image1)

![그림12-2 마름모-정사각형 알고리듬을 이용한 가상의 자기답음꼴 3차원 지형(산맥) 이미지 생성 과정.](image2)

재귀적 되풀이 과정에서 내삽된 값에 난수가 더해지는 덕분에 새로 생성된 값의 분포에도 불확실성이 가미된다. 이 때문에 [그림12-2]처럼 평 자연스러워 보이는 지형을 만능적으로 만들어내는 것도 가능하다. 예를 들어 이러한 알고리즘은 그러질듯 보이는 산맥을 생성하는데 활용할 수 있는데, 이 경우 각 격자 점에 배치되는 값을 일종의 해별 고도라고 생각할 수 있을 것이다. 당연히 배치되는 값은 음수도 가능하며(이 경우 해수면 아래라고 생각하면 된다) 높이의 변화는 허용되는 난수의 범위에 의해 결정된다. 사실 재귀적 되풀이 과정에서 사용되는 난수의 범위, 즉, 난수 분포의 표준편차를 어떻게 정할 것인지는 3차원 산맥 형태의 핵심 특성 차원을 정하는 변수가 될 수 있다. 2부에서는 이러한 변수 설정이 어떻게 패턴이나 데이터의 복잡도와 연결되는지 보다 자세한 내용을 소개할 것이다.
프랙탈 지형은 비단 2차원 평면에만 국한되지 않으며 3차원 구면 상에서도 구현될 수 있다. 예를 들어 격자 평면 대신 구면조화함수 Spherical harmonic function을 반복적으로 격리되 조화함수의 모드 mode와 변동폭을 선택하는 과정에 프랙탈 차원을 미리 설정하고 조화함수들이 격리되는 부분에 난수를 발생시켜 값을 배정하면, 되풀이 과정이 반복되면서 [그림 13]처럼 가상의 행성 지형도 만들어 낼 수 있다.

그림 13 구면 상 프랙탈 지형 생성 알고리듬을 응용한 가상의 행성 표면 지형 이미지 생성 과정.

인공적으로 생성된 프랙탈 지형이나 이미지를 논할 때 유의할 부분이 있다. 그것은 바로 자연의 생물체 혹은 지형에 자기답음률 특성이 있어도 그것이 완벽한 의미에서의 프랙탈은 아니라는 사실이다. 수학적 의미에서의 프랙탈 패턴은 척도 변화 범위가 무한이지만 실제 자연에서는 척도의 변화가 무한일 수는 없으므로 (즉, 어느 이상 확대하면 자기답 응답이 사라지고), 인공적으로 만들어내는 프랙탈 패턴과는 달리, 자연의 프랙탈 패턴은 준- 자기유사적 프랙탈 Proto-fractal이라고 불린다. 따라서 실제 자연의 이미지는 인공적으로 생성된 프랙탈 나무나 지형의 이미지와의 유사성이 일정 범위의 척도 내에서만 존재한다고 할 수 있다.

1부에서는 자연에서 관찰되거나 인공적으로 만들어진 패턴의 자기답음률과 프랙탈 특성을 분석하는 방법에 대해 소개했다. 2부에서는 일반적 맥락에서 패턴 혹은 데이터의 복잡도를 어떻게 측정할 수 있는지에 대해 엔트로피 위주로 개념을 소개하고, 실제로 어떻게 엔트로피를 활용하여 패턴과 데이터의 복잡도를 측정할 수 있는지 사례를 탐색해 볼 것이다.

참고문헌