A Natural Explanation for the Minimum Entropy Production Principle

Griselda Acosta[1], Eric Smith [2], and Vladik Kreinovich[3]
[1]Department of Electrical and Computer Engineering
[2]Department of Industrial, Manufacturing, and Systems Engineering
[3]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
gvacosta@miners.utep.edu, esmith2@utep.edu, vladik@utep.edu

Abstract

It is well known that, according to the second law of thermodynamics, the entropy of a closed system increases (or at least stays the same). In many situations, this increase is the smallest possible. The corresponding minimum entropy production principle was first formulated and explained by a future Nobelist Ilya Prigogine. Since then, many possible explanations of this principle appeared, but all of them are very technical, based on complex analysis of differential equations describing the system's dynamics. Since this phenomenon is ubiquitous for many systems, it is desirable to look for a general system-based explanation, explanation that would not depend on the specific technical details. Such an explanation is presented in this paper.

1 Formulation of the Problem

Minimum entropy production principle. It is well known that, according to the second law of thermodynamics, the entropy of any closed system – including the Universe as a whole – cannot decrease, it can only either increase or stay the same; see, e.g., [3, 23].

It is somewhat less well known that in many situation, this entropy increase is the smallest possible; this fact is known as the *minimum entropy production principle*. This principle was first formulated in 1945 by a future Nobelist Ilya Prigogine [20]; see also [6, 8, 9, 16, 17, 18, 21].

In contrast to the second law of thermodynamics – which is always true – the minimum entropy production principle is not always valid (see, e.g., [7]), but it is still valid in many practical situations. In particular, it explains why usually, a symmetric state, when perturbed, does not immediately turn into a state with no symmetries at all; usually, some symmetries are preserved – and the more symmetries are preserved, the more frequent are such transitions. For example, when heated, a highly symmetric solid-body state usually does not immediately turn into a completely symmetry-less gas state, it first transitions into a liquid state in which some symmetries are preserved. Sometimes, a solid state does turn directly into gas: e.g., dry ice used to keep ice cream cold goes directly into a gas state without becoming a liquid. However, usually, symmetries are broken sequentially, not all at once. This seemingly simple idea explains many physical phenomena: e.g., it explains the observable shapes of celestial bodies, relative frequency of different shapes, and how shapes change with time; see, e.g., [4, 5, 15].

Challenge: to provide a simple explanation for the minimum entropy production principle. While the principle itself sounds reasonable, all its available derivations are very technical and non-intuitive. Usually, in physics, no matter how complex the corresponding equations, there is a reasonably simple explanation – at least a qualitative one – of the observed phenomena [3, 23]. However, for the minimum entropy production principle, such an explanation has been lacking.

What we do in this paper. In this paper, we provide a general system-based explanation for the ubiquity of the minimum entropy production principle, an explanation which – unlike the existing ones – uses only simple easy-to-understand math.

In this explanation, we will first start with the analysis how complex problems are solved, and then we will explain how this analysis helps explain the minimum entropy production principle.

2 How Complex Problems Are Solved: Reminder and Related Analysis

NP-complete problems: a brief reminder. As we have mentioned, our explanation for the minimum entropy production principle starts not with physics, but with the known fact that in real life, we need to solve complex problems:

- we may need to find a path that leads from point A to point B,
- a mechanic needs to find a way to repair a broken car,
- a medical doctor needs to cure the patients.

In most such problems, it may be difficult to come up with a solution, but once we have a candidate for a solution, we can relatively easily check whether this is indeed a solution. For example, if may be difficult to find a way to repair a car, but if we follow some sequence of actions and the car starts running, we clearly have a solution – otherwise, if the car does not start running, the sequence is not a solution.

The class of all such problems, i.e., problems in which we can, in reasonable ("feasible") time check whether a given candidate for a solution is indeed a solution, is known as the class NP. Within this class, there is a subclass of all the problems that can be *solved* in reasonable time. This subclass is usually denoted by P; see, e.g., [12, 19] for details.

Most computer scientists believe that there are problems that cannot be solved in reasonable time, i.e., that P is different from NP; however, this has never been proven, it is still an open problem. What is known is that in the class NP, there are problems which are as hard as possible – in the sense that all other problems can be reduced to this one. Such problems are known as *NP-complete*.

Historically the first NP-complete problem was the following propositional satisfiability problem for 3-SAT formulas.

- We start with *Boolean* (*propositional*) variables $x_1, \ldots, x_n$, i.e., variables that can take only two values: true (1) and false (0).
- A *literal* is either a variable x_i, or its negation $\neg x_i$.
- A *clause* (*disjunction*) is an expression of the type $a \vee b$ or $a \vee b \vee c$, where a, b, and c are literals.
- Finally, a *3-SAT formula* is an expression of the type $C_1 \,\&\, C_2 \,\&\, \ldots, \&\, C_m$, where C_j are clauses.

An example is a 3-clauses formula

$$(x_1 \vee x_2) \,\&\, (\neg x_1 \vee x_2 \vee x_3) \,\&\, (x_1 \vee \neg x_2 \vee \neg x_3).$$

The general problem is:

- given a 3-SAT formula,
- check whether this formula is *satisfiable*, i.e., whether there exist values of the variables that make it true.

How NP-complete problems are solved now. If P≠NP, this means, in particular, that no feasible algorithm is possible that would solve all the instance of the general 3-SAT problem. So, in practice, when only feasible algorithms are possible, we have to use *heuristic* algorithms, i.e., algorithms which do not always lead to a solution.

Many such algorithms start by selecting a literal – i.e., equivalently, by selecting one of the Boolean variables x_i and selecting its truth value. Then, when we substitute this value into the original formula, we get a new propositional formula with one fewer variable. If the original formula was satisfiable and we selected the literal correctly, then the new formula is also satisfiable – and so, by repeating this procedure again and again, we will confirm that the formula is satisfiable (and also find the values of the variables x_i that make the formula true).

Which literal should we select? In general, a satisfying 3-SAT formula has several satisfying vectors. For example, by trying all 8 possible combinations of truth values, we can check that the above sample 3-SAT formula has four different solutions: (101), (110), (111), and (010).

By selecting a literal, we restrict the number of solutions, from the original number N to a new – usually smaller – number $N' \leq N$. A priori we do not know which vector of Boolean values are solutions, all 2^n such vectors are equally probable to be a solution. Thus, the more vectors remain, the higher the probability that by this restriction we do not miss a solution. It is therefore reasonable to select a literal for which the estimated number of satisfying vectors is the largest possible; see, e.g., [1, 2, 10, 11] and references therein.

For a general 3-SAT formula, the expected number of solutions can be estimated, e.g., as follows:

- a formula $a \vee b$ is satisfied by 3 out of 4 combinations of the values (a, b) (the only combination which does not make this formula true is $a = b =$ false); thus, the probability that this clause will be satisfied by a random Boolean vector is 3/4;

- a formula $a \vee b \vee c$ is satisfied by 7 out of 8 combinations of the values (a, b, c) (the only combination which does not make this formula true is $a = b = c =$ false); ; thus, the probability that this clause will be satisfied by a random Boolean vector is 7/8.

It is difficult to take into account correlation between the clauses, so, in the first approximation, we can simply assume that the clauses are independent, and thus, the probability that a random vector satisfies the formula is equal to the product of the corresponding probabilities – and the number of satisfying vectors can be estimated if we multiply the overall number 2^n of Boolean vectors of length n by this probability.

For example, for the above 3-SAT formula, the corresponding probability is $(3/4) \cdot (7/8) \cdot (7/8)$, and the estimates number of satisfying Boolean vectors is $(3/4) \cdot (7/8) \cdot (7/8) \cdot 2^3 \approx 4.6$. In this formula, we have three variables, so we have six possible literals. Which one should we select?

- if we select x_1 to be true, then the first and the third clauses are always satisfied, and the formula becomes $\neg x_2 \vee \neg x_3$; here, the estimated number of solutions is $(3/4) \cdot 2^2 = 3$;

- if we select a literal $\neg x_1$, i.e., we select x_1 to be false, then the second clause is satisfied, and the formula becomes $x_2 \,\&\, (\neg x_2 \vee \neg x_3)$; here, the estimated number of solutions is $(1/2) \cdot (3/4) \cdot 2^2 = 1.5$;

- if we select a literal x_2, then the formula becomes $x_1 \vee \neg x_3$; here, the estimated number of solutions is $(3/4) \cdot 2^2 = 3$;

- if we select a literal $\neg x_2$, then the formula becomes $x_1 \,\&\, (\neg x_1 \vee x_3)$; here, the estimated number of solutions is $(1/2) \cdot (3/4) \cdot 2^2 = 1.5$;

- if we select a literal x_3, then the formula becomes $(x_1 \vee x_2) \,\&\, (x_1 \vee \neg x_2)$; here, the estimated number of solutions is $(3/4) \cdot (3/4) \cdot 2^2 = 2.25$;

- finally, if we select a literal $\neg x_3$, then the formula becomes

$$(x_1 \vee x_2) \,\&\, (\neg x_1 \vee x_2);$$

  here, the estimated number of solutions is $(3/4) \cdot (3/4) \cdot 2^2 = 2.25$.

The largest estimate of remaining Boolean vectors is when we select x_1 or x_2. So, on the first step, we should select either the literal x_1 or the literal x_2. One can check that in both cases, we do not miss a solution (and in each of these cases, we actually get 3 solutions, exactly the number that we estimated).

General case. The same idea is known to be efficient for many other complex problems; see, e.g., [10]. For example, a similar algorithm has been successfully used to solve another NP-complete problem: a discrete optimization *knapsack problem*, where:

- given the resources $r_1, \ldots, r_n$ needed for each of n projects, the overall amount r of available resources, and the expected gain $g_1, \ldots, g_n$ from each of the projects,

- we need to select a set of projects $S \subseteq \{1, \ldots, n\}$ which has the largest expected gain $\sum\limits_{i \in S} g_i$ among all the sets that we can afford, i.e., among all the sets S for which $\sum\limits_{i \in S} r_i \le r$.

The corresponding algorithms are described, e.g., in [14, 22].

In general, it is important to keep as many solution options open as possible. In decision making, one of the main errors is to focus too quickly and to become blind to alternatives. This is a general problem-solving principle which the above SAT example illustrates very well.

3 How This Analysis Helps Explain The Minimum Entropy Production Principle

How is all this related to entropy. From the physical viewpoint, entropy is proportional to the logarithm of the number of micro-states forming a given macro-state; see, e.g., [3, 23]. In the case of the SAT problems, micro-states are satisfying vectors, so the number of micro-states is the number of such vectors. Similarly, in other complex problems, solution options are micro-states, and the number of micro-states is the number of such options.

As we solve each problem, the number of states decreases – but decreases as slowly as possible. Thus, the entropy – which is the logarithm of the number of states – also decreases, but decreases as slowly as possible, at the minimal possible rate.

So, if we consider the dependence of entropy on time, then, in the backward-time direction (i.e., in the direction in which entropy *in*creases), this increase is the smallest possible.

How is all this related to physics. At first glance, the above text may be more relevant for human and computer problem solving than for physics, since at first glance, nature does not solve problems.

However, in some reasonable sense it does; let us explain this. Traditionally, physical theories – starting from Newton's mechanics – have been formulated in terms of differential equations. In this formulation, there is no problem to solve: once we know the state at a given moment of time, we can compute the rate at which each variable describing the state changes with time. This computation may be tedious, may require a lot of computation time on a high-performance computer, but it does not constitute a challenging NP-complete problem.

At present, however, the most typical way to describe a physical theory is in the form of a variational principle, i.e., in the form of an objective function whose optimization corresponds to the actual behavior of the physical systems; see, e.g., [3, 13, 23]. This formulation is especially important if we take quantum effects into account:

- while in non-quantum physics, optimization is exact and is just another equivalent form of describing the corresponding differential equations,
- in quantum physics, optimization is approximate: a quantum system tries to optimize, but its result is close to (but not exactly equal to) the corresponding optimum.

In this formulation, what nature does *is* solving the complex optimization problem: namely, trying to optimize the value of the corresponding functional.

We therefore expect to see the same pattern of entropy changes as in general problem solving: in the direction in which entropy is increasing, this increase is the smallest possible.

Increasing entropy is exactly how we determine the direction of physical time. For example:

- if we see a movie in which a cup falls down and break, we understand that this is exactly the time direction, while
- if we see the same movie played backward, when the pieces of a broken cup mysteriously come together to form a whole cup, we realize that we saw this movie in reverse.

From this viewpoint, the above statement means that in the forward-time direction – i.e., in the direction in which entropy increases – the rate of the entropy increase is the smallest possible.

We thus have a natural systems-based explanation for the minimum entropy production principle.

Acknowledgments

This work was supported by the Institute of Geodesy, Leibniz University of Hannover. It was also supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence). This paper was partly written when V. Kreinovich was a visiting researcher at the Leibniz University of Hannover.

References

[1] O. Dubois, "Counting the number of solutions for instances of satistiability", *Theoretical Computer Science*, 1991, Vol. 81, pp. 49–64.

[2] O. Dubois and J. Carlier, "Probabilistic approach to the satisfiability problem", *Theoretical Computer Science*, 1991, Vol. 81, pp. 65–85.

[3] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[4] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", *International Journal of Theoretical Physics*, 1997, Vol. 36, No. 4, pp. 1009–1020.

[5] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: geometry explains shapes of celestial bodies", *Geombinatorics*, 1997, Vol. VI, No. 4, pp. 125–139.

[6] P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations*, Wiley-Interscience, London, 1971.

[7] W. T. Grandy Jr., *Entropy and the Time Evolution of Macroscopic Systems*, Oxford University Press, Oxford, 2008.

[8] E. T. Jaynes, "The minimum entropy production principle", *Annual Review of Physical Chemistry*, 1980, Vol. 31, pp. 579–601.

[9] M. J. Klein and P. H. E. Meijer, "Principle of minimum entropy production", *Physical Review*, 1954, Vol. 96, pp. 250–255.

[10] V. Kreinovich, "S. Maslov's Iterative Method: 15 Years Later (Freedom of Choice, Neural Networks, Numerical Optimization, Uncertainty Reasoning, and Chemical Computing)", In: V. Kreinovich and G. Mints (eds.), *Problems of Reducing the Exhaustive Search*, American Mathematical Society, Providence, Rhode Island, 1997, pp. 175–189.

[11] V. Kreinovich and O. Fuentes, "High-Concentration Chemical Computing Techniques for Solving Hard-To-Solve Problems, and Their Relation to Numerical Optimization, Neural Computing, Reasoning Under Uncertainty,

and Freedom Of Choice", In: E. Katz (ed.), *Molecular and Supramolecular Information Processing: From Molecular Switches to Logical Systems*, Wiley-VCH, Wienheim, Germany, 2012, pp. 209–235.

[12] V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.

[13] V. Kreinovich and G. Liu, "We live in the best of possible worlds: Leibniz's insight helps to derive equations of modern physics", In: R. Pisano, M. Fichant, P. Bussotti, and A. R. E. Oliveira (eds.), *The Dialogue between Sciences, Philosophy and Engineering. New Historical and Epistemological Insights, Homage to Gottfried W. Leibnitz 1646–1716*, College Publications, London, 2017, pp. 207–226.

[14] V. Kreinovich and S. Shukeilo, "A new probabilistic approach to the knapsack problem", *Proceedings of the Third USSR All-Union School on Discrete Optimization and Computers, Tashtagol, Russia, December 2–9, 1987*, Moscow, 1987, pp. 123–124 (in Russian).

[15] S. Li, Y. Ogura, and V. Kreinovich, *Limit Theorems and Applications of Set Valued and Fuzzy Valued Random Variables*, Kluwer Academic Publishers, Dordrecht, 2002.

[16] R. Livi and P. Politi, *Non-Equilibrium Statistical Physics: A Modern Perspective*, Cambridge University Press, Cambridge, UK, 2017.

[17] C. Maes and K. Netočný, "Minimum entropy production principle from a dynamical fluctuation law", *Journal of Mathematical Physics*, 2007, Vol. 48, Paper 053306.

[18] I. N. Martyushev, A. S. Nazarova, and V. D. Seleznev, "On the problem of the minimum entropy production in the nonequilibrium stationary state", *Journal of Physics A: Mathematical and Theoretical*, 2007, Vol. 40, No. 3, pp. 371–380.

[19] C. H. Papadimitriou, *Computational Complexity*, Addison-Wesley, San Diego, 1994.

[20] I. Prigogine, "Modération et transformations irréversibles des systémes ouverts", *Bulletin de la Classe des Sciences, Académie Royale de Belgique*, 1945, Vol. 31, pp. 600—606.

[21] I. Prigogine, *Etude Thermodynamique des phénoménes irréversibles*, Desoer, Liége, France, 1947.

[22] S. Shukeilo, *A New Probabilistic Approach to the Knapsack Problem*, Leningrad Electrical Engineering Institute (LETI), Master's Thesis, 1988 (in Russian).

[23] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.