Possible Geometric Explanations for Basic Empirical Dependencies of Systems Engineering

Francisco Zapata[1] and Vladik Kreinovich[2]
[1]Research Institute for Manufacturing & Engineering Systems
[2]Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
fazg74@gmail.com, vladik@utep.edu

Abstract

In this paper, we provide possible geometric explanation for basic empirical dependencies of system engineering: that a properly designed system should have no more than 7 ± 2 elements reporting to it, and that the relative cost of correcting a defect on different stages of the system's life cycle is 3–6 on the second (design) stage, 20–100 on the third (development) stage, and 250–1000 on the fourth (production and testing) stage.

1 Basic Empirical Dependencies of Systems Engineering

Systems Engineering. According to [1], "Systems Engineering (SE) is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceedings with design synthesis and system validation while considering the complete problem: operations, cost and schedule, performance, training and support, test, manufacturing, and disposal. SE considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs."

Basic empirical dependencies of systems engineering. The first numerical dependency listed in [1] is the "rule of thumb" describing system hierarchies: at any level of the system hierarchy, a properly designed system should have no more than 7 ± 2 elements reporting to it. To be more precise:

- a system with fewer than $5 = 7 - 2$ elements does not need any hierarchical structure, and

- in a system with more elements, there needs to be a hierarchical structure, so that each system has between 5 to 9 subsystems reporting to it.

The second numerical dependency listed in [1] described a relative cost of correcting a defect on different stages of the system life-cycle. If we take the average cost of correcting a defect at the concept stage as 1, then:

- at the first *concept* stage, the cost is 1;
- at the second *design* stage, the cost is 3–6;
- at the third *development* stage, the cost is 20–100;
- finally, at the fourth stage of *production and testing*, the cost is 500–1000 times larger than at the first stage.

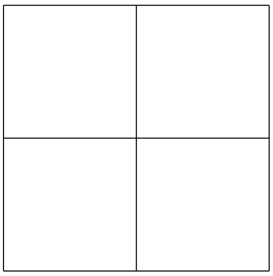
Need to explain these empirical dependencies. The above empirical dependencies hold in many different systems, which seems to indicate that there should be some fundamental explanation for these dependencies.

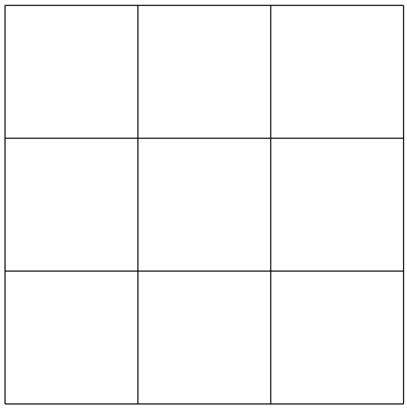
What we do in this paper. In this paper, we provide a geometric analysis of this problem, and we show that both dependencies, we can come up with possible simple geometric explanations.

2 Rule of Thumb Related to System Hierarchy: A Possible Geometric Explanation

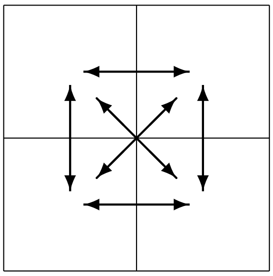
System hierarchy: a simple geometric interpretation. One of the main examples of a complex system is a manufacturing factory. Usually, the factory consists of several buildings. These building are usually rectangular in shape. For simplicity, we can ignore the difference between different sides of each building and between sizes of different buildings – and thus, assume that each building has the shape of a square (as we will see, our explanation does not depend on this simplifying assumption, but the assumption makes the geometry easier to understand).

These building are usually located close to each other. From this viewpoint, a factory consists of several adjacent square blocks from a grid. Two examples of such grids are given below.



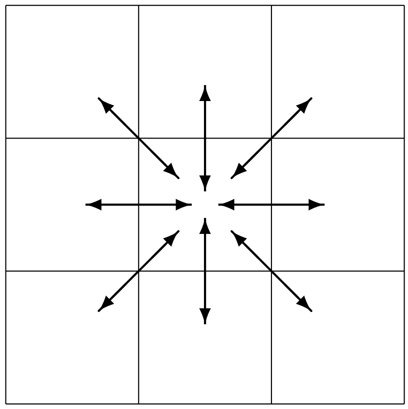


When is a hierarchical structure needed? If any two blocks can directly communicate with one another, there is no need for an intermediary, so there is any need for a hierarchical structure. The two blocks can directly communicate with each other if they are adjacent to each other, i.e., if they are direct neighbors – horizontally, vertically, or diagonally:



On a plane, we can place up to four blocks so that each of them directly communicates with all others. If we have more than four blocks, we need an intermediary, i.e., in other words, we need a hierarchical structure: instead of subsystems always communicating directly with one another, they sometimes communicate with the "central" system, and this central system forwards the communication to the recipient.

How many subsystems can be served by a central system? In our geometric model, a "central" block can successfully communicate with all its neighbors – horizontally, vertically, or diagonally:



The central system and its (direct) neighbors form nine subsystems. This is the maximum number of subsystems that a central system can serve. Thus, we arrive at the following conclusion.

Conclusion. Up to four subsystems can function without a hierarchical structure. If we have five or more subsystems, we need a hierarchical structure. One central system can handle up to nine subsystems. Thus, a central system can handle between five and nine subsystems. This is exactly the 7 ± 2 rule. So, *our simple geometric model provides a possible explanation for the empirical* 7 ± 2 *rule.*

Comment. Of course, the 7 ± 2 rule is only a heuristic approximate rule, it reflects what happens in general, but there are special cases when a central system can successfully handle more than $7 + 2 = 9$ subsystems (or, vice versa, handling even $7 - 2 = 5$ subsystems is sometimes too difficult for a single central system). This possibility is also in line with our geometric interpretation.

For example, if instead of squares, we consider triangles or hexagons or 3-D cubes, we can have more than 9 neighbors – and thus, we can potentially have a central system directly controlling more than 9 subsystems.

Vice versa, for circles of equal size, we can have at most three circles each of which neighbors every other. So, in this case, a configuration of 4 circular subsystems would already require a central unit.

3 Rule of Thumb Related to Relative Cost of Correcting a Defect at Different Stages of the System Life-Cycle

Defects are inevitable. In practice, complex systems are practically never produced in a flawless way: perfect design followed by a perfect implementation. Real-life systems have defects, and correcting these defects is an important part of the system design and testing. Most defects can traced to a single component of the corresponding system. Thus, correcting this defect means correcting this particular component.

Components influence each other. In general, however, components are related to each other. As a result, if we make change to one component, we often need to make changes to "neighboring" components, i.e., components which directly interact with this one. Otherwise, if we make changes to one component, and do not change neighboring components, this may cause the system to fail – and it is actually one of the known sources of system failure.

Reduction to a geometric problem: main idea. The more such neighboring components are involved, the more effort is needed to correct the original defect. To estimate this cost, we can thus estimate the corresponding number of neighbors. In this sense, estimating the cost can be reduced to a geometric problem.

Different stages of the system life-cycle: towards a geometric description. We consider different stages of system design and testing, i.e., different stages of the system life-cycle. The further we go into the life-cycle, the more realistic (and thus, more complex) is the resulting description of the system components are their interaction.

In qualitative terms, two things happen if we switch to a more realistic description:

- First, we add more parameters to the components' description. In the beginning, we may only take into account a few most important parameters. However, as we make our description more and more realistic, we need to take into account other parameters as well. In geometric terms, the number of parameters corresponds to the dimension of the corresponding state space. In these terms, as we move to more realistic descriptions of different components, we increase the dimension of the state space.

- Second, we need to more accurately take into account the dependence between the system's components – as a result of which we have to take into account the effect of components which are further away (and whose influence we could previously safely ignore).

Different stages of the system life-cycle: towards a quantitative description. As we have mentioned, the complexity of repairing a defect is proportional to the number of components that need to be modified – i.e., to the number of components in the appropriate neighborhood of the component which turned out to be defective. This number of components is proportional to the volume of this neighborhood.

This volume is determined by the dimension d and by the radius r of this neighborhood – i.e., by the distance from the found-to-be-defective component at which the nearby component is affected by the discovered defect and therefore, needs to be modified. In each of d dimensions, we have a neighborhood of size

$2r$. The overall volume of this neighborhood can be estimated as a product of linear sizes corresponding to different dimensions, i.e., as $(2r)^d$.

How is the dimension d increasing as we progress from one stage to another? An increase in dimension means that we take into account an additional quantity. This is not just a numerical change, this is clearly a qualitative change. We can thus say that every time the dimension increases, we get to a new stage in the system life-cycle.

In this approximate description, we start with the simplest case $d = 1$ when a single parameter is sufficient; at the next stage, we have $d = 2$ parameters, etc. Thus, in this description, sequential stages can be identified by the number of parameters d needed for an adequate description of components:

- on the 1st stage, we need $d = 1$ parameter;
- on the 2nd stage, we need $d = 2$ parameters;
- ...
- on the d-th stage, we need d parameters.

As we progress from each stage to the next one, the corresponding radius r increases. In other words, the radius $r(d)$ corresponding to the stage with dimension d is an increasing function of d. In the first approximation, we can expand the dependence of r on d in Taylor series and keep only the linear term in the expansion. As a result, we get a linearized dependence $r(d) = c \cdot d$, for some constant c. The linear size $2r$ has the form $2r(d) = a \cdot d$, where we denoted $a \stackrel{\text{def}}{=} 2c$.

The volume of the corresponding neighborhood is thus approximately equal to $(2r)^d = (a \cdot d)^d$. The number of components that need to be corrected and, therefore, the cost $C(d)$ of correcting a defect at stage d, are proportional to this volume, i.e., have the form $C(d) = C \cdot (a \cdot d)^d$ for some constant d.

In the usual analysis, the cost of correcting a defect at the first stage is taken as a unit. In other words, instead of the *actual* cost $C(d)$, we consider the *relative* cost

$$R(d) \stackrel{\text{def}}{=} \frac{C(d)}{C(1)} = \frac{C \cdot (a \cdot d)^d}{C \cdot a} = a^{d-1} \cdot d^d.$$

Resulting formulas.

- On the first stage $d = 1$, we get $R(1) = 1$.
- On the second stage $d = 2$, we get $R(2) = 2^2 \cdot a = 4a$.
- On the third stage, we get $R(3) = 3^3 \cdot a^2 = 27a^2$, and
- on the fourth stage, we get $R(4) = 4^4 \cdot a^3 = 256a^3$.

Comparison with the known empirical ranges. Already for the simplest value $a = 1$, we get $R(1) = 1$, $R(2) = 4$, $R(3) = 27$, and $R(4) = 256$, which is close to the desired ranges 1, 3–6, 20–100, and 500–1000. Specifically:

- for Stages 2 and 3, our estimates are closer to the lower end of these ranges, and
- our estimate goes below the empirical range for Stage 4.

Thus, to get a better fit, we need to slightly increase a. This indeed enables us to fit within the empirical ranges. For example:

- For $a = 1.2$, we get $R(1) = 1$, $R(2) \approx 5$ $R(3) \approx 40$, and $R(4) \approx 450$.
- For $a = 1.3$, we get $R(1) = 1$, $R(2) \approx 5$, $R(3) \approx 45$, and $R(4) \approx 550$.
- For $a = 1.5$, we get $R(1) = 1$, $R(2) = 6$, $R(3) \approx 60$, and $R(4) \approx 900$.

Conclusion. *The empirical ranges for the cost of correcting a defect at different stages of the system life-cycle can also be geometrically explained.*

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber- ShARE Center of Excellence) and DUE-0926721.

References

[1] C. Haskins (ed.), *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*, International Council on Systems Engineering INCOSE, Document INCOSE-TP-2003-002-03.2.2, 2011.