Enformation Theory — Part IV:
—
From *Stasis* back to *Chaos*★

Heikki Hyötyniemi

Aalto University School of Electrical Engineering
Department of Automation and Systems Technology
P.O. Box 15500, 00076 Aalto, Finland

Abstract. The *neocybernetic model structure* determines stiffening *constraints* even though living things are better characterized in terms of their *freedoms*. The ever continuing evolutionary optimization finally exhausts the freedoms, and the compromized robustness results in *fractal collapses*. As a consequence, there is a lesson to be learned: *everything goes in cycles.*

1 Introduction

In Parts II and III, it was observed that a system constructs *models* of its environment to implement *controls*; the former was the system's view of what happens, and the latter was the view of the environment. Thus, the coupling between the system and its environment is characterized by *circular causal chains*.

The optimized data mappings between the environment, or the vectors u and $\bar{u}$, and the system, or the vector $\bar{x}$, are captured in steady state by the formulas

$$\bar{x} = \Phi^{\mathrm{T}}\,\bar{u} = \mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}^{-1}\mathcal{E}\left\{\bar{x}u^{\mathrm{T}}\right\}\,u \tag{1}$$

and

$$\hat{u} = \Phi\,\bar{x} = \mathcal{E}\left\{u\bar{x}^{\mathrm{T}}\right\}\mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}^{-1}\,\bar{x}, \tag{2}$$

so that the residual is

$$\bar{u} = u - \hat{u} = u - \Phi\,\bar{x}. \tag{3}$$

For more thorough explanation, see Parts I, II, and II, and [1].

Is this all there is to say about what happens in systems, does this description capture the *essence*? — It can be claimed that the most interesting things do not take place among the *constraints* but among the remaining *freedoms*; in this paper, an *outside view* into neocybernetic systems is presented when the whole loop is seen from above as a fixed structure.

★ Submitted to *The Eleventh International Symposium on Intelligent Data Analysis* (IDA 2012), 25–27 October, 2012, in Helsinki, Finland; Part 4 of 5

2 Pattern view

It turns out that the behaviors in a cybernetic system can be studied also in terms of a quadratic *energy function*

$$J(x) = \frac{1}{2}\, x^{\mathrm{T}} \mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\} x - x^{\mathrm{T}} \mathcal{E}\left\{\bar{x}\bar{u}^{\mathrm{T}}\right\} u. \tag{4}$$

Starting from an arbitrary internal system state x, the system finds the appropriate optimal state $\bar{x}$, as given in (1), when the criterion becomes minimized. At this point the internal and external tensions, as determined by the vectors x and u, become exactly balanced; $J(x)$ can be seen as defining some kind of a *tension field* in the space of variables. It is interesting that the original enformation maximization task has been changed to minimization.

The criterion (4) also connects the time scales: it can be used for determining $\bar{x}$ (when minimizing $J(x)$, even for unoptimal $\mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}$ and $\mathcal{E}\{\bar{x}\bar{u}^{\mathrm{T}}\}$), and, on the higher level, for determining the model itself (when minimizing $\mathcal{E}\{J(\bar{x})\}$).

The criterion can be rewritten in other forms, too. Assuming that in steady state u and the estimate $\Phi\bar{x}$ match each other, one can write the energy function as

$$J(x) = \frac{1}{2}\, (u - \Phi x)^{\mathrm{T}}\, \Phi\mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\Phi^{\mathrm{T}}\, (u - \Phi x)\,. \tag{5}$$

Here the *reconstruction error* $u - \Phi x$ has a central role; it is being minimized. The reconstruction, or estimate Φx, can be interpreted so that the *features,* or columns Φ_i, together explain the changing *patterns* in u. The features are weighted by the variables x_i so that their sum maximally explains each individual input pattern. In some other application fields, rather than features, one could speak of *ecological/economical lockers*, etc., that together span the resources in the ecosystem. The matching process between the patterns and the weightings of features is an iteration where the balance between the pattern and its reconstruction is searched for. The model is then a storage for features.

It is not the individual variables u_j that are of interest any more: the neocybernetic system experiences the world as a combination of features. Autonomous abstraction has been taking place. For example, the human low-level visual cortex does not concentrate on individual intensity values but more complex visual hints (like the "strokes", when looking at drawings). It can be said that the original observation pattern is decomposed into a (low-level) internal "perception".

The weighting matrix in (5) deserves a closer look. It emphasizes those directions, where there is plenty of variation; clearly, enformation is now seen as an asset to be taken care of:

$$\Phi\mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}\Phi^{\mathrm{T}} = \mathcal{E}\left\{\hat{u}\hat{u}^{\mathrm{T}}\right\} \tag{6}$$

On the other hand, in *information theory*, and in traditional *identification*, etc., there is the *inverse* of the covariance matrix that is applied for weighting: variation is there seen as noise. The inverse covariance weighting is derived directly from the *maximum likelihood approach* as applied to normally distributed variables; in enformation theory one could perhaps speak of *maximum livelihood* instead! — The emphases are in very different directions, and the resulting interpretations differ as well.

3 Constraints and freedoms

The traditional way to model the world is to use *constraints*. When you have, say, m variables, they span an m-dimensional space. If nothing is known about their values, we have no model; but if there are some couplings among the variables, somehow constraining their behaviors, making them interdependent, this knowledge reduces our uncertainty. Each equation binds sets of variables together, making it possible to make predictions about the unknown ones. For example, the *laws of nature* are typically given in an equation form.

However, in complex systems there are typically huge numbers of elementary constraints, and when the systems develop, there will be even more of them. For example, implementing a new feedback in an industrial process introduces a coupling among variables. Writing the model, or collecting all those constraints together becomes a difficult task; and, specially, using the model is challenging when the constraints may have been defined in different forms (as static or differential equations, etc.).

What is more, the variable couplings become blurred in complex systems. Even if the constraints may be defined in a one-to-one fashion, from one variable to another, over time these effects are reflected in the whole network of constraints. Typically, in a system with all levels of feedbacks, all connections become *pancausal*. What is more, not all constraints are equally compelling, not all are equally relevant. There are no more exact identities, and intractable noise deteriorates the model.

The (linear) constraints are visible in the system data covariance structure: it is those eigenvector directions corresponding to the near-zero eigenvalues that are the most certain invariants in the system. Unfortunately, there are typically not many of them, and proceeding this way, modeling soon becomes very tedious.

Yet, the same dependencies among variables can be characterized also in the opposite way: one can concentrate on the remaining *freedoms*. If there are few freedoms left, the model that is based on freedoms is assumedly *better*, according to *Occam's razor*.

The neocybernetic models concentrate on the freedoms. Each variable $\bar{x}_i$ defines a *degree of freedom* of its own along its feature axis. Whereas constraints represent the rigid structure, freedoms represent the remaining variability therein — and it is here where the enformation is found, changes and variations. Loosely speaking, in monadic terms one can say that the freedom can be seen as the *axis of rotation*, and the corresponding variable x_i is then the "rotation speed" dictated by the outside environmental tensions.

By definition, freedoms escape rigid definitions (constraints). An example of what freedoms can be, is *rules of a game* versus *strategies*. Beginners follow the rules but experts see "through" them, being capable of recognizing *patterns*. This applies to all expertise: novices apply *declarative* reasoning, whereas experts "see it" in an associative manner. Such expertise cannot be explicated.

Even though the freedoms are difficult to explicitly define, they are visible in data as behavioral variations.

Freedoms-orientation means that very different tools are needed as compared to traditional modeling. The keyword is *multivariate statistics*, and specially *principal component analysis* (PCA). Indeed, the neocybernetic algorithm spans the same principal subspace as PCA does; however, the basis axes are rotated towards some kind of "robust components". What is more, the weightings are different: PCA does not (normally) employ such extra weighting as shown in (6).

In Part III it was observed that the coupled system deforms its experienced environment $\bar{u}$, as compared to the original u; it becomes somehow simpler and *homogenized*. The hypothesis here is that *the role of the lower-level systems is to make the world better comprehensible* for the higher-level systems, so that the relevant phenomena are always reducible to simple variables. Then, for example, at the highest level, when studying human decision making, the world is filtered through *values*. And it is always that enformation pursuit in the space of such variables that rules; the claim here is that *the same modeling principles* apply at all levels, and it is always *the same model structures*, trying to capture enformation in the space of freedoms, in physical and in mental systems alike.

It is no more possible to reduce the discussions to elementary variables, as there are now too long chains of intermediate steps; however, qualitatively the behaviors can best be understood in terms of the presented high-level concepts.

4 Mechanisms of evolution

Similarly, there is another thing that characterizes all living: it is evolution. This evolution takes place on all levels, and depending on the time scale, it may look different to human eye. In its simplest form, it is the familiar *Hebbian-style learning* on the high scale that looks like evolution: models get better, and controls get better correspondingly.

The other mechanism is *tightening of controls*: when enformation processes on the lower levels become more effective (because of different kinds of local innovations), this is reflected on the high level as the coupling factors q_i having higher values. After a certain threshold, though, increase in coupling results in *less* enformation being transferred from the environment to the system (this limit is $q_i = 4/\lambda_j$; see Part II).

For example study an example of a specific system where *human acts as an agent*, or *signal carrier*:

> In the perspective of universal life, a *company* is a living entity, where the driving force, or enformation pursuit, is realized in terms of *money*. There are many levels of monads, each worker, for example, constituting one "atom" in the living system; however, the only goal of the system is to increase the money flow. Evolution in the system takes place mainly in the form of *freedom elimination* to reach better control of behaviors.

> The new worker controls, for example, are implemented in the form of more supervision (to model behaviors), and applying this information to implement tighter rules. The goal is to make the "squirrel wheels" run faster; and the same intensification of functions takes place everywhere, under the names transparency, efficiency, and less ado and randomness.

Indeed, the basic mechanism of evolution is to slowly enhance the highest-level models to better capture the available enformation. This means that the observed degrees of freedom are exploited by implementing more efficient controls. When the models are complete, all enformation is eliminated in the observed environment, and what remains is a stiff structure: all freedoms have changed to a monolithic package of constraints because of the completed controls. Putting it boldly, *evolution eats life and excretes fixed dead structures*.

The claim here (again) is that the same principles are applicable at all levels of complex systems, as it is always the same enformation pursuit that determines the emergent structures. This all works also in "ideasphere", whereas the "semiosis" is more challenging then; it takes a human to evaluate the environment and to change the amorphous observations into a concrete more or less numeric form. A prototypical example is *science* as seen as a *memetic system*; whereas science has its autonomous dynamics, again, it is humans that act as *agents of evolution*:

> In science, the capture of environmental information/enformation is seen as an explicit goal, mismatches between theories and the world acting as the "variation" to be captured; in the noble spirit this is indeed seen as the proclaimed exclusive ideal (even though the scientific society of humans, following its own systemic principles, can temporarily distort it). But whenever some dilemma is understood, the tension vanishes. In the Kuhnian spirit, a *paradigm* can be seen as a freedom; when it becomes exhausted, the "standard science" starves, in the balance facing the unexplained — until the *antithesis* is found, where the new seemingly radical interpretations open up an unblemished degree of freedom. Clever researchers select their topics in a very "Hebbian" manner, in the direction where there seems to exist most to exploit, the old paradigms suffering as all "life" has already been squeezed, all feedbacks already being implemented.

If evolution is seen as slow dynamic process, system parameter adaptation being seen as convergence of state variables, one can understand the *saltationistic* evolution steps. Whenever a new degree of freedom is found, the balance of tensions along it is found relatively fast, but the sequence of peaceful balances is interrupted by sudden qualitative leaps from a balance to another.

Where do the fresh degrees of freedom pop up from? New horizons are innovations being related to system's enhanced internal "understanding", when the system semiosis becomes somehow extended, fresh enformation becoming available in terms of new uncorrelated resources.

When looking the evolution as the highest-level dynamic process characterizing the system's behavior, one can have a more complete view of the *feedback structure*.

— The lowest level, or the Hebbian learning, defines a *positive* feedback; this would make the system unstable without the next-level *negative* feedback through the environment. However, at the highest level *the evolutionary feedback is positive* again. When the developed system is no more at the mercy of the environment, the sparsity pursuit changes to extreme centralization. It is no more the overall ecosystem benefit that counts, but the individual gains of single subsystems only. The subsystems have their personal "tailored semiosis", resulting in increasing differentiation and different rules. There is no more balance among the actors, but the explicit competition exhaust the losers finally to extinction. If life were to continue without any limit, some subsystem sooner or later would invent the "winning strategy" of a *cancer cell*.

For too far developed systems, only *full-scale regeneration* helps in the dead ends. Luckily enough, extreme optimization results in loss of robustness ...

5 Faith of adaptive controls

How is the idea of ever continuing evolution, that is so essential in life, compatible with the monadic cycles that were also seen so essential in all life, is there not some contradiction? To understand this, one needs some *engineering intuition*.

> A few decades ago, *adaptive controllers* were seen as a universal solution in the field of control engineering: one does not need to construct exact models for processes to implement their controllers, because adaptive controllers are able to tune the models themselves. For this purpose, they analyze the available signals and identify them. — However, it soon turned out that such controllers did not keep the promises: their behaviors were too unpredictable. The reason for this is that good identification is possible only when the control is not yet good; only then there is relevant information available in the signals. For too good models, there was only noise present, misleading the model adaptation. This means that adaptive controllers behaved in a cyclic manner: after controls were very good, they soon collapsed, and after that the controls gradually started getting better again.

The same problem haunts nature's models: there is pursuit towards optimum, to reach maximum enformation capture, but then there is no more enformation left for constructing the models. — It needs to be noted that this problem does *not apply* to the original neocybernetic structure where there is that regularizing matrix Q, leaving some enformation in the signals and resulting in cautious controls. Optimality and robustness are opposite goals!

On the other hand, if the structures are allowed to stiffen in the optimum, they become fragile. There are disturbances coming from outside the system, and these shake the structures; because of the optimality there is no robustness, and in an extreme case, the structures can collapse altogether, sinking back to chaos. And patch fixes do not help in the long run; the hidden tensions just increase.

However, when seen in the correct way, all noise is just a possibility of fresh enfornation; whenever it becomes understood, or when its internal correlations become detected, it becomes exploitable. In this spirit, also the above disturbances must be seen as a source of higher-level enformation. This higher level is not yet known, and there is not enough knowledge of that level yet. Nature wants to "understand" that enformation to construct a model for it, and to exploit it; and the clever way to gain fresh data for that purpose, technically speaking, is *to excite the world* and *implement some impulse tests* on it. And it is such collapses that maximally excite the hidden dynamics in the world. Indeed, it can be said that *systems are probes in the unknown*, and letting them collapse is *the nature's way of implementing experiments*. When a large number of experiments is abstracted, in the enformation theoretic spirit, individual systems and their destinies are no more visible, and only the underlying principles truly reflecting the world properties remain in the resulting model.

There is an infinite number of different kinds of possible disturbances, and the exact reason for collapse cannot typically be determined; this has to be seen, again, as a benefit, because then new kinds of correlation structures among data can become visible. Nature wants to see it all: the collapses repeat, starting from chaos and ending in chaos over and over again, but always something different there in between; indeed, when seen in the long time perspective, there are *cyclic collapses*. The destiny of all systems seemingly is some kind of *Samsara*, the *wheel of life*, or eternal cycle between birth and death.

Intervals between the successive collapses are more or less constant. This means that there are correlations in the noise along the time axis as observed outside the collapsing system, so that new enformation is created in the world. What is more, complex systems consist of subsystems, and these behave qualitatively in the same periodically pulsating way; variation is produced fractally in all levels of systems. As seen from very far above (as compared to the local time scales), systems are seen as "spectral fingerprints" to be observed by others; *such spectra is systems' way to interact with other systems*. In addition to the *original noise* (some kind of "background noise") main part of the noise is *recycled*. As physical matter comes from recycled stars, observed data comes from recycled enformation, being there ready to be reused.

As examples of the repeating nature of collapses, one can study the faster-scale noise caused by *cell cycles* in biological matter, or *workers* changing in a company; slower-scale noise is caused by the deaths of the whole organisms, or companies. On very long time scales, one can recognize that also whole cultures and ecosystems collapse.

Another interesting example of cyclic bursts and resulting waves is the operation of the *brain*. On the lowest level, the neuron activities discharge rapidly: this is the starting point for something that can finally be measured as *brain waves*.

— This long explanation was needed to acquire intuition of data and its properties: frequencies on specific bands characterize the most relevant visible properties of systems, spectra revealing their "identities", and to recognize the frequencies, one must see over the time scales. The time axis has to be seen both from very near and very far; the key is to abstract the time axis away altogether. How to accomplish this ... this is studied closer in Part V.

6 Step aside: *analogies*

In addition to the selected path, there is another “research freedom” also available to follow. The cost criterion (4) is an extremely compressed *mathematical pattern* where only the essence remains; this makes it possible to see analogies among seemingly very different application fields. — First, some analysis is needed.

For a continuously differentiable function L, using *calculus of variations* one can minimize (or, actually, find the stationary solution for) the functional

$$\mathcal{J} = \int_a^b L(t, q, \dot{q})\, dt \tag{7}$$

where q is the vector of *generalized coordinates* (here q having nothing to do with the coupling factors), resulting in *Euler-Lagrange equations*

$$\frac{d}{dt}\left(\frac{\partial\, L(t, q, \dot{q})}{\partial \dot{q}}\right) - \frac{\partial\, L(t, q, \dot{q})}{\partial q} = 0. \tag{8}$$

If one now defines

$$L(t, q, \dot{q}) = \frac{1}{2}\, \dot{q}^{\mathrm{T}} A\, \dot{q} - \dot{q}^{\mathrm{T}} B\, U(q, t) \tag{9}$$

for some symmetric $n \times n$ matrix A and $n \times m$ matrix B, the “potential” $U(q, t)$ being a vector-valued function, one can write *generalized momenta*

$$\frac{\partial\, L(t, q, \dot{q})}{\partial \dot{q}} = A\, \dot{q} - B\, U(q, t), \tag{10}$$

so that the total derivative becomes

$$\frac{d}{dt}\left(\frac{\partial\, L(t, q, \dot{q})}{\partial \dot{q}}\right) = A\, \ddot{q} - \frac{dU(q, t)}{dq}\, B^{\mathrm{T}}\, \dot{q} - B\, \frac{dU(q, t)}{dt}. \tag{11}$$

On the other hand,

$$\frac{\partial\, L(t, q, \dot{q})}{\partial q} = -\frac{dU(q, t)}{dq}\, B^{T}\, \dot{q}. \tag{12}$$

Now (8) becomes

$$A\, \ddot{q} = B\, \frac{dU(q, t)}{dt}. \tag{13}$$

This can be interpreted as *Newton’s second law of motion* in vector form, with A being some (momentary) *inertial matrix* and $F(q, t) = B\, dU(q, t)/dt$ being the vector of (generalized) driving forces applying in the directions of generalized coordinates; thus, U can be seen as some kind of special “global potential” whose spatial changes over individual q do not matter.

An interesting observation is that (13) would have been found directly by setting the gradient of (9) to zero, solving for $\dot{q}$, and differentiating with respect to t; this means that "pointwise" optimization of a criterion of the form (7) gives the same result as global optimization, motivating the use of the same criterion regardless of the time scales. And the ordering of spatial samples does not matter; thus, analysis of the time points can just as well be truly *statistical*.

Indeed, this opens up interesting visions. One can identify the vector $\bar{x}$ with the generalized velocities $\dot{q}(t)$ (the set of generalized coordinates needs not be minimal), and the vector u with the external function $U(q,t)$; further, if the "visibility horizon" in semiosis extends from a to b, it turns out that criteria (4) and (7) can coincide. This means that *the neocybernetic model can be seen to implement Lagrangian mechanics*.

There are now interesting possibilities available: the neocybernetic principles can be applied for some kind of *structural optimization*. If one lets the matrices A and B locally adapt towards the available enformation, towards the matrices $\mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}$ and $\mathcal{E}\{\bar{x}u^{\mathrm{T}}\}$, respectively, the overall system dynamics should become more optimized on average.

The same criterion (4) is familiar from other domains, too. For example, the *deformation energy* in mechanics has essentially similar outlook: when variables in $\bar{x}$ are interpreted as *deformations* or *displacements* and u is the vector of acting forces, the former term in (4) becomes the *internal energy* and the latter term becomes the *external energy* (the matrices determining the internal "springs" within the structure). Neocybernetic instantaneous optimization then gives the steady state of the structure in the deformation energy minimum, and optimization of the matrices gives the overall average deformation minimum, being assumedly robust against deforming forces. Indeed, neocybernetic systems have also been called *elastic systems*, because they can be seen as trying to minimize the deformation energy of some hypothetical elastic membrane.

Mechanical cases are not the only ones when searching for analogies. For example, if the (squares of the) state variables are seen as negative charges, and if the fixed positive charges are seen as the "environment", the criterion (4) can be interpreted as the electrostatic energy, and based on that "electrons" can be seen to cybernetically organize around the nuclei — thus automatically determining some kind of emergent structures, or *cybernetic orbitals* along molecules (see [1]).

Finally, it can be observed that applying the *Legendre transform* to (4) with the *adjoint state* as $s = \frac{dJ}{dx} = \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}\, x - \mathcal{E}\{\bar{x}u^{\mathrm{T}}\}\, u = \mathcal{E}\{\bar{x}\bar{x}^{\mathrm{T}}\}\,(x - \bar{x})$ the transformed cost has a very intriguing form:

$$G(s) = \frac{1}{2}\left(s + \mathcal{E}\left\{\bar{x}u^{\mathrm{T}}\right\}u\right)^{\mathrm{T}} \mathcal{E}\left\{\bar{x}\bar{x}^{\mathrm{T}}\right\}^{-1} \left(s + \mathcal{E}\left\{\bar{x}u^{\mathrm{T}}\right\}u\right). \tag{14}$$

References

1. *Neocybernetics — Pragmatic Semiosis by Complex Adaptive Systems.* Research pages accessible in Internet through http://neocybernetics.com.