Level 3

Elasticity of Systems and *Goals of Evolution*

As it was observed in the previous chapter, the biological data can efficiently be modeled applying multivariate statistical tools. However, it seems that such data-oriented approaches do not suffice after the elementary levels. It is the same with "chemical pattern recognition" as it is also in other areas of data-based modeling: The statistical correlations are not enough to unambiguously determine the higher-level structures.

To get further, one has to apply more ambitious ways of limiting the available complexity. Traditionally, the approach is to introduce more stringent and complex model structures to direct the parameter matching. However, again following the neocybernetic ideas, no extra complexity is voluntarily integrated in the models — alternatives to increased complexity are searched for instead.

The alternative employed here is rather radical.

In the postmodern era, there should exist no taboos. However, one thing that has never been proposed in circles of modern serious science, is that of *finalism.* One should only answer the *how* questions, never the *why* questions. Yet, applying teleological assumptions, most compact problem settings are reached, and one can also study systems that *not yet exist.* The claim here is that appropriate finalistic arguments can also be given concrete contents, so that they become verifiable — or falsifiable. It is only the starting point that sounds radical: The discussions collapse back into well-established frameworks.

It has been said that nothing in biology can be explained without taking evolution into account. And here this observation is exploited by studying the question: What is it that evolution tries to accomplish? Such issues are studied in the neocybernetic perspective — balance pursuit is the only finalistic goal after all, together with extreme environment-orientedness.

3.1 Balancing between static and dynamic models

From now on, one needs to (for a moment) forget about the technically oriented derivations in the previous chapter. The emphasis is changed: It is not what the model designer intends that is relevant — the interesting things are those what the system naturally does. Again, there is the same starting point (1.7) that is assumed to hold for a thermodynamically consistent chemical balance system. Indeed, such a simple formulation can be written for any linear system, no matter what is the domain field, if the variables are selected appropriately. This set of equations can be interpreted so that it defines a *static balance* with *no structure,* and one first needs to extend the framework.

3.1.1 Restructuring data

Assume that the variables in z in (1.7) are divided in two parts: Vector u, dimension m, describes the environmental conditions, whereas vector $\bar{x}$, dimension n, contains the system-specific *internal variables,* somehow characterizing the equilibrium state of the system. The internal state is not assumed to be necessarily observable by an external observer. The "environment" here is not something external — it only consists of variables that are determined from outside, but essentially all variables (concentrations) still coexist in the same volume. Rewriting the constraints characterizing the system, one can distinguish between the variables:

$$A\bar{x} = Bu. \tag{3.1}$$

The construction of the matrices A and B is not uniquely determined by this expression — this issue, determination of the system matrices in a plausible way, is studied later. To keep the internal state of the system well-defined, it is assumed that there are as many constraints here as there are latent variables, so that A is square. Because of environment-orientedness, the internal variables are assumed to be directly determined by the environment, so that there assumedly is a (linear) dependency between $\bar{x}$ and u. Formula (3.1) is an implicit expression; assuming that A is invertible, one can explicitly solve the unique linear function from the environmental variables into the system state:

$$\bar{x} = A^{-1}B\,u, \tag{3.2}$$

so that one can define an explicit mapping matrix from u to $\bar{x}$

$$\phi^T = A^{-1}B. \tag{3.3}$$

However, the main motivation for the formulation in (3.1) is that one can formally extend the static model into a dynamic one. The formula (3.1) only characterizes the final visible balance in the system, but one has to remember that it is local operations only that exist — how can such uncoordinated local actions implement the global-level behaviors? Indeed, one needs to extend

studies beyond the final balance, and take into account the dynamic behaviors caused by the imbalances.

Formula (3.1) can be interpreted as a balance of tensions determined by forces $A\bar{x}$ and Bu, caused by the system itself and by the environment, respectively. If the forces are not in balance, there is a drift. Assuming that the data structures are selected appropriately, so that $-A$ is stable (eigenvalues having negative real parts), one can define a dynamic model to characterize the tensions as

$$\frac{d\,x}{\gamma\,d\tau} = -Ax + Bu. \tag{3.4}$$

The parameter γ can be used for adjusting the time axis. The steady state equals that of (3.2), so that $\lim_{\tau\to\infty} x = \bar{x}$ for constant u. Because of linearity, this steady state is unique, no matter what was the initial state. Using the above construction, the static pattern has been transformed into a dynamic pattern — the observed equivalences are just emergent phenomena reflecting the underlying dynamic equilibrium.

How can such a genuine extension from a static model into a dynamic one be justified? It needs to be observed that there *must* exist such an inner structure beyond the surface. The seemingly static dependencies of the form (1.7) have to be basically dynamic equilibria systems so that the equality can be restored after disturbances: The actors, or the molecules in this case, do not know the "big picture", and it is the interactions among the molecules that provide for the tensions resulting in the tendency towards balance. It is assumed here that the mathematical model represents *what a system really does.* The model is not only mathematically appropriate, but it explains the actual mechanisms taking place in the chemical system that is getting towards balance after a transient.

What causes the dynamics, then? Thinking of the mindless actors in the system, the only reasonable explanation for the distributed behaviors is *diffusion.* It is the concentration gradients that only are visible at the local scale in a chemical system. So, interpreting (3.4) as a (negative) gradient, there has to exist an integral — a criterion that is being minimized. By integration with respect to the variable x, it is found that

$$\mathcal{J}(x,u) = \frac{1}{2}x^T Ax - x^T Bu \tag{3.5}$$

gives a mathematical "pattern" that characterizes the system in a yet another way. Note that by employing the dynamic systems understanding, it was possible to escape the limits of the "dead" formulation and turn an originally static problem into another, more interesting static form. Such an optimization-oriented view of systems as proposed above combines the two ways of seeing systems: The criterion itself represents the pattern view, whereas the optimization process represents the process view. Similarly, there is also connection to the philosophies: Whereas Heraclitus emphasized the processes, Plato tried to capture the "ideals", or the patterns beyond the changes.

Now one can conclude that the chemical balance system formally implements pattern matching of the form (2.12) as studied in the previous chapter, with

variables being interpreted within the new structure:

$$\begin{aligned} J(x,u) &= \frac{1}{2}\left(u-\varphi x\right)^T W\left(u-\varphi x\right) \\ &= \frac{1}{2}x^T\varphi^T W\varphi x - x^T\varphi^T W u + \frac{1}{2}u^T W u, \end{aligned} \tag{3.6}$$

so that

$$J(x,u) = \mathcal{J}(x,u) + \frac{1}{2}u^T W u. \tag{3.7}$$

The two cost criteria (3.5) and (2.12) are equivalent what comes to the "tensions" imposed by them; constant factors (with respect to x) do not change the location of the minimum, nor the gradients, for given u. The correspondence between the cost criteria is reached when one defines the matrices as

$$\left\{ \begin{array}{lcl} A &=& \varphi^T W\varphi \\ B &=& \varphi^T W. \end{array} \right. \tag{3.8}$$

This connection between data structures is studied closer in Sec. 3.2.3. If (3.8) holds, one can see that all eigenvalues of A are non-negative, meaning that with such a selection the process (3.4) always remains stable.

Criterion (3.6) gives another view too see the same gradient-based minimization (3.4). When (3.6) is minimized using the steepest descent gradient approach, the continuous-time process implementing this minimization is

$$\frac{d\,x}{\gamma\,d\tau} = \varphi^T W\left(u-\varphi x\right). \tag{3.9}$$

It is the latter part $u-\varphi x$ that makes it possible to reach more sophisticated results in matching: For example, the adaptation can do the pattern matching even if the feature vectors in φ were non-orthogonal or unnormalized. This feedback structure will be studied later; now the key point is the basic structure of this formula (3.9). Whereas the matrix ϕ^T implements a mapping from the environmental variables u into the system variables $\bar{x}$, the feature matrix φ can be interpreted as an inverse mapping from the space of x into the space of u. Formally, simply for mathematical reasons, there must hold $\phi^T = \left(\varphi^T W\varphi\right)^{-1}\varphi^T W$, but more useful results can be found.

The effects φx and $\phi^T u$, or diffusion processes into and out from the system, eventually find their balance — it is not possible to determine the "original causes". One can even speak of a *holistic view* here. Because of pancausality, there exists a two-way connection: Changes in any variable causes changes in other variables, no matter whether the variable belongs to x or u. Just as the environmental variables can affect the system variables, the environment can be affected by the system. This two-way assumption blurs the traditional view of distinguishing between a system and its environment, there is no clear distinction between them. The "original" environment u is changed by x — but there does not exist any intact environment to start with. The vector u is the net effect of all accompanying subsystems, all of them together modifying their common substrate. A subsystem is an integral part of the whole, the

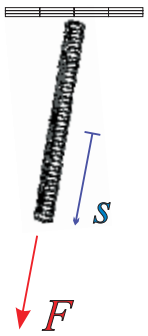

Figure 3.1: Prototypical spring being stretched

environment being a composition of subsystems. The whole system — indeed, the environment — consisting of a large number of subsystems can be dictated by none of the subsystems alone. The environment should not be seen as a distinct concept, or as something fundamentally intractable in the hierarchy of models. The deep connection between the mappings ϕ and φ is a key issue when trying to capture the behaviors of cybernetic systems.

However, such observations above have little value if the data structures ϕ, φ, and W (or A and B) cannot be determined. To attack this problem, a wider perspective is needed.

3.1.2 Elastic systems

Study the cost criterion (3.5) closer. It turns out that this cost criterion has a very familiar outlook, and employing new terminology, valuable intuitions are available. To see this, some facts need to be refreshed.

Study a *spring* having the spring constant k (the spring can also be torsional, etc.). When the spring is stretched by an amount s because of an external force F (see Fig. 3.1), there are external and internal stored energies in the spring:

- Due to the potential field: $W_{\text{ext}} = -\int_0^s F\,ds = -Fs$.
- Due to the internal tensions: $W_{\text{int}} = \int_0^s ks\,ds = \frac{1}{2}ks^2$.

This can be generalized, assuming that there are many forces, and many points being connected by springs, so that the internal tension between the points s_1 and s_2, for example, becomes

$$W_{\text{int}}(s_1, s_2) = \frac{1}{2}k_{1,2}\,(s_1 - s_2)^2 = \frac{1}{2}k_{1,2}s_1^2 - k_{1,2}s_1s_2 + \frac{1}{2}k_{1,2}s_2^2.$$

A matrix formulation is also possible for vectors s and F, when the interaction factors are collected in matrices A and B. It turns out that the expressions for potential energy components have familiar

outlooks:

$$W_{\text{int}}(s) = \frac{1}{2} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T A \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}, \tag{3.10}$$

and

$$W_{\text{ext}}(s, F) = - \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}^T B \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}. \tag{3.11}$$

For a moment, assume that vector u denotes *forces* acting in a (discretized) mechanical system, and x denotes the resulting *deformations.* Further, assume that A is interpreted as the *elasticity matrix* and B is *projection matrix* mapping the forces onto the deformation axes. Matrix A must be symmetric, and must be positive definite to represent stable structures sustaining external stresses — these conditions are fulfilled if (3.8) hold. Then, it turns out that (3.5) is the difference between the potential energies stored in the mechanical system. Principle of minimum potential (deformation) energy [19] states that a structure under pressure ends in minimum of this criterion, trying to exhaust the external force with minimum of internal deformations.

However, the same criterion can be seen to characterize all cybernetic balance systems. This means that in non-mechanical cybernetic systems, the same intuition concerning understanding of mechanical systems can be exploited. It does not matter what is the domain, and what is the physical interpretation of the "forces" u and of the "deformations" $\bar{x}$, the structure of the system behavior remains intact: As the system is "pressed", it yields in a more or less humble manner, but when the pressure is released, the original state is restored. Indeed, in chemical environments, this behavior is known as the *Le Chatelier principle:* If there is some disturbance acting on the system, the balance moves in such a direction where the effects are "eaten up". In this sense, one can generally speak of *elastic systems.*

In short: Neocybernetic systems are identical with elastic systems — systems that are characterized by dynamic equilibria rather than static equivalences. When rigid constraints are substituted by "soft" tensions, there is smoothness, and — by definition — local linearizability can be assumed also what comes to originally nonlinear models.

The effect of the environmental pressures on the system can be easily quantified: Just as in the case of a potential field, it is the product of the force and displacement that determines the change in potential energy. Similarly, regardless of the physical units of the variables, one can interpret the product $\bar{x}_i u_j$ in terms of *energy transferred from the environment into the system* through the pair of variables u_j and x_i. Correspondingly, if there are variables that can be interpreted as dissipative flows or rates, the energies are also effectively divided by time, so that it is some kind of *power* that is transferred. This concept deserves a name, or, actually, an old concept is renamed: In what follows, this "emergent level energy" is studied along the following definition:

> **Emergy** (a scalar dimensionless quantity) is the product of the (abstract) force and the corresponding (abstract) deformation.

As it turns out, this emergy is "information energy" that is the prerequisite for emergence of information structures. Emergy will here be a much more abstract thing and will have a broader scope than that used in [58].

Such energy flows have been studied before in more concrete terms in various contexts: *Bond Graphs* are used to model systems in terms of energies being transferred among system components [16]. It has been shown that this modeling strategy can be applied to a wide variety of tasks, so that this approach seems to be a rather general one. However, Bond Graphs are traditionally used for modeling different kinds of dissipative flows — and now the emphasis is on balances. Resulting models are very different.

However, it must be remembered that there is not only the effect from the external environment into the internal system — there is a symmetric two-way interaction that takes place. It is the matrices ϕ^T and φ that characterize the emergy transfer between the system and its environment. It is not only so that u should be seen as the "force" and $\bar{x}$ as the effect: $\bar{x}$ can be seen as the action and u as the reaction just as well. This duality makes it possible to tie the loose ends together.

3.1.3 Evolutionary fitness

It was mentioned above that the key challenge in this chapter is to determine the *goals of evolution.* Traditionally, one is facing paradoxes here: Remember that the layman intuition does not work. If the fitness criterion were the "maximum number of offspring", for example, there would only exist bacteria on earth. On the other hand, the "blind watchmaker" metaphor with random optimization [21] simply cannot be the mechanism beyond evolution.

Neocybernetic environment-orientedness suggests a criterion emphasizing some kind of *match with environment.* Indeed, applying the above discussion concerning energy/power transfer from the environment into the system and back, an intuitively appealing fitness criterion would be

> Maximize the average amount of emergy that is being transferred between the system and the environment.

No matter what is the physical manifestation of the environmental variables, a surviving system interprets them as resources, and exploits them as efficiently as possible. Note that it is not predetermined what should be done with the extracted energy: The metabolic products can change the environment to be further exploited. This makes it possible that evolutionary processes can proceed in many different ways — the relevance of the behaviors is later evaluated by the evolutionary selection. To begin with, the criterion is always the same — match with environment — no matter how some "master mind" would like the system to develop.

When there are resources available in the environment, it is also clever to utilize this abundance somehow. Typically, if the environmental "force" comes into

a yeast cell in the form of glucose steps, for example, it is different kinds of metabolic products that can be produced: In some cases it can be the mannose-production path that outperforms others, producing new cells; in some other cases, heat production is to promote — meaning that reproduction and survival are competing goals. In each case, the assumption here is that the cell most efficiently exploiting the available energy prospers in the long run.

Following the above lines of thought, the momentary emergy traversing from the environmental variable j to the state variable i can be written as $\bar{x}_i u_j$, or, when written in a matrix form simultaneously for all variables, $\bar{x}u^T$. Similarly, the momentary emergy traversing from the state variable i to the environmental variable j can be written as $u_j \bar{x}_i$, or, when written simultaneously for all variables, $u\bar{x}^T$. If evolution proceeds in a consistent manner, the differences among such variable pairs should determine the growth rates of the corresponding links between the variables; when the mapping matrices ϕ^T and φ are defined as shown above, one can assume that a stochastic adaptation process takes place, the observations of prevailing variable levels determining the stochastic gradient direction:

$$\begin{cases} \dfrac{d\,\phi^T}{dt} & \propto \quad \bar{x}(t)u^T(t) \\ \dfrac{d\,\varphi}{dt} & \propto \quad u(t)\bar{x}^T(t). \end{cases} \tag{3.12}$$

However, note that the matrix elements cannot be explicitly localized in the system. When (structural) changes take place in the underlying system, it is constraints that are being added or modified, and these changes are reflected in the elements of ϕ^T and φ in more or less random ways. All changes in the underlying structure typically affect the mappings — but all of the changes affect them only little, at least if the number of components in the system is high. The high number of discrete parameters are projected onto the low-dimensional set of more or less smooth "emergent parameters". When the discrete space of structures changes into a more continuous behavior of emergent parameters, more or less consistent evolutionary optimization becomes possible. What is more, the local optimizations are independent of each other — this makes it possible that various optimization processes can take place simultaneously, thus making the optimization a parallel process, relatively fast and robust. The time scales in (3.12) are much longer than in (3.4).

When looking at the formulas (3.3) and (3.12) together, for example, it is clear that such adaptation processes are unstable — high correlations between $\bar{x}_i$ and u_j eventually result in still higher correlations between them, thus making $\bar{x}_i$ (or u_j) grow without limit. Indeed, this adaptation principle is an extension of the *Hebbian learning rule,* where it is the correlation between the environmental signal in u_j and neuronal activity in $\bar{x}_i$ that has been shown to determine the synaptic adaptation in real neural cells [37].

There is a positive feedback in the adaptation law, and just as it is with the Hebbian neurons, the stability problem emerges if the trivial learning rule is applied (see [92]). Stabilization of the Hebbian learning model has been studied a lot — but, again, applying the neocybernetic simplicity ideal, one should not introduce new structures separately for stabilization purposes. For a moment, simply assume that $\bar{x}$ and u *for some reason* remain bounded; then it is rea-

sonable to assume that the processes (3.12) find a fixed state, and the solution for this fixed state can be assumed to be such that the matrix elements ϕ_{ji} are relative to the correlations between $\bar{x}_i$ and u_j, or

$$\phi^T = q\,\mathrm{E}\left\{\bar{x}u^T\right\}, \tag{3.13}$$

and in the backward direction,

$$\varphi = b\,\mathrm{E}\left\{u\bar{x}^T\right\}. \tag{3.14}$$

If the dynamics of x is rather fast, so that the system can be assumed to always be in dynamic balance, one can substitute $\bar{x}$ with x in the above formulas (and also in the formulas that follow). Here, the parameters q and b are some constants; the role of these *coupling coefficients* is studied later. Similarly, the relevance of the observation (3.15), or the role of the system as a *mirror image* of the environment, will be discussed later. This means that the matrices ϕ and φ should become proportional to each other:

$$\varphi = \frac{b}{q}\,\phi. \tag{3.15}$$

As it turns out, these factors scale the signal levels in the system and in the environment. When interpreting (3.15), it is quite natural to think that *exploitation means exhaustion* — it is those elements u_j that contribute most in the determination of the values of $\bar{x}$ that become exhausted the most.

It needs to be recognized that the adaptation in the system according to (3.13) is completely local for any element in the matrices ϕ and φ even though the assumed goal of the evolutionary process is presented in a collective matrix format. It is interesting to note here that the expressions for ϕ and φ are essentially symmetric. Remember that it was Heraclitus who said "the way up and the way down are the same" — whatever he meant.

3.2 Towards self-organization

The key question in complex systems is that of self-organization: How can anything qualitatively new emerge from non-centralized operations. For a system to self-organize, it must first self-regulate. In this section, the issue of self-regulation is first studied, and the issue of self-organization after that.

The basic solution to regulation is negative feedback. However, now there are no explicit control structures available, and no organized communication or signal transfer infrastructure within the system: How to implement the feedback structures? Again, some background analysis is first in place.

3.2.1 Feedback through environment

The traditional approach to avoid explosions is to include non-idealities in the originally idealized models. For example, an originally linear system can become

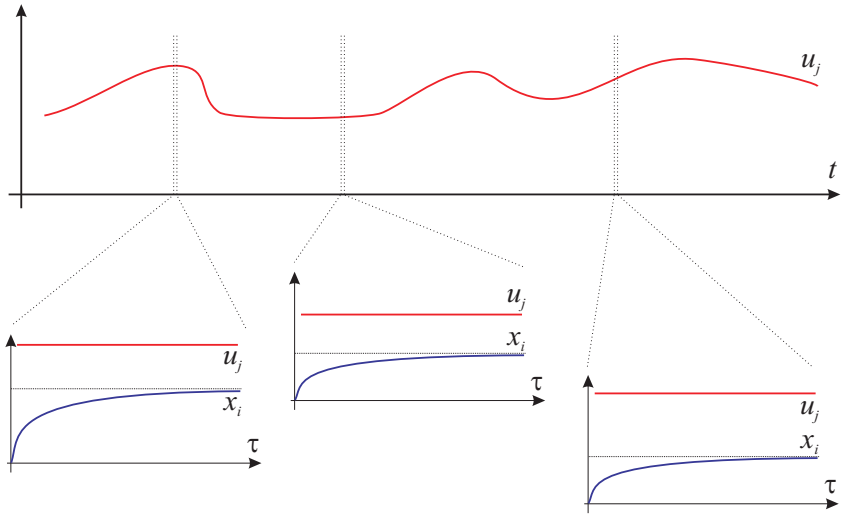

Figure 3.2: Illustration of two time scales. It is assumed that the dynamics of u (on the t scale) is much slower than that of x (τ scale)

stable if nonlinearities are added so that signals saturate. Here, non-idealities are again included in the model — however, these non-idealities are now located in an unorthodox place.

There are no unidirectional effects in real systems: Information flows cannot exist without physical flows that implement them. When emergy is being consumed by the system, this emergy is taken from the environment, or environmental "resources" are exhausted. To understand these mechanisms, study the pattern matching process (3.9). There are essentially two parts in this expression: First, in the front there is $\varphi^T W$ implementing parallel matching of data against the model, determining the directions of local diffusion processes; second, there is $u - \varphi x$ defining some kind of *virtual environment* that is being matched. The negative feedback structure $-\varphi x$ represents real material flow in from the system into the environment, the resources being exhausted. The changed environment becomes

$$\tilde{u} = \underbrace{u}_{\substack{\text{actual}\\ \text{environment}}} - \underbrace{\varphi x}_{\text{feedback}} . \tag{3.16}$$

The system never sees the original u but only the distorted $\tilde{u}$, where the momentary emergy consumption in the system, or φx, is taken into account. Clearly, as the environment affects the system and the system affects the environment, there exists a feedback structure; again, one is interested in the final balance after transients:

$$\bar{u} = u - \varphi \bar{x}. \tag{3.17}$$

Later on, real-life realism will be applied: Only $\bar{u}$ is visible, never u itself. The matrix ϕ^T is redefined here: It stands for the mapping from the effective environment to the state, however this environment is manifested — in this feedback case meaning that $\bar{x} = \phi^T \bar{u}$.

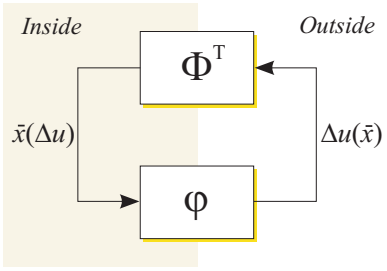

Figure 3.3: The algebraic loop between the environment and the system

Because the environment is disturbed by the system, the setting is nonideal, but this nonideality makes new functionalities possible — like self-organization, as shown in the next section. But the key issue here is that this negative feedback keeps the system in balance and signals bounded, as was assumed in the previous section. The feedback structure is implicit, through the environment, and the effects of this feedback will be studied below. To start with, no assumptions like (3.14) are made — φ is an arbitrary $m \times n$ mapping matrix.

When studying the steady state, there is efficiently an algebraic loop in the system (see Fig. 3.3), and this means that this structure has peculiar properties. Multiplying (3.17) from the right by $\bar{x}^T$, taking expectations, and reordering the terms, one receives

$$\mathrm{E}\{(u-\bar{u})\bar{x}^T\}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} = \varphi, \tag{3.18}$$

so that, when one defines a quantity for measuring the discrepancy between the undisturbed open-loop environment and the disturbed closed-loop environment,

$$\Delta u = u - \bar{u}, \tag{3.19}$$

the expression (3.17) can be written in the form

$$\Delta u = \mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\bar{x}. \tag{3.20}$$

Variables in $\bar{x}$ and Δu are mutually connected, they vary hand in hand, but together representing the same mapping as φ, but in terms of observation data, helping to see another view of the system properties. Indeed, this Δu can be seen as a "loop invariant" that helps to see properties of the feedback loop, and it turns out to offer a way to reach simplified analysis of the signals. Because Δu assumedly linearly dependent of u, one can interpret this variable as the actual input driving the whole loop, so that there exists a mapping Φ^T

$$\bar{x} = \Phi^T\,\Delta u. \tag{3.21}$$

Assuming that the feedback can implement stabilization, the system in Fig. 3.3 will search a balance so that

$$\bar{x} = \Phi^T\varphi\,\bar{x}. \tag{3.22}$$

To have not only trivial solutions (meaning $\bar{x} \equiv 0$), there must hold

$$\Phi^T\varphi = I_n, \tag{3.23}$$

so that the feedforward and feedback mappings have to be mutually orthogonal. This is a very stringent constraint, and it essentially determines the properties of the feedforward matrix Φ. Here, to determine Φ, assume symmetry with (3.18), and make the following attempt and study where it leads to:

$$\Phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}. \tag{3.24}$$

3.2.2 Back to principal subspace

Above, the balances of x were studied as the environment u was assumed fixed. However, to reach interesting results, the neocybernetic principles need to be exploited again: It is assumed that there exist various levels of seeing the system, and at each of the levels, the balances are exploited. Specially, see Fig. 3.2: Whereas u was assumed to remain constant this far, it only has much slower dynamics than x, and on the wider scale, the environment changes. But assuming stationarity of the environment, or balance on the higher scale, so that u has fixed statistical properties, one can find a "balance model of balances". A truly cybernetic model is a *second-order balance model,* or a *higher-order balance model* over the variations in the system — at these levels beyond the trivial first level balance, one can reach stronger views to see the systems, including *self-organization,* as shown below.

So, assume that dynamics of u is essentially slower than that of x and study the statistical properties over the range of $\bar{x}$, and, specially, construct the covariance matrix of it. From (3.24) one has

$$\bar{x}\bar{x}^T = \mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}\mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \Delta u\Delta u^T\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}. \tag{3.25}$$

When applying expectation operator on both sides,

$$\mathrm{E}\left\{\bar{x}\bar{x}^T\right\} = \mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}\mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \mathrm{E}\left\{\Delta u\Delta u^T\right\}\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^{-1}.$$

Multiply these from left and from right by $\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}$:

$$\mathrm{E}\left\{\bar{x}\bar{x}^T\right\}^3 = \mathrm{E}\left\{\bar{x}\Delta u^T\right\}\ \mathrm{E}\left\{\Delta u\Delta u^T\right\}\ \mathrm{E}\left\{\bar{x}\Delta u^T\right\}^T, \tag{3.26}$$

and, when observing the nature of Φ, this can be written

$$\left(\Phi^T\mathrm{E}\left\{\Delta u\Delta u^T\right\}\Phi\right)^3 = \Phi^T\mathrm{E}\left\{\Delta u\Delta u^T\right\}^3\Phi. \tag{3.27}$$

If $n = m$, any orthogonal matrix $\Phi^T = \Phi^{-1}$ will do; however, if $n < m$, so that x is lower-dimensional than u, the solution to the above expression is non-trivial (see [92]: Report 144, "Hebbian Neuron Grids: System Theoretic Approach", pages 12–15). It turns our that *any subset of the principal component axes of the data* Δu *can be selected to constitute* Φ*,* that is, the columns Φ_i can be any n of the m covariance matrix eigenvectors θ_j of this data. Further, these basis vectors can be mixed, so that $\Phi = \theta D$, where D is any orthogonal $n \times n$ matrix[1], so that $D^T = D^{-1}$. In any case, there holds

$$\Phi^T\Phi = I_n. \tag{3.28}$$

[1]Note that there is an *error* in that report in [92]: The matrix D is not whatever invertible matrix, it must be orthogonal (as becomes evident when going through the proof therein)

Now, return to the assumption in (3.24) — indeed, the above selection for Φ seems to fulfill the orthogonality claim (3.23):

$$\begin{aligned}
\Phi^T\varphi &= \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}\,\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
&= \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\Phi^T\mathrm{E}\{\Delta u\Delta u^T\}\,\mathrm{E}\{\Delta u\Delta u^T\}^T\Phi\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
&= \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\Phi^T\mathrm{E}\{\Delta u\Delta u^T\}^2\,\Phi\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
&= \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^2\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \\
&= I_n.
\end{aligned} \tag{3.29}$$

The above derivations show that any set of covariance matrix eigenvectors can be selected in Φ. However, in practice it is not whatever combination of vectors θ_j that can be selected: Some solutions are *unstable* when applying the iterative adaptation strategies. Indeed, following the lines of thought shown in [92], the only stable and thus relevant solution is such where it is the n most significant eigenvectors (as revealed by the corresponding eigenvalues) that constitute the matrix Φ in convergent systems. This means that the system implements *principal subspace analysis* for input data. Because of the mixing matrix D, the result is not unique in the sense of principal components, but the subspace spanned by them is identical, and exactly the same amount of input data variation is captured. Specially, if there were some further exploitation of the latent variables $\bar{x}$, reconstructions of $\hat{y}$ would be equally accurate no matter whether the principal components or the principal subspace only were used.

The above derivations apply to all feedback matrices φ: The system signals adapt to fulfill the equation (3.18). The results only apply to the subspace spanned by φ — that is, in the subspace where there is variation in Δu caused by the feedback — and within that subspace, the structure of maximum variation is found. If φ is adaptive and selected applying the evolutionary strategy, so that $\varphi^T = b\mathrm{E}\{\bar{x}\bar{u}^T\}$, it is the principal subspace of u that is spanned. These issues will be studied later.

Now one can conclude that completely local operations result in non-trivial structures that are meaningful on the global scale: Competitive learning without any structural constraints results in self-regulation (balance) and self-organization (in terms of principal subspace). Feedback through the environment, or competition for the resources, results in stabilization and organization of the system.

3.2.3 Closer look at the cost criteria

When comparing to (3.3) to (3.24), and when u in the formulas is substituted with Δu, one can see that an appropriate connection between the data structures is reached when one selects the matrices so that

$$\left\{\begin{array}{lcl} A &=& \mathrm{E}\left\{\bar{x}\bar{x}^T\right\} \\ B &=& \mathrm{E}\left\{\bar{x}\Delta u^T\right\}. \end{array}\right. \tag{3.30}$$

As presented in [92], essentially the same formulas were found in the neuronal system applying not only "Hebbian learning", but together with the "anti-Hebbian" structures, where the feedbacks were explicitly implemented. When the feedback through the environment is taken into account, simpler structures suffice, and the results are the same. However, there is a difference: Whereas

the explicitly implemented feedback structures analyze the original undisturbed environment, the feedbacks implemented through the environment analyze the disturbances in the environment. These differences between open-loop environment and closed-loop environment are measurable only after adaptation, Δu substituting the original u in analyses. The model with explicit feedback is not completely based on local information: There the matrix ϕ implements a mapping from u onto $\bar{x}$, essentially assuming that the feedback is implemented without affecting the environment itself. Such a feedback scheme is possible in systems where the actors are "intelligent agents" that are capable of seeing the environment in a wider perspective, as studied in the next chapter.

Yet another conclusion is in place here: Comparing expressions (3.8) and (3.30), it turns out that to avoid contradictions, one has to choose $W = \mathrm{E}\{\Delta u \Delta u^T\}$. If the feedback is explicit, on the other hand, the weighting matrix is $W = \mathrm{E}\{uu^T\}$. The implicit data weighting is also identical with that proposed in the context of emergent models. The technical manipulations in the previous chapter are essentially an appropriate way to characterize the behaviors also in the locally controlled, real (but idealized) system:

> Neocybernetic system implements the emergent model structure. The locally controlled system carries out modeling of the environment u applying principal subspace based feature extraction (slow process of determining ϕ) and pattern matching (fast process of determining $\bar{x}$).

Having compact formulations for the matrices, the cost criteria can also be studied closer. Defining $\mathcal{J}(u) = \mathcal{J}(\bar{x}, u)$, from (3.5) one has, assuming that there holds (3.24,

$$\begin{aligned}
\mathcal{J}(u) &= \frac{1}{2}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} - \bar{x}^T\mathrm{E}\{\bar{x}\Delta u^T\}\Delta u \\
&= \frac{1}{2}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} - \Delta u^T\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}\Delta u^T\}\Delta u \\
&= \frac{1}{2}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x} \\
&\quad - \underbrace{\Delta u^T\mathrm{E}\{\bar{x}\Delta u^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}}_{\bar{x}^T}\mathrm{E}\{\bar{x}\bar{x}^T\}\underbrace{\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}\mathrm{E}\{\bar{x}u^T\}\Delta u}_{\bar{x}} \\
&= -\frac{1}{2}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x},
\end{aligned}$$

so that the *average* of the criterion can be written as

$$\begin{aligned}
\mathrm{E}\{\mathrm{trace}\{\mathcal{J}(u)\}\} &= -\frac{1}{2}\mathrm{E}\left\{\mathrm{trace}\left\{\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\bar{x}\right\}\right\} \\
&= -\frac{1}{2}\mathrm{E}\left\{\mathrm{trace}\left\{\bar{x}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\right\}\right\} \\
&= -\frac{1}{2}\mathrm{trace}\left\{\mathrm{E}\left\{\bar{x}\bar{x}^T\mathrm{E}\{\bar{x}\bar{x}^T\}\right\}\right\} \\
&= -\frac{1}{2}\mathrm{trace}\left\{\mathrm{E}\{\bar{x}\bar{x}^T\}^2\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{n}\lambda_i^2.
\end{aligned} \tag{3.31}$$

The above simplification comes from the linearity of the operators, trace $\{\mathrm{E}\{\cdot\}\} = \mathrm{E}\{\mathrm{trace}\{\cdot\}\}$, and from the properties of matrix trace: Trace it is the sum of the diagonal elements, and simultaneously it is the sum of the matrix eigenvalues; for scalars there is naturally no effect. What is more, matrices within trace can be rotated, that is, $\mathrm{trace}\{M_1 M_2\} = \mathrm{trace}\{M_2 M_1\}$, if the matrices M_1 and M_2 are appropriately compatible. The above result means that the completely adapted system maximizes the sum of the n most significant eigenvalue squares as seen from within the system. Using the other criterion, the optimum reaches $\mathrm{E}\{J(u)\} = \sum_{j=n+1}^{m} \lambda_j^2$.

It has to be kept in mind that if the feedbacks are implemented through the environment, the eigenvalues λ_i are eigenvalues of $\mathrm{E}\{\Delta u \Delta u^T\}$. They are eigenvalues of $\mathrm{E}\{uu^T\}$ only if the feedbacks are implemented actively by some intelligent agent (as studied in later chapters).

3.2.4 Making it local

The above theoretical derivations were interesting, giving qualitative understanding of the properties of the feedback loop, but they were applicable only for the global scale analyses: From the point of view of the system, Δu is not known, as the original undisturbed u cannot be seen without disturbing it. So, from now on, assume that the system only sees the real, virtual environment as disturbed by the feedbacks, and, according to (3.13), define[2]

$$\bar{x} = \phi^T \bar{u}, \tag{3.35}$$

where $\phi^T = q\mathrm{E}\{\bar{x}\bar{u}^T\}$. Now it is the really measurable environment, as manifested in $\bar{u}$, that is only involved in local calculations. As it is the feedback that supplies for the basic functionality of a cybernetic system, spanning the principal subspace of the data, it is the role of the learning to make this data represent the external environment u as manifested in $\bar{u}$. There are two main functionalities in the studied system structure: Feedback implements principal subspace analysis, and adaptation in the form (3.13) and (3.14) implements match with environment, so that it is the signals Δu, and simultaneously the original u, that determine this principal subspace. Going towards maximum variation spans the principal subspace in the data when the latent variables are kept linearly independent.

[2]How is (3.35) related to (3.24), how can they represent the same system — specially, where does the inverse covariance matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ emerge in the formula? To have intuition on this, note that

$$\begin{aligned} \phi^T &= q\,\mathrm{E}\{\bar{x}\bar{u}^T\} \\ &= q\,\mathrm{E}\{\bar{x}(u - b/q\,\phi\bar{x})^T\} \\ &= q\,\mathrm{E}\{\bar{x}u^T\} - b\,\mathrm{E}\{\bar{x}\bar{x}^T\}\,\phi^T, \end{aligned} \tag{3.32}$$

and when solving this,

$$\phi^T = \left(\mathrm{E}\{\bar{x}\bar{x}^T\} + \frac{1}{b} I_n\right)^{-1} \frac{q}{b}\mathrm{E}\{\bar{x}u^T\}. \tag{3.33}$$

When letting b grow, the required functional structure emerges:

$$\phi^T = \mathrm{E}\{\bar{x}\bar{x}^T\}^{-1} \frac{q}{b}\mathrm{E}\{\bar{x}u^T\}. \tag{3.34}$$

The signals $\bar{x}$ and $\bar{u}$, as defined as in (3.35), have peculiar properties. For example, multiplying (3.35) from the right by $\bar{x}^T$ and taking expectation, one has an expression for the latent vector covariance:

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = q\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T. \tag{3.36}$$

This holds *if* the latent variables x_i do not fade away altogether (or explode). These issues are studied later — however, here it is assumed that the system is strictly cybernetic, all latent variables are occupied, and, for example, the matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ remains invertible. On the other hand, multiplying (3.35) from the right by $\bar{u}^T$ and taking expectation, one has

$$\mathrm{E}\{\bar{x}\bar{u}^T\} = q\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{u}\bar{u}^T\}. \tag{3.37}$$

Substituting this in (3.36),

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = q^2\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{u}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T, \tag{3.38}$$

or

$$\begin{aligned}\frac{1}{q}\,I_n &= \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}\,\mathrm{E}\{\bar{u}\bar{u}^T\}\,\mathrm{E}\{\bar{x}\bar{u}^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\,\sqrt{q}\\ &= \bar{\theta}'^T\,\mathrm{E}\{\bar{u}\bar{u}^T\}\,\bar{\theta}',\end{aligned}$$

where

$$\bar{\theta}'^T = \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}. \tag{3.39}$$

From (3.36), it is evident that there holds[3]

$$\bar{\theta}'^T\bar{\theta}' = I_n. \tag{3.40}$$

This means that the columns in $\bar{\theta}'$ span the subspace determined by n of the principal components of $\mathrm{E}\{\bar{u}\bar{u}^T\}$, so that $\bar{theta}' = \bar{\theta}D$, where $\bar{theta}$ is a matrix containing n of the covariance matrix eigenvectors, and D is some orthogonal matrix; as in Section 3.2.2, it can be assumed that this is the principal subspace spanned by the n most significant of them (this claim is confirmed by simulations). All eigenvalues $\bar{\lambda}_j$ in the *closed loop* equal $1/q$.

Assume that the coupling coefficients q_i vary between latent variables, so that one has $\phi^T = Q\mathrm{E}\{\bar{x}\bar{u}^T\}$ for some diagonal coupling matrix Q. Following the above guidelines, it is easy to see that the matrix of eigenvalues for $\mathrm{E}\{\bar{u}\bar{u}^T\}$ becomes Q^{-1}. What is more interesting, is that one can derive for the symmetric matrix $\mathrm{E}\{\bar{x}\bar{x}^T\}$ two expressions: Simultaneously there holds $\mathrm{E}\{\bar{x}\bar{x}^T\} = Q\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T$ and $\mathrm{E}\{\bar{x}\bar{x}^T\} = \mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^TQ$. For non-trivial Q, and if the eigenvalues are distinct, this can only hold if latent vector covariance is

[3]The property (3.36) has also practical consequences. Recognizing that the Hessian $d^2J(x)/dxdx^T$ of the criterion (3.5) becomes a scaled identity matrix, it is evident that the originally first-order convergence properties of the gradient descent process (3.4) change into second-order dynamics, the process becoming an implementation of Newton method towards reaching the balance $\bar{x}$ after a transient

diagonal; what is more, the vectors in $\bar{\theta}^T = \sqrt{Q}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\mathrm{E}\{\bar{x}\bar{u}^T\}$ now not only span the principal subspace, but they are the PCA basis vectors themselves (basis vectors not necessarily ordered in the order of significance). This means that the modes become separated from each other if they are coupled to the environment in different degrees.

The eigenvectors of u are the same as those of $\bar{u}$, but the eigenvalues are evidently not. Now study how the realizable mapping ϕ^T affects on the virtual mapping between u and $\bar{x}$. From (3.35) one has

$$\bar{x} = \sqrt{q}\,\mathrm{E}\{\bar{x}\bar{u}^T\}\,(u - b\varphi\bar{x})\,, \tag{3.41}$$

and, when solving for $\bar{x}$ and when recognizing (3.36),

$$\begin{array}{rcl} \bar{x} & = & \left(I_n + bq\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T\right)^{-1} q\mathrm{E}\{\bar{x}\bar{u}^T\}\,u \\ & = & \left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^{-1} q\mathrm{E}\{\bar{x}\bar{u}^T\}\,u, \end{array} \tag{3.42}$$

so that

$$\left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)\,\mathrm{E}\{\bar{x}\bar{x}^T\}\,\left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right) = q^2\,\mathrm{E}\{\bar{x}\bar{u}^T\}\mathrm{E}\{uu^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T,$$

or, utilizing (3.39),

$$\begin{array}{l} \left(I_n + b\mathrm{E}\{\bar{x}\bar{x}^T\}\right)^2 \\ \qquad = q\,\sqrt{q}\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\,\mathrm{E}\{\bar{x}\bar{u}^T\}\,\mathrm{E}\{uu^T\}\mathrm{E}\{\bar{x}\bar{u}^T\}^T\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1/2}\sqrt{q} \\ \qquad = q\,\bar{\theta}^T\,\mathrm{E}\{uu^T\}\,\bar{\theta}. \end{array}$$

This comes from the fact that $M\,f(M) = f(M)\,M$ for a square matrix M and a function f that is defined in terms of a matrix power series. From the fact that the eigenvectors $\bar{\theta}_j$ of $\mathrm{E}\{\bar{u}\bar{u}^T\}$ are also eigenvectors of $\mathrm{E}\{uu^T\}$, one has

$$\mathrm{E}\{\bar{x}\bar{x}^T\} = \frac{\sqrt{q}}{b}\,\bar{\theta}^T\mathrm{E}\{uu^T\}^{1/2}\bar{\theta} - \frac{1}{b}\,I_n. \tag{3.43}$$

The eigenvalues of $\mathrm{E}\{\bar{x}\bar{x}^T\}$ also can be expressed in terms of the n most significant eigenvalues λ_j of the undisturbed $\mathrm{E}\{uu^T\}$:

$$\frac{\sqrt{q\lambda_j} - 1}{b}. \tag{3.44}$$

As compared to the discussion in Section 3.2.2, the refined model has qualitatively very different properties: Whereas in the nominal principal component model the variation in input is maximally inherited by the latent structure, so that $\sum_{i=1}^n \mathrm{E}\{\bar{x}_{ii}^2\} = \sum_{j=1}^n \lambda_j$, now there is loss of variation within the system.

3.3 Analysis of elasticity

This section concludes the mathematical analysis of the generic neocybernetic framework. Intuitively, it is elasticity that will pop up every now and then in the subsequent analyses, and the conceptually farthest-ranging consequences come from the rigidity of the feedback structure: The environment changes its outlook because of the systems in it.

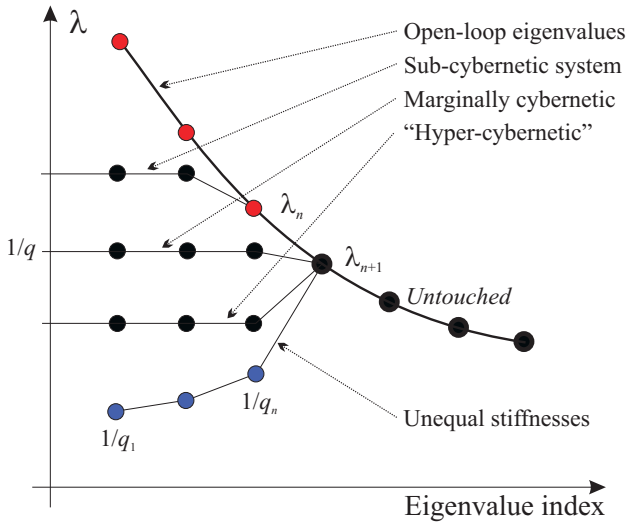

Figure 3.4: Schematic view of the cybernetic eigenvalues

3.3.1 Balance between system and environment

Because of the cybernetic scaling in the form $\mathrm{E}\{\bar{x}\bar{x}^T\}^{-1}$, the latent variables cannot go to zero, and a balance is found where the opposing drifting effects are compensated. In the directions dictated by the mapping matrices ϕ and φ (or $\mathrm{E}\{\bar{x}\bar{u}^T\}^T$), there is loss of excitation in the environment, as studied in Section 3.2.4, so that equalization of environmental variation takes place. This kind of "trivialization" of the environment is implemented not only through adaptation in the system, but also through changes in the environment. These results concerning "constant elasticity" are of extreme importance and they will be studied later.

It is also so that the environmental variation is suppressed, but simultaneously it is inherited by the system manipulating the environment. To reach such cybernetic situation where all n latent variables remain occupied, from (3.44) it is evident that there must hold $q\lambda_n > 1$. This means that there has to exist enough excitation to invoke the system, and make the adaptation process *without* the feedback unstable.

It is also possible to have separate values for q_i and b_i in different feedback loops, represented by different latent variables x_i, so that mappings $\phi^T = Q\mathrm{E}\{\bar{x}\bar{u}^T\}$ and $\varphi^T = B\mathrm{E}\{\bar{x}\bar{u}^T\}$ become "species-specific":

$$Q = \begin{pmatrix} q_1 & & 0 \\ & \ddots & \\ 0 & & q_n \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{pmatrix}. \tag{3.45}$$

Then it is not the principal subspace only that is constructed in the cybernetic process — it turns out that different eigenvalues are localized, and $\mathrm{E}\{\bar{x}\bar{x}^T\}$ becomes diagonal. The remaining covariance matrix corresponding to the cybernetic modes in the environment is, as projected onto the n-dimensional principal

subspace,

$$\begin{pmatrix} \frac{1}{q_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{q_n} \end{pmatrix}, \tag{3.46}$$

and the induced covariance matrix of the cybernetic modes in the system is

$$\begin{pmatrix} \frac{\sqrt{q_1\lambda_{j(1)}}-1}{b_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\sqrt{q_n\lambda_{j(n)}}-1}{b_n} \end{pmatrix}. \tag{3.47}$$

Here, notation $j(i)$ means that any permutation of the n most significant eigenvalues of $\mathrm{E}\{uu^T\}$ is possible. It turns out that all cross-correlations among system variables are eliminated, $\mathrm{E}\{\bar{x}\bar{x}^T\}$ being diagonal; the covariance $\mathrm{E}\{\bar{u}\bar{u}^T\}$ is not diagonal, though. It also turns out that when the feedback is implemented through the environment, one can have $n = m$ without losing the cybernetic properties of the system. To be sure that all modes are cybernetic, there must hold

$$q_i\lambda_n > 1. \tag{3.48}$$

In Figure 3.4, such situation where all modes fulfill the above constraint, is called (marginally) cybernetic, whereas cases where the "coupling" is too weak is called "sub-cybernetic". At least for some of the latent variables, in the closed-loop system one has $0 = q_i\mathrm{E}\{\bar{x}_i\bar{u}^T\}\bar{u}$ — the simplest solution for this is the trivial $\bar{x}_i \equiv 0$ and $\mathrm{E}\{\bar{x}_i\bar{u}^T\} = 0$, and the latent variable can fade away altogether. In real, converged systems, it can also be assumed that existent, non-vanishing latent structures cannot be sub-cybernetic. Further, looking at Fig. 3.4: If the (visible) variation structure changes so that the ordering of the eigenvalues becomes blurred, less significant variation directions outweighing the originally more significant ones, the situation is called "hyper-cybernetic". Note that the system still sees the original variation in u rather than the compensated in $\bar{u}$, so that there are no convergence problems however high the values of q_i are.

The parameters q_i and b_i remain free design parameters: Different kinds of system / environment combinations are instantiated for different selections of them, all of them equally valid, as long as (3.48) is fulfilled. Now it is possible to interpret these coupling coefficients in intuitive terms:

- **Stiffness ratio** q_i determines how tightly connected the system is into its environment, and how aggressively the system affects the environment, directly determining how "rigid" the corresponding direction in the data space is.
- **Dissipation rate** b_i determines how efficiently variation on the lower level (environment) is transferred onto the higher level (system itself). The non-transferred portion can be seen as loss of resources — see next chapters for closer analyses.

To assure cybernetic operation of the system, one can also make q_i adaptive. For example, local manipulations only are needed if one selects ($\nu > 1$ being some constant)

$$q_i = \frac{\nu}{\mathrm{E}\{\bar{x}_i^2\}}. \tag{3.49}$$

However, for a strictly cybernetic variable the above expression is automatically fulfilled, as the variance of the variable is relative to the inverse of the coupling factor, and other kinds of adaptation strategies for q_i can be proposed (see chapter 6).

3.3.2 Power of analogies

When applying linear models, the number of available structures is rather limited – indeed, there exist more systems than there are models. This idea has been applied routinely: Complicated systems are visualized in terms of more familiar systems with the same dynamics. In the presence of modern simulation tools, this kind of lumped parameter simplifications seem somewhat outdated — however, in the case of really complicated distributed parameter systems, such analogies may have reincarnation.

Mechanical associations

The original intuition concerning mechanical deformable systems in Sec. 3.1.2 can be extended. Think of a steel plate: If there are external forces acting on the plate, there is a continuum of smooth deformations on the surface. The plate is a distributed parameter system, but the distinct actors are like "probes", discretizing the state space, channeling the infinite-dimensional system onto the finite set of variables. Not all forces affecting the system can be detected, and not all deformations can be compensated — but what comes to the *visible* phenomena, projected through the observables onto the realm of concrete numbers, they can be mastered in the neocybernetic framework, exploiting the above observations: The variation structures become restructured.

The infinite complexity of the environment (the "forces") are mapped onto the measurements ("deformations"). A special case — but typical in practical systems — is the distributed case where individual observations and feedbacks are paired: Only the local environment can be observed, and it is this local environment that is mainly affected by the corresponding feedback. Now the structure of the environment is determined by this setup: No matter where the sensor/actuator pairs are located, the deformations in those locations become equalized and separated. The system variables are *a priori* fixed, and the whole infinite-dimensional "world" becomes anchored by the sensor/actuators.

In a mechanical system, such sensor-actuators are naturally separated in space. However, in more abstract systems, separation is not about spatial but about more complicated (higher-dimensional) dependency structures. The mechanical analogy makes the high-dimensional domain fields better graspable, projecting the wealth of simultaneous variables into the wealth of locations along the hypothetical plate, the interpretations of the semantics-loaded variables being made

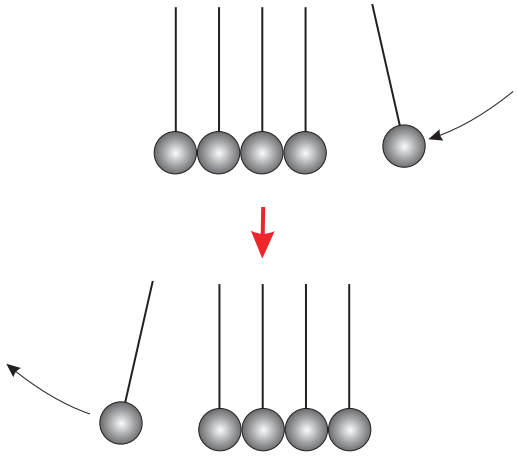

Figure 3.5: Maximum energy transfer is reached when impedances match. The physical units of "impedance" can vary (here it is the *masses* that need to be matched)

commeasurable — everything is only about "forces" and "deformations". It can be said that *dimensional complexity changes into spatial diversity.* These issues are studied closer in the subsequent chapters.

Electrical understanding

Another type of analogues are also routinely constructed: One can select *electrical current* and *voltage* rather than force and deformation. The external forces change to electrical loads disturbing the system: The deformation is the voltage drop, and the compensating action is the increased current supply (or vice versa). Traditionally, the non-idealities (output voltage drops when current is used) make it difficult to study interconnected groups of systems — the information flow is not unidirectional — but now the neocybernetic framework makes it possible to *exploit* these underlying feedbacks, even though they are implicit. Applying the distributed parameter framework instead the traditional lumped parameter one, one can reach again new intuitions, getting rid of SISO thinking.

The electric analogy makes it possible to extend the inner-system discussions onto the environmental level, to inter-system studies. When there are many connected systems interacting, one subsystem exhausting emergy supplied by the other subsystems — or providing emergy for the others, or transferring emergy between them — the criterion for system fitness can be based on the power transmission capabilities among the systems. And it is the product of current and voltage that has the unit of power, so that exactly the above discussions apply. Only the intuitions change: Now one can utilize the *inter-system* understanding supplied by electrical systems. Remember that the maximum throughput without "ringing" between electrical systems is reached when there is *impedance matching:* The output impedance in the former system and the input impedance of the latter one should be equal, otherwise not all of the power goes through but bounces back (however, in a non-mechanical/non-electrical system, there is not necessarily inertia, and no oscillatory tendency). This same bouncing metaphor can be applied efficiently also in non-electrical environments — the variables can have different interpretations but the qualitative behaviors

remain the same (for example, see Fig. 3.5). It is not only local agent-level optimization that results in global system-level optimization, it is local system-level optimization that finally results in global environment-level optimization. These intuitions will be exploited in the following chapter.

Again, it is natural to study systems where the pairs of input and output variables are localized. Assume that $m = n$ and the variables are coupled as pairs, that is, the mapping matrix ϕ is diagonal, and $\bar{u}_i$ and $\bar{x}_i$ go together. The electrical analogy makes it possible to interpret the role of the coupling coefficients q_i in the formulas in a new way. As it is this parameter that connects the input (voltage) and the output (current) for cybernetic systems, $\bar{x}_i = q_i \mathrm{E}\{\bar{x}_i \bar{u}_i\}\, \bar{u}_i$, it is $Z_i = 1/q_i \mathrm{E}\{\bar{x}_i \bar{u}_i\}$ that explicitly stands for impedance. It is also so that in an evolutionary surviving environment the corresponding impedances have to become equal. This means that there is yet another iterative optimization loop — this time not within one system, but between all pairs of systems within an environment.

The field of electrical engineering is a highly sophisticated branch of powerful mathematics, and there developed conceptual tools can directly be exploited also in the analysis of cybernetic systems. There are some extensions that are needed:

- This far, only real-valued variables have been seen reasonable, and the models have been constructed accordingly. However, if transpositions are always changed to Hermitean matrices, so that in addition to transposing the matrices the elements are also complex conjugated, all of the above analyses can directly be extended to complex domain, so that all variables and matrices can consist of real and imaginary parts.
- What is more, only scalar variables have this far been reasonable. However, the variables can be extended to function domain: The variables can be parameterized, so that the constructed models and data structures remain functions of these parameters. So, if the extra parameter is the angular frequency ω, the analyses can be carried out in frequency domain — and then one needs the complex variables.

The above extensions make it possible to study dynamic phenomena by applying essentially the same formulas. Impedances $Z_i(s)$ can be interpreted in terms of *dynamic filters* between Laplace-transformed signals $\bar{U}_i(s)$ and $\bar{X}_i(s)$, being transfer functions of the complex variable s. The explicit spectra of $Z_i(s)$ can be found for values $s = \mathrm{i}\omega$, and the inverse transforms as $\bar{x}(t) = \mathcal{L}^{-1}\bar{X}(s)$. This means that it is not only the final balance that can be studied in the neocybernetic framework but also the stationary non-balance phenomena — and, indeed, dynamic models are most appropriate for real life systems, where understanding of how they behave during transients is very relevant. To reach best possible power transfer it is also these frequency-domain functions $Z_i(s)$ that need to match in neighboring systems. If the system can efficiently affect its environment, there is an iteration process where all systems constituting the environment adapt to find the common balance, that is, the objective $Z_i(s)$ are not given *a priori*.

Further, if the whole environment is evolutionarily optimal, it is the above observations that characterize the behaviors: The matrix $\mathrm{E}\{\bar{X}(s)\bar{X}^H(s)\}$ can

be interpreted as a matrix of *autospectra* and *cross-spectra* for signals in $\bar{X}(s)$. Again, there are surprises:

> Comparing to (3.36), it is evident that the functions $q_i(s)$ must be selected so that they can be interpreted as autospectra, so that $q_i(s) = q_i'(s)q_i'(-s)$ for some valid transfer function $q_i'(s)$. The power spectrum must be real (and non-negative) for all frequencies, meaning that there exist no phase properties present in such spectra. This means that in $q_i(\mathrm{i}\omega)$ there are no phase properties, the transfer function containing no actual dynamics — meaning that $q_i(\mathrm{i}\omega)$ must have the same value for all frequencies: Coupling coefficient q_i must be constant and real.

The only remaining degrees of freedom in this extremum is the values of the interaction constants q_i and b_i. And, indeed, such questions are very relevant in everyday systems — or, actually, they may be relevant actually to "everything" ... see chapter 9 for more discussion. The system-internal frequency-domain considerations are elaborated on from another perspective again in chapter 5.

3.3.3 Applications in engineering systems

What happens if the evolutionary adaptation scheme is applied in technical systems? The discussions above were idealized, assuming absolute evolutionary optimality (as defined in terms of emergy transfer). However, to exploit the intuitions in real-life systems where the assumptions about maximum coupling with the environment do not hold, some more analysis is needed.

Assume that the system is man-made, meaning that the system state can freely be manipulated; the problem is that the inverse effect back from $\bar{x}$ to $\bar{u}$ typically is not optimized in the sense of emergy transfer. Still, it needs to be recognized that the property (3.36), and other observations therein, are general properties that always apply to the neocybernetic adaptation of ϕ in the form (3.13), so that the (visible) environmental signals are equalized. Assume that for physical reasons the feedback mapping is fixed F instead of adaptive φ.

If this inverse mapping F does not follow the "Hebbian learning" principle, so that (3.14) does *not* hold, does the whole cybernetic structure collapse? The answer to this question is *no.* Assuming that the feedback still can implement stabilization, the system in Fig. 3.3 will search a balance so that

$$\Phi^T F = I_n, \tag{3.50}$$

so that the feedforward and feedback mappings still have to be mutually orthogonal. Again, the feedback structure $\Delta u = F\bar{x}$ can be written as in (3.20); to make (3.50) hold, Φ again has to be given by (3.24), and it has to span the principal subspace of $\mathrm{E}\{\Delta u \Delta u^T\}$, because (3.29) still holds.

Even though the feedback structure in the system is fixed, the system properties remain essentially the same. The system cannot escape the subspace determined by F — but within that subspace, the model is optimized to tackle with observed variances. Even though the feedback matrix F perhaps cannot be

affected, statistical properties of the signals can; after adaptation F spans the principal subspace of the converged environment as seen in signals Δu, making the originally non-ideal system ideal after all, in its own narrow world. The adaptation strategy does not allow trivial solutions, but excites the system by force. When the properties of the environment change, it starts reflecting the peculiarities of the system and its non-idealities — in this sense, it becomes questionable whether it is the system itself or the environment that implements the cybernetic adaptation.

It turns out that the latent variables can also be selected freely: Assume that there exists some $x' = Dx$ for some invertible mapping matrix D. Then, the original formulation $\bar{x} = q\,\mathrm{E}\{\bar{x}u^T\}\bar{u}$, when multiplied by D from the left, is identical with a new one, where only the variable x' is employed:

$$\bar{x}' = q\,\mathrm{E}\{\bar{x}'\bar{u}^T\}\,\bar{u}. \tag{3.51}$$

Utilizing these observations, the cybernetic studies can be applied for analysis of non-ideal real-life systems, where complete reciprocity of the data transfer structures does not *originally* hold. For example, in some cases these $\bar{x}'_i$ can be selected as the actual control signals acting on the system, as studied below.

Distributed controls

The above results make it possible to implement, for example, new kinds of sensor/actuator networks. In traditional agent systems, the issues of co-operation and shared "ontologies" are difficult; in the current setting, such problems become trivial: Each agent just tries to exploit the available resources in the environment. There is no need for negotiation as the interactions and feedbacks are implemented automatically through exhaustion of the resources. From the engineering point of view, it is nice that the goal of the agents — exhaustion of the variation in the environment — is parallel with the the goal of regulatory control (see [92]).

If the agents share the measurement information, transmitting the local measurements to the neighbors, the principal components oriented control of the environment is implemented after adaptation. If this assumption of complete information does not hold, the model becomes distorted: For example, if an agent only knows its own measurement, if there is no communication whatsoever among the agents, the operation of the control network becomes highly localized, even though there still is feedback through the environment.

As studied closer in the next chapter, the set of sensor/actors implements discretization of infinite-dimensional partial differential equations, the sensor/actuator nodes acting as discretization centers. Simultaneously the active participation of these nodes transforms the environment to fit the cybernetic structures. This control scheme can be applied, in principle, in chemical systems (the actuators adding chemicals if the measurements are low), or in thermal systems (the actuators heating the environment). An especially good application example is offered by mechanical systems, where the deformations and interactions are manifested practically delaylessly when some external forces are applied. In [92], examples of cybernetic "stiffening" of a steel plate are presented. This scheme can be applied also for design of mechanical structures, as shown below.

Design of mechanical structures

If the sensor/actor network is (virtually) extended over the whole mechanical construction, the network of controllers becomes more or less continuous; such settings can be studied, for example, in mechanical design systems that are equipped with finite element method (FEM), or, perhaps more appropriately, boundary element method solvers. Then one can apply the assumed forces onto the construction, calculate the deformations (or, more appropriately, the strains along the surface), and adapt the local controllers to oppose those deformations. After adaptation, there should be constant stiffness over the whole structure (see Fig. 3.6). The nice thing about this scheme is that the controls are manifested as increased stiffness, and the final "controls" can be implemented in terms of passive elements, simply adding extra layers of material in the locations of high experienced stress.

Today's design methods only take into account the maximum loads, and safety factors in specifications are needed to cope with unanticipated phenomena. Still, catastrophes take place every now and then — and typically the reason is *fatigue.* When the metal structures are under fluctuating tensions, the structures may break even though the specifications are never exceeded. Fractures are related to "gnawing". The cybernetic design approach — effectively damping and equalizing the vibrations — could offer new perspectives here.

The same idea of cybernetic designs could also be applied in frequency domain: At least in principle, (active) damping of vibrations can be implemented in this way. Similarly as in the static case of mechanical constructs, the system needs to be studied as a whole, as local damping actions can make damping efforts in neighboring nodes redundant, and iterative adaptation hopefully results in damping and equalization of vibrations. Here, the extension of the cybernetic framework to modeling of (discretized) functions is needed: The sensor/actor nodes host a family of input variables, these variables characterizing the measured energies at separate frequency bands in the power spectrum. Simpler implementation of vibration damping is reached if one concentrates on the velocities: Then the "information" being exhausted, or average of velocity squared, is proportional to the kinetic energy.

Optimization of parameters

The idea of cybernetic adaptation and constant stiffness against environmental disturbances can also be extended to large-scale industrial plants where there also is elasticity: A reasonably designed system can sustain environmental disturbations and other changes in a more or less robust way. Smoothly changing of parameters in the system (control parameters) or in the environment (set points, etc.) pushes the operating point of the plant in an elastic manner. Today, the low-level controls are typically poorly tuned, and separate control loops can have very different time constants, others being sluggish and other ones being faster. Uneven stiffness is manifested exactly in such heterogeneity between subsystems, and one can assume that cybernetic adaptation of the control parameters could make the subsystems better compatible.

In technical systems, however, there are domain-specific goals for evolution.

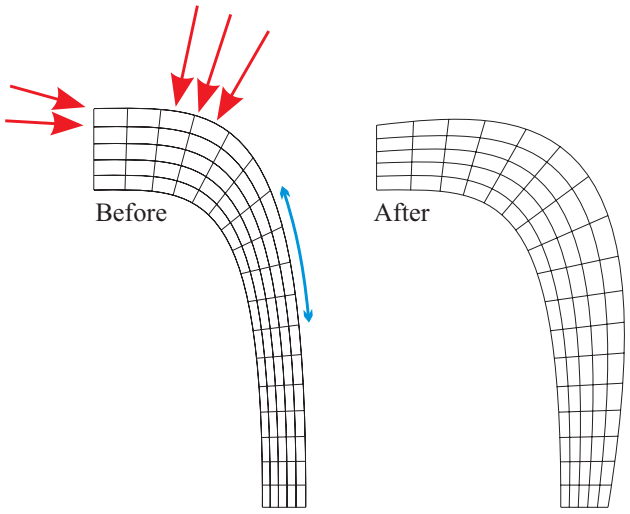

Figure 3.6: Iterative redesign of mechanical structures

One would not simply like to blindly adapt towards maximum emergy transfer, as assumed above, but one would like to maximize the match with the environment and the *intended* system functionalities. When the vector of system functionalities in x is predetermined by an external designer to contain some kind of quality measures, characterizing the "goodness" of operation, guided evolution is possible. The latent structure between u and x can technically be implemented in terms of not only correlations among variables in u (in the PCA style), but also in terms of cross-correlations between u and the intended x (in the PLS or CCR style, for example — see [42]). When the design parameters in u are seen as variables on the slower time scale, evolution towards better parameter values implementing higher values of $\bar{x}$ can locally be seen as pressing the elastic system into a desired direction along the determined axes of "quality freedom" (see [92], Report 139: "Process Performance Optimization Using Iterative Regression Tuning").

3.4 Towards *complex* complex systems

How are the above abstract assumptions about evolution related to real-life observations? Indeed, it seems that *increase of stiffness,* or *hyperplasia,* is the key behavior in natural adaptation processes. For example, skin becomes thicker if it is burdened, and a muscle becomes stronger if it is used (reactions of the neural system to signal activation are discussed in detail in chapter 7). Similarly, companies invest money and employ new staff if there is very much activity. This kind of trivial-looking behaviors, when boosted with self-regulation and self-adaptation, result in global-level system properties that can be described in terms of principal components. Because of the properties of PCA, adaptation in the assumed form maximally compensates the external disturbances.

This far, simple systems have been studied. The key observation is that, when seen at the correct level of abstraction, "all" complex reasonable systems are elastic. Elasticity offers tools to attack really complex, formless systems. From

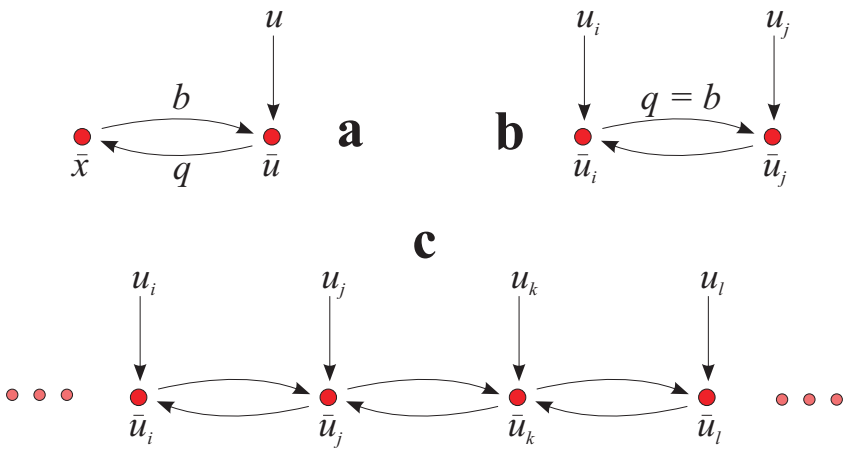

Figure 3.7: Extending the cybernetic framework: The studied cybernetic system structure (a) is symmetric what comes to the interactions, and the roles of the system and the environment can be shared (b) — and, finally, the environmental variables can be physically distributed (c)

now on, no accurate mathematical analyses are available any more: One just has to trust in the strong modeling principles, and intuition. These elasticity ideas are closer studied in the subsequent chapter.

As a brief introduction to extensions of elasticity considerations, look at Fig. 3.7. It turns out that there can exist a wealth of neighboring systems that are more or less tightly connected together; from the point of view of a single subsystem, the neighbors together constitute the environment. As it is various neighbors that see the same system as their environment, the coupling factors q in different subsystems must become identical as they see the same level of variation in their environments. As seen from outside, it is only the coupling coefficients that remain, determining the dynamic properties of the system. As the number of neighbors grows, dynamic transitions become diffusion processes among differential elements. In any case, the local adaptation as presented before still gives consistent results. In physical systems the interactions are concrete, but they need not be — it is all about information transfer. Interchange of the roles of the system and the environment is studied in more detail in the next chapter.

As a conclusion of this chapter, it can be observed that within the neocybernetic framework, local learning has globally meaningful results. As seen from functional point of view, new interpretations for cybernetic systems are available:

- **First-order cybernetic system** finds balance under external pressures, pressures being compensated by internal tensions. Any existing (complex enough) interacting system that can maintain its integrity in a changing environment is cybernetic in this sense. First-order cybernetic system momentarily implements *minimum (observed) deformation emergy in the system.*

- **Second-order cybernetic system** adapts the internal structures to better match the observed environmental pressures, towards maximum expe-

rienced stiffness. Any existing (competing) interacting system that has survived in evolution finally is cybernetic in this sense. Second-order cybernetic system additionally implements *minimum average observed deformation emergy in the system.*

- **Higher-order cybernetic system** adapts the external structures of the system to better match the observed environmental structures by adjusting the impedances. Evolutionarily optimal environment, or system of systems, assumedly only contains higher-order cybernetic systems. Higher-order cybernetic system implements *maximum average transfer of emergy through the environment.*