Chapter 4
Model-Driven Evaluation of the Emergent Complexity of Cooperative Work Based on Effective Measure Complexity

In this book, the main improvement on Grassberger's original definition of the effective measure complexity EMC, which is based on classic information-theoretic quantities like Shannon's information entropy that were developed to evaluate stochastic processes with discrete states, is the generalization of the theory and measures to continuous-state processes like that generated by the previously introduced VAR(1) model of cooperative work according to state Eq. 8. However, Li and Xie (1996), Bialek et al. (2001), de Cock (2002), Bialek (2003), Ellison et al. (2009) and others have already pioneered the generalization of Grassberger's concepts toward continuous systems in their works, and we can build upon their results. Their analyses show that we must primarily consider the so-called "differential block entropy" (Eq. 233) and the corresponding continuous-type mutual information (Eq. 234) as basic information-theoretic quantities.

In general, the differential entropy extends the basic idea of Shannon's information entropy as a universal measure of uncertainty about a discrete-type random variable with known probability mass function over the finite alphabet $\mathcal{X}$ to a p-dimensional continuous-type variable X with a probability density function $f[x]$ (*pdf*, see previous chapters) whose support is a set $\mathbb{X}^p$. The differential entropy is defined as:

$$H[X] := -\int_{\mathbb{X}^p} f[x] \log_2 f[x]\, dx. \tag{232}$$

The differential block entropy (cf. Eq. 219) is defined in an analogous manner as:

$$H(n) := H[X^n] = -\int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f[x_1, \ldots, x_n] \log_2 f[x_1, \ldots, x_n]\, dx_1 \ldots\, dx_n. \tag{233}$$

In the above equation $f[x_1, \ldots, x_n]$ denotes the joint *pdf* of the vectors $(X_1, \ldots, X_n)$ with support $\mathbb{X}^{np}$.

C. Schlick, B. Demissie, *Product Development Projects*, Understanding Complex Systems, DOI 10.1007/978-3-319-21717-8_4

The information entropy of a discrete-type random variable is non-negative and can be used as a measure of average surprisal. This is slightly different for a continuous-type random variable, whose differential entropy can take any value from $-\infty$ to ∞ and is only used to measure changes in uncertainty (Cover and Thomas 1991; Papoulis and Pillai 2002). For instance, the differential entropy of a continuous random variable X that is uniformly distributed from 0 to a (and whose *pdf* is therefore $f[x] = 1/a$ from 0 to a, and 0 elsewhere) is $\log_2 a$. For $a < 1$ the differential entropy is negative and can become arbitrarily small as a approaches 0. The differential entropy measures the entropy of a continuous distribution relative to the uniformly distributed one. For a Gaussian distribution with a variance of σ^2 the differential entropy is $H[X] = 1/2\log_2\sigma^2 + const$. Thus the differential entropy can be regarded as a generalization of the familiar notion of variance. With a normal distribution, the differential entropy is maximized for a given variance. An additional subtlety is that the differential entropy can be negative or positive depending on the coordinate system used for encoding the vectors. This also holds true for the differential block entropy. However, it can be proven that the complexity measure EMC calculated on the basis of dynamic entropies (cf. Eqs. 224 and 225) is always positive and may exist even in cases where the block entropies diverge. Under the assumption of an underlying VAR model, for instance, a closed-form solution for the EMC can be derived that is simply a logarithmic ratio of determinants of covariance matrices (cf. Eqs. 246 and 258), which in most industrial case studies is a real number that is much larger than zero. In this case, the generalized complexity measure can be interpreted similarly to discrete-state processes. Furthermore, it can be proven that for finite complexity values EMC is independent of the basis in which the state vectors of work remaining are represented, and is invariant under linear transformations of the state-space coordinates for any regular transformation matrix (Schneider and Griffies 1999). This invariance is due to the fact that the measure can be expressed as the continuous-type mutual information $I\left[X_{-\infty}^{-1}; X_0^{\infty}\right]$ between the infinite past and future histories of a stochastic process, where the base-independent mutual information $I[.;.]$ between the sequences $X_1^n = (X_1, \ldots, X_n)$ and $Y_1^m = (Y_1, \ldots, Y_m)$ of random vectors with support $\mathbb{X}^{nq}$ and $\mathbb{Y}^{mp}$ is defined as

$$I\left[X_1^n; Y_1^m\right] := \int_{\mathbb{X}^q} \cdots \int_{\mathbb{Y}^p} f[x_1, \ldots, x_n, y_1, \ldots, y_m]$$
$$\log_2 \frac{f[x_1, \ldots, x_n, y_1, \ldots, y_m]}{f[x_1, \ldots, x_n] f[y_1, \ldots, y_m]} dx_1 \ldots dx_n dy_1 \ldots dy_m \,. \tag{234}$$

For two random variables X and Y that are jointly normal with a correlation coefficient of ρ there is $I[X; Y] = 1/2\log_2(1 - \rho^2)$. As such, the mutual information can be viewed as a generalized covariance. Kraskov et al. (2004) published a simple proof that the mutual information as defined in Eq. 234 is not only invariant under linear transformations but also with respect to arbitrary reparameterizations based on smooth and uniquely invertible maps $x'_1 = x'_1(x_1), \ldots, x'_n = x'_n(x_n), y'_1 = y'_1(y_1), \ldots, y'_m = y'_m(y_m)$. Therefore, $I[.;.]$ provides a measure of statistical dependency structures between variables that is independent of the subjective choice of the measurement instrument. The analyses of Bialek et al. (2001) and other researchers show that this measure is a valid, expressive and consistent quantity for evaluating emergent complexity in open systems.

In the following chapters the generalization of the EMC to project organizations that are modeled by continuous state variables will be carried out step-by-step. Though some of the calculations are quite involved, the interested reader will find that they lay important groundwork for the complexity analysis of cooperative work in various kinds of open organizational systems, not only product development organizations. In Section 4.1, we start by calculating closed-form solutions with different strength for the vector autoregression models that were introduced in Sections 2.1, 2.2 and 2.4. These models do not have "hidden" state variables and therefore are quite easy to analyze in information-theoretic terms. To simplify the analysis a generalized solution for a VAR(1) process that does not refer to a specific family of *pdf*s of the unpredictable performance fluctuations is calculated in Section 4.1. We will use this generalized solution to derive closed-form solutions for the original state space (Section 4.1.1) and the spectral basis (Section 4.1.2) under the assumption of Gaussian behavior. Furthermore, a very compact closed-form solution will be obtained through a canonical correlation analysis (Section 4.1.3). For these three different approaches, we will also present the corresponding closed-form solutions of the persistent mutual information EMC(τ) (Eq. 229) according to Ball et al. (2010). Moreover, to clarify the concept of emergent complexity, polynomial-based solutions for simple processes with two and three tasks are presented in Section 4.1.4. This chapter also includes a short analytical study of minimizing emergent complexity subject to the constraint that the expected total amount of work done over all tasks is constant. Moreover, lower bounds are put on the EMC in Section 4.1.5. In Section 4.2.1, an additional explicit closed-form solution for a Markov process with hidden variables (a linear dynamical system, LDS, see Section 2.9) is calculated. This solution is, admittedly, complicated and difficult to interpret because the state variables of cooperative work that are not directly accessible can generate a significant number of long-range correlations between observations, and a great deal of linear algebra is needed to evaluate the associated infinite-dimensional integrals. Therefore, Section 4.2.2 will introduce two additional implicit formulations for the EMC. The first implicit solution is based on the seminal work of de Cock (2002) and allows analogical reasoning between the forward and backward innovation forms developed in Section 2.9, and the generated past-future mutual information. The second implicit solution is directly derived from the infinite-dimensional integrals and makes it possible for the interested reader to gain additional insights into the information-generating mechanisms by following the calculation step by step. Although the closed-form solutions for LDS are significantly more complicated, their derivations show that Grassberger's theory can, in principle, be applied in a straightforward manner to a larger model class that, thanks to its informational richness and predictive power, is especially attractive for applications in project management.

4.1 Closed-Form Solutions of Effective Measure Complexity for Vector Autoregression Models of Cooperative Work

To obtain analytical results, it is assumed that the parameterized VAR(1) process $\{X_t\}$ is strict-sense stationary (Puri 2010) and therefore all its statistical properties (especially the first and second moments) are invariant to a shift in the chosen time origin. Let $f_\theta[x_{t+1}, \ldots, x_{t+n}]$ $(t \in \mathbb{Z}, n \in \mathbb{N})$ be the joint *pdf* of the block of vectors $(X_{t+1}, \ldots, X_{t+n})$ generating the stochastic process, and let $f_\theta\left[x_{t+n} \middle| x_{t+1}, \ldots, x_{t+n-1}\right]$ be the conditional density of vector X_{t+n} given vectors $X_{t+1}, \ldots, X_{t+n-1}$. We use the shorthand notation $f[.]$ and $f\left[. \middle| .\right]$ in the following to denote these density functions. Due to strict sense stationarity the joint distribution of any sequence of samples does not depend on the sample's placement:

$$f[x_{t+1}, \ldots, x_{t+n}] = f[x_{t+1+\tau}, \ldots, x_{t+n+\tau}] \quad (t \in \mathbb{Z}, n \in \mathbb{N}, \tau \geq 0)\,.$$

We can use the index υ instead of t to express the shift-invariance. Therefore, $f[x_{v+1}, \ldots, x_{\upsilon+n}]$ denotes the joint *pdf* and $f\left[x_{\upsilon+n} \middle| x_{\upsilon+1}, \ldots, x_{\upsilon+n-1}\right]$ denotes the conditional density of the process in the steady state. The conditional density is given by (cf. Billingsley 1995):

$$f\left[x_{\upsilon+n} \middle| x_{\upsilon+1}, \ldots, x_{\upsilon+n-1}\right] = \frac{f[x_{\upsilon+1}, \ldots, x_{\upsilon+n}]}{f[x_{\upsilon+1}, \ldots, x_{\upsilon+n-1}]}\,.$$

Since the considered VAR(1) process is a Markov process (Eq. 18), the conditional density simplifies to

$$f\left[x_{\upsilon+n} \middle| x_{\upsilon+1}, \ldots, x_{\upsilon+n-1}\right] = f\left[x_{\upsilon+n} \middle| x_{\upsilon+n-1}\right] = \frac{f[x_{\upsilon+n-1}, x_{\upsilon+n}]}{f[x_{\upsilon+n-1}]}, \tag{235}$$

and the strict stationarity condition implies (Brockwell and Davis 1991)

$$\begin{gathered} f\left[x_{\upsilon+n} \middle| x_{\upsilon+n-1}\right] = f\left[x_{\upsilon} \middle| x_{\upsilon-1}\right] = f\left[x_2 \middle| x_1\right] \\ \text{and } \; f[x_{\upsilon+n-1}] = f[x_\upsilon] = f[x_1] \;\; \forall \upsilon \geq 2\,. \end{gathered} \tag{236}$$

Furthermore, we assume that ergodicity holds, and the complexity measure can be conveniently derived using stochastic calculus based on an ensemble average or an infinite number of realizations of the unpredictable performance fluctuations (see

Puri 2010). To compute the EMC for the introduced VAR(1) process in the steady state, please recall from Eq. 216 that

$$\mathrm{EMC} = I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] .$$

According to the definition of the mutual information $I[.;.]$ from Eq. 234, we can write the information that is communicated from the past to the future as

$$I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] = \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] \log_2 \frac{f\left[x_{-\infty}^{-1}, x_0^{\infty}\right]}{f\left[x_{-\infty}^{-1}\right] f\left[x_0^{\infty}\right]}\, dx_{-\infty}^{-1}\, dx_0^{\infty}. \tag{237}$$

In the above equation the shorthand notation $f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] = f[x_{-\infty}, x_{-\infty+1}, \ldots, x_{-1}, x_0, x_1, , \ldots, x_{\infty-1}, x_{\infty}]$, $f\left[x_{-\infty}^{-1}\right] = f[x_{-\infty}, x_{-\infty+1}, \ldots, x_{-1}]$, $f\left[x_0^{\infty}\right] = f[x_0, x_1, \ldots, x_{\infty-1}, x_{\infty}]$, $dx_{-\infty}^{-1} = dx_{-\infty} dx_{-\infty+1} \ldots dx_{-1}$ and $dx_0^{\infty} = dx_0 dx_1 \ldots dx_{\infty}$ was used. Due to the Markov property (Eqs. 235 and 236) the joint *pdfs* can be factorized:

$$\begin{aligned}
f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] &= f[x_{-\infty}] f\left[x_{-\infty+1} \middle| x_{-\infty}\right] \ldots f\left[x_{-1} \middle| x_{-2}\right] f\left[x_0 \middle| x_{-1}\right] f\left[x_1 \middle| x_0\right] \ldots f\left[x_{\infty} \middle| x_{\infty-1}\right] \\
f\left[x_{-\infty}^{-1}\right] &= f[x_{-\infty}] f\left[x_{-\infty+1} \middle| x_{-\infty}\right] \ldots f\left[x_{-1} \middle| x_{-2}\right] \\
f\left[x_0^{\infty}\right] &= f[x_0] f\left[x_1 \middle| x_0\right] \ldots f\left[x_{\infty} \middle| x_{\infty-1}\right].
\end{aligned}$$

Hence, we can simplify the mutual information:

$$\begin{aligned}
I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] &= \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] \log_2 \frac{f\left[x_0 \middle| x_{-1}\right]}{f[x_0]}\, dx_{-\infty}^{-1}\, dx_0^{\infty} \\
&= \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] \log_2 f\left[x_0 \middle| x_{-1}\right]\, dx_{-\infty}^{-1}\, dx_0^{\infty} \\
&\quad - \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] \log_2 f[x_0]\, dx_{-\infty}^{-1}\, dx_0^{\infty} \\
&= \int_{\mathbb{X}^p} \int_{\mathbb{X}^p} \log_2 f\left[x_0 \middle| x_{-1}\right]\, dx_0 dx_{-1} \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] dx_{-\infty} \ldots dx_{-2} dx_1 \ldots dx_{\infty} \\
&\quad - \int_{\mathbb{X}^p} \log_2 f[x_0]\, dx_0 \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] dx_{-\infty} \ldots dx_{-1} dx_1 \ldots dx_{\infty}.
\end{aligned} \tag{238}$$

On the basis of the definitions of the marginal density functions

$$f[x_0] = \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] \, dx_{-\infty} \ldots dx_{-1} dx_1 \ldots dx_{\infty}$$
$$f[x_{-1}, x_0] = \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^{\infty}\right] dx_{-\infty} \ldots dx_{-2} dx_1 \ldots dx_{\infty}$$

we can conclude that

$$\begin{aligned} I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] &= \int_{\mathbb{X}^p}\int_{\mathbb{X}^p} f[x_{-1}, x_0] \log_2 f\left[x_0 \middle| x_{-1}\right] dx_{-1} dx_0 - \int_{\mathbb{X}^p} f[x_0] \log_2 f[x_0] dx_0 \\ &= \int_{\mathbb{X}^p}\int_{\mathbb{X}^p} f\left[x_0 \middle| x_{-1}\right] \, f[x_{-1}] \log_2 f\left[x_0 \middle| x_{-1}\right] dx_{-1} dx_0 - \int_{\mathbb{X}^p} f[x_0] \log_2 f[x_0] dx_0, \end{aligned} \tag{239}$$

or equivalently

$$I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] = \int_{\mathbb{X}^p}\int_{\mathbb{X}^p} f\left[x_1 \middle| x_0\right] \, f[x_0] \log_2 f\left[x_1 \middle| x_0\right] dx_0 dx_1 - \int_{\mathbb{X}^p} f[x_0] \log_2 f[x_0] dx_0.$$

It is evident that the second summand is the differential entropy of the random variable X_0 with probability density function $f[x_0]$. The first summand represents the entropy of the random variable X_1 conditioned on the variable X_0 taking a value in the support $\mathbb{X}^p$. The first summand therefore represents a conditional entropy that is obtained by averaging over all possible values for X_0.

Before we proceed with calculating the EMC on the basis of the generalized solution from Eq. 239 in the coordinates of the original state space $\mathbb{R}^p$, we summarize five essential properties that hold completely independent of the stochastic model generating a strict-sense stationary Gaussian process $\{X_t\}$. A Gaussian process is a stochastic process whose realizations consist of random values associated with every time step such that each random variable in the sequence has a normal distribution. In addition, every finite ensemble of random variables generating the process has a multivariate normal distribution (Puri 2010).

The five essential properties are as follows (cf. Boets et al. 2007):

1) The EMC of a strict-sense stationary Gaussian process equals zero if and only if the process is temporally uncorrelated:

$$\begin{gathered} \mathrm{EMC} = I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] = 0 \Leftrightarrow X_t = v_t \text{ with} \\ v_t = \mathcal{N}(\eta; \mu, V) \text{ and } E\left[v_t v_s^{\mathrm{T}}\right] = V\delta_{ts} \ . \end{gathered} \tag{240}$$

μ denotes the mean of the process and $s \in \mathbb{Z}$ an arbitrary time step. δ_{ts} is the Kronecker delta according to Eq. 14. The implication EMC $= 0$ can be easily deduced as Gaussian random variables being uncorrelated is equivalent to statistical independence, i.e. $f\left(X_{-\infty}^{-1};X_0^{\infty}\right)=f\left(X_{-\infty}^{-1}\right)\cdot f\left(X_0^{\infty}\right)$. A proof of the implication that the process is temporally uncorrelated involves Jensen's inequality and can be found in elementary textbooks like Cover and Thomas (1991). Concerning the state and output equations of a LDS with additive Gaussian noise (Eqs. 136 and 137), this may be realized either by setting $H=0$ or with $A_0=0$.

2) The range of values of the Effective Measure Complexity is

$$\mathrm{EMC}\in[0,\,+\infty). \tag{241}$$

This property follows directly from the canonical correlation analysis of the past $\left(X_{-\infty}^{-1}\right)$ and future $\left(X_0^{\infty}\right)$ histories of the Gaussian process (see Eq. 265 in Section 4.1.3)

$$I\left[X_{-\infty}^{-1};X_0^{\infty}\right]=-\frac{1}{2}\log_2\prod_{i=1}^{q}\left(1-\rho_i^2\right)$$

and the fact that the canonical correlations ρ_i are confined to $\rho_i\in[0,1)$ (see e.g. de Cock 2002). The variable $q>p$ denotes the effective dimensionality of the process (see Section 4.2.1). The canonical correlation analysis was introduced by Hotelling (1935) and is often used for state-space identification. The goal is to find a suitable basis for cross-correlation between two random variables—in our case the infinite, one-dimensional sequences of random variables representing the past and future histories of the process. Based on the material of Creutzig (2008) we use a common variant of the canonical correlation analysis to provide a so-called balanced state-space representation (cf. Section 4.2). Given the ordered concatenation of the variables representing the past history

$$X_{past}=\begin{pmatrix}X_{-\infty}^{\mathrm{T}} & \cdots & X_{-2}^{\mathrm{T}} & X_{-1}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$$

and the future history

$$X_{fut}=\begin{pmatrix}X_0^{\mathrm{T}} & X_1^{\mathrm{T}} & \cdots & X_{\infty}^{\mathrm{T}}\end{pmatrix}^{\mathrm{T}}$$

of the Gaussian process we seek an orthonormal base $U=\left(U^{(1)},\ldots,U^{(m)}\right)$ for X_{past} and another orthonormal base $V=\left(V^{(1)},\ldots,V^{(n)}\right)$ for X_{fut} that have maximal correlations but are internally uncorrelated. Therefore, it must hold that $E\left[U^{(i)}V^{(j)}\right]=\rho_i\delta_{ij}$, for $i,j\le min(m,n)$. $U^{(i)}$ and $V^{(j)}$ are two zero-mean random variables of dimensions m and n, respectively. The resulting basis variables $(U^{(1)},\ldots,U^{(m)})$ and $(V^{(1)},\ldots,V^{(n)})$ are called canonical variates, and the correlation coefficients ρ_i between the canonical variates are called canonical correlations. The cardinalities of the bases must be chosen in a way that is compatible with the persistent informational structure of the process. The ρ_i's are not to be confused with the introduced ordinary correlations ρ_{ij} and ρ'_{ij} from Chapter 2.

To find the orthonormal bases, we normalize with Cholesky factors. The factors are given by

$$L_{past} \cdot L_{past}^{\mathrm{T}} := \left\{ \frac{1}{N} \right\} \cdot X_{past} \cdot X_{past}^{\mathrm{T}}$$
$$L_{fut} \cdot L_{fut}^{\mathrm{T}} := \left\{ \frac{1}{N} \right\} \cdot X_{fut} \cdot X_{fut}^{\mathrm{T}}.$$

N denotes the number of samples that are taken from the stochastic process. The sample size must be sufficiently large to uncover all canonical correlations. The normalized variables $\widehat{X}_{past}$ and $\widehat{X}_{fut}$ to determine the balanced state-space representation are computed by

$$\widehat{X}_{past} = L_{past}^{-1} \cdot X_{past}$$
$$\widehat{X}_{fut} = L_{fut}^{-1} \cdot X_{fut}.$$

A singular value decomposition is carried out (see e.g. de Cock 2002, and Section 4.1.3) to identify the orthonormal bases:

$$\left\{ \frac{1}{N} \right\} \cdot \widehat{X}_{fut} \cdot \widehat{X}_{past}^{\mathrm{T}} = \Sigma_{yu} \cdot \widehat{V} \cdot \widehat{\Sigma} \cdot \widehat{U}.$$

Σ_{yu} denotes the cross-covariance between X_{fut} and X_{past}. We compute the state space by

$$\widehat{X}_t := \widehat{V}^{\mathrm{T}} \cdot \widehat{X}_{past} = \widehat{V}^{\mathrm{T}} \cdot L_{past}^{-1} \cdot X_{past}$$

and balance

$$\widehat{X}'_t = \widehat{\Sigma}^{\frac{1}{2}} \cdot \widehat{X}_t$$

such that for the covariance matrix it holds that

$$\left\{ \frac{1}{N} \right\} \cdot \widehat{X}'_t \cdot \left(\widehat{X}'_t \right)^{\mathrm{T}} = \widehat{\Sigma}.$$

The requirement that the ρ_i's be nonnegative and ordered in decreasing magnitude makes the choice of bases unique if all canonical correlations are distinct. It is important to note that for a strict-sense stationary VAR(1) process $\{X_t\}$, only the p leading canonical correlations ρ_i of each pair $\left(X_{-\infty}^{-1}, X_0^{\infty}\right)$ of subprocesses are non-zero and therefore the cardinality of the base is equal to p (de Cock 2002; Boets et al. 2007). This is due to the simple fact that the process is Markovian and so the amount of information that the past provides about the future can always be encoded in the probability distribution over the p-dimensional present state (assuming an efficient coding mechanism is used). Furthermore, because of strict-sense stationarity, all ρ_i's are less than one.

3) EMC is a strictly increasing function of each of the canonical correlations. This property also follows directly from relation 265:

$$I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] = -\frac{1}{2}\sum_{i=1}^{q} \log_2\left(1-\rho_i^2\right). \tag{242}$$

4) The EMC is invariant under a transformation of the observations X_t by a nonsingular constant matrix $T \in \mathbb{R}^{p\times p}$. When we denote the transformed observations $Z_t = T \cdot X_t$, it holds that

$$I\left[Z_{-\infty}^{-1}; Z_0^{\infty}\right] = I\left[X_{-\infty}^{-1}; X_0^{\infty}\right]. \tag{243}$$

From the explicit result in Eq. 291 for the EMC of a process that is generated by a linear dynamical system with additive Gaussian noise, one can directly derive this invariance property. Similar to the notation in Section 4.2, $x_{t_1}^{t_2}$ denotes the vector obtained by stacking the observation sequence $X_{t_1}^{t_2}$ in a long vector of size $p(t_2 - t_1 + 1) \times 1$. We define the long vector $z_{t_1}^{t_2}$ in the same way. Then we can relate the transformed observations to the original ones via $Z_{t_1}^{t_2} = (I_{t_2-t_1+1} \otimes T) \cdot X_{t_1}^{t_2}$. The covariance of the history of transformed observations follows immediately and can be related to the covariance $(C_x)_{t_1}^{t_2} = E\left[X_{t_1}^{t_2}\left(X_{t_1}^{t_2}\right)^{\mathrm{T}}\right]$:

$$\begin{aligned}
(C_z)_{t_1}^{t_2} &= E\left[Z_{t_1}^{t_2}\left(Z_{t_1}^{t_2}\right)^{\mathrm{T}}\right] \\
&= E\left[(I_{t_2-t_1+1} \otimes T) X_{t_1}^{t_2}\left(X_{t_1}^{t_2}\right)^{\mathrm{T}} (I_{t_2-t_1+1} \otimes T)^{\mathrm{T}}\right] \\
&= (I_{t_2-t_1+1} \otimes T)(C_x)_{t_1}^{t_2}(I_{t_2-t_1+1} \otimes T)^{\mathrm{T}}.
\end{aligned}$$

It is straightforward to compute EMC by using the general expression for LDS (Eq. 291) with $H = I$ and $V = 0$ as

$$
\begin{aligned}
I\left[Z_{-\infty}^{-1};Z_0^{\infty}\right] &= \frac{1}{2}\log_2\frac{\mathrm{Det}(C_z)_{-\infty}^{-1}\mathrm{Det}(C_z)_0^{\infty}}{\mathrm{Det}(C_z)_{-\infty}^{\infty}} \\
&= \frac{1}{2}\lim_{\substack{t_p\to-\infty\\ t_f\to\infty}}\log_2\frac{\mathrm{Det}(C_z)_{t_p}^{-1}\mathrm{Det}(C_z)_0^{t_f}}{\mathrm{Det}(C_z)_{t_p}^{t_f}} \\
&= \frac{1}{2}\lim_{\substack{t_p\to-\infty\\ t_f\to\infty}}\left\{\log_2\mathrm{Det}\left(\left(I_{-t_p}\otimes T\right)(C_x)_{t_p}^{-1}\left(I_{-t_p}\otimes T\right)^{\mathrm{T}}\right)\right. \\
&\qquad + \log_2\mathrm{Det}\left(\left(I_{t_f+1}\otimes T\right)(C_x)_0^{t_f}\left(I_{t_f+1}\otimes T\right)^{\mathrm{T}}\right) \\
&\qquad \left. - \log_2\mathrm{Det}\left(\left(I_{t_f-t_p+1}\otimes T\right)(C_x)_{t_p}^{t_f}\left(I_{t_f-t_p+1}\otimes T\right)^{\mathrm{T}}\right)\right\} \\
&= \frac{1}{2}\lim_{\substack{t_p\to-\infty\\ t_f\to\infty}}\log_2\frac{\mathrm{Det}(C_x)_{t_p}^{-1}\mathrm{Det}(C_x)_0^{t_f}}{\mathrm{Det}(C_x)_{t_p}^{t_f}} \\
&= I\left[X_{-\infty}^{-1};X_0^{\infty}\right],
\end{aligned}
$$

where we have used the fact that $\mathrm{Det}(A\cdot B)=\mathrm{Det}(A)\cdot\mathrm{Det}(B)$ and that for matrices $A\in\mathbb{R}^{n\times n}, B\in\mathbb{R}^{m\times m}$ we have $\mathrm{Det}(A\otimes B)=(\mathrm{Det}(A))^n(\mathrm{Det}(B))^m$.

5) If the p-component vector of all observations X_t can be divided into two separate sets comprised of the vectors $X_t^{(1)}\in\mathbb{R}^{p_1}$ and $X_t^{(2)}\in\mathbb{R}^{p_2}$ with $p=p_1+p_2$, which are completely uncorrelated,

$$
C_{XX}(\tau)=E\left[X_tX_{t+\tau}^{\mathrm{T}}\right]=E\left[\begin{bmatrix}X_t^{(1)}\\X_t^{(2)}\end{bmatrix}\begin{bmatrix}X_{t+\tau}^{(1)}\\X_{t+\tau}^{(2)}\end{bmatrix}^{\mathrm{T}}\right]=\begin{bmatrix}C_{X^{(1)}X^{(1)}}(\tau) & 0\\ 0 & C_{X^{(2)}X^{(2)}}(\tau)\end{bmatrix},
$$

then the EMC of the whole sequence of observations equals the sum of the EMC of each set resulting from the partitioning:

$$
I\left[X_{-\infty}^{-1};X_0^{\infty}\right]=I\left[\left(X^{(1)}\right)_{-\infty}^{-1};\left(X^{(1)}\right)_0^{\infty}\right]+I\left[\left(X^{(2)}\right)_{-\infty}^{-1};\left(X^{(2)}\right)_0^{\infty}\right]. \tag{244}
$$

Since uncorrelated Gaussian random variables are independent, i.e. their joint *pdf* equals the product of the individual *pdfs*—in this case

$$
\begin{aligned}
f\left(x_{-\infty}^{-1},x_0^{\infty}\right) &= f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty},\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right] \\
&= f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]\cdot f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right],
\end{aligned}
$$

the above property of additivity for uncorrelated observations can be easily verified:

$$
\begin{aligned}
I\left[X_{-\infty}^{-1};X_0^{\infty}\right] &= \int f\left[x_{-\infty}^{-1},x_0^{\infty}\right]\log_2\frac{f\left[x_{-\infty}^{-1},x_0^{\infty}\right]}{f\left[x_{-\infty}^{-1}\right]f\left[x_0^{\infty}\right]}dx_{-\infty}^{-1}dx_0^{\infty}\\
&= \int f\left[x_{-\infty}^{-1},x_0^{\infty}\right]\log_2\frac{f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]}{f\left[\left(x^{(1)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(1)}\right)_0^{\infty}\right]f\left[\left(x^{(2)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(2)}\right)_0^{\infty}\right]}dx_{-\infty}^{-1}dx_0^{\infty}\\
&= \int f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]\\
&\qquad \log_2\frac{f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]}{f\left[\left(x^{(1)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(1)}\right)_0^{\infty}\right]}dx_{-\infty}^{-1}dx_0^{\infty}\\
&+\int f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]\\
&\qquad \log_2\frac{f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]}{f\left[\left(x^{(2)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(2)}\right)_0^{\infty}\right]}dx_{-\infty}^{-1}dx_0^{\infty}.
\end{aligned}
$$

In the first term, the integration with respect to $\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}$ yields one, and analogously in the second term the integration with respect to the first variable set yields one. Ultimately, we obtain:

$$
\begin{aligned}
I\left[X_{-\infty}^{-1};X_0^{\infty}\right] &= \int f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]\log_2\frac{f\left[\left(x^{(1)}\right)_{-\infty}^{-1},\left(x^{(1)}\right)_0^{\infty}\right]}{f\left[\left(x^{(1)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(1)}\right)_0^{\infty}\right]}d\left(x^{(1)}\right)_{-\infty}^{-1}d\left(x^{(1)}\right)_0^{\infty}\\
&\quad+\int f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]\log_2\frac{f\left[\left(x^{(2)}\right)_{-\infty}^{-1},\left(x^{(2)}\right)_0^{\infty}\right]}{f\left[\left(x^{(2)}\right)_{-\infty}^{-1}\right]f\left[\left(x^{(2)}\right)_0^{\infty}\right]}d\left(x^{(2)}\right)_{-\infty}^{-1}d\left(x^{(2)}\right)_0^{\infty}\\
&= I\left[\left(X^{(1)}\right)_{-\infty}^{-1}\left(X^{(1)}\right)_0^{\infty}\right]+I\left[\left(X^{(2)}\right)_{-\infty}^{-1};\left(X^{(2)}\right)_0^{\infty}\right].
\end{aligned}
$$

4.1.1 Closed-Form Solutions in Original State Space

To calculate the EMC on the basis of the generalized solution from Eq. 239 in the coordinates of the original state space $\mathbb{R}^p$, we must find the *pdf* of the generated stochastic process in the steady state. Let the p-dimensional random vector $X_{-\infty-\tau+1}$ be normally distributed with location $\mu_{-\infty-\tau+1} = A_0 \cdot x_{-\infty-\tau}$ and covariance

$\Sigma_{-\infty-\tau+1} = \Sigma_1$ (Eqs. 19 and 20), that is $X_{-\infty-\tau+1} \sim \mathcal{N}(x; A_0 \cdot x_{-\infty-\tau}, \Sigma_1)$. Starting with this random vector the project evolves according to state Eq. 8. As already shown in Section 2.2, the strictly stationary behavior for $t \to \infty$ means that a joint probability density is formed that is invariant under shifting the origin. Hence, for the locus we must have $\mu = A_0 \cdot \mu + E[\varepsilon_t] = A_0 \cdot \mu$, and for the covariance matrix the well-known Lyapunov criterion $\Sigma = A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}} + \mathrm{Var}[\varepsilon_t] = A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}} + C$ must be satisfied (Eqs. 4 and 27). It follows that μ must be an eigenvector corresponding to the eigenvalue 1 of the WTM A_0. Clearly, if the modeled project is asymptotically stable and the modulus of the largest eigenvalue of A_0 is less than 1, no such eigenvector can exist. Hence, the only vector that satisfies this equation is the zero vector 0_p, which indicates that there is no remaining work (Eq. 26).

Let $\lambda_1(A_0), \ldots, \lambda_p(A_0)$ be the eigenvalues of WTM A_0 ordered by magnitude. If $|\lambda_1(A_0)| < 1$, the solution of the Lyapunov Eq. 27 can be written as an infinite power series (Lancaster and Tismenetsky 1985):

$$\Sigma = \sum_{k=0}^{\infty} A_0^k \cdot C \cdot \left(A_0^{\mathrm{T}}\right)^k. \tag{245}$$

It can also be expressed using the Kronecker product $\otimes$:

$$vec[\Sigma] = \left[I_{p^2} - A_0 \otimes A_0\right]^{-1} vec[C] \, .$$

Σ is also positive-semidefinite. In the above equation it is assumed that $I_{p^2} - A_0 \otimes A_0$ is invertible, $vec[C\ ]$ is the vector function which was already used for the derivation of the least square estimators in Section 2.7, and I_{p^2} is the identity matrix of size $p^2 \times p^2$.

Under the assumption of Gaussian behavior, it is not difficult to find different closed-form solutions. Recalling that the random vector X_0 in steady state is normally distributed with location $\mu = 0_p$ and covariance Σ, it follows from textbooks (e.g. Cover and Thomas 1991) that the differential entropy as the second summand in Eq. 239 can be expressed as

$$\begin{aligned} -\int_{\mathbb{X}^p} f[x_0] \log_2 f[x_0] dx_0 &= -\int_{\mathbb{R}^p} \mathcal{N}(x_0; \mu, \Sigma) \log_2 \mathcal{N}(x_0; \mu, \Sigma) dx_0 \\ &= \frac{1}{2} \log_2 (2\pi e)^p \mathrm{Det}[\Sigma]. \end{aligned}$$

For the calculation of the conditional entropy (first summand in Eq. 239), the following insight is helpful. Given a value x_0, the distribution of X_1 is a normal distribution with location $A_0 \cdot x_0$ and covariance C. Hence, the conditional entropy is simply equal to minus the differential entropy of that distribution. For Gaussian distributions, the differential entropy is independent of the locus. Therefore, for the conditional entropy it holds that

$$\begin{aligned}
&\int_{\mathbb{X}^p}\int_{\mathbb{X}^p} f\left[x_1\middle|x_0\right] f[x_0]\log_2 f\left[x_1\middle|x_0\right] dx_0 dx_1\\
&= \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \mathcal{N}(x_1;A_0x_0,C)\mathcal{N}(x_0;\mu,\Sigma)\log_2\mathcal{N}(x_1;A_0x_0,C)dx_0dx_1\\
&= \int_{\mathbb{R}^p} \mathcal{N}(x_1;A_0x_0,C)\ \log_2\mathcal{N}(x_1;A_0x_0,C)dx_1\\
&= -\frac{1}{2}\log_2(2\pi e)^p\mathrm{Det}[C].
\end{aligned}$$

It follows for the VAR(1) model of cooperative work that

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}[\Sigma]}{\mathrm{Det}[C]}\right) = \frac{1}{2}\log_2\mathrm{Det}[\Sigma] - \frac{1}{2}\log_2\mathrm{Det}[C]\\
&= \frac{1}{2}\log_2\mathrm{Det}\left[\Sigma\cdot C^{-1}\right].
\end{aligned} \tag{246}$$

According to the above equation, the EMC can be decomposed additively into dynamic and pure-fluctuation parts. The dynamic part represents the variance of the process in steady state. If the fluctuations are isotropic, the dynamic part completely decouples from the fluctuations, as will be shown in Eqs. 250 and 251 (Ay et al. 2012). If the solution of the Lyapunov equation (Eq. 245) is substituted into the above equation, we can write the desired first closed-form solution as

$$\mathrm{EMC} = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\sum_{k=0}^{\infty} A_0^k \cdot C \cdot \left(A_0^{\mathrm{T}}\right)^k\right]}{\mathrm{Det}[C]}\right). \tag{247}$$

The determinant Det[Σ] of the covariance matrix $\Sigma = \sum_{k=0}^{\infty} A_0^k \cdot C \cdot \left(A_0^{\mathrm{T}}\right)^k$ in the numerator of the solution above can be interpreted as a generalized variance of the stationary process. In the same manner Det[C] represents the generalized variance of the inherent fluctuations. The inverse C^{-1} is the so-called "concentration matrix" or "precision matrix" (Puri 2010). Det[C] can also be interpreted as the intrinsic (mean squared) one-step prediction error that cannot be underrun, even if we condition the observation on infinite past histories to build a maximally predictive model. An analogous interpretation of Det[Σ] is to consider it as the (mean squared) error for an infinite-step forecast of the VAR(1) model that is parameterized by the optimizing parameters x_0, A_0 and C (Lütkepohl 1985). In this sense, EMC is the logarithmic ratio related to the mean squared errors for infinite-step and one-step forecasts of the process state. Another interesting interpretation of Det[Σ] is produced if we do not refer to the predictions of a parameterized VAR(1) model over an infinite forecast horizon but instead to the one-step prediction error of a naïve

VAR(0) model whose predictions are based on the (zero) mean of the stationary process. It is evident that this kind of model completely lacks the ability to compress past project trajectories into a meaningful internal configuration to denote the state of the project and therefore has zero complexity. Hence, EMC can also be interpreted as the logarithmic ratio of the (mean squared) one-step prediction error of a naïve VAR(0) model with zero complexity and a standard VAR(1) model with non-negligible complexity due to procedural memory that incorporates an effective prediction mechanism. In this context "effective" means that the state should be formed in a way that the mean squared prediction error is minimized at fixed memory (sensu Still 2014). In terms of information theory, the generalized variance ratio can be interpreted as the entropy lost and information gained when the modeled project is in the steady state, and the state is observed by the project manager with predefined "error bars", which cannot be underrun because of the intrinsic prediction error (Bialek 2003).

The covariance matrices Σ and C are positive-semidefinite. Under the assumption that they are of full rank, the determinants are positive, and the range of the EMC is $[0, +\infty)$. This was already mentioned in the discussion of the essential properties of EMC (see Eq. 241).

Interestingly, we can reshape the above solution so that it can be interpreted in terms of Shannon's famous "Gaussian channel" (cf. Eq. 262 and the associated discussion) as

$$\mathrm{EMC} = \frac{1}{2}\log_2 \mathrm{Det}\left[I_p + \left(\sum_{k=1}^{\infty} A_0^k \cdot C \cdot \left(A_0^{\mathrm{T}}\right)^k\right) \cdot C^{-1}\right]. \tag{248}$$

If the covariance C is decomposed into an orthogonal forcing matrix K and a diagonal matrix Λ_K as shown in Eq. 22, the determinant in the denominator of Eq. 247 can be replaced by $\mathrm{Det}[C] = \mathrm{Det}[\Lambda_K]$.

We can also separate the noise component $K \cdot \Lambda_K \cdot K^{\mathrm{T}}$ in the sum and reshape the determinant in the numerator as follows:

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\sum_{k=0}^{\infty} A_0^k \cdot K \cdot \Lambda_K \cdot K^{\mathrm{T}} \cdot \left(A_0^{\mathrm{T}}\right)^k\right]}{\mathrm{Det}[\Lambda_K]}\right)\\
&= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\sum_{k=1}^{\infty} A_0^k \cdot K \cdot \Lambda_K \cdot K^{\mathrm{T}} \cdot \left(A_0^{\mathrm{T}}\right)^k + K \cdot \Lambda_K \cdot K^{\mathrm{T}}\right]}{\mathrm{Det}[\Lambda_K]}\right)\\
&= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}[K] \cdot \mathrm{Det}\left[K^{\mathrm{T}} \cdot \left(\sum_{k=1}^{\infty} A_0^k \cdot K \cdot \Lambda_K \cdot K^{\mathrm{T}} \cdot \left(A_0^{\mathrm{T}}\right)^k\right) \cdot K + \Lambda_K\right] \cdot \mathrm{Det}\left[K^{\mathrm{T}}\right]}{\mathrm{Det}[\Lambda_K]}\right)\\
&= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[K^{\mathrm{T}} \cdot \left(\sum_{k=1}^{\infty} A_0^k \cdot K \cdot \Lambda_K \cdot K^{\mathrm{T}} \cdot \left(A_0^{\mathrm{T}}\right)^k\right) \cdot K + \Lambda_K\right]}{\mathrm{Det}[\Lambda_K]}\right).
\end{aligned}$$

Moreover, because Λ_K is diagonal, taking $\mathrm{Tr}[\log_2(\Lambda_K)]$ is equivalent to $\log_2(\mathrm{Det}[\Lambda_K])$ and we have

$$\mathrm{EMC} = \frac{1}{2}\log_2 \mathrm{Det}\left[A_0' + \Lambda_K\right] - \frac{1}{2}\sum_{i=1}^{p}\log_2 \lambda_i(C), \tag{249}$$

where $A_0' = K^{\mathrm{T}} \cdot \left(\sum_{k=1}^{\infty} A_0^k \cdot K \cdot \Lambda_K \cdot K^{\mathrm{T}} \cdot \left(A_0^{\mathrm{T}}\right)^k\right) \cdot K$.

If the noise is isotropic, that is, the variances along the independent directions are equal $\left(C = \left\{\sigma^2\right\} \cdot I_p\right)$, and therefore correlations ρ_{ij} (Eq. 43) between performance fluctuations do not exist, we obtain a surprisingly simple solution:

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2 \mathrm{Det}\left[\sum_{k=0}^{\infty} A_0^k \cdot \left(A_0^{\mathrm{T}}\right)^k\right] \\
&= \frac{1}{2}\log_2 \mathrm{Det}\left[\left(I_p - A_0 \cdot A_0^{\mathrm{T}}\right)^{-1}\right] \\
&= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_p - A_0 \cdot A_0^{\mathrm{T}}\right].
\end{aligned} \tag{250}$$

The above solution is based on the von Neumann series that generalizes the geometric series to matrices (cf. Section 2.2).

If the matrix A_0 is diagonalizable, it can be decomposed into eigenvectors $\vartheta_i(A_0)$ in the columns $S_{:i}$ of S (Eq. 35) and written as $A_0 = S \cdot \Lambda_S \cdot S^{-1}$. Λ_S is a diagonal matrix with eigenvalues $\lambda_i(A_0)$ along the principal diagonal. Hence, if $C = \left\{\sigma^2\right\} \cdot I_p$ and A_0 is diagonalizable, the EMC from Eq. 250 can be fully simplified:

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2 \prod_{i=1}^{p} \frac{1}{1 - \lambda_i(A_0)^2} \\
&= \frac{1}{2}\sum_{i=1}^{p} \log_2 \frac{1}{1 - \lambda_i(A_0)^2} \\
&= -\frac{1}{2}\sum_{i=1}^{p} \log_2 \left(1 - \lambda_i(A_0)^2\right).
\end{aligned} \tag{251}$$

Both closed-form solutions that were obtained under the assumption of isotropic fluctuations only depend on the dynamical operator A_0, and therefore the dynamic part of the project can be seen to decouple completely from the unpredictable performance fluctuations. Under these circumstances the argument $\left(1 - \lambda_i(A_0)^2\right)$ of the binary logarithmic function can be interpreted as the damping coefficient of design mode $\phi_i = (\lambda_i(A_0), \vartheta_i(A_0))$ (see Section 2.1).

Similarly, for a project phase in which all p development tasks are processed independently at the same autonomous processing rate a, the dynamic part completely decouples from the performance fluctuations under arbitrary correlation coefficients. In this non-cooperative environment with minimum richness of temporal and structure-organizational dependencies, we simply have $A_0 = \mathrm{Diag}[a, \ldots, a]$. For EMC, it therefore holds that

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\sum_{k=0}^{\infty}\left(\mathrm{Diag}[a, \ldots, a]\right)^k \cdot C \cdot \left(\mathrm{Diag}[a, \ldots, a]^{\mathrm{T}}\right)^k\right]}{\mathrm{Det}[C]}\right) \\
&= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[C \cdot \sum_{k=0}^{\infty}\left(\mathrm{Diag}[a, \ldots, a]\right)^k \cdot \left(\mathrm{Diag}[a, \ldots, a]\right)^k\right]}{\mathrm{Det}[C]}\right) \\
&= \frac{1}{2}\log_2\mathrm{Det}\left[\sum_{k=0}^{\infty}\left(\mathrm{Diag}\left[a^2, \ldots, a^2\right]\right)^k\right] \\
&= \frac{1}{2}\log_2\mathrm{Det}\left[\mathrm{Diag}\left[\frac{1}{1-a^2}, \ldots, \frac{1}{1-a^2}\right]\right] \\
&= -\frac{p}{2}\log_2\left(1-a^2\right).
\end{aligned} \tag{252}$$

An additional closed-form solution in which the EMC can be expressed in terms of the dynamical operator A_0 and a so-called prewhitened operator W was formulated by DelSole and Tippett (2007) and Ay et al. (2012). Using $\mathrm{Det}[A]/\mathrm{Det}[B] = \mathrm{Det}\left[A \cdot B^{-1}\right]$ and the Lyapunov Eq. 27 we can write

$$\frac{\mathrm{Det}[C]}{\mathrm{Det}[\Sigma]} = \mathrm{Det}\left[\left(\Sigma - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}}\right) \cdot \Sigma^{-1}\right] = \mathrm{Det}\left[I_p - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}} \cdot \Sigma^{-1}\right].$$

Defining

$$W := \Sigma^{-\frac{1}{2}} \cdot A_0 \cdot \Sigma^{\frac{1}{2}}$$

we obtain

$$\frac{\mathrm{Det}[C]}{\mathrm{Det}[\Sigma]} = \mathrm{Det}\left[I_p - W \cdot W^{\mathrm{T}}\right],$$

where $\mathrm{Det}\left[I_p - A \cdot N \cdot A^{-1}\right] = \mathrm{Det}\left[I_p - N\right]$ and $\Sigma = \Sigma^{\mathrm{T}}$ were used. Hence, we obtain the EMC also as

$$\mathrm{EMC} = -\frac{1}{2}\log_2 \mathrm{Det}\left[I_p - W \cdot W^{\mathrm{T}}\right]. \tag{253}$$

According to DelSole and Tippett (2007) the application of the dynamical operator W can be regarded as a whitening transformation of the state-space coordinates of the dynamical operator A_0 by means of the covariance matrix Σ.

Concerning the evaluation of the persistent mutual information—represented by the variable EMC(τ)—of a vector autoregressive process, Section 3.2.4 showed that this can be expressed by the continuous-type mutual information I[.;.] as

$$\begin{aligned}\mathrm{EMC}(\tau) &= I\left[X_{-\infty}^{-1}; X_\tau^\infty\right]\\ &= \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_\tau^\infty\right] \log_2 \frac{f\left[x_{-\infty}^{-1}, x_\tau^\infty\right]}{f\left[x_{-\infty}^{-1}\right] f\left[x_\tau^\infty\right]}\, dx_{-\infty}^{-1}\, dx_\tau^\infty.\end{aligned}$$

The independent parameter $\tau \geq 0$ denotes the lead time. The term $f\left[x_{-\infty}^{-1}\right]$ designates the joint *pdf* of the infinite one-dimensional history of the stochastic process. Likewise, $f\left[x_\tau^\infty\right]$ designates the corresponding *pdf* of the infinite future from time τ onward. We used the shorthand notation $f\left[x_{-\infty}^{-1}, x_\tau^\infty\right] = f[x_{-\infty}, x_{-\infty+1}, \ldots, x_{-1}, x_\tau, x_{\tau+1}, , \ldots, x_{\infty-1}, x_\infty]$, $f\left[x_{-\infty}^{-1}\right] = f[x_{-\infty}, \ldots, x_{-1}]$, $f\left[x_\tau^\infty\right] = f[x_\tau, \ldots, x_\infty]$, $dx_{-\infty}^{-1} = dx_{-\infty} \ldots dx_{-1}$ and $dx_\tau^\infty = dx_\tau \ldots dx_\infty$. Informally, for positive lead times the term $I\left[X_{-\infty}^{-1}; X_\tau^\infty\right]$ can be interpreted as the information that is communicated from the past to the future ignoring the current length-τ sequence of observations $X_0^{\tau-1}$. Assuming strict stationarity, the joint *pdfs* are invariant under shifting the origin. Due to the Markov property of the VAR (1) model they can be factorized as follows:

$$\begin{aligned}f\left[x_{-\infty}^{-1}, x_\tau^\infty\right] &= \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_0^\infty\right] dx_0 \ldots dx_{\tau-1}\\ &= f[x_{-\infty}] f\left[x_{-\infty+1}\middle|x_{-\infty}\right] \ldots f\left[x_{-1}\middle|x_{-2}\right] f\left[x_{\tau+1}\middle|x_\tau\right] \ldots f\left[x_\infty\middle|x_{\infty-1}\right]\\ &\quad \times \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_0\middle|x_{-1}\right] \ldots f\left[x_\tau\middle|x_{\tau-1}\right] dx_0 \ldots dx_{\tau-1}\\ f\left[x_{-\infty}^{-1}\right] &= f[x_{-\infty}] f\left[x_{-\infty+1}\middle|x_{-\infty}\right] \ldots f\left[x_{-1}\middle|x_{-2}\right]\\ f\left[x_\tau^\infty\right] &= f[x_\tau] f\left[x_{\tau+1}\middle|x_\tau\right] \ldots f\left[x_\infty\middle|x_{\infty-1}\right]\end{aligned}.$$

Hence, we can simplify the mutual information as follows:

$$\begin{aligned}I\left[X_{-\infty}^{-1}; X_\tau^\infty\right] &= \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_\tau^\infty\right]\\ &\log_2 \frac{\int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_0\middle|x_{-1}\right] \ldots f\left[x_\tau\middle|x_{\tau-1}\right] dx_0 \ldots dx_{\tau-1}}{f[x_\tau]} dx_{-\infty}^{-1}\; dx_\tau^\infty.\end{aligned}$$

According to the famous Chapman-Kolmogorov equation (Papoulis and Pillai 2002) it holds that:

$$\int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_0 \middle| x_{-1}\right] \ldots f\left[x_\tau \middle| x_{\tau-1}\right] dx_0 \ldots dx_{\tau-1} = f\left[x_\tau \middle| x_{-1}\right]$$

Hence, we have

$$\begin{aligned} I\left[X_{-\infty}^{-1}; X_\tau^\infty\right] &= \int_{\mathbb{X}^p}\int_{\mathbb{X}^p} \log_2 f\left[x_\tau \middle| x_{-1}\right] dx_{-1} dx_\tau \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_\tau^\infty\right] dx_{-\infty}^{-2}\, dx_{\tau+1}^\infty \\ &\quad - \int_{\mathbb{X}^p} \log_2 f[x_\tau] dx_\tau \int_{\mathbb{X}^p} \cdots \int_{\mathbb{X}^p} f\left[x_{-\infty}^{-1}, x_\tau^\infty\right] dx_{-\infty}^{-1}\, dx_{\tau+1}^\infty \\ &= \int_{\mathbb{X}^p}\int_{\mathbb{X}^p} f[x_{-1}, x_\tau] \log_2 f\left[x_\tau \middle| x_{-1}\right] dx_{-1} dx_\tau - \int_{\mathbb{X}^p} f[x_\tau] \log_2 f[x_\tau]\, dx_\tau \\ &= \int_{\mathbb{X}^p} f[x_{-1}] dx_{-1} \int_{\mathbb{X}^p} f\left[x_\tau \middle| x_{-1}\right] \log_2 f\left[x_\tau \middle| x_{-1}\right] dx_\tau - \int_{\mathbb{X}^p} f[x_\tau] \log_2 f[x_\tau]\, dx_\tau. \end{aligned}$$

For a VAR(1) process the transition function is defined as

$$f\left[x_\tau \middle| x_{-1}\right] = \mathcal{N}\left(x_\tau; A_0^\tau \cdot x_{-1}, C(\tau)\right),$$

with the lead-time dependent covariance

$$\begin{aligned} C(\tau) &= A_0 \cdot C(\tau - 1) \cdot A_0^{\mathrm{T}} + C \\ &= \sum_{k=0}^{\tau} A_0^k \cdot C \cdot \left(A_0^{\mathrm{T}}\right)^k. \end{aligned}$$

We find the solution

$$\begin{aligned} \mathrm{EMC}(\tau) &= \frac{1}{2}\log_2 \mathrm{Det}[\Sigma] - \frac{1}{2}\log_2 \mathrm{Det}[C(\tau)] \\ &= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}[\Sigma]}{\mathrm{Det}[C(\tau)]}\right) \\ &= \frac{1}{2}\log_2 \mathrm{Det}\left[\Sigma \cdot (C(\tau))^{-1}\right]. \end{aligned}$$

The solution can also be expressed as the logarithm of the variance ratio (Ay et al. 2012):

$$\mathrm{EMC}(\tau) = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}[\Sigma]}{\mathrm{Det}\left[\Sigma - A_0^{\tau+1} \cdot \Sigma \cdot \left(A_0^{\mathrm{T}}\right)^{\tau+1}\right]}\right), \tag{254}$$

noting that $C = \Sigma - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}}$. As in Section 4.1 we can rewrite the above solution on the basis of the dynamical operator A_0 and lead-time dependent prewhitened operator $W(\tau)$ (Eq. 253; DelSole and Tippett 2007; Ay et al. 2012) as

$$\mathrm{EMC}(\tau) = -\frac{1}{2}\log_2 \mathrm{Det}\left[I_p - W(\tau) \cdot W(\tau)^{\mathrm{T}}\right],$$

with

$$W(\tau) = \Sigma^{-\frac{1}{2}} \cdot A_0^{\tau+1} \cdot \Sigma^{\frac{1}{2}}. \tag{255}$$

Following the same principles, a closed-form solution can be calculated for the elusive information $\sigma_\mu(\tau) = I\left[X_{-\infty}^{-1}; X_\tau^\infty \middle| X_o^{\tau-1}\right]$ from Eq. 231. As explained in Section 3.2.4, the elusive information is one of two essential pieces of the persistent mutual information and represents the Shannon information that is communicated from the past to the future by the stochastic process, but does not flow through the currently observed length-τ sequence $X_0^{\tau-1}$ (James et al. 2011). The key distinguishing feature of the persistent mutual information is that it is nonzero for $\tau \geq 1$ if a process necessarily has hidden states (Marzen and Crutchfield 2014). Conversely, due to the Markov property of the VAR(1) model, the elusive information completely vanishes for positive length τ.

This statement is easy to prove by using the definitions for the conditional mutual information from Eq. 214 and the conditional entropy from Eq. 213. Based on these definitions, the following relationship can be expressed:

$$\begin{aligned} I\left[X;Y\middle|Z\right] &= H\left[X\middle|Z\right] + H\left[Y\middle|Z\right] - H\left[X,Y\middle|Z\right] \\ &= H[X,Z] - H[Z] + H[Y,Z] - H[Z] - H[X,Y,Z] + H[Z] \\ &= H[X,Z] + H[Y,Z] - H[Z] - H[X,Y,Z]. \end{aligned}$$

As it holds

$$I[X;Z,Y] = H[X] + H[Z,Y] - H[X,Y,Z]$$

we find

$$\begin{aligned} I\left[X;Y\middle|Z\right] &= H[X,Z] - H[Z] - H[X] + I[X;Z,Y] \\ &= I[X;Z,Y] - I[X;Z]. \end{aligned}$$

In particular, we have

$$\begin{aligned} \sigma_\mu(\tau) &= I\left[X_{-\infty}^{-1}; X_\tau^\infty \middle| X_0^{\tau-1}\right] \\ &= I\left[X_{-\infty}^{-1}; X_0^{\tau-1}, X_\tau^\infty\right] - I\left[X_{-\infty}^{-1}; X_0^{\tau-1}\right] \\ &= I\left[X_{-\infty}^{-1}; X_0^\infty\right] - I\left[X_{-\infty}^{-1}; X_0^{\tau-1}\right]. \end{aligned}$$

Using the Markov property (Eq. 235), we see from the calculations Eq. 237–238 that the emergent complexity does not depend on the future of the autoregressive process beyond the lead time τ, i.e.

$$I\left[X_{-\infty}^{-1}; X_0^{\infty}\right] = I\left[X_{-\infty}^{-1}; X_0^{\tau-1}\right].$$

This proves that it holds for $\tau \geq 1$:

$$\sigma_\mu(\tau) = 0.$$

This result is independent of the coordinate system in the vector autoregression model of cooperative work.

4.1.2 Closed-Form Solutions in the Spectral Basis

In this chapter, we calculate additional solutions in which the dependence of the EMC on the anisotropy of the performance fluctuations is made explicit. These solutions are much easier to interpret, and to derive them we work in the spectral basis (cf. Eq. 35). According to Neumaier and Schneider (2001), the steady-state covariance matrix Σ' in the spectral basis can be calculated on the basis of the transformed covariance matrix of the performance fluctuations $C' = S^{-1} \cdot C \cdot \left(\left[S^{\mathrm{T}}\right]^*\right)^{-1}$ (Eq. 41) as

$$\Sigma' = \begin{pmatrix} \dfrac{{c'_{11}}^2}{1-\lambda_1\overline{\overline{\lambda_1}}} & \dfrac{\rho'_{12}c'_{11}c'_{22}}{1-\lambda_1\overline{\overline{\lambda_2}}} & \cdots \\ \dfrac{\rho'_{12}c'_{11}c\prime_{22}}{1-\lambda_2\overline{\overline{\lambda_1}}} & \dfrac{{c'_{22}}^2}{1-\lambda_2\overline{\overline{\lambda_2}}} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{256}$$

In the above equation, the ρ'_{ij}'s are the transformed correlations, which were defined in Eq. 43 for a WTM A_0 with arbitrary structure and in Eq. 47 for A_0's that are symmetric. The ${c'_{ii}}^2$'s (cf. Eq. 10) and $\rho'_{ij}c'_{ii}c'_{jj}$'s (cf. Eq. 11) are the scalar-valued variance and covariance components of C' in the spectral basis:

$$C' = \begin{pmatrix} {c'_{11}}^2 & \rho'_{12}c'_{11}c'_{22} & \cdots \\ \rho'_{12}c'_{11}c'_{22} & {c'_{22}}^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{257}$$

The transformation into the spectral basis is a linear transformation of the state-space coordinates (see Eq. 41) and therefore does not change the mutual information being communicated from the infinite past into the infinite future by the stochastic process. Hence, the functional form of the closed-form solution from Eq. 246 holds, and the EMC can be calculated as the (logarithmic) variance ratio (Schneider and Griffies 1999; de Cock 2002):

$$\mathrm{EMC} = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}[\Sigma']}{\mathrm{Det}[C']}\right) = \frac{1}{2}\log_2 \mathrm{Det}\left[\Sigma' \cdot C'^{-1}\right]. \tag{258}$$

The basis transformation does not change the positive-definiteness of the covariance matrices. Under the assumption that the matrices are of full rank, the determinants are positive. As already shown in Section 4.1.1., the determinant $\mathrm{Det}[\Sigma']$ of the covariance matrix Σ' can be interpreted as a generalized variance of the stationary process in the spectral basis, whereas $\mathrm{Det}[C']$ represents the generalized variance of the inherent performance fluctuations after the basis transformation. The variance ratio can also be interpreted in a geometrical framework (de Cock 2002). It is well known that the volume Vol[.] of the parallelepiped spanned by the rows or columns of a covariance matrix, e.g. Σ', is equal to the value of its determinant:

$$\mathrm{Vol}[\mathrm{parallelepiped}[\Sigma']] = \mathrm{Det}[\Sigma'].$$

In this sense the inverse variance ratio $\mathrm{Det}[C']/\mathrm{Det}[\Sigma']$ represents the factor by which the volume of the parallelepiped referring to the dynamical part of the process can be collapsed due to the state observation by the project manager leading to a certain information gain.

An important finding is that the scalar-valued variance and covariance components of the fluctuation part are not relevant for the calculation of the EMC. This follows from the definition of a determinant (see Eq. 267). The calculated determinants of Σ' and C' just give rise to the occurrence of the factor $\prod_{n=1}^{p} {c'_{nn}}^2$, which cancels out:

$$\mathrm{Det}\left[\Sigma' \cdot C'^{-1}\right] = \mathrm{Det}[\Sigma'] \cdot \mathrm{Det}\left[C'^{-1}\right] = \frac{\mathrm{Det}[\Sigma']}{\mathrm{Det}[C']} = \frac{\mathrm{Det}\left[\Sigma'_N\right]}{\mathrm{Det}\left[C'_N\right]}.$$

Hence, we can also calculate with the "normalized" covariance matrices Σ'_N and C'_N:

$$\Sigma'_N = \begin{pmatrix} \frac{1}{1-|\lambda_1|^2} & \frac{\rho'_{12}}{1-\lambda_1\overline{\lambda_2}} & \cdots \\ \frac{\rho'_{12}}{1-\lambda_2\overline{\lambda_1}} & \frac{1}{1-|\lambda_2|^2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{259}$$

$$C'_N = \begin{pmatrix} 1 & \rho'_{12} & \cdots \\ \rho'_{12} & 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{260}$$

It can be proved that the normalized covariance matrices are also positive-semidefinite. If they are furthermore not rank deficient, inconsistencies of the complexity measure do not occur. According to Shannon's classic information-theory findings about the capacity of a Gaussian channel (Cover and Thomas 1991), the normalized covariance matrix Σ'_N can be decomposed into summands as follows:

$$\Sigma'_N = C'_N + \begin{pmatrix} \dfrac{1}{1-|\lambda_1|^2} - 1 & \dfrac{\rho'_{12}}{1-\lambda_1\overline{\lambda_2}} - \rho'_{12} & \cdots \\ \dfrac{\rho'_{12}}{1-\lambda_2\overline{\lambda_1}} - \rho'_{12} & \dfrac{1}{1-|\lambda_2|^2} - 1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

The second summand in the above equation is defined as Σ''_N. This matrix can be simplified:

$$\Sigma''_N = \begin{pmatrix} \dfrac{|\lambda_1|^2}{1-|\lambda_1|^2} & \rho'_{12}\dfrac{\lambda_1\overline{\lambda_2}}{1-\lambda_1\overline{\lambda_2}} & \cdots \\ \rho'_{12}\dfrac{\lambda_2\overline{\lambda_1}}{1-\lambda_2\overline{\lambda_1}} & \dfrac{|\lambda_2|^2}{1-|\lambda_2|^2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}. \tag{261}$$

We obtain the most expressive closed-form solution based on the signal-to-noise ratio $\text{SNR} := \Sigma''_N . {C'_N}^{-1}$:

$$\text{EMC} = \frac{1}{2}\log_2 \text{Det}\left[I_p + \Sigma''_N \cdot {C'_N}^{-1}\right]. \tag{262}$$

The SNR can be interpreted as the ratio of the variance Σ''_N of the signal in the spectral basis that is generated by cooperative task processing and the effective variance C'_N of the performance fluctuations. The variance of the signal drives the process to a certain extent and can be reinforced through the structural organization of the project. The effective fluctuations are in the same units as the input x_t. This is called "referring the noise to the input" and is a standard method in physics for characterizing detectors, amplifiers and other devices (Bialek 2003). Clearly, if one builds a photodetector it is not so useful to quote the noise level at the output in volts; one wants to know how this noise limits the ability to detect dim lights. Similarly, when we characterize a PD project that uses a stream of progress reports to document a quasicontinuous workflow, we don't want to know the variance in the absolute labor units; we want to know how variability in the performance of the developers limits precision in estimating the real work progress (signal), which amounts to defining an effective "noise level" in the units of the signal itself. In the present case, this is just a matter of "dividing" generalized variances, but in reality it is a fairly complex task. According to Sylvester's determinant theorem, we can swap the factors in the second summand:

$$\mathrm{Det}\left[I_p + \Sigma_N'' \cdot {C_N'}^{-1}\right] = \mathrm{Det}\left[I_p + {C_N'}^{-1} \cdot \Sigma_N''\right].$$

The obtained closed-form solution in the spectral basis has at most only $(p^2 - p)/2 + p = p(p+1)/2$ independent parameters, namely the eigenvalues $\lambda_i(A_0)$ of the WTM and the correlations ρ'_{ij} in the spectral basis, and not a maximum of the approximately $p^2 + (p^2 - p)/2 + p = p(3p+1)/2$ parameters encoded in both the WTM A_0 and the covariance matrix C (Eq. 248). In other words, through a transformation into the spectral basis we can identify the essential variables influencing emergent complexity in the sense of Grassberger's theory and reduce the dimensionality of the problem in many cases by the factor $(3p+1)/(p+1)$.

Furthermore, these independent parameters are easy to interpret, and at this point we can make a number of comments to stress the importance and usefulness of the analytical results. It is evident that the eigenvalues $\lambda_i(A_0)$ represent the essential temporal dependencies of the modeled project phase in terms of effective productivity rates on linearly independent scales determined by the eigenvectors $\vartheta_i(A_0)$ $(i = 1 \ldots p)$. The effective productivity rates depend only on the design modes ϕ_i of the WTM A_0 and therefore reflect the project's organizational design. The lower the effective productivity rates because of slow task processing or strong task couplings, the less the design modes are "damped," and hence the larger the project complexity. On the other hand, the correlations ρ'_{ij} model the essential dependencies between the unpredictable performance fluctuations in open organizational systems that can give rise to an excitation of the design modes and their interactions. This excitation can compensate for the damping factors of the design mode. The ρ'_{ij}'s scale linearly with the $\lambda_i(C)$ along each independent direction of the fluctuation variable ε'_t: the larger the $\lambda_i(C)$, the larger the correlations and the stronger the excitation (Eq. 43). However, the scale factors are determined not only by a linear interference between design modes ϕ_i and ϕ_j caused by cooperative task processing but also by the weighted interference with performance fluctuation modes Ψ_i and Ψ_j caused by correlations between performance variability (cf. Eqs. 43 and 47). In other words, the emergent complexity of the modeled project phase does not simply come from the least-damped design mode $\phi_1 = (\lambda_1(A_0), \vartheta_1(A_0))$ because this mode may not be sufficiently excited, but rather is caused (at least theoretically) by a complete interference between all design and performance fluctuation modes. Like the analytical considerations of Crutchfield et al. (2013) concerning stationary and ergodic stochastic processes whose measurement values cover a finite alphabet, the obtained closed-form solutions show that in a development process complexity is not just controlled by the "first spectral gap," i.e. the difference between the dominant eigenvalue and the eigenvalue with the second largest magnitude. Rather, the entire spectrum of eigenvalues is relevant and therefore all subspaces of the underlying causal-state process can contribute to emergent complexity (Crutchfield et al. 2013). In most practical case studies, only a few subspaces will dominate project dynamics. However, the closed-form solution from Eq. 262 in conjunction with Eqs. 260 and 261 shows that this is not generally the case. In Section 4.1.4, we will present fairly simple polynomial-based solutions for projects with only two or three tasks, and we will make the theoretical connections between the eigenvalues, the spectral gaps and the correlations very clear. The solution for two tasks will also allow us to identify simple scaling laws for real-valued eigenvalues. As a result, we see that emergent complexity in the sense of Grassberger's theory is a holistic property of the structure and process organization, and that, in most real cases, it cannot be reduced to singular properties of the project organizational design. This is a truly nonreductionist approach to complexity assessment insisting on the specific character of the organizational design as a whole.

Similarly to the previous chapter, we can obtain a closed-form solution for the persistent mutual information $\text{EMC}(\tau)$ in the spectral basis. The transformation into the spectral basis is a linear transformation of the state-space coordinates and therefore does not change the persistent mutual information communicated from the past into the future by the stochastic process. Hence, in analogy to Eq. 256 the variance ratio can also be calculated

$$\text{EMC}(\tau) = \frac{1}{2}\log_2\left(\frac{\text{Det}[\Sigma']}{\text{Det}\left[\Sigma' - \Lambda_S^{\tau+1} \cdot \Sigma' \cdot \left(\left[\Lambda_S^{\text{T}}\right]^*\right)^{\tau+1}\right]}\right)$$

in the spectral basis, where the diagonal matrix Λ_S is the dynamical operator (Eq. 39) as

$$\Lambda_S = \text{Diag}[\lambda_i(A_0)] \quad 1 \le i \le p\,.$$

Because Λ_S is diagonal, the solution in the spectral basis can be simplified to

$$\begin{aligned}
\mathrm{EMC}(\tau) &= -\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\left(\left[\Lambda_S^{\mathrm{T}}\right]^*\right)^{\tau+1}\right]}{\mathrm{Det}[\Sigma']}\right)\\
&= -\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}\right]}{\mathrm{Det}[\Sigma']}\right)\\
&= -\frac{1}{2}\log_2\mathrm{Det}\left[I_p - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}\cdot\Sigma'^{-1}\right]\\
&= -\frac{1}{2}\log_2\mathrm{Det}\left[I_p - \Sigma'(\tau)\cdot\Sigma'^{-1}\right],
\end{aligned}\tag{263}$$

with $\Sigma'(\tau) = \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}\ (\tau\geq 0)$.

As with the derivation of the expressive closed-form solution in Section 4.2, the generalized variance term $\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\left(\left[\Lambda_S^{\mathrm{T}}\right]^*\right)^{\tau+1} = \Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}$ in the denominator of the variance ratio can be written in an explicit matrix form:

$$\begin{aligned}
&\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}\\
&= \begin{pmatrix}
\frac{{c'_{11}}^2}{1-\lambda_1\overline{\lambda_1}} - \frac{{\lambda_1}^{\tau+1}{c'_{11}}^2\left(\overline{\lambda_1}\right)^{\tau+1}}{1-\lambda_1\overline{\lambda_1}} & \frac{\rho'_{12}c'_{11}c'_{22}}{1-\lambda_1\overline{\lambda_2}} - \frac{{\lambda_1}^{\tau+1}\rho'_{12}c'_{11}c'_{22}\left(\overline{\lambda_2}\right)^{\tau+1}}{1-\lambda_1\overline{\lambda_2}} & \cdots\\
\frac{\rho'_{12}c'_{11}c'_{22}}{1-\lambda_2\overline{\lambda_1}} - \frac{{\lambda_2}^{\tau+1}\rho'_{12}c'_{11}c'_{22}\left(\overline{\lambda_1}\right)^{\tau+1}}{1-\lambda_2\overline{\lambda_1}} & \frac{{c'_{22}}^2}{1-\lambda_2\overline{\lambda_2}} - \frac{{\lambda_2}^{\tau+1}{c'_{22}}^2\left(\overline{\lambda_2}\right)^{\tau+1}}{1-\lambda_2\overline{\lambda_1}} & \cdots\\
\vdots & \vdots & \ddots
\end{pmatrix}\\
&= \begin{pmatrix}
\left({c'_{11}}^2\right)\frac{1-|\lambda_1|^{2(\tau+1)}}{1-|\lambda_1|^2} & c'_{11}c'_{22}\frac{\rho'_{12}\left(1-{\lambda_1}^{\tau+1}\left(\overline{\lambda_2}\right)^{\tau+1}\right)}{1-\lambda_1\overline{\lambda_2}} & \cdots\\
c'_{11}c'_{22}\frac{\rho'_{12}\left(1-{\lambda_2}^{\tau+1}\left(\overline{\lambda_1}\right)^{\tau+1}\right)}{1-\lambda_2\overline{\lambda_1}} & \left({c'_{22}}^2\right)\frac{1-|\lambda_1|^{2(\tau+1)}}{1-|\lambda_1|^2} & \cdots\\
\vdots & \vdots & \ddots
\end{pmatrix}.
\end{aligned}$$

It can be proved that the covariances c'_{ij} in the above matrix form are not relevant for the calculation of $\mathrm{EMC}(\tau)$. This follows from the definition of a determinant (see Eq. 267). When calculating the determinants of Σ' and $\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}$ they just give rise to the occurrence of a factor $\prod_{n=1}^{p}{c'_{nn}}^2$, which cancels out in the variance ratio. Therefore, the persistent mutual information can also be calculated using normalized covariance matrices. The normalized covariance matrix of Σ', termed Σ'_N, was defined in Eq. 259. The normalized covariance matrix of $\Sigma' - \Lambda_S^{\tau+1}\cdot\Sigma'\cdot\Lambda_S^{*\tau+1}$ is simply

$$\Sigma'_N - \Lambda_S^{\tau+1} \cdot \Sigma'_N \cdot \Lambda_S^{*\tau+1}$$
$$= \begin{pmatrix} \frac{1-|\lambda_1|^{2(\tau+1)}}{1-|\lambda_1|^2} & \rho'_{12}\frac{\left(1-\lambda_1{}^{\tau+1}\left(\overline{\lambda_2}\right)^{\tau+1}\right)}{1-\lambda_1\overline{\lambda_2}} & \cdots \\ \rho'_{12}\frac{\left(1-\lambda_2{}^{\tau+1}\left(\overline{\lambda_1}\right)^{\tau+1}\right)}{1-\lambda_2\overline{\lambda_1}} & \frac{1-|\lambda_2|^{2(\tau+1)}}{1-|\lambda_2|^2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Hence,

$$\begin{aligned} \mathrm{EMC}(\tau) &= -\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\Sigma'_N - \Lambda_S^{\tau+1} \cdot \Sigma'_N \cdot \Lambda_S^{*\tau+1}\right]}{\mathrm{Det}\left[\Sigma'_N\right]}\right) \\ &= -\frac{1}{2}\log_2\mathrm{Det}\left[I_p - \Lambda_S^{\tau+1} \cdot \Sigma'_N \cdot \Lambda_S^{*\tau+1} \cdot \Sigma_N'^{-1}\right] \\ &= -\frac{1}{2}\log_2\mathrm{Det}\left[I_p - \Sigma'_N(\tau) \cdot \Sigma_N'^{-1}\right], \end{aligned} \qquad (264)$$

with $\Sigma'_N(\tau) = \Lambda_S^{\tau+1} \cdot \Sigma'_N \cdot \Lambda_S^{*\tau+1}$ $(\tau \geq 0)$.

4.1.3 Closed-form Solution through Canonical Correlation Analysis

If the matrix C'_N representing the intrinsic prediction error in the spectral basis is diagonal in the same coordinate system as the normalized covariance matrix Σ'_N contributed by cooperative task processing, then the matrix product $\Sigma'_N \cdot C_N'^{-1} = \left(I_p + \Sigma''_N \cdot C_N'^{-1}\right)$ is diagonal, and simple reduction of emergent complexity to singular properties of the design modes $\phi_i = (\lambda_i(A_0), \vartheta_i(A_0))$ and performance fluctuation modes $\Psi_i = (\lambda_i(C), k_i(C))$ will work. In this case, the elements along the principal diagonal are the signal-to-noise ratios along each independent direction. Hence, the EMC is proportional to the sum of the log-transformed ratios, and these summands are the only independent parameters. However, in the general case we have to diagonalize the above matrix product in a first step to obtain an additional closed-form solution. This closed-form solution has the least number of independent parameters. In spite of its algebraic simplicity, the solution is not very expressive, because the spatiotemporal covariance structures of the open organizational system are not revealed. We will return to this point after presenting the solution.

Unfortunately, the diagonalization of the matrix product $\Sigma'_N \cdot C_N'^{-1}$ cannot be carried out through an eigendecomposition, because the product of two symmetric matrices is not necessarily symmetric itself. Therefore, the left and right eigenvectors can differ and do not form a set of mutually orthogonal vectors, as they would if the product was diagonal. Nevertheless, we can always rotate our coordinate system in the space of the output to make the matrix product diagonal (Schneider and Griffies 1999). To do this, we decompose $\Sigma'_N \cdot {C'_N}^{-1}$ into singular values (singular value decomposition, see e.g. de Cock 2002) as

$$\Sigma'_N \cdot {C'_N}^{-1} = U \cdot \Lambda_{UV} \cdot V^{\mathrm{T}},$$

where

$$U \cdot U^{\mathrm{T}} = I_p \quad \text{and} \quad V \cdot V^{\mathrm{T}} = I_p$$

and

$$\Lambda_{UV} = \mathrm{Diag}\left[\sigma'_i\right] \quad 1 \le i \le p\,.$$

The columns of U are the left singular vectors; those of V are the right singular vectors. The columns of V can be regarded as a set of orthonormal "input" basis vectors for $\Sigma'_N \cdot {C'_N}^{-1}$; the columns of U form a set of orthonormal "output" basis vectors. The diagonal values σ'_i in matrix Λ_{UV} are the singular values, which can be thought of as scalar "gain controls" by which each corresponding input is multiplied to give a corresponding output. The σ'_i's are the only independent parameters of the following closed-form solution (see). The relationship between the singular values σ'_i of $\Sigma'_N \cdot {C'_N}^{-1}$ and the canonical correlations ρ_i (see summary of properties of EMC at the end of Section 4.1) in our case is as follows (de Cock 2002):

$$\sigma'_i = \frac{1}{1-\rho_i^2} \quad 1 \le i \le p\,.$$

Under the assumption that $\mathrm{Det}\left[\Sigma'_N \cdot {C'_N}^{-1}\right] > 0$, it is possible to prove that $\mathrm{Det}[U] \cdot \mathrm{Det}[V] = 1$. We can obtain the desired closed-form solution as follows:

$$\begin{aligned}
\mathrm{EMC} &= \frac{1}{2}\log_2 \det\left[\Sigma'_N \cdot {C'_N}^{-1}\right] \\
&= \frac{1}{2}\log_2 \det\left[U \cdot \Lambda_{UV} \cdot V^{\mathrm{T}}\right] \\
&= \frac{1}{2}\log_2 (\mathrm{Det}[U] \cdot \mathrm{Det}[\Lambda_{UV}] \cdot \mathrm{Det}[V]) \\
&= \frac{1}{2}\log_2 \det[\Lambda_{UV}] \\
&= \frac{1}{2}\mathrm{Tr}[\log_2(\Lambda_{UV})] \\
&= \frac{1}{2}\sum_{i=1}^{p}\log_2 \sigma'_i \\
&= \frac{1}{2}\sum_{i=1}^{p}\log_2\left(\frac{1}{1-\rho_i^2}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{p}\log_2\left(1-\rho_i^2\right).
\end{aligned} \tag{265}$$

In spite of its algebraic simplicity, a main disadvantage of this closed-form solution with only p parameters σ_i' or ρ_i^2 is that both the temporal dependencies of the modeled work process in terms of essential productivity rates (represented by the λ_i's), and the essential cooperative relationships exciting fluctuations (represented by the ρ_{ij}''s) are not explicit, but are compounded into correlation coefficients between the canonical variates. Therefore, it is impossible for the project manager to analyze and interpret the spatiotemporal covariance structures of the organizational system and to identify countermeasures for coping with emergent complexity.

A canonical correlation analysis over τ time steps leads to the following solution of the persistent mutual information:

$$\begin{aligned}
\mathrm{EMC}(\tau) &= \frac{1}{2}\log_2 \mathrm{Det}\left[\Sigma_N' \cdot \left(\Sigma_N' - \Lambda_S^{\tau+1} \cdot \Sigma_N' \cdot \Lambda_S^{*\tau+1}\right)^{-1}\right] \\
&= \frac{1}{2}\log_2 \mathrm{Det}\left[U(\tau)\cdot \Lambda_{UV}(\tau)\cdot V(\tau)^{\mathrm{T}}\right] \\
&= \frac{1}{2}\log_2 (\mathrm{Det}[U(\tau)]\cdot \mathrm{Det}[\Lambda_{UV}(\tau)]\cdot \mathrm{Det}[V(\tau)]) \\
&= \frac{1}{2}\log_2 \mathrm{Det}[\Lambda_{UV}(\tau)] \\
&= \frac{1}{2}\mathrm{Tr}[\log_2(\Lambda_{UV}(\tau))] \\
&= \frac{1}{2}\sum_{i=1}^{p}\log_2 \sigma_i'(\tau) \\
&= \frac{1}{2}\sum_{i=1}^{p}\log_2\left(\frac{1}{1-(\rho_i(\tau))^2}\right) \\
&= -\frac{1}{2}\sum_{i=1}^{p}\log_2\left(1-(\rho_i(\tau))^2\right).
\end{aligned} \tag{266}$$

The term $U(\tau)\cdot\Lambda_{UV}(\tau)\cdot V(\tau)^{\mathrm{T}}$ represents the product of the matrices resulting from a decomposition of $\Sigma_N' \cdot \left(\Sigma_N' - \Lambda_S^{\tau+1} \cdot \Sigma_N' \cdot \left(\Lambda_S{}^*\right)^{\tau+1}\right)^{-1}$ as a function of the lead time τ:

$$(U(\tau), \Lambda_{UV}(\tau), V(\tau)) = \mathrm{SVD}\left[\Sigma_N' \cdot \left(\Sigma_N' - \Lambda_S^{\tau+1} \cdot \Sigma_N' \cdot \left(\Lambda_S{}^*\right)^{\tau+1}\right)^{-1}\right],$$

where the matrix-valued function SVD[.] represents the singular value decomposition of the argument. The $\sigma_i'(\tau)$'s and $\rho_i(\tau)$'s represent, respectively, the singular values and canonical correlations given the lead time.

4.1.4 Polynomial-Based Solutions for Processes with Two and Three Tasks

We can also analyze the spatiotemporal covariance structure of Σ_N' (Eq. 259) in the spectral basis explicitly by recalling the definition of a determinant. If $B = \left(b_{ij}\right)$ is a matrix of size p, then

$$\mathrm{Det}(B) = \sum_{\beta \in R_p} \mathrm{sgn}(\beta) \prod_{i=1}^{p} b_{i,\beta(i)} \tag{267}$$

holds. R_p is the set of all permutations of $\{1,\ldots,p\}$. Thus, because of the regular structure of the matrix Σ'_N, $\mathrm{Det}[\Sigma'_N]$ is a sum of $p\,!$ summands. Each of these summands is a fraction, because it is a product of elements from Σ'_N, where exactly one entry is chosen from each row and column. The denominator of those fractions is a product consisting of p factors of $1-\lambda_i(\mathrm{A}_0)\lambda_i\overline{\lambda_{\mathrm{j}}(\mathrm{A}_0)}$. The numerator is a product of $2, 3, \ldots, p$ factors ρ'_{ij}, or simply 1 if the permutation is the identity. (The case of one factor cannot occur, because the amount of factors equals the amount of numbers changed by the permutation β, and there is no permutation that changes just one number). The coefficients (i,j) of the factor $1-\lambda_i(\mathrm{A}_0)\overline{\lambda_{\mathrm{j}}(\mathrm{A}_0)}$ in the denominator correspond to the coefficients (k,l) of the factor ρ'_{kl} in the numerator, i.e. $i=l$ and $j=k$, if $i \neq k$ holds. Otherwise, in the case that $i=k$, no corresponding factor is multiplied in the numerator, because the appropriate entry of Σ'_N lies on the principal diagonal. Moreover, $1-\lambda_i(A_0)\overline{\lambda_{\mathrm{j}}(A_0)} = 1-|\lambda_i(A_0)|^2$ holds in that case.

These circumstances are elucidated for project phases with only $p=2$ and $p=3$ fully interdependent tasks. For $p=2$ we have

$$\Sigma'_N = \begin{pmatrix} \frac{1}{1-|\lambda_1|^2} & \frac{\rho'_{12}}{1-\lambda_1\overline{\lambda_2}} \\ \frac{\rho'_{12}}{1-\lambda_2\overline{\lambda_1}} & \frac{1}{1-|\lambda_2|^2} \end{pmatrix},$$

hence,

$$\mathrm{Det}\left[\Sigma'_N\right] = \frac{1}{\left(1-|\lambda_1|^2\right)\left(1-|\lambda_2|^2\right)} - \frac{{\rho'_{12}}^2}{\left(1-\lambda_2\overline{\lambda_1}\right)\left(1-\lambda_1\overline{\lambda_2}\right)}.$$

For $p=3$ we have

$$\Sigma'_N = \begin{pmatrix} \frac{1}{1-|\lambda_1|^2} & \frac{\rho'_{12}}{1-\lambda_1\overline{\lambda_2}} & \frac{\rho'_{13}}{1-\lambda_1\overline{\lambda_3}} \\ \frac{\rho'_{12}}{1-\lambda_2\overline{\lambda_1}} & \frac{1}{1-|\lambda_2|^2} & \frac{\rho'_{23}}{1-\lambda_2\overline{\lambda_3}} \\ \frac{\rho'_{13}}{1-\lambda_3\overline{\lambda_1}} & \frac{\rho'_{23}}{1-\lambda_3\overline{\lambda_2}} & \frac{1}{1-|\lambda_3|^2} \end{pmatrix},$$

hence,

$$\begin{aligned}\mathrm{Det}\left[\Sigma'_N\right] =& \frac{1}{\left(1-|\lambda_1|^2\right)\left(1-|\lambda_2|^2\right)\left(1-|\lambda_3|^2\right)}\\ &-\frac{\rho'_{23}2}{\left(1-|\lambda_1|^2\right)\left(1-\lambda_3\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_3}\right)}-\frac{\rho'_{13}2}{\left(1-|\lambda_2|^2\right)\left(1-\lambda_3\overline{\lambda_1}\right)\left(1-\lambda_1\overline{\lambda_3}\right)}\\ &-\frac{\rho'_{12}2}{\left(1-|\lambda_3|^2\right)\left(1-\lambda_1\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_1}\right)}+\frac{\rho'_{12}\rho'_{13}\rho'_{23}}{\left(1-\lambda_1\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_3}\right)\left(1-\lambda_3\overline{\lambda_1}\right)}\\ &+\frac{\rho'_{12}\rho'_{13}\rho'_{23}}{\left(1-\lambda_2\overline{\lambda_1}\right)\left(1-\lambda_3\overline{\lambda_2}\right)\left(1-\lambda_1\overline{\lambda_3}\right)}.\end{aligned}$$

The results for C'_N are much simpler. From Eqs. 259 and 260 it follows that the numerator is the same, whereas the denominator is simply 1.

For $p = 2$ we have

$$C'_N = \begin{pmatrix} 1 & \rho'_{12} \\ \rho'_{12} & 1 \end{pmatrix},$$

hence,

$$\mathrm{Det}\left[C'_N\right] = 1 - {\rho'_{12}}^2.$$

For $p = 3$ we have

$$C'_N = \begin{pmatrix} 1 & \rho'_{12} & \rho'_{13} \\ \rho'_{12} & 1 & \rho'_{23} \\ \rho'_{13} & \rho'_{23} & 1 \end{pmatrix},$$

hence,

$$\mathrm{Det}\left[C'_N\right] = 1 + 2\rho'_{12}\rho'_{13}\rho'_{23} - {\rho'_{12}}^2 - {\rho'_{13}}^2 - {\rho'_{23}}^2.$$

These results readily yield the closed-form expression

$$\mathrm{EMC} = \frac{1}{2}\log_2\left[\frac{1}{1-{\rho'_{12}}^2}\left(\frac{1}{\left(1-|\lambda_1|^2\right)\left(1-|\lambda_2|^2\right)} - {\rho'_{12}}^2\frac{1}{\left(1-\lambda_2\overline{\lambda_1}\right)\left(1-\lambda_1\overline{\lambda_2}\right)}\right)\right] \tag{268}$$

for $p = 2$ tasks and

$$\mathrm{EMC} = \frac{1}{2}\log_2\left[\frac{1}{1 + 2\rho'_{12}\rho'_{13}\rho'_{23} - {\rho'_{12}}^2 - {\rho'_{13}}^2 - {\rho'_{23}}^2}\mathrm{Det}\left[\Sigma'_N\right]\right] \tag{269}$$

for $p = 3$ tasks, where the simplified determinant $\mathrm{Det}[\Sigma'_N]$ of the normalized covariance matrix Σ'_N is given by

$$\begin{aligned}\mathrm{Det}\left[\Sigma'_N\right] = &\frac{1}{\left(1-|\lambda_1|^2\right)}\left(\frac{1}{\left(1-|\lambda_2|^2\right)\left(1-|\lambda_3|^2\right)} - \frac{\rho'_{23}}{\left(1-\lambda_3\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_3}\right)}\right)\\ &+\rho'_{12}\rho'_{13}\rho'_{23}\left(\frac{1}{\left(1-\lambda_1\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_3}\right)\left(1-\lambda_3\overline{\lambda_1}\right)} + \frac{1}{\left(1-\lambda_1\overline{\lambda_3}\right)\left(1-\lambda_2\overline{\lambda_1}\right)\left(1-\lambda_3\overline{\lambda_2}\right)}\right)\\ &-\frac{{\rho'_{12}}^2}{\left(1-|\lambda_3|^2\right)\left(1-\lambda_1\overline{\lambda_2}\right)\left(1-\lambda_2\overline{\lambda_1}\right)} - \frac{{\rho'_{13}}^2}{\left(1-|\lambda_2|^2\right)\left(1-\lambda_3\overline{\lambda_1}\right)\left(1-\lambda_1\overline{\lambda_3}\right)}.\end{aligned} \tag{270}$$

Now, we suppose that all eigenvalues $\lambda_i(A_0)$ are real. Under this assumption EMC can be expressed by the spectral gaps $\left(\lambda_i - \lambda_j\right)_{i\neq j}$ between eigenvalues as

$$\begin{aligned}\mathrm{EMC} &= \frac{1}{2}\log_2\left[\frac{1}{\left(1-{\lambda_1}^2\right)\left(1-{\lambda_2}^2\right)} + \frac{{\rho'_{12}}^2}{1-{\rho'_{12}}^2}\,\frac{(\lambda_1-\lambda_2)^2}{\left(1-{\lambda_1}^2\right)\left(1-{\lambda_2}^2\right)(1-\lambda_1\lambda_2)^2}\right]\\ &= \frac{1}{2}\log_2\left[\frac{1}{\left(1-{\lambda_1}^2\right)\left(1-{\lambda_2}^2\right)}\left(1 + \frac{{\rho'_{12}}^2}{1-{\rho'_{12}}^2}\,\frac{(\lambda_1-\lambda_2)^2}{(1-\lambda_1\lambda_2)^2}\right)\right]\\ &= -\frac{1}{2}\log_2\left[1-{\lambda_1}^2\right] - \frac{1}{2}\log_2\left[1-{\lambda_2}^2\right] + \log_2\left[1 + \frac{{\rho'_{12}}^2}{1-{\rho'_{12}}^2}\,\frac{(\lambda_1-\lambda_2)^2}{(1-\lambda_1\lambda_2)^2}\right],\end{aligned} \tag{271}$$

for $p = 2$, and as

$$\begin{aligned}\mathrm{EMC} = &-\frac{1}{2}\log_2\left[1-{\lambda_1}^2\right] - \frac{1}{2}\log_2\left[1-{\lambda_2}^2\right] - \frac{1}{2}\log_2\left[1-{\lambda_3}^2\right]\\ &+\frac{1}{2}\log_2\left[1 + \frac{{\rho'_{12}}^2}{\rho'}\,\frac{(\lambda_1-\lambda_2)^2}{(1-\lambda_1\lambda_2)^2} + \frac{{\rho'_{13}}^2}{\rho'}\,\frac{(\lambda_1-\lambda_3)^2}{(1-\lambda_1\lambda_3)^2} + \frac{{\rho'_{23}}^2}{\rho'}\,\frac{(\lambda_2-\lambda_3)^2}{(1-\lambda_2\lambda_3)^2}\right.\\ &\left.+\frac{2{\rho'_{12}}^2{\rho'_{13}}^2{\rho'_{23}}^2}{\rho'}\,\frac{\left(1-{\lambda_1}^2\right)\left(1-{\lambda_2}^2\right)\left(1-{\lambda_3}^2\right) - (1-\lambda_1\lambda_2)(1-\lambda_1\lambda_3)(1-\lambda_2\lambda_3)}{(1-\lambda_1\lambda_2)(1-\lambda_1\lambda_3)(1-\lambda_2\lambda_3)}\right]\end{aligned} \tag{272}$$

for $p = 3$ using analogous simplifications. The factor ρ' equals the determinant of the covariance matrix of a standard trivariate normal distribution taking variances $c_{11}^2 = c_{22}^2 = c_{33}^2 = 1$ and is given by

$$\rho' = 1 + 2\rho'_{12}\rho'_{13}\rho'_{23} - {\rho'_{12}}^2 - {\rho'_{13}}^2 - {\rho'_{23}}^2.$$

Hence, if the dynamical operator A_0 has only real eigenvalues $\lambda_i(A_0)$, EMC can be decomposed into simple additive complexity factors and the factor related to the correlations between the covariance components of C' in the spectral basis is a simple function of the spectral gap(s).

For a process with $p = 2$ tasks that is asymptotically stable in the sense of Lyapunov (Eq. 4), it is evident that the first, second and third summand in the last row of Eq. 271 can only take values in the range $[0, +\infty)$, and for different correlations $\rho'_{12} \in [-1; 1]$ the sum of the first and second summand $-1/2\log_2\left[1 - {\lambda_1}^2\right] - 1/2\log_2\left[1 - {\lambda_2}^2\right]$ is a lower bound. To gain additional insights into the scaling behavior of EMC in the spectral gap $\Delta\lambda = (\lambda_1 - \lambda_2)$ and the correlation coefficient ρ'_{12}, we define another variable $\varsigma = (\lambda_1 + \lambda_2)$ that is orthogonal to $\Delta\lambda$. The Taylor series expansion of EMC in the spectral gap $\Delta\lambda$ about the point $\Delta\lambda = 0$ to order $\Delta\lambda^2$ leads to:

$$\mathrm{EMC} = -\frac{1}{2}\log_2\left[\left(1 - \frac{\varsigma^2}{4}\right)^2\right] + \frac{4\left(\left(1 + {\rho'_{12}}^2\right) + \varsigma^2 - 3{\rho'_{12}}^2\varsigma^2\right)}{\left({\rho'_{12}}^2 - 1\right)\left(\varsigma^2 - 4\right)\log_{10}(2)}\Delta\lambda^2 + o[\Delta\lambda]^3.$$

For the correlation coefficient ρ'_{12} we obtain the series expansion

$$\begin{aligned}\mathrm{EMC} = -\frac{1}{2}\log_2\left[\left(1 - \frac{(\varsigma - \Delta\lambda)^2}{4}\right)\left(1 - \frac{(\varsigma + \Delta\lambda)^2}{4}\right)\right] \\ + \frac{2\Delta\lambda^2}{\left(4 + \Delta\lambda^2 - \varsigma^2\right)\log_{10}(2)}{\rho'_{12}}^2 + o\left[\rho'_{12}\right]^3\end{aligned}$$

about the point $\rho'_{12} = 0$ to order ${\rho'_{12}}^2$.

For $p = 3$ tasks it can also be proved that the fourth summand in Eq. 272 can only take values in the range $[0, +\infty)$ in view of the definition of the covariance matrix. The sum of the first, second and third summands is also a lower bound.

Interestingly, the coefficient ${\rho'_{12}}^2/\left(1 - {\rho'_{12}}^2\right)$ in Eq. 271 is equivalent to Cohen's f^2, which is an effect size measure that is frequently used in the context of an F-test for ANOVA or multiple regression. By convention, in the behavioral sciences effect sizes of 0.02, 0.15, and 0.35 are termed small, medium, and large, respectively (Cohen 1988). The squared product-moment correlation ${\rho'_{12}}^2$ can also be easily interpreted within the class of linear regression models. If an intercept is included in a linear regression model, then ${\rho'_{12}}^2$ is equivalent to the well known coefficient of determination R^2. The coefficient of determination provides a measure of how well future outcomes are likely to be predicted by the statistical model.

Moreover, interesting questions arise from the identification of these lower bounds. The answers will improve the understanding of the unexpectedly rich dynamics that even small open organizational systems can generate. The identified lower bounds can be reached, if and only if either the performance fluctuations are isotropic, that is, for the corresponding covariance matrix in the original state-space coordinates the expression $C = \{\sigma^2\} \cdot I_p$ holds (see Eq. 250), or the dynamical operator A_0 is symmetric and the column vectors of the forcing matrix K are "aligned," in the sense that $A_0 = \{c\} \cdot K$ holds ($c \in \mathbb{R}$ or $c = \text{Diag}[c_i]$ in general). More details about the interrelationship between A_0 and K were presented earlier in Section 2.3. In the following, we focus on the question of how to identify the "optimal" spectrum of eigenvalues λ_i, in the sense that emergent complexity according to the metric $\text{EMC} = \frac{1}{2}\log_2 \text{Det}\left[\Sigma'_N \cdot C'^{-1}_N\right]$ is minimized subject to the constraint that the expected total amount of work $x_{tot} \in \mathbb{R}^+$ done over all tasks in the modeled project phase is constant. This constrained optimization problem will be solved under the assumptions that all eigenvalues $\lambda_i(A_0)$ are real, it holds that $\lambda_i(A_0) > 0$ and the performance fluctuations are isotropic, i.e. for a process consisting only of relaxators (see Fig. 2.6) and in which the design modes are excited as little as possible. We therefore need to find project organization designs that could, on average, process the same amount of work while leading to minimum emergent complexity. A closed-form solution of the mean vector $\bar{x}$ of the accumulated work for distinct tasks in an asymptotically stable process given the initial state x_0 can be calculated across an infinite time interval as $\bar{x} = \left(I_p - A_0\right)^{-1} \cdot x_0$ (see Section 2.2). The expected total amount of work $x_{tot} = \text{Total}[\bar{x}]$ is simply the sum of the vector components (Eq. 16). For two tasks, the above question can be formulated as the following constrained optimization problem:

$$\min_{(a_{11},a_{12},a_{21},a_{22})} \frac{1}{2}\log_2 \text{Det}\left[\left(\left(1-\lambda_1\left[\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22}\end{pmatrix}\right]^2\right)\left(1-\lambda_2\left[\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22}\end{pmatrix}\right]^2\right)\right)^{-1}\right]$$

$$\text{subject to Total}\left[\begin{pmatrix} 1-a_{11} & -a_{12} \\ -a_{21} & 1-a_{22}\end{pmatrix}^{-1} \cdot \begin{pmatrix} x_{01} \\ x_{02}\end{pmatrix}\right] = x_{tot}.$$

For three tasks, the corresponding formulation would be:

$$\min_{\left(\{a_{ij}\}_{(i,j)\epsilon\{1,2,3\}^2}\right)} \frac{1}{2}\log_2 \text{Det}\left[\prod_{i=1}^{3}\left(1-\left(\lambda_i\left[\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{pmatrix}\right]^2\right)\right)^{-1}\right]$$

$$\text{subject to Total}\left[\begin{pmatrix} 1-a_{11} & -a_{12} & -a_{13} \\ -a_{21} & 1-a_{22} & -a_{23} \\ -a_{31} & -a_{32} & 1-a_{33}\end{pmatrix}^{-1} \cdot \begin{pmatrix} x_{01} \\ x_{02} \\ x_{03}\end{pmatrix}\right] = x_{tot}.$$

In these equations, $\lambda_i[.]$ represents the i-th eigenvalue of the argument matrix. To solve the constrained optimization problems, the method of Lagrange multipliers is used. Unfortunately, this method leads to simple and expressive closed-form solutions that this book can only present and discuss under additional constraints. The first additional constraint is that only two development tasks are processed. Furthermore, both tasks have to be "uncoupled" and the corresponding off-diagonal elements $a_{12} = 0$ and $a_{21} = 0$ indicate the absence of cooperative relationships. Finally, the initial state is constrained to a setting in which both tasks are 100% to be completed, that is $x_0 = [1\ 1]^{\mathrm{T}}$, and in this case the total amount of work must be larger than $2 (x_{tot} > 2)$. Under these constraints, it follows that the eigenvalues $\lambda_1(A_0)$ and $\lambda_2(A_0)$ are equal to the autonomous task processing rates:

$$\lambda_1\left[\begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}\right] = a_{11} \quad \text{and} \quad \lambda_2\left[\begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}\right] = a_{22}\,.$$

The closed-form solution of the constrained optimization problem is the piecewise-defined complexity function:

$$\mathrm{EMC}_{min} = \begin{cases} \log_2\left(\dfrac{{x_{tot}}^2}{(x_{tot}-1)}\right) - 2 & \text{if } 2 < x_{tot} \le 2+\sqrt{2} \\ \dfrac{1}{2}\log_2(2x_{tot}-1) - 1 & \text{if } 2+\sqrt{2} < x_{tot}. \end{cases}$$

The corresponding equations for the autonomous task processing rates (alias eigenvalues) are

$$a_{11}^{min} = \lambda_1^{min} = \begin{cases} \dfrac{x_{tot}-2}{x_{tot}} & \text{if } 2 < x_{tot} \le 2+\sqrt{2} \\ \dfrac{1}{x_{tot}-1-\sqrt{2+(x_{tot}-4)x_{tot}}} & \text{if } 2+\sqrt{2} < x_{tot} \end{cases}$$

$$a_{22}^{min} = \lambda_2^{min} = \begin{cases} \dfrac{x_{tot}-2}{x_{tot}} & \text{if } 2 < x_{tot} \le 2+\sqrt{2} \\ \dfrac{1}{x_{tot}-1+\sqrt{2+(x_{tot}-4)x_{tot}}} & \text{if } 2+\sqrt{2} < x_{tot}. \end{cases}$$

When we analyze the above solutions, an interesting finding is that the value $x_{tot}^1 = 2+\sqrt{2} \sim 3.414$ of the total amount of work indicates a kind of "bifurcation point" in the complexity landscape. Below that point, minimum complexity values are assigned for equal autonomous task processing rates (or eigenvalues); above it, minimum complexity values are attained, if and only if the difference between rates (the spectral gap $\Delta\lambda^{min} = \lambda_1^{min} - \lambda_2^{min}$) is

$$a_{11}^{min} - a_{22}^{min} = \lambda_1^{min} - \lambda_2^{min} = \frac{2\sqrt{2+(x_{tot}-4)x_{tot}}}{2x_{tot}-1}.$$

This bifurcation behavior of an open organizational system in which only two uncoupled tasks are concurrently processed was unexpected. Figure 4.1 shows the bifurcation point in detail.

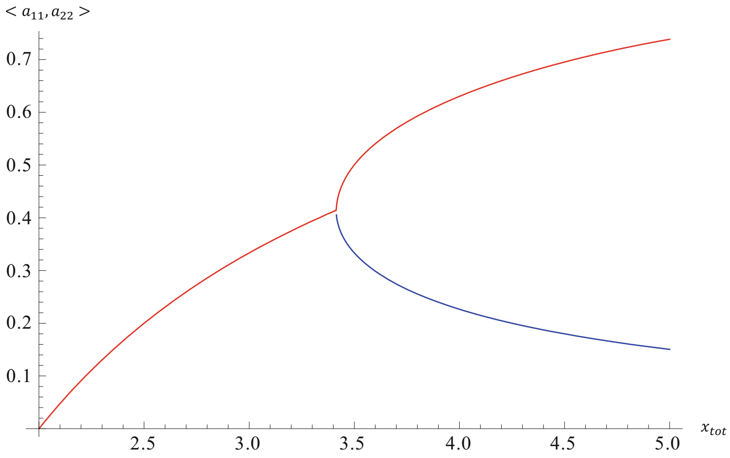

Fig. 4.1 Plot of autonomous task processing rates a_{11} and a_{22} leading to a minimum *EMC* subject to the constraint that the expected total amount of work x_{tot} is constant. The underlying closed-form solution was calculated based on Lagrange multipliers. Note that the solution only holds under the assumption that the tasks are uncoupled and the initial state is $x_0 = [1\ 1]^{\mathrm{T}}$, in which case x_{tot} must be larger than 2

We also found analytical results for the constrained optimization problem in the more general case of two uncoupled overlapping tasks, i.e. a bundle of independent tasks where, initially, only the second task has to be fully completed, while the first task is already completed to a level of $x\,\%$ and we therefore have an initial state $x_0 = [(1-x)/100\ 1]^{\mathrm{T}}$. However, the closed-form solutions are very complicated and, due to space limitations, cannot be presented here. It is important to note that the piecewise-defined complexity function and the corresponding bifurcation point are completely independent of the degree of task overlapping and only depend on the dynamics of task processing. This is a highly desirable property of the preferred complexity metric.

When we relax the constraint that both tasks have to be uncoupled, and consider all four matrix entries of the WTM A_0 as free parameters, we find another simple analytical solution to the constrained optimization problem. The initial state is constrained to be $x_0 = [1\ 1]^{\mathrm{T}}$ as before. However, the obtained solution is not very structurally informative, as all four elements of A_0 are supposed to be equal to $(x_{tot} - 2)/2x_{tot}$, and we have the symmetric matrix representation

$$A_0^{min} = \begin{pmatrix} \frac{1}{2} - \frac{1}{x_{tot}} & \frac{1}{2} - \frac{1}{x_{tot}} \\ \frac{1}{2} - \frac{1}{x_{tot}} & \frac{1}{2} - \frac{1}{x_{tot}} \end{pmatrix}.$$

From a practical point of view, this kind of project organizational design seems to be rather "pathological" because the relative couplings between tasks are extremely strong and one must expect a large amount of additional work in the iterations. The corresponding complexity solution is

$$\mathrm{EMC}_{min} = \frac{1}{2}\log_2\left(\frac{{x_{tot}}^2}{4(x_{tot}-1)}\right) \quad \text{if } 2 < x_{tot} \ .$$

It is evident that the above minimum of emergent complexity scales for $x_{tot} > 2.5$ almost linearly in the expected total amount of work.

4.1.5 Bounds on Effective Measure Complexity

To calculate the lower bounds on EMC for an arbitrary number of tasks we can make use of Oppenheim's inequality (see Horn and Johnson 1985). Let M and N be positive-semidefinite matrices and let $M \circ N$ be the entry-wise product of these matrices (so-called "Hadamard product"). The Hadamard product of two positive-semidefinite matrices is again positive-semidefinite. Furthermore, if M and N are positive-semidefinite, then the following equality based on Oppenheim holds:

$$\mathrm{Det}[M \circ N] \geq \left(\prod\nolimits_{i=1}^{p} M_{[[i,i]]}\right)\mathrm{Det}[N].$$

Let $M = \left(M_{[[i,j]]}\right) = \left(1/\left(1-\lambda_i(A_0)\overline{\lambda_j(A_0)}\right)\right)$ be a Cauchy matrix $(1 \leq i, j \leq p)$. The elements along the principal diagonal of this matrix represent the "damping factor" $1-\left|\lambda_i\right|^2$ of design mode ϕ_i, and the off-diagonal elements $1-\lambda_i\overline{\lambda_i}$ are the damping factors between the interacting modes ϕ_i and ϕ_j. We follow the convention that the eigenvalues are ordered in decreasing magnitude in rows. Let $N = C'_N$ be the normalized covariance matrix of the noise, as defined in Eq. 260. Then the normalized covariance matrix of the signal Σ'_N from Eq. 259 can be written as the Hadamard product $\Sigma'_N = M \circ C'_N$. According to Oppenheim's inequality, the following inequality holds:

$$\begin{aligned}\mathrm{EMC} = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\Sigma'_N\right]}{\mathrm{Det}\left[C'_N\right]}\right) = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[M \circ C'_N\right]}{\mathrm{Det}\left[C'_N\right]}\right) &\geq \frac{1}{2}\log_2\left(\frac{\left(\prod\nolimits_{i=1}^{p} M_{[[i,i]]}\right)\mathrm{Det}\left[C'_N\right]}{\mathrm{Det}\left[C'_N\right]}\right)\\ &= \frac{1}{2}\log_2\left(\prod\nolimits_{i=1}^{p}\frac{1}{1-\left|\lambda_i\right|^2}\right)\\ &= -\frac{1}{2}\sum_{i=1}^{p}\log_2\left(1-\left|\lambda_i\right|^2\right).\end{aligned} \tag{273}$$

The lower bound according to the above equation is equal to the closed-form solution for EMC that was obtained under the assumptions of isotropic noise $\left(C = \left\{\sigma^2\right\} \cdot I_p\right)$ and A_0 being diagonalizable (see Eq. 251). In other words, emergent complexity in PD projects can be kept to a minimum, if the variances of the unpredictable performance fluctuations are equalized by purposeful interventions of the project manager and correlations between vector components are suppressed.

Next, because of the commutativity of the Hadamard product, it holds that

$$\mathrm{EMC}=\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\Sigma'_N\right]}{\mathrm{Det}\left[C'_N\right]}\right)=\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[C'_N\circ M\right]}{\mathrm{Det}\left[C'_N\right]}\right)\geq\frac{1}{2}\log_2\left(\frac{\left(\prod_{i=1}^{p}C'_{N[[i,i]]}\right)\mathrm{Det}[M]}{\mathrm{Det}\left[C'_N\right]}\right)$$
$$=\frac{1}{2}\log_2\left(\frac{\mathrm{Det}[M]}{\mathrm{Det}\left[C'_N\right]}\right).$$

The determinant of the Cauchy matrix M in the numerator can be written as (Krattenthaler 2005)

$$\mathrm{Det}[M]=\mathrm{Det}\begin{bmatrix}\frac{1}{1-|\lambda_1|^2} & \frac{1}{1-\lambda_1\overline{\lambda_2}} & \cdots\\ \frac{1}{1-\lambda_2\overline{\lambda_1}} & \frac{1}{1-|\lambda_2|^2} & \cdots\\ \vdots & \vdots & \ddots\end{bmatrix}=\frac{\prod_{i<j}^{p}\left(\lambda_i-\lambda_j\right)\left(\overline{\lambda_i}-\overline{\lambda_j}\right)}{\prod_{i,j}^{p}\left(1-\lambda_i\overline{\lambda_j}\right)}.$$

Hence,

$$\begin{aligned}\mathrm{EMC}&=\frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[C'_N\circ M\right]}{\mathrm{Det}\left[C'_N\right]}\right)\\&\geq\frac{1}{2}\log_2\left(\frac{\prod_{i<j}^{p}\left(\lambda_i-\lambda_j\right)\left(\overline{\lambda_i}-\overline{\lambda_j}\right)}{\prod_{i,j}^{p}\left(1-\lambda_i\overline{\lambda_j}\right)\mathrm{Det}\left[C'_N\right]}\right)\\&=\frac{1}{2}\left(\sum_{i<j}^{p}\left(\log_2\left(\lambda_i-\lambda_j\right)+\log_2\left(\overline{\lambda_i}-\overline{\lambda_j}\right)\right)-\sum_{i,j}^{p}\log_2\left(1-\lambda_i\overline{\lambda_j}\right)-\log_2\mathrm{Det}\left[C'_N\right]\right).\end{aligned}\tag{274}$$

The lower bound on the EMC in the above equation is only defined for a dynamical operator A_0 with distinct eigenvalues. Under this assumption, a particularly interesting property of the bound is that it includes not only the damping factors $\left(1-\lambda_i\overline{\lambda_i}\right)$ inherent to the dynamical operator A_0 (as does the bound in Eq. 273) but also the spectral gap between eigenvalues $\left(\lambda_i-\lambda_j\right)$ and their complex conjugates $\left(\overline{\lambda_i}-\overline{\lambda_j}\right)$. We can draw the conclusion that under certain circumstances, differences among effective productivity rates (represented by the λ_i's) stimulate emergent complexity in PD (cf. Eqs. 271 and 272). Conversely, small complexity scores are assigned if the effective productivity rates are similar.

Additional analyses have shown that the lower bound defined in Eq. 273 is tighter when the eigenvalues of the dynamical operator A_0 are of similar magnitudes. Conversely, the lower bound defined in Eq. 274 comes closer to the true complexity values if the magnitudes of the eigenvalues are unevenly distributed.

Finally, it is also possible to put both upper and lower bounds on the EMC that are explicit functions of the dynamical operator A_0 and its dimension p. To find these bounds, we considered results for the determinant of the solution of the Lyapunov equation (Eq. 27, cf. Mori et al. 1982). Let Σ be the covariance matrix of the process in the steady state, and let the dominant eigenvalue $\rho(A_0) = \max|_i|$ of A_0 be less than 1 in magnitude (see Section 2.1). Then we have

$$\mathrm{Det}[\Sigma] \geq \frac{\mathrm{Det}[C]}{\left(1 - (\mathrm{Det}[A_0])^{\frac{p}{2}}\right)^p} .$$

Moreover, if A_0 is diagonalizable and $\rho\left(A_0^{\mathrm{T}} \cdot A_0\right) \cdot C - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}}$ is positive-semidefinite, then

$$\mathrm{Det}[\Sigma] \leq \frac{\mathrm{Det}[C]}{\left(1 - \rho\left(A_0^{\mathrm{T}} \cdot A_0\right)\right)^p},$$

where $\rho\left(A_0^{\mathrm{T}} \cdot A_0\right)$ denotes the dominant eigenvalue of $A_0^{\mathrm{T}} \cdot A_0$. Based on Eq. 246 we can calculate the following bounds:

$$-\frac{p}{2}\log_2\left(1 - (\mathrm{Det}[A_0])^{\frac{p}{2}}\right) \leq \mathrm{EMC} \leq -\frac{p}{2}\log_2\left(1 - \rho\left(A_0^{\mathrm{T}} \cdot A_0\right)\right). \tag{275}$$

The upper bound only holds if A_0 is diagonalizable and $\rho\left(A_0^{\mathrm{T}} \cdot A_0\right) \cdot C - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}}$ is positive-semidefinite. If C is diagonal, then $\rho\left(A_0^{\mathrm{T}} \cdot A_0\right) \cdot C - A_0 \cdot \Sigma \cdot A_0^{\mathrm{T}}$ is always positive-semidefinite. Both bounds grow strictly monotonically with the dimension of the dynamical operator A_0 and it is evident that the EMC assigns larger complexity values to projects with more tasks, if the task couplings are similar. One can also divide the measure by the dimension p of the state space and compare the complexity of project phases with different cardinalities.

4.1.6 Closed-Form Solutions for Higher-Order Models

It is also not difficult to calculate the EMC of stochastic processes in steady state that are generated by higher-order autoregressive models of cooperative work in PD projects. In Section 2.4, we said that a vector autoregression model of order n, abbreviated as VAR(n) model, without an intercept term is defined by the state equation (see Neumaier and Schneider 2001 or Lütkepohl 2005):

$$X_t = \sum_{i=0}^{n-1} A_i \cdot X_{t-i-1} + \varepsilon_t.$$

The probability density function of the vector ε_t of performance fluctuations is given in Eq. 13. It is evident that due to the autoregressive behavior involving n instances of the process in the past, the generated stochastic process $\{X_t\}$ does not possess the Markov property (cf. Eq. 18) and therefore neither the generalized complexity solution from Eq. 239 nor the closed-form solution for a VAR(1) process from Eq. 247 can be used to evaluate emergent complexity. However, as we showed in Section 2.5, we can make the stochastic process Markovian by "augmenting" the state vector and rewriting the state equation as a first-order recurrence relation (Eq. 59):

$$\tilde{X}_t = \tilde{A} \cdot \tilde{X}_{t-1} + \tilde{\varepsilon}_t \quad t = 1, \ldots, T, \tag{276}$$

where $\tilde{X}_t$ is the augmented state vector (Eq. 60)

$$\tilde{X}_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-n+1} \end{pmatrix},$$

$\tilde{\varepsilon}_t$ is the augmented noise vector (Eq. 61)

$$\tilde{\varepsilon}_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and $\tilde{A}$ is the extended dynamical operator (Eq. 62)

$$\tilde{A} = \begin{pmatrix} A_0 & A_1 & \cdots & A_{n-2} & A_{n-1} \\ I_p & 0 & \cdots & 0 & 0 \\ 0 & I_p & \cdots & 0 & 0 \\ & & & & \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \ldots & I_p & 0 \end{pmatrix}.$$

The covariance matrix $\tilde{C}$ can be written as

$$\begin{aligned}\tilde{C} &= E\left[\tilde{\varepsilon}_t \tilde{\varepsilon}_t^{\mathrm{T}}\right] \\ &= \begin{pmatrix} C & 0 & \cdots & 0 \\ 0 & 0 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.\end{aligned} \tag{277}$$

The partial covariance $C = E\left[\varepsilon_t \varepsilon_t^{\mathrm{T}}\right]$ represents the intrinsic one-step prediction error of the original autoregressive process.

In light of the mutual information that is communicated from the infinite past to the infinite future (by storing it in the present) the problem with this kind of order reduction by state-space augmentation is that the augmented state vector $\tilde{X}_t$ has vector components that are also included in the previous state vector $\tilde{X}_{t-1}$ and therefore the past and future are not completely shielded in information-theoretic terms, given the present state. To be able to apply the closed-form complexity solution from Eq. 247 directly to the higher-order model in the coordinates of the original state space $\mathbb{R}^p$, we have to find a state representation with disjoint vector components. This can be easily done by defining the combined future and present project state $\tilde{X}_{t+n-1}$ to be the block of random vectors

$$\tilde{X}_{t+n-1} = \begin{pmatrix} X_{t+n-1} \\ X_{t+n-2} \\ \vdots \\ X_t \end{pmatrix}$$

and the past project state $\tilde{X}_{t-1}$ to be the block of vectors

$$\tilde{X}_{t-1} = \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-n} \end{pmatrix}.$$

The calculation of the n-th iterate of $\tilde{X}_{t+n-1}$ leads to the higher-order recurrence relation

$$\begin{aligned}\tilde{X}_{t+n-1} &= \tilde{A} \cdot \tilde{X}_{t+n-2} + \tilde{\varepsilon}_{t+n-1} \\ &= \tilde{A}\left(\tilde{A} \cdot \tilde{X}_{t+n-3} + \tilde{\varepsilon}_{t+n-2}\right) + \tilde{\varepsilon}_{t+n-1} \\ &= \tilde{A}^2 \cdot \tilde{X}_{t+n-3} + \tilde{A} \cdot \tilde{\varepsilon}_{t+n-2} + \tilde{\varepsilon}_{t+n-1} \\ &\vdots \\ &= \tilde{A}^n \cdot \tilde{X}_{t-1} + \sum_{i=1}^{n} \left(\tilde{A}\right)^{n-i} \cdot \tilde{\varepsilon}_{t+i-1} \quad t = 2-n, \ldots, T-n+1.\end{aligned} \tag{278}$$

Under the assumption of strictly stationary behavior of $\left\{\tilde{X}_t\right\}$ for $t \to \infty$, we can utilize the complexity solution from eq. 247 and express the mutual information that is communicated by the VAR(n) model from the infinite past to the infinite future through the present project state by the logarithmic generalized variance ratio as follows:

$$\mathrm{EMC} = \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\sum_{k=0}^{\infty}\left(\tilde{A}^{n}\right)^{k}\cdot\left(\sum_{i=1}^{n}\left(\tilde{A}\right)^{n-i}\cdot\tilde{C}\cdot\left(\left(\tilde{A}\right)^{n-i}\right)^{\mathrm{T}}\right)\cdot\left(\left(\tilde{A}^{n}\right)^{\mathrm{T}}\right)^{k}\right]}{\mathrm{Det}\left[\sum_{i=1}^{n}\left(\tilde{A}\right)^{n-i}\cdot\tilde{C}\cdot\left(\left(\tilde{A}\right)^{n-i}\right)^{\mathrm{T}}\right]}\right)$$

$$= \frac{1}{2}\log_2\left(\frac{\mathrm{Det}\left[\tilde{\Sigma}\right]}{\mathrm{Det}\left[\sum_{i=1}^{n}\left(\tilde{A}\right)^{n-i}\cdot\tilde{C}\cdot\left(\left(\tilde{A}\right)^{n-i}\right)^{\mathrm{T}}\right]}\right), \tag{279}$$

where the steady-state covariance $\tilde{\Sigma}$ in the denominator is given by the infinite sum

$$\tilde{\Sigma} = \sum_{k=0}^{\infty}\left(\tilde{A}^{n}\right)^{k}\cdot\left(\sum_{i=1}^{n}\left(\tilde{A}\right)^{n-i}\cdot\tilde{C}\cdot\left(\left(\tilde{A}\right)^{n-i}\right)^{\mathrm{T}}\right)\cdot\left(\left(\tilde{A}^{n}\right)^{\mathrm{T}}\right)^{k}.$$

As an alternative to this solution, we can calculate the mutual information between infinite past and future histories using the additive factors method of Li and Xie (1996). In this method, the total mutual information is decomposed into additive components which can be expressed as a ratio of conditional (auto)covariances of the steady-state process. This method is very appealing as it allows us to interpret the additive components in terms of the universal learning curve $\Lambda(m)$ that was formulated by Bialek et al. (2001, see Eq. 224) and is explained in detail in Section 3.2.4. EMC is simply the discrete integral of $\Lambda(m)$ with respect to the block length m, which controls the speed at which the mutual information converges to its limit (Crutchfield et al. 2010). When we use the block length as a natural order parameter of the additive components, we can also easily evaluate the speed of convergence. If convergence is slow, it is an indicator of emergent complexity (see discussion in Section 3.2.4).

Let

$$\boldsymbol{C}_m = \begin{pmatrix} C_{\tilde{X}\tilde{X}}(0) & C_{\tilde{X}\tilde{X}}(1) & \ldots & C_{\tilde{X}\tilde{X}}(m-1) \\ C_{\tilde{X}\tilde{X}}(1) & C_{\tilde{X}\tilde{X}}(0) & \ldots & C_{\tilde{X}\tilde{X}}(m-2) \\ \vdots & \vdots & \ddots & \vdots \\ C_{\tilde{X}\tilde{X}}(m-1) & C_{\tilde{X}\tilde{X}}(m-2) & \ldots & C_{\tilde{X}\tilde{X}}(0) \end{pmatrix} \tag{280}$$

be a $mp \times mp$ $(m \in \mathbb{N})$ Toeplitz matrix (Li and Xie 1996) storing the values of the autocovariance functions (Eq. 159)

$$\begin{aligned} C_{\tilde{X}\tilde{X}}(\tau) &= E\left[\left(\tilde{X}_t - \mu_{\tilde{X}}\right)\left(\tilde{X}_{t+\tau} - \mu_{\tilde{X}}\right)^{\mathrm{T}}\right] \\ &= E\left[\tilde{X}_t \tilde{X}_{t+\tau}^{\mathrm{T}}\right] - \mu_{\tilde{X}} \mu_{\tilde{X}}^{\mathrm{T}} \\ &= E\left[\tilde{X}_t \tilde{X}_{t+\tau}^{\mathrm{T}}\right] \end{aligned}$$

of the steady-state process generated by state Eq. 59 (and not Eq. 276) for lead times $\tau = 0, 1, \ldots, m-1$. We know from Section 2.9 that, in steady state, the autocovariance $C_{\tilde{X}\tilde{X}}(\tau)$ and the autocorrelation $R_{\tilde{X}\tilde{X}}(\tau)$ (Eq. 160) are equal and that we have $C_{\tilde{X}\tilde{X}}(\tau) = R_{\tilde{X}\tilde{X}}(\tau)$. Note that the matrix elements $C_{\tilde{X}\tilde{X}}(\tau)$ are defined to be $p \times p$ block autocovariance matrices of the corresponding subspaces. Furthermore, let

$$\tilde{\Sigma}_{\mu} = \mathrm{Det}\left[E\left[\left(\tilde{X}_t - E\left[\tilde{X}_t \middle| \tilde{X}_{t-1}, \ldots, \tilde{X}_{-\infty}\right]\right)\left(\tilde{X}_t - E\left[\tilde{X}_t \middle| \tilde{X}_{t-1}, \ldots, \tilde{X}_{-\infty}\right]\right)^{\mathrm{T}}\right]\right]$$

be the (mean squared) one-step prediction error with respect to the steady-state process and

$$\tilde{\Sigma}_{(m)} = \mathrm{Det}\left[E\left[\left(\tilde{X}_t - E\left[\tilde{X}_t \middle| \tilde{X}_{t-1}, \ldots, \tilde{X}_{t-m}\right]\right)\left(\tilde{X}_t - E\left[\tilde{X}_t \middle| \tilde{X}_{t-1}, \ldots, \tilde{X}_{t-m}\right]\right)^{\mathrm{T}}\right]\right] \tag{281}$$

be the one-step prediction error of order m (cf. Eq. 66 in Section 2.4). According to these definitions $\tilde{\Sigma}_{\mu}$ can be interpreted as the inherent prediction error of the process that cannot be underrun, even if we condition our observations on the infinite past to build a maximally predictive model. $\tilde{\Sigma}_{(m)}$ represents the prediction error resulting from conditioning the observations on only m past instances of the process to build a maximally predictive model, and not on all instances that were theoretically possible. In this sense a certain error component of $\tilde{\Sigma}_{(m)}$ does not result from the inherent unpredictability because of limited knowledge or chaotic behavior, but because of the unpredictability resulting from a limit of the length of the observation window on the state evolution. Under the assumption that $\boldsymbol{C}_m$ is invertible, the one-step prediction error $\tilde{\Sigma}_{(m)}$ of order m can be expressed as the generalized variance ratio (Li and Xie 1996):

$$\tilde{\Sigma}_{(m)} = \frac{\mathrm{Det}[\boldsymbol{C}_{m+1}]}{\mathrm{Det}[\boldsymbol{C}_m]}. \tag{282}$$

The zeroth-order prediction error can be derived from the autocovariance for zero lead time, and it holds that:

$$\tilde{\Sigma}_{(0)} = \mathrm{Det}\left[C_{\tilde{X}\tilde{X}}(0)\right]. \tag{283}$$

Following the information-theoretic considerations of a VAR(n) model that were carried out in Section 4.1 (cf. Eq. 238) it is not difficult to show that, for any autocovariance matrix representation $\boldsymbol{C}_m$ with $m \geq n$, it holds for the one-step prediction errors that

$$\tilde{\Sigma}_{(m)} = \tilde{\Sigma}_{(n)} = \tilde{\Sigma}_{\mu} \quad \forall m \geq n\,.$$

In other words, due to the limited "memory depth" of the generative VAR(n) model, conditioning the current observation on sequences larger than the regression order does not, on average, lead to further reductions of the one-step prediction error in steady state. Under these circumstances of severely limited procedural memory the prediction error of order n equals the intrinsic prediction error. As a consequence of this behavior, higher-dimensional matrix representations than $\boldsymbol{C}_n$ must not be considered when evaluating the past-future mutual information. An additional theoretical analysis of the vector autoregression model in the original state-space coordinates allows us to conclude that the inherent prediction error equals the determinant of the expectation $E[\varepsilon_t \varepsilon_t^{\mathrm{T}}]$ and that it can be simply expressed as

$$\tilde{\Sigma}_{\mu} = \mathrm{Det}\left[E\left[\varepsilon_t \varepsilon_t^{\mathrm{T}}\right]\right] = \mathrm{Det}[C].$$

Furthermore, in steady state the $np \times np$ matrix $\boldsymbol{C}_n$ storing all relevant autocovariances up to lead time $\tau = n - 1$ equals the steady-state covariance of the process generated by state Eq. 59, and we have (Eq. 245, Lancaster and Tismenetsky 1985):

$$\boldsymbol{C}_n = \sum_{k=0}^{\infty} \tilde{A}^{k} \cdot \tilde{C} \cdot \left(\tilde{A}^{\mathrm{T}}\right)^{k},$$

where $\tilde{A}$ is the extended dynamical operator from Eq. 62, and $\tilde{C}$ is the corresponding covariance matrix from Eq. 277. If needed, the autocovariances for smaller lead times can be easily extracted as block matrices from this large representation. Based on these theoretic considerations and the material of Li and Xie (1996), the mutual information between infinite past and future histories can be conveniently expressed by n additive components as

$$\begin{aligned} \mathrm{EMC} &= \frac{1}{2}\left(\sum_{i=0}^{n-1}\left(\log_2 \tilde{\Sigma}_{(i)} - \log_2 \tilde{\Sigma}_{\mu}\right)\right) \\ &= \frac{1}{2}\left(\sum_{i=0}^{n-1} \log_2 \tilde{\Sigma}_{(i)} - n \log_2 \tilde{\Sigma}_{\mu}\right) \\ &= \frac{1}{2}\sum_{i=0}^{n-1} \log_2 \tilde{\Sigma}_{(i)} - \frac{1}{2} n \log_2 \mathrm{Det}[C]. \end{aligned} \tag{284}$$

Each summand $1/2\left(\log_2 \tilde{\Sigma}_{(i)} - \log_2 \tilde{\Sigma}_{\mu}\right) = 1/2\left(\log_2 \tilde{\Sigma}_{(i)} - \log_2 \mathrm{Det}[C]\right)$ can be used to evaluate the local predictability of the process. The corresponding local "over-estimates" of the intrinsic prediction error allow us to define a universal learning curve $\Lambda(i)$ in the sense of Bialek et al. (2001) with respect to block length i as (cf. Eq. 224)

$$\Lambda(i) = \log_2 \tilde{\Sigma}_{(i-1)} - \log_2 \mathrm{Det}[C], \quad i = 1, 2, \ldots, n,$$

where the maximum block length is determined by the autoregression order of the generative model. As already explained in Section 3.1.4, in light of a learning curve, EMC measures the amount of apparent randomness at small order i, which can be "explained away" by considering correlations between sequences with increasing length $i+1, \ i+2, \ldots$.

Returning to state Eq. 276 for informationally separated instances of past and future histories, we can use the first-order recurrence relation to apply the solution principles that were introduced at the end of Section 4.1.1 and find a simple expression for the persistent mutual information $\mathrm{EMC}(\tau)$ (Eq. 229) as a function of the lead time $\tau \geq 0$. Substituting the steady-state covariance and the dynamical operator in Eq. 254, we can express $\mathrm{EMC}(\tau)$ as the logarithmic generalized variance ratio (Ay et al. 2012):

$$\mathrm{EMC}(\tau) = \frac{1}{2} \log_2 \left(\frac{\mathrm{Det}\left[\tilde{\Sigma}\right]}{\mathrm{Det}\left[\tilde{\Sigma} - \left(\tilde{A}^{\,n}\right)^{\tau+1} \cdot \tilde{\Sigma} \cdot \left(\left(\tilde{A}^{\,n}\right)^{\mathrm{T}}\right)^{\tau+1}\right]} \right), \tag{285}$$

As one would expect, the steady-state covariances in the numerators of the variance ratios related to both measures of emergent complexity are equal (Eqs. 279 and 285). We note that for the inherent one-step prediction error it holds that $E\left[\tilde{\varepsilon}_t \tilde{\varepsilon}_t^{\mathrm{T}}\right] = \tilde{C} = \tilde{\Sigma} - \tilde{A}^{\,n} \cdot \tilde{\Sigma} \cdot \left(\tilde{A}^{\,n}\right)^{\mathrm{T}}$.

Applying the principles and techniques introduced in Section 4.1.2 and 4.1.3, it is also not difficult to derive additional closed-form solutions in the spectral basis and other coordinate systems. We leave this as an exercise for the interested reader.

With the previous complexity considerations of higher-order autoregressive models of cooperative work in PD projects, it is possible to analyze in detail the differences between the EMC as originally developed by Grassberger (1986) and the persistent mutual information $\mathrm{EMC}(\tau)$ according to Eq. 229, proposed recently by Ball et al. (2010) as a complexity measure. In order to clarify the differences between both measures we refer to the seminal work of Li (2006) and evaluate both the emergent complexity of a strict-sense stationary process $\{X_t\}$ generated by a VAR(n) model, and the emergent complexity related to the model in conjunction with a causal finite impulse response (FIR) filter (see e.g. Puri 2010) of order $m (m \geq 1)$. Each of the output sequences of such a filter is a weighted sum of the most recent m filter input values:

$$y_t = \sum_{i=0}^{m} b_i \cdot x_{t-i}.$$

The b_i's denotes the filter coefficients. The transfer function of the FIR filter is denoted by $H(z)$ (cf. Section 4.2.1). It is assumed that the filter has all its roots on the unit circle. We pass the VAR(n) model outputs x_t through the filter to obtain the output sequence y_t. If $m \geq 1$, according to Li (2006) it holds that the EMC^y related to the stationary filter output y_t is not finite:

$$\mathrm{EMC}^y \to \infty.$$

However, the corresponding persistent mutual information EMC^y(m) is finite and equal to the effective measure complexity EMC^x of the steady-state process that is filtered:

$$\mathrm{EMC}^y(m) = \mathrm{EMC}^x.$$

Li (2006) proved these properties for arbitrary stationary Gaussian processes. His theorems also show that zeros on the unit circle can easily cause EMC to be infinite. For instance, even for a simple first-order moving average process $\{X_t\}$ (a so-called MA(1) process, see Section 4.2.1) generated by state equation

$$X_t = \varepsilon_t - \varepsilon_{t-1}$$

the corresponding effective measure complexity

$$\mathrm{EMC} \to \infty$$

grows over all given limits (Li 2006). Nevertheless, the persistent mutual information

$$\mathrm{EMC}(1) < \infty$$

for lead time one is finite. Hence, in cases where we have a transfer function in the form of a polynomial of degree m that has all its roots on the unit circle, the persistent mutual information EMC(τ) according to Eq. 229 should be used instead of the original formulation of the EMC. However, these cases are extremely rare in project management.

4.2 Closed-Form Solutions of Effective Measure Complexity for Linear Dynamical System Models of Cooperative Work

4.2.1 Explicit Formulation

According to the analytical considerations set out at the beginning of Section 4.1, the EMC of a linear dynamical system (LDS, see Section 2.9.) as an advanced model of cooperative work in PD projects, which is defined by the system of equations

$$X_{t+1} = A_0 X_t + \varepsilon_t$$
$$Y_t = H X_t + \nu_t$$

with $\varepsilon_t = \mathcal{N}\left(\xi; 0_q, C\right)$ and $\nu_t = \mathcal{N}\left(\eta; 0_p, V\right)$, can be expressed by the continuous-type mutual information $I[.;.]$ as

$$\begin{aligned} \mathrm{EMC} &= I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] \\ &= \int f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] \log_2 \frac{f\left[y_{-\infty}^{-1}, y_0^{\infty}\right]}{f\left[y_{-\infty}^{-1}\right] f\left[y_0^{\infty}\right]} dy_{-\infty}^{\infty} \; . \end{aligned} \tag{286}$$

In contrast to the previous chapters we have not written the multiplication symbol "·" between a matrix and a vector explicitly in the above equations. We will use this more compact notation here and in the following chapter to save space and simplify the interpretation of longer terms. Their meaning should always be clear from the context.

The function $f\left[y_{-\infty}^{-1}\right]$ designates the joint *pdf* of the observable infinite one-dimensional history. Similarly, the function $f\left[y_0^{\infty}\right]$ represents the corresponding *pdf* of the observable infinite future.

It is important to point out that if, and only if, the joint *pdf* of the past $f\left[y_{-\infty}^{\infty}\right]$ and future $f\left[y_0^{\infty}\right]$ histories of observations reach the same steady state, the evaluation of the infinite-dimensional integral yields a finite value. Otherwise, the integral will diverge, as will become clear below. This is possible if the covariance for the initial state in the infinite past with *pdf* given by $f[x_{-\infty}] = (x_t; \mu, \Sigma_0)$ equals the one in the steady state Σ, i.e. the one that satisfies the Lyapunov criterion

$$\Sigma = A_0 \Sigma A_0^{\mathrm{T}} + C$$

from Eq. 27. If the initial state is in steady state, then its expected value is the zero vector $\mu = 0$.

In what follows we will therefore assume that the hidden Markov process $\{X_t\}$ is strict-sense stationary and that in steady state a stable distribution $f[x_\nu]$ is formed. From the state-space model, the following normal distributions can be deduced in steady state:

$$\begin{aligned} f[x_\nu] &= \mathcal{N}(x_\nu;\mu,\Sigma) \\ f\left[x_\nu \middle| x_{\nu-1}\right] &= \mathcal{N}(x_\nu;A_0 x_{\nu-1},C) \\ f\left[y_\nu \middle| x_\nu\right] &= \mathcal{N}(y_\nu;Hx_\nu,V). \end{aligned}$$

Before we proceed, note of the following: if only the observations y_t are available, it is always possible to introduce an arbitrary invertible transform T so that the model for the observations $Y_t = HX_t + \nu_t$ remains unchanged if $H' = HT, X'_t = T^{-1}X_t$, as

$$\begin{aligned} Y_t &= H'X'_{\mathrm{t}} + \nu_t \\ &= HTT^{-1}X_t + \nu_t \\ &= HX_t + \nu_t. \end{aligned} \tag{287}$$

For example, one could choose a whitening transform, cf. Eqs. 156, 157 and 158, $X'_t = \Lambda_u^{-1/2}U^{\mathrm{T}}X_t$ for the hidden-state process which leads to a covariance of the performance fluctuations equal to the identity matrix $C = I_q$. However, in the subsequent derivations, we will continue to use a general covariance C to clarify the interrelationships between the random performance fluctuations and emergent complexity. Following the notation introduced in Section 2.9, we will use the (long) vector $\boldsymbol{y}_{-\infty}^{\infty}$ of the stacked variables $y_{-\infty}^{\infty}$, i.e. $y_{-\infty}^{\infty} = \left(y_{-\infty}^{T}, \cdots, y_{\infty}^{T}\right)^{T}$ in what follows. The vectors $\boldsymbol{y}_{-\infty}^{-1}$ and $\boldsymbol{y}_{0}^{\infty}$ are defined accordingly. We also add subscripts and superscripts to the quantities $V, C, \Delta t, b$ to mark the corresponding time step. The three joint *pdf*'s in the general definition of the EMC are given for the Gaussian density model (see Eq. 134 in Section 2.9):

$$f\left[\boldsymbol{y}_{-\infty}^{\infty}\right] = c_{\boldsymbol{y}_{-\infty}^{\infty}} \mathrm{Exp}\left[-\frac{1}{2}\left(\boldsymbol{y}_{-\infty}^{\infty}\right)^{\mathrm{T}} \mathcal{V}_{-\infty}^{\infty} \boldsymbol{y}_{-\infty}^{\infty}\right] \frac{(2\pi)^{\Delta t_{-\infty}^{\infty} q/2}}{\sqrt{\mathrm{Det}\mathcal{C}_{-\infty}^{\infty}}} \mathrm{Exp}\left[\frac{1}{2}\left(\boldsymbol{b}_{-\infty}^{\infty}\right)^{\mathrm{T}} \left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1} \boldsymbol{b}_{-\infty}^{\infty}\right] \tag{288}$$

$$f\left[\boldsymbol{y}_{-\infty}^{-1}\right] = c_{\boldsymbol{y}_{-\infty}^{-1}} \mathrm{Exp}\left[-\frac{1}{2}\left(\boldsymbol{y}_{-\infty}^{-1}\right)^{\mathrm{T}} \mathcal{V}_{-\infty}^{-1} \boldsymbol{y}_{-\infty}^{-1}\right] \frac{(2\pi)^{\Delta t_{-\infty}^{-1} q/2}}{\sqrt{\mathrm{Det}\mathcal{C}_{-\infty}^{-1}}} \mathrm{Exp}\left[\frac{1}{2}\left(\boldsymbol{b}_{-\infty}^{-1}\right)^{\mathrm{T}} \left(\mathcal{C}_{-\infty}^{-1}\right)^{-1} \boldsymbol{b}_{-\infty}^{-1}\right] \tag{289}$$

$$f\left[\boldsymbol{y}_{0}^{\infty}\right] = c_{\boldsymbol{y}_{0}^{\infty}} \mathrm{Exp}\left[-\frac{1}{2}\left(\boldsymbol{y}_{0}^{\infty}\right)^{\mathrm{T}} \mathcal{V}_{0}^{\infty} \boldsymbol{y}_{0}^{\infty}\right] \frac{(2\pi)^{\Delta t_{0}^{\infty} q/2}}{\sqrt{\mathrm{Det}\mathcal{C}_{0}^{\infty}}} \mathrm{Exp}\left[\frac{1}{2}\left(\boldsymbol{b}_{0}^{\infty}\right)^{\mathrm{T}} \left(\mathcal{C}_{0}^{\infty}\right)^{-1} \boldsymbol{b}_{0}^{\infty}\right]. \tag{290}$$

Within a direct calculation of the EMC, given by the integral 286, here are two possible paths: One involves splitting the integral into two parts:

$$
\begin{aligned}
I\left[Y_{-\infty}^{-1};Y_0^{\infty}\right] &= \int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2\frac{f\left[y_{-\infty}^{-1},y_0^{\infty}\right]}{f\left[y_{-\infty}^{-1}\right]f\left[y_0^{\infty}\right]}dy_{-\infty}^{\infty}\\
&= \int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 f\left[y_{-\infty}^{-1},y_0^{\infty}\right]dy_{-\infty}^{\infty}\\
&\quad -\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 f\left[y_{-\infty}^{-1}\right]f\left[y_0^{\infty}\right]dy_{-\infty}^{\infty}.
\end{aligned}
$$

The other involves leaving the integral as a whole, computing first the ratio $f\left[y_{-\infty}^{\infty}\right]/\left(f\left[y_{-\infty}^{-1}\right]f\left[y_0^{\infty}\right]\right)$, and then carrying out the integration at the end. The latter approach will lead to an implicit formulation of the EMC. We will pursue this in Section 4.2.2.

For now, we will follow the first path, which will lead us to a result for the EMC in an expressive form given by the (logarithmic) ratio of the product of the determinants of the covariances of the joint *pdf*s of the past and future histories and the determinant of the covariance for the whole history. These covariances are infinite-dimensional in principle, but we will see, numerically, that low-dimensional approximations come very close to the asymptotic result. The smallest possible dimension, i.e. if only two time steps are involved, leads to a simple yet meaningful approximation, which will be discussed in more detail below.

For the first term we can use the result for the differential entropy of a Gaussian variable, see e.g. Cover and Thomas (1991),

$$
\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 f\left[y_{-\infty}^{-1},y_0^{\infty}\right]dy_{-\infty}^{\infty} = -\frac{1}{2}\log_2\left((2\pi e)^{p\Delta t_{-\infty}^{\infty}}\right) - \frac{1}{2}\log_2\left(\mathrm{Det}\left(\mathcal{C}_y\right)_{-\infty}^{\infty}\right).
$$

The second term can be computed as follows:

$$
\begin{aligned}
&\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 f\left[y_{-\infty}^{-1}\right]f\left[y_0^{\infty}\right]dy_{-\infty}^{\infty}\\
&=\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2\frac{\mathrm{Exp}\left[-\frac{1}{2}\left(\mathbf{y}_{-\infty}^{-1}\right)^{\mathrm{T}}\left(\left(\mathcal{C}_y\right)_{-\infty}^{-1}\right)^{-1}\mathbf{y}_{-\infty}^{-1}-\frac{1}{2}\left(\mathbf{y}_0^{\infty}\right)^{\mathrm{T}}\left(\left(\mathcal{C}_y\right)_0^{\infty}\right)^{-1}\mathbf{y}_0^{\infty}\right]}{\sqrt{2\pi}^{p\Delta t_{-\infty}^{\infty}}\sqrt{\mathrm{Det}\left(\mathcal{C}_y\right)_{-\infty}^{-1}\mathrm{Det}\left(\mathcal{C}_y\right)_0^{\infty}}}dy_{-\infty}^{\infty}\\
&=-\frac{1}{2}\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\frac{1}{\ln 2}\left(\mathbf{y}_{-\infty}^{\infty}\right)^{\mathrm{T}}\widehat{\mathcal{C}}\,\mathbf{y}_{-\infty}^{\infty}dy_{-\infty}^{\infty}-\frac{1}{2}\log_2\left((2\pi)^{p\Delta t_{-\infty}^{\infty}}\right)\\
&\quad-\frac{1}{2}\log_2\left(\mathrm{Det}\left(\mathcal{C}_y\right)_{-\infty}^{-1}\mathrm{Det}\left(\mathcal{C}_y\right)_0^{\infty}\right),
\end{aligned}
$$

with

$$
\widehat{\mathcal{C}} = \begin{pmatrix} \left(\left(\mathcal{C}_y\right)_{-\infty}^{-1}\right)^{-1} & 0 \\ 0 & \left(\left(\mathcal{C}_y\right)_0^{\infty}\right)^{-1} \end{pmatrix}.
$$

The integral in the first summand of the above equation yields

$$\begin{aligned}\frac{1}{2}\int f\left[y_{-\infty}^{-1}, y_0^{\infty}\right]\frac{1}{\ln 2}\left(\mathbf{y}_{-\infty}^{\infty}\right)^{\mathrm{T}}\widehat{\mathcal{C}}\,\mathbf{y}_{-\infty}^{\infty}dy_{-\infty}^{\infty} &= \frac{1}{2}\frac{1}{\ln 2}\mathrm{Tr}\left(\widehat{\mathcal{C}}\cdot\left(\mathcal{C}_y\right)_{-\infty}^{\infty}\right)\\ &= \frac{1}{2}\frac{1}{\ln 2}p\Delta t_{-\infty}^{\infty}\\ &= \frac{1}{2}\log_2 e^{p\Delta t_{-\infty}^{\infty}},\end{aligned}$$

where we used the fact that $\left(\mathcal{C}_y\right)_{-\infty}^{\infty}$ can be partitioned in a 2×2 block matrix, in which the upper left block equals $\left(\mathcal{C}_y\right)_{-\infty}^{-1}$ and the lower right block equals $\left(\mathcal{C}_y\right)_0^{\infty}$, as can be seen from the block Toeplitz structure of the covariance of the observations. The matrix product $\widehat{\mathcal{C}}\cdot\left(\mathcal{C}_y\right)_{-\infty}^{\infty}$ then has only ones on the diagonal and it is easy to evaluate the trace. Finally, by combining the individual results, we obtain

$$I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] = \frac{1}{2}\log_2\frac{\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{-1}\mathrm{Det}\,\left(\mathcal{C}_y\right)_0^{\infty}}{\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{\infty}}. \tag{291}$$

Note that this result has been obtained in a more general context by de Cock (2002), see Eq. 295. The matrices are infinite dimensional, which makes this result impractical for direct use. However, we found in simulations that for a moderately small number of time steps $\Delta t = t_2 - t_1 + 1$ of either the past or the future (the total number of time steps involved is then $2\Delta t$), the value for the EMC tends to its asymptotic value (see Fig. 4.2).

As we have shown in Section 2.10, the likelihood of the observation sequence $\{y_t\}_{t_1}^{t_2}$ is invariant under an arbitrary invertible transform $\Psi\in\mathbb{R}^{q\times q}$ with $\mathrm{Det}(\Psi) = 1$ transforming the set of parameters as $x_t' = \Psi x_t$, $\pi_0' = \Psi\,\pi_0$, $A_0' = \Psi\,A_0\,\Psi^{-1}$, $C' = \Psi\,C\,\Psi^T$, $\Pi_0' = \Psi\,\Pi_0\,\Psi^T$, $H' = H\,\Psi^{-1}$ and $V' = V$. Therefore, the system matrices can not be identified uniquely.

However, the emergent complexity is invariant under this parameter transform, as easily proved by using the expression for the EMC from Eq. 291:

$$\begin{aligned}C_y' &= I_{\Delta t}\otimes V + (I_{\Delta t}\otimes H')C_x'\left(I_{\Delta t}\otimes H'^T\right)\\ &= I_{\Delta t}\otimes V + \left(I_{\Delta t}\otimes H\Psi^{-1}\right)(I_{\Delta t}\otimes\Psi)C_x\left(I_{\Delta t}\otimes\Psi^T\right)\left(I_{\Delta t}\otimes\Psi^{-T}H^T\right)\\ &= I_{\Delta t}\otimes V + (I_{\Delta t}\cdot I_{\Delta t})\otimes\left(H\Psi^{-1}\Psi\right)C_x(I_{\Delta t}\cdot I_{\Delta t})\otimes\left(\Psi^T\Psi^{-T}H^T\right)\\ &= I_{\Delta t}\otimes V + (I_{\Delta t}\otimes H)C_x\left(I_{\Delta t}\otimes H^T\right)\\ &= C_y.\end{aligned}$$

This general result holds for the covariance of any observation interval, in particular for $\left(C_y\right)_{-\infty}^{-1}$, $\left(C_y\right)_0^{\infty}$, $\left(C_y\right)_{-\infty}^{\infty}$, and, therefore, the EMC remains unchanged.

Surprisingly, the smallest possible value $\Delta t = 1$, i.e. if we consider that just

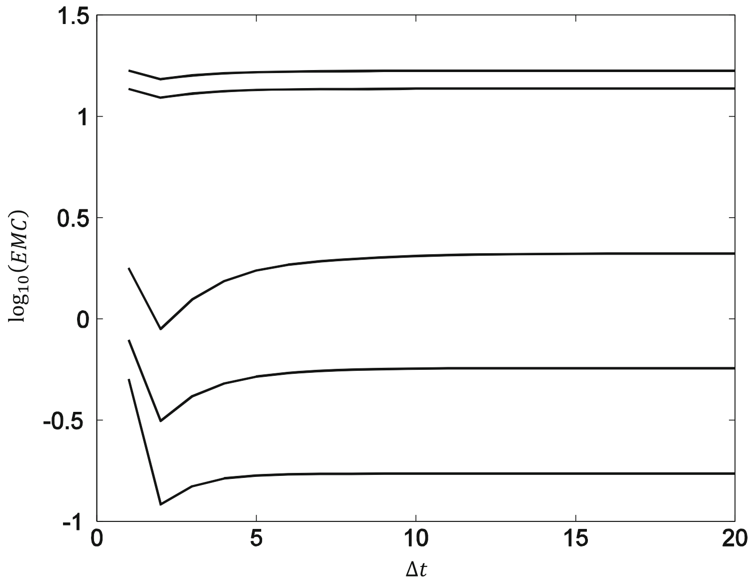

Fig. 4.2 Calculated values of $\log_{10}(EMC)$ for five different, randomly chosen system matrices A_0 and H and varying number of time steps Δt

$$\mathrm{EMC}^{(1)} = \frac{1}{2}\log_2 \frac{\mathrm{Det}\ \left(\mathcal{C}_y\right)_{-1}^{-1} \mathrm{Det}\ \left(\mathcal{C}_y\right)_0^0}{\mathrm{Det}\ \left(\mathcal{C}_y\right)_{-1}^0}$$

leads to a result that is very close to the asymptotic value (see Fig. 4.2). In this case, which we can call a first-order approximation, a very simple closed-form expression for the EMC can be derived. The covariances for a single time step are given by

$$\begin{aligned}\left(\mathcal{C}_y\right)_{-1}^{-1} = \left(\mathcal{C}_y\right)_0^0 &= I_1 \otimes V + (I_1 \otimes H)\mathcal{C}_x\left(I_1 \otimes H^{\mathrm{T}}\right)\\ &= V + H\mathcal{C}_x H^{\mathrm{T}}\\ &= V + H\Sigma H^{\mathrm{T}}.\end{aligned}$$

For two time steps we have

$$\begin{aligned}\left(\mathcal{C}_y\right)_{-1}^0 &= I_2 \otimes V + (I_2 \otimes H)\mathcal{C}_x\left(I_2 \otimes H^{\mathrm{T}}\right)\\ &= \begin{pmatrix} V & 0 \\ 0 & V \end{pmatrix} + \begin{pmatrix} H & 0 \\ 0 & H \end{pmatrix}\begin{pmatrix} \Sigma & \Sigma A_0^{\mathrm{T}} \\ A_0\Sigma & V \end{pmatrix}\begin{pmatrix} H^{\mathrm{T}} & 0 \\ 0 & H^{\mathrm{T}} \end{pmatrix}\\ &= \begin{pmatrix} V + H\Sigma H^{\mathrm{T}} & H\Sigma A_0^{\mathrm{T}} H^{\mathrm{T}} \\ HA_0\Sigma H^{\mathrm{T}} & V + H\Sigma H^{\mathrm{T}} \end{pmatrix}.\end{aligned}$$

The determinant of covariance of the two time steps can be simplified using a formula for the determinant of block-matrices,

$$\mathrm{Det}\left(\mathcal{C}_y\right)_{-1}^{0} = \mathrm{Det}\left(V + H\Sigma H^{\mathrm{T}}\right)\mathrm{Det}\left(V + H\Sigma H^{\mathrm{T}} - H\Sigma A_0^{\mathrm{T}} H^{\mathrm{T}}\left(V + H\Sigma H^{\mathrm{T}}\right)^{-1} HA_0\Sigma H^{\mathrm{T}}\right),$$

and we obtain the following first-order approximation for the EMC:

$$\begin{aligned}
\mathrm{EMC}^{(1)} &= \frac{1}{2}\log_2 \frac{\mathrm{Det}\left(V + H\Sigma H^{\mathrm{T}}\right)}{\mathrm{Det}\left(V + H\Sigma H^{\mathrm{T}} - H\Sigma A_0^{\mathrm{T}} H^{\mathrm{T}}\left(V + H\Sigma H^{\mathrm{T}}\right)^{-1} HA_0\Sigma H^{\mathrm{T}}\right)} \\
&= -\frac{1}{2}\log_2 \mathrm{Det}\left(I_p - H\Sigma A_0^{\mathrm{T}} H^{\mathrm{T}}\left(V + H\Sigma H^{\mathrm{T}}\right)^{-1} HA_0\Sigma H^{\mathrm{T}}\left(V + H\Sigma H^{\mathrm{T}}\right)^{-1}\right).
\end{aligned} \tag{292}$$

Interestingly, this approximate result can be derived if one starts from the assumption that the Markov property holds in steady state for the observable process, and we have

$$\begin{aligned}
f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] &= f[y_{-\infty}] f\left[y_{-\infty+1}\middle|y_{-\infty}\right] \ldots f[y_{-1}|y_{-2}] f[y_0|y_{-1}] f[y_1|y_0] \ldots f[y_{\infty}|y_{\infty-1}] \\
f\left[y_{-\infty}^{-1}\right] &= f[y_{-\infty}] f\left[y_{-\infty+1}\middle|y_{-\infty}\right] \ldots f[y_{-1}|y_{-2}] \\
f\left[y_0^{\infty}\right] &= f[y_0] f[y_1|y_0] \ldots f[y_{\infty}|y_{\infty-1}].
\end{aligned}$$

Hence, the expression for the EMC reduces to

$$\begin{aligned}
I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= \int_{\mathbb{Y}^p} \cdots \int_{\mathbb{Y}^p} f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] \log_2 \frac{f[y_0|y_{-1}]}{f[y_0]} dy_{-\infty}^{-1} dy_0^{\infty} \\
&= \int_{\mathbb{Y}^p} \cdots \int_{\mathbb{Y}^p} f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] \log_2 f[y_0|y_{-1}] dy_{-\infty}^{-1} dy_0^{\infty} \\
&\quad - \int_{\mathbb{Y}^p} \cdots \int_{\mathbb{Y}^p} f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] \log_2 f[y_0] dy_{-\infty}^{-1} dy_0^{\infty} \\
&= \int_{\mathbb{Y}^p}\int_{\mathbb{Y}^p} \log_2 f[y_0|y_{-1}] dy_{-1} dy_0 \left(\int_{\mathbb{Y}^p} \cdots \int_{\mathbb{Y}^p} f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] dy_{-\infty} \ldots dy_{-2} dy_1 \ldots dy_{\infty}\right) \\
&\quad - \int_{\mathbb{Y}^p} \log_2 f[y_0] dy_0 \left(\int_{\mathbb{Y}^p} \cdots \int_{\mathbb{Y}^p} f\left[y_{-\infty}^{-1}, y_0^{\infty}\right] dy_{-\infty} \ldots dy_{-1} dy_1 \ldots dy_{\infty}\right).
\end{aligned}$$

Exploiting the relations for the marginal probability densities, we obtain:

$$\begin{aligned}
I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= \int_{\mathbb{Y}^p} f[y_{-1},\ y_0] \log_2 f[y_0|y_{-1}] dy_{-1} dy_0 - \int_{\mathbb{Y}^p} f[y_0] \log_2 f[y_0] dy_0 \\
&= \int_{\mathbb{Y}^p} f[y_{-1}] dy_{-1} \int_{\mathbb{Y}^p} f[y_0|y_{-1}] \log_2 f[y_0|y_{-1}] dy_0 - \int_{\mathbb{Y}^p} f[y_0] \log_2 f[y_0] dy_0.
\end{aligned}$$

The probability density for the observable variable Y_t can then be expressed as:

$$\begin{aligned} f[y_t] &= \int_{\mathbb{X}^p} f[y_t, x_t] dx_t \\ &= \int_{\mathbb{X}^p} f[y_t|x_t] f[x_t] dx_t \\ &= \int_{\mathbb{X}^p} \mathcal{N}(y_t; Hx_t, V)\mathcal{N}(x_t; \mu, \Sigma) dx_t. \end{aligned}$$

In order to solve the above integral, it is useful to apply the following transformation formula for normal distributions:

$$\mathcal{N}(y; Hx, V)\mathcal{N}(x; \mu, \Sigma) = \mathcal{N}(y; H\mu, S)\mathcal{N}\left(x; \mu + W(y - H\mu), \Sigma - WSW^{\mathrm{T}}\right)$$

with

$$S = H\Sigma H^{\mathrm{T}} + V \ \text{ and } \ W = \Sigma H^{\mathrm{T}} S^{-1}.$$

Hence, we obtain:

$$f[y_t] = \mathcal{N}\left(y_t; H\mu, H\Sigma H^{\mathrm{T}} + V\right).$$

For the calculation of $f\left[y_0 \middle| y_{-1}\right] = f[y_{-1}, y_0]/f[y_{-1}]$ we insert the hidden states x_{-1} and x_0 and exploit the Markov property

$$\begin{aligned} f\left[y_0 \middle| y_{-1}\right] &= \int_{\mathbb{X}^q}\int_{\mathbb{X}^q} \frac{f[y_{-1}, y_0, x_{-1}, x_0]}{f[y_{-1}]} dx_{-1} dx_0 \\ &= \int_{\mathbb{X}^q}\int_{\mathbb{X}^q} f\left[x_{-1} \middle| y_{-1}\right] f[x_0|x_{-1}] f[y_0|x_0] dx_{-1} dx_0. \end{aligned}$$

Because of Bayes theorem

$$f\left[x_{-1} \middle| y_{-1}\right] = \frac{f[y_{-1}|x_{-1}] f[x_{-1}]}{f[y_{-1}]},$$

we find

$$\begin{aligned} & f\left[y_0 \middle| y_{-1}\right] \\ &= \frac{1}{f[y_{-1}]}\int_{\mathbb{X}^q}\int_{\mathbb{X}^q} f[x_{-1}] f[y_{-1}|x_{-1}] f[x_0|x_{-1}] f[y_0|x_0] dx_{-1} dx_0 \\ &= \frac{1}{f[y_{-1}]}\int_{\mathbb{R}^q}\int_{\mathbb{R}^q} \mathcal{N}(x_{-1};\mu,\Sigma)\mathcal{N}(y_{-1};Hx_{-1},V)\mathcal{N}(x_0;A_0x_{-1},C)\mathcal{N}(y_0;Hx_0,V) dx_{-1} dx_0. \end{aligned}$$

First, we transform the first two Gaussians as:

$$\mathcal{N}(x_{-1};\mu,\Sigma)\mathcal{N}(y_{-1};Hx_{-1},V) = \mathcal{N}\left(y_{-1};H\mu,H\Sigma H^{\mathrm{T}}+V\right)$$
$$\times\mathcal{N}\left(x_{-1};\mu+W(y_{-1}-H\mu),\Sigma-WSW^{\mathrm{T}}\right),$$

with

$$S = H\Sigma H^{\mathrm{T}} + V \text{ and } W = \Sigma H^{\mathrm{T}} S^{-1}.$$

The first Gaussian on the right hand side cancels $f[y_{-1}]$. The second Gaussian on the right hand side together with the third Gaussian $\mathcal{N}(x_0;A_0x_{-1},C)$ from the previous expression for $f\left[y_0|y_{-1}\right]$ yields:

$$\begin{aligned}&\mathcal{N}(x_0;A_0x_{-1},C)\mathcal{N}\left(x_{-1};\mu+W(y_{-1}-H\mu),\Sigma-WSW^{\mathrm{T}}\right)\\&=\mathcal{N}\left(x_0;A_0(\mu+W(y_{-1}-H\mu)),\ A_0\left(\Sigma-WSW^{\mathrm{T}}\right)A_0^{\mathrm{T}}+C\right)\mathcal{N}(x_{-1};\bar{x}_{-1},C')\end{aligned}$$

with some inconsequential mean $\bar{x}_{-1}$ and covariance C'. After integration with respect to x_{-1} we have:

$$f\left[y_0|y_{-1}\right]=\int_{\mathbb{R}^q}\mathcal{N}\left(x_0;A_0(\mu+W(y_{-1}-H\mu)),A_0\left(\Sigma-WSW^{\mathrm{T}}\right)A_0^{\mathrm{T}}+C\right)\mathcal{N}(y_0;Hx_0,V)dx_0.$$

Again, by transforming the two Gaussians we can carry out easily the integration with respect to x_0 and obtain:

$$f\left[y_0|y_{-1}\right]=\mathcal{N}\left(y_0;HA_0(\mu+W(y_{-1}-H\mu)),H\left(A_0\left(\Sigma-WSW^{T}\right)A_0^{\mathrm{T}}+C\right)H^{\mathrm{T}}+V\right).$$

Using the fact that the differential entropy of a multivariate Gaussian distribution $\mathcal{N}(x;\mu,C)$ is given by

$$-\int_{\mathbb{R}^q}\mathcal{N}(x;\mu,C)\log_2\mathcal{N}(x;\mu,C)dx=\frac{1}{2}\log_2(2\pi e)^p\mathrm{Det}[C],$$

we arrive at the known first-order approximation from Eq. 292 for the EMC:

$$\mathrm{EMC}^{(1)}=\frac{1}{2}\log_2\mathrm{Det}\left[H\Sigma H^{\mathrm{T}}+V\right]-\frac{1}{2}\log_2\mathrm{Det}[D],$$

with

$$D=V+H\Sigma H^{\mathrm{T}}-H\Sigma A_0^{\mathrm{T}}H^{\mathrm{T}}\left(V+H\Sigma H^{\mathrm{T}}\right)^{-1}HA_0\Sigma H^{\mathrm{T}}\ .$$

It is evident that in the case of $H=I$ and $V=\{0\}I_q$, we obtain the same result as we did for the VAR(1) model (see Eq. 246).

For small covariance V, i.e. if the eigenvalues of $\left(H\Sigma H^{\mathrm{T}}\right)^{-1}V$ lie inside the unit circle, we can expand

$$\begin{aligned}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}} &= \left(H\Sigma H^{\mathrm{T}}\right)^{-1}\left(I_p+\left(H\Sigma H^{\mathrm{T}}\right)^{-1}V\right)^{-1}\\ &\approx \left(H\Sigma H^{\mathrm{T}}\right)^{-1}\left(I_p-\left(H\Sigma H^{\mathrm{T}}\right)^{-1}V\right)\end{aligned}$$

and arrive at an approximate expression for D:

$$D=H\left(A_0H^{-1}H^{-\mathrm{T}}\Sigma^{-1}H^{-1}VH\Sigma A_0+C\right)H^{\mathrm{T}}+V.$$

Assuming furthermore $V=\left\{\sigma_v^2\right\}I_p$, we obtain:

$$D=H\left(\left\{\sigma_v^2\right\}A_0H^{-1}H^{-\mathrm{T}}A_0+C\right)H^{\mathrm{T}}+\left\{\sigma_v^2\right\}I_p.$$

Following the procedure from Section 4.1.2, we can also express the first-order approximation for the EMC as the signal-to-noise ratio:

$$\begin{aligned}\mathrm{EMC}^{(1)} &= -\frac{1}{2}\log_2\mathrm{Det}\left[I_p-HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}\right]\\ &= \frac{1}{2}\log_2\left(\mathrm{Det}\left[I_p-HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}\right]\right)^{-1}\\ &= \frac{1}{2}\log_2\mathrm{Det}\left[\left(I_p-HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}\right)^{-1}\right]\\ &= \frac{1}{2}\log_2\mathrm{Det}\left[\sum_{k=0}^{\infty}\left(HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}\right)^k\right]\\ &= \frac{1}{2}\log_2\mathrm{Det}\left[I_p+\sum_{k=1}^{\infty}\left(HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}\right)^k\right].\end{aligned} \tag{293}$$

The above derivation is based on the von Neumann series generated by the operator $HA_0\Sigma H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-\mathrm{T}}H\Sigma^{\mathrm{T}}A_0^{\mathrm{T}}H^{\mathrm{T}}\left(H\Sigma H^{\mathrm{T}}+V\right)^{-1}$. The von Neumann series generalizes the geometric series (cf. Section 2.1). The infinite sum represents the signal-to-noise ratio.

The closed-form solution from Eq. 286 also allows us to develop homologous vector autoregression models for linear dynamical systems. For $t\to\infty$ these models generate stochastic processes with equivalent effective measure complexity, but the state variables are completely observable. In this sense, the homologous models reveal all correlations and dynamical dependency structures during the observation time and do not possess any kind of crypticity (Ellison et al. 2009). To make this possible we usually have to use a higher dimensionality $p>q$. We start by focusing on homologous VAR(1) models with dynamical operator A_0^h and covariance matrix C^h that are defined over a p-dimensional space $\mathbb{R}^p$ of observable states X_t^h:

$$X_t^h = A_0^h X_{t-1}^h + \varepsilon_t^h \quad t = 1, \ldots, T\,,$$

with

$$\varepsilon_t^h \sim \mathcal{N}\left(0_p, C^h\right).$$

Assuming that the performance fluctuations represented by the homologous model are isotropic and temporally uncorrelated, i.e. $\varepsilon_t^h \sim \mathcal{N}\left(0_p, \left\{\sigma_v^2\right\} I_p\right)$ and $E\left[\varepsilon_t^h \left(\varepsilon_s^h\right)^{\mathrm{T}}\right] = C^h \delta_{ts}$, we can construct a dynamical operator A_0^h representing a large variety of cooperative relationships. The preferred structure of relationships must be determined in the specific application context of complexity evaluation. According to the analysis in Section 4.1.1 only two constraints must be satisfied: (1) A_0^h must be diagonalizable and (2) for the weighted sum of eigenvalues $\lambda_i(A_0^h)$, it must hold that (cf. Eq. 251):

$$\begin{aligned} -\frac{1}{2}\sum_{i=1}^{p} \log_2\left(1 - \lambda_i\left(A_0^h\right)^2\right) &= \frac{1}{2}\log_2 \frac{\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{-1} \mathrm{Det}\,\left(\mathcal{C}_y\right)_0^{\infty}}{\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{\infty}} \\ &= \frac{1}{2}\left(\log_2 \mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{-1} \mathrm{Det}\,\left(\mathcal{C}_y\right)_0^{\infty} - \log_2 \mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{\infty}\right). \end{aligned} \tag{294}$$

It is evident that the most simple homologous model can be constructed by setting the autonomous task processing rates as diagonal elements of A_0^h to the same rate a, i.e. $A_0^h = \mathrm{Diag}[a, \ldots, a]$. For this structurally non-informative model, the corresponding stationary stochastic process communicates the same amount of information from the infinite past to the infinite future, if

$$a = \sqrt{1 - 2^{-\frac{1}{p}\left(\log_2 \mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{-1} \mathrm{Det}\,\left(\mathcal{C}_y\right)_0^{\infty} - \log_2 \mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{\infty}\right)}}.$$

The above equation also holds for homologous models with non-isotropic fluctuations, because all tasks are processed at the same time scale.

Finally, we can develop a homologous model that is defined over a one-dimensional state space. This model is termed an auto-regressive moving average (ARMA) model and is characterized by the following linear difference equation (see e.g. Puri 2010):

$$Y_t = \sum_{i=1}^{p} a_i Y_{t-i} + \sum_{j=1}^{q} b_i U_{t-j}.$$

The input of the model is Gaussian white noise with variance $\sigma^2 = 1$, i.e. $U \sim \mathcal{N}(0, 1)$. This model is notated ARMA(p, q) in the literature (note that in this notation the variable q does not denote the dimensionality of the observation vectors Y_t; it denotes the number of inputs U_{t-j} driving the process). It is evident that an ARMA(p, q) model can be rewritten as either a VAR(p) model of order p (Section 2.4) or an LDS$(p, 1)$ model (Section 2.9) (see e.g. de Cock 2002). It is not difficult to show that for a stable and strictly minimum phase ARMA(p, q) model the effective measure complexity is given by

$$\mathrm{EMC} = \frac{1}{2}\log_2 \frac{\prod_{i,\,j=1}^{p,q} \left|1-\alpha_i\overline{\beta}_j\right|}{\prod_{i,\,j=1}^{p} \left|1-\alpha_i\overline{\alpha}_j\right| \prod_{i,\,j=1}^{q} \left|1-\beta_i\overline{\beta}_j\right|}$$
$$= \frac{1}{2}\left(\sum_{i,\,j}^{p,q}\log_2\left|1-\alpha_i\overline{\beta}_j\right| - \sum_{i,\,j}^{p}\log_2\left|1-\alpha_i\overline{\alpha}_j\right| + \sum_{i,\,j}^{q}\log_2\left|1-\beta_i\overline{\beta}_j\right|\right),$$

where the variables $\alpha_1, \ldots, \alpha_p$ denote the roots of the polynomial $a(z) = z^p + a_1 z^{p-1} + \ldots + a_p$ and $\beta_1, \ldots, \beta_q$ the roots of the polynomial $b(z) = z^q + b_1 z^{q-1} + \ldots + b_q$ (see e.g. de Cock 2002). These polynomials are the results of the z-transform of the difference equation of the ARMA(p, q) model. The well-known transfer function $H(z)$ from control theory is the quotient of these polynomials. Since the polynomials are real, the roots are all real or come in conjugate pairs. Hence, for the poles $\alpha_1, \ldots, \alpha_p$ and the zeros $\beta_1, \ldots, \beta_q$ of the transfer function $H(z)$ of the homologous ARMA(p, q) model, it must hold that

$$\frac{1}{2}\left(\sum_{i,\,j}^{p,q}\log_2\left|1-\alpha_i\overline{\beta}_j\right| - \sum_{i,\,j}^{p}\log_2\left|1-\alpha_i\overline{\alpha}_j\right| + \sum_{i,\,j}^{q}\log_2\left|1-\beta_i\overline{\beta}_j\right|\right)$$
$$= \frac{1}{2}\left(\log_2\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{-1}\mathrm{Det}\,\left(\mathcal{C}_y\right)_{0}^{\infty} - \log_2\mathrm{Det}\,\left(\mathcal{C}_y\right)_{-\infty}^{\infty}\right).$$

4.2.2 Implicit Formulation

Interestingly, the sophisticated closed-form solution of EMC from Eq. 286 that was obtained through the evaluation of the infinite-dimensional integral of the continuous-type mutual information (Eq. 286) can also be written in a structurally rich implicit form. This form is based on the seminal work of de Cock (2002). The implicit form is especially easy to interpret because its independent parameters can be derived from solutions of fundamental equations. In order to derive the implicit form of de Cock (2002) we work with the “forward innovation model” from Section 2.9 (Eqs. 164 and 165):

$$X_{t+1}^f = A_0 X_t^f + K\eta_t$$
$$Y_t = HX_t^f + \eta_t.$$

According to de Cock (2002), the effective measure complexity can be expressed as

$$\begin{aligned}\mathrm{EMC} &= I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] \\ &= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_q - \Sigma_f\left(G_z^{-1} + \Sigma^f\right)^{-1}\right].\end{aligned} \tag{295}$$

The covariance matrix Σ^f is the solution of the Lyapunov equation (cf. Eq. 167)

$$\Sigma^f = A_0 \Sigma^f A_0^{\mathrm{T}} + KSK^{\mathrm{T}}.$$

In the above Lyapunov equation

$$K = \left(G^f - A_0\Sigma^f H^{\mathrm{T}}\right)\left(C_{YY}(0) - H\Sigma^f H^{\mathrm{T}}\right)^{-1}$$

is the Kalman gain (Eq. 169) and

$$S = C_{YY}(0) - H\Sigma^f H^{\mathrm{T}}.$$

is the covariance $S_{t+1|t}$ (Eq. 168) of the single-source performance fluctuations η_t for $t \to \infty$. Hence, we have the following algebraic Riccati equation for Σ^f (van Overschee and de Moor 1996):

$$\Sigma^f = A_0\Sigma^f A_0^{\mathrm{T}} + \left(G^f - A_0\Sigma^f H^{\mathrm{T}}\right)\left(C_{YY}(0) - H\Sigma^f H^{\mathrm{T}}\right)^{-1}\left(\left(G^f\right)^{\mathrm{T}} - H\Sigma^f H^{\mathrm{T}}\right). \tag{296}$$

The additional covariance matrix G_z from Eq. 295 satisfies the Lyapunov equation

$$G_z = (A_0 - KH)^{\mathrm{T}} G_z (A_0 - KH) + H^{\mathrm{T}} S^{-1} H. \tag{297}$$

An important finding of de Cock (2002) is that the inverse aggregated covariance matrix $\left(G_z^{-1} + \Sigma^f\right)^{-1}$ is the solution of another Lyapunov equation

$$\begin{aligned}\overline{\Sigma}^b &= \overline{A}_0 \overline{\Sigma}^b \overline{A}_0^{\mathrm{T}} + \overline{K}\,\overline{S}\,\overline{K}^{\mathrm{T}} \\ &= A_0^{\mathrm{T}} \overline{\Sigma}^b A_0 + \overline{K}\,\overline{S}\,\overline{K}^{\mathrm{T}},\end{aligned}$$

which is related to the backward innovation representation of the corresponding backward model (Eqs. 178 and 179):

$$\begin{aligned}\overline{X}_{t-1}^b &= \overline{A}_0 \overline{X}_t^b + \overline{K}\,\overline{\eta}_t \\ Y_t &= \overline{H}\,\overline{X}_t^b + \overline{\eta}_t.\end{aligned}$$

Substituting the Kalman gain $\overline{K}$ (Eq. 181) and the fluctuations covariance $\overline{S} = E\left[\overline{\eta}_t \overline{\eta}_t^{\mathrm{T}}\right]$ (Eq. 182) in the Lyapunov equation for the backward innovation representation leads to the following algebraic Riccati equation for the backward state covariance matrix:

$$\begin{aligned}\overline{\Sigma}^b &= A_0^{\mathrm{T}}\overline{\Sigma}^b A_0 + \left(H^{\mathrm{T}} - A_0^{\mathrm{T}}\overline{\Sigma}^b G\right)\left(\left(H^{\mathrm{T}} - A_0^{\mathrm{T}}\overline{\Sigma}^b G\right)\left(C_{YY}(0) - G^{\mathrm{T}}\overline{\Sigma}^b G\right)^{-1}\right)^{\mathrm{T}} \\ &= A_0^{\mathrm{T}}\overline{\Sigma}^b A_0 + \left(H^{\mathrm{T}} - A_0^{\mathrm{T}}\overline{\Sigma}^b G\right)\left(C_{YY}(0) - G^{\mathrm{T}}\overline{\Sigma}^b G\right)^{-\mathrm{T}}\left(H^{\mathrm{T}} - A_0^{\mathrm{T}}\overline{\Sigma}^b G\right)^{\mathrm{T}} \\ &= A_0^{\mathrm{T}}\overline{\Sigma}^b A_0 + \left(H^{\mathrm{T}} - A_0^{\mathrm{T}}\overline{\Sigma}^b G\right)\left(C_{YY}(0) - G^{\mathrm{T}}\overline{\Sigma}^b G\right)^{-1}\left(H - G^{\mathrm{T}}\overline{\Sigma}^b A_0\right).\end{aligned} \tag{298}$$

Hence, the most intuitive solution is obtained (de Cock 2002):

$$\begin{aligned} I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_q - \Sigma^f\left(G_z^{-1} + \Sigma^f\right)^{-1}\right] \\ &= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_q - \Sigma^f\overline{\Sigma}^b\right]. \end{aligned} \tag{299}$$

According to Sylvester's determinant theorem, this solution can equivalently be expressed based on the signal-to-noise ratio $\mathrm{SNR} = G_z\left(\left(\Sigma^f\right)^{-1}\right)^{-1}$, and we have:

$$\begin{aligned} I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_q - \overline{\Sigma}^b\Sigma^f\right] \\ &= \frac{1}{2}\log_2 \mathrm{Det}\left[I_q + G_z\Sigma^f\right] \\ &= \frac{1}{2}\log_2 \mathrm{Det}\left[I_q + \Sigma^f G_z\right]. \end{aligned}$$

The standard numerical approach to solve the forward Riccati Eq. 296 is to solve the generalized eigenvalue problem

$$\begin{pmatrix} A_0^{\mathrm{T}} - H^{\mathrm{T}}(C_{YY}(0))^{-1}G^f & 0 \\ -G^f(C_{YY}(0))^{-1}\left(G^f\right)^{\mathrm{T}} & I_q \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} I_q & -H^{\mathrm{T}}(C_{YY}(0))^{-1}H \\ 0 & A_0 - G^f(C_{YY}(0))^{-1}H \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}\Lambda$$

and compute the covariance matrix Σ^f as

$$\Sigma^f = W_2 W_1^{-1},$$

see, e.g. van Overschee and de Moor (1996). The complementary backward Riccati Eq. 298 can be tackled by solving

$$\begin{pmatrix} A_0 - G(C_{YY}(0))^{-1}H & 0 \\ -H^{\mathrm{T}}(C_{YY}(0))^{-1}H & I_q \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} I_q & -G(C_{YY}(0))^{-1}H^{\mathrm{T}} \\ 0 & A_0^{\mathrm{T}} - H^{\mathrm{T}}(C_{YY}(0))^{-1}\left(G^f\right)^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} W_1 \\ W_2 \end{pmatrix}\Lambda$$

and computing the covariance matrix $\overline{\Sigma}_b$ as

$$\overline{\Sigma}^b = W_2 W_1^{-1}.$$

It is evident that the same numerical function can be used in the preferred programming language to solve the above generalized eigenvalue problems. This function must be called for the forward Riccati equation with the argument $(A_0, H, G^f, C_{YY}(0))$, whilst for the backward Riccati equation the argument must be $(A_0^{\mathrm{T}}, (G^f)^{\mathrm{T}}, H^{\mathrm{T}}, C_{YY}(0))$.

Similar to the canonical correlation analysis of the basic VAR(1) process in Section 4.1.3, we can diagonalize the forward and backward state covariance matrices obtained by solving the algebraic Riccati Eqs. 296 and 298 simultaneously and bring them in a form called "stochastic balanced realization" (Desai and Pal 1984). A stochastic balanced representation is an innovations representation with state covariance matrix equal to the canonical correlation coefficient matrix for the sequence of observations. Let the eigendecomposition (cf. Eq. 22) of the product of the state covariance matrices $\Sigma^f \overline{\Sigma}^b$ be given by the representation

$$\Sigma^f \overline{\Sigma}^b = M \Lambda_M^2 M^{-1}$$
$$\Lambda_M^2 = \mathrm{Diag}\left[\lambda_i\left(\Sigma^f \overline{\Sigma}^b\right)\right] \quad 1 \leq i \leq q,$$

where the eigenvector matrix M is picked as

$$M = U_{\overline{\Sigma}^b} \Lambda_{\overline{\Sigma}^b}^{-1/2} U_{\Sigma^f} \Lambda_{\mathrm{M}}^{1/2}.$$

The matrices $U_{\overline{\Sigma}^b}$ and $\Lambda_{\overline{\Sigma}^b}$ can be specified by the eigendecomposition of $\overline{\Sigma}_b$ as

$$U_{\overline{\Sigma}^b} \Lambda_{\overline{\Sigma}^b} U_{\overline{\Sigma}^b}^{-1} = \overline{\Sigma}^b,$$

and for U_{Σ^f} it holds that

$$U_{\Sigma^f} \Lambda_M^2 U_{\Sigma^f}^{-1} = \Lambda_{\overline{\Sigma}^b}^{1/2} U_{\overline{\Sigma}^b}^{-1} \Sigma^f \Lambda_{\overline{\Sigma}^b}^{1/2}.$$

Furthermore, let the forward state that is subject to the simultaneous diagonalization of the state covariance matrices be

$$X_t^d = T X_t^f$$

and the corresponding backward state be

$$\overline{X}_t^d = T^{-1} \overline{X}_t^b$$

with the coefficient of the similarity transformation

$$T = M^{\mathrm{T}},$$

then in steady state it holds for the expectations (Desai and Pal 1984) that:

$$E\left[X_t^d \left(X_t^d\right)^{\mathrm{T}}\right] = \Lambda_M = E\left[\overline{X}_t^d \left(\overline{X}_t^d\right)^{\mathrm{T}}\right].$$

Hence, the stochastic balanced representation allows us to make the dependency of the effective measure complexity on the eigenvalues of the product of the state covariance matrices $\Sigma^f \overline{\Sigma}^b$ explicit:

$$\begin{aligned}
I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_p - \Sigma^f \overline{\Sigma}^b\right] \\
&= -\frac{1}{2}\log_2 \mathrm{Det}\left[I_q - \Lambda_M^2\right] \\
&= -\frac{1}{2}\log_2 \prod_{i=1}^{q}\left(1 - \lambda_i\left(\Sigma^f \overline{\Sigma}^b\right)\right) \\
&= -\frac{1}{2}\log_2 \prod_{i=1}^{q}\left(1 - {\rho_i}^2\right) \\
&= -\frac{1}{2}\sum_{i=1}^{q}\log_2\left(1 - {\rho_i}^2\right).
\end{aligned} \tag{300}$$

In the last line of the above equation the ρ_i's represent the canonical correlations, which were already introduced in Section 4.1.3 (cf. Eq. 265) to analyze emergent complexity based on a reduced number of independent parameters. In other words, the eigenvalues of $\Sigma^f \overline{\Sigma}^b$ are simply the squares of the canonical correlation coefficients between the canonical variates. However, it is important to note that in contrast to Section 4.1.3 the infinite random sequences representing the past $X_{past} = \left(X_{-\infty}^{\mathrm{T}} \;\cdots\; X_{-2}^{\mathrm{T}} \;\; X_{-1}^{\mathrm{T}}\right)^{\mathrm{T}}$ and future $X_{fut} = \left(X_0^{\mathrm{T}} \;\; X_1^{\mathrm{T}} \;\cdots\; X_{\infty}^{\mathrm{T}}\right)^{\mathrm{T}}$ histories of the hidden state process are not the subject of the canonical correlation analysis, but rather the canonical correlations between the pair $\left(\left(Y_{-\infty}^{\mathrm{T}} \;\cdots\; Y_{-2}^{\mathrm{T}} \;\; Y_{-1}^{\mathrm{T}}\right)^{\mathrm{T}}, \left(Y_0^{\mathrm{T}} \;\; Y_1^{\mathrm{T}} \;\cdots\; Y_{\infty}^{\mathrm{T}}\right)^{\mathrm{T}}\right)$ of past and future histories of the observation process $\{Y_t\}$ are considered to evaluate complexity explicitly. Due to the potentially higher dimensionality of the state space of the hidden state process $(q > p)$, all q complexity-shaping summands $\log_2(1 - {\rho_i}^2)$ that can give rise to correlations between observations of the project state must therefore be considered. The reduced dimension of the observation process is usually not sufficient, because apart from organizationally retarded cases not only the p but also the q leading canonical correlations are non-zero. The observation process is not necessarily Markovian and therefore the amount of information that the past provides about the future usually cannot be "stored" in the p-dimensional present state. However, because of strict-sense stationarity of the state process, all ρ_i's are less than one. The canonical correlations ρ_i's should not be confused with the ordinary correlations ρ_{ij} and ρ'_{ij}, which were introduced in Chapter 2.

As an alternative to the use of the stochastic balanced representation of Desai and Pal (1984), a minimum phase balancing based on the scheme of McGinnie (1994) could be carried out. The minimum phase balancing scheme allows us to find a forward innovation form of the LDS model in which the state covariance matrix Σ^f (Eq. 296) and the covariance matrix G_z (Eq. 297) are equal and diagonal. Let

$$\Lambda_P = \mathrm{Diag}[\sigma_i] \quad 1 \le i \le q$$

be this diagonal matrix and σ_i the minimum phase singular values of the dynamical system. Under these circumstances, we simply have

$$I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] = -\frac{1}{2}\sum_{i=1}^{q} \log_2\left(1 - \sigma_i{}^2\right). \tag{301}$$

A structurally different implicit formulation for the EMC can be obained when we compute the integral in formula 286 directly. Plugging the results for the joint *pdf*s of the past, the future and the whole observation sequence into the general expression for the EMC from Eq. 286, the ratio of the whole *pdf* to the ones of the past and the future histories is given by:

$$\begin{aligned}
\frac{f\left[y_{-\infty}^{\infty}\right]}{f\left[y_{-\infty}^{-1}\right] f\left[y_0^{\infty}\right]} &= \frac{c_{y_{-\infty}^{\infty}}}{c_{y_{-\infty}^{-1}}\, c_{y_0^{\infty}}} \cdot \frac{\sqrt{\mathrm{Det}\, \mathcal{C}_{-\infty}^{-1}}\sqrt{\mathrm{Det}\, \mathcal{C}_0^{\infty}}}{\sqrt{\mathrm{Det}\, \mathcal{C}_{-\infty}^{\infty}}} \\
&\cdot \mathrm{Exp}\left[\frac{1}{2}\left(\mathbf{b}_{-\infty}^{\infty}\right)^{\mathrm{T}}\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}\mathbf{b}_{-\infty}^{\infty} - \frac{1}{2}\left(\mathbf{b}_{-\infty}^{-1}\right)^{\mathrm{T}}\left(\mathcal{C}_{-\infty}^{-1}\right)^{-1}\mathbf{b}_{-\infty}^{-1} - \frac{1}{2}\left(\mathbf{b}_0^{\infty}\right)^{\mathrm{T}}\left(\mathcal{C}_0^{\infty}\right)^{-1}\mathbf{b}_0^{\infty}\right] \\
&= c_1 \cdot \mathrm{Exp}\left[\frac{1}{2}\left(\mathbf{y}_{-\infty}^{\infty}\right)^{\mathrm{T}}\mathcal{B}\mathbf{y}_{-\infty}^{\infty}\right].
\end{aligned} \tag{302}$$

The constant c_1 is defined accordingly. As we can write

$$\mathbf{b}_{t_1}^{t_2} = \left(I \otimes H^{\mathrm{T}}V^{-1}\right)\mathbf{y}_{t_1}^{t_2},$$

we defined the covariance matrix

$$\mathcal{B} = \left(I \otimes V^{-1}H\right)\left(\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1} - \begin{bmatrix} \left(\mathcal{C}_{-\infty}^{-1}\right)^{-1} & 0 \\ 0 & \left(\mathcal{C}_0^{\infty}\right)^{-1} \end{bmatrix}\right)\left(I \otimes H^{\mathrm{T}}V^{-1}\right).$$

Inserting Eq. 302 into the general Eq. 286 leads to

$$\begin{aligned} I\left[Y_{-\infty}^{-1};Y_0^{\infty}\right] &= \int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 c_1 \mathrm{Exp}\left[\frac{1}{2}\left(\mathbf{y}_{-\infty}^{\infty}\right)^{\mathrm{T}}\mathcal{B}\mathbf{y}_{-\infty}^{\infty}\right]dy_{-\infty}^{\infty} \\ &= \int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\log_2 c_1 dy_{-\infty}^{\infty} + \frac{1}{\ln 2}\int f\left[y_{-\infty}^{-1},y_0^{\infty}\right]\left(\mathbf{y}_{-\infty}^{\infty}\right)^{\mathrm{T}}\mathcal{B}\mathbf{y}_{-\infty}^{\infty}dy_{-\infty}^{\infty}. \end{aligned}$$

Using the fact that the joint *pdf* is normalized to one and some well-known results for Gaussian integrals, we obtain

$$I\left[Y_{-\infty}^{-1};Y_0^{\infty}\right] = \log_2 c_1 + \frac{1}{\ln 2}\mathrm{Tr}\left[\mathcal{B}\left(\mathcal{V}_{-\infty}^{\infty} - \mathcal{B}_{-\infty}^{\infty}\right)^{-1}\right], \tag{303}$$

where

$$\mathcal{B}_{-\infty}^{\infty} = \left(I\otimes V^{-1}H\right)\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}\left(I\otimes H^{\mathrm{T}}V^{-1}\right).$$

Using the Woodbury matrix identity (Higham 2002, Eq. 148)

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1}+VA^{-1}U\right)^{-1}VA^{-1},$$

we can calculate

$$\begin{aligned} \left(\mathcal{V}_{-\infty}^{\infty} - \mathcal{B}_{-\infty}^{\infty}\right)^{-1} &= \left(\left(I\otimes V^{-1}\right) - \left(I\otimes V^{-1}H\right)\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}\left(I\otimes H^{\mathrm{T}}V^{-1}\right)\right)^{-1} \\ &= \left(I\otimes V^{-1}\right)^{-1} - \left(I\otimes V^{-1}\right)^{-1}\left(I\otimes V^{-1}H\right) \\ &\quad \cdot\left(-\mathcal{C}_{-\infty}^{\infty} + \left(I\otimes H^{\mathrm{T}}V^{-1}\right)\left(I\otimes V^{-1}\right)^{-1}\left(I\otimes V^{-1}H\right)\right)^{-1} \\ &\quad \cdot\left(I\otimes H^{\mathrm{T}}V^{-1}\right)\left(I\otimes V^{-1}\right)^{-1}. \end{aligned}$$

Using the identities $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$ and $(A\otimes B)(C\otimes D) = AC\otimes BD$, we find

$$\begin{aligned} \left(\mathcal{V}_{-\infty}^{\infty} - \mathcal{B}_{-\infty}^{\infty}\right)^{-1} &= I\otimes V + (I\otimes H)\left(\mathcal{C}_{-\infty}^{\infty} - I\otimes H^{T}V^{-1}H\right)^{-1}\left(I\otimes H^{\mathrm{T}}\right) \\ &= I\otimes V + (I\otimes H)\left((\mathcal{C}_2)_{-\infty}^{\infty}\right)^{-1}\left(I\otimes H^{\mathrm{T}}\right). \end{aligned}$$

We note that the above expression equals the inverse of the covariance of the observed states $y_{-\infty}^{\infty}$.

The first part of the constant c_1 can be evaluated directly:

$$c_1 = \sqrt{\frac{\mathrm{Det}\ \Sigma}{\mathrm{Det}\ C}}\cdot\frac{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{-1}}\sqrt{\mathrm{Det}\ \mathcal{C}_0^{\infty}}}{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{\infty}}}, \tag{304}$$

whereas the second part containing the determinants can be solved as follows: first, we observe that we have

$$\sqrt{\text{Det } \mathcal{C}_{-\infty}^{-1}}\,\sqrt{\text{Det } \mathcal{C}_0^{\infty}} = \sqrt{\text{Det } \tilde{\mathcal{C}}_{-\infty}^{\infty}},$$

where the covariance $\tilde{\mathcal{C}}_{-\infty}^{\infty}$ is given by

$$\tilde{\mathcal{C}}_{-\infty}^{\infty} = \begin{pmatrix} & 0 & \cdots & 0 \\ \mathcal{C}_{-\infty}^{-1} & 0 & & \vdots \\ & A & 0 & 0 \\ & & & \\ 0 & & \mathcal{C}_0^{\infty} & \end{pmatrix}$$

$$= \begin{pmatrix} B_1 & A & & & & & & & & \\ A^{\mathrm{T}} & B & \ddots & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \ddots & B & A & & & & & \\ & & & A^{\mathrm{T}} & B_{\Delta t} & A & & & & \\ & & & & & B_1 & A & & & \\ & & & & & A^{\mathrm{T}} & B & \ddots & & \\ & & & & & & \ddots & \ddots & \ddots & \\ & & & & & & & \ddots & B & A \\ & & & & & & & & A^{\mathrm{T}} & B_{\Delta t} \end{pmatrix}.$$

$\tilde{\mathcal{C}}_{-\infty}^{\infty}$ differs from $\mathcal{C}_{-\infty}^{\infty}$ given by

$$\mathcal{C}_{-\infty}^{\infty} = \begin{pmatrix} B_1 & A & & & & & & & & \\ A^{\mathrm{T}} & B & \ddots & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \ddots & B & A & & & & & \\ & & & A^{\mathrm{T}} & B & A & & & & \\ & & & & A^{\mathrm{T}} & B & A & & & \\ & & & & & A^{\mathrm{T}} & B & \ddots & & \\ & & & & & & \ddots & \ddots & \ddots & \\ & & & & & & & \ddots & B & A \\ & & & & & & & & A^{\mathrm{T}} & B_{\Delta t} \end{pmatrix}$$

only in three blocks in the center. If we introduce a vector $e_t = (0\cdots 010\cdots 0)^{\mathrm{T}}$ which has only a one at the position corresponding to time step t, then $e_i e_j^{\mathrm{T}}$ is a matrix that has zeros everywhere except at postion i, j. Accordingly, $\left(e_i e_j^{\mathrm{T}}\right) \otimes A$ is a block-matrix where only the block i, j contains the matrix A. With these prerequisites we can write

$$\tilde{\mathcal{C}}_{-\infty}^{\infty} = \mathcal{C}_{-\infty}^{\infty} + \left(e_0 e_0^{\mathrm{T}}\right) \otimes \begin{pmatrix} (B_{\Delta t} - B) & 0 \\ -A^{\mathrm{T}} & B_1 - B \end{pmatrix},$$

where e_0 is chosen to select the central four blocks of $\tilde{\mathcal{C}}_{-\infty}^{\infty}$. Using the identity for Kronecker products

$$AC \otimes BD = (A \otimes B)(\mathrm{C} \otimes \mathrm{D})$$

we get

$$\tilde{\mathcal{C}}_{-\infty}^{\infty} = \mathcal{C}_{-\infty}^{\infty} + \left(e_0 \otimes \begin{pmatrix} (B_{\Delta t} - B) & 0 \\ -A^{\mathrm{T}} & B_1 - B \end{pmatrix}\right)\left(e_0^{\mathrm{T}} \otimes I_{2q}\right).$$

Using Sylvester's determinant theorem, which states that for matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}, X \in \mathbb{R}^{m\times m}$ it holds that

$$\mathrm{Det}\ (X + AB) = \mathrm{Det}\ (X)\mathrm{Det}\ \left(I_n + BX^{-1}A\right),$$

we obtain

$$\mathrm{Det}\left(\tilde{\mathcal{C}}_{-\infty}^{\infty}\right) = \mathrm{Det}\left(\mathcal{C}_{-\infty}^{\infty}\right)\mathrm{Det}\left(I_{2q} + \left(e_0^{\mathrm{T}} \otimes I_{2q}\right)\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}\left(e_0 \otimes \begin{pmatrix} (B_{\Delta t} - B) & 0 \\ -A^{\mathrm{T}} & B_1 - B \end{pmatrix}\right)\right).$$

In the second term of the right determinant, only the central blocks of $C_{-\infty}^{\infty}$ contribute. They are denoted as

$$\left(\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}\right)_{i=\{-1,0\},\, j=\{-1,0\}} := \begin{pmatrix} X_0 & X_1 \\ X_1^{\mathrm{T}} & X_0 \end{pmatrix}.$$

Note that $C_{-\infty}^{\infty}$ is symmetric and that the blocks along the diagonal are contant in the asymptotic regime. Finally, we get

$$\frac{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{-1}}\sqrt{\mathrm{Det}\ \mathcal{C}_0^{\infty}}}{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{\infty}}} = \sqrt{\mathrm{Det}\left(I_{2q} + \begin{pmatrix} X_0 & X_1 \\ X_1^{\mathrm{T}} & X_0 \end{pmatrix}\begin{pmatrix} B_{\Delta t} - B & 0 \\ -A^{\mathrm{T}} & B_1 - B \end{pmatrix}\right)}. \quad (305)$$

Using the fact that $B_1 - B = \Sigma^{-1} - C^{-1}, B_{\Delta t} - B = -A_0^{\mathrm{T}} C^{-1} A_0$, and $A^{\mathrm{T}} = -C^{-1} A_0$ an expression for the constant c_1 in terms of the system matrices then reads:

$$c_1 = \sqrt{\frac{\text{Det}\,\Sigma}{\text{Det}\,C}}\sqrt{\text{Det}\left(I_{2q} + \begin{pmatrix} (X_1 - X_0 A_0^{\text{T}})C^{-1}A_0 & X_1\left(\Sigma^{-1} - C^{-1}\right) \\ (X_0 - X_1^{\text{T}} A_0^{\text{T}})C^{-1}A_0 & X_0\left(\Sigma^{-1} - C^{-1}\right) \end{pmatrix}\right)}. \quad (306)$$

Next, we turn to the second term in the expression for the EMC, Eq. 303. The matrix $\mathcal{B}$ consists of the difference of $\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}$ for the whole time axis and $\left(\mathcal{C}_{-\infty}^{-1}\right)^{-1}$ $+\left(\mathcal{C}_{0}^{\infty}\right)^{-1}$ for the past and the future of the observed process. Evidently, the matrix for the whole time history $\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}$ coincides with the one for the past $\left(\mathcal{C}_{-\infty}^{\infty}\right)^{-1}$ at least from the infinite past until a certain point of time in the past, where the later matrix elements are still in the asymptotic regime. For later times until $t = -1$, the matrix $\left(\mathcal{C}_{-\infty}^{-1}\right)^{-1}$ is characterized by the transition regime due to the transient phase in the recursions for $\mathbf{N}_j$. This is illustrated in Fig. 4.3 where the transition regimes are shown in yellow. Similarly, the matrix corresponding to the future observations $\left(\mathcal{C}_{0}^{\infty}\right)^{-1}$ deviates significantly from the one for the whole time history only in the beginning and up to some finite point in time in the future (we assumed that the whole process is in steady state and that the covariances of the initial states are equal). Furthermore, the matrix elements decay exponentially in the direction perpendicular to the diagonal. Therefore, the contributions to the sum in Eq. 303 only come from the finite area enclosed by the red lines in Fig. 4.3.

Altogether, in order to numerically compute the EMC given the system matrices, one has to calculate the invers matrices of $\mathcal{C}$ and $\mathcal{C}_2$ for some sufficiently large order Δt, where the asymptotic regime has been reached in the center. The corresponding matrix elements can then be plugged directly into the final result.

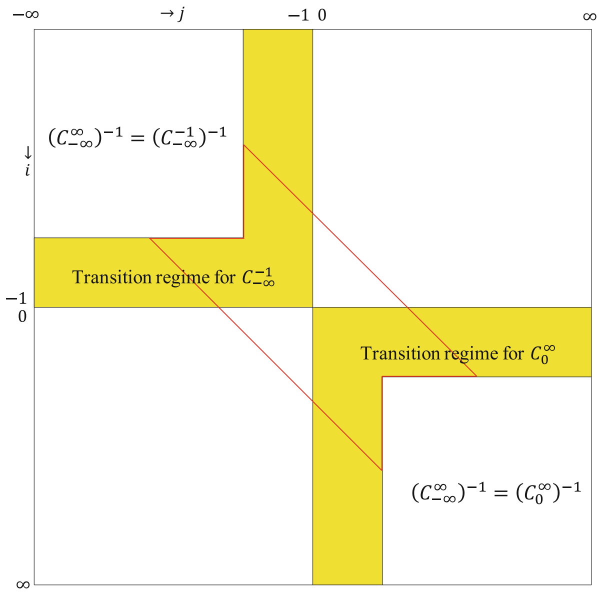

Fig. 4.3 Structure of inverses of $\mathcal{C}_{-\infty}^{\infty}$, $\mathcal{C}_{-\infty}^{-1}$, and $\mathcal{C}_{0}^{\infty}$: only in the area enclosed by the red lines do the inverses differ and contribute to the *EMC*

Using some of the results obtained so far, we can now simplify the general result Eq. 286: We use the results for the normalization of the joint *pdf*, which we obtained by integrating over the hidden states. The ratio of the normalization constants of the joint *pdf*'s was denoted as

$$c_1 = \frac{c_{y_{-\infty}^{\infty}}}{c_{y_{-\infty}^{-1}} c_{y_0^{\infty}}} \cdot \frac{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{-1}}\sqrt{\mathrm{Det}\ \mathcal{C}_0^{\infty}}}{\sqrt{\mathrm{Det}\ \mathcal{C}_{-\infty}^{\infty}}}.$$

Alternatively, the ratio can be expressed directly in terms of the determinants of the corresponding covariances:

$$c_1 = \frac{\sqrt{\mathrm{Det}\ (\mathcal{C}_y)_{-\infty}^{-1}}\sqrt{\mathrm{Det}\ (\mathcal{C}_y)_0^{\infty}}}{\sqrt{\mathrm{Det}\ (\mathcal{C}_y)_{-\infty}^{\infty}}}. \tag{307}$$

Therefore, we can finally write the closed-form solution as

$$\begin{aligned} I\left[Y_{-\infty}^{-1}; Y_0^{\infty}\right] &= \log_2 c_1 \\ &= \log_2 \sqrt{\frac{\mathrm{Det}\ \Sigma}{\mathrm{Det}\ C}} \sqrt{\mathrm{Det}\left(I_{2q} + \begin{bmatrix} \left(X_1 - X_0 A_0^{\mathrm{T}}\right) C^{-1} A_0 & X_1\left(\Sigma^{-1} - C^{-1}\right) \\ \left(X_0 - X_1^{\mathrm{T}} A_0^{\mathrm{T}}\right) C^{-1} A_0 & X_0\left(\Sigma^{-1} - C^{-1}\right) \end{bmatrix}\right)} \\ &= \frac{1}{2}\log_2 \frac{\mathrm{Det}\ \Sigma}{\mathrm{Det}\ C} \mathrm{Det}\left(I_q + X_0\left(\Sigma^{-1} - C^{-1}\right)\right) \\ &\cdot \mathrm{Det}\left(I_q + \left(X_1 - X_0 A_0^{\mathrm{T}}\right) C^{-1} A_0 - X_1\left(X_0 + \left(\Sigma^{-1} - C^{-1}\right)\right)^{-1}\left(X_0 - X_1^{\mathrm{T}} A_0^{\mathrm{T}}\right) C^{-1} A_0\right). \end{aligned} \tag{308}$$

References

Ay, N., Bernigau, H., Der, R., Prokopenko, M.: Information driven self-organization: the dynamic system approach to autonomous robot behavior. Theory Biosci. Special issue on Guided Self-Organisation (GSO-2010), **131**, 161–179 (2012)

Ball, R.C., Diakonova, M., MacKay, R.S.: Quantifying emergence in terms of persistent mutual information. (2010). Advances in Complex Systems **13**(1), 327–338 (2010)

Bialek, W.: Some background on information theory. Unpublished Working Paper. Princeton University, (2003)

Bialek, W., Nemenman, I., Tishby, N.: Predictability, complexity and learning. Neural Comput. **13** (1), 2409–2463 (2001)

Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York, NY (1995)

Boets, J., de Cock, K., de Moor, B.: A mutual information based distance for multivariate Gaussian processes. In: Chiuo, A., et al. (eds.) Modeling, Estimation and Control, LNCIS 364, pp. 15–33. Springer, Berlin (2007)

Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York, NY (1991)

Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum Associates, Mahwah, NJ (1988)

Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley and Sons, New York, NY (1991)

Creutzig, F.: Sufficient encoding of dynamical systems—from the grasshopper auditory system to general principles. Ph.D. thesis, Humboldt-Universität zu Berlin (2008)

Crutchfield, J.P., Ellison, C.J., James, R.G., Mahoney, J.R.: Synchronization and control in intrinsic and designed computation: an information-theoretic analysis of competing models of stochastic computation. Santa Fe Institute Working Paper 2010-08-015, (2010)

Crutchfield, J.P., Ellison, C.J., Riechers, P.M.: Exact complexity: the spectral decomposition of intrinsic computation. Santa Fe Institute Working Paper 2013-09-028, (2013)

de Cock, K.D.: Principal angles in system theory, information theory and signal processing. Ph.D. thesis, Katholieke Universiteit Leuven (2002)

DelSole, T., Tippett, M.K.: Predictability: recent insights from information theory. Rev. Geophys. **45**, RG4002 (2007)

Desai, U.N., Pal, D.: A transformation approach to stochastic model reduction. IEEE Trans. Autom. Control **29**(12), 1097–1100 (1984)

Ellison, C.J., Mahoney, J.R., Crutchfield, J.P.: Prediction, retrodiction, and the amount of information stored in the present. Santa Fe Institute Working Paper 2009-05-017, (2009)

Grassberger, P.: Toward a quantitative theory of self-generated complexity. Int. J. Theor. Phys. **25** (9), 907–938 (1986)

Higham, N.J.: Accuracy and Stability of Numerical Algorithms, 2nd edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2002)

Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge U.K (1985)

Hotelling, H.: The most predictable criterion. J. Educ. Psychol. **26**, 139–142 (1935)

James, R., Ellison, C.J., Crutchfield, J.P.: Anatomy of a bit: information in a time series observation. Santa Fe Institute Working Paper 2011-05-019, (2011)

Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating Mutual Information. (2004). arXiv:cond-mat/0305641v1 [cond-mat.stat-mech]

Krattenthaler, C.: Advanced determinant calculus: A complement. Linear Algebra Appl. **411**(2), 68–166 (2005)

Lancaster, P., Tismenetsky, M.: The Theory of Matrices, 2nd edn. Academic, Orlando, FL (1985)

Li, L.: Some notes on mutual information between past and future. J. Time Ser. Anal. **27**(2), 309–322 (2006)

Li, L., Xie, Z.: Model selection and order determination for time series by information between the past and the future. J. Time Ser. Anal. **17**(1), 65–84 (1996)

Lütkepohl, H.: Comparison of criteria for estimating the order of a vector autoregressive process. J. Time Ser. Anal. **6**, 35–52 (1985)

Lütkepohl, H.: New Introduction to Multiple Time Series Analysis. Springer, Berlin (2005)

Marzen, S., Crutchfield, J.P.: Circumventing the curse of dimensionality in prediction: causal rate-distortion for infinite-order Markov processes. Santa Fe Institute Working Paper 2014-12-047, (2014)

McGinnie, B.P.: A balanced view of system identification. PhD thesis, Cambridge University, Cambridge, UK (1994)

Mori, T., Fukuma, N., Kuwahara, M.: On the discrete Lyapunov matrix equation. IEEE Trans. Autom. Control **27**(2), 463–464 (1982)

Neumaier, A., Schneider, T.: Estimation of parameters and Eigenmodes of multivariate autoregressive models. ACM Trans. Math. Softw. **27**, 27–57 (2001)

Papoulis, A., Pillai, S.U.: Probability, Random Variables and Stochastic Processes. McGraw-Hill, Boston, MA (2002)

Puri, N.N.: Fundamentals of Linear Systems for Physical Scientists and Engineers. CRC Press, Boca Raton, FL (2010)

Schneider, T., Griffies, S.M.: A conceptual framework for predictability studies. J. Clim. **12**, 3133–3155 (1999)

Still, S.: Information bottleneck approach to predictive inference. Entropy **16**(1), 968–989 (2014)

Van Overschee, P., de Moor, B.: Subspace Identification for Linear Systems: Theory, Implementations, Applications. Kluwer Academic Publishers, Boston, MA (1996)